Phonetically Based Phonology

EDITED BY
Bruce Hayes, Robert Kirchner
and Donca Steriade

CAMBRIDGE

CAMBRIDGE www.cambridge.org/9780521825788

This page intentionally left blank

Phonetically Based Phonology

Phonetically Based Phonology is centered around the hypothesis that phonologies are determined by phonetic principles; that is, phonetic patterns involving ease of articulation and perception are expressed linguistically as grammatical constraints. This book brings together a team of leading scholars to provide wide-ranging study of phonetically based phonology. It investigates the role of phonetics in many phonological phenomena – such as assimilation, vowel reduction, vowel harmony, syllable weight, contour tone distribution, metathesis, lenition, sonority sequencing, and the Obligatory Contour Principle (OCP) – exploring in particular the phonetic bases of phonological markedness in these key areas. The analyses also illustrate several analytical strategies whereby phonological sound patterns can be related to their phonological underpinnings. Each chapter includes a tutorial discussion of the phonetics on which the phonological discussion is based.

Diverse and comprehensive in its coverage, *Phonetically Based Phonology* will be welcomed by all linguists interested in the relationship between phonetics and phonological theory.

BRUCE HAYES is a Professor of Linguistics at UCLA. His research in phonology covers many areas, among them the metrical theory of stress, the role of phonetics in phonology, poetic metrics, and learnability. He is the author of *Metrical Stress Theory: Principles and Case Studies* (1995). Professor Hayes's web site is at http://www.linguistics.ucla.edu/people/hayes/

ROBERT KIRCHNER is an Assistant Professor of Linguistics at the University of Alberta. He has published widely in the areas of Optimality Theory and phonetically based phonology. He is the author of *An Effort-Based Approach to Consonant Lenition* (2001). Professor Kirchner's web site is at http://www.ualberta.ca/∼kirchner

DONCA STERIADE is a Professor of Linguistics at MIT. She has published widely in several areas of phonology, including segment licensing, syllable structure, autosegmental theory, the theory of morphological bases, and the theory of phonetically based phonology. Professor Steriade's web site is at http://mit.edu/linguistics/www/steriade.home.html

Phonetically Based Phonology

edited by

Bruce Hayes, Robert Kirchner, and Donca Steriade

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521825788

First published in print format 2004

ISBN-13 978-0-511-21106-5 eBook (EBL)
ISBN-10 0-511-21283-6 eBook (EBL)

ISBN-13 978-0-521-82578-8 hardback
ISBN-10 0-521-82578-4 hardback

Contents

Contributors

JULIETTE BLEVINS *Research Centre for English and Applied Linguistics, Cambridge*

KATHERINE M. CROSSWHITE *Department of Linguistics, University of Rochester*

EDWARD FLEMMING *Department of Linguistics, Stanford University*

STEFAN A. FRISCH *Department of Communication Sciences and Disorders, University of South Florida*

ANDREW GARRETT *Department of Linguistics, University of California at Berkeley*

MATTHEW GORDON *Department of Linguistics, University of California at Santa Barbara*

BRUCE HAYES *Department of Linguistics, University of California, Los Angeles*

JONGHO JUN *Department of English, Yeungnam University*

ABIGAIL KAUN *Annenberg School of Communication, University of Southern California*

ROBERT KIRCHNER *Department of Linguistics, University of Alberta*

DONCA STERIADE *Department of Linguistics, Massachusetts Institute of Technology*

RICHARD WRIGHT *Department of Linguistics, University of Washington*

JIE ZHANG *Department of Linguistics, University of Kansas*

Abbreviations

BLS	*Proceedings of the Annual Meeting, Berkeley Linguistics Society*
CLS	*Papers from the Annual Regional Meeting, Chicago Linguistic Society*
IJAL	*International Journal of American Linguistics*
JASA	*Journal of the Acoustical Society of America*
JL	*Journal of Linguistics*
JPhon	*Journal of Phonetics*
Lg	*Language*
LI	*Linguistic Inquiry*
NELS	*Papers from the Annual Meeting of the North East Linguistic Society*
NLLT	*Natural Language & Linguistic Theory*
WCCFL	*Proceedings of the West Coast Conference on Formal Linguistics*

1 Introduction: the phonetic bases of phonological Markedness

Bruce Hayes and Donca Steriade

If phonological systems were seen as adaptations to universal performance constraints on speaking, listening and learning to speak, what would they be like?
Lindblom (1990: 102)

1 Introduction

Our starting point is a hypothesis central to contemporary phonology: that the markedness laws characterising the typology of sound systems play a role, as grammatical constraints, in the linguistic competence of individual speakers. From this assumption, a basic question follows: How are grammars structured, if markedness laws actively function within them as elements of linguistic competence? We find the answer offered by Optimality Theory (Prince and Smolensky 1993) worth investigating: the grammatical counterparts of markedness laws are ranked and violable constraints and the latter form 'the very substance from which grammars are built: a set of highly general constraints which, through ranking, interact to produce the elaborate particularity of individual languages' (Prince and Smolensky 1993: 217). With qualifications, this view is adopted by many of the contributions to this volume.

The focus of our book is on a different, complementary question: Where do markedness laws come from? Why are sound systems governed by these laws and not by some conceivable others? What is the source of the individual's knowledge of markedness-based constraints? The hypothesis shared by many writers in this volume is that phonological constraints can be rooted in *phonetic knowledge* (Kingston and Diehl 1994), the speakers' partial understanding of the physical conditions under which speech is produced and perceived. The source of markedness constraints as components of grammar is this knowledge. The effect phonetic knowledge has on the typology of the world's sound systems stems from the fact that certain basic conditions governing speech perception and production are necessarily shared by all languages, experienced by all speakers, and implicitly known by all. This shared knowledge leads learners to postulate independently similar constraints. The activity of similar constraints is a source of systematic similarities among grammars and generates a structured phonological typology.

In this introduction, we explain why it is useful to explore the hypothesis that knowledge of markedness derives from phonetic knowledge: how one's view of markedness changes under this hypothesis and what empirical results come from this change of perspective. We outline first how research on phonetically based markedness can be beneficially explored in the framework of Optimality Theory (section 2); and how the OT search for the right constraint set can be speeded up on the view that markedness is phonetically based (sections 3 and 4). We then discuss a specific example of a phonetically based Markedness constraint that illustrates several options in mapping the facts of phonetic difficulty to the elements of grammar (section 5). In the remaining sections, we relate the general discussion of markedness to the specific contents of the book, noting that despite differences of analytical strategy or general theoretical outlook, the diverse phenomena analysed by our contributors can be viewed in a unified fashion.

2 Phonetically based Markedness and Optimality Theory

The idea that phonological Markedness has phonetic roots has particular antecedents in *The sound pattern of English* (Chomsky and Halle 1968), in the theory of Natural Phonology (Stampe 1973), and in the more recent work on Grounded Phonology by Archangeli and Pulleyblank (1994). Optimality Theory makes it worth returning to these issues, since it provides tools with which the questions can be addressed in novel ways. OT takes on a difficulty that held back earlier approaches to naturalness: the *what* is phonetically difficult is not the same as the *how to fix it*. In a rule-based framework, one must provide the theory with multiple fixes, all of which address the same phonetic difficulty. OT separates the problem (embodied in the Markedness constraints) from the solution; the latter is the general procedure at the core of OT, namely creation of a large candidate set by GEN, with the choice from among them determined by the relative ranking of the Markedness constraints with respect to Faithfulness and each other. As a result, OT allows the phonetic principles that drive the system to be expressed directly (Myers 1997): a constraint can embody a particular form of phonetic difficulty, with the issue of how and whether the difficulty is avoided relegated to other parts of the grammar. For a clear case of this sort, see the discussion of postnasal voicing in Pater (1999) and Kager (1999).

The separation of Markedness and Faithfulness also provides a cogent response to an ancient canard: If phonetic optimality is important, why don't sound systems contain nothing but the Jakobsonian optimal [ba]? The answer is that not all the constraints can be satisfied at once. Faithfulness and Markedness constraints conflict; and moreover, there are conflicts between different types of Markedness constraints (notably, those grounded in production vs those grounded in perception). There is no reason to expect the resolution of these conflicts to be uniform across languages. The postnasal voicing example just mentioned is a plausible case of multiple resolutions of the same difficulty.

The more direct argument for OT is that phonetically based constraints discussed here are frequently both active and violated, yielding Emergence of the Unmarked effects (McCarthy and Prince 1994) which require explicit ranking. Kirchner's, Kaun's, and Crosswhite's chapters provide extensive evidence of this type, as does a voicing example discussed below.

3 Markedness

The term *markedness* is ambiguous. It can be used in a strictly typological sense, to identify structures that are infrequently attested or systematically missing, as in *Active use of* [–*ATR*] *is marked* (Archangeli and Pulleyblank 1994: 165 and passim). The term can also refer to an element of a formal linguistic theory, as in OT, where the term *markedness* characterises a constraint type, often disambiguated by capitalisation: Markedness constraints penalise particular structures in surface forms, whereas Faithfulness constraints evaluate dimensions of similarity between specified pairs of lexically related structures, such as the underlying and surface representations.

The definition of markedness in OT is also sometimes related to the hypothesis that Markedness constraints are universal and innate. This claim is logically independent of the central tenets of OT about constraint interaction.[1] Accordingly we are free to assume that a constraint need not be universal or innate to qualify as a Markedness constraint; rather, we use the term in the purely technical sense of a constraint whose violations are evaluated solely on surface forms. We use the term *markedness law* to denote patterns found in typological data, which Markedness constraints are often meant to explain. We may add that the correspondence conditions themselves are formulated with the intention of deriving key aspects of phonological typology.[2]

The terms thus clarified, we turn now to the options available to phonologists who study markedness in either of these two senses.

4 Inductive and deductive approaches to the study of Markedness

Lindblom (1990: 46)[3] observes that the study of distinctive features can proceed in two ways: inductively and deductively. The inductive approach in the study of features is to introduce a new feature whenever the descriptive need arises. The deductive approach, for example Stevens' Quantal Theory (1989) or Lindblom's Dispersion Theory (1986), proceeds not from a question of description ('What are the features used in language?') but from a principled expectation: 'What features should we expect to find given certain assumptions about the conditions [under which] speech sounds are likely to develop?' (Lindblom and Engstrand 1989: 107). The deductive approach can thus hope to provide not only an empirically verifiable feature theory, in the form of principles from which feature sets derive, but may also yield answers to further questions, such as 'Why are the mental representations of speech sounds feature-based (and likewise segment-, syllable-, foot-based)?' These questions simply do not arise under approaches that take for granted the existence of such units and merely aim to discover in the data a basis for their classification.

The distinction between inductive and deductive approaches applies equally to research on markedness. Most attempts to discover markedness principles in phonology have proceeded, until recently, in inductive fashion: phonologists accumulate factual observations about languages and, in due course, a cluster of such observations coheres into a law. The law may be absolute ('There are no initial or final systems in which all obstruent combinations are heterogeneous with regard to voicing'; Greenberg 1978: 252), or implicational ('The presence of syllabic [h̩] implies the presence of syllabic fricatives'; Bell 1978: 183), or only a trend ('If a nasal vowel system is smaller than the corresponding basic vowel system, it is most often a mid vowel that is missing from the nasal vowels'; Crothers 1978: 136). But in most cases the laws originate as generalisations over known languages, not as principles explaining why these laws should be expected to hold. A set of such laws, when they survive peer review, forms a proposed theory of markedness.

The markedness questions asked in earlier typological work seem to have been those for which evidence happened to be available. We cannot exclude the possibility that a priori principles have guided the search for typological generalisations, as reported in the classic work of Trubetzkoy 1938, Jakobson 1941, Hockett 1955, and Greenberg 1978, but these guiding principles were not spelled out and cannot be reconstructed. One may ask, for instance, why the search for clustering universals (Greenberg 1978) proceeds by asking some questions (*Is there an implicational relation between initial* [ln] *and initial* [lt]?) but not others (*Is there an implicational relation between initial* [ht] *and initial* [th]?).

There is an issue of research strategy here. The number of conceivable typological observations is so vast that our results will be haphazard if we examine the data in arbitrary order. Without a general conception of what makes a possible markedness principle, there is no more reason to look into the markedness patterns of, say, initial retroflex apicals (a useful subject, as it turns out; see section 6.1) than into those of prenasal high tones (a topic whose interest remains unproven). The researcher has to take a stab in the dark. In light of this, it seems a sensible research strategy to hypothesise general principles concerning why the constraints are as they are, and let these principles determine a structured search for markedness patterns. We also see below that pursuing the deductive strategy can yield a completely different picture of markedness in several empirical domains.

The work reported in this volume proceeds deductively – as advocated by Lindblom (1990) and Ohala (1983, and much later work) – by asking at the outset variants of the following question: Are there general properties distinguishing marked from unmarked phonological structures, and, if so, what are they? Earlier work in phonetics[4] and phonology[5] suggests that a connection can be found between constraints governing the production and perception of speech and markedness patterns. Certain processes (cluster simplification, place assimilation, lenition, vowel reduction, tonal neutralisation) appear to be triggered by demands of articulatory simplification, while the specific contexts targeted by simplification (e.g. the direction of place assimilation, the segment types it tends to target) are frequently attributable to perceptual factors.

Deductive research on phonological markedness starts from the assumption that markedness laws obtain across languages not because they reflect structural properties of the language faculty, irreducible to non-linguistic factors, but rather because they stem from speakers' shared knowledge of the factors that affect speech communication by impeding articulation, perception, or lexical access. Consider the case discussed below, that of the cross-linguistic dispreference for voiced geminates. The deductive strategy starts from the assumption that this dispreference cannot reflect an innate constraint that specifically and arbitrarily bans [bː dː gː], but must be based on knowledge accessible to individual speakers of the factors that might interfere with the production and perception of voicing. This knowledge and its connection to the grammar have then to be spelled out.

Is the deductive strategy reductionist? Clearly so, but in specific respects. The research presented here bears only on the possibility of systematically deducing the contents of phonological constraints from knowledge of grammar-external factors. This is not the same as deducing the grammar itself: on the contrary, structural properties of the grammar may well filter phonetic knowledge and limit the ways it is mapped onto grammatical statements, as suggested by Gordon (chapter 9) and summarised below (section 5.7). Further, none of the contributions addresses systematically the nature of phonological representations or deduces *their* properties from extra-grammatical factors or discusses whether such reduction is feasible (Gafos 1999). The same goes for the nature of constraint interaction. On the issue of external grounding for all of these components, see Pierrehumbert's overview (2000), and the discussion of representations and constraint interactions by Flemming (2001).

5 Markedness from phonetics: a constraint and its phonetic basis

We now examine a specific example of the deductive strategy. This section introduces a markedness scale and points out its sources in the aerodynamics of speech.

In the phonological analysis of a number of languages, a constraint is needed that penalises voiced obstruent geminates; (1) is a first approximation.

(1)
$$*\begin{bmatrix} \text{+voice} \\ \text{–son} \end{bmatrix}$$

X X

Variants of (1) are active in Ancient Greek (Lupas 1972), Ossetic (Abaev 1964), Nubian (Bell 1971), Lebanese Neo-Syrian (Ohala 1983), Tamil (Rajaram 1972), Yakut (Krueger 1962), Limbu (van Driem 1987), Seleyarese and Buginese (Podesva 2000), and Japanese (Ito and Mester 1995). No language known to us bans just the *voiceless* geminates.[6] The constraint in (1) thus has a typological counterpart, the implicational law in (2):

(2) The presence of a voiced obstruent geminate in a given language implies, in any context, that of the corresponding voiceless geminate.[7]

If a Markedness constraint like (1) reflects, directly or not, an implicational law like (2), then we must consider the possibility that the constraint is universal, in the sense of being potentially active in any grammar. In the next section we explore the hypothesis that some version of (1) is universal in the sense of being inferable from generally available phonetic knowledge.

5.1 From phonetics to grammar

As indicated earlier, we assume that constraints may be universal without being innate (cf. Lindblom 1990; Donegan 1993; Boersma 1998; Hayes 1999). We view Universal Grammar (UG) primarily as a set of abstract analytical predispositions that allow learners to induce grammars from the raw facts of speech, and not as a – dauntingly large – collection of a priori constraints. The project then is to understand how constraints like (1) are induced from evidence about the conditions under which voicing is perceived and produced and what form they take if they are so induced. It is useful here to make the four-way distinction shown below:

(3) a. Facts of phonetic difficulty
b. Speakers' implicit knowledge of the facts in (a)
c. Grammatical constraints induced from the knowledge in (b)
d. Sound patterns reflecting the activity of the constraints in (c)

Facts about phonetic difficulty (3a) and sound patterns (3d) are, in principle, accessible; they are obtainable from experiment, vocal tract modelling, and descriptive phonological work. But the precise contents of (3b) and (3c) have to be guessed at. We see no alternative to drawing these distinctions and making some inferences.

With Prince and Smolensky (1993), we assume that constraint organisation, (3c), reflects transparently the structure of markedness scales, (3b).[8] We also assume that the correspondence between the facts of phonetic difficulty (3a) and the markedness scales (3b) is necessarily indirect: the crucial question is how indirect.

The markedness scales phonologists have mainly relied on so far do not, in their current formulations, explicitly relate to scales of articulatory or perceptual difficulty. Examples are: (a) the nucleus goodness scale in Prince and Smolensky 1993; (b) a place optimality scale like ({*Labial, Dorsal*} $\prec$ *Coronal* $\prec$ *Pharyngeal*), where $\prec$ denotes 'worse than'; Lombardi (in press); and (c) syllabic markedness scales like *CVCC*, *CCVC* $\prec$ *CVC* $\prec$ *CV*. This may reflect the fact that there is no connection between Markedness constraints and phonetic scales or that the exact ways in which phonetic scales map onto phonological markedness has no consequences for the functioning of the phonology. However, the research reported in this book as well as in earlier work indicates that there is often evidence for a much closer connection.

In the next subsections we summarise the articulatory difficulties involved in sustaining vocal cord vibration in different obstruents and consider ways in which speakers can encode knowledge of these difficulties in markedness scales. Our point will be that among several types of mapping (3a) onto (3b)–(3c), a more direct one yields more predictive and more successful models of grammar.

5.2 *Aerodynamics of voicing*

Phonetic studies (Ohala and Riordan 1979; Westbury 1979; Westbury and Keating 1986) have located the rationale for the markedness law in (2) in the aerodynamics of voicing production:

(4) a. Voicing requires airflow across the glottis.
    b. In obstruents, the supraglottal airflow is not freely vented to the outside world.

For these reasons, active oral tract expansion (for example, by tongue root advancement or larynx lowering) is necessary to maintain airflow in an obstruent. These manoeuvres cannot be continued indefinitely or controlled tightly. It is therefore more difficult to sustain production of voicing in long obstruents. The difficulty is directly witnessed in languages like Ossetic, whose speakers attempt to maintain a voicing distinction in long obstruents but nonetheless lose 'part or all of the voiced quality' (Abaev 1964: 9) in [bː dː gː]. No comparable difficulty exists in sustaining voicelessness in [pː tː kː] or voicing in long sonorants, while the problem of maintaining voicing in singleton stops is necessarily one of shorter duration. So far the discussion motivates a simple voicing difficulty scale of the form $D_iː \prec D_i$ where $D_iː$ is a geminate voiced obstruent, and D_i is the corresponding singleton.

Consider now a second factor that influences phonetic difficulty in obstruents, namely place of articulation. As Ohala and Riordan (1979) observe, the size of the cavity behind the oral constriction affects the aerodynamics of voicing. The time interval from the onset of stop closure to the point where passive devoicing will set in varies with the site of the oral constriction: in one experiment, voicing was observed to continue in [b] for 82 ms, but for only 63 and 52 ms respectively in [d] and [g]. This is because the larger cavity behind the lips offers more compliant tissue, which allows the cavity to continue for a longer time to expand passively in response to airflow. A consequence of this is the known asymmetry (Gamkrelidze 1978) between voicing markedness in singleton bilabials as against alveolars and velars: [g] implies [d] which implies [b].[9] This asymmetry holds, according to Ohala (1983), among voiced geminates as well: a geminate [bː]'s duration will certainly exceed 82 ms, and thus some active expansion of the oral tract must be taking place, just as for [dː] and [gː]. But a difference in ease of voicing maintenance persists among the voiced geminates, because there are more options for expansion available in front than in back articulations.

5.3 From aerodynamics to markedness to constraints

There are then at least two sources of articulatory (and indirectly perceptual) difficulty in maintaining voicing: the duration of oral closure and the size of the cavity behind the oral constriction. Phonologically, these are completely different, yet at the level of phonetic difficulty, they are essentially the same thing: in both [g] (a singleton with a small cavity behind the constriction) and [bː] (a geminate with a large cavity) there is difficulty in maintaining voicing past the point where passive devoicing normally sets in. Thus at the phonetic level we can posit a single scale of difficulty that includes both singletons and geminates.

(5) *[+voice]: {gː ≺ dː ≺ bː ≺ g ≺ d ≺ b}

The scales we formulate henceforth distinguish a shared target property – [+voice] in (5) – and the set of contexts in which this property is realised with greater or lesser difficulty: (5) states that the [+voice] feature is hardest to realise in [gː], next hardest in [dː], and so on, and easiest to realise in [b].

The scale in (5) identifies [bː], the best voiced geminate, as harder to voice than short [g], the worst singleton. The difference between a singleton and a geminate consonant is typically much more than the 30 ms that separate the onset of passive devoicing in [b] vs [g] (Lehiste 1970; Smith 1992). Thus the difficulty involved in sustaining voicing should be far more extreme for any geminate obstruent than it would be for any voiced singleton: (5) reflects this point.

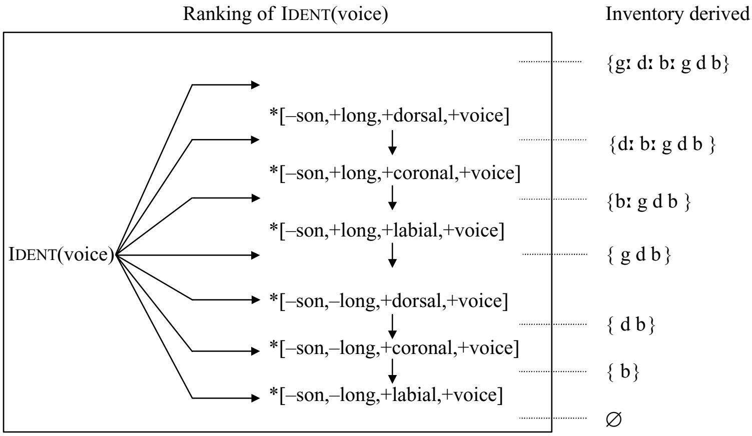
If knowledge about the difficulty of sustaining voicing in obstruents resembles the scale in (5), then its grammatical counterpart cannot be a single constraint; nor can the constraints against voiced geminates remain unrelated to those against voicing in singletons. This is because the voicing difficulty in [gː dː bː] is of the same type – if not of the same magnitude – as that involved in [g d b]. We need a constraint set that reflects the whole scale in (5), not just its upper region. The more general point is that knowledge of markedness, when viewed as phonetic knowledge, generates constraint families and rankings whose structure reflects a broader map of phonetic difficulty, as the learner understands it, rather than isolated points and relations on this map.

As a specific proposal to this end, consider the set of Markedness constraints in (6). These constraints are assumed to be ranked a priori, according to the phonetic difficulty of the segments that they ban (but see fn. 8 above on the issue of fixed rankings).

(6)
a. *[−son, +long, +dorsal, +voice] ‘no voiced long dorsal obstruents’ ≫
b. *[−son, +long, +coronal, +voice] ‘no voiced long coronal obstruents’ ≫
c. *[−son, +long, +labial, +voice] ‘no voiced long labial obstruents’ ≫
d. *[−son, –long, +dorsal, +voice] ‘no voiced short dorsal obstruents’ ≫
e. *[−son, –long, +coronal, +voice] ‘no voiced short coronal obstruents’ ≫
f. *[−son, –long, +labial, +voice] ‘no voiced short labial obstruents’

If the rankings in (6) are fixed, then the relative ranking of this constraint family with respect to the Faithfulness constraint IDENT(voice) determines the inventory of voiced obstruents, as shown in (7):

(7)



An interesting aspect of the constraint set in (6) is that it uses very fine categories, each embodying information about both place and length. Phonologists characteristically judge that constraints are based on rather broader categories. One thus could imagine a more modular characterisation of voicing markedness, as in (8):

(8)	a.	*[−son, +dorsal, +voice]	‘no voiced dorsal obstruents’	≫
		*[−son, +coronal, +voice]	‘no voiced coronal obstruents’	≫
		*[−son, +labial, +voice]	‘no voiced labial obstruents’	
	b.	*[−son, +long, +voice]	‘no long voiced obstruents’	≫
		*[−son, –long, +voice]	‘no short voiced obstruents’	

The constraints in (8) are simpler than those of (6), and involve separate chains of a priori rankings for the dimensions of place and length. As a result, this constraint set is silent on how closure duration and cavity size interact – that is, on the [bː] vs [g] comparison – and thus makes rather different predictions. Notably, we find that in ranking IDENT(voice) amid the chains of (8) (interleaving the chains freely), we cannot derive the inventories for two of the crucial cutoff points in (5): {bː g d b} (forbidding *[dː] and harder) and {dː bː g d b} (forbidding just *[gː]).[10]

5.4 From scales to sound patterns: some language data

The special possibilities implied by (6) (i.e., the constraint set that embodies a unitary scale of voicing difficulty) are confirmed by examples from real languages. The chart in (9) illustrates patterns of selective voicing neutralisation, on a scale like (5), defined by length and place categories: shaded cells in the chart indicate that the voiced obstruent in the column header does not occur. As we compare the three scales introduced earlier with the chart in (9), we observe first that there exist languages that draw a cutoff on all seven possible points of (5):

(9) Place and length constraints on voicing contrasts

	b	d	g	bː	dː	gː
a. Delaware (Maddieson 1984)	■	■	■	■	■	■
b. Dakota (Maddieson 1984)		■	■	■	■	■
c. Khasi (Maddieson 1984)			■	■	■	■
d. Various (citations under (1) above)				■	■	■
e. Kadugli (Abdalla 1973), Sudan Nubian (derived environments; Bell 1971)					■	■
f. Cochin Malayalam (Nair 1979), Udaiyar Tamil (Williams & Jayapaul 1977), Sudan Nubian (root-internal only: Bell 1971)						■
g. Fula (Maddieson 1984)						

The cases of greatest interest here are (9e) and (9f), which show languages that allow all of the voiced singletons but only some of the voiced geminates. These cases are crucial to the comparison at hand (they are allowed by (6) but not (8)), so we discuss them in greater detail.

A dialect of Sudanese Nubian (Nilo-Saharan; Bell 1971), first discussed in this connection by Ohala (1983), disallows [dʒː] and [gː] root-internally but does allow [bː dː]. Derived geminates pattern differently: derived [bː] but not [dː] is preserved as such, with only occasional devoicing of [bː], as seen below in (10).

(10)	Stem	Stem + /goːn/ 'and'	Gloss
	[fag]	[fakːoːn]	'and goat'
	[kadʒ]	[katʃːoːn]	'and donkey'
	[kid]	[kitːoːn]	'and rock'
	[fab]	[fabːoːn], occasional [fapːoːn]	'and father'

As (10) shows, suffixes like /-goːn/ cause gemination of a preceding non-continuant. Gemination entails obligatory devoicing for non-labial stops. There is a difference, then, between the obligatory devoicing of derived [dː] (cf. [kitːoːn] from /kid-goːn/, expected *[kidːoːn]) and the preservation of root-internal [dː] (e.g. [edːi] 'hand'). The devoicing of [dː] in derived environments can be interpreted as an emergence of the unmarked effect (McCarthy and Prince 1994, McCarthy, in press):[11] hence the markedness ranking [dː] ≺ [tː]. The fact that derived [bː] normally surfaces intact suggests a markedness difference relative to derived [dː], which must devoice: this supports the further scale fragment [dː] ≺ [bː]. Moreover, since non-derived [bː] and [dː] are preserved, while [gː] is impossible across the board, a further scale section is established: [gː] ≺ [dː] ≺ [bː]. Finally, singletons are not subject to even optional devoicing, unlike [bː]. We can infer from this that [bː] ≺ [g, d, b]. The Nubian data thus supports a voicing markedness scale that distinguishes at least four intervals: [gː] ≺ [dː] ≺ [bː] ≺ [g d b].

The Nubian pattern of selective voicing neutralisation in geminates is not isolated. A closely related system appears in Kadugli (Niger-Congo; Abdalla 1973): here all voiced singletons are permitted, as well as [bː] and the implosives [ɓː ɗː]. No other voiced geminate obstruents occur. Voiceless geminates are found at all points of articulation, including [pː t̪ː tː kː], but voiced counterparts of the non-labials [dː d̪ː gː] are impossible. Note the *[dː] vs [ɗː] difference: larynx lowering in [ɗː] sustains voicing. Moreover, as seen in (9), some languages exclude just geminate [gː], allowing [bː], [dː], and all singleton voiced C's.

Of related interest to the discussion of voicing markedness is the fact that Nubian lacks [p], a gap related to aerodynamic factors reviewed by Ohala (1983). A short [p] must be actively devoiced, unlike stops at other points of articulation. But [pː] and [p] differ, because the longer duration of [pː] allows it to reach unassisted the point of passive devoicing. In Nubian, this explains why [p] is absent, while [pː] is allowed to arise. We return to this point in section 5.7.

The patterns reviewed in this section and the overall picture in (9) exceed the predictive powers of the most modular statement of voicing difficulty examined, the duo of scales in (8). This is because (8), by hypothesis, limits markedness comparisons to very simple, minimally different pairs of abstract phonological categories: geminates vs singletons and labials vs coronals vs dorsals. This argues that the mapping from voicing difficulty to markedness scales must be more direct and consequently that the scales, and thus the grammars, reflect in greater detail the complexity of phonetic difficulty. The same conclusion is echoed in this volume in the chapters by Kirchner and Zhang.

5.5 *Markedness scales and language-specific phonetics*

In comparing (6) and (8), we found that (6), an approach that sacrifices some degree of formal simplicity in order to reflect more closely the asymmetries of production and perception, achieves better descriptive coverage, notably of asymmetrical systems like Nubian. Yet even (6) is not a purely phonetically based system: it uses standard phonological categories, and refers to only two of the many factors that can influence obstruent voicing. A more thoroughgoing option would be to state that any factor whatsoever that influences difficulty of voicing can be reflected in the constraints and their ranking. This is outlined in the phonetic scale of (11):

(11) [+voice] $\{x \prec y\}$, where x, y is any pair of voiced segments or voiced sequences, such that, without active oral tract expansion, the ratio of voiced closure to total closure duration is less in x than in y.

This is not a fixed list of sounds but a schema for generating phonetic difficulty scales based on knowledge about the phonetic factors that contribute to voicing maintenance. Such a schema would be expected to respond to fine-grained differences in how particular phonological categories are realised phonetically in individual languages.

Suppose, for instance, that in some particular language [d] is a brief flap-like constriction and [b] is a full stop. In such a case, (11) may predict, depending on the specifics of the durational difference, that [+voice] {[b] ≺ [d]}, contrary to (6) and (8). There are in fact languages that allow [d] but not [b] (Maddieson 1984); but the comparative duration of these [d] relative to other voiced stops is not known to us.

There is some evidence that languages indeed deploy phonological constraints based on the conditions set up by language-specific phonetic factors. Zhang's chapter provides an interesting case, which we review here. In Standard Thai, CVR syllables (V = short vowel, R = sonorant consonant) have richer tone-bearing possibilities than CVːO (Vː = long vowel, O = obstruent). In particular, CVːO in Thai cannot host LH or M tones, whereas CVR can host any of the five phonemic tones of the language. The Navajo pattern is close to being the opposite: CVːO can host any phonemic tone (H, L, HL, LH), but CVR cannot host HL or LH.

To explain this type of language-specific difference, Zhang proposes that what licenses contour tones is a combination of length and sonority: vowels make better contour hosts than consonantal sonorants, but, at equal sonority levels, the longer sonorous rhyme is the better carrier. In Zhang's Navajo data, CVR and the Vː portion of CVːO are very close in duration. Thus, the sonority difference of R in CVR versus the second half of the long vowel in CVːO implies that it should be CVːO that is the better tone bearer, and the phonology bears this out: CVːO can host more contours.

In contrast, for Thai, it is CVR that is tonally free and CVːO that is restricted. The source of this reversal vis-à-vis Navajo is evidently a pattern of allophony present only in Thai: long vowels are dramatically shorter in closed syllables. As a result, Thai CVːO has considerably less sonorous rhyme duration than CVR, and the difference is plausibly enough to compensate for CVR's inferior sonority profile. The upshot is that a language-specific difference of allophonic detail – degree of shortening in closed syllables – is apparently the source of a major phonological difference, namely in the tone-bearing ability of different syllable types.

This example is striking evidence for the view that at least some of the markedness scales relevant to phonology must be built on representations that contain language-specific phonetic detail: there is, as Zhang argues at length, a cross-linguistically unified theory of optimal contour carriers, based on a single scale of sonorous rhyme duration. But specific rhymes can be ranked on this scale only when their (non-contrastive, language-specific) durations are specified, not by comparing more schematic representations like CVR to CVːO.

Similar conclusions on the nature of markedness scales follow from Gordon's work on weight (chapter 9), which demonstrates that the typology of optimal stress-bearing syllables is generated by scales of total perceptual energy (integration of acoustic energy over time within the rhyme domain). Gordon shows that language-specific facts about coda selection explain why some languages (e.g. Finnish) count VC and VV rhymes as equally heavy, while others (like Mongolian) rank VV as heavier. Relevant in the present context is that Gordon's results, like Zhang's, do not support universal scales composed of fixed linguistic units (say fixed rhyme types like $V\!ːC_0 > VCC_0 > V$) but rather schemas for generating, on the basis of language-specific information, scales of weight or stressability. The advantage of this approach in Gordon's case is that it reveals the basis on which specific languages choose to count specific rhyme types as heavy or light, a choice long believed to be arbitrary.

5.6 *The stabilisation problem*

If phonetic factors that are allophonic matter to phonological patterning, we must consider the fact that a great deal of allophonic variation is optional and gradient. If such variation bears on phonology, we would expect to see a number of phonological effects that seem to be missing. For example, we are not aware of any sound system in which slowed-down speech, or phrase-level lengthening, causes categorical obstruent devoicing, for either geminates or singletons. Conversely we know of no case in which fast speech allows voicing distinctions to emerge that are absent at normal rates.

These are instances of what we call the *stabilisation problem*: maintaining a (relatively) stable phonology in the face of extensive variation in the phonetic factors that govern the phonological constraints. The stabilisation problem arises in all markedness domains that one might plausibly link to perception and production factors: most types of articulatory and perceptual difficulties are exacerbated by either excessive or insufficient duration, yet variation in speech rate is seldom associated with phonological neutralisation.

The stabilisation problem can be addressed in a number of ways. One possibility, suggested by Steriade (1999), is to suppose that the computation of optimal candidates is carried out relative to a standard speaking rate and style; stabilisation arises when outputs at other rates and styles are bound to the standard outputs by correspondence constraints. Another approach, suggested by Hayes (1999), posits that phonological learning involves testing candidate constraints against aggregated phonetic experience, stored in a kind of map; those phonological constraints are adopted that achieve a relatively good match to aggregated phonetic experience; thus all speaking rates and style contribute together to constraint creation. For further discussion of stabilisation, see Boersma 1998, Kirchner 1998, Flemming 2001, Pierrehumbert 2001, and Zhang's chapter.

We have compared so far the predictions of three different ways of encoding voicing markedness, making the assumption that the set of Markedness constraints reflects directly properties of phonetic difficulty scales. We have seen that simple statements of markedness like (8), which break down continua of phonetic difficulty into multiple unrelated scales, are unable to reflect cross-class markedness relations such as [dː] ≺ [bː] ≺ [g] or [dː] ≺ [ɗː]. For the voicing example considered, the evidence suggests that adherence to a tight-fisted criterion of formal simplicity is therefore untenable. Moreover, we have seen evidence that phonetically based constraints cannot be stated with a priori phonological categories, as in (6), because the phonetic details of how phonological categories are implemented in particular languages turn out to matter to the choice of constraints and their ranking.

5.7 *The tension between formal symmetry and phonetic effectiveness*

Cases like the Nubian voicing phenomena are perhaps eye-opening to many phonologists. Nubian appears to pursue the goal of a good phonetic fit despite the phonological asymmetry that is involved: the set of voiced stops that is allowed in the derived contexts of Nubian is the unnatural class [b d g bː]. Such cases lead one to wonder whether adherence to phonetic factors can give rise to phonological asymmetry on an unlimited basis.

In addressing this question, we should remember that the complexity seen in Nubian only scratches the surface. There are other factors besides gemination and place of articulation that influence voicing, notably whether an obstruent follows another obstruent, or whether it is postnasal or not. Since these factors all impinge on the crucial physical parameter of transglottal airflow, they trade off with one another, just as place and gemination do. Each factor geometrically increases the space of logical possibilities that must be considered in formulating constraints.

Evidence from vocal tract modelling (Hayes 1999), which permits phonetic difficulty to be estimated quantitatively, indicates that pursuing the imperative of good phonetic fit can give rise to hypothetical phonological patterns considerably more complex than Nubian. Consider, for instance, a hypothetical language in which the conditions of (12) hold true

(12) a. [b] is illegal only after obstruents;
  b. [d] is illegal after obstruents and initially; and
  c. [g] is illegal anywhere other than postnasal position

Modelling evidence indicates that this is a system that has a very close fit to the patterns of phonetic difficulty. However, a pattern with this level of complexity has not been documented.

The question of whether there is an upper complexity limit for phonological constraints has also been explored by Gordon (chapter 9), who fitted a large set of logically possible phonological criteria to amplitude and duration measurements made on a variety of languages. Gordon's goal was to assess how well these criteria can classify the syllables of individual languages into groups whose rhymes maximally contrast for total acoustic energy, which appears to be the primary phonetic basis of syllable weight. Gordon finds that the best-distinguished classification often can be achieved by employing a formally very complex phonological distinction – which is never the distinction actually used by the languages in question. Instead, languages evidently adopt whichever of the formally simpler distinctions best matches the patterns of total rhyme energy seen in their syllables. Gordon's conclusion is that formal simplicity places a limiting role on how closely phonetic effectiveness can define phonological constraints.

A puzzle arises here. On the one hand, Gordon found a rather strict limit on the complexity of weight criteria (essentially, two phonological predicates). On the other hand, in the area of segment inventories, languages seem to tolerate complex and asymmetrical systems like Nubian (see scale in (6), which employs minimally four predicates per constraint). Why is the drive for formal simplicity stronger in weight computation? We conjecture that this has to do with the relatively greater difficulty in learning syllable weight categories as compared to segmental categories. Syllables are not actually *heard* as heavy or light; they are *categorised* as such, and this knowledge can only come from an understanding of the prosodic phenomena of the language that depend on weight. Moreover, the primary system reflecting weight, namely stress, is often itself rather complex and difficult to learn. Therefore, any hypothesis about syllable weight is itself dependent for its verification on another complex system, that of stress. Things are different in the case of segmental inventories;

if the grammar under consideration predicts that particular segments should or should not exist, this can be verified fairly directly. Perhaps for this reason, simplicity in computation is not at a premium for inventories and alternations.

Does formal symmetry nevertheless sometimes play a role in determining segment inventories? A possibly relevant case again involves obstruent voicing. We noted in section 5.4 above that the conditions permitting voicelessness in obstruents are essentially the opposite of those for voiced obstruents: [p] is the most difficult obstruent to keep voiceless (particularly in voicing-prone environments, such as intervocalic position); it is followed in order by [t k pː tː kː]. In light of this it is puzzling that Arabic bans geminate [pː], but allows [t k], thus permitting the more difficult sounds and disallowing the easier.

We can interpret this pattern along lines parallel to (8), with [−voice] substituted for [+voice]. There are two families of constraints for [−voice], one based on place, the other on length, with each ranked a priori according to phonetic difficulty. IDENT(voice) is ranked with respect to them as shown in (13); this derives the voiceless inventory [t k tː kː]:[12]

(13)

*[–son, +labial, –voice]		
	← IDENT(voice) →	
*[–son, +coronal, –voice]		*[–son, –long, –voice]
*[–son, +dorsal, –voice]		*[–son, +long, –voice]

Thus, it is possible that languages can vary according to whether the constraints that regulate any particular phenomenon are detailed and closely tailored to phonetics, as in (6), or more general and related to phonetics more abstractly, as in (8) or (13). At present, it appears that both hypotheses like (6) and hypotheses like (8/13) undergenerate, suggesting we cannot account for all the facts unless both are allowed.

6 Markedness scales beyond voicing

The voicing example has outlined some of the issues that arise when we pursue systematically the hypothesis that knowledge of Markedness constraints stems from knowledge of phonetic difficulty. We now connect these issues to the contents of the book, outlining the empirical domains covered by the other chapters and pointing out formal parallels to the voicing case.

6.1 Scales of perceptibility

A central ingredient in the analyses of segmental phonology are the scales of perceptibility. Certain featural distinctions are more reliably perceived in some contexts than in others. Rounding is better perceived in high, back, and long vowels than in non-high, front, and short vowels (Kaun, chapter 4). Place distinctions in consonants are better perceived in fricatives than in stops; in prevocalic or at least in audibly released consonants than in unreleased ones; in preconsonantal position, a consonant's major place features are better perceived if followed by an alveolar than by a velar or labial (Wright, chapter 2; Jun, chapter 3). All vocalic distinctions are better perceived among longer or stressed vowels, than in short stressless ones (Crosswhite, chapter 7).

Relative lack of perceptibility triggers two kinds of changes: the perceptually fragile contrast is either *enhanced* (Stevens and Keyser 1989) – by extending its temporal span or increasing the distance in perceptual space between contrast members – or it is *neutralised*. Kaun's chapter explores enhancement. She argues that rounding harmony is a contrast enhancement strategy: a vowel whose rounding is relatively harder to identify extends it to neighbouring syllables. In this way, what the feature lacks in inherent perceptibility in its original position it gains, through harmony, in exposure time. The key argument for harmony as a strategy of contrast enhancement – and thus for linking the phonology of rounding to the phonetics of perceptibility – comes from observing systems in which only the harder-to-perceive rounded vowels act as triggers. Thus in some languages only the short vowels trigger harmony, in others just the non-high vowels, or just the front vowels, or just the non-high front vowels. More generally, when specific conditions favour certain harmony triggers, these conditions pick out that subset of vowels whose rounding is expected a priori to be less perceptible compared to the rounding of non-triggers. It is these generalisations on triggers that support the idea of harmony as perceptual enhancement.

According to Crosswhite (chapter 7), enhancement and neutralisation of perceptually difficult contrasts are not incompatible strategies. Certain types of vowel reduction display both. Crosswhite notes that the lowering of stressless mid vowels (as in Belarussian) creates a stressless vowel inventory [a i u] whose elements are maximally distinct acoustically. The lowering of [e o] to [a] neutralises the mid-low contrast, but contrast enhancement is also needed to explain why the non-high vowels fail to shift to [ə] (an option exercised by a different reduction type), but rather lower to [a].

Better documented are cases in which the less perceptible features are eliminated altogether. The class of phenomena discussed in Jun's chapter (see also Jun 1995; Myers 1997; Boersma 1998: ch. 11) involve perceptibility scales for consonantal place. Jun argues that place assimilation is just one more consequence of the general conflict between effort avoidance – whose effect is to eliminate or reduce any gesture – and perceptibility sensitive preservation. The latter corresponds, in Jun's analysis, to a set of constraint families whose lower-ranked members identify less perceptible gestures as more likely to disappear. Thus corresponding to the scales in (14), Jun proposes the families of correspondence constraints in (15):

(14) a. perceptibility of C-place: {(strident) fricative ≻ stop ≻ nasal}
b. perceptibility of C-place: {velar ≻ labial ≻ coronal}
c. perceptibility of C-place: {before V ≻ before coronal C ≻ before non-coronal C}

(15) a. $\text{P}{\small\text{RES}}(\text{pl}(\underset{[+\text{cont}]}{\underline{\quad}}\text{C})) \gg \text{P}{\small\text{RES}}(\text{pl}(\underset{[\text{stop}]}{\underline{\quad}}\text{C})) \gg \text{P}{\small\text{RES}}(\text{pl}(\underset{[\text{nasal}]}{\underline{\quad}}\text{C}))$
b. $\text{P}{\small\text{RES}}(\text{pl}(\text{dorsal})) \gg \text{P}{\small\text{RES}}(\text{pl}(\text{labial})) \gg \text{P}{\small\text{RES}}(\text{pl}(\text{coronal}))$
c. $\text{P}{\small\text{RES}}(\text{pl}(\_\text{V})) \gg \text{P}{\small\text{RES}}\left(\text{pl}\left(\_\left[\begin{smallmatrix}\text{C}\\ +\text{coronal}\end{smallmatrix}\right]\right)\right) \gg \text{P}{\small\text{RES}}\left(\text{pl}\left(\_\left[\begin{smallmatrix}\text{C}\\ -\text{coronal}\end{smallmatrix}\right]\right)\right)$

Unlike the voicing scales discussed above, the three scales in (14) represent independent dimensions of perceptibility, hence do not seem to be reducible to a single scale: the scales in (14b, c) reflect the effect of the external context (duration of vocalic transitions; masking effect of following segment) on the perceptibility of C-place, while (14a) ranks the effectiveness of place cues internal to the segment. Correspondingly, Jun observes variation in the typology of place assimilation, suggesting that the manner of the target consonant, the place of the target, and the context of assimilation do not interact and are not mutually predictable. This is what one might expect given the option of intersecting at different points the distinct constraint hierarchies in (15).

The phonological relevance of the perceptibility scales is strengthened by the broader correlation between perceptibility and neutralisation (Steriade 1999). Normally, place distinctions are better identified in pre- than post-V position (Fujimura, Macchi and Streeter 1978; Ohala 1990). However, one-place contrast (that between apico-alveolars like [t] and retroflexes like [ʈ]) concentrates essential place cues in the V-to-C transitions and thus is best perceived if the apicals are postvocalic. Indeed, confusion rates among apicals – but not other C-places – rise steeply in contexts where V-to-C transitions are absent (Ahmed and Agrawal 1969; Anderson 1997). The phonology of place neutralisation is sensitive to this difference in the contextual perceptibility of different place contrasts. In a VC_1C_2V sequence, assimilation for major place features (dorsal, coronal, labial) targets C_1. This follows, as Jun notes, from the fact that, in VC_1C_2V, C_1 occupies a lower rank in the place perceptibility scale relative to the C_2. But this only follows for major place and not for the apical place contrast [ʈ] vs [t]: apicals in C_2 position of VC_1C_2V, should be less perceptible, hence more likely to neutralise, than postvocalic apicals in C_1 position. This is indeed what happens: non-assimilatory neutralisation always targets C_2 apicals in VC_1C_2V strings (Hamilton 1996; Steriade 1995); and moreover place assimilation in apical clusters is predominantly progressive (Steriade 2001): we find mostly /VʈtV/ → [VʈʈV] and /VtʈV/ → [VttV] assimilations.[13] As before, this observation suggests that phonological constraints track the phonetic difficulty map rather faithfully: we do not observe the adoption of any general-purpose context of place licensing, employed for all contrasts, regardless of differences in their contextual perceptibility.

One of the many questions left open by the study of perceptibility on segmental processes relates to the choice between the strategies of place enhancement and place neutralisation. Thus Jun's study of C-place neutralisation, when read in the light of Kaun's results on V-place enhancement, invites the speculation that there exists a parallel typology of C-place enhancement that affects preferentially those Cs whose place specifications are either inherently or contextually weaker. Thus, if every perceptually weak segment is equally likely to be subject to either place enhancement (say via V-epenthesis) or to place neutralisation, then the preferential targets of C-place assimilation identified by Jun should correspond, in other systems, to preferential triggers of epenthesis. We are unaware of cases that fit exactly this description; however, Wright (1996) and Chitoran, Goldstein and Byrd (2002) have documented timing differences among CC clusters tied to differences in perceptibility: the generalisation emerging from these studies is that C_1C_2 clusters containing a less perceptible oral constriction in C_1 typically tolerate less overlap. Further research is needed to determine whether the polar strategies of enhancement and neutralisation are equally attested across all contrast types.

6.2 Scales of effort

One option we did not explore in the earlier discussion of voicing scales like (5) ($\{\text{gː} \prec \text{dː} \prec \text{bː} \prec \text{g} \prec \text{d} \prec \text{b}\}$) was to identify more directly the difficulty posed by voicing maintenance with biomechanical articulatory effort. This is the strategy pursued by Kirchner (chapter 10) in analysing consonant lenition. Kirchner draws several comparisons, some of which are outlined below, and which suggest a global connection between patterns of lenition and effort avoidance.

(16) Effort avoidance and lenition patterns: three comparisons

(a) Vertical displacement of articulators active in C constriction	Greater displacement	Lesser displacement
	stop	approximant
(b) Rate of change	Faster displacement	Slower displacement
	V-stop-V (fast rate)	V-stop-V (slow rate)
(c) Jaw displacement in C constriction relative to neighbouring V	Greater displacement	Lesser displacement
	$\begin{bmatrix}\text{V}\\ +\text{low}\end{bmatrix}$//stop	$\begin{bmatrix}\text{V}\\ +\text{high}\end{bmatrix}$//stop
(d) Number of jaw displacement gestures	Two gestures: C-to-V and V-to-C	One gesture C-to-V or V-to-C
	VCV	(C)CV, VC(C)

Lenition typically turns stops into approximants and, as the comparison in (16a) suggests, this substitutes a less extreme displacement for a more extreme one. Lenition is also more likely at faster rates, a point Kirchner exemplifies with evidence from Tuscan Italian: (16b) suggests that at faster rates the articulators active in Cs have to accelerate in order to cover the same distance to the constriction site in less time. Thus the faster rate makes it more urgent that a less effortful approximant constriction be substituted for the more effortful stop. In Tuscan (and elsewhere: cf. Kirchner 1998) lenition is more likely when one or both flanking vowels are low or at least non-high; less likely if both vowels are high. (16c) suggests that the lower jaw position of low vowels adds to the distance that the articulators must cover in order to generate a stop constriction. Again, the additional effort required here makes it more likely that the active consonantal articulators will fall short of the target, and thus more likely that an approximant will be substituted for the stop. Finally, lenition in one-sided V contexts (pre- or post-V) implies lenition in double sided V_V. This can be tied, as (16d) suggests, to the larger number of jaw displacement gestures required in V_1CV_2 (jaw raising from V_1C and lowering from C to V_2) relative to (C)CV or VC(C).

Rather than recognise as many isolated scales of articulatory difficulty as there are comparisons like (16) – a safe but less interesting move – Kirchner opts for a single scale of biomechanical effort, which underlies all of them. This scale generates a single constraint family – LAZY – whose members penalise articulations in proportion to the degree of effort exertion they entail. This makes it possible to compare disparate gestures, not just oral constrictions, as realised in diverse contexts: the common grounds for the comparison between them being the level of effort expenditure required of each. (Faithfulness constraints cut down on the range of possible articulatory substitutions.) The clear benefit is that, when an independent method for identifying effort costs is found, a precise and elaborate system of predictions will be generated about the circumstances under which one articulation replaces another.

6.3 Scales combining effort and perceptibility

Zhang's study of contour tone licensing (chapter 6) offers an additional possibility: instead of constraint families based exclusively on articulatory or perceptual difficulty, there may be constraints that simultaneously reflect both factors. The formal apparatus Zhang develops relies ultimately on a quantitative measure one could call *steepness*: of two otherwise identical contour tones *x* and *y*, *x* is steeper than *y* if *x*'s duration is shorter than *y*'s, or else if the pitch range covered in *x* is greater than in *y*. Thus, for example, HL on [a] is steeper than HL on [aː], as well as ML on [a].

The steepness comparisons among contour tones are similar to those drawn by Kirchner between sequences more or less likely to undergo lenition: thus the same articulatory trajectory from the low jaw/dorsum position in [a] to the high position needed for [k] is steeper if it has to be completed in less time, for instance at a faster speech rate. Responses of the system to excessive steepness are likewise similar: tonal contour flattening and stop lenition (as well as vowel reduction; see Crosswhite and Flemming's chapters) all reduce steepness.

However, Zhang's point is that, at least for contour tones, steepness is not simply a measure of articulatory difficulty: adequate duration is not only needed for the speaker to complete an articulatory trajectory but also for the listener to identify what contour tone has been articulated. Thus the steepness measure for contour tones should be neutral between articulation and perception. It remains to be seen if Zhang's effort/perceptibility scales are appropriate strictly for contour tones (and diphthongs; Zhang 2001) or whether they extend to facts now analysed by reference to scales that refer to effort or to perceptibility alone.

7 How the picture changes

In a reply to Natural Phonology (Donegan and Stampe 1979) and phonetic determinism (Ohala 1979), Anderson writes: 'the reason [to look for phonetic explanations] is to determine what facts the linguistic system proper is *not* responsible for: to isolate the core of features whose arbitrariness from other points of view makes them a secure basis for assessing properties of the language faculty itself' (1981: 497). Any scholar's interest in the phonetic components of phonological markedness could in principle grow out of an Andersonian belief that we will gain a better understanding of phonology proper once we learn to extract the phonetics out of it. But the project of extracting the phonetics out of phonology can take unexpected turns: in trying to discover those aspects of phonological markedness that are 'arbitrary from other points of view', our views of phonological organisation have changed. Here we outline two changes of this nature that relate to the contents of this volume.

7.1 Segment licensing: syllables vs perceptibility

An important role of syllable structure in contemporary phonology is to deliver compact statements of permissible segment sequences. The hope is that an explicit description of minimal syllabic domains, like onsets and rhymes, should suffice to predict the phonotactic properties of larger domains, like the phonological word. Syllables look like good candidates for Anderson's 'core of features whose arbitrariness from other points of view makes them a secure basis for assessing properties of the language faculty itself', because the choice between different syllable structures seems to be simultaneously central to phonology and unrelated to any extra-grammatical consideration: what phonetic or processing factors could determine the choice between parses like [ab.ra] and [a.bra]? Syllables are also invoked as predicates in the statement of segmental constraints. Thus the fact that final or pre-C consonants are more likely to neutralise place and laryngeal contrasts is attributed (Ito 1986; Goldsmith 1990) to the idea that codas license fewer features than onsets do. Thus contexts like 'in the onset' or 'in the coda' come to play a critical rule in constraints and rules alike. The licensing ability of onsets is of interest to us precisely because it is 'arbitrary from other points of view': nothing about perception, articulation or processing leads us to expect any licensing asymmetry among syllable positions.

As shown in Wright's chapter, the content of the onset-licensing theory can be reconstructed on a phonetic basis. Steriade (1999, 2001) argues that languages tend to license segmental contrasts where they are maximally perceptible. For segments of low sonority this is harder to do, because the perceptibility of a low-sonority segment depends not on its own internal acoustic properties (e.g., all stops sound alike during closure), but on the *external cues* present on neighbouring high-sonority segments, which are created by coarticulation. Thus, there is strong pressure for low-sonority segments to occur adjacent to high-sonority segments. Moreover, not all forms of adjacency are equal: for the psychoacoustic reasons outlined in Wright's chapter, external cues are more salient at CV transitions than at VC transitions.

When incorporated into phonetically based constraints, these principles largely recapitulate the traditional syllable-based typology: branching onsets and codas, which are assumed to be marked, always include consonants that are suboptimally cued: $\underline{C}CV$, $VC\underline{C}$. Moreover, the preference for cues residing in the CV transitions takes over the burden of the traditional arbitrary postulate that onsets are better licensers than codas. Thus, in the following cases, C_1 normally is better cued than C_2: $\#\underline{C}_1V\underline{C}_2\#$, $VC.\underline{C}_1V\underline{C}_2.CV$.

A cue-based theory not only recapitulates the syllabic theory in non-arbitrary form, but outperforms the syllabic theory when we move beyond the broad outlines to the specific details. Thus, for instance, a preconsonantal nasal in onset position (as in many Bantu languages) is very unlikely to have place of articulation distinct from the following consonant; nor are onset obstruents that precede other obstruents (as in Polish [ptak] 'bird') likely to take advantage of their putatively privileged onset position to take on phonologically independent voicing values (as in '[btak]'). Both cases fall out straightforwardly from the cue-based theory. Wright's chapter further notes that sibilant-stop initials should be preferred to other obstruent clusters, on the grounds that sibilants, unlike stops, are recoverable from the frication noise alone. In terms of sonority sequencing, sequences like [spa] are as bad or worse than [tpa], but in terms of perceptual recovery of individual oral constrictions, there is a clear difference that favours [spa]. The typology of word initial clusters (Morelli 1999) clearly supports Wright's approach.[14]

Jun's survey of place neutralisation (chapter 3) also bears on the issue of onset vs coda licensing, by showing that not all codas are equally likely to neutralise: recall from (14) that nasals assimilate more than stops and stops more than fricatives, even when all three C-types are codas. What does distinguish the codas that are more likely targets of assimilation from less likely ones are the scales of perceptibility discussed earlier. Importantly, these factors explicate the entire typology of place neutralisation, with no coverage left for onset licensing. Recall further that C-place neutralisation targets onsets, not codas, whenever the C-place contrast is cued primarily by V-to-C transitions (Steriade 1999 and above): assimilation is strictly progressive in combinations of apicals and retroflexes, because these sounds are more confusable in post-C than post-V position. In this respect too a syllable-based theory of C-neutralisation simply cannot generate the right predictions.

7.2 *Contrast and contrast-based constraints*

Flemming's chapter shows that the deductive approach to markedness leads to a fundamental rethinking of the ways in which constraints operate. The issue is whether constraints evaluate individual structures – sounds or sequences – or systemic properties, such as the co-existence of certain sequences in the same language. Flemming starts from the simple observation that perceptibility conditions cannot be evaluated by considering single sounds or single sequences: when we say, for instance, that [ĩ] and [ẽ] are more confusable than [i] and [e], we mean that [ĩ] and [ẽ] are more confusable with each other, not that they are confusable with unspecified other sounds or with silence. It matters, then, what exactly is the set of mutually confusable sounds that we are talking about.

From this it follows that if there exist phonological constraints that evaluate perceptibility, the candidates considered by those constraints consist of sets of contrasting sequences, not of individual sequences. This implies a quite radical conclusion, that OT grammars must evaluate abstract phonotactic schemata, rather than candidates for particular underlying forms, since no one individual form specifies the other entities with which it is in contrast.

Since the implications of this conclusion are daunting, it is important to determine if Flemming's proposal is empirically warranted. For instance, does it make a difference in terms of sound patterns predicted whether we say that nasalised vowels are avoided *because* they are mutually confusable (a perceptibility constraint that requires evaluating whole nasal vowel sets) or whether we say that nasalised vowels are just marked, with no rationale supplied?

Flemming's fundamental argument is that traditional OT constraints, based on Markedness and Faithfulness, simply misgenerate when applied to areas where the effect of contrast is crucial. For instance, the languages that maintain a backness distinction among high vowels could be analysed with a constraint banning central vowels ('*[ɨ]'), letting only [i] and [u] survive to the surface. The seemingly sensible *[ɨ] constraint becomes a great liability, however, when we consider vertical vowel systems, which maintain no backness contrasts. It is a liability because it predicts the existence of vertical systems in which the only vowel is [i] or [u]; such cases are systematically missing. Evidently, it is the perceptually salient contrast (maximal F2 difference) of [i] and [u], and not any inherent advantage of either of these two vowels alone, that causes them to be selected in those languages that maintain a backness contrast. Thus, it is the entire system of contrasts (at least in this particular domain) that the grammar must select, not the individual sounds. The constraints of conventional OT, which reward or penalise individual segments, cannot do this. Parallel results can be obtained, as Flemming shows, in the study of contrastive and non-contrastive voicing and nasality and (Padgett 2001; Lubowicz 2003) in other phonological domains as well.

8 Other areas

8.1 The role of speech processing

Frisch's chapter makes the important point that what we have been calling 'phonetic difficulty' characterises only the periphery of the human sound processing apparatus; that is, the physical production of sound by the articulators and the initial levels of processing within the auditory system. The deeper levels of the system, such as those that plan the execution of the utterance, or that access the lexicon in production or perception, are just as likely to yield understanding of how phonology works. Frisch covers a number of areas where we might expect to find such effects, focusing in particular on how the widely attested OCP-Place effects might reflect a principle of phonological design that helps avoid 'blending of perceptual traces', and thus avoid misperception.

8.2 The diachronic view of phonetics in phonology

Blevins and Garrett's chapter takes a sharply and intriguingly different approach to the role of phonetics in phonology. Their view[15] is that articulatory ease and perceptual recoverability channel historical sound changes in certain directions, but lack counterparts in the synchronic grammar. Whatever the constraints may be that learners actually internalise, they are believed not to impose articulatory ease or perceptual recoverability on phonological structure.

The core of Blevins and Garrett's approach is the phenomenon of 'innocent misapprehension' (Ohala 1981, 1990). First, phonetic factors determine a pattern of low-level variation. Then, language learners assign to the forms that they are mishearing a novel structural interpretation, differing from that assigned by the previous generation; at this point, phonological change has occurred. To call this process 'innocent misapprehension' emphasises its lack of teleology: phonology is phonetically effective, not because grammars tend to be designed that way, but because innocent misapprehension allows only phonetically effective phonologies to survive.

Various other authors in our volume (Kaun, Frisch) also take the view that diachrony helps explain some aspects of phonological naturalness, and we believe there is clear empirical support for this possibility. But the heart of the controversy, and what makes it interesting to us, lies with Blevins and Garrett's view that the diachronic account suffices entirely, and that we can adopt a theory of phonology (whatever that ends up being) that is entirely blind to phonetically based markedness principles; or perhaps to any markedness principles at all.

Large differences of viewpoint are scientifically useful because they encourage participants on both sides to find justification for their opinions. In this spirit, to further the debate, we offer the following attempted justification of our own position.

First, the study of child phonology shows us many phonological phenomena that could not originate as innocent misapprehensions. Child phonology is characteristically endogenous (Menn 1983): the child inflicts her own spontaneous changes on the adult forms, which in general have been heard accurately (Smith 1973). Child-originated phonological changes often constitute solutions to specific phonetic difficulties, and include phenomena such as cluster simplification, sibilant harmony, and [f]-for-[θ] substitution. Child-originated changes are often adopted by other children and carried over into the adult language (Wells 1982: 96).[16] If children can deploy phonetically natural constraints on their own, it becomes a puzzle that this very useful capacity is not employed in acquiring the adult language.

Our second objection rests on our doubt that innocent misapprehension is capable of driving systematic phonological changes (Steriade 2001). Consider, for instance, the possible roots of regressive place assimilation (/Vŋ+bV/ → [VmbV]) in the misapprehension of the place feature of a preconsonantal nasal. Hura et al. (1992), who have investigated the phenomenon of perceptual assimilation, report that the nasals in stimuli like [VŋbV] are indeed misperceived, but not primarily in an assimilatory fashion. They suggest, then, that simple confusion cannot alone explain the typological fact that nasals frequently assimilate in place to a following obstruent. Confusion alone would predict some form of nonassimilatory neutralisation. Thus, unless there is some factor present in real language-change situations that was absent in Hura et al.'s experiments, 'innocent misapprehension' seems to lack the directional stability that would be needed for it to drive diachronic change.

Lastly, we consider the typology of stop-sibilant metathesis (Hume 1997, Steriade 2001, and Blevins and Garrett's chapter) as supporting the teleological approach to phonology assumed in phonetically based OT. The crucial observation is that stop-sibilant metathesis acts to place the stop – which requires external cues more strongly than the sibilant does – in a position where the best external cues will be available. Usually, this means that the stop is placed in prevocalic (or merely released) position; thus /VksV/ → [VskV] is phonetically optimising. The single known exception (Blevins and Garrett, section 3.4) occurs in a strong-stress language, where it is plausible to assume that post-tonic position provides better cues than pre-atonic position; hence /ˈVskV/ → [ˈVksV]. This cross-linguistic bias in metathesis is unexpected if stop-sibilant metathesis is merely random drift frozen in place by innocent misapprehension, but makes sense if it is implemented 'deliberately' in language, as a markedness-reducing operation.

We believe that most of the evidence that could bear on either side's position remains to be gathered or considered, and thus that further attention to this debate could lead to research progress.

References

Abaev, Vasilii I. (1964). *A grammatical sketch of Ossetic*. The Hague: Mouton.

Abdalla Ibrahim Abdalla (1973). *Kadugli language and language usage*. Institute of African and Asian Studies. University of Khartoum, Sudan.

Ahmed, R. and S. S. Agrawal (1969). Significant features in the perception of (Hindi) consonants. *JASA* 45: 758–63.

Anderson, Stephen R. (1981). Why phonology isn't 'natural'. *LI* 12: 493–539.

Anderson, Victoria B. (1997). The perception of coronals in Western Arrernte. In G. Kokkinakis et al. (eds.), *Proceedings of the 5th European conference on speech communication and technology*, Vol. 1. University of Patras, Greece, 389–92.

(2000). *Giving weight to phonetic principles: The case of place of articulation in Western Arrernte*. PhD dissertation, University of California, Los Angeles.

Archangeli, Diana and Douglas Pulleyblank (1994). *Grounded phonology*. Cambridge, MA: MIT Press.

Barnes, Jonathan (2002). *The phonetics and phonology of positional neutralization*. PhD dissertation, University of California, Berkeley.

Baroni, Marco (2001). How do languages get crazy constraints? Phonetically-based phonology and the evolution of the Galeata Romagnolo vowel system. *UCLA Working Papers in Phonology* 5: 152–78. http://home.sslmit.unibo.it/∼baroni/research.html.

Beckman, Jill N. (1998). *Positional Faithfulness.* PhD dissertation, University of Massachusetts, Amherst.

Beddor, Patrice, Rena Krakow, and Stephanie Lindemann (2001). Patterns of perceptual compensation and their phonological consequences. In Elizabeth Hume and Keith Johnson (eds.), *The role of speech perception in phonology*. San Diego: Academic Press, 55–78.

Bell, Alan (1978). Syllabic consonants. In Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik (eds.), *Universals of human language*. Stanford: Stanford University Press, vol. 2, 153–201.

Bell, Herman (1971). The phonology of Nobiin Nubian. *African Language Review* 9: 115–39.

Boersma, Paul (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Browman, Catherine P. and Louis Goldstein (1990). Tiers in articulatory phonology, with some implications for casual speech. In John Kingston and Mary Beckman (eds.), *Papers in Laboratory Phonology I.* Cambridge: Cambridge University Press, 341–76.

Calabrese, Andrea (1995). A constraint-based theory of phonological markedness and simplification procedures. *LI* 26: 373–463.

Casali, Roderic (1997). Vowel elision in hiatus contexts: Which vowel goes? *Lg* 73: 493–533.

Chitoran, Ioanna, Louis Goldstein and Dani Byrd (2002). Gestural overlap and recoverability: Articulatory evidence from Georgian. In Carlos Gussenhoven and Natasha Warner (eds.), *Laboratory Phonology 7*. Berlin: Mouton de Gruyter, 419–48.

Chomsky, Noam and Morris Halle (1968). *The sound pattern of English*. New York: Harper and Row.

Crothers, John (1978). Typology and universals of vowel systems. In Joseph H. Greenberg et al. (eds.), *Universals of human language*. Stanford: Stanford University Press, vol. 2, 93–152.

De Lacy, Paul (2002). *The formal expression of markedness*. PhD dissertation, University of Massachusetts, Amherst.

Donegan, Patricia (1993). On the phonetic basis of phonological change. In Charles Jones (ed.), *Historical linguistics: Problems and perspectives*. London: Longman, 98–130.

Donegan, Patricia J. and David Stampe (1979). The study of natural phonology. In Daniel A. Dinnsen (ed.), *Current approaches to phonological theory*. Bloomington: Indiana University Press.

Flemming, Edward (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18: 7–46.

(2002). *Auditory representations in phonology*. New York: Routledge.

Fujimura, Osamu, Marian J. Macchi, and L. Streeter (1978). Perception of stop consonants with conflicting transitional cues: A cross-linguistic study. *Language and Speech* 21: 337–46.

Gafos, Adamantios (1999). *The articulatory basis of locality in phonology*. New York: Garland Publishers.

Gamkrelidze, T. V. (1978). On the correlation of stops and fricatives in a phonological system. In Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik (eds.), *Universals of human language*. Stanford: Stanford University Press, vol. 2, 9–46.

Goldsmith, John (1990). *Autosegmental and metrical phonology*. Oxford: Basil Blackwell.

Grammont, Maurice. (1933). *Traité de phonétique.* Paris: Delagrave.

Greenberg, Joseph (1978). Some generalizations concerning initial and final consonant clusters. In Joseph H. Greenberg,Charles A. Ferguson, and Edith A. Moravcsik (eds.) *Universals of human language*. Stanford: Stanford University Press, vol. 2, 243–80.

Guion, Susan G. (1996). *Velar palatalization: Coarticulation, perception, and sound change*. PhD dissertation, University of Texas, Austin.

Halle, Morris and Kenneth Stevens (1971). A note on laryngeal features. In *Quarterly Progress Report 101*. Cambridge, MA: Research Laboratory of Electronics, Massachusetts Institute of Technology, 198–213.

Hamilton, Philip (1996). *Phonetic constraints and markedness in the phonotactics of Australian aboriginal languages*. PhD dissertation, University of Toronto.

Hansson, Gunnar (2001). *Theoretical and typological issues in consonant harmony*. PhD dissertation, University of California, Berkeley.

Hayes, Bruce (1999). Phonetically-driven phonology: The role of Optimality Theory and inductive grounding. In Michael Darnell, Edith Moravcsik, Michael Noonan, Frederick Newmeyer, and Kathleen Wheatly (eds.), *Functionalism and formalism in linguistics, volume I: general papers*. Amsterdam: John Benjamins, 243–85.

Hockett, Charles F. (1955). *A manual of phonology*. Baltimore: Waverly Press.

Hume, Elizabeth (1997). The role of perceptibility in consonant/consonant metathesis. *WCCFL* 17: 293–307.

Hume, Elizabeth and Keith Johnson (2001a), *The role of speech perception in phonology*. San Diego: Academic Press.

(2001b). A model of the interplay of speech perception and phonology. In Elizabeth Hume and Keith Johnson (eds.), *The role of speech perception in phonology*. San Diego: Academic Press, 3–26.

Hura, S., B. Lindblom, and R. Diehl (1992). On the role of perception in shaping phonological assimilation rules. *Language and Speech* 35. 1–2: 59–72.

Hyman, Larry M. (2001). The limits of phonetic determinism in phonology: *NC revisited. In Elizabeth Hume and Keith Johnson (eds.), *The role of speech perception in phonology*. San Diego: Academic Press, 141–85.

Ito, Junko (1986). *Syllable theory in prosodic phonology*. PhD dissertation, University of Massachusetts, Amherst.

Ito, Junko and Armin Mester (1995). Japanese phonology. In John Goldsmith (ed.), *The handbook of phonological theory*. Oxford: Blackwell, 817–38.

Jakobson, Roman (1941). *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Almqvist & Wiksell.

Jun, Jongho (1995). *Perceptual and articulatory factors in place assimilation: An optimality theoretic approach*. PhD dissertation, University of California, Los Angeles.

Kager, René (1999). *Optimality Theory*. Cambridge: Cambridge University Press.

Kavitskaya, Darya (2002). *Compensatory lengthening: Phonetics, phonology, diachrony*. New York: Routledge.

Keating, Patricia A. (1985). Universal phonetics and the organization of grammars. In Victoria Fromkin (ed.), *Phonetic linguistics*. Orlando, FL: Academic Press, 115–32.

Kingston, John and Randy Diehl (1994). Phonetic knowledge. *Lg* 70: 419–53.

Kirchner, Robert (1997). Contrastiveness and Faithfulness. *Phonology* 14: 83–111.

(1998). *An effort-based approach to consonant lenition*, PhD dissertation, University of California, Los Angeles. Rutgers Optimality Archive 276, http://roa.rutgers.edu.

Kochetov, Alexei (2002). *Production, perception, and emergent phonotactic patterns: A case of contrastive palatalization*. London: Routledge.

Krueger, John R. (1962). *Yakut manual*. Bloomington: Indiana University.

Lehiste, Ilse (1970). *Suprasegmentals.* Cambridge, MA: MIT Press.

Lindblom, Björn (1986). Phonetic universals in vowel systems. In John J. Ohala and Jeri Jaeger (eds.), *Experimental phonology*. Orlando: Academic Press, 13–44.

(1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle and A. Marchal (eds.), *Speech production and speech modelling*. Dordrecht: Kluwer, 403–39.

Lindblom, Björn and Olle Engstrand (1989). In what sense is speech quantal? *JPhon* 17: 107–21.

Lombardi, Linda (in press). Coronal epenthesis and markedness. *Phonology* 19: 219–51. Rutgers Optimality Archive 579.

Lupas, L. (1972). *Phonologie du grec attique*. Mouton: The Hague.

Lubowicz, Ania (2003). *Contrast preservation in phonological mappings*. PhD dissertation, University of Massachusetts, Amherst.

Maddieson, Ian (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

McCarthy, John (in press). Comparative markedness. *Theoretical Linguistics*.

McCarthy, John and Alan Prince (1994). The emergence of the unmarked: Optimality in prosodic morphology. *NELS* 24: 333–79.

Menn, Lise (1983). Development of articulatory, phonetic, and phonological capabilities. In Brian Butterworth (ed.), *Language production*. London: Academic Press, vol. 2, 3–50.

Morelli, Frida (1999). *The phonotactics and phonology of obstruent clusters in Optimality Theory*. PhD dissertation, University of Maryland, College Park.

Myers, Scott (1997). Expressing phonetic naturalness in phonology. In Iggy Roca (ed.), *Derivations and constraints in phonology*. Oxford: Clarendon Press, 125–52.

Nair, Somasekharan P. (1979). *Cochin dialect of Malayalam*. Trivandrum: Dravidian Linguistic Association.

Ohala, John J. (1979). Universals of labial velars and de Saussure's chess analogy. In *Proceedings of the Ninth International Congress of Phonetic Sciences*, Copenhagen, vol. 2, 41–7.

(1981). The listener as a source of sound change. In Carrie S. Masek, Roberta A. Hendrick, and Mary Frances Miller (eds.), *Papers from the parasession on language and behavior*. Chicago: Chicago Linguistic Society, 178–203.

(1983). The origin of sound patterns in vocal tract constraints. In Peter F. MacNeilage (ed.), *The production of speech*. New York: Springer Verlag, 189–216.

(1990). The phonetics and phonology of aspects of assimilation. In John Kingston and Mary Beckman (eds.), *Papers in laboratory phonology I: Between the grammar and the physics of speech*. Cambridge: Cambridge University Press, 258–75.

Ohala, John J. and Carol Riordan (1979). Passive vocal tract enlargement during voiced stops. In J. J. Wolf and D. H. Klatt (eds.), *Speech communication papers*. New York: Acoustical Society of America, 89–92.

Padgett, Jaye (2001). Contrast dispersion and Russian palatalization. In Elizabeth Hume and Keith Johnson (eds.), *The role of speech perception in phonology*. San Diego: Academic Press, 187–218.

Passy, Paul (1890). *Etude sur les changements phonétiques et leurs caractères généraux*. Paris: Firmin-Didot.

Pater, Joe (1999). Austronesian nasal substitution and other NC effects. In Harry van der Hulst, René Kager, and Wim Zonneveld (eds.), *The prosody morphology interface*. Cambridge: Cambridge University Press, 310–43.

Pierrehumbert, Janet (2000). The phonetic grounding of phonology. *Les cahiers de l'ICP*, Bulletin de la Communication Parlée, 5: 7–23.

(2001). Why phonological constraints are so coarse-grained. *Language and Cognitive Processes* 16: 691–98.

Podesva, Robert (2000). Constraints on geminates in Buginese and Selayarese. *WCCFL* 19: 343–56.

Prince, Alan (2001). Invariance under re-ranking. Talk given at WCCFL XX, University of Southern California, 24 February 2001. http://ling.rutgers.edu/gamma/talks/WCCFL2001.pdf

Prince, Alan and Paul Smolensky (1993). *Optimality Theory: Constraint interaction in generative grammar*. Rutgers Optimality Archive 537, http://roa.rutgers.edu/.

Rajaram, S. (1972). *Tamil phonetic reader* Mysore: CIIL.

Smith, Caroline (1992). *The temporal organization of vowel and consonant gestures*. PhD dissertation, Yale University.

Smith, Neilson (1973). *The acquisition of phonology: A case study*. Cambridge: Cambridge University Press.

Smolensky, Paul (1993). Harmony, markedness, and phonological activity. Rutgers Optimality Archive 537, http://roa.rutgers.edu/.

Stampe, David (1973). *A dissertation on natural phonology*. PhD dissertation, University of Chicago. Distributed 1979 by Indiana University Linguistics Club, Bloomington.

Steriade, Donca (1995) Positional neutralization. Ms., Department of Linguistics, UCLA. http://www.linguistics.ucla.edu/people/steriade/papers/PositionalNeutralization.pdf.

(1999). Phonetics in phonology: the case of laryngeal neutralization. In Matthew Gordon (ed.), *Papers in Phonology 3*. (UCLA Working Papers in Linguistics 2) Los Angeles: Department of Linguistics, UCLA, 25–146. http://www.linguistics.ucla.edu/people/steriade/papers/phoneticsinphonology.pdf.

(2001). Directional asymmetries in place assimilation: A perceptual account. In Elizabeth Hume and Keith Johnson (eds.), *The role of speech perception in phonology.* New York: Academic Press, 219–50.

Stevens, Kenneth N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In P. B. Denes and E. E. David, Jr (eds.), *Human communication, a unified view*. New York: McGraw-Hill, 51–66.

(1989). On the quantal nature of speech. *JPhan* 17: 3–46.

Stevens, Kenneth N. and S. Jay Keyser (1989). Primary features and their enhancement in consonants. *Lg* 65: 81–106.

Suomi, Kari (1983). Palatal vowel harmony: A perceptually-motivated phenomenon? *Nordic Journal of Linguistics* 6: 1–35.

Trubetzkoy, Nikolai S. (1938). *Grundzüge der Phonologie*. (Travaux du cercle linguistique de Prague, 7) Prague: Cercle linguistique de Prague.

van Driem, George (1987). *A grammar of Limbu*. Berlin: Mouton de Gruyter.

Wells, John (1982). *Accents of English I: An introduction*. Cambridge: Cambridge University Press.

Werner, Roland (1987). *Grammatik des Nobiin*. Hamburg: Helmut Buske Verlag.

Westbury, John (1979). *Aspects of the temporal control of voicing in consonant clusters in English*. (Texas Linguistic Forum, 14) Austin, TX: Dept. of Linguistics, University of Texas, Austin.

Westbury, John and Patricia Keating (1986). On the naturalness of stop consonant voicing. *JL* 22: 145–66.

Williams, T. Edward and V. Y. Jayapaul (1977). *Udaiyar dialect of Tamil*. Annamalainagar: Annamalai University.

Wright, Richard (1996). *Consonant clusters and cue preservation in Tsou*. PhD dissertation, University of California, Los Angeles.

Zhang, Jie (2001). The contrast-specificity of positional prominence – evidence from diphthong distribution. Paper presented at the 75th annual meeting of the Linguistic Society of America, Washington, DC.

Notes

1. Indeed, the view that all the substantive elements of phonological theory are innate is not unique to OT; cf. Calabrese 1995 or Archangeli and Pulleyblank 1994.
2. See in particular work on 'positional faithfulness', such as Jun 1995, Casali 1997, Beckman 1998, Steriade 1995, Steriade 2001.
3. Cf. Lindblom and Engstrand 1989, Lindblom 1990.
4. See Passy 1890, Grammont 1933, Ohala 1983, 1990, Lindblom 1990, Browman and Goldstein 1990, Halle and Stevens 1973, Keating 1985, and Stevens and Keyser 1989.
5. See Chomsky and Halle 1968, Stampe 1973, and Archangeli and Pulleyblank 1994.

6. Maddieson (1984) lists Wolof as such a case; this is evidently an error; cf. forms like *japp* ‘do one’s ablutions’, *wacc* ‘leave behind’ (personal communications from Pamela Munro, Russell Schuh, and Mariam Sy).
7. See discussion of Arabic below for a possible counterexample and ways of analysing it.
8. Formally, the link between markedness scales and Optimality-Theoretic grammar can be achieved in (at least) two ways. Consider a markedness hierarchy $M(S_1) > M(S_2) > \ldots > M(S_n)$, where S_1–S_n are phonological structures and M(S) refers to their relative degrees of markedness. This hierarchy can correspond to a universally fixed ranking in which $^*S_1 \gg {^*S_2} \gg \ldots \gg {^*S_n}$, as in Prince and Smolensky 1993. Alternatively (Prince 2001), the constraints on $S_1 \ldots S_n$ are formulated so that each one bans all elements on the scale at the same markedness level or higher: thus $^*S_2$ penalises S_2 as well as the more marked S_1 structures, whereas $^*S_1$ penalises just S_1. In this system, less marked structures like S_2 are penalised by a proper subset of the constraints that ban more marked ones S_1: no fixed ranking is needed. Empirical arguments favouring the second approach are outlined in Prince 2001 and De Lacy 2002.
9. Maddieson (1984) reports seven languages with a voicing contrast limited to labials; and seventeen where labials and coronals contrast in voicing but velars do not. For discussion, see section 5.5.
10. Moreover, the constraints of (8) derive two inventories that those of (6) cannot derive: {dː bː d b} and {bː b}. We return to the question of such unnatural-but-symmetrical inventories in section 5.7 below.
11. Comparable avoidance of derived-only voiced geminates is documented for Egyptian Nubian (Werner 1987) and Buginese (Podesva 2000).
12. An alternative interpretation of the missing [pː] in Arabic could invoke the fact that a majority of geminates arise through gemination of underlying singletons: if [p] is prohibited and if IDENT(voice) between correspondent segments is undominated, there will be few occasions for the geminate [pː]’s to arise. This predicts that there will be few [p:]’s in this type of system; the fact that there are none does not directly follow.
13. /VtʈV/ → [VʈʈV] and /VʈtV] → [VttV] are limited to cross-word boundary cases, where greater faithfulness plausibly protects C_2; cf. Casali 1997.
14. Initial [mb], [pt], and [sp] are sometimes considered not to consist of a single onset; rather, the initial consonant is said to be under an appendix node, attached directly to the prosodic word, or stray. Such theories must add stipulations for why these structural configurations occur where they do, and why they behave differently in licensing richer ([st]) or more impoverished (*[nb], *[bt]) phonotactic possibilities.
15. Other work along these general lines includes Ohala 1983, 1990, Suomi 1983, Guion 1996, Baroni 2001, Beddor et al. 2001, Hansson 2001, Hume and Johnson 2001b, Hyman 2001, Kavitskaya 2002, Kochetov 2002, and Barnes 2002.
16. Such changes imply the possibility of a theory that is both diachronically based (in agreement with Blevins and Garrett) and phonologically teleological (in disagreement with them).

2 A review of perceptual cues and cue robustness

Richard Wright

1 Introduction

Much recent work in phonological theory has highlighted the role that perception plays in phonological processes. The perceptual basis of contrasts and features has been explored by Flemming (1995/2002), Gordon (1999), and Kirchner (1997) among others. Other work has highlighted the role that perception can play in motivating constraints. In this approach, phonological processes such as positional neutralisation, gemination, and assimilation are motivated by interactions between the strength of perceptual cues and some notion of articulatory ease (Jun 1995; Kirchner 2000; Silverman 1997; Steriade 1995). While some would argue that permitting perceptually motivated constraints to play a role in phonological grammar introduces a prohibitive level of complexity into analyses, the (re)introduction of functionally motivated constraints permits the unification of previously disparate analyses, and in the work cited above it has reduced the number of ad hoc stipulations and exceptions necessary to capture the pattern.

The time seems ripe to revise one of the most widely used, and yet one of the most problematic, constraints in phonological theory: the Sonority Sequencing Constraint. It is almost universally recognised that the Sonority Sequencing Constraint is plagued with exceptions (see Clements 1990 for discussion) and yet most of the efforts to reform it have ended in stipulative patches rather than real improvement. The lack of success in reformulating the Sonority Sequencing Constraint stems at least in part from its greatest flaw: it lacks an explicit, unified phonetic characterisation. Previous efforts at motivating the Sonority Sequencing Constraint on typological grounds have been criticised for their lack of explanatory power and can be seen as circular in their reasoning. For example, Ohala (1992) argues that if the Sonority Sequencing Constraint is typologically motivated, then it cannot be used to explain the typological patterns that inspire it. While the analyses that posit the Sonority Sequencing Constraint as a part of an innate underlying grammar escape circularity, they fail to address the problems associated with its current formulations. Finally, although it has gone more or less unaddressed, many researchers intuitively recognise the similarities between phonotactic constraints (Sonority Sequencing Constraint, No-Coda, Onset, Syllable Contact, etc.) and the phonological phenomena associated with them (contrast neutralisation, metathesis, vowel epenthesis, glide formation, consonant epenthesis, consonant deletion, etc.). However, in their current formulation there is no principled way of motivating phonotactic constraints in a unified way.

This chapter will propose, as a first step, reformulating the Sonority Sequencing Constraint as a perceptually motivated and scalar constraint in which an optimal ordering of segments is one that maximises robustness of encoding of perceptual cues to the segmental makeup of the utterance. Robustness of encoding can be defined along several dimensions: redundancy of cues, the auditory impact of cues, the perceptual distance between cues, and the resistance of cues to environmental masking. As will become clear in the discussion below, a particularly weak encoding can result from strings of consonants, especially stops or nasals that are not flanked by vowels, and a particularly robust encoding results from alternating strings of consonants and vowels. Moreover, in general terms the ordering of consonants by degree of constriction, typical in articulatory descriptions of the Sonority Sequencing Constraint, results in an encoding that is nearly as robust as that produced by consonant-vowel alternations. Finally, certain exceptions that have proven especially problematic for sonority sequencing involving /s/ and other sibilants are no longer exceptions to robust sequencing version of the constraint because they generate an acoustic signal in which the cues to phonological contrasts can be recovered by the listener in the absence of a flanking vowel. The idea of perceptually motivating segmental organisation and phonotactic constraints is not new, but rather this proposal pieces together observations of earlier work on the interaction of perception and the organisation of speech sounds that includes the work of Bladon (1986), Byrd (1996a, 1996b), Flemming (1995/2002), Kawasaki (1982), Lindblom (1983, 1990), Mattingly (1981), Ohala (1992), Silverman (1997), Steriade (1995), and Wright (1996, 2001), among others.

It might be argued that in a constraint-based phonology, violations of a constraint are merely that, and no further research need be done. Problems with this approach are immediately apparent – the exceptions are neither uniform across segment types nor random across languages, nor are they predicted by the Sonority Sequencing Constraint ranking itself. It could also be argued that the Sonority Sequencing Constraint and its exceptions are merely reflections of markedness. In essence, the goal of this chapter is to motivate phonotactic markedness from a perceptual perspective. No doubt articulatory factors, among others, will have to be brought into play to fully motivate phonotactic markedness, but it is worth the effort to see how much can be motivated by perception alone. Developing a perceptually motivated sequencing constraint to replace the Sonority Sequencing Constraint has an added advantage: it allows for the interaction of perceptually driven phonological constraints with other types of constraints such as production-driven constraints, or more traditional phonological constraints.

2 Survey of auditory cues

To understand how the presence, absence, or weakness of a cue will affect the reliable recovery of the segmental sequence, it is first necessary to have a clear idea of where perceptual cues are found in the signal. Therefore, a brief survey of auditory cues follows. Though it is by no means exhaustive, the following survey will give the reader a sense of cue distribution and its relation to segmental organisation. This chapter will follow Wright (2001) in defining ‘cue’ to mean information in the acoustic signal that allows the listener to apprehend the existence of a phonological contrast. This is a very narrow definition, which is used because of the focus of this chapter. For most purposes, it is probably too narrow to be useful because it artificially isolates phonological speech perception from other processes involved in word recognition and spoken-language comprehension, such as apprehending the indexical characteristics of the speaker (Abercrombie 1967), interpreting the sociolinguistic variables at work, and understanding the discourse structure of the utterance. However, for the purposes of understanding the relationship of information in the speech signal and phonotactic markedness, a narrow definition is preferred.

The acoustic signal is produced by articulations that are continuous and overlapping to a greater or lesser degree; therefore, the resulting acoustic cues vary with context. This contextual variation is a factor that contributes to the redundancy in the signal that makes speech perception possible even in difficult listening environments. The variation in the signal due to coarticulation is desirable also because it provides the listener with valuable information about speaking rate or about the relative novelty or predictability of semantic information within the signal (Lieberman 1963), or about the lexical characteristics of a spoken word (Wright 2004). Therefore, the cues that are discussed in this chapter are not intended to be considered invariant, or absolute values, but rather context sensitive and interactive. Although this study will focus exclusively on auditory cues, it should be noted that contributions from other modalities such as vision should be included for a work of this sort to be truly comprehensive. The exact manner in which the cues in the acoustic signal are integrated with each other and with information from other modalities such as vision is a subject that falls outside the domain of this work (for discussion see Bregman 1990, Liberman and Mattingly 1989, Massaro 1987, and Nygaard and Pisoni 1995).

2.1 *Cues to place contrasts in consonants*

There are several potential sources of cues to the place of articulation of a consonant, including second formant transitions, stop release bursts, nasal pole-zero patterns, and fricative noise. The strongest places of articulation cues are found in the brief transitional period between a consonant and an adjacent segment, although there are some that are found internally to the consonant itself (as in fricatives), or distributed over an entire syllable (as with some types of laterals).

2.1.1 Formant transitions In the production of speech the vocal tract can be seen as a time-varying filter with resonances that shapes the spectrum of the sound source (Fant 1960). When a consonant constriction is superimposed on an adjacent vowel, the deformation of the vocal tract results in localised perturbations of the vowel's formant structure as the vocal tract changes shape. Although they are contained within the vowel's portion of the acoustic signal, the formant changes (referred to as *transitions*) provide strong cues to the identity of the adjacent constriction. The second formant (F2) transition and to a lesser degree the third formant (F3) transition provide the listener with cues to the place of articulation of consonants with oral constrictions, particularly stops, affricates, nasals, and fricatives (Delattre, Liberman and Cooper 1955). Transitions are periodic with formant structure. As they are the result of rapid movements, they are transient and dynamic, with the rapidity of the transitions depending on the manner and, to a lesser degree, the place of the consonant. Unlike other consonants, glides and liquids have clear formant structure throughout their durations. Glides are distinguished from each other by the distance between the first and second formant value at the peak of the consonant constriction, and they are distinguished from vowels by rapidity of the formant movement and by the duration of the steady-state portion. Because their cues are dynamic in nature, approximants are highly dependent on the presence of a neighbouring vowel to carry the transitional information. Within the approximants, the lateral [l] is distinguished from the approximant [ɹ] by the unusually low F3 of [ɹ] which is typically below 2000 Hz (Espy-Wilson 1992; Hagiwara 1995; O'Connor, Gerstman, Liberman, Delattre and Cooper 1957).

2.1.2 Fricative noise Fricatives are characterised by a narrow constriction that results in noise either at the place of the constriction or at an obstruction downstream from the constriction (Shadle 1991; Stevens 1998). Frication noise is aperiodic with a relatively long duration. Its spectrum is shaped primarily by the cavity in front of the noise source (Heinz and Stevens 1961). The spectrum of the frication noise is sufficient for listeners to reliably recover the place of articulation in the sibilant fricatives. However, for fricatives with lower amplitude and more diffuse spectra ([θ], [ð], [f], and [v]), the F2 transition has been found to be necessary for listeners to reliably distinguish the place of articulation (Harris 1958). Of these, the voiced fricatives ([ð] and [v]) are the least reliably distinguished (Miller and Nicely 1955). The intensity of frication noise and the degree of front cavity shaping affects the recoverability of place cues for other fricatives as well as making sibilants particularly easy to recover as a group.

As fricatives have continuous noise that is shaped by the cavity in front of the constriction, they can convey information about adjacent consonants in a fashion that is similar, though inferior, to vowels. The overlap results in changes in the spectral shape of a portion of the frication noise, most markedly when the constriction is in front of the noise source. The offset frequency of the fricative spectrum in fricative + stop clusters serves as a cue to place of articulation of a following stop (Bailey and Summerfield 1980; Repp and Mann 1981).

2.1.3 Stop release bursts In stop articulations there is complete occlusion of the vocal tract and a resulting build-up of pressure behind the closure. The sudden movement away from complete stricture results in brief high amplitude noise. Release bursts are aperiodic with a duration of approximately 5–10 ms. They have been shown to play an important role in the perception of place of articulation of stop consonants (e.g. Dorman, Studdert-Kennedy and Raphael 1977; Kewley-Port et al. 1983). Although the release burst or the formant transitions alone are sufficient cues to place, the transitions have been shown to dominate place perception; that is, if the release burst spectrum and the F2 transition provide conflicting place cues, listeners perceive place according to the F2 transition (Walley and Carrell 1983). Listeners show the greatest reliance on the transition in distinguishing the velar place in stops (Kewley-Port, Pisoni and Studdert-Kennedy 1983). While release bursts are reliable cues under artificially quiet circumstances, they are particularly susceptible to even small amounts of masking noise, while fricative noise is much less so (Wright 2001).

Affricates are similar to both stops and fricatives. In their stop portion they have a complete closure, a build-up of pressure and the resultant release burst at release. The release is followed by a brief period of frication. Both the burst and the homorganic frication provide place cues.

2.1.4 Nasal cues Like the oral stops, nasals have an oral constriction that results in formant transitions in the adjacent vowels. In addition, nasals show a marked weakening in the upper formants due to the antiresonance (zero) and a low frequency resonance (pole) below 500 Hz. The nasal pole-zero pattern serves as a place cue (Kurowski and Blumstein 1984). It is most reliable in distinguishing /n/ and /m/, and less so for /ŋ/ (House 1957). Listeners identify the place of articulation more reliably from formant transitions than the nasal portion of the signal; therefore the F2 transition is considered the more powerful

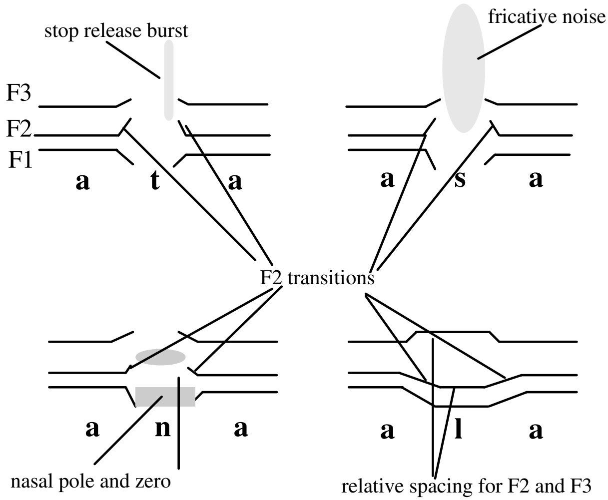


Figure 2.1 Schematic illustration of *place* cues in four VCV sequences (After Wright, Frisch, and Pisoni 1999)

cue (Malécot 1956). Figure 2.1 schematically illustrates the distribution of cues to place contrasts.

2.2 *Cues to manner contrasts in consonants*

All oral constrictions will result in an attenuation of the signal, particularly in the higher frequencies. The relative degree of attenuation is a strong cue to the manner of a consonant. An abrupt attenuation of the signal in all frequencies (excepting the F0 frequency in voiced stops) is a cue to the presence of a stop. Insertion of a period of silence in a signal, either between vowels or between a fricative and a vowel, results in the listener perceiving a stop (Bailey and Summerfield 1980). A complete attenuation of the harmonic signal but with fricative noise provides the listener with cues to the presence of a fricative; the higher the intensity of frication, the more reliably a fricative is heard instead of a stop. A less severe drop in amplitude accompanied by nasal murmur and a nasal pole and zero are cues to nasal manner (Hawkins and Stevens 1985). Nasalisation of the preceding vowel (weakening of the higher formants, broadening of formant bandwidths, and the introduction of a nasal formant) provides look-ahead cues to the nasal manner (Ali, Gallager, Goldstein, and Daniloff 1971; Hawkins and Stevens 1985). Glides and liquids maintain formant structure throughout their peak of stricture. Glides are additionally differentiated from consonants that impose rapid spectral changes during the formant transitions

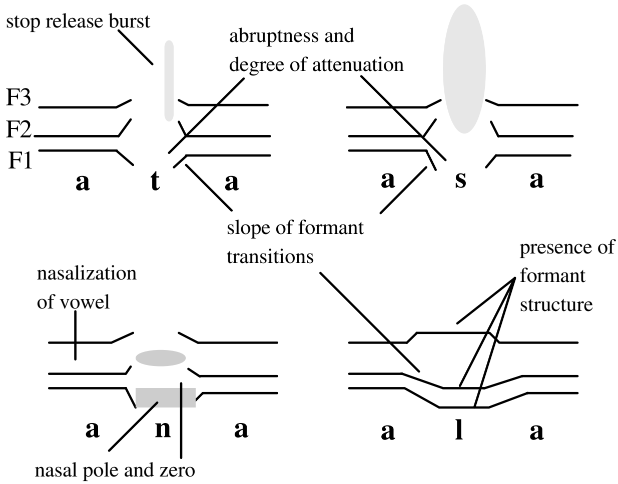


Figure 2.2 Schematic illustration of *manner* cues in four VCV sequences (After Wright, Frisch, and Pisoni 1999)

by the relative gradualness of the transitions into and out of the peak of stricture. Lengthening the duration of synthesised formant transitions changes the listener's percept of manner from stop-transition to glide (Liberman, Delattre, Gerstman, and Cooper 1956). A similar cue is found in the amplitude envelope at the point of transition: stops have the most abrupt and glides have the most gradual rise time (Shinn and Blumstein 1984). Manner cues tend to be more robust in masking noise than place cues, although distinguishing stop from fricative manner is less reliable with the weaker fricatives (Miller and Nicely 1955). Figure 2.2 schematically illustrates the distribution of cues to manner contrasts.

2.3 Cues to voicing contrasts in consonants

Periodicity in the signal is an obvious cue to voicing; however there are several other important cues, such as VOT, the presence and the amplitude of aspiration noise, and duration cues (see figure 2.3). In fricatives, the presence or absence of periodicity during the frication noise is a strong cue to voicing (Cole and Cooper 1975). In English and many other languages voiced obstruents, especially stops, often have no vocal-fold activity. This is particularly true in syllable final position. This means that the listener must rely on other cues to voicing. For syllable initial stops, the primary cue appears to be VOT lag, the time between the release burst and the onset of voicing (Lisker and Abramson 1970). This is true even in languages that maintain voicing during stop closure

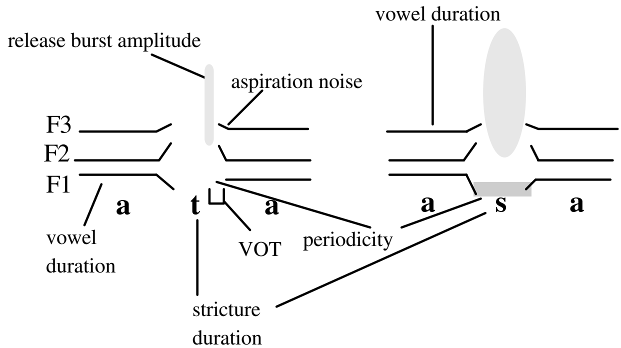


Figure 2.3 Schematic illustration of the voicing cues in two VCV sequences. (After Wright, Frisch, and Pisoni 1999)

(Van Dommelen 1982). Although the relationship between VOT and voicing is, in part, language dependent, generally a short or negative VOT is a cue to voicing, a long VOT is a cue to voicelessness, and a very long VOT is a cue to aspiration (in languages with an aspiration contrast). For English speakers, the presence or absence, or the relative amplitude, of aspiration noise is a contributing cue to voicing (Repp 1979). An additional cue to voicing in syllable onset stops is the relative amplitude of the release burst: a low amplitude burst cues voiced stops while a high amplitude burst cues voiceless stops (Repp 1979).

The duration and spectral properties of the preceding vowel provide a cue to voicing in postvocalic stops and fricatives (Soli 1981). When the vowel is short and has a lower durational proportion of formant steady-state to offset transitions, voicelessness is perceived. The duration of the consonant stricture peak is also a cue to both fricative and stop voicing: longer duration cues voicelessness (Massaro and Cohen 1983).

2.4 *Cues to vowel quality*

Unlike consonants, vowels are made with a relatively open vocal tract and the main cues to vowel contrasts are found in the resonances of the vocal tract. Vowel distinctions are generally thought to be based on the relative spacing of the fundamental frequency (F0) and the first three vocal tract resonances (F1, F2, F3) (Syrdal and Gopal 1986). In very clear (hyperarticulated) speech, vowels have relatively long steady-state portions where the relative spacing of the formants remains fixed and the F0 remains relatively flat. Under these conditions steady-state formant values suffice for vowel perception (Gerstman 1968). However, in naturally spoken language, formants rarely achieve a steady state. Rather vowels that are flanked by consonants have formants that often fall short of values seen in hyperarticulated speech as a result of undershoot (Fant 1960;

Stevens and House 1963). Therefore, listeners must be able to retrieve vowel information from the formant transitions as well. Under these conditions, identification of vowels from formant transitions is more reliable than identification based on steady-state values (Lindblom and Studdert-Kennedy 1967; Strange, Jenkins, and Johnson 1983).

3 Cue robustness as a principle of segmental organisation

In an ideal setting, there is no background noise or distractions, and the listener is so riveted by what the talker is saying that he/she gives the signal undivided attention. Under normal conditions it is rare for speech to occur in the absence of at least some form of environmental masking. What this means for speech is that a robustly encoded phonological contrast is more likely to survive signal degradation or interference in reception. Robustness involves cue redundancy, resistance of cues to environmental masking, the ability of cues to survive momentary distractions on the part of the listener, and the exploitation of the auditory system's tendency to boost certain aspects of the signal. These are not mutually exclusive conditions, but rather are largely overlapping.

3.1 Auditory influences on cue robustness

To understand the role of perception in the shaping of cross-linguistic patterns of segmental organisation, the ways in which the auditory periphery shapes the acoustic signal must be taken into consideration. That is, not only should the distribution of perceptual cues in the signal be considered, but also how the auditory system can change a particular portion of the speech signal. For example, the onset-offset asymmetry that is characteristic of many stages of the auditory pathway can effectively boost the signal-to-noise ratio of certain portions of the signal, while forward masking can obscure other portions of the signal. This means that not all acoustic features that can be discerned in a spectrogram or waveform will necessarily have an equal impact on the listener. It also means that even cues that have been established through perceptual experiments under ideal conditions will have a varying degree of impact on the listener depending on their acoustic properties and on dynamic properties of the signal in which they are found.

Temporal asymmetries are largely a result of the response of the auditory nerve fibres and of certain processing cells in the nuclei of the auditory pathway. Although the auditory nerve fibre response and higher level activity are studied experimentally in mammals other than humans (cat, guinea pig, rat), the results are thought to be representative of humans as well. An auditory nerve fibre's response exhibits a dynamic nonlinear response that depends on the environmental context and the rise-time characteristics of the signal itself. It has been frequently observed that there is a marked burst of activity of the auditory nerve fibres in response to the onset of a stimulus signal (eg. Kiang, Watanabe, Thomas, and Clark 1965; Smith and Zwislocki 1979; Sinex and Geisler 1983). The initial peak in response is followed by a very rapid decay in response during the first 5 ms of the stimulus onset (rapid adaptation), characterised by a return to a much lower level of response. Rapid adaptation is followed by a slower decay during the next 50 ms (short-term adaptation), settling thereafter into a steady pattern. At levels typical of speech, saturation takes place approximately at the end of the short-term adaptation, after which a change in stimulus intensity will not result in an equivalent change in the fibre's firing rate. The point at which saturation occurs in a particular fibre depends on the spontaneous firing rate of the fibre, and the frequency and intensity of the stimulus. In the absence of saturation, response rate is equated with signal intensity; thus, the transient boost in the firing rate at signal onset is seen as effectively amplifying a brief period of the stimulus (Delgutte and Kiang 1984a). At its peak, the onset boost results in an increase in activity to a level many times higher than the level after short-term adaptation has taken place.

The magnitude of the peak response at onset depends largely on the levels of activity in the frequency regions of the stimulus signal immediately preceding the stimulus onset: the less the activity and the longer the period of inactivity, up to approximately 50 ms, preceding the onset, the greater the initial response (Delgutte and Kiang 1984a). A secondary factor involved in the magnitude of response is the rise time at the onset of the signal: the shorter the rise time, the greater the response (Delgutte and Kiang 1984b).

The onset asymmetry in the auditory nerve response is mirrored by similar nonlinearities as the signal ascends the auditory pathway. For example, in the cochlear nucleus there are cells, principally the onset units, that respond with a burst of activity to the onset portion of a signal, particularly to frequencies above 1 kHz (Greenberg and Rhode 1987). A similar effect has been seen for frequency modulation: the more rapid the frequency change in the characteristic frequency of a unit, the greater the response (Møller 1977). This emphasis on the portions of the signal with sharp discontinuities continues up to the auditory cortex (Clarey, Barone, and Imig 1992). Taking into account the auditory transformations will not change what are commonly considered to be cues, particularly those that have been tested in perceptual experiments, but it will help to explain why a cue's impact is greater in some positions than in others, and why some cues are more robust than others in a noisy environment.

3.1.1 The boost at onset The auditory nerve fibre encoding of consonants has been studied for both CV onset consonants (Sinex and Geisler 1983; Miller and Sachs 1983; Delgutte and Kiang 1984a, b) and for VC word-final consonants (Sinex 1995). Overall, the results from such experiments indicate that the onset

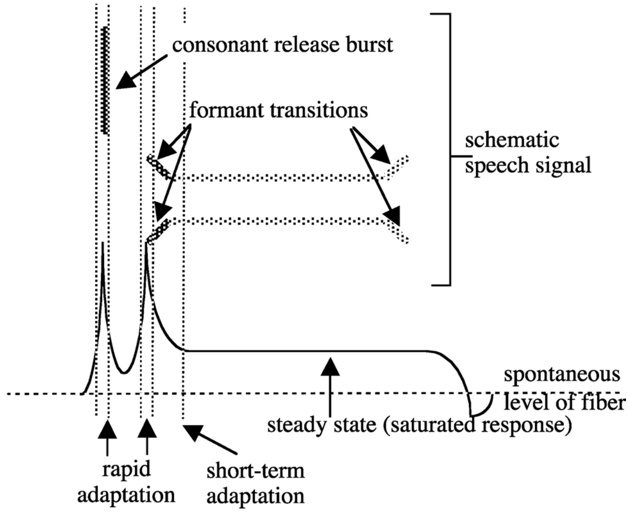


Figure 2.4 Schematic illustration of the onset-offset response asymmetry in the sequence [dat˥] showing rapid and short-term adaptation. The portions of the signal that benefit from the boost in auditory nerve response are darkened, while portions that are rendered less salient appear more faintly (after Wright 2001)

peak is present for complex speech signals as it is for simple signals, and that the formant transitions out of a consonant closure receive a boost as does the release burst or fricative noise. This is particularly true of the stops, fricatives, and affricates, but less so for the nasal consonants, which have high amplitude responses in the fibres with lower characteristic frequencies of long enough duration for saturation to occur before the nasal consonant ends. There is no equivalent boost of activity at the speech signal offset (postvocalic closure), thus the formant transitions into a closure are not amplified in the way that onset transitions are. In the case of postvocalic stops, however, the reduced amplitude (voiced) or silence (voiceless) that results from the stop closure provides recovery time for the auditory nerve. Therefore, if the postvocalic stop is released, there will be a boost in response during the release. Sinex (1995) found that both voiced and voiceless postvocalic stops showed a peak in activity in response to the stop's release burst. This indicates that the closure time is sufficient for recovery in the auditory nerve fibres. Figure 2.4 schematically illustrates the onset-offset asymmetry in response to a speech signal.

There is an interaction of the onset boost and the consonant manner (Delgutte and Kiang 1984a, b). The greatest response increase is seen following the period of silence resulting from a voiceless stop closure or at an onset following a pause. A less marked increase in activity is seen after a low amplitude period in the signal such as following a voiced stop closure, and even less increase is seen following the higher amplitude voicing during a nasal. In the case of the fricatives, the intensity and frequency of the frication noise will determine the degree to which the fricative will drive down the boost in activity at the onset of voicing of the vowel. The greater the intensity in the frequencies up to 2000 Hz, the more decremental the effect on the relative boost at the onset of voicing. A stop consonant release burst is brief enough that, even at relatively high intensity levels, it will not drive down the response to the onset of voicing. There is also a differential internal response to the fricatives, affricates, and nasals that depends on the rise time of the signal. Affricates, with the sharpest rise time, result in the most marked peak at the onset of frication (virtually the same as that seen for consonant release bursts), while fricatives with the slowest rise time showed very little boost at the onset of frication (Delgutte and Kiang 1984b).

3.1.2 Robustness in noise Two gross factors that determine the relative response of the auditory system are the frequency and periodicity of a signal. First, the auditory system is most sensitive to lower frequencies both in terms of overall sensitivity (perceived loudness) and in terms of response to changes (critical bands). Second, a signal that is periodic and below 5000 Hz can be more reliably encoded in the presence of masking noise because of phase locking in the auditory nerve fibres and cells of the processing nuclei.

It is well known that aperiodic signals are more easily masked by other periodic or aperiodic noise than periodic signals are, and that this has a direct impact on the type of cues that are most easily masked (e.g. Miller and Nicely 1955). Cues that are found in aperiodic noise, such as the place cues in the noise of fricatives, affricates, and consonant release bursts, are highly vulnerable. The intensity of the noise and the sharpness of the onset will determine the degree to which the information is lost due to masking. The sibilant fricatives are the most resistant to masking, despite their high-frequency concentration of energy, because they are more intense than non-sibilant fricatives. Onset release bursts are more resistant to masking than the non-sibilant fricatives: the release is preceded by relatively little activity in the signal, and they have a sharp rise time, both of which lead to an increase in activity in the auditory response. However, their transience makes them more vulnerable to masking, an effect that offsets some of the benefits they derive from the heightened auditory response. This means that if a sibilant fricative is heard in isolation the listener will be more readily able to identify it than any of the other fricatives, and because of its duration it may survive masking better than a stop release. The stops and non-sibilant fricatives will be much more dependent on the presence of a formant transition.

Place cues in the formant transitions are more resistant to loss through masking because they are periodic. However, even formant place cues are more susceptible to masking than some of the broad cues to manner, such as signal attenuation, because they are transient. This vulnerability is offset in the syllable-onset position by the onset boost in the auditory system. Syllable-offset transitions (into a consonant closure) on the other hand are probably degraded slightly by forward masking in the lower frequencies. Figures 2.1–2.4 above schematically illustrate the onset boost that makes onset transitions less vulnerable to environmental masking than offset transitions. The relative advantage that the syllable-onset transitions have relies heavily on the manner of the preceding consonant with the voiceless stops having the greatest benefit and the fricatives having the least. This is not to say that coda transitions necessarily contain less information, but rather that in any but the best listening environments (i.e. listening in quiet with a pair of headphones) they will stand a poorer chance of reliably transmitting information about a particular contrast.

Place and manner cues in the nasal pole and zero are resistant to masking; however, they tend to be weaker place cues than the actual formant transitions (Malécot 1956). Therefore, while nasal manner is highly recoverable from the nasal spectrum, nasal place is not as reliably recovered.

3.2 Overlap, redundancy, and parallel transmission

Mattingly (1981) proposed that a driving force in the organisation of speech is the need for both a maximal speed of signal transmission and a robust encoding of the information in the signal. This is achieved through gestural overlap, in production that increases the transmission rate and robustly encodes information about the articulations. The degree to which overlap results in greater robustness and increased speed of transmission depends on the makeup of the segmental string. The greatest benefit is derived from an organisation of speech in which constrictions are released in decreasing order of stricture and constrictions are applied in increasing order of stricture. To derive the greatest perceptual benefit from overlap, there will be intervening vowels that represent the peak of aperture.

Mattingly's proposal is based on the well-known tendency for articulations to overlap in the production of speech. The articulatory (or gestural) overlap results in an acoustic signal in which information about one articulation may show extensive temporal overlap with information about another. The parallel nature of the transmission results in a compression of the information and a decreased transmission time. In addition to increasing the speed of transmission, articulatory overlap potentially leads to redundancy of information in the signal. The word 'potentially' is used here because whether or not overlap results in an increase or a decrease in the number of cues depends on the characteristics of adjacent segments whose articulations are overlapping. For example, when two stop consonants overlap, there is very little increase in information and a loss of the release burst of C1 if the degree of overlap is great enough. On the other hand, formant transitions out of a consonant closure and into a vowel provide information about the preceding stop's place and manner of articulation as well as providing information about the vowel. The same is true of transitions out of the vowel and into a following consonant closure. Thus, the articulatory overlap between consonants and vowels both speeds up the transmission of the signal in terms of production and encodes the information in the signal redundantly.

Under optimal laboratory listening circumstances any one of an array of cues may be sufficient for a listener to recover a phonological contrast. However, while any one cue may in principle suffice, not all cues will be equally effective in conveying their information to the listener in all environments. The inequality of cues comes from the fact that spoken language must rely on the transmission through an acoustic medium and on the reception of the signal by the listener. Under these conditions opportunities abound for the introduction of noise into the process. Redundancy of encoding results in a signal that is perceptually robust in noisy environments that are typical of spoken language.

Even though increased overlap may result in perceptual benefits given the proper segmental sequence, too great a degree of overlap can result in information loss and a degradation of the signal. Mattingly points out that complete overlap may result in a loss of cues: if a gesture is co-extensive with another gesture with a higher degree of stricture, the gesture with a lesser degree of stricture may have no impact on the acoustic signal because it will be fully obscured. He uses the example of the sequence /pla/ in which extensive overlap between the /p/ and the /l/ leads to the desired increase in rate of the transmission, but in which complete overlap between the /p/ closure and the /l/ constriction could lead to a loss of information about /l/.

3.3 Modulation and increased salience

Kawasaki (1982) and Ohala (1992) have proposed a different perceptual factor in segment sequencing that makes somewhat overlapping predictions than Mattingly's. Their proposal is based on the assumption that change (modulation) along an acoustic dimension, such as frequency or amplitude, will result in increased salience of the cues in the portion of the signal where the change occurs. Therefore, the greater the modulation and the more dimensions that are involved, the greater the salience. The more modulation there is in the acoustic signal, the better the segmental organisation. It turns out that modulation alone is a rather poor predictor of many common phonotactic patterns (Kawasaki 1982) and that additional factors, such as cue enhancement and confusability, need to be taken into consideration (see Flemming 1995/2002 for a discussion).

Nevertheless, in a broad sense, modulation does appear to play a role in the relative auditory impact of a cue.

The increased salience may be rooted in both attentional factors and in the functioning of the auditory periphery. Changes in frequency and amplitude result in a dramatic increase in activity in the auditory nerve fibres and in certain cells in the nuclei of the auditory pathway. The amount of increase that a change provokes will vary depending on the context and, in certain cases, the direction of change. Overall, however, the link between modulation and salience appears to be borne out at least to a degree.

The benefits of modulation can be seen in the alternating CV sequencing that is typical of most languages. The alternating closures and apertures create a large modulation along the amplitude dimension, and smaller abrupt changes in frequency in the formant transitions into and out of the closures. As most abrupt changes occur either at a consonant closure or aperture, these cue-rich portions of the signal are guaranteed an increase in salience. As the articulators move from the consonant constriction towards the vowel target, the greater the distance the formants travel, the greater the increase in salience. According to this model, the transition from a labial constriction with a relatively low second formant (F2) into a front vowel with a relatively high F2 will be more salient than the transition into a back vowel with a relatively low F2. The relative lack of modulation may underlie the co-occurrence restriction on labialised consonants or labial glides plus rounded vowels, as well as the similar restriction on palatalised consonants or palatal glides and high front vowels. These sequences of segments provide little modulation along the frequency or amplitude dimensions. Stevens (1989), Stevens and Keyser (1989), and Flemming (1995/2002) explore in detail the ways in which segmental makeup can enhance or diminish phonological contrasts.

To a large degree, the sequences that result in a greater degree of signal modulation are the same ones that result in increased cue redundancy. Thus, the two can be seen as overlapping factors that increase the robustness of encoding in the signal. For example, the alternation of consonants and vowels (CV) results in a robust encoding of information both because it results in the greatest perceptual benefit from overlap, and because it creates an optimal signal modulation. In addition, sequences with a similar degree of aperture (stop + stop, fricative + fricative, nasal + nasal, etc.) result in a poor encoding both because they result in very little perceptual benefit from overlap, and also because they result in very little signal modulation. However, the two principles should not be seen as predicting identical results. For example, because it refers mainly to degree of stricture, the overlap principle predicts that all sequences of glide + vowel should be equally robustly encoded in the signal; thus a sequence like /ja/ is as good (or bad) as a sequence like /ji/. On the other hand, the modulation principle alone does not take into account the benefit of cue redundancy in determining the relative perceptual benefit of a particular sequence. In the sequence /pla/ information is robustly encoded because /p/, with complete closure, precedes the /l/ with partial closure, which, in turn, precedes the /a/ with an even greater degree of aperture. This ordering ensures that information about the liquid is not lost as a result of overlap (as might be the case if it were to precede the stop), since a portion of it overlaps with the following vowel, and it creates a signal in which information about any one segment is distributed redundantly throughout the signal. The p-release will be shaped by the overlapping /l/ closure but not completely obscured by it since the lateral does not make a complete closure, thus it will bear information both about the place of articulation and manner of the /p/ as well as the /l/. The lateral may be partially devoiced and its acoustic onset will be shaped by the preceding labial closure, thus it will bear information about both the /p/ and the /l/. The onset of the vowel will have transitions that will provide place and manner information about the preceding lateral as well as information about the vowel's quality. If the lateral were to precede the labial stop, increased overlap might result in loss of information about its manner and place (depending of course on the segmental makeup of the surrounding utterance).

4 Preferred segmental sequences

The alternating consonant and vowel pattern (CVCV), which is by far the most common pattern in the world's languages, is also the best pattern in terms of the sheer number and redundancy of cues in the signal. At each transition from vowel to consonant there are numerous cues to both the vowel's quality and to the consonant's place, manner, and voicing. This results in a high degree of redundancy in the signal so that the signal can be considerably degraded before there is a critical loss of information about any of the segments. The pattern of consonant + glide + vowel + glide + consonant (and variations thereof) is only a slightly worse sequencing pattern than the CV pattern above. How good or bad these clusters are will depend on how similar the glide transitions are to the vowel formants and the formant transitions of the consonant. It should be remembered that glides are dynamic so that the usefulness of their internal cues depends on the listener identifying the formant movement as resulting from a glide and not from the transition from consonant to vowel. The greater the similarity between the consonant formant transitions and the glide formant transitions, the more easily the glide will be lost. The more similar the glide's formants are to the vowels, the less movement there will be and the poorer the identification of the glide will be. Word-final and preconsonantal consonants (coda consonants) are in a slightly poorer situation vis-à-vis the cues for auditory perception.

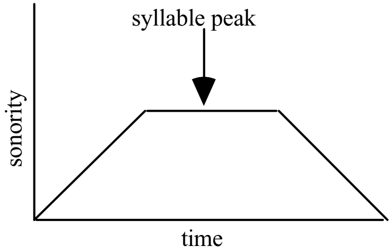


Figure 2.5 Schematic illustration of sonority sequencing

Of course, how poor the situation is from a perceptual point of view depends a great deal on the acoustic characteristics of the consonant in question. In general it can be said that consonants with internal cues to place, manner, and voicing will suffer less in the coda. For example, nasals will survive better in a coda from the point of view of voicing (since nasals are generally voiced) and manner. However, from the point of view of place, they are in only slightly better a position than stops, because their internal cues to place are fairly weak. Thus we expect a loss of place contrasts in coda nasals and preconsonantal nasals. Fricatives should also be better in preconsonantal and word-final positions than stops. They have strong cues to place in their frication noise (particularly the sibilants) and they also have good internal cues to voicing. In word-initial position, fricative + stop and nasal + stop sequences are predicted to be much more common than stop + stop and even stop + fricative sequences because they are not as reliant on vowel transitions for perceptual cues. The nasals will be even more susceptible to loss of place than they are word internally, and the fricatives will be limited to the sibilants that have strong internal cues.

The sonority sequencing principle has proven a particularly useful tool in phonology for describing many of the patterns that result from perceptual factors such as these as well as articulatory factors that are outside the domain of this discussion. One common form of the sonority sequencing principle states that within a syllable, segments are ordered first in order of increasing sonority from a sonority minimum to a sonority peak, and in order of decreasing sonority thereafter (e.g. Bell and Hooper 1978; Selkirk 1984). Most works on sonority cite Sievers (1881), Jespersen (1913), and de Saussure (1914) as the earliest uses of sonority-based rankings to explain the ordering of segments within a syllable. Ohala (1992) also cites earlier work including de Brosches (1765) and Whitney (1874). Figure 2.5 schematically illustrates the increase in sonority over the course of the syllable.

In this view, the sonority value of a segment is considered an inherent property of a segment. Traditionally a segment's inherent sonority was based on stricture. The greater the degree of stricture the lower the sonority value. Other

highest sonority	lowest sonority

Vowels > Glides > Liquids > Nasals > Fricatives > Stops

Figure 2.6 The sonority hierarchy

articulatory measures of sonority include the relative jaw height (Lindblom 1983). Definitions of sonority have also included a measure of audibility or energy particularly in the lower frequencies: the greater the energy, the higher the sonority (Bloch and Trager 1942). These two criteria sometimes conflict, particularly in the case of fricatives where an increased stricture may lead to an increase in amplitude (Keating 1983). However, in general the wider the openness of the vocal tract, the greater the energy in the signal and the higher the sonority. While there is not full agreement about the exact definition of the sonority hierarchy, nor about the amount of detail necessary in its divisions, most current explanations of segment organisation are based on it. Figure 2.6 illustrates a common version of the sonority hierarchy. This organisation results in many of the most commonly attested patterns of segment sequencing. The following are examples from English that illustrate the patterns:

CV, CVC (where C is a consonant and V is a vowel): 'mama' /mama/ or 'tool' /tul/

CGV, CGVC, etc. (where G is a glide): 'pew' /pju/, 'twin' /twɪn/, etc.

CLV, CLVC, etc. (where L is a liquid) for example 'tree' /tri/, 'clear' /klɪr/, etc.

In all of the examples sonority rises and falls regularly so the segment sequencing is said to follow the sonority-sequencing principle. In a revised version of the Sonority Sequencing Constraint that is based on perceptual robustness the traditional patterns hold as well. The following examples list the perceptual basis for the most commonly attested patterns (in descending order of preference; ANF = auditory nerve fibre).

CV, CVC	ANF boost, increased redundancy, increased perceptual distance
CGV, CGVC, etc.	ANF boost, increased redundancy, increased perceptual distance
CLV, CLVC, etc.	ANF boost, increased redundancy, increased perceptual distance
sCV	internal cues (in frication), increased perceptual distance (as long as C2 is not another fricative)

(where s = sibilant fricative)

In past formulations of the Sonority Sequencing Constraint there was no principled way of accounting for the special status of sibilant fricatives; instead they were labelled 'special' or 'exceptional'. In a Sonority Sequencing Constraint that is based on perceptual robustness, a stranded consonant (one without a flanking vowel, liquid, or glide) is dispreferred unless it has sufficiently robust internal cues to survive in the absence of formant transitions. The only consonants that have reliable cues at their peaks of stricture are sibilant fricatives. This is arguably the reason why syllables of the type sCV are cross-linguistically common, even though they have a sonority reversal. The cues for /s/ in sCV are weaker than those for C in CV, so there are also many languages that have CV but lack sCV. If place is sacrificed, then nasals may also occur in sonority reversals: their manner cues are clear even in isolation, but their place cues will disappear making them homorganic with the following obstruent.

5 Conclusion

It should be obvious now that organising the consonants and vowels of speech in certain patterns is going to produce a signal that will encode enough information for the listener to recover place, manner, and voicing more reliably than other patterns will. Robustness can be defined as the redundancy of the cues minus the vulnerability of those cues. That is, the more cues point to a contrast and the less susceptible to masking or loss those cues are the more likely the contrast is to survive. The necessity of maintaining a particular contrast is weighed against other articulatory demands on the system such as the desirability of overlapping adjacent articulations or minimisation of effort (LAZY from Kirchner 2000).

Given the relative resistance to masking of consonant cues that are either encoded in vocalic (vowel and glide) transitions or in gross signal changes (such as manner cues), the ideal segmental organisations will be those that are most commonly attested cross-linguistically: CV, CGV, and so on. These patterns allow maximum degree of overlap between segments and the least risk to portions of the signal that contain vulnerable cues such as consonant bursts or weak frication. Figure 2.7 illustrates the advantages of alternating CVCV over CVC1C2V (where C1 and C2 are both stops). While word onset does not provide the same benefits as the intervocalic positions in terms of cue redundancy, it is still relatively better than coda positions: there is greater redundancy of cues in the CV transition than in the VC transition and the cues have the auditory onset advantage. This is particularly true of stop consonants, which always have a release burst in CV position, and have VOT and a greater attenuation of the signal preceding the burst and the formant transitions resulting in a greater onset advantage.

Segments that we expect to survive without the benefits of flanking vowels, and thus be found at syllable edges with intervening stops, are the sibilant fricatives, potentially other fricatives (depending on intensity of frication), and nasals (particularly if they are homorganic with a following consonant or not otherwise contrasting in place). The same type of pattern should be seen for

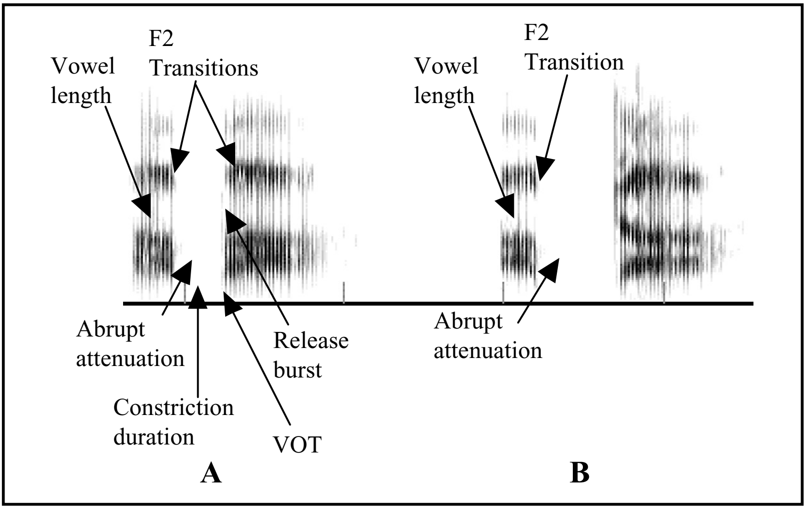


Figure 2.7 Spectrograms of the two nonsense words /aba/ and /abda/ illustrating increased redundancy in V + stop + V (**A**) as opposed to V + stop + stop + V (**B**) sequences. The acoustic cues for the bilabial stop are indicated by the arrows (after Wright forthcoming)

word-final and preconsonantal (coda) consonants. But even for these segments, coda position is worse than the onset position overall. The coda transitions get no boost and suffer from forward masking (saturation of the ANF response). The best codas will be glides and liquids, nasals that do not contrast for place, stops that do not contrast for place, and fricatives and affricates with high intensity energy.

A Sonority Sequencing Constraint that is based on robust sequencing has the added benefit of relating transparently to many other phonotactic constraints. The OCP can be partially motivated (for local effects) because of the perceptual benefit of increased perceptual distance between adjacent segments (Ohala 1992; Flemming 1995/2002). The No Coda constraint can be motivated by the preference for maximal redundancy through overlap and by the auditory onset boost. The Contact Law can be motivated by the optimisation of segments for auditory advantages and recoverability in codas: approximants > fricatives > nasals.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Chicago: Aldine.

Ali, L., T. Gallagher, J. Goldstein, and R. Daniloff (1971). Perception of coarticulated nasality. *JASA* 49: 538–40.

Bailey, P. J. and Q. Summerfield (1980). Information in speech: Observations on the perception of [s]-stop clusters. *Journal of Experimental Psychology: Human Perception and Performance* 6: 536–63.

Bell, A. and J. B. Hooper (eds.) (1978). *Syllables and Segments*. Amsterdam: North-Holland.

Bladon, Anthony (1986). Phonetics for hearers. In Graham McGregor (ed.), *Language for hearers*. Oxford: Pergamon Press, 1–24.

Bloch, B. and G. Trager (1942). *Outlines of linguistic analysis*. Baltimore: LSA Special Publications.

Bregmen, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

Brosches, C. de (1765). *Traité de la formation méchanique de langues, et de principes physiques de l'étymologie*. Paris: Chez Saillant, Vincent, Desaint.

Byrd, D. (1996a). A phase window framework for articulatory timing. *Phonology* 13: 139–69.

(1996b). Influences on articulatory timing in consonant sequences. *JPhon* 24: 209–44.

Clarey, J. C., P. Barone, and T. J. Imig (1992). Physiology of thalamus and cortex. In A. N. Popper and R. R. Fay (eds.), *The mammalian auditory pathway: Neurophysiology*. New York: Springer Verlag.

Clements, G. N. (1990). The role of the sonority hierarchy in core syllabification. In J. Kingston and M. E. Beckman (eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*. Cambridge: Cambridge University Press, 283–333.

Cole, R. A. and W. E. Cooper (1975). Perception of voicing in English affricates and fricatives. *JASA* 58: 1280–7.

Delattre, P. C., A. M. Liberman, and F. S. Cooper (1955). Acoustic loci and transitional cues for consonants. *JASA* 27: 769–73.

Delgutte, B. and N. Y. S. Kiang (1984a). Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics. *JASA* 75: 897–907.

(1984b). Speech coding in the auditory nerve: III. Voiceless fricative consonants. *JASA* 75: 887–96.

Dorman, M. F., M. Studdert-Kennedy, and L. J. Raphael (1977). Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception and Psychophysics* 22: 109–22.

Espy-Wilson, C. Y. (1992). Acoustic measures for linguistic features distinguishing the semi-vowels /w j r l/ in English. *JASA* 92: 736–51.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Flemming, E. (1995). *Auditory representations in phonology*. PhD dissertation, University of California, Los Angeles.

(2002). *Auditory representations in phonology*. New York: Routledge.

Gerstman, L. J. (1968). Classification of self-normalized vowels. *IEEE Transactions on Audio and Electroacoustics* 16: 78–80.

Gordon, M. (1999). *Syllable weight: Phonetics, phonology and typology*. PhD dissertation, University of California, Los Angeles.

Greenberg, S. and S. Rhode (1987). Periodicity coding in the cochlear nerve and ventral cochlear nucleus. In W. A. Yost and C. S. Watson (eds.), *Auditory processing of complex sounds*. Hillsdale, NJ: Lawrence Erlbaum.

Hagiwara, R. (1995). *Acoustic realizations of American /r/ as produced by women and men*. PhD, University of California, Los Angeles.
Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech* 1: 1–7.
Hawkins, S. and K. N. Stevens (1985). Acoustic and perceptual correlates of the non-nasal-nasal distinction for vowels. *JASA* 77: 1560–75.
Heinz, J. M. and K. N. Stevens (1961). On the properties of voiceless fricative consonants. *JASA* 33: 589–96.
House, A. S. (1957). Analog studies of nasal consonants. *Journal of Speech and Hearing Disorders* 22: 190–204.
Jespersen, O. (1913). *Lehrbuch der Phonetik*. Leipzig: B. G. Teubner.
Jun, J. (1995). *Perceptual and articulatory factors in place assimilation: An optimality theoretic approach*. PhD dissertation, University of California, Los Angeles.
Kawasaki, H. (1982). *An acoustic basis for universal constraints on sound sequences*. PhD dissertation, University of California, Berkeley.
Keating, P. A. (1983). Comments on the jaw and syllable structure. *JPhon* 11: 401–6.
Kewley-Port, D., D. B. Pisoni, and M. Studdert-Kennedy (1983). Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. *JASA* 73: 1779–93.
Kiang, N. Y. S., T. Watanabe, E. C. Thomas, and L. F. Clark (1965). *Discharge patterns of single fibres in the cat's auditory nerve*. Cambridge, MA: MIT Press.
Kirchner, R. (1997). Contrastiveness and faithfulness. *Phonology* 14: 83–111.
(2000). Geminate inalterability and lenition. *Lg* 76: 509–45.
Kurowski, K. M. and S. E. Blumstein (1984). Acoustic properties for place of articulation in nasals. *JASA* 81: 1917–27.
Liberman, A. M. and I. G. Mattingly (1989). A specialization for speech perception. *Science* 243: 489–94.
Liberman, A. M., P. C. Delattre, L. J. Gerstman, and F. S. Cooper (1956). Tempo of frequency change as a cue for distinguishing classes of speech sounds. *Journal of Experimental Psychology* 52: 127–37.
Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech* 6: 172–87.
Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (ed.), *The production of speech*. New York: Springer-Verlag, 217–45.
(1990). Explaining phonetic variation: A sketch of the H and H theory. In W. J. Hardcastle and A. Marchal (eds.), *Speech production and speech modelling*. Dordrecht: Kluwer Academic Publishers, 403–39.
Lindblom, B. E. and M. Studdert-Kennedy (1967). On the rôle of formant-transitions in vowel recognition. *JASA* 42: 830–43.
Lisker, L. and A. S. Abramson (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the 6th International Congress of Phonetic Sciences*. Prague: Academia, 563–67.
Malécot, A. (1956). Acoustic cues for nasal consonants: An experimental study involving tape-splicing techniques. *Lg* 32: 274–84.
Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.

Massaro, D. W. and M. M. Cohen (1983). Consonant/vowel ratio: An improbable cue in speech. *Perception and Psychophysics* 33: 501.

Mattingly, I. G. (1981). Phonetic representation and speech synthesis by rule. In T. Myers, J. Laver, and J. Anderson (eds.), *The cognitive representation of speech*. Amsterdam: North Holland, 415–20.

Miller, G. and P. Nicely (1955). An analysis of perceptual confusions among English consonants. *JASA* 27: 338–52.

Miller, M. I. and M. B. Sachs (1983). Representation of stop consonants in the discharge patterns of auditory nerve fibres. *JASA* 74: 502–17.

Møller, A. R. (1977). Coding of time-varying sounds in the cochlear nucleus. *Audiology* 17: 446–68.

Nygaard, L. C. and D. B. Pisoni (1995). Speech perception: New directions in research and theory. In J. L. Miller and P. D. Eimas (eds.), *Speech, language, and communication*. San Diego: Academic Press, 63–96.

O'Connor, J. D., L. J. Gerstman, A. M. Liberman, P. C. Delattre, and F. S. Cooper (1957). Acoustic cues for the perception of initial /w,j,r,l/ in English. *Word* 13: 24–43.

Ohala, J. J. (1992). Alternatives to the sonority hierarchy for explaining segmental constraints. In *Papers from the parasession on the syllable*. Chicago: Chicago Linguistic Society, 319–38.

Repp, B. H. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech* 22: 173–89.

Repp, B. H. and V. A. Mann (1981). Perceptual assessment of fricative-stop coarticulation. *JASA* 69: 1154–63.

Saussure, F. de (1914). *Cours de linguistique générale*. Paris: Payot.

Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Shadle, C. H. (1991). The effect of geometry on source mechanisms of fricative consonants. *JPhon* 19: 409–24.

Shinn, P. and S. E. Blumstein (1984). On the role of the amplitude envelope for the perception of [b] and [w]. *JASA* 75: 1243–52.

Sievers, E. (1881). *Grundzüge der Phonetic*. Leipzig: Breitkopf & Hartel.

Silverman, D. (1997). *Phasing and recoverability*. New York: Garland.

Sinex, D. G. (1995). Auditory nerve fibre representation of cues to voicing in syllable-final stop contexts. *JASA* 90: 2441–9.

Sinex, D. G. and C. D. Geisler (1983). Responses of auditory-nerve fibres to consonant-vowel syllables. *JASA* 73: 602–15.

Smith, R. L. and J. J. Zwislocki (1979). Short-term adaptation and incremental responses of single auditory-nerve fibres. *Biological Cybernetics* 17: 169–82.

Soli, S. D. (1981). Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. *JASA* 70: 976–84.

Steriade, Donca (1995). Positional neutralization. Ms., Department of Linguistics, UCLA. http://www.linguistics.ucla.edu/people/steriade/papers/PositionalNeutralization.pdf.

Stevens, K. N. (1989). On the quantal nature of speech. *JPhon* 17: 3–45.

(1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Stevens, K. N. and A. S. House (1963). Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research* 6: 111–28.

Stevens, K. N. and S. J. Keyser (1989). Primary features and their enhancement in consonants. *Lg* 65: 81–106.

Strange, W. J., J. J. Jenkins, and T. L. Johnson (1983). Dynamic specification of coarticulated vowels. *JASA* 74: 695–705.

Syrdal, A. K. and H. S. Gopal (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *JASA* 79: 1086–100.

Van Dommelen, W. (1982). A contrastive investigation of vowel duration in German and Dutch. *Phonetica* 39: 23–35.

Walley, A. C. and T. D. Carrell (1983). Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. *JASA* 73: 1011–22.

Whitney, W. D. (1874). *Oriental and linguistics studies*. New York: Charles Scribner's Sons.

Wright, Richard (1996). *Consonant clusters and cue preservation in Tsou*. PhD dissertation, University of California, Los Angeles.

(2001). Perceptual cues in contrast maintenance. In E. Hume and K. Johnson (eds.), *The role of speech perception phenomena in phonology*. San Diego: Academic Press, 251–77.

(2004). Factors of lexical competition in vowel articulation. In J. Local (ed.), *Papers in laboratory phonology VI*. Cambridge: Cambridge University Press.

Wright, R., S. Frisch, and D. B. Pisoni (1999). Speech perception. In J. G. Webster (ed.), *Wiley encyclopedia of electrical and electronic engineering*. New York: John Wiley and Sons, Vol. 20, 175–95.

3 Place assimilation*

Jongho Jun

1 Introduction

The present study is concerned with place assimilation in consonant clusters. In such assimilations, one of two neighbouring consonants takes on the place of articulation of another. This occurs, for example, in Diola-Fogny. In (1), /m/ takes on the velar place of the following /g/, becoming [ŋ]:

(1) ni+ga*m*+gam [nigaŋgam] 'I judge' (Sapir 1965)

It is not difficult to derive place assimilation within the framework of previous theories such as classical generative theory, feature geometry, or underspecification theory. This is also true in Optimality Theory (Prince and Smolensky 1993). For instance, if there is a markedness constraint of the type *HETERORGANICCLUSTER, and faithfulness constraints for place of articulation are ranked below it, then place assimilation will be derived, as shown below.

(2) A familiar Optimality-Theoretic analysis of place assimilation

input = / ni+gam+gam /	*HETERORGANICCLUSTER	MAX-PLACE
a. [nigamgam]	*!	
b. ☞ [nigaŋgam]		*

However, such an account leaves unanswered some crucial questions. One wonders, for instance, why languages should have a constraint like *HETERORGANICCLUSTER in the first place – what is it about heterorganicity such that languages should so often avoid it? Moreover, positing a constraint like *HETERORGANICCLUSTER does nothing to account for the interesting typological patterns that have been found to govern place assimilation (Mohanan 1993; Jun 1995; and below). These include the characteristically regressive direction of place assimilation, the greater propensity of nasals to undergo it, the greater propensity of coronals to undergo it, and the greater tendency of non-coronals to trigger it. These patterns are elaborated further below. Any reasonable theoretical account of place assimilation should account for these patterns, which are quite robust cross-linguistically.

In this chapter, I will lay out an approach to place assimilation that aims to achieve these goals. In my approach, there is no role for a constraint like *HeterorganicCluster. Instead, place assimilation is treated as the consequence of interactions among constraints that have phonetic teleologies, both articulatory and perceptual. As a result, the cross-linguistic patterns mentioned above receive explanations that emerge from properties of human speech production and perception. The crucial basis of the explanation lies in asymmetries in the perceptibility of place of articulation in different segments and in different contexts.

The proposal described here represents a minor revision of ideas originally put forth in Jun 1995, 1996a. Subsequent work that has followed similar lines of inquiry includes Steriade 1995, 2001, Myers 1997, and Boersma 1998.

2 Perception of place

There are many potential perceptual cues for phonological place contrasts. For instance, cues for place may be located in the formant transitions of adjacent vowels, in stop release bursts, and in fricative noise. Different consonant types have different cues, and the salience of the cues may be different. Accordingly, different consonant types vary in the degree of perceptibility of place contrasts. In addition, not all potential sources of perceptual cues for each segment type are available in actual speech; the cues may be present or absent (or weakened or enhanced) depending on the context. Such contextual variation in cue distribution is the focus of this section.

To begin, here is a brief overview of the most important place cues and their distribution; for more detailed discussion, see below, as well as Borden, Harris and Raphael 1994, Wright 1996, this volume; and Wright, Frisch, and Pisoni 1996. Place cues may be either *internal* or *transitional*, depending on where the cues are located in time relative to the constriction of a consonant. Internal cues can be found during the acoustic interval corresponding to the consonantal constriction, as in the frication noise of a fricative. Transitional cues are found during the period of coarticulation between a consonant and its neighbouring segments. The most important transitional cues are vowel formant transitions. In CV and VC sequences, articulators move from a consonantal position to a vocalic position or vice versa, changing the vocal tract shape. This shape is acoustically reflected in formant changes during the vowel offset (VC) or onset (CV). Thus, although the cues occur ‘during’ the vowel, they serve as a major source of information concerning the place of articulation of the consonant.

Let us now consider what potential sources of place cues are available for each manner of articulation. Stops bear no cues for place during closure, but when adjacent to vowels they induce formant transitions; many stops also possess a release burst, which results from the sudden venting of high intraoral pressure at release. Earlier studies of the perception of place in stops show that the release burst alone can provide salient place cues (Malécot 1958; Winitz, Scheib, and Reeds 1972). However, formant transitions, especially CV transitions, also provide a good source of place cues, and it is hard to determine the relative contribution of transitions and burst to the identification of place (Manuel 1991; Byrd 1992; Wright this volume; and others).

For fricatives, the spectrum of noise provides an internal place cue. The noise spectrum is highly reliable for the identification of place in sibilant fricatives, but less so in nonsibilant fricatives (Wright et al. 1996). Thus, formant transitions on an adjacent vowel play an important role in the perception of nonsibilant fricatives, compared with sibilants.

Nasals are produced with complete oral closure just like stops, and bear cues in the formant transitions of neighbouring vowels similar to those of stops. Nasals lack bursts, but have internal cues in the form of the nasal resonance (murmur) during the period of closure. However, the nasal resonance cues are less reliable in identifying the place of a nasal than the vowel formant transition cues (Malécot 1956).

Unlike other consonants, approximants (liquids and glides) have a formant structure that serves as an internal cue. They also benefit from transition cues on the neighbouring vowels; typically, formants change fairly gradually in the transition between vowels and approximants.

The potential sources of consonant place cues just discussed are summarised in (3).

(3) Sources of consonant place cues

	Cue types	
Segment types	Internal	Transitional
stops	none	CV, VC formant transitions, release burst
nasals	nasal resonance	CV, VC formant transitions
fricatives	frication noise	CV, VC formant transitions
liquids and glides	formant structure	CV, VC formant transitions

For present purposes, it is the *relative* perceptibility of consonants that is crucial. Below, I compare the perceptibility of consonants under four different conditions:

1. Comparison between preconsonantal consonants with different manner features (*target manner*)
2. Comparison between preconsonantal consonants with different place features (*target place*)
3. Comparison between preconsonantal and prevocalic consonants (*target position*)
4. Comparison between consonants occurring before consonants with different place features (*trigger place*)

Variation in perceptibility in each condition will be seen to explain a number of patterns of assimilation.

2.1 Target manner

This section discusses the relative perceptibility of place cues in stops, nasals, fricatives, liquids, and glides. For reasons that will be clear later on, I focus here on the effects of manner in consonants that occur as the first member of a consonant cluster; specifically, as C_1 in $V_1C_1C_2V_2$. Consonants in this position often lack salient place cues; in particular, CV formant transitions are not available. In a $V_1C_1C_2V_2$ sequence, V_2 normally bears little information about C_1 due to the intervening C_2; the formant transitions in the onset of V_2 will mostly have cues for C_2. In addition, when C_1 is a stop, it is often unreleased due to overlap with C_2 (especially when C_2 is a nasal or obstruent) and thus lacks a release burst. Thus, the pattern of cues seen earlier in (3) is reduced in this context to (4):

(4) Sources of consonant place cues in pre-C position (esp. C_2 = a nasal or obstruent)

	Cue types	
Segment types	Internal	Transitional
stops	none	VC formant transitions
nasals	nasal resonance	VC formant transitions
fricatives	frication noise	VC formant transitions
liquids and glides	formant structure	VC formant transitions

The crucial comparison here will be between noncontinuants (stops and nasals) vs continuants (fricatives and approximants) in preconsonantal position. The crucial point will be that noncontinuants suffer a proportionately greater disadvantage from occurring in this context relative to continuants.

Stops in preconsonantal position lack prominent cues in the release burst and CV transitions. VC formant transitions may provide the only place cues for stops in this location; and as will be discussed below, cues in VC formant transitions are much less prominent than those in CV transitions. Therefore, perceptibility of the stop place is drastically reduced in preconsonantal position, compared to prevocalic position.

This drastic weakening of place cues is also true of the nasals. According to Kurowski and Blumstein's (1984) perceptual experiments, the nasal murmur and transitions surrounding the nasal release provide the most reliable place cues, but neither the murmur immediately preceding the release nor the transition immediately following the release provides a salient place cue for the nasal by itself. Since in preconsonantal position nasals are normally unreleased, and unreleased nasals lose the most prominent cues around the release, the perceptibility of nasals' place cues thus is drastically reduced in preconsonantal position.

In contrast, if the consonant occurring preconsonantally is a continuant, it maintains its internal cues in addition to the cues in the VC formant transitions. Thus, even if a fricative or non-nasal sonorant overlaps with a consonant in C_2, its place cues can be well preserved. A short period of non-overlapping frication at the beginning of a fricative, especially in the case of sibilants, can provide somewhat stable place cues. Thus, fricatives have more robust place cues than stops and nasals in preconsonantal position. Hura, Lindblom, and Diehl (1992) carried out perception tests to compare English fricatives, stops, and nasals in confusability when occurring before a stop. Their results indicate that fricatives were less confusable than stops and nasals. Also, as discussed above, the critical place cues for glides and liquids are located in relatively gradual frequency changes. Especially in case of (English) liquids, the frequency changes may persist for much of the preceding vowel's duration. Such gradual changes can provide robust perceptual cues. In conclusion, continuant consonants – fricatives and non-nasal sonorants – have more robust place cues than stops and nasals when occurring as the first member of a consonant cluster.

Nasals and stops differ with respect to perceptibility of place information. Both suffer from a lack of release in preconsonantal position, but nasals have an additional handicap. In a VN sequence, the vowel is characteristically nasalised, either throughout, or at least during the crucial period of the consonant transitions. Nasalised vowels are perceptually difficult in comparison with oral vowels, and their amplitudes are decreased by anti-formants (Johnson 1997). Thus, nasalised vowels are normally less distinct than oral vowels. This acoustic and perceptual difference between nasalised and oral vowels may lead us to assume that formant transitions of nasalised vowels are less distinct with respect to the place of the following consonant than those of oral vowels. In other words, nasals have less prominent place cues in the VC transitions than stops, suggesting the relative perceptibility of stops over nasals in preconsonantal position. This argument is consistent with results of the perception experiment by Hura et al. mentioned above: it was found that when occurring before a stop, stops were less confusable than nasals, although this difference was not statistically significant.

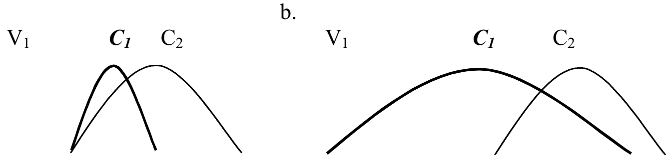
To summarise this section: in preconsonantal position, nasals have weaker place cues than stops, which in turn have weaker cues than continuants.

2.2 *Target place*

Consider the relative perceptibility of place cues of coronals, labials, and velars when they are unreleased, occurring as the first member of a consonant cluster. Experimental evidence (Byrd 1994, citing Öhman 1967, Kuehn and Moll 1976, and Winitz et al. 1972) has established that a coronal, specifically [d], has perceptually weaker cues compared to noncoronals. This weakness can be explained as follows (cf. Browman and Goldstein 1990; Kang 1999). Typically, the underlying gesture with which coronals are realised is articulated more rapidly. That is, tongue tip gestures are rapid and thus have rapid transition cues; whereas tongue dorsum and lip gestures are more sluggish and thus give rise to long transitions.

This has important consequences for the acoustic effects of articulatory overlap. Consider the difference between a longer and a shorter consonant articulation gesture, when the consonant occurs in preconsonantal position:

(5) a. b.



rapid $\mathbf{C_1}$ gesture   slow $\mathbf{C_1}$ gesture

When the gesture for C_1 is articulated rapidly, the formant transitions at the end of V_1 will be affected not just by C_1, but also by the overlapping C_2. When the gesture for C_1 is made slowly, however, the transitions in V_1 will result almost entirely from C_1. Thus, a longer-gesture consonant in C_1 (i.e. a velar or labial) will be rendered more identifiable by the preceding formant transitions than will a shorter-gesture consonant (such as a coronal).

The claim for the perceptual weakness of unreleased coronals relative to labials and velars is supported by results of Winitz, Scheib, and Reeds's (1972) perception study of the identification of English voiceless stops. In this study, the release burst of a stop and 100 msec of the adjacent vowel were isolated from English monosyllable words taken from running speech. When stop bursts alone, isolated from word-final VC sequences, were presented to subjects, alveolars were found to be most accurately identified. However, when 100 msec of a vowel segment preceding the stop was additionally employed, alveolars became least accurately identified. These results indicate not only that alveolars are the most confusable stops at the end of the word, but also that vowel transitions into an alveolar stop are least informative of its place of articulation. Consequently, the results support the claim that coronals have weaker place cues than noncoronals when they lack cues in the release burst.

Other results of Winitz et al.'s perception study may suggest an additional distinction among noncoronals. In the comparison between two perception tests, one employing stimuli with stop bursts alone and the other with stop bursts plus 100 msec of adjacent vowel, the accuracy of identification greatly improved for final velars compared to labials (some improvement) and alveolars (almost no improvement). My interpretation of these results is that vowel transitions into a velar stop are more informative of the stop's place of articulation than those for labials and coronals. If this interpretation is correct, velars have stronger place cues than coronals and labials when they are unreleased.

This difference may be attributed to an acoustic characteristic property of velars, that is, compactness (Jakobson, Fant, and Halle 1963). Velars can be characterised by a noticeable convergence of F2 and F3 of neighbouring vowels. These two formants can form a prominence in the mid-frequency range. As argued by Stevens (1989: 17–18), such a mid-frequency prominence of velars can form a robust acoustic cue for place of articulation. Listeners do not have to know the exact target points of F2 and F3 transitions to identify velars; mere convergence of two formants will provide a sufficient cue, regardless of where the convergence occurs.

The general conclusion of this section is that place cues of unreleased dorsals are more perceptible than those of unreleased labials, which in turn are more perceptible than those of unreleased coronals.

2.3 *Target position*

This section concerns the relative perceptibility of consonants in prevocalic and preconsonantal positions. It has been shown in the literature (Bladon 1986; Manuel 1991; Ohala 1990, 1992 among others) that prevocalic or syllable-initial consonants are acoustically/perceptually stronger than preconsonantal or syllable-final consonants. In section 2.1, I related this claim to patterns of release: stops and nasals are released before vowels but typically unreleased before consonants, especially obstruents and nasals. Thus, in preconsonantal position, both stops and nasals have only weak place cues, losing the prominent release burst cues and cues around the nasal release, respectively.

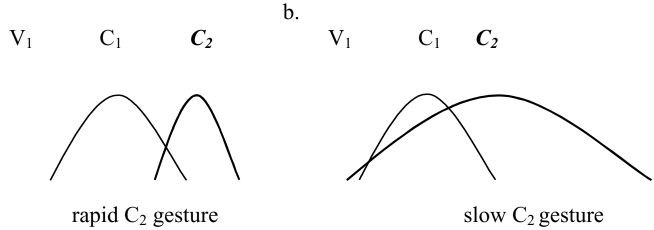
Another important factor is the relative perceptibility of CV and VC formant transitions. Experimental results consistently show that CV transitions provide more prominent place cues than VC transitions (Repp 1978; Fujimura, Macchi, and Streeter 1978; Wright 1996; Ohala 1990 and references cited therein). One possible explanation for this difference lies in the auditory response system (see Wright this volume; Côté 2000): the onset of the speech signal is emphasised compared to its offset. Because of this, CV transitions out of a consonant closure are effectively amplified, whereas VC transitions into a consonant closure are not.

Thus, for two reasons – patterns of release and auditory system asymmetries – the place cues of a consonant are stronger in prevocalic position than in preconsonantal position.

2.4 Trigger place

The perceptibility of place cues of C_1 in a cluster C_1C_2 also varies as a function of the place of C_2. Due to coarticulation, the formant transitions of V_1 are affected by both C_1 and C_2, although C_1's influence is dominant (Byrd 1992; Zsiga 1992). Thus, C_2 with different place features may obscure place cues of the C_1 to differing degrees. I claim that this depends on the inherent velocity of the articulator involved in C_2: a slower gesture for C_2 obscures the place cues for C_1 more easily. The schematic representations in (6) can illustrate this point.

(6) a. b.



In (6a), a rapid C_2 gesture (marked in bold) slightly overlaps with the preceding C_1 gesture; so that C_2's influence onto VC formant transitions would be minimal. Example (6b) illustrates that a slow gesture of C_2 may begin even before the C_1 gesture does; thus it will greatly obscure the VC transition cues to identification of the C_1. The slower the gesture for C_2, the greater this effect will be. As discussed above, coronals are characterised by rapid articulatory gestures, whereas noncoronals are characterised by slow gestures. From this, it follows that the place cues of C_1 can be obscured more easily before noncoronals than coronals.

2.5 Summary

The four perceptibility scales proposed in this section are repeated below.

(7) Perceptibility Scales for Place
  a. Target manner: [+cont]/__C > [stop]/__C > [nasal]/__C
  b. Target place: unreleased dorsal > unreleased labial > unreleased coronal
  c. Target position: __V > __C
  d. Trigger place: __coronal > __noncoronal

It will be shown in section 4 that typological patterns of place assimilation follow from these perceptibility scales.

3 Typology

Place assimilation patterns display language-specific variability. However, the variability is not unconstrained; there are systematic gaps and asymmetries. This section discusses a data survey, building on Mohanan 1993, which confirms this claim. The discussion below is based on data from Brussels Flemish, Catalan, Diola-Fogny, English, German, Hindi, Inuktitut dialects, Keley-i, Korean, Lithuanian, Malay, Malayalam, Nchufie, Thai, Toba Batak, Yakut, and Yoruba. However, the assimilatory patterns that emerge are the same as those discussed in a much larger data set employed in Jun 1995. I will classify the attested patterns according to the same four categories used in the preceding section.

3.1 Target manner

The following table shows a summary of surveyed patterns of place assimilation with respect to the manner of the target:[1]

(8) Patterns for target manner

Language list	Nasal	Stop	Continuant
Catalan[2], English, German, Korean, Malay, Thai, Yakut	yes	yes	no
Brussels Flemish, Diola Fogny, Hindi, Keley-I, Lithuanian, Malayalam, Nchufie, Toba Batak, Yoruba	yes	no	no

(yes = ‘targeted’, no = ‘untargeted’ and ‘undetermined/unknown’)

In the above table, ‘yes’ indicates that a consonant produced with the corresponding manner of articulation can be targeted in place assimilation, and ‘no’ indicates that the consonant cannot be targeted. In some cases it is impossible to determine whether the relevant consonant can be targeted or not; for instance, this is true of fricatives in Korean, since only codas can be targeted in Korean place assimilation but fricatives never appear in coda position. There are also cases whose patterns are not known to me. These cases are included in the ‘no’ category.

From the above list, two observations can be made. First, continuants (i.e. fricatives, liquids, and glides) virtually never undergo place assimilation. (There are cases, e.g. Japanese, in which continuants can undergo place assimilation as part of a process of total assimilation; such cases are excluded from consideration here.)

Second, among the noncontinuants, nasals are more likely to undergo place assimilation than stops, and there is an implicational relation: if in some languages stops can be targeted for assimilation, so can nasals. Thus, languages such as Malayalam allow only nasals to be targeted, and languages like English target both oral stops and nasals, but there are no languages in which only oral stops can be targeted.

3.2 Target place

The following table sorts the surveyed data according to the place of articulation of the target:

(9) Patterns for target place

Language list	Coronal	Labial	Velar
Diola Fogny, Malay, Nchufie, Yoruba, Thai	yes	yes	yes
Korean	yes	yes	no
Hindi, Malayalam	yes	yes	
Catalan, English, German, Toba Batak, Yakut	yes	no	no
Brussels Flemish, Keley-I, Lithuanian	yes	no	

(‘yes’ = ‘targeted’, ‘no’ = ‘untargeted’, blank = ‘undetermined or unknown’)

‘Yes’ indicates that a consonant produced at the corresponding place of articulation can be targeted in place assimilation; ‘no’ indicates that it cannot. Unlike in (8), undetermined (and unknown) cases are unmarked. This table shows that consonants produced at different places of articulation tend to be targeted in place assimilation to a different degree. First, coronals are a target in all surveyed languages. The prevalence of coronals among place assimilation targets has been noted earlier by Bailey (1970), Kiparsky (1985), Cho (1990), Paradis and Prunet (1991), and others. In Brussels Flemish, Catalan, English, German, Keley-I, Lithuanian, Toba Batak, and Yakut, only coronals are targeted in place assimilation. There are also languages such as Diola-Fogny, Hindi, Korean, and Malayalam in which not only coronals but also noncoronals can be targeted. However, I am not aware of any languages where only noncoronals are targets. Thus, it seems that in any language, if noncoronals can be targets of place assimilation, so can coronals.

Among the languages surveyed, Korean is the only language that shows a clear asymmetry among noncoronals, that is, between labials and velars: Korean place assimilation targets both coronals and labials, but not velars. However, additional data in Dorais' (1986) survey of Inuktitut dialects (not included in my own database) involve a number of cases in which coronals and labials, but not velars, assimilate.[3] The Korean and Inuktitut data together suggest that labials are more likely to undergo place assimilation than velars. We propose, then, the implicational statements in (10):

(10) Target Place
  a. If velars are targets of place assimilation, so are labials.
  b. If labials are targets of place assimilation, so are coronals.

3.3 *Position of the target consonant*

It is well known that in intervocalic position regressive assimilation is much more common than progressive assimilation. Beckman (1997: 22) suggests that this is not just a tendency: where all else is equal between C_1 and C_2, 'progressive assimilation in consonant clusters is virtually unattested'. The preference for C_1 target is confirmed in my survey data: all the patterns involve C_1, not C_2, as a target. Webb's (1982) survey, in which two hundred languages were examined as part of the Stanford Archiving Project, yielded similar results.

However, there are two exceptions to the pattern of regressive assimilation, both of which find independent explanations. First, in cases like Kambata (Hudson 1980: 105), Musey (Shryock 1993), and Dutch (Trommelen 1983), the initial consonant of a suffix assimilates to the final consonant of a stem. This appears to reflect the widely noted tendency that stems resist alternation in comparison to affixes (see, e.g., Casali 1996; Silverman 1995). The second exception is that in word-final VC_1C_2, C_2 sometimes assimilates to C_1, as in German /va:gən/ ➔ [va:gŋ̩] 'Wagen' (Kohler 1990: 83). This 'exception' is only an apparent one: since the word-final C_2 lacks most of the perceptual cues of C_2 in VC_1C_2V, none of the discussion above is relevant to such cases.

In sum, except for these special contexts, place assimilation appears to be always regressive.

3.4 *Trigger place*

Most surveyed patterns do not show any asymmetries with respect to the place of articulation of the trigger consonant. A relevant case, however, can be found in Korean place assimilation.

(11) Korean (from Jun 1996b)[4]
a. Noncoronals are triggers.
/ip+ko/ → [ikko] 'wear and'
b. Coronals are not triggers.
/ip+tolok/ → [iptolok], *[ittolok] 'wear + causative marker'

In Korean, labials do not assimilate to following coronals (11b), but they do assimilate to following velars (11a). A similar asymmetry is seen in Latin stop assimilation (Kühner and Holzweissig 1966): stops assimilate to following velars and labials (e.g. su/b+k/urro → su[kk]urro 'I rush up to') but they do not assimilate to following coronals (scri/b+t/um → scri[pt]um, not *scri[tt]um 'write-neuter passive participle'). In the absence of counterexamples, I will assume the following descriptive statement: If coronals trigger place assimilation, so do noncoronals.

3.5 *Summary*

The typological survey of place assimilation presented above yields the following implicational statements.

(12) Implicational statements of place assimilation
(a) Target manner
(i) If continuants are targets of place assimilation, so are stops.
(ii) If stops are targets of place assimilation, so are nasals.
(b) Target place
(i) If velars are targets of place assimilation, so are labials.
(ii) If labials are targets of place assimilation, so are coronals.
(c) Position of target
In C_1C_2 cluster, if C_2 is a target of place assimilation, so is C_1 (exceptions: C_2 in suffixes; C_2 not prevocalic)
(d) Trigger place
If coronals are triggers, so are noncoronals.

In the following section, I develop a formal analysis intended to capture these generalisations.

4 Analysis

To begin, it will be useful to consider the phonetic events that occur in place assimilation. Two articulatory gestures, encoding the place of articulation of the target and trigger segments, are involved. Occurrence of place assimilation means loss of the target gesture, with concomitant extension in time of the trigger gesture so as to occupy the slot formerly held by the target.[5]

It has been suggested by Browman and Goldstein (1990) that loss of gestures does not occur in assimilation, and that assimilation is simply the submergence of one articulatory gesture under another. My own data suggest otherwise. In Jun 1996b I explored the patterns of gestural overlap and reduction in Korean and English labial-initial stop clusters, for example /ipku/, using intraoral pressure data. I found that mere gestural overlap does not yield perceptual place assimilation; overlapped /pk/ is in fact heard as [pk]. Assimilation is perceived only when there is gestural reduction (partial or complete) of the target consonant. Moreover, the observed gestural reduction occurs in exactly the contexts where traditional phonological analysis posits a process of assimilation; that is, for labials preceding velars, not coronals. Nolan's (1992) study of assimilation of coronals in English obtained similar results. Thus, the articulatory process that corresponds at the phonetic level to phonological assimilation appears to be more complex than mere overlap. Rather, the correct mechanism appears to be gestural reduction of the target consonant, with concomitant temporal extension of the trigger consonant. The analysis below proceeds under this assumption.

Why do speakers reduce articulatory gesture of the target segment? A widely held view of phoneticians (e.g. Lindblom 1983, 1990) is that speech production is the result of reconciling two conflicting requirements: ease of articulation and ease of perception. Under this view, gestural reduction occurs in order to satisfy the requirement of ease of articulation. (We will see below, however, that perceptual requirements also play a role.)

Let us consider how assimilation can be treated formally within this general view. I assume the general theoretical framework of Optimality Theory (Prince and Smolensky 1993). My proposal is that the constraints predicting the patterns of place assimilation represent the grammatical reflexes of the requirements of articulation and perception (cf. Kohler 1990; Mohanan 1993).

A constraint motivated by ease of articulation, which I call WEAKENING, formulates the minimisation of articulatory effort, as stated in (13):[6]

(13) WEAKENING: Conserve articulatory effort.

Following Kirchner (1998, this volume), I assume that violations of WEAKENING are assessed based on the effort cost (i.e. 'a mental estimate of the biomechanical energy') required for the articulation of each candidate. For present purposes a simplified computation of the effort cost may be assumed: an arbitrary positive value is assigned to all complete closure gestures, while a zero value is assigned in cases of elimination of gestures. (In section 5, a finer distinction between the effort costs will be made in the analysis of gradient assimilation.) A violation mark of WEAKENING is incurred whenever an articulatory gesture occurs in a candidate. WEAKENING thus has the effect of reducing or eliminating articulatory gestures, leading to place assimilation in consonant clusters.

WEAKENING conflicts with faithfulness constraints, stated here as mandating the preservation of the perceptual cues to features (Flemming 1995):

(14) PRESERVE: Preserve perceptual cues for input features.

I follow Steriade (1997) in assuming that faithfulness constraints are evaluated relative to a hypothesised phonetic interpretation of the input string; this is a mental representation that includes the articulatory realisation of the input features and the acoustic-auditory consequences of the articulations. Faithfulness to the input is assessed in terms of maximal preservation of these acoustic-auditory properties of the phonetically interpreted input. PRESERVE has the indirect effect of preserving gestures insofar as these gestures have acoustic/auditory consequences; thus gestural reduction or elimination is penalised, and therefore, so is place assimilation.

The tableaux in (15) illustrate how the occurrence and absence of place assimilation follow from the interaction of the WEAK(ENING) and PRES(ERVE) constraints.

(15) Occurrence and absence of place assimilation

a.

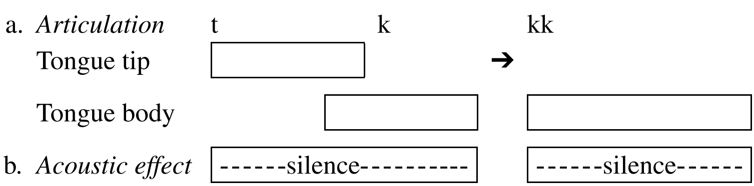
/tk/	Articulation	Perception	WEAK	PRES(place)
(i)	coronal dorsal	[tk]	**!	
(ii) ☞	coronal dorsal	[kk]	*	*

b.

/tk/	Articulation	Perception	PRES(place)	WEAK
(i) ☞	coronal dorsal	[tk]		**
(ii)	coronal dorsal	[kk]	*!	*

For each output candidate, articulatory gestures encoding the input place features are represented with boxes, and their corresponding perceptual consequences are represented with phonetic symbols. In candidate (15i), two gestures are made and thus two violation marks of Weakening are provided, whereas in candidate (15ii) only a single gesture is made, thus incurring a single violation of Weakening.[7] Consider next the evaluation of Preserve(place). Candidate (15i) is a hypothesised phonetic interpretation of the input in which all the input features are phonetically realised. As such, it preserves all the perceptual information of the input places, and therefore incurs no violations of Preserve(place). In contrast, in candidate (15ii), the coronal gesture is not made, and thus its perceptual information is not preserved, so that Preserve(place) is violated.[8] When Weakening outranks the Preserve constraint for place features, place assimilation occurs (15a); otherwise, no assimilation results (15b).

In assimilation, deletion of the target gesture is accompanied by the lengthening of the trigger gesture, which is another aspect of the process that a phonological analysis must account for. I treat this process as a kind of faithfulness effect: the goal is to maintain in the output the manner cues of the input target gesture. For example, when the target is a voiceless stop, the period of silence obtained by lengthening the trigger gesture will have the effect of preserving the manner cues of the target gesture.

(16)


In (16a), the boxes represent stop closure gestures which result acoustically in silence, as represented in (16b). By comparing the left and right sides of the arrow, we observe that the tongue tip closure gesture is completely reduced and that the tongue body closure gesture lengthens. The stop closure is acoustically silent; thus, there would not be any loss of the stop manner cue, that is, silence.

In terms of the formal analysis, place assimilation – as opposed to deletion – occurs under the following ranking: {Preserve constraints for manner cues} ≫ Weakening ≫ {Preserve constraints for place cues}. This can be seen in tableau (17). (In the remainder of this paper, candidates in tableaux will include only phonetic symbols indicating the perceptual consequences of the articulatory gestures involved.)

(17) Occurrence of place assimilation, not deletion

Input = /tk/	PRES(manner)	WEAK	PRES(place)
a. ☞ kk (assimilation)		*	*
b. tk (no change)		**!	
c. k (deletion)	*!	*	*

Let us consider how this analysis can be used to explain the typological results given above. The crucial principle will be a universal ranking for PRESERVE constraints, which is a formal implementation of the hypothesis in (18) (cf. Kohler 1990, 1991, 1992; Steriade 1993; and Byrd 1994):

(18) Production Hypothesis
Speakers make more effort to preserve the articulation of speech sounds with relatively more powerful acoustic cues.

According to this hypothesis, speakers exert less effort on articulations that present inherent acoustic weaknesses. They exert more effort on articulations with acoustically salient consequences, since the effort will pay off in enhancing the perceptibility of the segment.

The Production Hypothesis provides a general basis for ranking PRESERVE constraints: constraints preserving perceptually more salient segments must be ranked above those preserving perceptually less salient segments. I propose the following formalisation of this idea:

(19) PRES(X(Y)): Preserve perceptual cues for X (place or manner of articulation) of Y (a segmental class).
Universal ranking: PRES(M(N)) ≫ PRES (M(R)),
where N's perceptual cues for M are stronger than R's cues for M.

Assuming (19), the following sets of position- or segment-specific PRESERVE constraints and their internal rankings can be directly projected from the perceptibility scales proposed in section 2.

(20) Position/segment-specific PRESERVE constraints and universal rankings
a. Target manner: PRES(pl($\underset{\text{[+cont]}}{\rule{1.5em}{0.4pt}}$C)) ≫ PRES(pl($\underset{\text{[stop]}}{\rule{1.5em}{0.4pt}}$C)) ≫ PRES(pl($\underset{\text{[nasal]}}{\rule{1.5em}{0.4pt}}$C))
b. Target place: PRES(pl(dorsal$^{\neg}$)) ≫ PRES(pl(labial$^{\neg}$)) ≫ PRES(pl(coronal$^{\neg}$))

c. Target position: PRES(pl(_V)) ≫ PRES(pl(_C))
d. Trigger place: PRES(pl(__coronal)) ≫ PRES(pl(__noncoronal))

I will now show how the interaction of these ranked faithfulness constraints and WEAKENING explains typological patterns of place assimilation discussed in section 3.

4.1 Target manner

The ranking in (20a) indicates that place cues of continuant consonants must be preserved in preference to stop place cues, which are, in turn, preserved in preference to nasal place cues. This ranking mirrors the implicational statements about manner of articulation of the target consonant, shown in (12a). More specifically, the ranked constraints in (20a) may produce different assimilation patterns depending on their ranking relative to WEAKENING:

(21) *Possible language-specific rankings*

a. ***WEAKENING***≫PRES(pl($\underset{\text{[+cont]}}{\rule{2em}{0.4pt}}$C)) ≫ PRES(pl($\underset{\text{[stop]}}{\rule{2em}{0.4pt}}$C)) ≫ PRES(pl($\underset{\text{[nasal]}}{\rule{2em}{0.4pt}}$C))
→ Continuants, stops, and nasals are all targets.

b. PRES(pl($\underset{\text{[+cont]}}{\rule{2em}{0.4pt}}$C))≫***WEAKENING***≫PRES(pl($\underset{\text{[stop]}}{\rule{2em}{0.4pt}}$C)) ≫ PRES(pl($\underset{\text{[nasal]}}{\rule{2em}{0.4pt}}$C))
→ Stops and nasals are targets but continuants are not.

c. PRES(pl($\underset{\text{[+cont]}}{\rule{2em}{0.4pt}}$C)) ≫ PRES(pl($\underset{\text{[stop]}}{\rule{2em}{0.4pt}}$C)) ≫ ***WEAKENING*** ≫ PRES(pl($\underset{\text{[nasal]}}{\rule{2em}{0.4pt}}$C))
→ Only nasals are targets.

All attested patterns, discussed in section 3.1, can be derived from the rankings above. For example, (21b) explains patterns in which stops and nasals can be targeted but continuants cannot, as in English, German, Malay, and Yakut. Ranking (21c) derives patterns in which only nasals can be targeted, as in Brussels Flemish, Hindi, Keley-I, Malayalam, and Toba Batak. Moreover, there are no possible rankings for unattested patterns. For instance, patterns in which stops, but not nasals, can be targeted, do not exist. Such a pattern would require that WEAKENING should outrank PRESERVE for stops but at the same time be outranked by the PRESERVE for nasals. This is impossible according to (20a). To demonstrate how attested patterns can be analysed by the proposed mechanism, let us consider cases in which only nasals are targeted by assimilation. The following tableaux illustrate an analysis of Malayalam data (taken from Mohanan 1993).

(22) a. Nasals are targets: /awan+kaɾaññu/ ➔ [awaŋkaɾaññu] ‘he cried’

Input = /awan+kaɾaññu/	PRES(pl(—[stop] C))	WEAK	PRES(pl(—[nasal] C))
(i) awankaɾaññu		**!	
(ii) ☞ awaŋkaɾaññu		*	*

b. Stops are not targets: /ut̪karʂ am/ –x-> *[ukkarʂ am] ‘progress’

Input = /ut̪karʂam/	PRES(pl(—[stop] C))	WEAK	PRES(pl(—[nasal] C))
(i) ☞ ut̪karʂam		**	
(ii) ukkarʂam	*!	*	

(In the above and the remainder of this chapter, only a crucial consonant cluster is considered in assessing violations of constraints)

In (22a), the fully faithful candidate incurs double violations of WEAKENING, since both coronal and dorsal gestures are made. In contrast, candidate (22a.ii), with place assimilation, incurs a single violation of WEAKENING, since the coronal gesture is reduced and so only a single gesture, that is, dorsal, is made. This candidate also violates the PRESERVE constraint for place of articulation of a nasal; that is, it does not preserve the place cues of the coronal nasal, due to the reduction of the coronal gesture. But since the PRESERVE constraint is lower ranked than WEAKENING, candidate (22a.ii) is the optimal output. Consequently, it is shown that if C_1 is a nasal, place assimilation occurs. In (22b), a candidate displaying place assimilation violates a higher-ranked PRESERVE constraint for place of articulation of an oral stop. In contrast, the faithful candidate violates only the lower-ranked WEAKENING, and thus is optimal. Thus, if C_1 is a stop, no assimilation occurs. In conclusion, asymmetric typological patterns of place assimilation with respect to the manner of the target can be analysed by the interaction of WEAKENING and internally ranked PRESERVE constraints, as proposed above.

There is an alternative account for some of the typological patterns of place assimilation analysed in this section. One might argue that the reason why fricatives, liquids, and glides are rarely targeted in place assimilation is because the continuants are often limited to the small number of places of articulation, not because their place cues are prominent. For instance, if [s] is the only acceptable fricative in a certain language, /sk/ would not become /xk/ even if /tk/ became [kk]. This can be analysed by ranking a structure preservation constraint *x and faithfulness constraints for manner, PRES(mnr), over WEAKENING:

(23) No fricative target: /sk/ ➔ [sk] (with no general occurrence of [x])

Input = /sk/	*x	PRES(mnr)	WEAK	PRES(place)
a. ☞ sk			**	
b. xk	*!		*	*
c. kk		*!	*	*

Here, the assimilating candidate *[xk] is ruled out by undominated *x. However, an analysis that relies on structure preservation constraints to block assimilation of continuants cannot provide a general account for the asymmetries under consideration. There are languages in which continuants with a variety of places of articulation can occur in general but fricative assimilation is still blocked. For instance, in German, the fricative /s/ cannot yield [x] through assimilation to velars (e.g. au[sg]eben, *au[xg]eben 'to spend'), even though [x] would be legal in this position. It seems that only the perception-based approach proposed in the present study can provide a general account for manner asymmetries in place assimilation.

4.2 *Target place*

The ranking in (20b), PRES(pl(dorsal¬)) ≫ PRES(pl(labial¬)) ≫ PRES(pl(coronal¬)), captures the implicational statements about place of articulation of the target consonant in (12b), namely, that if velars are targets of place assimilation, so are labials, and that if labials are targets of place assimilation, so are coronals. The following shows the interactions of WEAKENING and PRESERVE constraints that are consistent with (20b):

(24) *Possible language-specific rankings*
a. ***WEAKENING*** ≫ PRES(pl(dor¬)) ≫ PRES(pl(lab¬)) ≫ PRES(pl(cor¬))
→ Coronals and noncoronals are all targets.
b. PRES(pl(dor¬)) ≫ ***WEAKENING*** ≫ PRES(pl(lab¬)) ≫ PRES(pl(cor¬))
→ Labials and coronals are targets but velars are not.
c. PRES(pl(dor¬)) ≫ PRES(pl(lab¬)) ≫ ***WEAKENING*** ≫ PRES(pl(cor¬))
→ Only coronals are targets.

All patterns predicted from the interaction of the relevant constraints are identical with the typological patterns of place assimilation discussed in section 3.2.

Consider, for instance, the cases in which only coronals are the target. The following tableaux are for the English pattern, where (for example) /t/ assimilates to /k/ but /p/ does not.

(25) a. Coronals are targets: /let kɔl/ ➔ [lek kɔl] ‘late call’

Input = /let kɔl/	PRES(pl(lab¬))	WEAK	PRES(pl(cor¬))
(i) let kɔl		**!	
(ii) ☞ lek kɔl		*	*

b. Noncoronals are not targets: /lip kwɪkli/ -x-> [lik kwɪkli] ‘leap quickly’

Input = /lip kwɪkli/	PRES(pl(lab¬))	WEAK	PRES(pl(cor¬))
(i) ☞ lip kwɪkli		**	
(ii) lik kwɪkli	*!	*	

Here, WEAKENING is ranked between PRES(pl(lab¬)) and PRES(pl(cor¬)). Thus a coronal gesture must be reduced to obey WEAKENING, sacrificing lower-ranked PRES(pl(cor¬)) (25a). In contrast, a labial gesture must be maintained to obey the dominant PRES(pl(lab¬)), sacrificing WEAKENING (25b). Similar rankings, not given here, derive the other possible patterns in analogous ways. There are no possible rankings for unattested patterns: for example, only labials as the target. In order for labials but not coronals to be the target of place assimilation, WEAKENING must outrank PRES(pl(lab¬)) and at the same time be outranked by PRES(pl(cor¬)). This is impossible according to the proposed universal ranking in (20b).

4.3 Target position

Here the crucial ranking is the one given in (20c): PRES(pl(__V)) ≫ PRES(pl(__C)). Placing WEAKENING in the relevant locations for this ranking, we obtain the following:

(26) *Possible language-specific rankings*

a. PRES(pl(__V)) ≫ ***WEAKENING*** ≫ PRES(pl(__C))
➔ Only C_1 is the target.

b. ***WEAKENING*** ≫ PRES(pl(__V)) ≫ PRES(pl(__C))
➔ Both C_1 and C_2 are the target.

The following tableau illustrates regressive assimilation in Yakut.

(27) Yakut: /at+ka/ ➔ [akka], *[atta] 'to a horse' (Krueger 1962)

Input = /at+ka /	PRES(pl(__V))	WEAK	PRES(pl(__C))
a. atka (no change)		**!	
b. ☞ akka (regressive)		*	*
c. atta (progressive)	*!	*	

Cases of progressive assimilation in suffixes were discussed in section 3.3. To account for these, I adopt the view (e.g. McCarthy and Prince 1995) that faithfulness constraints for stem material are always ranked above faithfulness constraints for affixes.

The ranking in (26b) produces cases in which both progressive and regressive assimilations occur: for instance, in a hypothetical language where only coronals can be targeted, both /atka/ and /akta/ would become [akka]. I have not yet discovered any cases of this sort.[9]

An important aspect of the analysis is that contexts are defined in terms of neighbouring segments, rather than in terms of prosodic positions such as coda. As a number of authors have argued (Padgett 1995; Steriade 1999, 2000, 2001; Côté 2000), syllable or prosodic positions are not appropriate for characterising typical target positions of place assimilation or indeed of other segmental phenomena. To give an example from place assimilation, nasals that are preconsonantal but in the onset characteristically do assimilate, as in Luganda (e.g. [lu.gaː.nda]). The same is true for syllabic nasals, as in Kpelle (e.g. [m̩ bolu]; forms from Padgett 1995); these presumably occupy the nucleus, not the coda.

4.4 Trigger place

The ranking in (20d), PRES(pl(__coronal)) ≫ PRES(pl(__noncoronal)), indicates that place cues of a consonant must be preserved before coronals in preference to before noncoronals. The possible rankings of WEAKENING within this hierarchy derive the observed typology:

(28) *Possible language-specific rankings*
   a. PRES(pl(__cor)) ≫ ***WEAKENING*** ≫ PRES(pl(__noncor))
      → Only noncoronals are the trigger.
   b. ***WEAKENING*** ≫ PRES(pl(__cor)) ≫ PRES(pl(__noncor))
      → Both coronals and noncoronals are the trigger.

Ranking (28a) is attested in Latin and Korean; here is a Korean example:

(29) a. Noncoronals are triggers.

Input = /ip+ko/	PRES(pl(__cor))	WEAK	PRES(pl(__noncor))
(i) ipko		**!	
(ii) ☞ ikko		*	*

b. Coronals are not triggers.

Input = /ip+tolok/	PRES(pl(__cor))	WEAK	PRES(pl(__noncor))
(i) ☞ iptolok		**	
(ii) ittolok	*!	*	

Because the ranking PRES(pl(__noncor)) ≫ PRES(pl(__cor)) is disallowed, there is no ranking that permits assimilation to coronals only.

5 Gradient assimilation

Patterns of place assimilation fall into two major types, categorical and gradient. In categorical assimilation, loss of a target gesture is always complete. An example is found in the English prefix *in-,* in such morphophonemic alternations as i[n]+ept vs i[m]+possible. Categorical alternation is characteristically insensitive to speech rate and style.

In gradient assimilation, a residual gesture corresponding to C_1 appears on the surface. Such remnants of the target gesture have been observed in German (Kohler 1976), English (Barry 1985; Browman and Goldstein 1990; Nolan 1992; and Byrd 1994), and Korean (Jun 1996b). Gradient assimilation is characteristically sensitive to rate and style.

A complete analysis of assimilation must provide a way of treating the gradient/nongradient distinction. In the account proposed here, all gestural reduction comes from the high ranking of WEAKENING, which requires the conservation of articulatory effort. An output that displays no target gesture at all will best satisfy this constraint. To analyse gradient assimilation, we need to establish what prevents the reduction from going all the way to zero.

I claim that partial reduction is the result of an attempt to preserve small remnants of the perceptual cues for the original segment. Specifically, I propose that the PRESERVE(place) is a constraint family that can be decomposed into constraints distinguished by the amount of place-of-articulation information, that is, PRESERVE_n(place) = 'Preserve at least *n* per cent of the perceptual cues for place of articulation', where $1 \leq n \leq 100$. PRESERVE_{100}(place) requires maximum preservation of the perceptual cues for place; thus, when this constraint is undominated, a complete closure gesture is produced. In contrast, PRESERVE_1(place) . . . PRESERVE_{99}(place) require preservation of lesser degrees of perceptual cues; thus the dominance of any of these constraints guarantees the production of some kind of partially reduced closure. To implement the idea that at least some of the perceptual information may need to be preserved when the maximum preservation is not possible, the following internal ranking is proposed: PRESERVE_1(place) $\gg$ . . . PRESERVE_{50}(place) . . . $\gg$ PRESERVE_{100}(place) (cf. Hayes 1995; Boersma 1998).

WEAKENING likewise can be decomposed into continuous constraints distinguished by the amount of effort cost, that is, $\text{WEAKENING}_\text{m}$ = 'Do not produce an articulatory gesture whose effort cost is at least *m*' (cf. Kirchner 1998, this volume). Following Kirchner's assumption that 'the impetus to lenite more effortful gestures is stronger than the impetus to lenite easier gestures', the universal ranking of the decomposed constraints must be the following: WEAKENING_{1x} $\gg$ . . . $\text{WEAKENING}_{0.5x}$ . . . $\gg$ $\text{WEAKENING}_{0.1x}$ where 1*x* is the effort cost required for the production of a complete closure gesture (cf. Boersma 1998).

To show how various reduction patterns of the target segment follow from the interaction of the proposed decomposed constraints, consider ten ranked constraints selected from each of the PRESERVE and WEAKENING families: {PRES_{10}(place), PRES_{20}(place) . . . PRES_{100}(place)} and {WEAK_{1x}, $\text{WEAK}_{0.9x}$ . . . $\text{WEAK}_{0.1x}$}. I assume that the PRESERVE and WEAKENING constraints conflict in such a way that $\text{WEAK}_{1x} \leftrightarrow \text{PRES}_{100}$(place), $\textit{WEAK}_{0.9x} \leftrightarrow \text{PRES}_{90}$(place) . . . $\text{WEAK}_{0.1x} \leftrightarrow \text{PRES}_{10}$(place) (where $\leftrightarrow$ indicates 'conflict with'). For instance, WEAK_{1x} prohibits the occurrence of a complete closure gesture, which would provide 100 per cent of perceptual cues for place; whereas PRES_{100}(place) requires the maximum preservation of perceptual cues, which can be achieved only by making a complete closure. Then there is no way to satisfy the two constraints at the same time.

I assume further that as the speech becomes faster and more informal, the ranking of the WEAKENING constraints relative to the PRESERVE constraints increases; whereas as the speech becomes slower and more formal, the ranking of PRESERVE constraints increases.

The table in (30) shows how the interaction of these constraints produce different reduction forms of the target segment in place assimilation depending on the speech style and speed.

(30) Ranking variation and reduction patterns

speech speed/style	← slow & formal					informal & fast →
	P_{10}	P_{10}	P_{10}		W_{1x}	W_{1x}
	P_{20}	P_{20}	P_{20}		$W_{0.9x}$	$W_{0.9x}$
	...	...	...		$W_{0.8x}$	$W_{0.8x}$
	P_{70}	P_{70}	P_{70}		$W_{0.7x}$	$W_{0.7x}$
	P_{80}	P_{80}	W_{1x}		...	...
Constraint	P_{90}	$\mathbf{P_{90}}$	$\mathbf{P_{80}}$		$\mathbf{W_{0.2x}}$	$W_{0.2x}$
Ranking	$\mathbf{P_{100}}$	$\mathbf{W_{1x}}$	$\mathbf{W_{0.9x}}$	...	$\mathbf{P_{10}}$	$\mathbf{W_{0.1x}}$
	$\mathbf{W_{1x}}$	$\mathbf{P_{100}}$	$\mathbf{P_{90}}$		$\mathbf{W_{0.1x}}$	$\mathbf{P_{10}}$
	$W_{0.9x}$	$\mathbf{W_{0.9x}}$	$\mathbf{W_{0.8x}}$		$\mathbf{P_{20}}$	P_{20}
	$W_{0.8x}$	$W_{0.8x}$	P_{100}		...	...
	$W_{0.7x}$	$W_{0.7x}$	$W_{0.7x}$		P_{70}	P_{70}
	...	...	...		P_{80}	P_{80}
	$W_{0.2x}$	$W_{0.2x}$	$W_{0.2x}$		P_{90}	P_{90}
	$W_{0.1x}$	$W_{0.1x}$	$W_{0.1x}$		P_{100}	P_{100}
reduction degree	0	10	20	...	90	100

(P_n = PRESERVE$_n$(place), W_m = WEAKENING$_m$; crucial parts are shaded)

In the table in (30), each column represents the ranking of WEAKENING and PRESERVE constraints for a given speech style/speed: the higher in the column, the higher the ranking. In the first column, which describes the most formal/slowest speech, all PRESERVE constraints outrank all WEAKENING constraints, preventing even a small amount of reduction. In the second column, for slightly more informal/faster speech, W_{1x} outranks P_{100}, thus preventing the occurrence of a full closure gesture. However, since P_{90} still outranks its conflicting partner $W_{0.9x}$, the reduction must not be so drastic that more than 10 per cent of perceptual cues would be lost. As a result, an optimal output form will show at most 10 per cent reduction of the target gesture. As the speech rate/style becomes faster and more informal, the ranking of the WEAKENING constraints relative to PRESERVE constraints becomes higher, thus increasing the degree of reduction in the output. In the last column, for the most informal/fastest speech, all WEAKENING constraints outrank all PRESERVE constraints, thus producing zero closure. Consequently, in addition to forms with full and completely reduced closures, the proposed mechanism may produce various semi-reduced forms, which can be observed in gradient assimilation.

In summary, the proposed analysis incorporates gradient assimilation by expanding the PRESERVE and WEAKENING constraint families to cover multiple, quantitative values. Categorical assimilation results when the families are completely non-overlapped in the ranking, with all WEAKENING constraints dominating all PRESERVE constraints.

6 Conclusion

The two most crucial aspects of the present study of place assimilation are as follows. First, the analysis can deal with both language-specific and universal patterns. This is made possible by the use of Optimality Theory: language-specific assimilation patterns result from language-specific constraint rankings, while the scope of cross-linguistic variability is limited by universal rankings. Second, I have appealed to phonetic research to give the analysis a non-arbitrary basis: the proposed constraints and their universal rankings are determined by principles and properties that are supported empirically by research in speech perception and production.

References

Bailey, Charles-James N. (1970). Toward specifying constraints on phonological metathesis. *LI* 1: 347–9.

Barry, Martin (1985). A palatographic study of connected speech processes. *Cambridge Papers in Phonetic and Experimental Linguistics* 4: 1–16.

Beckman, Jill N. (1997). *Positional faithfulness*. PhD dissertation, University of Massachusetts, Amherst.

Bladon, Anthony (1986). Phonetics for hearers. In Graham McGregor (ed.), *Language for hearers*. Oxford: Pergamon Press, 1–24.

Bobaljik, Jonathan David (1996). Assimilation in the Inuit languages and the place of the uvular nasal. *IJAL* 62.4: 323–50.

Boersma, Paul (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Borden, Gloria J., Katherine S. Harris, and Lawrence J. Raphael (1994). *Speech science primer: Physiology, acoustics, and perception of speech*. 3rd edn, Philadelphia: Williams & Wilkins.

Browman, Catherine P. and Louis M. Goldstein (1990). Tiers in articulatory phonology, with some implications for casual speech. In John Kingston and Mary E. Beckman (eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*. Cambridge: Cambridge University Press, 341–76.

Byrd, Dani (1992). Perception of assimilation in consonant clusters: A gestural model. *Phonetica* 49: 1–24.

(1994). *Articulatory timing in English consonant sequences*. PhD dissertation, University of California, Los Angeles.

Casali, Roderic (1996). *Resolving hiatus*. PhD dissertation, University of California, Los Angeles.

Cho, Young-mee Yu (1990). *Parameters of consonantal assimilation*. PhD dissertation, Stanford University.

Côté, Marie-Hélène (2000). *Consonant cluster phonotactics: A perceptual approach.* PhD dissertation, MIT.

Dorais, Louis-Jacques (1986). Inuktitut surface phonology: A trans-dialectal survey. *IJAL* 52.1: 20–53.

Flemming, Edward (1995). *Auditory representations in phonology*. PhD dissertation, University of California, Los Angeles.

Fujimura, O., M. J. Macchi and L. A. Streeter (1978). Perception of stop consonants with conflicting transitional cues: A cross-linguistic study. *Language and Speech* 21: 337–46.

Hardcastle, W. J. and A. Marchal (eds.) (1990). *Speech production and speech modeling*. Dordrecht: Kluwer.

Hayes, Bruce (1995). A phonetically-driven, optimality-theoretic account of post-nasal voicing. Handout presented at Tilburg Derivationality Residue Conference. Rutgers Optimality Archive 93, http://roa.rutgers.edu/.

Hudson, G. (1980). Automatic alternations in non-transformational phonology. *Lg* 56: 94–125.

Hura, Susan, Björn Lindblom, and Randy Diehl (1992). On the role of perception in shaping phonological assimilation rules. *Language and speech* 35.1–2: 59–72.

Jakobson, R., G. Fant, and M. Halle (1963). *Preliminaries to speech analysis*. Cambridge, MA: MIT Press.

Johnson, Keith (1997). *Acoustic and auditory phonetics*. Oxford: Blackwell.

Jun, Jongho (1995). *Perceptual and articulatory factors in place assimilation: An optimality theoretic approach*. PhD dissertation, University of California, Los Angeles.

(1996a). Place assimilation as the result of conflicting perceptual and articulatory constraints. *WCCFL* 14: 221–37.

(1996b). Place assimilation is not the result of gestural overlap: Evidence from Korean and English. *Phonology* 13: 377–407.

Kang, Yoonjung (1999). The phonetic basis of coronal phonotactics. *WCCFL* 18: 232–45.

Kim-Renaud, Young-Key (1986). *Studies in Korean linguistics*. Seoul: Hanshin.

Kingston, John and Mary E. Beckman (eds.) (1990). *Papers in laboratory phonology I: Between the grammar and physics of speech*. Cambridge: Cambridge University Press.

Kiparsky, Paul (1985). Some consequences of lexical phonology. *Phonology Yearbook* 2: 85–138.

Kirchner, Robert (1998). *An effort-based approach to consonant lenition*. PhD dissertation, University of California, Los Angeles. Rutgers Optimality Archive 276, http://roa.rutgers.edu.

Kohler, K. J. (1976). Die Instabilität wortfinaler Alveolarplosive im Deutschen: Eine elektropalatographische Untersuchung. *Phonetica* 33: 1–30.

(1990). Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In W. J. Hardcastle and A. Marchal (eds.), *Speech production and speech modeling.* Dordrecht: Kluwer, 69–92.

(1991). The phonetics/phonology issue in the study of articulatory reduction. *Phonetica* 48: 180–92.

(1992). Gestural reorganisation in connected speech: A functional viewpoint on ‘articulatory phonology’. *Phonetica* 49: 205–11.

Krueger, John R. (1962). *Yakut manual*. Bloomington: Indiana University.

Kuehn, D. P. and M. L. Moll (1976). A cineradiographic study of VC and CV articulatory velocities. *JPh* 4: 303–20.

Kurowski, Kathleen and Sheila Blumstein (1984). Perceptual integration of the murmur and formant transitions for place of articulation in nasal consonants. *JASA* 76: 383–90.

Kühner, Raphael and Friedrich Holzweissig (1966). *Ausführliche Grammatik der lateinischen Sprache*. Hannover: Verlag Hahnsche Buchhandlung.

Lindblom, B. (1983). Economy of speech gestures. In Peter F. MacNeilage (ed.), *The production of speech*. New York: Springer Verlag, 217–45.

(1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle and A. Marchal (eds.), *Speech production and speech modeling*. Dordrecht: Kluwer, 403–39.

Malécot, André (1956). Acoustic cues for nasal consonants: An experimental study involving tape-splicing technique. *Lg* 32: 274–84.

(1958). The role of releases: In the identification of released final stops. *Lg* 34: 370–80.

Manuel, S. Y. (1991). Some phonetic bases for the relative malleability of syllable-final versus syllable-initial consonants. *Proceedings of the XIIth International Congress of Phonetic Sciences* 5: 118–21.

Mascaró, Joan (1978). *Catalan phonology and the phonological cycle*. Bloomington: Indiana University Linguistics Club.

McCarthy, John and Alan Prince (1995). Faithfulness and reduplicative identity. In Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk (eds.), *University of Massachusetts occasional papers in linguistics 18: Papers in Optimality Theory*. Amherst: GLSA, 249–384.

Mohanan, K. P. (1993). Fields of attraction in phonology. In John Goldsmith (ed.), *The last phonological rule: Reflections on constraints and derivations*. Chicago: University of Chicago Press, 61–116.

Myers, Scott (1997). Expressing phonetic naturalness in phonology. In Iggy Roca (ed.), *Derivations and constraints in phonology*. Oxford: Oxford University Press, 125–52.

Nolan, F. (1992). The descriptive role of segments: Evidence from assimilation. In G. J. Docherty and D. R. Ladd (eds.), *Papers in laboratory phonology II: Gesture, segment, prosody*. Cambridge: Cambridge University Press, 261–80.

Ohala, John J. (1990). The phonetics and phonology of aspects of assimilation. In John Kingston and Mary E. Beckman (eds.), *Papers in laboratory phonology I: Between the grammar and the physics of speech.* Cambridge: Cambridge University Press, 258–75.

(1992). Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In *Papers from the Parasession on the Syllable*. Chicago: Chicago Linguistic Society, 319–38.

Öhman, S. E. G. (1967). Numerical model of coarticulation. *JASA* 41: 310–20.

Padgett, Jaye (1995). Partial class behavior and nasal place assimilation. Rutgers Optimality Archive 113, http://roa.rutgers.edu/.

Paradis, Carole and Jean-Francois Prunet (eds.) (1991). *Phonetics and phonology: The special status of coronals: Internal and external evidence*. Orlando: Academic Press.

Prince, A. and P. Smolensky (1993). *Optimality Theory: Constraint interaction in generative grammar*. Rutgers Optimality Archive 537, http://roa.rutgers.edu/.

Repp, Bruno H. (1978). Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception and Psychophysics* 24: 471–85.

Sapir, David J. (1965). *A grammar of Diola-Fogny*. Cambridge: Cambridge University Press.

Shryock, Aaron (1993). Assimilation in Musey. Ms., University of California, Los Angeles.

Silverman, Daniel (1995). *Phasing and recoverability*. PhD dissertation, University of California, Los Angeles.

Steriade, Donca (1993). Neutralization and the expression of contrast. Paper presented at *NELS* 24, University of Massachusetts, Amherst.

(1995). *Positional neutralization*. Ms. Department of Linguistics, University of California, Los Angeles. http://www.linguistics.ucla.edu/people/steriade/papers/ Positional Neutralization.pdf.

(1997). Phonetics in phonology. Handout presented at SICOL 1997, Seoul, Korea.

(1999). Phonetics in phonology: The case of laryngeal neutralization. In M. Gordon (ed.), (UCLA Working Papers in Linguistics 2) Los Angeles: Department of Linguistics, UCLA, *Papers in Phonology 3*. http://www.linguistics.ucla.edu/ people/steriade/papers/phonetics in phonology.pdf. 25–146.

(2000). The phonology of perceptibility effects: The P-map and its consequences for constraint organisation. Ms., University of California, Los Angeles.

(2001). Directional asymmetries in place assimilation: A perceptual account. In Elizabeth Hume and Keith Johnson (eds.), *The role of speech perception in phonology*. New York: Academic Press, 219–50.

Stevens, Kenneth N. (1989). On the quantal nature of speech. *JPhon* 17: 3–45.

Trommelen, M. (1983). *The syllable in Dutch*. Dordrecht: Foris.

Webb, Charlotte (1982). A constraint on progressive consonantal assimilation. *Linguistics* 20: 309–21.

Winitz, H., M. E. Scheib, and J. A. Reeds (1972). Identification of stops and vowels for the burst portion of /p,t,k/ isolated from conversational speech. *JASA* 51: 1309–17.

Wright, Richard (1996). *Consonant clusters and cue preservation in Tsou*. PhD dissertation, University of California, Los Angeles.

Wright, Richard, Stefan Frisch, and David B. Pisoni (1996). Speech perception. In *Research on Spoken Language Processing, Progressive Report* 21, Indiana University.

Zsiga, E. (1992). Acoustic evidence for gestural overlap in consonant sequences. *Haskins Laboratories Report on Speech Research*, SR-111/112: 1–20.

Notes

* This paper has greatly benefited from the advice and comments of Bruce Hayes and Donca Steriade. I am also grateful to Robert Kirchner and Jie Zhang for their helpful comments.

1. See Jun 1995 for the discussion of the typological data and their original sources.
2. In Catalan, among continuants, only /l/ may assimilate in place to a following palatal or velar but it cannot to a following labial (Mascaró 1978). This limited case is not specified in (8) for simplicity's sake.
3. The assimilation of labials in the relevant dialects is a well-established diachronic fact; however, because of the lack of labial-final stems (Bobaljik 1996: 323), it is harder to show synchronically. Preconsonantal labials are phonologically illegal except before another labial, a gap that could easily be accounted for by assuming assimilation.
4. Broad phonetic transcriptions are employed for these examples. For instance, actual phonetic forms are subject to the regular process of post-obstruent fortition, whereby lenis obstruents become fortis after an obstruent (Kim-Renaud 1986).
5. See section 5 below for gradient assimilation, in which the loss can be partial.
6. Kirchner (1998, this volume) and Boersma (1998) adopt the same type of constraint.
7. In assessing violations of the WEAKENING, short and long gestures are not differentiated.
8. For the analysis shown in (15), a perceptual definition of faithfulness provided in (14) does not seem to be necessary; a conventional articulatory definition will be sufficient. However, in section 5, the role of the perceptual faithfulness will be evident in the analysis of gradient place assimilation.
9. If the gap is genuine, it cannot be explained by a universal ranking PRESERVE(pl(__V)) $\gg$ WEAKENING, since this ranking would exclude all cases of progressive assimilation in suffixes (section 3.3).

4 The typology of rounding harmony

Abigail R. Kaun

1 Rounding harmony typology

Rounding harmony is a phonological process whereby certain vowels surface as rounded under the influence of a neighbouring rounded vowel. What is striking about rounding harmony is the fact that the simplest possible statement – 'a vowel must be rounded when preceded by/followed by a rounded vowel' – fails to characterise the great majority of rounding harmony systems. In most cases, conditions referring to tongue body position (height and/or backness) are imposed on either the triggering element, the target, or both. I argue that this interaction among vowel features renders traditional rule-based accounts of the typological patterns nonexplanatory. Within a constraint-based framework such as Optimality Theory (Prince and Smolensky 1993), however, the interaction of rounding with these other phonological dimensions can be modelled in a straightforward manner that allows for the characterisation of all attested rounding harmony patterns, while making falsifiable predictions regarding the logically possible but cross-linguistically unattested patterns. Central to my analysis is the claim that phonological systems are organised around principles of articulation and perception. These principles are encoded in the formal grammar as Optimality-Theoretic constraints.

The goals of the chapter are as follows: (1) to exemplify the range of attested rounding harmony patterns, (2) to identify the perceptual and articulatory principles that give rise to these patterns, and (3) to propose a formal model to characterise the role of these principles in grammar. In the final section, I outline the results of a recent experiment involving loanwords in Turkish. The experimental results indicate that the model of rounding harmony developed here, while motivated by evidence from typology, is also an appropriate model of individual grammars.

For data, I rely on a number of earlier typological studies of rounding harmony (Bogoroditskij 1953; Korn 1969; Kaun 1994, 1995), as well as sources describing rounding harmony in Mongolian (Svantesson 1985), Tungusic (especially Li 1996), and a variety of non-Altaic languages. In total, thirty-three languages were surveyed.

Table 4.1. *Height-conditioned typology*

	Same-height harmony		Cross-height harmony		
Type	high target	nonhigh target	high target	nonhigh target	Sample language, number of languages
1	✓	✓	✓	✓	Kirgiz, 2
2	✓	✓			Yokuts, 1
3	✓				Hixkaryana, 3
4		✓			Khalkha Mongolian, 8
5	✓		✓		Nawuri, 7
6	✓	✓	✓		Yakut, 2

What all of these languages have in common is that rounding harmony is nearly always constrained so as to apply only when certain conditions are met: conditions that refer to phonological dimensions other than lip rounding.

1.1 Rounding harmony and height

Both vowel height and backness play a role in rounding harmony. I will begin by addressing harmony systems in which it is only the height of the trigger and/or target that determines whether harmony will apply. Table 4.1 gives the typology of height-sensitive systems. For each type, a representative language is listed, followed by the number of languages in the sample instantiating that particular pattern. A check mark indicates that rounding harmony is observed in the relevant configuration. In 'same-height' harmony the trigger and target agree in height, whereas in 'cross-height' harmony the trigger and target disagree in height.

In the Type 1 system, rounding harmony is triggered by any rounded vowel and targets any vowel. In the majority of attested rounding harmony systems, however, harmony is constrained by vowel height. Same-height harmony is more frequently observed than cross-height harmony (nine types vs four), and where cross-height harmony arises, it typically involves a nonhigh trigger and a high target, that is, *oCu/öCü* sequences rather than *uCo/üCö* sequences, where the first vowel represents the trigger. From this, we may conclude that the effects of height on rounding harmony are limited to the adoption (within particular languages) of one or more of the following principles:

(1) i. The trigger must be nonhigh.
ii. The target must be high.
iii. The trigger and target must agree in height.

In the Yokuts type (Type 2), rounding harmony occurs as long as condition (iii) is met, that is, as long as the trigger and target agree in height. In the Type 3 languages (Hixkaryana, Kachin Khakass, and Tsou), both conditions (ii) and (iii) must be met in order for harmony to apply; hence only high vowels trigger harmony, and only high vowels undergo it. Type 4, widely observed in the Mongolian and Tungusic languages families, requires both (i) and (iii), so that harmony is observed only among nonhigh vowels. Harmony in Type 5 applies as long as (ii) is met. Finally, harmony applies in the Type 6 system as long as either (ii) or (iii) is met, so that the only configuration where harmony is blocked is where the trigger is high and the target is nonhigh, ruling out the widely dispreferred *uCo/üCö* sequences.

Let us examine data from two of these types by way of illustration. The most general pattern, that which imposes none of the height restrictions from (3), is Type 1. The dialect of Kirgiz reported in Comrie (1981) instantiates this system. Kirgiz has a fully symmetrical vowel inventory of sixteen vowels, classified as front/back, rounded/nonrounded, high/nonhigh, and long/short; thus /i, iː, ü, üː, ı, ıː, u, uː, e, eː, ö, öː, a, aː o, oː/. The quality of vowels in noninitial syllables is to a large extent predictable on the basis of the quality of the vowel occurring in the first syllable: vowels agree with the vowel of the initial syllable in both backness and rounding. The effects of backness harmony and rounding harmony are most readily apparent in suffixal vowel alternations, although the vowels of native polysyllabic roots display the same distributional patterns.

Let us consider first the ordinative suffix, which has the surface variants {*-(i)ntʃi*, *-(ı)ntʃı*, *-(u)ntʃu*, *-(ü)ntʃü*}. The vowels of this suffix are in all instances high; however, their rounding and backness is variable. When the root contains front unrounded vowels the alternant *-(i)ntʃi* surfaces. Following back unrounded vowels the suffix contains back unrounded vowels and the alternant *-(ı)ntʃı* surfaces: thus *bir* 'one', *bir-intʃi* 'one-ord.'; *beʃ* ~ *beʃ-intʃi* 'five'; *altı* ~ *altı-ntʃı* 'six'; *ʒıjırmı* ~ *ʒıjırmı-ntʃı* 'twenty'. The vowels of this suffix are rounded following roots containing rounded vowels, as in the following examples: *ütʃ* ~ *ütʃ-üntʃü* 'three', *tört* ~ *tört-üntʃü* 'four', *toguz* ~ *toguz-untʃu* 'nine', *on* ~ *on-untʃu* 'ten'.

To demonstrate the effects of backness and rounding harmony in nonhigh vowels, consider the ablative suffix, which has the surface variants {*-t/den*, *-t/dan*, *-t/dön*, *-t/don*}. The nonhigh vowel of this suffix also agrees in both backness and rounding with the vowels of the root. (Additionally, consonants agree in voicing with a preceding consonant): *iʃ* 'work' ~ *iʃ-ten* 'work-ABL', *et* ~ *et-ten* 'meat', *ʒıl* ~ *ʒıl-dan* 'year', *alma* ~ *alma-dan* 'apple', *üj* ~ *üj-dön* 'house', *köl* ~ *köl-dön* 'lake', *tuz* ~ *tuz-don* 'salt', *tokoj* ~ *tokoj-don* 'forest'. This effect is pervasive across sequences of suffixes, as illustrated in polymorphemic words such as *köz-ün-dö* 'eye-poss.-loc.'

This pattern, while simple and symmetric, is in fact very unusual. More typical is the familiar pattern instantiated in Turkish (Type 5), where vowel height plays a role in determining the application of harmony. The vowels of

Table 4.2. *Backness-sensitive typology (partial)*

Type	Front trigger	Back trigger: Same-height high target	Back trigger: Same-height nonhigh target	Back trigger: Cross-height high target	Back trigger: Cross-height nonhigh target	Sample language, number of languages
7	✓	✓	✓	✓		Altai, 2
8	✓	✓		✓		Karakalpak, 3
9	✓	✓				Kyzyl Khakass, 1

Turkish, like those of Kirgiz, contrast for frontness, rounding and height: /i, ü, ı, u, e, ö, a, o/. Also like Kirgiz, suffix vowels undergo both backness and rounding harmony. While backness harmony functions entirely independently of vowel height, height plays a role in rounding harmony. High suffixes, such as the first singular possessive, undergoes rounding harmony, as in *ip* 'rope', *ip-im* 'rope-1.sg.poss.', *süt* ∼ *süt-üm* 'milk', *ev* ∼ *ev-im* 'house, *tʃöp* ∼ *tʃöp-üm* 'garbage', *kız* ∼ *kız-ım* 'girl', *buz* ∼ *buz-um* 'ice', *at* ∼ *at-ım* 'horse', *gol* ∼ *gol-um* '(football) goal'.

Nonhigh vowel suffixes, such as the dative suffix, do not undergo rounding harmony: *ip* 'rope' ∼ *ip-e* 'rope-dat.', *süt* ∼ *süt-e* (**süt-ö*) 'milk-', *ev* ∼ *ev-e* 'house', *tʃöp* 'garbage' ∼ *tʃöp-e* (**tʃöp-ö*) 'garbage', *kız* ∼ *kız-a* 'girl', *buz* ∼ *buz-a* (**buz-o*) 'ice', *at* ∼ *at-a* 'horse', *gol* ∼ *gol-a* (**gol-o*) '(football) goal'.

Data from languages exhibiting the remaining height-sensitive types are presented in Kaun 1995.

1.2 Rounding harmony and backness

In addition to vowel height, backness also emerges as a conditioning factor in the typology of rounding harmony. In a number of Turkic languages, height conditions of the sort just described are imposed when the trigger is a back vowel, but are suspended when the trigger is front, thus yielding across-the-board harmony among front vowels. Adding these cases to our typology yields three additional types, given in table 4.2. Check marks in the first column indicate that when the trigger is a front vowel, no height restrictions are imposed. Each of these types can be thought of as a version of one of the height-sensitive types from table 4.1. For example, Type 7 can be thought of as a variant of Type 6 with the stipulation that all height conditions are suspended when the trigger is front. Similarly, Type 8 is a variant of Type 2, in which harmony triggered by back vowels may only target high vowels. Type 9 can be thought of as a variant

Table 4.3. *Backness-sensitive typology (continued): back targets preferred*

		Back trigger				
		Same-height		Cross-height		
Type	Front trigger	high target	nonhigh target	high target	nonhigh target	Sample language, number of languages
10	(*)	✓	✓	✓		Sibe (roots), 1
11	(*)	✓		✓		Sibe (affixes), 1
12	(*)		✓			Shuluun Höh, 1

of Type 3 in that the operative height constraints are the requirement that the trigger and target agree in height, and that the target must be high.

In Shuluun Höh (Svantesson 1985), like other Mongolian languages, rounding harmony is observed as long as the trigger and target agree in height and the trigger is nonhigh. Shuluun Höh imposes the additional requirement that the target of harmony must be [+back]. Rounding harmony in Sibe, a Tungusic language of China (Li 1996), imposes this condition as well, and will be discussed in greater detail in section 3. Shuluun Höh and Sibe provide three additional types (Sibe itself exhibiting two: one within roots and one across morpheme boundaries).

Through extensive fieldwork on Turkic languages in Siberia, Harrison (2001) has uncovered several rounding harmony systems that appear to be in transition. In Tofà, the trigger of rounding harmony must be a front vowel, and the target must be high. Some speakers exhibit harmony consistently, others sporadically, and still others not at all. The same type of variability is evidenced in Tuha, where the target of rounding harmony is always high, and harmony is consistently applied when the trigger and target agree in height, yielding *uCu* and *üCü* sequences. When the trigger and target are distinct in height, harmony is applied only variably. Thus in Tuha, one hears harmonic *oCu/öCü* sequences as well as nonharmonic *oCı/öCi* sequences. In Altai Tuvan, rounding harmony may occur in any trigger/target pair, but is apparently obligatory only when the trigger is the nonhigh front vowel [ö].

I will treat Altai Tuvan as essentially an exemplar of Type 1, in which harmony applies without regard to the backness or height of the trigger and target, though taking note of this case as an additional instantiation of the preference for nonhigh and front triggers, as it is only the nonhigh front vowel [ö] that obligatorily triggers harmony. Tuha can be thought of as an exemplar of Type 5, exemplified most familiarly in Standard Turkish; however, this language provides an additional case in which same-height harmony is preferred over cross-height harmony. Tofà constitutes a new type (labelled Type 13 in

Table 4.4. *Backness-sensitive typology (completed)*

		Front trigger				
		Same-height		Cross-height		
Type	Back trigger	high target	nonhigh target	high target	nonhigh target	Sample language, number of languages
13		✓		✓		Tofà, 1

table 4.4), in which only front vowels trigger harmony and only high vowels are available as rounding harmony targets:

1.3 Summary

To summarise, all of the observed types can be characterised by means of the five height and backness conditions given in (2):

(2) Conditions favouring rounding harmony
   i. The trigger is nonhigh.
   ii. The trigger is front.
   iii. The target is high.
   iv. The target is back.
   v. The trigger and target agree in height.

From this list we should take note of the fact that the preferred rounding harmony *targets*, namely the high vowels and the back vowels, are the typologically dispreferred rounding harmony *triggers*. I will show that this semi-complementary distribution is not an accident.

In the remainder of this chapter, I discuss the phonetic underpinnings of this typology and present a formal account that incorporates these general phonetic principles as its foundation. Section 2 lays out the phonetic properties of rounded vowels, the key observation being that when lip rounding is combined with diverse tongue shapes, its articulatory, acoustic, and perceptual manifestations are nonuniform. In this section, I also propose that vowel harmony is perceptually driven and that perceptual differences associated with vowels of differing tongue shapes are the source of the observed trigger/target asymmetries. In section 3, I posit Optimality-Theoretic constraints to account for the typology and demonstrate that the proposed analysis provides a good fit with typological data. In the final section, I describe the results of an experiment on harmony in loanwords in Turkish, and suggest that those experimental results provide support for the analysis not only as a metagrammar for the typology as a whole, but also as a model of individual grammars.

2 Phonetic underpinnings of the rounding harmony typology

We turn now to the question of phonetic grounding. The crucial questions are: Why should nonhigh vowels be preferred as triggers, and high vowels as targets? Why is harmony preferred when the trigger and target agree in height? And, more generally, why is rounding harmony sensitive to vowel features other than rounding, namely to height and backness? In seeking an answer to these questions, I will first discuss the articulatory, acoustic, and perceptual properties of rounded vowels.

2.1 Articulation

Linker (1982) studied labial activity in vowels for five genetically diverse languages: English, Cantonese, Finnish, French, and Swedish, with the goal of identifying the linguistically significant parameters of lip position. With the exception of English, both back and front rounded vowels are present in the inventories of all of the languages studied, and all of the languages exhibit both high and nonhigh rounded vowels. Thus, the data set allowed for the comparison of labial activity in back versus front rounded vowels, as well as a comparison of labial activity in high versus nonhigh rounded vowels.

Linker's study included measurements of twenty-four distinct dimensions taken from still photographs of the side and front view of the mouth.[1] Using the factor analysis algorithm PARAFAC (Harshman 1970; Harshman, Ladefoged, and Goldstein 1977; Harshman and Berenbaum 1980), Linker identified the articulatory dimensions of lip position that appear to be relevant for distinguishing vowels within each of the languages studied. These dimensions involved horizontal opening, vertical opening, and lip protrusion, or some combination thereof. Additionally, using CANON (Goldstein n.d.), Linker isolated a set of canonical factors of lip position relevant to all of the languages studied.

Her results yielded two articulatory factors associated with lip rounding: one based largely on horizontal opening, and the second based on vertical opening and lip protrusion. While all of the rounded vowels that were studied clustered on the higher end of both scales relative to the unrounded vowels, there were systematic differences of degree among the rounded vowels themselves. For all languages studied, the high rounded vowels fell on the higher end of the scales relative to their nonhigh counterparts, thus the high vowels were in a sense 'more rounded' than the nonhigh vowels. The back vowels tended to fall on the higher end of the scales relative to their front counterparts, the notable exceptions being the nonhigh vowels of Cantonese and French. In those cases, the front [ö]-type vowel was slightly more protruded than the back [o]-type vowel. Aside from this exception, Linker's results indicate that the magnitude of lip rounding is relatively greater for high vowels than for nonhigh vowels and, though with less uniformity, for back vowels than for front vowels.

2.2 Acoustics

This generalisation is consistent with the acoustic patterning associated with rounded vowels. Stevens (1998: 294) makes the following observation: '[ɪ]increased prominence of the principal spectral peak, together with a lowered centre of gravity of the peak, can be considered as primary acoustic correlates of the rounding feature'. When rounded vowels are compared in terms of both of these acoustic properties, it is evident that the acoustic consequences of adding lip rounding to back vowels is greater than that achieved by adding lip rounding to front vowels. It is also evident that lip rounding has more dramatic acoustic consequences for high vowels than for mid vowels (see Stevens' figure 6.22 (1998: 293)).

2.3 Perception

The difference in lip activity associated with high vowels vs nonhigh vowels and back vowels vs front vowels (Linker), alongside the observed acoustic differences (Stevens), apparently gives rise to perceptual differences among rounded vowels as well. Terbeek (1977) presents an investigation of perceptual distances in the vowel space. In his study, Terbeek examined the perceptual distance among ten monophthongs {i, ü, e, ö, ı, u, a, o, æ, ɚ} with the goal of identifying the perceptual attributes according to which listeners perceive differences among vowels.

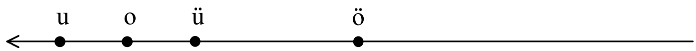
Speakers of English, German, Thai, Turkish, and Swedish served as subjects. For each of these subjects, some but not all of the monophthongs were similar to vowels occurring in the listener's native language. The data consisted of triadic comparisons of the test vowels in the context [bəb__] and the task was to determine which of the three stimuli sounded the most distinct from the others. From the responses collected, dissimilarity matrices were constructed, which then were submitted to a PARAFAC factor analysis algorithm (Harshman 1970; Harshman, Ladefoged, and Goldstein 1977; Harshman and Berenbaum 1980). Terbeek's PARAFAC analysis yielded a six-dimensional solution, indicating that six factors are relevant to the identification of vowels within a multidimensional space. These six dimensions correlate more or less with the standard phonological oppositions shown in (3):

(3) Dimensions of vowel identification (Terbeek 1977)

Dimension 1:	back vs nonback (1)
Dimension 2:	back vs nonback (2)[2]
Dimension 3:	low vs nonlow
Dimension 4:	high vs nonhigh
Dimension 5:	round vs nonround
Dimension 6:	peripheral vs central

The results of Terbeek's investigation indicate that along the round vs nonround continuum, the rounded vowels are arranged as shown schematically in (4):

(4) The round vs nonround continuum (Terbeek 1977)



This arrangement indicates that nonhigh vowels and front vowels are perceived as relatively less rounded: the high vowels lie on the higher end of the scale relative to the nonhigh vowels, and the back vowels lie on the higher end of the scale relative to the front vowels.

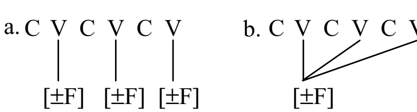
2.4 Explaining the trigger/target asymmetries

We have seen from Linker's and Terbeek's studies, as well as Stevens' discussion of the tongue shape-dependent acoustic consequences of lip rounding, that the manifestations of lip rounding are not the same for all rounded vowels. I will argue that these phonetic differences amongst rounded vowels gives rise to the differences exhibited in their phonological patterning.

2.5 What makes a good trigger?

From the preceding discussion, it appears that those vowels for which the addition of lip rounding induces a relatively weak acoustic effect are the typologically preferred triggers of harmony, whereas those for which the acoustic effect of lip rounding is relatively dramatic are the typologically preferred rounding harmony targets. This can be understood under the assumption that vowel harmony is essentially a perceptually driven phenomenon, an approach first put forth by Suomi (1983). Suomi proposes that harmony is best regarded as a means by which to enhance the probability that a given contrast or set of contrasts will be accurately perceived by the hearer. The key idea is that harmony gives rise to an extension of the temporal span associated with some perceptually vulnerable quality, represented below as [±F]. By increasing the listener's exposure to the quality in question, harmony increases the probability that the listener will accurately identify that quality. Suppose that two competing representations for a given string are available, those given in (5a) and (5b):

(5)



The decision to prefer (b) over (a) has the positive consequence that it provides the listener with increased exposure to the feature value in question. Harmony gives rise to the perceptual enhancement of the [±F] contrast by extending its duration, although it does so at the cost of reducing the number of possible words in the language.

The harmonic structure in (b) has an additional advantage over the structure in (a). Suppose the listener knows that a given feature is harmonic and thus that over some span the value of that feature will remain constant. Over that span, then, the value of [F] must be identified only once. If the identification is made early on in the string, the acoustic dimension associated with the harmonic feature need no longer be attended to, and attention may be focused on other aspects of the acoustic signal. If only a tentative identification of the harmonic feature value is made early on, additional input is available in the remainder of the string for verification. Finally, if the acoustic cues of the feature in question are somehow obscured in the early portion of the string, the feature value is still potentially recoverable from information carried in the latter portion of the string.

There is a second way in which harmony could be argued to facilitate the correct identification of the triggering vowel. It is well known that vowels exert a coarticulatory effect on neighbouring vowels. Both anticipatory and carry-over coarticulatory effects have been documented for languages such as English (Bell-Berti and Harris 1976), Russian (Purcell 1979), and Catalan (Recasens 1984). In a given V_iCV_j utterance, the articulation of V_i will typically affect that of V_j, and vice versa. It seems reasonable to assume that in VCV utterances in which the vowels are identical or similar, coarticulatory effects will be either nonexistent or fairly minor. If the goal is to maximise the perceptibility of a given vowel, then by insisting that vowels in neighbouring syllables be identical or similar to that vowel, the effects of coarticulation will be eliminated or at least reduced.

Under a perceptual approach to harmony, we would expect that the perceptual advantage of having harmony would be greater when the contrastive value is particularly vulnerable to misidentification. This perceptual motivation underlies the observed trigger preferences. Due to the relative phonetic weakness of rounding in front and nonhigh rounded vowels, harmony triggered by these vowels creates a greater perceptual advantage than harmony triggered by their phonetically more stable high and back counterparts.

To summarise, then, the proposal is that vowel harmony is a perceptually driven phenomenon that serves to prolong the duration of a given feature or quality. As such, its utility is greater when the prolonged quality is one that is particularly at risk for misidentification. Rounding harmony triggered by nonhigh vowels and front vowels (i.e. ‘bad’ rounded vowels) will thus perform a more critical function than harmony triggered by the perceptually more stable high and back rounded vowels (i.e. the ‘good’ rounded vowels). Similarly, as I will discuss below in section 2.8, the utility of harmony initiated by a prosodically short trigger will be greater than that of harmony triggered by a prosodically long trigger.

2.6 What makes a good target?

Rounding harmony is frequently blocked when its application would give rise either to a nonhigh rounded vowel or to a front rounded vowel. One might attribute this pattern to the relative markedness of these vowels, as formalised in Chomsky and Halle's marking convention #XI (1968: 405), and later in Archangeli and Pulleyblank's theory of Grounded Phonology (1994: 78). Under a markedness analysis, the claim would be something to the effect that rounding harmony rules are less highly valued when they generate marked feature combinations. Such an approach relies on the proposition that markedness is a property of individual segments, a view that is challenged by Flemming (1995, this volume), who claims that markedness should be understood as a property of contrasts, not segments.[3] He argues, for instance, that the relative rarity of front rounded vowels and back unrounded vowels within vowel inventories is not due to their inherent markedness. Rather, it is 'the contrast between front rounded and back vowels that is marked because it is less distinct than a contrast between a front unrounded vowel and a back vowel' (Flemming 1995: 26).

Under Flemming's view, there is indeed a connection between markedness and the likelihood that a particular feature will be harmonic within a given language (e.g. rounding in languages in which rounding and backness are independently contrastive, nasality in languages with nasal and oral vowels, and ATR/RTR). Rather than attributing the typological preference for unmarked harmony targets to markedness per se, I believe that a more plausible explanation is one based on the role of perceptual influences in the evolution of rounding harmony systems.

We may assume that when the acoustic consequences of adding rounding to a given tongue shape are strongest, rounding will be most reliably recovered by listeners. Those vowels that are less reliably perceived as rounded might be interpreted as not having undergone harmony, despite the speaker's production of lip rounding during the articulation of these vowels. Over successive generations of speakers, the less robust targets may be reinterpreted as nontargets and a system might emerge in which harmony targets only those vowels whose roundedness has been most reliably perceived. The claim, then, is that the trigger preferences and the target preferences share a common perceptual origin. Harmony triggered by perceptually less salient vowels will perform a greater functional service, and those vowels that are perceptually more salient will be more likely to be interpreted by listeners as having undergone harmony.

2.7 Why should triggers and targets agree in height?

I focus now on the avoidance of cross-height harmony, that is the relative rarity of harmony in configurations where the trigger and target disagree in height. I propose that this pattern reflects a requirement that a given articulatory instruction, or autosegment, have a uniform execution mechanism throughout its span of association. In other words, a single autosegment should be interpreted phonetically as an instruction to achieve a single target articulatory posture. Cross-height harmony is thus avoided because the lip rounding gesture is not equivalent for high and nonhigh rounded vowels: typically, high vowels are more rounded than nonhigh, and there are also sometimes effects of backness as well. In this section I will consider some of the phonetic literature that supports this point.

Goldstein (1991), examining data from Linker (1982), concludes that the articulatory goal for rounded vowels is contact along the sides of the upper and lower lips: '[w]hat is specified is whether or not the *upper and lower lips touch along their sides*' (Goldstein 1991: 98). This single factor very clearly separates Linker's lip-activity measurements for rounded vs unrounded vowels. Rounded vowels involve side contact of the lips; unrounded vowels do not. Within the class of rounded vowels, there is a relation between jaw height (and consequently lip aperture) and amount of side contact:

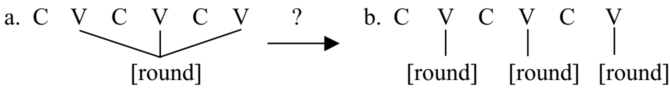
> [w]hen the lips are touching there will be an inherent relation between LA [lip aperture] and LW [lip width]. As LA decreases (everything else being equal), the length of the lips' contact region along the sides will increase, and the side-to-side width of the *opening* decreases. (Goldstein 1991:100)

A harmony span including vowels mismatched for height would necessarily involve re-adjustments in lip aperture and lip width. The avoidance of cross-height harmony can thus be construed as the avoidance of this kind of articulatory readjustments – a principle that I will label Gestural Uniformity.

Note that while there is a clear difference in side contact between high vs nonhigh rounded vowels, there is no systematic difference in length of side contact among front vs back rounded vowels (Goldstein p.c.). This may suggest that a gestural uniformity violation should be assigned to a trigger-target pair of differing heights, but not to a trigger-target pair disagreeing in backness. This is consistent with the typological patterns described here, although the prediction is difficult to test because many rounding harmony languages (the Turkic languages in particular) also exhibit backness harmony.

A putative phonological constraint that refers to the phonetic realisation of a phonologically multiply-linked structure can only be accepted if the geometry of the phonological representation is directly reflected in the phonetic outcome, as in (6a). This approach would be invalidated if one could show that the actual phonology-to-phonetics mapping is one in which the phonetics reconfigures the phonological output, conferring a rounding target on each vowel as in (6b).

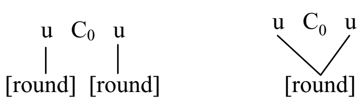
(6) Possible phonology-to-phonetics mappings



Boyce's (1988) study of coarticulation in English and Turkish provides experimental evidence that bears on this question. Boyce studied vowel-to-vowel coarticulation in English and Turkish *uC₀u* utterances. These two languages were chosen for comparison because there is good reason to believe that segmentally identical sequences may be assigned distinct phonological representations in these languages. Turkish, as a rounding harmony language, arguably represents uC_0u sequences as containing a single [round] autosegment multiply linked to both vocalic positions. English, which lacks rounding harmony, would be expected to represent the same sequence with two independent [round] specifications:

(7) Hypothesised phonological representation of uC_0u sequences for English and Turkish

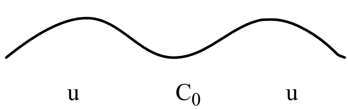
a. English b. Turkish



The question investigated by Boyce was whether the distinct phonological representations shown in (7) correspond to distinct articulatory patterns.

The English articulatory pattern, based upon measurements of lip activity and position, yielded a 'trough'-like pattern, shown schematically in (8). The tracing represents lip protrusion (especially of the lower lip):

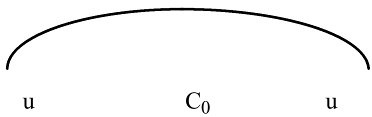
(8) English 'trough' pattern



As indicated, the lips attained a position of protrusion in the articulation of the first rounded vowel, then receded during the articulation of the consonantal sequence, then once again attained a position of protrusion for the second rounded vowel.

The Turkish articulatory pattern was qualitatively different. The results obtained by Boyce showed a ‘plateau’-like pattern in the articulation of uC_0u sequences by Turkish speakers. This is shown schematically in (9):

(9) Turkish ‘Plateau’ pattern



In the Turkish articulation, as shown, the lips attained a position of protrusion during the articulation of the first rounded vowel and remained protruded throughout the utterance.

These experimental findings suggest that whereas the English speakers executed two lip rounding movements, the Turkish speakers executed only one, indicating that the distinct phonological representations appropriate for English and Turkish give rise to distinct phonetic behaviour. If the Turkish pattern is representative of harmony languages in general, then we have reason to reject the remapping in (6b) and conclude that a single [round] autosegment in the phonology corresponds to a single lip rounding gesture in the phonetics. Gestural uniformity regulates the production of such multiply linked structures, dictating that the execution of a single autosegment or articulatory instruction should be achievable by the assumption of a uniform articulatory posture.

2.8 *Length-based trigger asymmetries*

If the tendency of front and low vowels to trigger harmony has a perceptual origin, then we would expect other factors that impede perception of vowel quality likewise to be involved in the typology of harmony triggers. One such factor, discussed inter alia by Crosswhite (this volume), is length: the quality of shorter vowels is harder to perceive. If a prosodically short vowel is more prone to misidentification than the corresponding long vowel, the perceptual account of harmony would predict that, in cases in which the length of the trigger is relevant to the applicability of harmony, short vowels should be the preferred triggers. In fact, there is evidence that vowel length can play a role in rounding harmony in just this way.

I am aware of three cases in which short vowels trigger harmony while their long counterparts do not. As far as I know, the reverse asymmetry, in which rounding harmony is triggered by long rounded vowels but not by their short counterparts, is unattested.[4] A particularly convincing case of this pattern is the Southern Tungusic language Baiyina Orochen (Li 1996), in which rounding harmony is triggered only by short vowels. Baiyina Orochen exhibits the general Tungusic pattern of harmony in which vowels within a word agree with respect to tongue root advancement. Alongside this general tongue root harmony exists a more restrictive harmony system involving rounding. Only nonhigh vowels trigger and undergo rounding harmony, also the typical Tungusic pattern. Consider the words in (10), where progressive rounding harmony is triggered by a short vowel. Targets are underlined:

(10) Short triggers (Li 1996: 126)

a. tɔɣɔ	'fire'	c. tʃolpon	'morning star'
b. ɔpɔɔ	'rocky hillock'	d. oŋtot	'strange'

Long vowels do not trigger rounding harmony, as shown in (11):

(11) Long nontriggers (Li 1996: 126)

a. kɔɔxan	'child'	c. koorgə	'bridge'
b. ɔɔŋan	'mountain pass'	d. oodən	'velvet'

It is important to note that while long vowels do not *initiate* a rounding harmony domain, they may both undergo and, more importantly, *propagate* harmony. This is shown in (12):

(12) Long vowels propagate harmony (Li 1996: 131)

a. gɔlɔɔ-tkɔɔɪ	'log; direct.'
b. ɔpɔɔ-lɔɔ	'rocky hillock; destin.'
c. sokkoo-mɲo	'muddy (water); contem.'
d. oloo-no-tʃo-	'to cook; intent. asp.; pt.t.'

This pattern is consistent with the perceptual account of rounding harmony outlined above. If the functional advantage of harmony is to increase the span of a particular distinctive (and hence important) quality, then the fact that only short vowels initiate harmony can be attributed to their relative perceptual vulnerability. The fact that both long and short vowels may serve to propagate the harmonic feature reflects the fact that these vowels do not carry distinctive information for the feature in question, hence no length-based asymmetry is expected in their harmonic behaviour.

A similar pattern is observed in the Northern Tungusic language Evenki (Nedjalkov 1998). In Evenki, only short vowels trigger rounding harmony; long vowels do not. So when the accusative definite morpheme /–vA/ is added to 'fish', the resulting form is *goro-vo* with a rounded suffix vowel. When this suffix follows a long vowel, rounding harmony does not occur, as in the affixed form of 'tree', realised as *moo-va*, rather than **moo-vo*. Additionally, Harrison (1999a) presents a case from Tuvan, a Turkic language of Siberia, in which short vowels trigger rounding harmony while their long counterparts do not.

2.9 Summary

I have argued in this section that harmony is fundamentally a perceptually driven phenomenon that serves to prolong the duration of some contrastive quality. To summarise:

- Acoustic, articulatory, and perceptual data were introduced to support the claim that contrastive rounding is particularly subtle for both nonhigh and front vowels. These are the vowels that are typologically preferred as rounding harmony triggers.
- Nonhigh and front vowels are typologically *dispreferred* as rounding harmony targets. This dispreference was attributed to the greater salience of rounding in the preferred targets which, over time, might lead to the retention of only those vowels as legitimate grammatical targets of harmony.
- The avoidance of cross-height harmony was linked to the difference in lip postures associated with high vs nonhigh rounded vowels, by means of the proposed gestural uniformity constraint.

3 A constraint-based account

It is simple enough to characterise each of the attested rounding harmony systems within a linear, rule-based framework such as that developed in Chomsky and Halle 1968, or within an autosegmental framework such as that employed in Clements and Sezer 1982. For instance, for Type 5 languages one could posit the rule in (13), which indicates that a high vowel assimilates in rounding to a preceding rounded vowel.

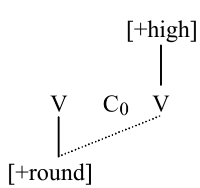
(13) SPE-style rule for Type 5

$$\begin{bmatrix}+\text{syl}\\+\text{high}\end{bmatrix} \rightarrow [+\text{round}] / \begin{bmatrix}+\text{syl}\\+\text{round}\end{bmatrix} C_0 \text{—}$$

A high vowel is rounded following a rounded vowel.

Alternatively, one might represent the vocalic features autosegmentally, and posit a rule such as that in (14):

(14) Autosegmental rule for Type 5



Rounding spreads from a rounded vowel onto a following high vowel.

Similarly, Type 2 could be represented with the rule in (15) (or some autosegmental analogue of it):

(15) $\begin{bmatrix} +\text{syl} \\ \alpha\text{high} \end{bmatrix} \rightarrow [+\text{round}] / \begin{bmatrix} +\text{syl} \\ \alpha\textit{high} \\ +\text{round} \end{bmatrix} C_0 \_\_$

A vowel is rounded if it is preceded by a rounded vowel of the same height.

These frameworks are capable of describing all of the attested rounding harmony patterns. However, given the mechanisms made available by these two formal systems, there is no way of distinguishing the attested rounding harmony rules from many formally similar but typologically unattested ones. For instance, the linear rule in (16) is no less complex than those in (13) and (15), yet it is not known to play a role in the grammar of any language.

(16) Formally similar rules

$\begin{bmatrix} +\text{syl} \\ -\text{high} \end{bmatrix} \rightarrow [+\text{round}] / \begin{bmatrix} +\text{syl} \\ +\text{round} \end{bmatrix} C_0 \_\_$

A nonhigh vowel is rounded when it is preceded by a rounded vowel.

The same point can be made with regard to the formal mechanisms of the autosegmental model. Thus, the rule-based approaches in no way limit the range of predicted rounding harmony systems.

More problematic for rule-based accounts is the fact that the formally most simple rounding harmony rule – one in which a rounded vowel triggers rounding of a neighbouring vowel without regard to height or backness – is typologically quite rare. If a maximally simple rule is assumed to be more highly valued than a more complex rule, then these models make the incorrect prediction that the harmony system characterised by the rules in (17) – i.e. Type 1 harmony – should be the most widely attested.

(17) Formally simplest, but typologically very rare rounding harmony rules

$[+\text{syl}] \rightarrow [+\text{round}] / \begin{bmatrix} +\text{syl} \\ +\text{round} \end{bmatrix} C_0 \_\_$

A vowel is rounded when preceded by a rounded vowel.

Optimality Theory (hereafter OT) replaces the rules of earlier generative models with constraints. The constraints are understood to be supplied by Universal Grammar (UG), while their relative importance or ranking is determined on a language-specific basis. OT is thus implicitly a model of linguistic typology in that a possible grammar is any ranking of the (fixed) set of universal constraints. A model such as this should allow for the characterisation of the conditions from (2) in the form of explicit grammatical statements or constraints.

3.1 Constraints

Following Smolensky (1993), I will assume that the grammatical expression of harmony is by means of Alignment (McCarthy and Prince 1993). Alignment constraints call for the coordination of domain edges. In the case of rounding harmony, the phonological domain, or feature, will be [round], while the morphological domain with which it is aligned will typically be the Prosodic Word (though see the discussion of Sibe in section 3.3).

(18) The pro-harmony constraint: ALIGN

ALIGN-L/R([RD], PRWD) *The autosegment [round] is aligned with the L/R edge of the Prosodic Word.*

One violation of this constraint is assessed for each docking site (vowel) following the last docking to which a [round] autosegment is linked.

I will also assume two further constraints, whose empirical effects tend to involve the selection of specific triggers for harmony. The alignment constraints in (19) and (20) each introduce a particular refinement to that given in (18):

(19) Nonhigh trigger

ALIGN-L/R ([RD/–HI], PRWD) *The autosegment [round], when co-occurring with [–high], is aligned with the L/R edge of the Prosodic Word.*

(20) Front trigger

ALIGN-L/R ([RD/–BA], PRWD) *The autosegment [round], when co-occurring with [–back] is aligned with the L/R edge of the Prosodic Word.*

These more specific alignment constraints, if ranked above general [round]-alignment, will give the effect of promoting harmony when the potential trigger is one of the preferred trigger-types, either a front rounded vowel or a nonhigh rounded vowel. Thus, the formal account of the observed trigger preferences lies in the absence in UG of constraints specifically promoting harmony triggered by the perceptually more salient high and back vowels.

The grammatical mechanism that suppresses harmony, that is, the constraint that overrides the alignment constraints in languages lacking rounding harmony, will be represented by means of a Faithfulness constraint from the DEP family (McCarthy and Prince 1995):

(21) The anti-harmony constraint: DEP

DEP(LINK) *The output may contain no association line absent in the input.*

To account for the vast majority of vowel harmony systems, in which harmony fails to apply in certain contexts, the grammar must also contain constraints whose effect is to block harmony when it would apply to one of the dispreferred target configurations.

First, I assume a general constraint, labelled *RoLo, which states a dispreference for nonhigh rounded vowels, thus constituting the grammatical manifestation of the articulatory and perceptual bias against lower rounded vowels discussed in section 2. This constraint, or some related grammatical principle, is instrumental in shaping vowel inventories, which commonly lack low rounded vowels. In harmony, *RoLo serves to block harmony when the output would contain a nonhigh rounded vowel not present in the input, that is, harmony targeting nonhigh vowels:[5]

(22) *RoLo[6] *Nonhigh rounded vowels are avoided.*

In addition, also in keeping with cross-linguistic vowel inventory patterns, I will posit a similar constraint against front rounded vowels. This constraint, like *RoLo, reflects the articulatory and perceptual bias against front rounded vowels:

(23) *RoFro *Front rounded vowels are avoided.*

Finally, we need a constraint forcing vowels within a rounding harmony span to share the same height specification. I argued above that this reflects a phonetic imperative to avoid the need for articulatory adjustments in the execution of a single gesture. This preference presumably has as its grammatical expression a family of gestural uniformity constraints. In rounding harmony systems, gestural uniformity will reject any instance of the autosegment [round] linked to positions with distinct height:

(24) Gestural uniformity

GestUni [(round)] *A multiply-linked [round] autosegment corresponds to a uniform mechanism for the execution of [round].*

The constraint inventory just laid out suffices to characterise all of the rounding harmony types identified in section 2. Space does not permit the full analysis to be given, but the crucial ranking needed to generate all thirteen types are listed in table 4.5. To illustrate how the constraints interact, I will present an account of one of the more complex languages surveyed, namely Sibe. This language brings together most of the relevant phenomena into a single example.

Table 4.5. *A hierarchical representation of the typology*

Type 1	Align[rd] » *others*
Type 2	Uni[rd] » Align[rd] » *others*
Type 3	Uni[rd], *RoLo » Align[rd] » *others*
Type 4	Uni[rd] » Align[rd/–hi] » Dep(Link) » *others*
Type 5	*RoLo » Align[rd] » *others*
Type 6	Align[rd/–hi] » Uni[rd] » Align[rd] » *others*
Type 7	Align[rd/–ba], Align[rd/–hi] » Uni[rd] » Align[rd] » *others*
Type 8	Align[rd/–ba] » Uni[rd] » Align[rd] » *others*
Type 9	Align[rd/–ba] » Uni[rd], *RoLo » Align[rd] » *others*
Type 10	*RoFro » Align[rd/–hi] » *RoLo » Align[rd] » *others*
Type 11	*RoFro, *RoLo » Align[rd] » *others*
Type 12	*RoFro » Uni[rd] » Align[rd/–hi] » Dep(Link) » *others*
Type 13	*RoLo » Align[rd/–ba] » Dep(Link) » *others*

3.2 Sibe

In Sibe, a south-west Tungusic language described in Li 1996, a rounding harmony pattern similar to that observed in Type 6 is exhibited within roots: as with other Type 6 languages, cross-height harmony is tolerated only with a preferred (i.e. high) target. Nonhigh vowels undergo rounding harmony as long as the trigger is also nonhigh, while high vowels undergo harmony without regard to the height of the trigger. Sibe is more complicated, however, in that while the language has front rounded vowels, only back vowels surface as the product rounding harmony. Sibe thus evidences both target preferences – the preference that targets be high and the preference that targets be back. Examples are shown in (25). In (a–b), the target is high. In (c–d), the target is nonhigh and agrees with the trigger in height. In (e–g), harmony fails to apply either because of height disagreement and a nonhigh target (e), or because the potential target is front (f–g):

(25) Root harmony in Sibe (Li 1996: 195–6)[7]

a.	fulxu	'root'
b.	ɕögu	'vegetable'
c.	ɔmɔl	'grandson'
d.	ölχɔ	'cowardly'
e.	uva (*uvɔ)	'flour'
f.	utɕi (*utɕü)	'door'
g.	ɕöbɛ (*ɕöbö)	'subsidiary'

Harmony targeting suffixes is even more restricted. As is the case within a root, while high back vowels undergo rounding harmony (26a–c), front vowels in suffixes are never targeted by rounding harmony (26d–e):

(26) Affix harmony in Sibe (Li 1996: 199–204)

a.	batur-lu	'to be heroic'
b.	bɔmbɔn-nu	'to form a cluster'
c.	gö-χu	'to hit (the target)'
d.	bu-kin (*bu-kün)	'to give'
e.	χɔsu-tɕi (*χɔsu-tɕü)	'corner'

Where suffix harmony differs from root harmony is in the fact that only high suffix vowels undergo harmony. Li's discussion clearly indicates that nonhigh vowels in suffixes never undergo harmony, even if they agree in height with the potential trigger. He does not, however, include any specific examples of a suffixal /a/ following a nonhigh rounded vowel. Some examples of suffixal /a/ are listed in (27). An invented root in (27e) demonstrates the crucial context in which harmony reportedly fails to apply:

(27) Suffixal /a/ Sibe (Li 1996: 199–204)

a.	ɔvu-ma-χɪ	'to wash-pres.prog.'
b.	vɣili-rtan	'to work-nom.'
c.	suxu-maq	'axe-instr.'
d.	is-maq	'soap-instr.'
e.	ɔs-maq (*ɔs-mɔq)	*invented*-instr.'

The Sibe pattern is instructive in a number of respects. First, different harmony patterns are exhibited within different morphological domains. We saw that nonhigh vowels are targeted by harmony only within roots. This pattern is easily captured by means of alignment, which calls for the co-incidence of phonological and morphological domain edges.

To account for the Sibe system, we may posit the constraint sub-hierarchy in (28):

(28) Constraint sub-hierarchy for Sibe[8]
*RoFro » Align([rd/–hi], Root) » *RoLo » Align([rd], PrWd) » *others*

The hierarchy in (28) may be interpreted as follows. First, the fact that front rounded vowels never arise as the result of harmony is reflected in the high ranking of *RoFro. Next, by virtue of its ranking above the constraint that requires harmony within the domain of the prosodic word, *RoLo blocks the application of harmony when it would target a low suffix vowel. The root harmony constraint (Align([rd/–hi], Root)) outranks *RoLo, however, thus allowing nonhigh vowels within roots to undergo harmony. Tableaux are shown in (29) and (30) to demonstrate how the constraints interact to characterise the Sibe rounding harmony patterns:

(29) Root-internal harmony in Sibe

	*RoFro	Align([rd/–hi], Root)	*RoLo	Align([rd], PrWd)	Uni[Rd]
☞ utçi				*	
utçü	*!				
fulxı				*!	
☞ fulxu					
çögı		*!		*	
☞ çögu					*
ɔmal		*!		*	
☞ ɔmɔl			*		
☞ uva				*	
uvɔ			*!		*

(30) Suffix harmony in Sibe

	*RoFro	Align([rd/–hi], Root)	*RoLo	Align([rd], PrWd)	Uni[Rd]
☞ bu-kin				*	
bu-kün	*!				
batur-lı				*!	
☞ batur-lu					
gö-χı				*!	
☞ gö-χu					*
☞ ɔs-maq				*	
ɔs-mɔq			*!		
☞ suxu-maq				*	
suxu-mɔq			*!		*

3.3 Factorial typology

Using the program OTSoft (Hayes et al. 2000), I generated a factorial typology based on the seven proposed constraints.[9] The input assumed a generic Turkic-style system with a symmetric eight-vowel system and backness harmony. Thus, for each of the four rounded vowels as potential harmony triggers, two trigger-target pairs were considered: one in which the potential target was high, and one in which the potential target was nonhigh. This yielded eight two-way choices:

(31) Input to the factorial typology

Trigger	Target	Harmonic		Nonharmonic
u	high	u-u	*or*	u-ı
u	nonhigh	u-o	*or*	u-a
o	high	o-u	*or*	o-ı
o	nonhigh	o-o	*or*	o-a
ü	high	ü-ü	*or*	ü-i
ü	nonhigh	ü-ö	*or*	ü-e
ö	high	ö-ü	*or*	ö-i
ö	nonhigh	ö-ö	*or*	ö-e

Given eight two-way choices, there were 2^8 possible outcomes, including the case in which no harmonic pairs were selected. There were thus $2^8 - 1$ (= 255) possible harmony systems. Of these, the proposed constraint set generated only thirty-six. All of the thirteen observed types from section 1 were generated. The predicted typology is thus reasonably small, and the cases of overgeneration in general look to be accidental gaps, not systematic ones. Thus, for instance, the pattern in which rounding applies to *uI* and *oI* sequences is the same as the familiar Type 5 pattern (cf. Turkish), except that harmony applies only amongst back vowels. Similarly, the pattern in which harmony applies to *üI* and *öA* is essentially the same-height harmony of Type 2, except that only front vowels participate.[10]

3.4 Cases of free variation

In assessing the typological validity of the constraint set proposed here, we also need to address the cases where the harmony pattern exhibits free variation. We have seen three cases here, all in section 1: Tuha, Tofà, and Altai Tuvan. As a model for free variation in Optimality Theory, I adopt the approach of Anttila (1997a, 1997b), under which certain constraints can be critically freely ranked. In such a grammar, the outcomes that arise for all permutations of the free ranked constraints are considered to be possible outputs.

In Tuha, the constraints that must be freely ranked are gestural uniformity and ALIGN[RD]: where the former dominates the latter, cases like /o I/ emerge as [o ɨ]; whereas under the opposite ranking /o I/ surfaces as [o u]. The case of Tofà can be analysed similarly, with a critical free ranking of ALIGN[RD/–BA] and DEP(LINK).

The case of Altai Tuvan is harder. Using the constraints proposed here, it is possible to achieve a rough statistical match to Harrison's (2001) observations, adopting the gradientised version of free ranking discussed in Boersma 1997 and Boersma and Hayes 2001. To achieve an exact match requires the addition of a new constraint to the system, which would require alignment for the 'best' possible triggers, namely vowels that are *both* front and low.[11] The typological consequences of such a constraint remain unexplored, however.

4 Optimality Theory as a model of typology or grammar, or both?

One central claim of OT is that all languages share a common set of (universal) constraints, but differ from one another with respect to how those constraints are ranked. This model entails that for any given grammar, a great number of ***inactive*** constraints must be present – constraints ranked too low to have any decisive effect in the determination of output forms. The evidence for these constraints is purely typological; that is, that together they form an inventory that generates the typology of existing languages. However, it is a fairly strong claim to say that the constraints are actually present in the grammars of languages in which they are inactive.

I conclude this chapter by presenting experimental results suggesting that speakers do in fact carry around inactive constraints, and that speakers recruit such constraints in the evaluation of novel phonological contexts such as those introduced in loanwords. Ross (1996) has presented a similar argument for Tagalog, as have Ringen and Heinämäki (1999) for Finnish. The experiment described here involves loanwords in Turkish.

4.1 Regressive harmony in Turkish loanwords

Yavas (1980) and Clements and Sezer (1982) demonstrate that in addition to the progressive harmony observed in native Turkish words, regressive harmony is exhibited in loanwords. Such harmony targets epenthetic vowels introduced to break up an initial consonant cluster. These epenthetic vowels are always high, but vary in backness and rounding on the basis of the quality of the first full vowel in the word. For example, backness harmony results in an epenthetic [i] in *siteno*, 'steno', in which the first full vowel is front, but an epenthetic [ı] appears in a word like *sıtar* 'star', where the full vowel is back.

The epenthesis phenomenon is of interest here because the epenthetic vowels sometimes undergo regressive rounding harmony, triggered by the stem; thus *flüt* 'flute', is realised as *fülüt*, with a rounded epenthetic vowel. Reports differ as to the conditions under which regressive harmony will apply. In the variety reported by Yavas, regressive harmony may be triggered only by high vowels, whereas in the variety described by Clements and Sezer, rounding harmony is consistently triggered by high vowels but may sometimes also be triggered by nonhigh vowels.

In Kaun 1999, I described the results of an experiment originally designed to resolve this discrepancy. Nine native speakers of Turkish[12] were presented with 107 loanwords taken from Özgüler 1989. Each of these words contained an initial consonant cluster, and the subjects were asked to indicate the appropriate quality (or qualities) of the epenthetic vowel. The task seemed to present little difficulty to the subjects; however, a great deal of subject-to-subject variation was observed.

All subjects agreed with the Yavas and Clements and Sezer patterns in one respect: High rounded vowels consistently triggered rounding harmony, as in words like *buluz* 'blouse',[13] and *fülüt* 'flute'. When the potential trigger was nonhigh, six distinct rounding harmony patterns emerged. These are presented in (32). In this chart, solid lines enclose consistent rounding harmony triggers, while dashed lines enclose optional triggers. Unenclosed vowels never trigger harmony for the pattern in question. The number of subjects instantiating each pattern is indicated in parentheses:

(32) Six rounding harmony patterns

Group A	(1)	ü	u	ö	o
Group B	(1)	ü	u	ö	o
Group C	(2)	ü	u	ö	o
Group D	(1)	ü	u	ö	o
Group E	(2)	ü	u	ö	o
Group F	(2)	ü	u	ö	o

The chart in (33) includes some of the words used in the experiment:

(33) Sample words

	'flute'	'blouse'	'flirt'	'block'
Group A	fülüt	buluz	filört	bılok
Group B	fülüt	buluz	fılört/fülört	bılok
Group C	fülüt	buluz	fülört	bılok
Group D	fülüt	buluz	fılört/fülört	bılok/bulok
Group E	fülüt	buluz	fülört	bılok/bulok
Group F	fülüt	buluz	fülört	bulok

Two generalisations can be made on the basis of these patterns. First, we note that, while subjects consistently exhibited same-height harmony (*fülüt* and *buluz*), cross-height harmony was not unanimously applied. Second, with harmony involving a height mismatch, front vowels were preferred as harmony triggers, that is, regressive harmony is observed more frequently in *flört*-type words, where the potential trigger is front, than in *blok*-type words, in which a back vowel serves as the potential trigger of regressive rounding harmony.

These patterns are not predictable on the basis of the progressive rounding harmony pattern of native Turkish words. As noted above in section 1, Standard Turkish is essentially a Type 5 language in which harmony may be triggered by any rounded vowel, but targets only high vowels. Thus, in the native pattern of progressive harmony, while vowel height does play a role in determining the applicability of harmony, it is the *target* whose height is relevant, rather than the trigger. Moreover, backness never serves to restrict the applicability of native rounding harmony.

While the observed patterns of rounding harmony share no common features with the native progressive harmony pattern, they do resemble the general cross-linguistic patterns of rounding harmony. In particular, the avoidance of cross-height harmony along with the preference for front rounded triggers are familiar from the general typology of the phenomenon. I argued in Kaun 1999 that the behaviour of the experimental subjects can be modelled as involving the recruitment of constraints that are inactive in native Turkish, but predicted to exist on the basis of the general typology of rounding harmony. These are among the constraints posited in section 3.

4.2 Discussion and conclusions

The results of the Turkish loanword experiment support the choice of OT rather than a rule-based system to account for the typology of rounding harmony. A rule-based account, which might characterise native Turkish harmony with a rule like that shown above in (13), makes no predictions regarding the realisation of rounding harmony in the loanword context. To the extent that such analyses could be said to offer any predictions with respect to regressive rounding harmony in Turkish, they would either predict that rounding harmony should not occur (because rounding harmony is progressive in native Turkish), or that rounding harmony will always proceed as in pattern F (example (32)), where any rounded vowel triggers rounding of the high epenthetic vowel. Neither of these predictions characterises the observed facts.

The loanword data also indicate that while OT allows for the characterisation of typologies in a direct and falsifiable manner (i.e. an analysis is wrong if a language can be shown to exhibit a pattern that cannot be generated by some ranking of the proposed constraint set), it also appears to be an appropriate model of individual grammars. The Turkish facts support the claim that constraints that never play a decisive role in determining surface structure are nonetheless present in grammars. Alternatively, these results could be interpreted as indicating that, when confronted with a novel phonological configuration, speakers can invent constraints 'on the spot', and that when they do so, they are guided by the same phonetic pressures that govern grammatical systems in general, in this case, the phonetic principles that underlie the phenomenon of rounding harmony.

I have argued that the account proposed here is functionally grounded. The functional underpinning of the typology can be held up as a means of understanding the evolution of rounding harmony systems. The Turkish loanword pattern further suggests that, in our effort to understand and model phonological systems, we should look upon substantive principles not just as a means of explaining the recurrence of phonological patterns post hoc, but as fundamental components of grammar, accessed and deployed by speakers of human language.

References

Anderson, Gregory and K. David Harrison (2000). *Tyvan*. Munich: Lincom Europa.

Anttila, Arto (1997a). *Variation in Finnish phonology and morphology*. PhD dissertation, Stanford University.

(1997b). Deriving variation from grammar: A study of Finnish genitives. In Frans Hinskens , Roeland van Hout, and Leo Wetzels (eds.), *Variation, change and phonological theory*. Amsterdam: John Benjamins.

Archangeli, Diana and Douglas Pulleyblank (1994). *Grounded phonology.* Cambridge: MIT Press.

Bell-Berti, Fredericka and Katherine S. Harris (1976). Some aspects of coarticulation. *Haskins Laboratories Status Report on Speech Research* SR-45/46: 197–204.

Boersma, Paul. (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 21: 43–58.

Boersma Paul and Bruce Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *LI* 32: 45–86.

Bogoroditskij, V. A. (1953). *Vvedenie v tatarskoe yazykoznanie* (Introduction to Tatar linguistics). Kazan: Tatgosizdat.

Casali, Roderic (1993). Cross-height vowel harmony in Ghanaian. Ms, University of California, Los Angeles.

Chomsky, Noam and Morris Halle (1968). *The sound pattern of English*. New York: Harper and Row.

Clements, George N. and Engin Sezer (1982). Vowel and consonant disharmony in Turkish. In Harry van der Hulst and Norval Smith (eds.), *Structure of phonological representations (part II)*. Dordrecht: Foris Publications, 213–55.

Comrie, Bernard (1981). *Languages of the Soviet Union.* Cambridge: Cambridge University Press.

Crothers, John (1978). Typology and universals of vowel systems. In Joseph H. Greenberg et al. (eds.), *Universals of human language*. Stanford: Stanford University Press, vol. 2, 93–152.

Derbyshire, Desmond (1985). *Hixkaryana and linguistic typology*. Austin, Tx: SIL publications and University of Texas, Austin.

Flemming, Edward (1995). *Auditory representations in phonology.* PhD dissertation, University of California, Los Angeles.

Goldstein, Louis (1991). Lip rounding as side contact. In *Actes du XIIème congrès international des sciences phonétiques (Vol I/5: Sessions plénières et semi-plénières).* Aix-en-Provence: Université de Provence, vol. 1, 97–101.
Harrison, K. David (1999a). Harmony and underspecification in Tuvan. In *MIT Working Papers in Linguistics: Proceedings of SCIL XI.* Cambridge, MA: MIT.
(1999b). Tuvan reduplication and harmony. *BLS* 25: *Parasession on Caucasian, Dravidian and Turkic Linguistics*.
Harshman, Richard (1970). Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multi-modal factor analysis. *UCLA Working Papers in Phonetics* 16.
Harshman, Richard and S. Berenbaum (1980). Basic concepts underlying the PARAFAC-CANDECOMP three-way factor analysis model and its application to longitudinal data. In D. H. Eichorn et al. (eds.), *Present and past in middle life*. New York: Academic Press, 435–59.
Harshman, Richard, Peter Ladefoged, and Louis Goldstein (1977). Factor analysis of tongue shapes. *JASA* 62: 693–707.
Hayes, Bruce, Bruce Tesar, and Kie Zuraw (2000). 'OTSoft', software package, http://www.linguistics.ucla.edu/people/hayes/otsoft/.
Herbert, Raymond and Nicholas Poppe (1963). *Kirghiz manual*. Bloomington: Indiana University.
Homer, Molly Kathryn (1998). *The role of contrast in nasal harmony*. PhD dissertation, *University of Illinois*.
Hsu, Chai-Shune (1993). Tsou phonology. Ms, University of California, Los Angeles.
Kaun, Abigail R. (1994). An Optimality Theoretic account of rounding harmony typology. *WCCFL* 13: 78–92.
(1995). *The typology of rounding harmony: An optimality theoretic approach*. PhD dissertation, University of California, Los Angeles.
(1999). Epenthesis-driven harmony in Turkish. *BLS* 25: *Parasession on Caucasian, Dravidian and Turkic Linguistics*.
Korn, David (1969). Types of labial vowel harmony in the Turkic languages. *Anthropological Linguistics* 11: 98–106.
Li, Bing (1996). *Tungusic vowel harmony: Description and analysis.* The Hague: Holland Academic Graphics.
Linker, Wendy (1982). *Articulatory and acoustic correlates of labial activity in vowels: A cross-linguistic study*. PhD dissertation, University of California, Los Angeles.
Maddieson, Ian (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.
McCarthy, John (1979). *Formal problems in Semitic phonology and morphology*. PhD dissertation, MIT.
McCarthy John and Alan Prince (1993). *Generalized alignment.* Ms., University of Massachusetts and Rutgers University. Rutgers Optimality Archive 7, http://ruccs.rutgers.edu/roa.html.
(1995). Faithfulness and reduplicative identity. In Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk (eds.), *University of Massachusetts occasional papers in linguistics 18: Papers in Optimality Theory*. Amherst: GLSA, 249–384.
Nedjalkov, Igor (1998). *Evenki*. Cambridge: Cambridge University Press.
Özgüler, Ali (1989). *Türçedemizdeki yabanci kelimeler sözlük.* N.P.: Sim Matbaacılık.
Prince, Alan and Paul Smolensky (1993). *Optimality Theory: Constraint interaction in generative grammar*. Rutgers Optimality Archive 537, http://roa.rutgers.edu/.

Purcell, Edward T. (1979). Formant frequency patterns in Russian VCV utterances. *JASA* 66: 1691–702.
Recasens, Daniel (1984). Vowel-to-vowel coarticulation in Catalan VCV sequences. *JASA* 76: 1624–35.
Ringen, C. and Ovokki Heinämäki (1999). Variation in Finnish vowel harmony: An OT account. *NLLT* 17: 303–37.
Ross, Kie (1996). Floating phonotactics: Variability in infixation and reduplication of Tagalog loanwords. MA thesis, University of California, Los Angeles.
Smolensky, Paul (1993). Harmony, markedness, and phonological activity. Paper delivered at the Rutgers Optimality Workshop 1.
(1995). On the internal structure of the constraint component *CON* of UG. Rutgers Optimality Archive 86, htt://roa.rutgers.edu.
Steriade, Donca (1995). Underspecification and Markedness. In John. A. Goldsmith (ed.), *The handbook of phonological theory*. Oxford: Blackwell, 114–74.
Stevens, Kenneth N. (1998). *Acoustic phonetics.* Cambridge, MA: MIT Press.
Suomi, Kari (1983). Palatal harmony: A perceptually motivated phenomenon? *Nordic Journal of Linguistics* 6: 1–35.
Svantesson, Jan-Olaf (1985). Vowel harmony shift in Mongolian. *Lingua* 67: 283–327.
Terbeek, Dale (1977). *A cross-language multidimensional scaling study of vowel perception*. PhD dissertation, University of California, Los Angeles.
Yavas, Mehmet (1980). *Borrowing and its implications for Turkish phonology*. PhD dissertation, University of Kansas.

Notes

1. For each language, data from eight male subjects were obtained. Photographs were taken simultaneously with audio recordings.
2. Back vs nonback (1) separated the front and back vowels with one exception. On the basis of this factor, the phonetically back vowel [ı] was grouped with the front vowel cluster. The vowel [ɚ] fell between the back and front vowel clusters. Back vs nonback (2) grouped the vowel [ı] with the cluster of back vowels and placed [ɚ] at the low end of the scale along with {ö, i, e, and ü}.
3. This general notion of contrast is applied to the phenomenon of nasal harmony in Homer 1998.
4. In Maltese (McCarthy 1979), backness/rounding harmony targets short vowels, but not long vowels. One explanation for this pattern might be the fact that short vowels are more subject to gestural overlap than longer vowels. The historical scenario would therefore be that at an earlier stage of the language, no harmony existed. Over time, due to coarticulation from a preceding rounded vowel, the short vowels came to be perceived as phonologically rounded, and a system of rounding/backness harmony entered the grammar.
5. The languages in which nonhigh rounded vowels may not undergo rounding harmony all possess nonhigh rounded vowels on the surface. What they do not tolerate is the occurrence of nonhigh rounded vowels as the output of rounding harmony. In these systems, we may assume that the retention of nonhigh rounded vowels that are not the product of the harmony is insured by an input-output Faithfulness constraint that outranks *RoLo. The same account will apply to those languages

that allow front rounded vowels on the surface, but do not allow their appearance as the product of harmony.

6. Both *RoLo and *RoFro are equivalent to Archangeli and Pulleyblank's (1994: 78) grounded path conditions on the feature [+round] and features referring to height and backness. In particular, *RoLo should be functionally equivalent to the LO/RD Condition and the RD/LO Condition, while *RoFro should do the work of the RD/BK Condition and the FR/RD Condition.
7. Disharmonic *uCı* sequences occur, but only in fifteen roots from Old Manchu. Relatively uneducated Sibe speakers are reported by Li to be unfamiliar with these roots (Li 1996: 195).
8. The grammars here omit constraints that ensure that the lexical specifications for rounding in stem vowels are respected. Just what these constraints are probably depends on the language. For languages in which the principles of harmony are respected even within stems, a constraint of the type Ident([Rd], initial syllables) is probably appropriate. For languages like Standard Turkish, in which the lexicon is very rich in disharmonic roots, probably the correct constraint is Ident([Rd], root). In either case, the relevant constraint is undominated, so I will simply omit candidates from consideration that alter the rounding of the trigger.
9. I wish to thank Bruce Hayes and Kie Zuraw for help in generating and interpreting the factorial typology described here.
10. The full set of plausible-but-unattested patterns is as follows. In each pattern, {} encloses the vowel pairs that undergo harmony: (1) {uI, öA, oA}, (2) {uI}, (3) {uI, oI}, (4) {uI, oI, uA, oA}, (5) {oI, oA}, (6) {oI}, (7) {öI, uI, oI, öA, oA}, (8) {öI, uI, oI, öA, uA, oA}, (9) {öI, uI, oI}, (10) {öI, oI, öA, oA}, (11) {öI, oI}, (12) {üI, uI, öA}, (13) {üI, öA, oA}, (14) {üI, öA}, (15) {üI}, (16) {üI, öI, uI}, (17) {üI, öI, uI, üA, öA, oA}, (18) {üI, öI, üA, öA, oA}, (19) {üI, öI, üA, öA}, (20) {üI, öI, oI, öA, oA}, (21) {üI, öI, oI}, (22) {üI, öI, oI, üA, öA, oA}, (23) {üI, öI, oI, üA, öA}.
11. On 'conjoined' constraints of this type, see Smolensky 1995.
12. The subjects ranged in age from eighteen to thirty-five years. All resided in the New Haven, CT area at the time of the experiment, but had been raised in urban settings including Ankara, Istanbul, and Izmir.
13. Due to a fronting effect of borrowed [l], this word is transcribed as *bül'uz* in both Yavas 1980 and Clements and Sezer 1982. The subjects in the experiment reported here did not produce front vowels before [l], supplying *buluz* instead.

5 The evolution of metathesis

Juliette Blevins and Andrew Garrett

1 Introduction

Our object of study in this chapter is metathesis, which we define as any reordering of segments or features within the phonological string.[1] Representative cases, discussed in more detail below, are illustrated in (1).

(1) a. Rotuman: /mo**fa**/ → *mo**af*** 'rubbish'; cf. (20) below
Proto-Indo-European > Armenian: **kʲ**u**bʰr**os* > *su**rb*** 'holy' (initial **kʲ* > *s*); cf. (6) below

b. Nxilxcín (Colville): *s**ʕ**áy* 'they are noisy' vs *sy-m-ənc**ʕ**àt* 'they make noise'; cf. (15) below
Classical > South Italian Greek: *gamb**r**ós* > *g**r**ambó* 'son-in-law'; cf. (10) below

c. Marathi: *õʈʰ* > ***h**õʈ* 'lip'; cf. (17) below

Adjacent segments seem to exchange positions in the common pattern seen in (1a), while (1b) shows examples of nonlocal movement. A case of feature metathesis is shown in (1c); such cases are relevant because feature and segment metatheses differ in their phonological effects but not their underlying causes.

Metathesis has long posed problems for phonological theory. These problems are of two main types: metathesis has resisted analysis in terms of phonetically natural or motivated sound change, and the reordering of sounds in metathesis has required extensions of otherwise highly restrictive phonological formalisms. We will argue here that metathesis can, despite these problems, be explained in a phonetically natural way based on precisely the same assumptions required to understand other phonological phenomena.

We also have a more programmatic goal. In recent years, phonologists have increasingly come to accept the view that phonological patterns, both within and across languages, can be explained by reference to the findings of experimental phonetics. As yet, however, there is no consensus as to the precise explanatory nexus between the two areas. In this chapter we will contrast two views of the relationship between phonetics and phonology, for which we will use the short-hand terms *phonetic optimisation* and *evolutionary phonology*. The first approach seeks to explain phonological patterns as the result of optimisation of some aspect of phonetics such as articulatory ease or perceptual salience. On this view, sound patterns are caused by (can be explained by) the phonetic optimisation they yield, and sound changes occur because their output is phonetically 'better' in some way – for example, easier to articulate or perceive. The phonetic optimisation approach has been advocated by numerous scholars, including many contributors to this volume (here and elsewhere).

We will suggest a very different approach here. Our view is that diachronic regularities play a major role in determining phonological typology. Since actual phonological systems have evolved diachronically, their properties reflect constraints on sound change as well as constraints on the nature of phonological systems. Explanations for phonological patterns may reside in synchronic analysis or diachronic evolution. Which explanation will emerge in any case is a matter to be resolved based on the evidence, but since historical accounts permit simpler grammatical models, they are preferable wherever possible.

Certain sound patterns are cross-linguistically frequent as a consequence of convergent evolution: the intrinsic properties of speech perception and production result in certain frequent sound changes; these in turn yield common sound patterns. We maintain that if sound patterns can be explained as the result of convergent evolution in this sense, the burden of proof falls on those who choose to duplicate such explanations in the synchronic domain. In short, one goal of the evolutionary phonology approach is to help simplify synchronic models by developing phonetically plausible diachronic explanations for phonological patterns.

Two problems that any model of phonological diachrony must confront are the mechanism of sound change and the cause of its typical regularity. In our view sound change is mainly caused by listener-based reinterpretation. This in turn may arise in several ways. For example, the actual phonetic string may present a listener with multiple potential phonological analyses; or a listener may simply misperceive the utterance due to biases in the perceptual system; or a listener may confront a choice of phonological analyses due to speaker variation on a continuum from hyperarticulated listener-oriented 'clear' speech to reduced, hypoarticulated 'casual' speech. In the last case, reanalysis reflects ambiguity presented by multiple phonetic forms in the input, not the ambiguous nature of a single phonetic form.

Sound change is regular for the same reason that language learners consistently categorise contextually determined phonetic categories with parallel phonological categories. English *pit*, *pat*, *pet*, *pot*, *put* are all 'learned' with initial /p/ and final /t/ because ranges of values for some set of cues (e.g. VOT, closure duration, burst properties, CV transition formant values) are interpreted as defining a single linguistic category. Because the sources of sound change all involve categorical perception, a shift in phonological representation for one lexeme will result in the same shift for another lexeme containing the same phonetic category. For example, in an English dialect where final /t/ is realised as [ʔt], the phonetic properties of this realisation may allow reinterpretation as /ʔ/. If this happens, since the cues now interpreted as defining /ʔ/ are found in *pit*, *pat*, *pet*, *pot*, *put*, a /t/ > /ʔ/ shift will occur in all these words. The regularity of sound change is thus a special case of the regularity of phonological category acquisition.

In comparing different approaches to sound change, it should be emphasised that the question of optimisation – are sound patterns functionally motivated? – is logically distinct from the question of whether phonetic explanations for sound patterns belong in the diachronic or synchronic arena. This suggests a four-way typology along the lines in (2).

(2) a. Synchronic + nonfunctionalist
b. Synchronic + functionalist (e.g. Flemming 1996; Hume 1997, 2001; Boersma 1998; Steriade 2001)
c. Diachronic + functionalist (e.g. Grammont 1950; Vennemann 1988)
d. Diachronic + nonfunctionalist (e.g. Ohala 1974, 1981, 1993; Blevins and Garrett 1998)

Various scholars' work is crudely classified in (2b–d); the nonfunctionalist synchronic approach in (2a) has been standard in phonological theory. The view we will defend here is diachronic and nonfunctionalist: phonetic explanations play an important diachronic role in explaining sound patterns, but (at least for the phenomena we investigate) optimisation is irrelevant.

Several forms of the phonetic optimisation approach can be envisioned. A relatively strong position is that optimisation is a property of *all* sound change (or all sound changes of a particular structural type). Arguing against the view that misperception causes metathesis sound changes, Steriade (2001: 234–5) writes as follows:

> [C]onfusability is, in principle, symmetric . . . [If] sound change is initiated as misperception, there would be no reason to expect metathesis in one direction and not in the other. In fact, however, the direction of metathesis is highly constrained. Only certain types of reversal, which can be identified as perception-optimising, are frequent and systematic . . .

The claim that all 'frequent and systematic' types of metathesis optimise perception represents a strong form of the phonetic optimisation approach.[2]

An alternative weaker position, as Donca Steriade reminds us, is simply that *some* optimising sound changes exist. Yet this weaker position is problematic. To refute the hypothesis that all sound change is optimising, it suffices to identify non-optimising sound changes, but it is harder to find evidence bearing on the weaker hypothesis that just some sound changes are motivated by optimisation.

On our account, some sound changes will have the effect of optimising aspects of phonetics simply by chance; indeed, in numerous cases, misperception leads directly to optimised phonetics. Therefore, the mere existence of optimising changes, even in great numbers, is not evidence for the phonetic optimisation model. To defend this model in its weaker form, one must argue either that there exist optimising sound changes that cannot involve perceptual reinterpretation (changes whose input and output cannot be related via misperception) or that the overall typology of sound changes follows from the optimisation model but not from our model. Our position is the reverse: the typology of sound changes follows from the evolutionary phonology model but not from the phonetic optimisation model.

For all these questions metathesis is of special interest. In traditional phonology, especially among the neogrammarians and structuralists, metathesis was treated as marginal precisely because it seemed to contradict standard doctrines separating phonetics and phonology. Essentially, all scholars who studied the matter came to conclusions like those of Grammont (1923), according to whom CC metathesis arises in order to avoid 'unpronounceable' clusters. It is also governed phonotactically, according to Grammont: less sonorous consonants (those with smaller 'aperture') are always positioned closer to a syllable boundary and more sonorous consonants closer to the syllable nucleus. In other words, unlike most other processes (e.g. assimilation), metathesis was seen as an output-driven phonological process.

In this chapter, extending earlier work on consonant-vowel metathesis (Blevins and Garrett 1998), we present a comprehensive and restrictive typology of regular metathesis in the world's languages. We identify four main types of metathesis, with specific phonetic characteristics. We list these metathesis types in (3), together with the phonetic features that allow us to explain and categorise them. We should emphasise that our names for these metathesis types are partly arbitrary labels, serving mainly to distinguish them from each other; coarticulation, perception, and audition play a role in all four types.

(3) Metathesis type	Phonetic feature
a. Perceptual metathesis (§3.1)	Elongated phonetic cues (§2.1)
b. Compensatory metathesis (§3.2)	Stress-induced temporal shifts (§2.2)
c. Coarticulatory metathesis (§3.3)	CC coarticulation (§2.3)
d. Auditory metathesis (§3.4)	Auditory-stream decoupling (§2.4)

The first type of metathesis involves features of intrinsically long duration (e.g. pharyngealisation); in multisegmental strings, such features are spread out over the entire sequence, allowing them to be reinterpreted in nonhistorical positions. The second type is prosodically conditioned: within a foot, features in a weak syllable undergo temporal shifts into the strong syllable. The third type of metathesis arises in clusters of consonants with the same manner of articulation but different places of articulation; the place cues do not necessarily have long duration, and we will suggest that metathesis results from coarticulation facilitated by shared articulatory gestures. The fourth type of metathesis results from the auditory segregation of sibilant noise from the rest of the speech stream.

Our typology is both restrictive and predictive. A segment or feature may undergo metathesis (be reinterpreted in a nonhistorical position) only if the phonetic signal is ambiguous or otherwise presents difficulties in feature or segment localisation. The phonetic properties underlying such ambiguities or difficulties are discussed in section 2. In section 3 we detail our metathesis typology. In section 4 we address some general issues, summarise our findings, and discuss the general role of phonetics in phonology: phonology is phonetically driven, but only in the diachronic dimension.

2 Phonetic background

In this section we outline the phonetics underlying the metathesis types to be surveyed in section 3.

2.1 Elongated phonetic cues

Segmentation is a long-standing problem in phonetic theory. For example, it is well known that consonant and vowel articulations, or their acoustic consequences, overlap in CV and VC contexts. Accurate perception of place of articulation for a prevocalic oral stop consonant is based primarily on information from the CV transition (Liberman 1970); the place features of the consonant are cued by information that co-occurs with the periodic waveform characteristic of a vowel, making it difficult to say where the consonant ends and the vowel begins.

Perceptual metathesis is closely linked to the segmentation problem as follows. As emphasised by Ohala (e.g. 1993) in his discussions of dissimilation, certain perceptual features are typically realised over relatively short time durations, whereas others are typically realised over relatively long durations. For example, irrespective of its phonological association with a consonant, vowel, or glide, pharyngealisation is typically phonetically realised over a minimal CV or VC domain. Listeners thus confront a problem if an entire CVC sequence is pharyngealised. If features are associated at some level with unique segments, there are at least seven logical possibilities for the phonological representation of the pharyngealised CVC sequence: any of the three segments could carry a secondary pharyngealisation feature ($C^{ʕ}VC$, $CV^{ʕ}C$, $CVC^{ʕ}$), or a pharyngeal could be the source of ambient pharyngealisation (ʕCVC, CʕVC, CVʕC, CVCʕ).[3] If the historical source of pharyngealisation is a pharyngeal glide and the listener posits a pharyngeal glide in a nonhistorical position, metathesis has occurred.

Phonetic studies show that many features have multisegmental domains spanning CV or VC strings, entire syllables, or strings of syllables. For example, West (1999, 2000) has established significant long-distance coarticulatory effects of English rhotics and laterals by replacing these segments with progressively longer sequences of noise; speakers can accurately identify the contrast between [ɹ] and [l] based on coarticulatory effects up to two syllables away from their phonological position in the string. This perceptual evidence is consistent with the articulatory findings of Kelly and Local (1986) and Kelly (1989). For at least one English dialect (Kelly and Local 1986), electropalatography data show that velar closure in *came* is significantly more front after *ballet* in (4a) than after *Barry* in (4b).

(4) a. Ballet came to my mind.
    b. Barry came to my mind.

The perceptual evidence is also consistent with acoustic studies. Kelly and Local (1986) show that the 'domain of resonance' of a liquid (i.e. its acoustic consequences) is measurable in all subsequent unaccented syllables. Tunley (1999) shows via measurements of vowel formants that English rhotics have significant long-distance effects on unstressed vowels, both perseveratively (on V_2 in rV_1CV_2 strings) and anticipatorily (on V_1 in V_1CV_2r strings). The effects documented by Tunley involve lowering of F2 and F3. She also shows that incorporating this sort of coarticulatory detail into synthetic speech can improve segmental intelligibility by 7–28 per cent, again providing evidence for long-distance coarticulation as a natural feature of speech which, when present, is perceptually accessible.[4]

In table 5.1 we list phonetic features with demonstrated drawn-out domains in one or more languages, along with their common phonological realisations and salient acoustic characteristics. Acoustic and articulatory data show that all these features have long domains spanning minimal VC/CV domains, entire syllables, or sequences of syllables. As mentioned above, long-domain effects of rhotics and laterals in English have been found to span domains up to three syllables long. Lip rounding and protrusion have been found to span multisyllabic domains in French and English (Lubker and Gay 1982; Benguerel and Cowan 1974). Palatalisation and velarisation with both vocalic and consonantal phonological sources have been shown to colour multisegmental domains in many different languages, including Catalan (Recasens 1984, 1987), English (Hawkins and Slater 1994), Japanese (Magen 1984), Marshallese (Choi 1992), Russian (Keating 1988), and several Bantu languages (Manuel 1987). In at least two Arabic dialects, pharyngealisation or tongue backing has been measured across multisyllabic domains while showing gradient properties typical

Table 5.1. *Features with typically long durations*

Feature	Segmental realisations	Acoustic property with long duration
rhoticity	rhotics, rhotic Vs	lowered F3 (LM: 244, 313)
laterality	laterals, lateral Vs	lateral formants (LM: 193–7)
rounding	rounded Cs, rounded Gs, round Vs	lowering of all formants (LM: 356–8)
palatalisation	palatalised Cs, palatal Gs, high front Vs	raised F2 (LM: 364)
velarisation	velarised Cs, velar Gs and high back Vs	lowered F2 (LM: 361–2)
pharyngealisation	pharyngealised Cs, Gs and Vs, ʕ, ħ	lowered F3, raised F1 (LM: 307)
laryngealisation	laryngealised Cs, Gs and Vs, ʔ	more energy in F1, F2 more jitter (LMJ)
aspiration	aspirated/breathy Cs, Gs and Vs, ɦ, h	more energy in F0; more noise (LMJ);
retroflexion	retroflex Cs and Vs	lowered F3, F4; clustering of F2, F3, F4 (L: 203, LM: 28)
nasalisation	nasals, nasalised vowels and glides	spectral zero/nasal anti-resonance (LM: 116)

(L = Ladefoged 1993; LM = Ladefoged and Maddieson 1996; LMJ = Ladefoged, Maddieson, and Jackson 1988.)

of phonetic coarticulation as opposed to phonological harmony (Ghazeli 1977; Card 1979); see Bessell 1992, 1997, 1998a, 1998b for phonetic analysis of multisegmental pharyngealisation domains in Interior Salish languages. Measurements of laryngealisation (creaky voice) and aspiration (voicelessness) in Cayuga by Dougherty (1993) show CV or VC domains that inform our analysis of metathesis in that language (Blevins and Garrett 1998: 509–12). The acoustic correlates of retroflexion typically have a minimal VC domain, as has been shown for Gooniyandi (McGregor 1990), Gujarati (Dave 1977), Hindi (Stevens and Blumstein 1975), Malayalam (Dart 1991), and Tiwi (Anderson and Maddieson 1994). Long-domain effects of nasalisation are also well documented; see Cohn 1990 and Walker 2000 for summaries of the vast phonetic literature on this subject. Finally, other features that have no standard phonological representation also show drawn-out domains, such as the jaw movement required for low front vowels (Amerman, Daniloff, and Moll 1970).

By 'typically long duration' in table 5.1 we mean that, in the majority of cases where the phonetic correlates of the features have been measured, they have been found to extend minimally across entire CV or VC strings. We do not claim that these features always take multisegmental domains, but simply that they can, and that they do so in the linguistic systems that give rise to metathesis sound changes. Missing in table 5.1 are major consonantal place of articulation (coronal, labial, dorsal), voicing, frication, continuancy, and the major class features. These features, unlike the features in table 5.1, typically show temporal alignment with single segments; on our approach they are not expected to take part in regular metathesis (though see sections 2.4 and 3.4 on the status of fricative noise).

For some phonetic features in table 5.1 a multisegmental coarticulatory domain has been phonologised, resulting in syllable-, foot-, and word-based harmonies. For example, though pharyngealisation is a feature of pharyngeal glides or coronal consonants in many Arabic dialects, it takes the syllable as its minimal domain in Cairene Arabic (Hoberman 1995 with further references) and has even broader domains in other dialects (Watson 1999); cf. Bessell 1992, 1998a, 1998b on Interior Salish pharyngealisation harmonies. Similarly, in at least two Australian languages, Mayali and Murrinhpatha, a retroflex coda consonant yields surface retroflex syllables (Evans 1995: 739–40). Word-level retroflex harmony is found in Yurok (Robins 1958: 12–13); this may be the long-distance effect of a formerly local coarticulatory effect found in Yurok's relative Wiyot, where a retroflexed affricate induced retroflexion on preceding low vowels (Reichard 1925: 8).[5] Labialisation and palatalisation/velarisation are well known from the word-domain harmony systems of Yokuts and Turkic languages respectively, and the typology of nasal harmony systems with syllable, foot, and word domains is detailed in Walker 2000.

At the same time, extended domains for certain features in table 5.1 are blocked in particular phonetic contexts where an incompatible phonetic feature abuts the one in question. This is important in understanding apparent exceptions to regular metathesis, or phonetic conditioning factors for particular metatheses. For example, though laryngeal metatheses of *h* and *ʔ* are common, and seem to result from the elongated phonetic cues of breathiness and laryngealisation often associated with these segments, laryngeal metathesis is typically blocked adjacent to a segment with conflicting laryngeal specifications. Thus, in Cayuga, the laryngeals (*h*, *ʔ*) metathesise with preceding vowels unless the output would be an *hʔ* or *ʔh* cluster (Foster 1982).[6] The anticipatory seepage of laryngealisation is blocked by a preceding segment that involves breathiness, and vice versa, since these two features involve antagonistic glottal gestures of constriction and spreading respectively; as a result, a vowel is not fully laryngealised, the signal is unambiguous, and metathesis does not occur. Contextual blocking effects of this type are widespread in perceptual metathesis; elsewhere (Blevins and Garrett 1998) we have discussed Cayuga in more detail together with similar cases of contextual blocking in Birom, Latin, and Le Havre French.

Now consider the nasalisation associated with a nasal stop. Spreading of this phonetic feature onto a preceding or following vowel is quite widespread and unremarkable. In phonetic terms, vowels and glides undergo coarticulatory nasalisation to a much greater extent than oral stops (Cohn 1990). We would therefore not be surprised to find nasal metathesis conditioned by an adjacent vowel, but perceptual metathesis of nasals and (oral) stops should not exist. (On putative counterexamples to this prediction see section 4.1 below.)

In sum, the phonetic features listed in table 5.1 are often characterised by long durations spanning multisegmental domains. A result of this many-to-one association between phonetic features and segments is ambiguity in segmentation. If a listener attributes the spread-out feature to a nonhistorical position, (perceptual) metathesis occurs. On this approach, exceptions to metathesis are expected just in case an adjacent phonetic feature conflicts with the spread-out feature. Coarticulation is blocked in such cases, and there is no ambiguity in segmentation.

2.2 Stress-induced temporal shifts of V-V coarticulation

Coarticulation between sequential vowels across an intervening consonant appears to occur in all spoken languages, where, in a VCV sequence, transitions from vowel to consonant and from consonant to vowel are significantly influenced by the quality of the transconsonantal vowel. In the word-final cases of compensatory metathesis we cite, there is extreme anticipatory coarticulation. This is consistent with acoustic and articulatory evidence suggesting that articulatory movement for V_2 in a V_1CV_2 sequence may begin during V_1 (Bell-Berti and Harris 1976; Fowler 1981a, 1981b; Manuel and Krakow 1984).

We have argued elsewhere (Blevins and Garrett 1998) that prosodically conditioned cases of CV metathesis (compensatory metathesis) involve temporal shifts whereby the unstressed (word-peripheral) vowel comes to be coarticulated more and more into the stressed (word-internal) position, eventually leaving no trace. This model of prosodically conditioned CV metathesis implies a relationship between stress and coarticulation in which the duration and perceptual prominence of the stressed vowel can give rise to extreme anticipatory coarticulation. Other factors that may facilitate this extreme anticipatory coarticulation include size and distribution of vowel inventory, degree of vowel variation, absence of secondary consonant articulations, absence of long consonants and consonant clusters, increased duration of stressed syllables, and relatively steady-state vowels (Blevins and Garrett 1998: 548). Phonetic studies show that, independent of prosodic effects, size of vowel inventory affects V-to-V coarticulation. As suggested by Manuel and Krakow (1984) and Manuel (1987), the size and distribution of a phonemic inventory may determine the limits of phoneme variability: in a language with a relatively small vowel system, formant frequencies of a vowel are more likely to be influenced by a vowel in an adjacent syllable than in languages with larger vowel systems where acoustic variability is not as great. All cases of compensatory metathesis known to us are found in languages with small vowel systems (three to five vowels), steady-state vowels, and simple CV syllable structure.

2.3 CC coarticulation

Consonant clusters are typically subject to variation in casual or fast speech. This variation has been argued to follow, to a great extent, from coarticulatory effects (Kohler 1976; Barry 1984; Browman and Goldstein 1990). Coarticulation within consonant clusters can have dramatically different acoustic effects depending on the extent to which the articulatory gestures involved are independent of each other. As demonstrated by Browman and Goldstein (1990), deletion, insertion, and assimilation can all be attributed to gestural overlap, with acoustic consequences following from the nature of the independent gestures and the extent to which they overlap.

We suggest in section 3.3 that the most common types of stop metathesis (PK > KP, TP > PT) are the result of extreme gestural overlap. There are several logically possible patterns of gestural overlap in VC_1C_2V sequences. One possible pattern is medial overlap of gestures with VC_1 and C_2V transitions intact. Linear order remains constant, and an excrescent segment may emerge if the two consonants differ in laryngeal or manner features (Ohala 1974). A second possibility is 'swallow-up' overlap, with the closure and release of one of the two consonants containing the closure and release of the other. If the two consonants share laryngeal and manner features, then one completely hides the other, with the surface effect of total assimilation or deletion. Finally, if closure and/or release of two consonants with distinct articulatory gestures are nearly simultaneous, place of articulation cues become difficult to recover. If the righthand cluster edge contains unambiguous release cues, it is possible to reanalyse C_1C_2 as C_2C_1. This possibility appears to be entirely dependent on the perception of nearly simultaneous closure of C_1 and C_2 as an instance of C_2, and in cases known to us it is limited to certain combinations of place features. In one subtype the clusters in question are labial-velar stop sequences; another involves coronal-noncoronal stop sequences. In both cases, coarticulation can result in nearly simultaneous closure, with labial release following velar release in the first case and with coronal release following noncoronal release in the second case. These metatheses are both unidirectional, respectively yielding velar-labial and noncoronal-coronal stop sequences.

We suggest that the unidirectionality of velar-labial stop metathesis reflects the same factors that underlie the phonetics of labial-velar stops (Connell 1994), which represent the extreme case of gestural overlap. In all labial-velar stops, the acoustic and articulatory record shows that velar release occurs before labial release (usually by 30–60 ms). Velar closure always precedes labial closure or is synchronous with it, but as noted by Connell (1994: 451), even where velar and labial closures are synchronised, 'the auditory impression is of an earlier velar closure'. In short, KP > PK metathesis is unattested because extreme coarticulation in such clusters leads naturally to a KP percept.

Gestural overlap may also explain unidirectionality in coronal-noncoronal stop metathesis, as well as illuminating an asymmetry in coronal-noncoronal stop assimilation patterns. As Bailey (1969, 1970) and Blust (1979) observe, there are striking parallels between possible place assimilation and attested metathesis in coronal-noncoronal clusters. In metathesis, coronal-noncoronal clusters invert position, giving rise to noncoronal-coronal clusters, but the reverse metathesis is unattested. A parallel asymmetry in the assimilation of heterorganic stop clusters in English is illustrated in (5) with examples from Blust (1979: 103).

(5)	Assimilation		No assimilation	
	tp	foo**tp**rint, hi**t p**arade	pt	ri**pt**ide
	tk	sui**tc**ase, ca**tc**all	kt	co**ckt**ail
	db	goo**db**ye	bd	ru**bd**own
	dg	hea**dg**ear	dg	do**gd**ays
	nm	fa**nm**ail, gu**nm**an	mn	roo**m n**umber
			ŋn	ha**ngn**ail

Regressive assimilation is possible and common in coronal-noncoronal clusters, but not perceptually salient for noncoronal-coronal clusters. These observations now have ample acoustic and articulatory support (e.g. Zsiga 1994; Byrd 1996), allowing us to conclude for English that gestural overlap in coronal-noncoronal stop clusters is greater than in noncoronal-coronal clusters, and that in coronal-noncoronal stop clusters the lips or tongue body often move toward closure in production of a noncoronal before closure for the coronal stop is achieved. We hypothesise that, as with labial-velars, the percept of simultaneous coronal and noncoronal closure can be one in which noncoronal closure features prevail.[7]

2.4 Auditory-stream decoupling

A number of regular metatheses involve sibilant-stop or stop-sibilant sequences. The primary acoustic cue for fricative manner of articulation, irrespective of place of articulation and voicing, is the presence of aperiodic noise in the spectrum (Delattre, Liberman, and Cooper 1962). Jongman (1989) demonstrates that the duration of this noise should be at least 20 ms. (In natural speech it is usually much longer, around 100 ms.) This noise is most intense for sibilant fricatives.

While there is still much work to be done on the acoustics and perception of sibilant noise, a number of studies suggest that, in consonant clusters containing sibilants, the sibilant noise somehow distracts the listener, leading to high confusion rates with respect to the linear order of segments (Bregman 1990). Specifically, there is a tendency to decouple sibilant noise from the rest of the speech stream, and this decoupling can result in dramatic misperceptions.[8]

An additional and possibly contributing factor is the misperception of fricatives as affricates or stops, and vice versa. Several studies demonstrate that the onset of noise must be fairly gradual for a segment to be perceived as a fricative. If it is too abrupt, the stimulus will be perceived as an affricate or a stop (Gerstman 1957; Cutting and Rosner 1974; Keating and Blumstein 1978; Cutting 1982). Even expert listeners have been found to perceive short intervals of sibilant noise as stops (Whalen 1991).

3 Typology of metathesis

In this section we will discuss the four different metathesis types identified in (3), showing how their properties receive natural explanations in the evolutionary phonology framework.

3.1 Perceptual metathesis

In cases of perceptual metathesis, a segment (or feature) with elongated phonetic cues as discussed in section 2.1 shifts its linear position in a phonological string. Our view is that this partly reflects the perceptual difficulty of localising the origin of a phonetic cue with long-distance effects. The result of perceptual metathesis is a 'mistake' from the point of view of the previous linguistic system: a segment (or feature) is reinterpreted as originating in a new position within the elongated span. This will involve the transposition of adjacent elements in some cases, and in other cases the metathesis will be nonlocal. Since we have already surveyed perceptual metathesis in adjacent CV sequences (Blevins and Garrett 1998), we focus here on other perceptual metathesis contexts: local CC metathesis and long-distance metathesis.[9]

The 'disproportionately high (and widespread) frequency of occurrence of liquids in metathesis' is called 'proverbial' by Ultan (1978: 375), and we begin with several cases involving rhotics. First, in the prehistory of Classical Armenian (Grammont 1908; Schmidt 1981; Ravnæs 1991), the linear order of stop (or affricate) + *r* clusters was regularly inverted, in initial as well as medial position. This is shown in (6).

(6)	Indo-European		Armenian			
a.	*k^{j}u**b^{h}r**os	>	*su**br**(V)	>	su**rb**	‘holy’
	*b^{h}i**dr**os	>	*b^{h}i**tr**n(V)	>	bi**rt**	‘rigid, rude’
	*me**g^{jh}r**i	>	*med͡z**r**(V)	>	me**rd͡z**	‘near’
b.	*d^{h}a**b^{h}r**os	>	*da**br**in	>	da**rb**in	‘smith’
	*swi**dr**os	>	*k^{h}i**tr**n	>	k^{h}i**rt**n	‘sweat’
c.	***b^{h}r**ātēr	>	***br**ājr	>	eł**b**ajr	‘brother’
	***b^{h}r**ēwr	>	***br**ewr	>	ał**b**ewr	‘spring, well’
	***dr**ak^{j}u	>	***tr**asu-	>	a**rt**asu-	‘tear(s)’
	***g^{w}r**āwōn	>	***kr**an	>	e**rk**an	‘millstone’

Note that a prothetic vowel *e* (or *a*, if *u* or *w* follows) arose in Armenian words beginning with a rhotic. In the first two forms in (6c), *ł* < **r* as a dissimilatory effect of the following *r*.[10]

A comparable sound change has occurred in Rendille (a Cushitic language spoken in Kenya) and is still manifested in synchronic alternations involving underlying obstruent-*r* and nasal-*r* sequences (Heine 1976; Oomen 1981; Sim 1981). The metathesis is shown in (7a) and (8a); the forms in (7b) and (8b) are for comparison.

(7)		2 sg. = 3 sg. fem.	1 sg. = 3 sg. masc.
a.	‘see’	á**gar**-te	á**rg**-e
	‘shiver’	ħá**mar**-te	ħá**rm**-e
	‘sleep’	ú**dur**-te	ú**rd**-e
b.	‘be full’	ɖá**rag**-te	ɖá**rg**-e

(8)		Singular	Plural
a.	‘bag’	**ugár**	**urg**-ó
	‘clothing’	da**fár**	da**rf**-ó
	‘mother’	a**bár**	a**rb**-ó
b.	‘gate’	**arít**	**art**-ó

In the Armenian and Rendille metatheses, an original *Cr* sequence inverts its order: *Cr* > *rC*. While common, this is not the only pattern for rhotic metathesis. A regular *rð* > *ðr* sound change has occurred in several eastern dialects of Judeo-Spanish (Ladino). This is shown in (9) with data from the Istanbul dialect (Subak 1906: 171–2) as well as standard Spanish for comparison.

(9)	Standard Spanish	Istanbul Judeo-Spanish	
	ta**rd**e	la ta**ðr**e	‘evening’
	basta**rd**o	bastá**ðr**o	‘bastard’
	ve**rd**ura	ve**ðr**úra	‘verdure’
	cue**rd**a	kwé**ðr**a	‘cord’
	co**rd**ero	ko**ðr**éro	‘lamb’
	so**rd**o	só**ðr**o	‘deaf’

Note that standard *rd* is [ɾð] and Judeo-Spanish *r* is [ɾ].[11]

Two interesting points about the directionality of metathesis emerge from the patterns in (6–9). First, directionality is independent of pre-existing phonotactics, since Armenian, Rendille, and Spanish all had both *Cr* and *rC* clusters prior to metathesis. Second, the directionality of the Armenian and Rendille metatheses, affecting various *Cr* cluster types, differs from that of the Judeo-Spanish metathesis, affecting only [ɾð] clusters. We suggest that the perseverative nature of the Judeo-Spanish rhotic shift may be a consequence of coarticulatory effects.

The three rhotic metatheses discussed above operate locally, transposing adjacent segments only. Long-distance liquid metathesis has occurred as a sound change in South Italian dialects of Greek (Rohlfs 1950, 1964).[12] In these dialects, prevocalic *r* or *l* in a noninitial syllable has been transposed into the initial syllable in certain circumstances. This occurred whenever (i) the liquid was positioned after an obstruent and either (iia) the initial syllable had a prevocalic noncoronal obstruent or (iib) the liquid was *r* and the initial syllable had a prevocalic *t*. If these conditions were satisfied, the liquid moved into prevocalic position in the initial syllable. As shown in (10), this resulted in word-initial *(s)Cr* clusters.[13]

(10)		Classical Greek	South Italian Greek	
	a.	*bót^{h}**r**akos	v**r**úθako	'frog' (Rohlfs 1924: 15–16; 1933: 19)
		februā́rius (L)	f**r**evári (O)	'February'
		gamb**r**ós	g**r**ambó	'son-in-law'
		kóp**r**os	k**r**ópo (O)	'dung'
		k^{h}ond**r**ós	x**r**ondó	'thick'
		past**r**ikós	p**r**ástiko	'clean'
		pik**r**ós	p**r**ikó	'bitter'
		tágist**r**on	t**r**ástina	'food bag'
	b.	fáku**l**a (L) > *fák**l**a	f**l**áka	'torch'
		*fúsk**l**a	*f**l**úska	'chaff' (Rohlfs 1933: 74–5)
		spéku**l**a (L) > *spék**l**a	sp**l**éka	'elevated place'

Contexts where the metathesis fails to occur are illustrated in (11).

(11)		Classical Greek	South Italian Greek	
	a.	ka**l**ós	ka**l**ó	'attractive'
		ka**r**día	ka**r**día	'heart'
		pa**r**at^{h}ý**r**a	pa**r**aθí**r**a	'side door'
	b.	ánt^{h}**r**ōpos	áθ**r**opo	'man'
		lūt**r**ón	lut**r**ó	'bath' (place name)
		mét**r**on	mét**r**o (O)	'measure'

	nep^{h}**r**ós	nef**r**ó	'kidney'
	*pléktra	pléθt**r**a	'plait'
c.	dák**r**yon	ðák**r**i	'tear'
	dénd**r**on	ðend**r**ó	'(oak) tree'
	dip**l**ȗs	ðip**l**ó	'doubled'
	kyk**l**íon	t͡ʃik**l**í	'small circle'
	*sék**l**ion	sék**l**i	'beet greens'
	tábula (L) > táb**l**a	táv**l**a	'table'

The data in (11a) show that intervocalic and preconsonantal liquids are unaffected, and (11b) shows that liquids are transposed only into initial syllables with prevocalic obstruents. The data in (11c) show that metathesis never yields clusters consisting of a coronal obstruent plus *l* (e.g. *tl*, *sl*, *ðl*, *t͡ʃl*) or consisting of a coronal fricative or affricate plus *r* (e.g. *ðr*, *t͡ʃr*). This is interesting because some inherited clusters of these types do exist. Compare, for example, the first two forms in (11c) with the forms in (12).

(12)	Classical Greek	South Italian Greek	
	drákōn	ð**r**áko	'dragon'
	dráks	ð**r**áka	'handful'
	drómos	ð**r**ómo	'way'

The failure of metathesis in *ðákri* 'tear' (not **ðráki*) cannot be attributed to structure preservation and has no obvious interpretation in the phonetic optimisation approach.

Perceptual metathesis involving labialisation and palatalisation is also well attested (Blevins and Garrett 1998). Here the difference between CV and CC metathesis is minimal, being essentially the positionally determined difference between *VCu* > *VwC* and *VCwV* > *VwCV* metathesis. Comparable long-distance cases are found among the Ethiopian Semitic labialisation and palatalisation processes described by Hetzron (1971; 1977: 45–9), Rose (1997), and other authors.

Perceptual metathesis also involves pharyngeals. For example, a synchronic adjacent-element pharyngeal metathesis has been reconstructed for Proto-Indo-European, and regular pharyngeal interpolation into adjacent vowels has occurred in the history of Cypriot Greek; both cases are cited in Blevins and Garrett 1998. A local pharyngeal metathesis is said to exist in Rendille, where 'the pharyngeal fricative switches with an adjacent consonant when preceded by the low vowel /a/' (Hume 1997: 294). This is illustrated in (13) by three plural nouns and verbs (Heine 1976: 214; 1978: 73; Oomen 1981: 50, 63).

(13)	Non-prevocalic	Prevocalic	
	aħa**m** (sg.)	a**m**ħ-a (pl.)	'eat!'
	baħá**b** (sg.)	ba**b**ħ-ó (pl.)	'armpits'
	saħa**b** (sg.)	sa**b**ħ-o (pl.)	'clap of hand'

Other forms, however, apparently fail to undergo this metathesis. Thus the word *sáħta* 'tomorrow' (Heine 1976: 222) has a surface [aħC] sequence in an apparently underived context, and the prevocalic forms cited in (14) from Heine (1976: 213, 220) lack metathesis despite being generally comparable to those in (13).

(14)	Non-prevocalic	Prevocalic	
	bí**ħi**-	bá**ħc**-	'remove'
	na**ħas** (sg.)	na**ħs**-ó (pl.)	'breast'

A possible explanation suggests itself if, as these data suggest, the real generalisation is that *aħC* → *aCħ* metathesis occurs only when C is voiced. Since only voiced segments are compatible with the spread of the voiced pharyngeal articulatory gesture, this extended feature is blocked by a following voiceless consonant. Restriction of the vocalic context to *a* is natural too, since an extended feature is especially likely to be mislinearised (to undergo metathesis) if it is hard to perceive in its original location.[14]

A long-distance pharyngeal metathesis occurs in the Interior Salish language Nxilxcín (Colville). In Nxilxcín, roots whose citation forms begin with a *Cʕ* cluster surface as such in forms with root stress, but in forms with suffix stress the pharyngeal instead surfaces immediately before the stressed vowel. In (15) this process, called 'pharyngeal movement' by Mattina (1979), is illustrated with contrasting forms derived from four different roots. In each case, the first form cited has stress on the root while the second form has stress on a suffix; the pharyngeal regularly precedes the stressed vowel.[15] For clarity we underline the root in all forms.

(15)		Root (and suffix)	Forms with root vs suffix stress			
	a.	ɬʕac	'soak(ed), drip'	c-k-ɬʕàc-p	'(it) still had a drop'	(S #308)
		(-əp)		c-ɬə-ɬc-ʕ**á**p	'(it) had water on'	(S #346)
	b.	q'ʷʕáy	'black, soiled'	q'ʷʕáy-lqs	'preacher' (<'black robe')	(S #655)
		(-**í**c'aʔ)		i-s-t-q'ʷəy-ʕ**á**c'aʔ	'I am dirty'	(S #753)
	c.	sʕáy	'make noise'	sʕáy	'they are noisy'	(S #890)
		(-əncú**t**)		sy-m-əncʕ**à**t	'they make noise'	(S #563)
	d.	χʕál	'day(light)'	s-χəl-χʕàl-t	'day'	(S #8)
		(-**ú**l'axʷ)		χəl-p-ʕ**á**l'axʷ	'it's daylight'	(Mattina 1979)

This case too is readily analysed in our framework. Unstressed vowels in Interior Salish languages are typically either reduced or deleted – an effect which could only enhance the intrinsic difficulty of hearing pharyngeals.[16] For the prehistory of Nxilxcín, we assume that pharyngealisation had an extended phonetic domain, that it was hard to perceive, and therefore sometimes not perceived in its original (root-internal) position when the root was unstressed, and that the linear position of this feature was then reinterpreted as being in the position where it was perceived, namely in the stressed syllable. This change *results* in perceptual optimisation, for the natural reason that what is harder to hear is sometimes not heard, but our account does not invoke perceptual optimisation as a mechanism or cause of the change.

Local and long-distance glottalisation metathesis is widespread. An interesting long-distance case is found in the Interior Salish language Secwepemctsín (Shuswap). According to Kuipers (1974), Secwepemctsín nonsyllabic glottalised sonorants do not surface as such in postconsonantal position. The sample data in (16) show various surface forms associated with a single suffix containing a glottalised sonorant. If an underlyingly glottalised sonorant is postconsonantal and to the right of the main accent, then its glottalisation shifts either leftward onto an immediately post-tonic sonorant (as in 'priest', 'break off boughs', 'I heat stones'), if there is one, or rightward onto an immediately following syllabic sonorant (as in 'to heat stones').[17]

(16)	Roots and suffixes		Derived forms	
a.	-é**l'**qs	'clothing'	t-k^{w}ltk-é**l'**qs	'underwear'
	q'^{w}e**y-**	'black'	q'^{w}é**y'-l**qs	'priest'
b.	-í**l'**əp	'foundation'	c'lx^w-í**l'**əp	'chair'
	q'i**w-**	'break'	c-q'í**w'-l**əp	'break off boughs for bedding'
c.	-ésx**n'**	'rock'	t-χy-ésx**n-m'**	'to heat stones'
	χe**y**-	'heat'	t-χyé**y'**sx**n**-m-kn	'I heat stones'

This case is of special interest not just because it involves a long-distance metathesis, but because the metathesis is strictly featural: glottalisation is detached from its segmental source.

A comparable long-distance featural metathesis has occurred in the history of Marathi, where aspiration (or breathy voice) has regularly shifted to word-initial position from the onset of a second syllable. This can be seen in (17), comparing Marathi forms with their Sanskrit ancestors and in some cases with more proximately related Prakrit forms (Bloch 1915; Turner 1962–66).

(17)	Sanskrit	Prakrit	Marathi	
a.	du**h**itr̩-	du**h**ia-	d$^{\mathbf{h}}$uːv	‘daughter’
	gr̩hɳāti	gaɳ**h**aːti	g$^{\mathbf{h}}$eɳẽ	‘takes, seizes’
	jab$^{\mathbf{h}}$ati		d͡ʒ$^{\mathbf{h}}$aɳẽ	‘copulates’
	kakʂa-	kakk$^{\mathbf{h}}$a-	kãːk$^{\mathbf{h}}$, k$^{\mathbf{h}}$ãːk	‘armpit’
	kaʈ$^{\mathbf{h}}$ina-	kaɖ$^{\mathbf{h}}$iɳa-	kaɖ$^{\mathbf{h}}$iːɳ, k$^{\mathbf{h}}$aɖiːɳ	‘hard’ > ‘difficult’
	ma**h**attara-		m$^{\mathbf{h}}$aːtaːraː	‘greater’ > ‘old’
	ma**h**iʂa-	ma**h**isa-	m$^{\mathbf{h}}$ais	‘buffalo’
b.	ast$^{\mathbf{h}}$i	att$^{\mathbf{h}}$i-	**h**aːd	‘bone’
	oʂʈ$^{\mathbf{h}}$a-	oʈʈ$^{\mathbf{h}}$a-	õʈ$^{\mathbf{h}}$, **h**õʈ	‘lip’

Aspiration in (17a) has shifted to a word-initial consonant or consonant cluster; in the originally vowel-initial words in (17b), aspiration has shifted to initial position. Note that by-forms with and without metathesis are said to exist in several cases.

Two general issues arise in the analysis of long-distance perceptual metathesis. The first concerns directionality effects. As we note elsewhere (Blevins and Garrett 1998), in cases known to us a segment or feature moves either into an initial syllable or into a position defined by proximity to stress. Examples of the stress type include the Nxilxcín and Secwepemctsín metatheses in (15–16) above; examples of the initial-syllable type include the South Italian Greek and Marathi metatheses in (10) and (17) above, Romani aspiration metathesis (Matras 2002: 35–6), and *r* metathesis in Luchonnais Gascon (n. 12) and Sardinian (Geisler 1994; Molinu 1999). Both patterns involve movement into what is plausibly regarded as a relatively prominent position. In phonetic optimisation approaches, this could be related to ease of perception: a liquid, pharyngeal, or laryngeal surfaces in a position where perception is optimised. The same patterns can also easily be explained on our approach: if a segment (or feature) has extended cues of the sort responsible for perceptual metathesis, then if its linear origin is misperceived it is likelier to be misperceived as originating in a more perceptually salient (prominent) position.

A second general issue concerns blocking in long domains. A referee notes that our analysis predicts that ‘when there is a blocker (a gesturally incompatible segment that blocks coarticulation or the long cue extension) there should be no metathesis’. This prediction distinguishes our account from the phonetic optimisation approach, and it seems to be the correct prediction. For example, as seen in (18a), the South Italian Greek liquid metathesis in (10) above was not restricted to adjacent-syllable transpositions.

(18)		Classical Greek	South Italian Greek	
	a.	kapíst**r**ion	k**r**apísti	‘halter’
		konū́ku**l**a (L) > *konū́k**l**a	k**l**onúka	‘distaff’
		pédik**l**on	p**l**étiko (O)	‘fetter’

b.	skólumbros	sku**l**ímb**ri**	‘wild artichoke sp.’
	*spé**l**end**r**on	spé**l**end**r**o	‘watercress sp.’
	kʰ**ar**ád**ra**	xa**r**áð**ra**	‘fissure’

As shown in (18b), this transposition did not occur if there was an intervening liquid.[18] Local metatheses too, as noted in section 2.1, show blocking effects in the form of contextual constraints.[19]

3.2 Compensatory metathesis

We use the term ‘compensatory metathesis’ for the sound changes schematised in (19), where a vowel at the edge of the phonological domain undergoes phonetic weakening in quality and duration, with compensation for this weakening in terms of anticipatory or perseverative coarticulation of the original peripheral vowel quality in nonperipheral stressed position.

(19) Right edge: $\cdots \acute{V}_1CV_2] > \cdots \acute{V}_1V_2C\breve{V}_2] > \cdots \acute{V}_1V_2C]$
Left edge: $[V_1C\,\acute{V}_2 \cdots > [\breve{V}_1CV_1\acute{V}_2 \cdots > [CV_1\acute{V}_2 \cdots$

Our diachronic analysis of compensatory metathesis is simple. VCV sequences undergo extreme V-to-V coarticulation, with one vowel persevering or anticipating itself in full as the unstressed vowel gradually shifts its temporal alignment to the stressed syllable. Relevant phonetic literature was summarised in section 2.2.

Rotuman, an Oceanic language, instantiates the right-edge sequence in (19), which occurs within a final trochee; Ngkot̪, a Northern Paman language of Australia, exemplifies these sound changes occurring within word-initial iambs.[20] Representative examples are cited in (20–21).

(20)	Rotuman		
	ses**é**v**a**	→ ses**éa**v	‘erroneous’
	t**í**k**o**	→ t**ío**k	‘flesh’
	f**ú**ti	→ f**ý**t	‘to pull’
	m**ó**s**e**	→ m**ǿ**s	‘to sleep’
(21)	Ngkot̪		
	***alí**-	> l**áj**-	‘to go’
	***amí**-	> m**áj**-	‘up’
	***i**·n**á**-	> n**já**-	‘to sit’
	***ulá**n	> l**wá**n	‘possum’

All cases of compensatory metathesis known to us are identified and described in Blevins and Garrett 1998: 527–39. Compensatory metathesis has occurred independently in several Austronesian languages and in five branches of Pama-Nyungan. This attested limitation to the Austronesian and Pama-Nyungan families is unsurprising in the context of our analysis: within both language families, the requisite prosodic contours are found, vowel systems are small, diphthongs are for the most part absent, and secondary consonantal articulations are relatively uncommon.

3.3 Coarticulatory metathesis

Coarticulatory metathesis is a type of metathesis with articulatory origins. As outlined in section 2.3, extreme coarticulation is possible in a sequence of stops, each of which involves closure of a distinct articulator. When C_1C_2 gestural overlap results in nearly simultaneous closure, with C_1 released after C_2, a C_2C_1 cluster may be perceived. There are two identifiable subtypes, labial-velar stop sequences and coronal-noncoronal stop sequences, which we discuss in turn.

We begin with labial-velar stop sequences. We are aware of at least four independent cases of a PK > KP sound change, but no cases of a KP > PK sound change.[21] As suggested in section 2.3, the unidirectional nature of this metathesis may be related to the phonetic properties of coarticulated labial and velar stops. In at least one language, the coarticulation of labial-velar sequences appears to be optional, resulting in optional metathesis. In the Micronesian language Mokilese, as seen in (22), all /pk/ sequences are optionally realised as [kp] (Harrison 1976: 45). No such reordering occurs with any other consonant clusters, nor are Mokilese /kp/ sequences (as in /likpia/ 'flying fish with eggs') ever realised as [pk].

(22)	/a**pk**as/	[a**pk**as], [a**kp**as]	'now'
	/ka**pk**iːla/	[ka**pk**iːla], [ka**kp**iːla]	'to drop'
	/di**pk**elkel/	[di**pk**elkel], [di**kp**elkel]	'to stumble'

A PK > KP metathesis is also found in some Bisayan languages. For example, according to Zorc (1977: 54), Aklanon has no surface *bg* clusters; historical **bg* and underlying /bg/ clusters surface with metathesis as *gb*. The two examples in (23) are given with Cebuano comparanda to show surface *bg* in another Bisayan language.

(23)	Cebuano	Aklanon	
	lí**bg**us	lí**gb**us	'mushroom'
	palí**bg**a	palí**gb**a (/pa-libug-a/)	'confuse him'

Finally, in two more poorly documented cases, a similar Klamath metathesis is cited by Barker (1964: 97) and a **pk > kp* change is suggested by the comparison of Wiyot *kbad* /kpat/ 'pitchwood' and Yurok *pkenc* 'pitch' from

Proto-Ritwan **pkanc* (Berman 1990: 432–3; Algonquian cognates show that the original sequence was **pk*).

Further support for our coarticulatory account comes from the common type of sound change in which a coarticulated labiovelar becomes a velar-labial sequence: *w* > *ɣ*^{w}, *w* > *g*^{w}; *p*^{ɣ} > *k͡p*, *b*^{ɣ} > *g͡b*, *m*^{ɣ} > *ŋ͡m*; *p*^{ɣ} > *k*^{w}, *b*^{ɣ} > *g*^{w}, *m*^{ɣ} > *ŋ*^{w}. Changes like the first two (e.g. *w* > *g*^{w}) are found in several early Indo-European languages, while the last six changes are found in some Oceanic languages. For instance, Proto-Oceanic **p*^{ɣ} , **b*^{ɣ} , and **m*^{ɣ} respectively are reflected as *k͡p*, *g͡b*, and *ŋ͡m* in Mwotlap and as *k*^{w}, *g*^{w}, and *ŋ*^{w} in Western Fijian (Ross 1998: 16–17). If the labiovelars are segments whose independent gestures are phonologically unordered, then their phonologisation as velar-labial sequences likely reflects the same phonetic factors referred to above: the velar closure prior to labial closure as the jaw closes, and simultaneous or nearly simultaneous closure having the percept of velic closure.

The unidirectionality of the Mokilese, Klamath, Bisayan, and Wiyot changes, as well as the variation characteristic of the first two cases, both support our view of these alternations as coarticulatory metathesis. As a coarticulatory effect KP > PK would not be expected, since coarticulated velar-labial stop clusters would be expected to maintain their linear sequencing properties or to show (perceptual) reanalysis to KK, PP, K, or P. The variation described for these phonological sequences parallels the variation inherent in other effects of gestural overlap, like the assimilatory effects noted for English in section 2.3.[22]

We turn now to coronal-noncoronal stop sequences. We know four examples of metathesis affecting such sequences, two of which are in closely related Austronesian languages. As suggested in section 2.3, the unidirectional nature of TP > PT and TK > KT changes seems to be related to the degree of gestural overlap in coronal-noncoronal clusters as opposed to noncoronal-coronal clusters. We hypothesise that, as with nearly simultaneous velar and labial closures, nearly simultaneous coronal and noncoronal closures provide a percept that is noncoronal. Such a case is found in the prehistory of ancient Greek, where **tk* > *kt* and **tp* > *pt* regularly, though the relevant clusters only occurred in the two words **k*^{w}*íd-pe* > **k*^{w}*ítpe* > *típte* (a particle) and **títkō* > *tíktō* 'I bear' (Lejeune 1972: 70; Rix 1992: 96). As with PK > KP changes, stops agree in all manner and laryngeal features and differ only with respect to place of articulation, with distinct articulators involved, allowing for gestural overlap.

Similar metatheses have occurred in the history of some Central Philippines languages. Blust (1979) discusses data from Tagalog and Cebuano Bisayan, languages in which T{P,K} > {P,K}T can also be viewed as a regular sound change. Representative Cebuano Bisayan data are cited in (24) from Blust (1979: 110).

(24)

Metathesis		No metathesis	
nm	i**num** : i**mn**-a 'drink'	mn	da**mán** : da**mn**-un 'talk, walk in one's sleep'
nŋ	tu**núŋ** : tu**ŋn**-a 'directly at a point'	ŋn	i**ŋún** : i**ŋn**-un 'say, tell'
tp	a**túp** : a**tp**-an, a**pt**-an 'roof'	pt	sá**put** : s-al-a**pt**-un 'bad temper'
tk	li**tik** : li**tk**-an, li**kt**-an 'snap the fingers'	kt	la**kát** : la**kt**-un 'walk'
		kp	da**kúp** : da**kp**-an 'arrest'

The Cebuano Bisayan facts are especially interesting because, as in Mokilese and Klamath, there are both metathesised and unmetathesised variants for obstruent clusters, suggesting that metathesis is directly related to degree of gestural overlap in the phonetic component. Metathesis with nasal clusters appears to be obligatory. It is also interesting to note that Bisayan languages show both PK > KP and T{P,K} > {P,K}T, since we are suggesting the same articulatory phonetic explanation for both phenomena.

A final case described by Blust (1979) is found in the historical phonology of Leti and Moa, two Austronesian languages of the Lesser Sunda group. Leti is well known for its synchronic CV metathesis alternations (van der Hulst and van Engelenhoven 1995; Hume 1998), which arose historically from the telescoping of final vowel copying and medial vowel syncope (Mills and Grima 1980; Blevins and Garrett 1998: 541–7). Blust shows that regular CC metathesis has also occurred just in case syncope results in a coronal-noncoronal cluster with shared manner and laryngeal features; cf., e.g., **saRman* > Leti *se**mn**a* 'outrigger float' vs ****inum*** > Leti ***emnu*** 'drink' and **ta**nem*** > Leti *to**mn**a*, Moa *ta**mn**a* 'to plant'.

Table 5.2. *Some attested regular sibilant metatheses*

Language	Metathesis	Source
Old English	sk > ks	(25) below
Faroese	sk > ks /___ t	Lockwood 1955: 23–4
Lithuanian	coronal fricative + velar stop > k + fricative / ___t	Seo and Hume 2001
Colloquial French	ks > sk /___ #	(26) below
Savoyard	*ts > st /#___	Ultan 1978
Classical Aramaic languages	*t + sibilant > sibilant + t / V___ V	Malone 1971, 1985, 1999
Ancient Greek	*dz > zd	Lejeune 1972: 113–16
Calabrian Greek	ps > sp	Rohlfs 1950: 74–6
Dutch	ps > sp	Stroop 1981– 2

3.4 Auditory metathesis

As discussed in section 2.4, auditory-stream decoupling leads to metathesis involving sibilants. Regular sibilant-stop and stop-sibilant metatheses are listed in table 5.2; we will discuss two examples here. The first is a well-documented case in the late West Saxon dialect of Old English (Weyhe 1908; Campbell 1959: 177–8; Luick 1921–40: 913–14; Jordan 1974: 168–70). In this dialect *sk* clusters regularly inverted their linear order and became *ks* clusters. The examples in (25a) show word-final metathesis; intervocalic metathesis is shown in (25b); and (25c) shows metathesis between a vowel and a sonorant.

(25)		Old English	Late West Saxon	
	a.	fro**sk**	fro**ks**	'frog'
		hu**sk**	hu**ks**	'insult'
		ma**sk**	ma**ks**	'meshes' (neut. pl.)
		tu**sk**	tu**ks**	'tooth' (cf. *tusk*)
	b.	a**sk**e	a**ks**e	'ash'
		aː**sk**ian	aː**ks**ian	'to ask'
		fi**sk**as	fi**ks**as	'fishes'
		hne**sk**ian	hne**ks**ian	'to soften'
		to**sk**a	to**ks**a	'frog'
		wa**sk**an	wa**ks**an	'to wash'
	c.	hor**sk** ('quick')	hor**ks**lic	'dirty'
		mu**sk**le	mu**ks**le	'mussel'
		θer**sk**an	θer**ks**an	'to thresh'
		θer**sk**old	θer**ks**old	'threshold'

In all the examples in (25), stress fell on the vowel immediately preceding the metathesising cluster (i.e. the root syllable). It would therefore be possible in principle to say that the change was restricted to immediately post-tonic position (though the more general statement also remains possible in principle).

By contrast, according to Grammont (1923: 73), a *ks* > *sk* change has occurred word-finally in colloquial French. The status of this example is somewhat unclear, since Grammont gives no phonetic or dialectological details, but in (26) we cite examples he mentions.

(26)	Standard	Colloquial	
	fi**ks**	fi**sk**	'fixed' (*fixe*)
	ly**ks**	ly**sk**	'luxury' (*luxe*)
	sɛ**ks**	sɛ**sk**	'sex' (*sexe*)
	a**ks**	a**sk**	'axis' (*axe*)
	feli**ks**	feli**sk**	'Félix' (*Félix*)

The Old English and French changes are apparently mirror images, *sk* > *ks* and *ks* > *sk*, with both occurring in final position. Citing several of the cases in table 5.2, Steriade (2001: 234–5) argues that all systematic ST > TS reorderings result in postvocalic stops, while all TS > ST metatheses result in prevocalic stops. In each case, she argues, stop cues are improved by providing a previously lacking VC or CV transition respectively; as she points out, French final stops are released. Yet the Old English change yields intervocalic *ks* clusters, and thus seems to contradict Steriade's claim.

We suggest that the crucial difference between the Old English and French examples may be prosodic. French has (weak) final stress, and this final stress could result in final sibilants being longer than medial ones; in (25), by contrast, the affected *sk* clusters were preceded by the strong initial stress of (Old) English.[23] The general pattern, we speculate, is that longer sibilants may induce a greater confusion effect on segmental order and are thus more likely to undergo metathesis with an adjacent stop. The apparent mirror-image effect that arises in comparing these two examples may thus be a by-product of independent differences in the languages' prosodic systems.

4 Phonetic explanations in phonology

At least two general issues emerge from the typological survey of metathesis in section 3. Before summarising our findings in section 4.3, we discuss gaps in the metathesis typology in section 4.1 and the general issue of directionality in section 4.2.

4.1 Typological gaps in metathesis patterns

Our approach to sound change predicts that certain logically possible metathesis types should not exist. One such type is the inversion of sequences consisting of a nasal and an oral stop. Given the articulatory requirements of nasal and oral stops, there is no way for nasality or orality to migrate across a neighbouring segment without directly affecting it; in such clusters assimilation is natural, but not metathesis. We thus predict that local nasal-obstruent metathesis should not occur as a sound change.[24] By contrast, the phonetic optimality approach predicts that at least intervocalic TN > NT metatheses are well motivated: TN clusters are rare while NT clusters are common (the only nongeminate clusters in some languages); and stop contrasts are relatively easy to perceive in prevocalic position.

For these reasons, the question of whether nasal-obstruent metathesis exists as a sound change offers a way of testing the two approaches. The literature does contain several cases where nasal-obstruent metatheses have been proposed, but in all such cases, we contend, other explanations not involving metathesis are available. The examples fall into two classes. The first class consists of three cases where there is no phonological process (neither a sound change nor a synchronic process) involving nasal-obstruent metathesis, and where metathesis has simply been erroneously proposed.[25] The second class consists of cases where there are synchronic metathesis patterns which, however, did not arise via any metathesis sound change. For these cases, restricted to a set of East Cushitic and South Omotic languages, we have shown elsewhere that the relevant patterns originated via a morphological process we call 'analogical morphophonology'.[26] In short, contrary to the assumptions of earlier work, careful analysis reveals that there are no cases in which local nasal-obstruent metathesis can be shown to have occurred as a sound change. This is predicted by our view of metathesis, while the phonetic optimisation view not only fails to predict it but predicts the opposite pattern for certain phonetic contexts.[27]

Other unattested metathesis sound changes include the inversion of velar-labial and noncoronal-coronal stop sequences, despite well-attested PK > KP and T{P,K} > {P,K}T changes (section 3.3). The phonetic optimisation model of sound change also apparently predicts the existence of *pg* > *gp* changes (n. 22) or a hypothetical $V_1npV_2 > V_1nV_2p$ metathesis (in which the place cues of a nonhomorganic nasal are optimised by intervocalic positioning).

In addition to metathesis types that should not occur as sound changes, our approach predicts the possible existence of some metathesis patterns that we have not yet encountered. Such predicted but unattested metatheses include *ɾV* > *Vɾ* (or the reverse). The articulation of taps typically involves transitory vowels preceding and following the brief constriction; if a phonetically predictable transition is reinterpreted as a full vowel, and a historical vowel is reinterpreted as a transition, metathesis will have occurred. (This potential metathesis type is not easily situated in our current typology.) The auditory-stream decoupling we have suggested as an explanation for sibilant metatheses (sections 2.4, 3.4) also predicts the possibility of similar metatheses for other noisy segment types such as [ɬ] and clicks, though, again, no such examples are yet known to us.

Apparent counterexamples to observed typological patterns highlight important provisos on the general role of phonetics in phonology. Such examples demonstrate that regular synchronic phonological metatheses are a superset of those that can arise through purely phonetic sound change, and thereby contribute to the literature on phonological alternations that do *not* reflect phonetic naturalness or phonological markedness.[28] Three known pathways other than sound change by which metathesis alternations may arise are listed in (27).

(27) Sources of metathesis alternations other than sound change
  a. Loan adaptation, e.g. Spanish (n. 25)
  b. Telescoping (epenthesis + deletion), e.g. Leti, Najdi Arabic (Blevins and Garrett 1998), Classical Mandaic (Malone 1995)
  c. Analogical morphophonology (Garrett and Blevins in press)

This list excludes erroneous analyses as well as cases whose diachrony remains unclear.

4.2 Directionality patterns

Many specific types of metathesis show directionality effects; a (schematic) metathesis XY > YX may be well attested while the reverse metathesis YX > XY is undocumented in the world's languages. Our model of sound change would fail to predict such asymmetries if all misperception patterns were symmetric, but in fact the various articulatory, acoustic, and perceptual factors underlying misperception and sound change are often intrinsically asymmetric (Guion 1996, 1998; Plauché 2001). Our metathesis survey in section 3 shows several directionality patterns that follow from our general model. We have already mentioned the unidirectionality of PK > KP and T{P,K} > {P,K}T metatheses, which follows from the intrinsic articulatory properties of stops of various places of articulation (sections 2.3, 3.3). Another pattern apparent from our research is that long-distance metathesis shifts liquids, pharyngeals, and laryngeal segments into relatively prominent (i.e. initial or stressed) positions but not into less prominent positions, a pattern that follows from the greater likelihood of not perceiving phonetic cues in positions where they are relatively hard to perceive (section 3.1). Similarly, a common metathesis pattern is AXY > AYX (or YXA > XYA), where A and X share features (and A and Y need not); examples include Rendille pharyngeal metathesis (13), Old English *r* metathesis (n. 14), and Le Havre French *r* metathesis (Blevins and Garrett 1998). The reverse pattern (e.g. AYX > AXY) is undocumented. This asymmetry ('like elements repel each other') is easily explained: an extended phonetic feature is less likely to be perceived in a position adjacent to a segment that possesses the same feature; it is more likely to be perceived (and then reinterpreted as originating) in a position farther away from such a segment.[29]

4.3 Summary and conclusion

We have had three main goals in this chapter. First, we have offered an empirically motivated typology of metathesis sound changes in the languages of the world. Diachronic metathesis sound changes are summarised in section 3.3, table 5.2, and Blevins and Garrett 1998; synchronic metathesis patterns for which we posit other origins are cited in notes 25–26, (27), and Garrett and Blevins in press. Our second goal, especially in section 2, has been to relate the typology of metathesis to the findings of experimental phonetics; this should be useful both to those who accept and to those who may doubt our overall argument. This overall argument has been our third goal: based on our analysis of metathesis and its phonetic roots, we contend that reinterpretations of the ambiguities in real speech are the main force driving sound change. In particular, the majority of attested regular historical metatheses in the world's languages can be explained as the result of phonetically natural sound changes in which coarticulation leads to a segment or feature being perceived in some nonhistorical position. Perceptual metathesis, compensatory metathesis, and coarticulatory metathesis are all of this type. We have also argued that sibilant-stop metatheses result from inherent perceptual difficulties in recovering sequential order from sibilant-stop and stop-sibilant clusters.

Just as phonetic studies can inform phonology, our phonological typology of metathesis suggests directions for further phonetic research. Attested compensatory metatheses suggest that directionality of V-to-V coarticulation in languages with unreduced vowels can be determined by the position of stress, with the unstressed vowel anticipated or persevering into the stressed vowel. Attested perceptual metatheses should encourage further research into possible long-domain effects of underdocumented phonetic features like aspiration, breathiness, and glottalisation. Examples of coarticulatory metathesis raise many interesting questions concerning complementarity between a percept of assimilation or deletion in CC clusters and a percept of metathesis. Finally, our account of attested sibilant/stop metatheses invokes a disruptive effect of sibilant noise on perception of linear order, and the percept of a short or abrupt sibilant (transition) as a stop: both hypotheses need to be rigorously tested by a range of perceptual experiments.

We conclude that metathesis can indeed be explained in a phonetically natural way based on the same assumptions required to understand other phonological phenomena. Future studies of perception and acquisition can test our hypotheses by investigating more precisely the conditions under which phonetic strings are phonologically ambiguous or subject to reanalysis. Insofar as our explanations are well founded, they suggest that phonetics determines emergent sound patterns. The typology of metathesis largely follows from convergent evolution, demonstrating the extent to which phonology is phonetically determined in the diachronic dimension.

References

Alexander, James D. (1985). *R*-metathesis in English: A diachronic account. *Journal of English Linguistics* 18: 33–40.

Amerman, James, Raymond Daniloff, and Kenneth Moll (1970). Lip and jaw coarticulation for the phoneme /æ/. *Journal of Speech and Hearing Research* 13: 147–61.

Anderson, Stephen R. (1981). Why phonology isn't 'natural'. *LI* 12: 493–539.

Anderson, Victoria B. and Ian Maddieson (1994). Acoustic characteristics of Tiwi coronal stops. *UCLA Working Papers in Phonetics* 87: 131–62.

Anttila, Raimo (1972). *An introduction to historical and comparative linguistics.* New York: Macmillan.

Bach, Emmon and Robert T. Harms (1972). How do languages get crazy rules? In Robert P. Stockwell and Ronald K. S. Macaulay (eds.), *Linguistic change and generative theory*. Bloomington: Indiana University Press, 1–21.

Bailey, Charles-James N. (1969). A possible explanation for an assimilation curiosity. *University of Hawaii Working Papers in Linguistics* 1: 187–98.

(1970). Toward specifying constraints on phonological metathesis. *LI* 1: 347–9.

Barker, M. A. R. (1964). *Klamath grammar*. (University of California Publications in Linguistics, 32) Berkeley: University of California Press.

Barry, M. C. (1984). Connected speech: Processes, motivations, models. *Cambridge Papers in Phonetics and Experimental Linguistics* 3: 1–16.

Bell-Berti, Fredericka and Katherine S. Harris (1976). Some aspects of coarticulation. *Haskins Laboratories Status Reports on Speech Research* SR-45/46. 197–204.

Benguerel, A.-P. and H. Cowan (1974). Coarticulation of upper lip protrusion in French. *Phonetica* 30: 41–55.

Berman, Howard (1990). New Algonquian-Ritwan cognate sets. *IJAL* 56: 431–4.

Bessell, Nicola J. (1992). *Towards a phonetic and phonological typology of post-velar articulation.* PhD dissertation, University of British Columbia.

(1997). St'át'imcets vowels and consonants: A phonetic study. *Papers for the 32nd International Conference on Salish and Neighbouring Languages, Port Angeles, Washington, August 7–9, 1997*, compiled by Timothy Montler, 64–85.

(1998a). Phonetic aspects of retraction in Interior Salish. In Ewa Czaykowska-Higgins and M. Dale Kinkade (eds.), *Salish language and linguistics*. The Hague: Mouton, 125–52

(1998b). Local and non-local consonant-vowel interaction in Interior Salish. *Phonology* 15: 1–40.

Bhaskararao, Peri (1980). *Koṇekor Gadaba: A Dravidian language.* Pune: Deccan College, Post-Graduate Institute.

Blevins, Juliette (1993). Klamath laryngeal phonology. *IJAL* 59: 237–79.

(1997). Rules in Optimality Theory: Two case studies. In Iggy Roca (ed.), *Derivations and constraints in phonology.* Oxford: Clarendon Press, 227–60.

(2002). The independent nature of phonotactic constraints: An alternative to syllable-based approaches. In Caroline Féry and Ruben van de Vijver (eds.), *The syllable in Optimality Theory*. Cambridge: Cambridge University Press, 375–403.

Blevins, Juliette and Andrew Garrett (1993). The evolution of Ponapeic nasal substitution. *Oceanic Linguistics* 32: 199–236.

(1998). The origins of consonant-vowel metathesis. *Language* 74: 508–56.

Bloch, Jules (1915). *La formation de la langue marathe.* (Bibliothèque de l'Ecole des hautes études, sciences historiques et philologiques, 215) Paris: E. Champion.

Blust, Robert (1979). Coronal-noncoronal consonant clusters: New evidence for markedness. *Lingua* 47: 101–17.

Boersma, Paul (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives.* The Hague: Holland Academic Graphics.

Bregman, A. S. (1990). *Auditory scene analysis.* Cambridge, MA: MIT Press.

Brilioth, Börje (1913). *A grammar of the dialect of Lorton (Cumberland), historical and comparative.* London: Oxford University Press.

Browman, Catherine P. and Louis Goldstein (1990). Tiers in articulatory phonology, with some implications for casual speech. In John Kingston and Mary E. Beckman (eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*. Cambridge: Cambridge University Press, 341–76.

Byrd, Dani (1996). Influences of articulatory timing in consonant sequences. *JPh* 24: 209–44.

Campbell, A. (1959). *Old English grammar.* Oxford: Clarendon Press.

Card, Elizabeth (1979). *A phonetic and phonological study of Arabic emphasis*. PhD dissertation, Cornell University.

Choi, John D. (1992). *Phonetic underspecification and target interpolation: An acoustic study of Marshallese vowel allophony*. UCLA Working Papers in Phonetics, 82.

Cohn, Abigail (1990). *Phonetic and phonological rules of nasalisation*. (UCLA Working Papers in Phonetics, 76) PhD dissertation, University of California, Los Angeles.

Connell, Bruce (1994). The structure of labial-velar stops. *JPh* 22: 441–76.

Crews, C. M. (1935). *Recherches sur le judéo-espagnol dans les pays balkaniques.* (Société de Publications Romanes et Françaises, 16) Paris: E. Droz.

Cutting, J. E. (1982). Plucks and bows are categorically perceived, sometimes. *Perception & Psychophysics* 31: 462–76.

Cutting, James E. and B. Rosner (1974). Categories and boundaries in speech and music. *Perception and Psychophysics* 16: 564–70.

Dart, Sarah (1991). *Articulatory and acoustic properties of apical and laminal articulations.* UCLA Working Papers in Phonetics, 79.

Dave, Radhekant (1977). Retroflex and dental consonants in Gujerati: A palatographic and acoustical study. *Annual Report of the Institute of Phonetics, University of Copenhagen* 11: 27–156.

Delattre, P. C., A. M. Liberman, and F. S. Cooper (1962). Formant transitions and loci as acoustic correlates of place of articulation in American fricatives. *Studia Linguistica* 16: 104–21.

Dougherty, Brian (1993). *The acoustic-phonetic correlates of Cayuga word stress*. PhD dissertation, Harvard University.

Dumenil, Annie (1987). A rule-account of metathesis in Gascon. *Lingvisticae Investigationes* 11: 81–113.

Durie, Mark (1985). *A grammar of Acehnese on the basis of a dialect of North Aceh.* Dordrecht: Foris.

Elworthy, Frederic Thomas (1875). *The dialect of west Somerset*. (English Dialect Society, Series D, Miscellaneous, 7) London: Trübner. (Also printed in *Transactions of the Philological Society* 1875–6: 197–272.)

Evans, Nick (1995). Current issues in the phonology of Australian languages. In John A. Goldsmith (ed.), *The handbook of phonological theory*. Cambridge, MA: Blackwell, 723–61.

Fay, Warren H. (1966). *Temporal sequence in the perception of speech.* (Janua Linguarum, Series minor, 45) The Hague: Mouton.

Flemming, Edward S. (1996). Laryngeal metathesis and vowel deletion in Cherokee. In Pamela Munro (ed.), *Cherokee papers from UCLA*. (UCLA Occasional Papers in Linguistics, 16) Los Angeles: Department of Linguistics, UCLA, 23–44.

Foster, Michael K. (1982). Alternating weak and strong syllables in Cayuga words. *IJAL* 48: 59–72.

Fougeron, Cécile and Sun-Ah Jun (1998). Rate effects on French intonation: Prosodic organisation and phonetic realisation. *JPh* 26: 45–69.

Fougeron, Cécile and Patricia A. Keating (1997). Articulatory strengthening at edges of prosodic domains. *JASA* 101: 3728–40.

Fowler, Carol A. (1981a). A relationship between coarticulation and compensatory shortening. *Phonetica* 38: 35–50.

(1981b). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research* 24: 127–39.

Fowler, Carol A. and Mary R. Smith (1986). Speech perception as vector analysis: An approach to the problem of invariance and segmentation. In Joseph S. Perkell and Dennis H. Klatt (eds.), *Invariance and variability in speech processes*. Hillsdale, NJ: Lawrence Erlbaum Associates, 123–39.

Garrett, Andrew (1991–93). A note on the morphosyntax of Lycian and Anatolian possession. *Die Sprache* 35: 155–62.

Garrett, Andrew and Juliette Blevins (in press). Analogical morphophonology. In Kristin Hanson and Sharon Inkelas (eds.), *The nature of the word*. Cambridge, MA: MIT Press.

Geisler, Hans (1994). Metathese im Sardischen. *Vox Romanica* 53: 106–37.

Gerken, LouAnn, W. D. Murphy, and R. N. Aslin (1995). Three- and four-year-olds' perceptual confusions for spoken words. *Perception and Psychophysics* 57: 475–86.

Gerstman, L. (1957). *Cues for distinguishing among fricatives, affricates and stop consonants*. PhD dissertation, New York University.

Ghazeli, Salem (1977). *Back consonants and backing coarticulation in Arabic*. PhD dissertation, University of Texas at Austin.

Grammont, Maurice (1905–1906). La métathèse dans le parler de Bagnères-de-Luchon. *Mémoires de la Société de Linguistique de Paris* 13: 73–90.

(1908). La métathèse en arménien. In *Mélanges de linguistique offerts à M. Ferdinand de Saussure*. Paris: Honoré Champion, 231–43.

(1923). L'interversion. In ANTIΔΩPON: *Festschrift Jacob Wackernagel zur Vollendung des 70. Lebensjahres am 11. Dezember 1923*. Göttingen: Vandenhoeck & Ruprecht, 72–7.

(1950). *Traité de phonétique.* 4th edn. Paris: Delagrave.

Guion, Susan Guignard (1996). *Velar palatalization: Coarticulation and phonologization.* PhD dissertation, University of Texas at Austin.

(1998). The role of perception in the sound change of velar palatalization. *Phonetica* 55: 18–52.

Hansson, Gunnar (2001). *Theoretical and typological issues in consonant harmony.* PhD dissertation, University of California, Berkeley.

Harrison, Sheldon P. (1976). *Mokilese reference grammar.* Honolulu: University of Hawaii Press.

Hawkins, S. and Andrew Slater (1994). Spread of CV and V-to-V coarticulation in British English: Implications for the intelligibility of synthetic speech. *Proceedings of ICSLP 94* (1): 57–60.

Hayward, Dick (1984). *The Arbore language: A first investigation, including a vocabulary.* (Kuschitische Sprachstudien, 2) Hamburg: H. Buske.

Heine, Bernd (1976). Notes on the Rendille language (Kenya). *Afrika und Übersee* 59: 176–223.

(1978). The Sam languages: A history of Rendille, Boni and Somali. *Afroasiatic Linguistics* 6: 23–115.

(1980). *The non-Bantu languages of Kenya.* (Language and Dialect Atlas of Kenya, 2) Berlin: Dietrich Reimer.

(1981). *The Waata dialect of Oromo: Grammatical sketch and vocabulary.* (Language and Dialect Atlas of Kenya, 4) Berlin: Dietrich Reimer.

Herbert, Robert K. (1986). *Language universals, markedness theory, and natural phonetic processes.* Berlin: Mouton de Gruyter.

Hetzron, Robert (1971). Internal labialisation in the *tt*-group of Outer South-Ethiopic. *Journal of the American Oriental Society* 91: 192–207.

(1977). *The Gunnän-Gurage languages.* (Studi di semitistica e del vicino oriente antico, ricerche, 12) Naples: Istituto orientale di Napoli.

Hinton, Leanne and Margaret Langdon (1976). Object-subject pronominal prefixes in La Huerta Diegueño. In Margaret Langdon and Shirley Silver (eds.), *Hokan studies: Papers from the First Conference on Hokan Languages*. The Hague: Mouton, 113–28.

Hirsh, I. J. (1959). Auditory perception of temporal order. *JASA* 31: 759–67.

Hoberman, Robert D. (1989). Parameters of emphasis: Autosegmental analyses of pharyngealization in four languages. *Journal of Afroasiatic Languages* 2: 73–97.

(1995). Current issues in Semitic phonology. In John A. Goldsmith (ed.), *The handbook of phonological theory*. Cambridge, MA: Blackwell, 839–47.

Hock, Hans Henrich (1985). Regular metathesis. *Linguistics* 23: 529–46.

Hume, Elizabeth (1997). The role of perceptibility in consonant/consonant metathesis. *WCCFL* 17: 293–307.

(1998). Metathesis in phonological theory: The case of Leti. *Lingua* 104: 147–86.

(2001). Metathesis: Formal and functional considerations. In Elizabeth Hume, Norval Smith and Jeroen van de Weijer (eds.), *Surface syllable structure and segment sequencing*. (HIL Occasional Papers) Leiden: HIL, 1–25.

Hume, Elizabeth and Keith Johnson (eds.) (2001). *The role of speech perception in phonology.* San Diego: Academic Press.

Hume, Elizabeth, Norval Smith, and Jeroen van de Weijer (eds.) (2001). *Surface syllable structure and segment sequencing*. (HIL Occasional Papers) Leiden: HIL.

Hyman, Larry M. (1972). Nasals and nasalisation in Kwa. *Studies in African Linguistics* 3: 167–205.

(2001). The limits of phonetic determinism in phonology: *NC revisited. In Elizabeth Hume and Keith Johnson (eds.), *The role of speech perception in phonology*. San Diego: Academic Press, 141–85.

Jaberg, K. and J. Jud (1928–40). *Sprach- und Sachatlas Italiens und der Südschweiz.* 9 vols. Zofingen: Ringier.

Jongman, Allard (1989). Duration of frication noise required for identification of English fricatives. *JASA* 85: 1718–25.

Jordan, R. (1974). *Handbook of Middle English grammar: Phonology.* Translated and revised by E. J. Crook. The Hague: Mouton.

Kahn, Margaret (1976). How abstract is pharyngology: Evidence from Kurmanji. *CLS* 12: 313–20.

Keating, Patricia (1988). Underspecification in phonetics. *Phonology* 5: 275–92.

Keating, Patricia and Sheila E. Blumstein (1978). Effects of transition length on the perception of stop consonants. *JASA* 64: 57–64.

Kelly, John (1989). On the phonological relevance of some non-phonological elements. In Tamás Szende (ed.), *Proceedings of the Speech Research '89 International Conference, June 1–3, 1989, Budapest*. (Magyar Fonetikai Füzetek, 21) Budapest: Linguistics Institute of the Hungarian Academy of Sciences, 56–9.

Kelly, John and John K. Local (1986). Long-domain resonance patterns in English. *International Conference on Speech Input/Output; Techniques and Applications. IEE Conference Publication* 258: 304–9.

Kingston, John and Mary E. Beckman (eds.) (1990). *Papers in laboratory phonology I: Between the grammar and physics of speech.* Cambridge: Cambridge University Press.

Koch, Harold (1997). Pama-Nyungan reflexes in Arandic languages. In Darrell T. Tryon and Michael Walsh (eds.), *Boundary rider: Essays in honour of Geoffrey O'Grady* (Pacific Linguistics, C-136) Canberra: Research School of Pacific Studies, The Australian National University, 271–302.

Kohler, K. J. (1976). Die Instabilität wortfinaler Alveolarplosive im Deutschen: Eine elektropalatographische Untersuchung. *Phonetica* 33: 1–30.

Krishnamurti, Bhadriraju (1955). The history of vowel-length in Telugu verbal bases. *Journal of the American Oriental Society* 75: 237–52.

(1961). *Telugu verbal bases: A comparative and descriptive study.* (University of California Publications in Linguistics, 24) Berkeley: University of California Press.

(1978). Areal and lexical diffusion of sound change: Evidence from Dravidian. *Lg* 54: 1–20.

Kuipers, Aert H. (1974). *The Shuswap language: Grammar, texts, dictionary.* The Hague: Mouton.

Ladefoged, Peter (1993). *A course in phonetics.* Fort Worth, TX: Harcourt Brace Jovanovich.

(2001). *Vowels and consonants: An introduction to the sounds of languages.* Malden, MA: Blackwell.

Ladefoged, Peter and Ian Maddieson (1996). *The sounds of the world's languages.* Oxford: Blackwell.

Ladefoged, Peter, Ian Maddieson, and Michel T. T. Jackson (1988). Investigating phonation types in different languages. In O. Fujimura (ed.), *Vocal physiology: Voice production, mechanisms and functions*. Raven: New York, 297–317.

Langdon, Margaret (1976). Metathesis in Yuman languages. *Lg* 52: 866–83.

Laycock, Donald (1982). Metathesis in Austronesian: Ririo and other cases. In Amran Halim, Lois Carrington and S. A. Warm (eds.), *Papers from the Third International Conference on Austronesian Linguistics, vol. 1, Currents in Oceanic*. (Pacific Linguistics, C-74) Canberra: Research School of Pacific Studies, Australian National University, 269–81.

Lejeune, Michel (1972). *Phonétique historique du mycénien et du grec ancien.* New edn. (Tradition de l'humanisme, 9) Paris: Klincksieck.

Liberman, A. M. (1970). The grammars of speech and language. *Cognitive Psychology* 1: 301–23.

Lipski, John M. (1990/1991). Metathesis as template matching: A case study from Spanish. *Folia Linguistica Historica* 11 (1990): 89–104 = 12 (1991): 127–45.

Lockwood, W. B. (1955). *An introduction to modern Faroese.* Copenhagen: Ejnar Munskgaard.

Lubker, J. and T. Gay (1982). Anticipatory labial coarticulation: Experimental, biological and linguistic variables. *JASA* 71(2): 437–48.

Luick, Karl (1921–40). *Historische Grammatik der englischen Sprache.* Leipzig: Tauchnitz.

Lydall, Jean (1976). Hamer. In Marvin Lionel Bender (ed.), *The non-Semitic languages of Ethiopia*. East Lansing, MI: Michigan State University, 393–438.

(1988). Gender, number, and size in Hamar. In Marianne Bechhaus-Gerst and Fritz Serzisko (eds.), *Cushitic-Omotic: Papers from the International Symposium on Cushitic and Omotic languages, Cologne, January 6–9, 1986*. Hamburg: Helmut Buske Verlag, 77–90.

Magen, Harriet S. (1984). Vowel-to-vowel coarticulation in English and Japanese. *JASA*, Supplement 75, S41.

Majors, Tivoli Jane (1998). *Stress dependent harmony: Phonetic origins and phonological analysis.* PhD dissertation, University of Texas at Austin.

Malkiel, Yakov (1950). La derivación de *rebelde*, *rebeldia* y las fuentes del grupo de consonantes *-ld-* en iberorrománico. In Patronato Marcelino Menendez y Pelayo (ed.), *Estudios dedicados a Menendez Pidal*. Madrid: n.p., vol. 1, 91–124.

Malone, Joseph (1971). Systematic metathesis in Mandaic. *Lg* 47: 394–415.

(1985). Classical Mandaic radical metathesis, radical assimilation, and the devil's advocate. *General Linguistics* 25: 92–122.

(1995). La circonscription prosodique en mandéen classique. *Langues Orientales Anciennes* 5–6: 233–57.

(1999). Metathesis and antiwedging in Classical Mandaic. *General Linguistics* 36: 227–55.

Manuel, Sharon Y. (1987). *Acoustic and perceptual consequences of vowel-to-vowel coarticulation in three Bantu languages*. PhD dissertation, Yale University.

Manuel, Sharon Y. and Rena A. Krakow (1984). Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories Status Reports on Speech Research* 77/78: 69–78.

Matras, Yaron (2002). *Romani: A linguistic introduction.* Cambridge: Cambridge University Press.

Mattina, Anthony (1979). Pharyngeal movement in Colville and related phenomena in the Interior Salish languages. *IJAL* 45: 17–24.

(1999). Anticipatory and progressive vowel lowering in Interior Salish, with notes on consonant retraction. In Mercedes Q. Hinkson (comp.), *Papers for the 34th International Conference on Salish and Neighbouring Languages*. Vancouver: Secwepemc Cultural Education Society, Simon Fraser University, 185–214.

McCarthy, John J. (1991). Synchronic rule inversion. In Laurel A. Sutton, Christopher Johnson, and Ruth Shields (eds.), *BLS 1991: General session and parasession on the grammar of event structure*. Berkeley, CA: Berkeley Linguistics Society, 192–207.

(2000). The prosody of phase in Rotuman. *NLLT* 18: 147–97.

McGregor, William (1990). *A functional grammar of Gooniyandi.* Amsterdam: John Benjamins.

Menéndez Pidal, Ramón (1958). *Manual de gramática histórica española.* 10th edn. Madrid: Espasa-Calpe.

Mills, Roger F. and John Grima (1980). Historical developments in Lettinese. In Paz Buenaventura Naylor (ed.), *Austronesian studies: Papers from the Second Eastern Conference on Austronesian Languages*. (Michigan Papers on South and Southeast Asia, 15) Ann Arbor: University of Michigan, Center for South and Southeast Asian Studies, 273–83.

Molinu, Lucia (1999). Métathèse et variation en sarde. *Cahier de Grammaire* 24: 153–81.

Newman, Stanley (1944). *Yokuts language of California.* (Viking Fund Publications in Anthropology, 2) New York: n.p.

Newton, David E. (1996). The nature of resonance in English: An investigation into lateral articulations. *York Papers in Linguistics* 17: 167–90.

Ohala, John J. (1974). Experimental historical phonology. In John M. Anderson and Charles Jones (eds.), *Historical linguistics.* Amsterdam: North-Holland, vol. 2, 353–89.

(1981). The listener as a source of sound change. In Carrie S. Masek, Roberta A. Hendrick, and Mary Frances Miller (eds.), *Papers from the Parasession on Language and Behavior, Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, 178–203.

(1993). The phonetics of sound change. In Charles Jones (ed.), *Historical linguistics: Problems and perspectives.* London: Longman, 237–78.

Oomen, Antoinette (1981). Gender and plurality in Rendille. *Afroasiatic Linguistics* 8/1: 35–75.

Penny, Ralph (1991). *A history of the Spanish language.* Cambridge: Cambridge University Press.

Pickett, J. M. (1958). Perception of compound consonants. *Language and Speech* 1: 288–304.

Plauché, Madelaine C. (2001). *Acoustic cues in the directionality of stop consonant confusions.* PhD dissertation, University of California, Berkeley.

Ravnæs, Erling (1991). *The chronology of the sound changes from Proto-Indo-European to Classical Armenian.* PhD dissertation, University of Oslo.

Recasens, Daniel (1984). Vowel-to-vowel coarticulation in Catalan VCV sequences. *JASA* 76: 1624–35.

(1987). An acoustic analysis of V-to-C and V-to-V coarticulatory effects in Catalan and Spanish VCV sequences. *JPh* 15: 299–312.

Reichard, Gladys A. (1925). *Wiyot grammar and texts.* (University of California Publications in American Archaeology and Ethnology, 22) Berkeley: University of California Press.

Rix, Helmut (1992). *Historische Grammatik des Griechischen: Laut- und Formenlehre.* Darmstadt: Wissenschaftliche Buchgesellschaft.

Robins, R. H. (1958). *The Yurok language: Grammar, texts, lexicon.* (University of California Publications in Linguistics, 15) Berkeley: University of California Press.

Rohlfs, Gerhard (1924). *Griechen und Romanen in Unteritalien: Ein Beitrag zur Geschichte der unteritalienischen Gräzität.* (Biblioteca dell' 'Archivum Romanicum', Serie 2, 7) Geneva: Leo S. Olschki.

(1933). *Scavi linguistici nella Magna Grecia.* Trans. Bruno Tomasini. Rome: Collezione Meridionale Editrice.

(1950). *Historische Grammatik der unteritalienischen Gräzität.* (Sitzungsberichte der Bayerischen Akademie der Wissenschaften, Philosophisch-historische Klasse, 4) Munich: Verlag der Bayerischen Akademie der Wissenschaften.

(1964). *Lexicon graecanicum Italiae inferioris: Etymologisches Wörterbuch der unteritalienischen Gräzität.* 2nd edn. Tübingen: Max Niemeyer.

Rose, Sharon (1997). *Theoretical issues in comparative Ethio-Semitic phonology and morphology.* PhD dissertation, McGill University.

Ross, Malcolm (1998). Proto-Oceanic phonology and morphology. In Malcolm Ross, Andrew Pawley, and Meredith Osmond (eds.), *The lexicon of Proto Oceanic*, vol. 1, *Material culture*. (Pacific Linguistics, C-152) Canberra: Australian National University, 15–35.

Sala, Marius (1970). *Estudios sobre el judeoespañol de Bucarest.* (Universidad Nacional Autónoma de México, Facultad de filosofia y letras, ediciones filosofia y letras, 73) Mexico city: Universidad Nacional Autónoma de México.

(1971). *Phonétique et phonologie du judéo-espagnol de Bucarest.* (Janua Linguarum, Series Practica, 142) The Hague: Mouton.

Sawyer, Jesse Otto (1959). *Achinese historical phonology.* PhD dissertation, University of California, Berkeley.

Schlee, Günther (1978). *Sprachliche Studien zum Rendille.* (Hamburger Philologische Studien, 46) Hamburg: Helmut Buske.

Schmidt, Rüdiger (1981). *Grammatik des klassisch-armenischen mit sprachvergleichenden Erläuterungen.* (Innsbrucker Beiträge zur Sprachwissenschaft, 32) Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Seo, Misun and Elizabeth Hume (2001). A comparative account of metathesis in Faroese and Lithuanian. In Elizabeth Hume, Norval Smith, and Jeroen van de Weijer (eds.), *Surface syllable structure and segment sequencing.* (HIL Occasional Papers) Leiden: HIL, 210–29.

Seymour, Peter J. (1985). *The golden woman: The Colville narrative of Peter J. Seymour*. (ed. Anthony Mattina, trans. Anthony Mattina and Madeline deSautel) Tucson: University of Arizona Press.

Sim, Ronald J. (1981). Morphophonemics of the verb in Rendille. *Afroasiatic Linguistics* 8/1: 1–33.

Steriade, Donca (2001). Directional asymmetries in place assimilation: A perceptual account. In Elizabeth Hume and Keith Johnson (eds.), *The role of speech perception in phonology.* New York: Academic Press, 219–50.

Stevens, Kenneth N. and Sheila Blumstein (1975). Quantal aspects of consonant production and perception: A study of retroflex consonants. *JPh* 3: 215–33.

Stroop, Jan (1981–82). Metathesis van *s* en *p*. *Spektator* 11: 224–48.

Subak, Jul (1906). Zum Judenspanischen. *Zeitschrift für romanisches Philologie* 30: 129–85.

Subrahmanyam, P. S. (1983). *Dravidian comparative phonology.* (Annamalai University, Department of Linguistics, Publication No. 74) Annamalainagar: Annamalai University.

Tryon, Darrell T. (1995). *Comparative Austronesian dictionary: An introduction to Austronesian studies*. Berlin: Mouton de Gruyter.

Tunley, Alison (1999). *Coarticulatory influences of liquids on vowels in English*. Doctoral dissertation, University of Cambridge.

Turner, R. L. (1962–66). *A comparative dictionary of the Indo-Aryan languages*. London: Oxford University Press.

Tuttle, Edward F. (1997). Preferential and pseudo-metathesis (in Italo-Romance dialects). In Pier Marco Bertinetto, Livio Gaeta, Georgi Jetchev, and David Michaels (eds.), *Certamen phonologicum III: Papers from the Third Cortona Phonology Meeting, April 1996*. Turin: Rosenberg & Sellier, 267–91.

Ultan, Russell (1978). A typological view of metathesis. In Joseph H. Greenberg (ed.), *Universals of human language*, vol. 2, *Phonology*. Stanford, CA: Stanford University Press, 367–402.

van der Hulst, Harry and Aone van Engelenhoven (1995). Metathesis effects in Tutukeian–Letinese. In Harry van der Hulst and Jeroen van de Weijer (eds.), *Leiden in last*. (HIL Phonology Papers, 1) The Hague: Holland Academic Graphics, 243–67.

Vennemann gen. Nierfeld, Theo (1972). Rule inversion. *Lingua* 29: 209–42.

(1988). *Preference laws for syllable structure and the explanation of sound change*. Berlin: Mouton de Gruyter.

(1996). The development of reduplicating verbs in Germanic. In Irmengard Rauch and Gerald F. Carr (eds.), *Insights in Germanic linguistics II: Classic and contemporary*. Berlin: Mouton de Gruyter, 297–336.

Walker, Rachel (2000). *Nasalisation, neutral segments, and opacity effects*. New York: Garland.

Wang, William S.-Y. (1968). Vowel features, paired variables, and the English vowel shift. *Lg* 44: 695–708.

Watson, Janet C. E. (1999). The directionality of emphasis spread in Arabic. *LI* 30: 289–300.

West, Paula (1999). Perception of distributed coarticulation properties of English /l/ and /ɹ/. *JPh* 27: 405–26.

(2000). *Long-distance coarticulatory effects of English /l/ and /r/*. DPhil thesis, University of Oxford.

Weyhe, Hans (1908). Zur Palatisierung von in- und auslautendem *sk* im Altenglischen. *Englische Studien* 39: 161–68.

Whalen, D. H. (1991). Perception of the English [s]-[ʃ] distinction relies on fricative noises and transitions, not on brief spectral slices. *JASA* 90: 1776–85.

Williamson, Kay (1973). More on nasals and nasalisation in Kwa. *Studies in African Linguistics* 4: 115–38.

Zoll, Cheryl. n.d. Positional asymmetries and licensing. Unpublished paper downloaded from Rutgers Optimality Archive, June 2002 (http://roa.rutgers.edu).

Zorc, David Paul (1977). *The Bisayan dialects of the Philippines: Subgrouping and reconstruction*. (Pacific Linguistics, Series C, 44) Canberra: Australian National University.

Zsiga, Elizabeth C. (1994). Acoustic evidence for gestural overlap in consonant sequences. *JPh* 22: 121–40.

Notes

For valuable comments, criticism, and discussion we are grateful to Tom Field, Bruce Hayes, Larry Hyman, Sharon Inkelas, Joe Malone, Ian Maddieson, Donca Steriade, and participants in the second author's Fall 1999 seminar at Berkeley. We use the following abbreviations for segment classes: C = consonant, G = glide, K = velar stop, N = nasal, P = labial stop, S = sibilant, T = stop or coronal stop (depending on context), and V = vowel.

1. This corresponds to both 'interversion' and 'metathesis' (i.e. respectively local and nonlocal metathesis) as defined in some earlier work (e.g. Grammont 1950).
2. Steriade is discussing stop-sibilant metathesis in particular (cf. section 3.4 below), and her general claim may be restricted to that subtype. Note that it is not true in principle that confusability is symmetric. For instance, Guion (1996, 1998) has shown that English /ki/ is misperceived as /t͡ʃi/ (in certain experimental contexts) significantly more often than /t͡ʃi/ is misperceived as /ki/; she argues that this asymmetry is related to the well-known asymmetry in sound change whereby *ki* > *t͡ʃi* is common but *t͡ʃi* > *ki* is not; see also Plauché 2001. It is not the case, as we will show in detail for metathesis, that an asymmetry in sound patterns or sound changes necessarily disproves a misperception account of their origins.
3. The possibilities increase if pharyngealisation can be associated with multisegmental phonological domains, or if multiple pharyngeal glides are posited at the phonological level.
4. See also Newton 1996 for references to other phonetic studies of English liquids and their contrast.
5. Retroflex harmonies (dental > retroflex C assimilation across intervening V) are also documented in several Dravidian languages (Subrahmanyam 1983: 361–3). For general discussion of retroflex and other consonant harmonies, see Hansson 2001.
6. The Cayuga alternations clearly involve spreading of laryngeal features, though metathesis as a phonological process may not yet be complete. It is irrelevant that Cayuga lacks /hʔ/ or /ʔh/ clusters, since sound change need not be structure-preserving; Cayuga laryngeal metathesis is demonstrably not structure-preserving (Blevins and Garrett 1998: 519–20).
7. Bailey's (1969, 1970) original argument was that the assimilation and metathesis facts are evidence for the marked status of coronal-noncoronal clusters. Blust provides further evidence, and considers the possibility that this arises from perceptual factors – in particular, a backward masking effect of noncoronal on coronal consonants (1979: 116). After considering weaknesses of the perceptual account, Blust (1979: 117) anticipates the kind of analysis we present when he concludes that 'on the basis of present evidence, it seems best to assume that the facts in question result from an innate limitation on the production of speech'.
8. We are grateful to Bruce Hayes for helping us formulate these statements. The common misperception of the positioning of [s] is demonstrated in a perception experiment by Ladefoged (2001: 175), with accompanying CD. Consistent with Ladefoged's experimental results, there is at least one example of long-distance sibilant metathesis in Ilokano, a Northern Philippine language (Anttila 1972: 75; Tryon 1995); a representative example is Ilokano ***saaŋit*** 'weep' vs Aklanon ***taaŋis***.

9. All published reports of perceptual metathesis known to us are cited here or in other general studies of metathesis (Grammont 1950; Ultan 1978; Hock 1985; Hume 1997, 2001; Blevins and Garrett 1998), or are included in this list of examples: liquid metatheses in Bisayan (Zorc 1977: 54), Chawchila Yokuts (Newman 1944: 32), Old English dialects (Alexander 1985), Somerset English (Elworthy 1875: 74–5), West Midland Middle English (Jordan 1974: 158), Koṇekor Gadaba (Bhaskararao 1980: 12–13), Oromo dialects (Heine 1980, 1981), Old Spanish (Malkiel 1950), and South-Central Dravidian 'apical displacement' (Krishnamurti 1955, 1961, 1978; Subrahmanyam 1983: 225–44); glide or high-vowel metatheses in Calabrian Greek (Rohlfs 1950: 82) and Old Spanish (Menéndez Pidal 1958; Penny 1991); laryngeal metatheses in Acehnese (Sawyer 1959: 143–4, 148; Durie 1985: 95–6, 146–7), Arbore (Hayward 1984: 72), Arnhem Land languages (Evans 1995: 738–9), Bisayan (Zorc 1977: 53), La Huerta Diegueño (Hinton and Langdon 1976), Ener and Endegeñ (Hetzron 1977: 39), Western Munster Irish (Malone 1971: 413 n. 69), Kiliwa (Langdon 1976: 874), Lycian (Garrett 1991–93), Classical Mandaic (Malone 1971, 1985), and Yokuts languages (Newman 1944: 15); and pharyngeal metathesis in Kurmanji Kurdish (Kahn 1976).
10. Intermediate forms in (6) are meant to clarify the historical developments; they may not be accurate, since the relative chronology of some sound changes is unknown.
11. We believe it is uncontroversial that the metathesis postdates Spanish *d* > ð lenition (i.e. that lenition occurred before 1492). For examples from other dialects, see Crews 1935 and Sala 1970: 171–2; 1971: 154.
12. On similar Romance patterns, see Lipski 1990/1991, Tuttle 1997, and, for the well-known example of Luchonnais Gascon, Grammont 1905–1906, Dumenil 1987, and Blevins and Garrett 1998; we mistakenly called this a French dialect. Such long-distance liquid metathesis is dubbed 'slope displacement' by Vennemann (1988, 1996), who calls its cause 'straightforward: like all language change, slope displacement is language improvement' (1996: 318).
13. Forms are cited from the dialect of Bova (unmarked) or Otranto ('O'); words originally borrowed from Latin are so noted ('L'; many South Italian Greek words with obstruent-*l* clusters are loanwords).
14. Compare the Northumbrian Old English metathesis whereby *Vr* > *rV*/__ *h*, e.g. *berht* > *breht* 'bright' (Luick 1921–40: 917–18); it is well established that Old English *r* shared a velar or other back constriction with *h* ([x]). In (13) and (14) note that the symbol ħ follows Heine, who notes that it varies for some speakers with ʕ, which he classifies as a stop. The phonetics of these symbols is unclear from his discussion.
15. 'S' citations in (15) refer to sentence numbers in Seymour 1985. Bessell 1998a, 1998b and Mattina 1999 are the most recent discussions of pharyngeal movement in Nxilxcín and related processes in other Interior Salish languages.
16. Nxilxcín pharyngeals are called 'difficult to hear' in one phonetic study (Bessell 1992: 159).
17. If no glottalisation shift is possible, the glottalised sonorant surfaces as a *Ceʔ* sequence (e.g. *ʔíχʷ-**le**ʔp* 'broom' < *ʔíχʷ-* 'sweep' + -í ***l'****əp*).
18. The Sardinian long-distance metathesis shows comparable blocking effects; cf., e.g., ***fr**enúku* < *fēnuk**l**um < Latin *fēnuku**l**um* 'fennel', ***pr**eðúku* < *pēduk**l**um <

*pēduku**l**um* ‘louse’ vs *fa**rr**ík**r**u* < **fa**rr**ik**l**um* < *fa**rr**iku**l**um* ‘spelt cake’, *ke**rr**ík**r**u* < **ke**rn**ik**l**um* < *ke**rn**iku**l**um* ‘sieve’ (Geisler 1994: 112, 123).

19. A referee suggests that we might not expect Nxilxcín pharyngealisation to shift across a segment requiring tongue body fronting, e.g. perhaps the glide *y* in (15b), and that Secwepemctsín glottalisation might be expected not to shift across the fricative cluster *sx* in the second example in (16c). With respect to Nxilxcín, note that precisely comparable pharyngealisation harmonies are well documented cross-linguistically, even among cognate harmonies elsewhere in Interior Salish (Bessell 1998a, 1998b). In Secwepemctsín, where the long-distance movement shows no blocking effects, we must assume phonologisation of an originally phonetically motivated sound change.
20. In Rotuman, the original V-to-V coarticulation has been obscured by further changes (e.g. **ui* > *y*, **oe* > ø). Note that it is possible in compensatory metathesis that the timing shifts between adjacent unstressed and stressed syllables need not be analysed as foot-internal, though we are aware of no evidence against such an analysis. On Rotuman metathesis, see now also McCarthy 2000.
21. We thus disagree with the tentative conclusion of Hume (2001) that the expected pattern is KP > PK metathesis (which, she contends, is perception-optimising). Her conclusion is based on metathesis patterns in South-Central Dravidian languages, which we analyse as the result of analogical change, not phonetically based sound change (Garrett and Blevins in press). Apart from these Dravidian patterns, which did not arise via genuine metathesis, the typical pattern is PK > KP metathesis.
22. A phonetic optimisation account might explain the unidirectional nature of this change as enhancement of the weak burst of the labial through prevocalic positioning. One difference between the two accounts is that the optimisation account predicts metathesis with segments whose laryngeal and/or manner features are not shared. Our articulatory account does not involve a ‘shift’ of voicing or manner features, but simple overlap of gestures by the major articulators, ruling out a sound change like *pg* > *gp*.
23. Fougeron and Jun (1998) have shown that French accentual-phrase-final syllables are significantly longer than non-accentual-phrase-final syllables; cf. Fougeron and Keating 1997. For French we cannot exclude an alternative analysis, invited by Grammont’s brief description, on which metathesis is a loan adaptation in semi-learned vocabulary (and therefore not the result of any metathesis sound change).
24. In languages where nasality is associated with sequential vowels and affects intervening consonants’ onset closure or release, the appearance of metathesis may arise due to variation in timing of velic closure. This is our interpretation of the reconstructed TN > NT changes in several Kwa languages discussed by Hyman (1972) and Williamson (1973), an interpretation supported by the fact that free variation of this sort actually occurs in the Kolokuma dialect of Ịjọ.
25. Unfortunately we lack space here to discuss these three examples in detail: Latin (apparent **dn* > *nd* via nasal infixation not sound change), Mutsun (erroneous analysis by Hume 1997), and Spanish (apparent *tn* > *nd* via loan adaptation not sound change).
26. See Garrett and Blevins in press, where we explicitly discuss only the East Cushitic examples; our analysis is equally applicable to the Hamer (South Omotic) example described by Lydall (1976, 1988) and recently discussed by Zoll (n.d.).

27. Our analysis thus resolves a paradox noted by Herbert (1986: 195): ‘Although no other cases of similar metatheses [i.e. CN > NC metatheses other than the Kwa cases cited in our n. 24 above] are reported in the literature, we might expect that they should occur. The basis for this expectation is the statistical fact that nasal-oral sequences occur much more frequently in the world’s languages than oral-nasal sequences.’ As Herbert writes, ‘reference to “ease of articulation” gives the wrong prediction in this case’.
28. See Wang 1968, Bach and Harms 1972, Vennemann 1972, Anderson 1981, and more recently, e.g., McCarthy 1991, Blevins and Garrett 1993, Blevins 1997, 2002, Hyman 2001, and Garrett and Blevins in press.
29. Such directionality patterns do not contradict the phonetic optimisation approach, of course, since the result of metathesis is that segments appear in relatively more perceptible positions. The point is that perceptual optimisation is a natural by-product on our analysis, and need not be posited as a mechanism or cause of change.

6 The role of contrast-specific and language-specific phonetics in contour tone distribution

Jie Zhang

1 Introduction

In some tone languages, contour tones (pitch changes within a syllable) may be used contrastively. The phonological distribution of contour tones has been of much theoretical interest, as it sheds light on both the representation of tone (Woo 1969; Leben 1973; Goldsmith 1976; Bao 1990; Duanmu 1990, 1994a; Yip 1989, 1995) and the relation between phonetics and phonology (Duanmu 1994b; Gordon 1998; Zhang 1998, 2002a).

Contour tones are commonly restricted to phonemic long vowels, as in Navajo (Young and Morgan 1987), or to stressed syllables, as in Xhosa (Lanham 1958). I ask the following questions in this chapter: What is the link between these two contexts? Should they be accounted for by independent mechanisms, based on contrastive vowel length and stress respectively, or by some unified mechanism? Drawing from typological and instrumental data, I argue that the unifying factor for contour tone licensing is sonorous rhyme duration, and that the distribution of contour tones is determined by phonetic categories for duration and sonority rather than abstract structural categories based on contrastive vowel length or stress.

The argument goes as follows. The articulation and perception of contour tones determine that they need a sufficient sonorous rhyme duration to be implemented. Thus, a long sonorous rhyme duration is the unifying factor for privileged contour tone licensers. Examining contour tone distribution cross-linguistically, we find that (a) the types of syllables on which contour tones are more likely to occur are exactly those that independently have a longer duration and higher sonority in the rhyme; and (b) the privileged contour tone licensers include not only long-vowelled and stressed syllables, but also syllables in other contexts shown in the literature to undergo lengthening, namely phrase-final position and shorter words. Moreover, instrumental studies show that in languages with two lengthening factors, the factor that induces greater lengthening of the sonorous portion of the rhyme is always the one that is more likely to license contour tones. This could not be explained unless the principles of contour licensing make direct reference to sonorous rhyme duration.

2 The phonetics of contour tones

2.1 *Sonorous rhyme duration is the carrier of contour tones*

2.1.1 The importance of sonority The main perceptual correlate of tone is the fundamental frequency (f_0). All harmonics serve as a cue to f_0, since they occur at integral multiples of f_0. However, as shown by Plomp (1967) and Ritsma (1967), the spectral region containing the second, third and fourth harmonics is especially important in the perception of fundamental frequencies in the range of speech sounds. Since sonorants possess richer harmonic structures than obstruents, including the crucial second to fourth harmonics, sonorants are better tone bearers than obstruents. Moreover, vowels typically have greater energy, and thus stronger acoustic manifestation of harmonics, in the crucial region, than sonorant consonants. These differences will be crucial in the analysis below.

2.1.2 The importance of duration Tone-bearing ability depends not just on sonority, but also duration. This is determined by factors involving both production and perception.

The production of contour tones is different from that of other contour segments (e.g. labial-velars like [k͡p] or clicks) in that for contour tones, the acoustic change results from the state change of a single articulator, the vocal folds. Laryngeal muscle contraction and relaxation, which determine vocal fold tension (Ohala 1978), must be sequenced to produce the pitch variation in a contour tone. Thus, unlike a complex segment whose different oral constrictions can be overlapped, a contour tone requires greater duration to be implemented. This duration depends on the tone's complexity (e.g. MLH[1] vs LH) as well as its pitch range (e.g. MH vs LH) (Sundberg 1979). Moreover, because the muscles responsible for pitch falls are both more numerous and more robust than those that execute a rise, it takes longer to implement a pitch rise than a pitch fall of the same extent (Ohala 1978; Sundberg 1979).

Contour tones also differ from other contour segments such as prenasalised stops and affricates in auditory terms. Although the production of the latter group of sounds also requires an articulator to go from one position to another, the acoustic consequence of such change is sudden; for example, the frication noise is formed the moment the oral occlusion is loosened, and the transition between the two states has no perceptual consequence. But for contour tones, the gradual stretching or relaxation of the vocal folds has a continuous acoustic effect, and the transition from the beginning state to the end state carries a significant perceptual weight in the identification of the tonal contour (Gandour 1978; Gandour and Harshman 1978). Correspondingly, Greenberg and Zee (1979) show that, given the same pitch excursion, the longer the duration of the vowel, the more 'contour-like' the tone is perceived by the listener. They also show that listeners cannot perceive pitch changes reliably when the duration is below 90ms.

2.2 *Two phonetic scales – contour tone-bearing ability and tonal complexity*

We are now in a position to define two relevant phonetic scales: contour tone-bearing ability and tonal complexity. The preceding section established that the realisation of contour tones relies on two aspects of the rhyme: sonority and duration. Therefore, we may hypothesise that it is a weighted sum of these two factors that determines the contour tone-bearing ability of the syllable. I term this weighted sum $C_{CONTOUR}$. Suppose that Dur(V) and Dur(R) represent the duration of the vowel and the sonorant consonant in the rhyme respectively. Then $C_{CONTOUR}$ can be defined as follows:

(1) $C_{CONTOUR} = a \cdot \text{Dur(V)} + \text{Dur(R)}$

Clearly, *a* must be greater than one, since vowels facilitate tonal realisation more than coda sonorants. However, *a* cannot be huge, since the sonorants do make a non-negligible contribution to tone-bearing ability. For more discussion of *a*, see Zhang 2002a.

$C_{CONTOUR}$ can be used to construct a tonal complexity scale, as in (2).

(2) Tonal complexity scale
For any two tones T_1 and T_2, let C_1 and C_2 be the minimum $C_{CONTOUR}$ values required for the production and perception of T_1 and T_2 respectively. T_1 is more tonally complex than T_2 iff $C_1 > C_2$.

From the discussion of contour tone phonetics, we already know that the following three parameters of a tone influence its position in the tonal complexity scale: the number of pitch targets, the pitch excursion between two targets, and the direction of the slope. The influence of these three parameters is stated more rigorously in (3).

(3) For any two tones T_1 and T_2, suppose T_1 has *m* pitch targets and T_2 has *n* pitch targets; the cumulative falling excursions for T_1 and T_2 are Δf_{F_1} and Δf_{F_2} respectively, and the cumulative rising excursions for T_1 and T_2 are Δf_{R_1} and Δf_{R_2} respectively. T_1 has a higher tonal complexity than T_2 if:
a. $m > n$, $\Delta f_{F_1} \geq \Delta f_{F_2}$, and $\Delta f_{R_1} \geq \Delta f_{R_2}$;

b. $m = n$, $\Delta f_{F_1} \geq \Delta f_{F_2}$, and $\Delta f_{R_1} \geq \Delta f_{R_2}$ ('=' holds for at most one of the comparisons);
c. $m = n$, $\Delta f_{F_1} + \Delta f_{R_1} = \Delta f_{F_2} + \Delta f_{R_2}$, and $\Delta f_{R_1} \geq \Delta f_{R_2}$.

Condition (3a) states that if T_1 has more pitch targets and T_1's cumulative falling excursion and rising excursion are both no smaller than those of T_2's, then T_1 is more tonally complex than T_2 (cf. section 2.1.1); for example, 534 is more complex than 53.[2] Condition (3b) states that if T_1 and T_2 have the same number of pitch targets, and one of T_1's cumulative falling excursion and rising excursion is greater than that of T_2's, and the other one is no smaller than that of T_2's, then T_1 is more complex than T_2. For example, 535 has a higher tonal complexity than 545, 534, or 435. Condition (3c) states that if T_1 and T_2 have the same number of pitch targets and the same overall pitch excursion, but the cumulative rising excursion in T_1 is greater than that in T_2, then T_1 is more complex than T_2. For example, 435 is more complex than 534, since $m = n = 3$, $\Delta f_{F_1} + \Delta f_{R_1} = \Delta f_{F_2} + \Delta f_{R_2} = 3$, and $\Delta f_{\mathrm{R}_1} = 2 > \Delta f_{R_2} = 1$.

2.3 *Phonological factors that influence $C_{CONTOUR}$*

Since C_{CONTOUR} has a major effect on tonal complexity, it is important to discuss the phonological factors that influence it. I identify four such factors here: segmental composition, stress, phrase-final position, and the number of syllables in the word to which the rhyme belongs.

Segmental composition factor refers to the length of vowels and the [sonorant] value of coda consonants. According to (1), all else being equal, VV has a greater C_{CONTOUR} value than V; a VR (R = sonorant) has a greater C_{CONTOUR} value than VO (O = obstruent); and VV has a greater C_{CONTOUR} value than VR, provided they have comparable duration.

Together with pitch and amplitude, duration is often one of the key phonetic correlates of stress; for references see Gordon's chapter (this volume) and Zhang 2002a. Therefore it is reasonable to assume that all else being equal, a stressed syllable has a greater C_{CONTOUR} value than an unstressed one.

Final lengthening is the basis for considering the phrase-final position as a relevant parameter. The phonetic literature has shown that the final syllable of a prosodic unit is subject to lengthening (Klatt 1975; Wightman et al. 1992, among others). We thus expect that, all else being equal, a final syllable in a prosodic unit has a greater C_{CONTOUR} value than a non-final syllable in the same prosodic unit.

Lastly, a syllable in a shorter word has a greater C_{CONTOUR} value than an otherwise comparable syllable in a longer word. This is motivated by a series of phonetic studies (Lehiste 1972; Lindblom et al. 1981; Lyberg 1977, among others) that documents that a syllable has a longer duration when it is in a shorter word than in a longer word.

2.4 *Predictions about contour tone distribution by two competing approaches*

As discussed in the introduction to this book, if we acknowledge that constraints on speech production and perception have an effect on phonological markedness, then typological research in phonology can proceed deductively; that is, we may lay out specific hypotheses about possible phonological patterns or implicational laws based on our knowledge of articulation and perception, and test these hypotheses against language data.

If the typology of contours is determined by ease of articulation and perception, we are led to the predictions in (4).

(4) a. The only syllables that selectively license contour tones are those with greater $C_{CONTOUR}$, i.e. long-vowelled, sonorant-closed, stressed, phrase-final, or found in shorter words.
b. Within a language, multiple factors can induce a greater $C_{CONTOUR}$ value, and their contour tone licensing ability corresponds to the degree of $C_{CONTOUR}$ increase.

These predictions are made within a general view of the role of phonetics in phonology that I will call the *direct approach*. This approach embodies two assumptions. The first is that positional licensing is contrast-specific. Since different contrasts require the support of different phonetic properties, they are preferentially licensed in different positions, which reflects the phonetics (Steriade 1993). Specifically, positional licensing of contour tones is tied to the duration and sonority of the rhyme, which are crucial to their production and perception. This may be contrasted with an alternative view, which I will call the *structure-only approach*, in which certain specified phonological positions are hypothesised to host phonological contrasts of any sort, rather than just the contrasts that they are phonetically well suited to host. On this view we expect, for instance, that contour tones should in some languages have a special license to occur on word-initial syllables, since this context has been shown to be privileged for many other phonological features (Steriade 1993, 1995; Beckman 1997). Moreover, since short words and phrase-final position are not contexts that license phonological contrast in general, a structure-only approach predicts that they should not be licensers for tonal contrasts either.

The second assumption of the direct approach is that, for a particular contrast, its positional licensing behaviour should be tuned to language-specific phonetics. That is, a language that has a greater quantitative amount of the relevant phonetic property in a position is expected to be more able or likely to use that position to license the relevant contrast. The structure-only approach differs in predicting that language-particular quantitative differences should play no role in phonological licensing.

A clear way to compare the direct and structure-only approaches is to consider cases in which there are multiple factors present that induce a greater C_{CONTOUR}. In such a case, the direct approach predicts which one should be the better contour tone licenser, but the structure-only approach does not. Let me spell out the argument in detail. Consider a language L in which two distinct properties of a syllable, P_1 and P_2, can both induce a greater C_{CONTOUR} value. Assume further that L has contour tones with distributional restrictions related to P_1 and P_2, and that $C_{\text{CONTOUR}}(P_1) > C_{\text{CONTOUR}}(P_2)$. Let us first see what predictions the structure-only approach makes. Since P_1 and P_2 are properties that increase the syllable's contour tone-bearing ability, we may posit two positional Markedness constraints that penalise the realisation of contour tones on syllables without these properties, as defined in (5).

(5) a. *CONTOUR($\neg P_1$): no contour tone is allowed on syllables without property P_1.
  b. *CONTOUR($\neg P_2$): no contour tone is allowed on syllables without property P_2.

Since P_1 and P_2 are distinct properties of the syllable, there are two possible scenarios for the ranking of (5a) and (5b). Either: (i) there is no universal ranking between them; or (ii) there is such a universal ranking, but it is not based on the phonetic characteristics of P_1 and P_2, so there is no a priori reason to believe that this ranking agrees with the C_{CONTOUR} comparison between P_1 and P_2. In either case, we cannot rule out the ranking *CONTOUR($\neg P_2$) $\gg$ *CONTOUR($\neg P_1$) in a principled way.

To explore this question further, let us complete the analysis by adding more constraints and computing the factorial typology (Prince and Smolensky 1993), which is the full set of language types that are possible under any ranking of the constraints. We first need the two general constraints given in (6).

(6) a. *CONTOUR: no contour tone is allowed on a syllable.
  b. IDENT(Tone): let α be a syllable in the input, and β be any syllable corresponding to α in the output; if α has tone T, then β has tone T.

Calculating the factorial typology of the four constraints given so far, we find that it includes five distinct patterns of contour tone realisation:

(7) Factorial typology (structure-only approach)

	Constraint ranking	Contour tone restriction predicted
a.	*Contour($\neg P_1$), *Contour($\neg P_2$), *Contour ⇓ **Ident(Tone)**	No contour tone on any syllable
b.	*Contour($\neg P_1$), *Contour($\neg P_2$) ⇓ **Ident(Tone)** ⇓ *Contour	Contour tone only on syllables with both P_1 and P_2
c.	*Contour($\neg P_1$) ⇓ **Ident(Tone)** ⇓ *Contour($\neg P_2$), *Contour	Contour tone only on syllables with P_1
d.	*Contour($\neg P_2$) ⇓ **Ident(Tone)** ⇓ *Contour($\neg P_1$), *Contour	Contour tone only on syllables with P_2
e.	**Ident(Tone)** ⇓ *Contour($\neg P_1$), *Contour($\neg P_2$), *Contour	Contour tone on all syllable types

When Ident(Tone) is ranked at the bottom, no contour is allowed on any syllable ((7a)); when Ident(Tone) is ranked between the positional Markedness and general Markedness constraints, contours are only allowed on syllables with P_1&P_2 simultaneously ((7b)), since all other combinations ($\neg P_1$&P_2, P_1&$\neg P_2$, $\neg P_1$&$\neg P_2$) violate at least one of the highly ranked *Contour($\neg P_1$) and *Contour($\neg P_2$); when Ident(Tone) is ranked between the two positional Markedness constraints, contours are only allowed on syllables with P_1 ((7c)) or on syllables with P_2 ((7d)); and finally, when Ident(Tone) is ranked on top, contours are allowed on all syllable types ((7e)).

A slight complication, which will be relevant later on, is the possibility that the grammar could include the disjoined constraint *CONTOUR(¬P_1) ∪ *CONTOUR(¬P_2), which is violated only when both *CONTOUR(¬P_1) and *CONTOUR(¬P_2) are violated (for the theory of constraint disjunction, see Smolensky 1995; Kirchner 1996; and Crowhurst and Hewitt 1997). If such a constraint is included, the factorial typology will expand slightly, to include cases in which contours are licensed by the presence of either P_1 or P_2; the critical ranking is *CONTOUR(¬P_1) ≫ *CONTOUR(¬P_2) ≫ IDENT(Tone) ∪ *CONTOUR(¬P_1), *CONTOUR(¬P_2), *CONTOUR. For simplicity, I will assume this six-member factorial typology in the comparison below; for further discussion of this pattern, see Zhang 2002b.

The really crucial prediction made by the structure-only approach is (7c): contour tones could in principle surface only on syllables with P_2, despite the fact that syllables with P_1 are phonetically better contour tone bearers.

Consider now the predictions the direct approach makes for the same situation in which $C_{CONTOUR}(P_1) > C_{CONTOUR}(P_2)$. I assume that the effect of the $C_{CONTOUR}$ value increase is additive; that is if a syllable has both properties P_1 and P_2, then its $C_{CONTOUR}$ value is even greater.[3] Therefore, we arrive at the following phonetic scale: $C_{CONTOUR}(P_1\&P_2) > C_{CONTOUR}(P_1) > C_{CONTOUR}(P_2)$. This gives rise to the following positional Markedness constraints:

(8) Positional Markedness constraints in a direct approach
   a. *CONTOUR(¬$C_{CONTOUR}(P_1\&P_2)$): no contour tone is allowed on syllables whose $C_{CONTOUR}$ value is less than $C_{CONTOUR}(P_1\&P_2)$.
   b. *CONTOUR(¬$C_{CONTOUR}(P_1)$): no contour tone is allowed on syllables whose $C_{CONTOUR}$ value is less than $C_{CONTOUR}(P_1)$.
   c. *CONTOUR(¬$C_{CONTOUR}(P_2)$): no contour tone is allowed on syllables whose $C_{CONTOUR}$ value is less than $C_{CONTOUR}(P_2)$.

Since these constraints refer to a unified phonetic scale, $C_{CONTOUR}$, and we know that $C_{CONTOUR}(P_1\&P_2) > C_{CONTOUR}(P_1) > C_{CONTOUR}(P_2)$, we can project a universal constraint ranking (cf. Prince and Smolensky 1993: 67), as shown in (9).

(9) *CONTOUR(¬$C_{CONTOUR}(P_2)$) ≫ *CONTOUR(¬$C_{CONTOUR}(P_1)$) ≫ *CONTOUR(¬$C_{CONTOUR}(P_1\&P_2)$)

The basis of this ranking is that *CONTOUR constraints for lower $C_{CONTOUR}$ values are always ranked above *CONTOUR constraints for higher $C_{CONTOUR}$ values.

With this ranking and the general constraints *CONTOUR and IDENT(Tone), the factorial typology predicted by the direct approach can be computed in (10).

(10) Factorial typology (direct approach)

	Constraint ranking	Contour tone restriction predicted
a.	*CONTOUR(¬$C_{CONTOUR}$(P_2)), *CONTOUR(¬$C_{CONTOUR}$(P_1)), *CONTOUR(¬$C_{CONTOUR}$(P_1&P_2)), *CONTOUR ⇓ **IDENT(Tone)**	No contour tone on any syllable
b.	*CONTOUR(¬$C_{CONTOUR}$(P_2)), *CONTOUR(¬$C_{CONTOUR}$(P_1)), *CONTOUR(¬$C_{CONTOUR}$(P_1&P_2)) ⇓ **IDENT(Tone)** ⇓ *CONTOUR	Contour tone only on syllables with P_1&P_2 simultaneously
c.	*CONTOUR(¬$C_{CONTOUR}$(P_2)), *CONTOUR(¬$C_{CONTOUR}$(P_1)) ⇓ **IDENT(Tone)** ⇓ *CONTOUR(¬$C_{CONTOUR}$(P_1&P_2)), *CONTOUR	Contour tone only on syllables with P_1
d.	*CONTOUR(¬$C_{CONTOUR}$(P_2)) ⇓ **IDENT(Tone)** ⇓ *CONTOUR(¬$C_{CONTOUR}$(P_1)), *CONTOUR(¬$C_{CONTOUR}$(P_1&P_2)), *CONTOUR	Contour tone only on syllables with P_1 or syllables with P_2
e.	**IDENT(Tone)** ⇓ *CONTOUR(¬$C_{CONTOUR}$(P_2)), *CONTOUR(¬$C_{CONTOUR}$(P_1)), *CONTOUR(¬$C_{CONTOUR}$(P_1&P_2)), *CONTOUR	Contour tone on all syllable types

This factorial typology turns out to be smaller (five members instead of six). In particular, it is lacking the case in which contours are allowed on only syllables with property P_2. For this outcome to arise, we would need some constraint in the system that penalised contours on P_1 but not P_2. However, no such constraint exists. Indeed, there could be no such constraint, since the constraints are based not on phonological contexts per se, but rather on the values for $C_{CONTOUR}$; and by hypothesis $C_{CONTOUR}(P_1) > C_{CONTOUR}(P_2)$.

In summary, the direct approach and the structure-only approach make two different predictions. First, the direct approach predicts that contour tones specifically gravitate to positions with greater $C_{CONTOUR}$ values, that is, ones with longer sonorous rhyme duration, and in the case of equal sonorous rhyme duration, the position with a longer vocalic component. The structure-only approach, however, is insensitive to phonetic properties specific to contour tones, and thus predicts that word-initial position should be privileged for contour tones, while phrase-final syllables and syllables in shorter words should not be. Second, the structure-only approach predicts that it is possible to have contour tones only on syllables with P_2, despite the fact that syllables with P_1 have a greater contour tone-bearing ability; the direct approach, however, predicts an implicational relation that allows contour tones on P_2 only if contour tones on P_1 are allowed.

To test these different predictions, I carried out a typological survey of contour tone distribution to see if contour tones are indeed more likely to surface on syllables with a greater $C_{CONTOUR}$ value. In addition, I conducted phonetic studies of duration in languages with multiple lengthening factors to see if there is an implicational relation between the stronger and weaker lengthening factors in their contour tone licensing ability.

3 The role of contrast-specific phonetics in contour tone distribution: a survey

3.1 Overview of the survey

The survey was composed of 187 genetically diverse tone languages with contour tones. The full details of the survey are reported in Zhang 2002a; here I will only give a very brief summary. Of the 187 languages, 22 have no restrictions on the distribution of contour tones; 159 have restrictions on contours that accord with the predictions of the direct approach; that is, they were related to the factors given in section 2.3 that increase the $C_{CONTOUR}$ value of the rhyme. Five languages have restrictions in both the expected and unexpected directions. These languages are Lealao Chinantec, Margi, Zengcheng Chinese, Lao, and Saek; for full discussion of these cases, see Zhang 2002a.[4] No languages imposed restrictions solely in the unexpected direction.

3.2 Implicational laws

More specifically, we can make the following observations regarding contour tone distribution from the survey, as in (11). 'Occurs more freely' in a context here means any of the following: (a) contour tones can occur in this context, but not other contexts; (b) the contour tones that occur in this context are a superset of those that occur in other contexts; (c) the pitch excursion of the contour tones that occur in this context is greater than that in other contexts; (d) rising tones can occur in this context, but not others. These scenarios are based on the definition for 'tonal complexity' in section 2.2.

(11) Contour tones occur more freely:
  a. on CVV in 38 languages (e.g. Somali, Navajo, Ju|'hoasi)
  b. on CVV and CVR in 66 languages (Kiowa, Nama, Fuzhou Chinese)
  c. on stressed syllables in 21 languages (Xhosa, Jemez, Lango)
  d. on the final syllable of words or utterances in 45 languages (Etung, Luganda, Beijing Chinese)
  e. on syllables in shorter words in 19 languages (Mende, Ngamambo, Shanghai Chinese)

Through these observations, the following implicational laws can be established:

(12) All else being equal,
  a. if CV can carry contour tones, then CVV can carry contour tones with equal or greater complexity;
  b. if CVO can carry contour tones, then CVR and CVV can carry contour tones with equal or greater complexity;
  c. if an unstressed syllable can carry contour tones, then a stressed syllable can carry contour tones with equal or greater complexity;
  d. if non-final syllables in a prosodic domain can carry contour tones, then the final syllable of the same prosodic domain can carry contour tones with equal or greater complexity;
  e. if syllables in a word having *n* syllables can carry contour tones, then syllables in a word having *n–1* syllables can carry contour tones with equal or greater complexity.

The limiting case of contextual limitation is complete absence: there are many languages in which the more complex contour tones simply do not occur. These gaps may also be phonetically based. Contour tones with higher complexity are disfavoured since they place a higher demand on the duration and sonority of the rhyme. The relevant observations are as follows. First, of all the 187 languages in the survey, only two do not have level tones: Guiyang (Li 1997) and Pingyao (Hou 1980), both Chinese dialects. Second, of the 46 languages that allow complex contours, all allow simple contours. Third, the number of languages that have stricter surface restrictions on rising tones far exceeds the number of languages that have them for falling tones. Thirty-seven languages belong to the former category and only three to the latter.

To this end, three strong implicational tendencies can also be established, as shown in (13).

(13) a. If a language has contour tones, then it also has level tones.
b. If a language has complex contour tones, then it also has simple contour tones.
c. If a language has rising tones, then it also has falling tones.

3.3 Discussion of the survey

Our survey leads to the following conclusion: only factors that systematically influence the duration or sonority of the rhyme can influence the distribution of contour tones; contour tones gravitate to the rhymes with greater C_{CONTOUR} values. The hypothesis in (4a) is supported.[5] Two observations are particularly striking. First, phrase-final syllables and syllables in shorter words are preferred bearers of contour tones, even though they are usually not privileged for other phonological contrasts. Moreover, word-initial syllables, which have been shown to selectively license many other phonological contrasts (Steriade 1993, 1995; Beckman 1997), do not show up on our list of privileged contour tone bearers. The positional licensing behaviour of contour tones is thus sensitive to the phonetic properties that are crucial to contour tones per se, namely duration and sonority. Phrase-final syllables and syllables in short words are privileged contour tone licensers because they are lengthened. Word-initial syllables, on the other hand, fail to license contours because lengthening of the initial rhyme is cross-linguistically very rarely attested.

Contour tones behave like other phonological structures in requiring context-specific licensers. For example, for obstruent place contrasts, Steriade (1993, 2001a) argues that, even though most place contrasts are more likely to be maintained in prevocalic position, the contrast between an alveolar and a retroflex is more likely to be maintained postvocalically, since unlike other place distinctions that primarily benefit from C-to-V formant transitions, their distinction resides in the V-to-C formant transitions. For diphthongs, Zhang (2001) shows that, like contour tones, they gravitate to positions with longer inherent duration, and for essentially the same reasons.

4 The role of language-specific phonetics in contour tone distribution: instrumental studies

This section brings experimental data to bear on the question of whether the formal theory needs to directly encode phonetic properties such as C_{CONTOUR} into the phonological constraints, as the direct approach claims. The data are from languages with contexts whose durational properties fit the description of P_1 and P_2 in the factorial typology study above (section 2.4). To recapitulate the argument, if we find languages in which the privileged factor for contour bearing is P_1, despite the fact that syllables endowed with P_1 but not P_2 have a smaller C_{CONTOUR} value than those endowed with P_2 but not P_1, then we must conclude that the structure-only approach is the correct one. If, however, the privileged factor is always the one that induces a greater C_{CONTOUR}, this supports the direct approach.

4.1 Identifying the languages

If we take stress to be P_1 and final position to be P_2, then we can find both syllables with only P_1 and syllables with only P_2, provided the target language includes words with non-final stress. Xhosa is such a language (Lanham 1958), and many Northern Chinese dialects (e.g. Beijing Chinese) also qualify. In these languages, all syllables are equally stressed, but some monosyllabic reduplicative morphemes and functional words can be destressed, and they can occur word-finally.

The second type of relevant languages has both vowel length and coda sonorancy contrasts, both of which influence the sonorous duration of the rhyme. If we take the [+long] feature of the vowel as property P_1 and the [+son] feature of the coda consonant as property P_2, then syllable CVVO has property P_1 but not P_2, and syllable CVR has property P_2 but not P_1. Among the languages that fit this description, Standard Thai (Abramson 1962; Gandour 1974) and Cantonese (Kao 1971) allow fewer contour tones on CVVO, while Navajo (Young and Morgan 1987) does not allow contour tones on CVR.

I summarise the relevant phonetics of these languages next. The detailed methods and word lists are documented in Zhang 2002a.

4.2 Instrumental Studies

4.2.1 Xhosa Xhosa has penultimate word stress. Vowel length is non-contrastive except in a few grammatical morphemes. All syllables are open: the apparent coda /m/ is in fact syllabic. There are three tones: H, L, and HL (falling). There are no distributional restrictions for H and L, but HL is generally restricted to the penult of a content word. A few monosyllabic grammatical

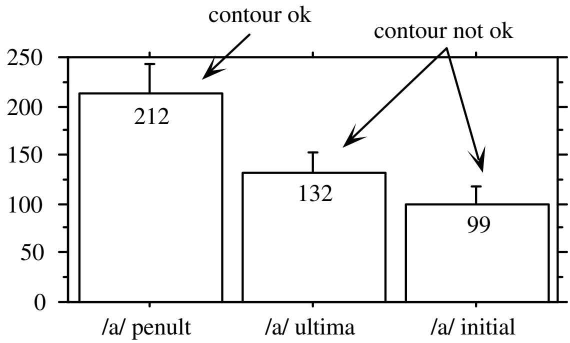


Figure 6.1 Xhosa vowel duration (ms)

prefixes and suffixes can also bear HL, and they do not necessarily occur in penultimate position. But the vowel in these morphemes is lengthened. In an utterance, especially when spoken quickly, some words lose their penultimate stress, creating a tonal alternation HL → H if the penult originally carried a HL (Lanham 1958).

The focus here is on the fact that HL is restricted to the penult of a word. The two relevant durational factors are stress and final position; thus, the two types of syllables of interest are the penult and the ultima. The penult is subject to lengthening by virtue of stress, but not by virtue of being at a prosodic boundary. The opposite is true for the ultima. Given that all syllables are open, the vowel alone constitutes the sonorous portion of the rhyme. The direct approach leads to the hypothesis that in Xhosa, the penult has greater vowel duration than the ultima.

Phonetic data for Xhosa were extracted from a forty-five-minute analogue tape in the UCLA Language Archive. It consists mainly of trisyllabic or quadrisyllabic words read in isolation by one female speaker. Each word has two repetitions. All words extracted for digitisation and measurement were trisyllabic. All target syllables – initial, penult, and ultima – were open with a level-toned /a/ as the nucleus. Forty-four words were used for initial syllables, thirty-four for penultimate, and fifty-four for final.

The mean duration of /a/ for the three positions is shown in figure 6.1. The error bars indicate one standard deviation. One way ANOVA shows that the effect of position is highly significant ($F(2,131) = 242.98$, $p < 0.0001$). Fisher's PLSD post-hoc tests show that all pairs of comparison – penult vs ultima, penult vs initial, and ultima vs initial – have a significant effect at the level of $p < 0.0001$.

The hypothesis that it is the phonetically longest rhymes of Xhosa that support contours is therefore supported by the experimental results. And since it is exactly stress that defines the contour restriction in Xhosa, the data in Xhosa are consistent with the direct approach.

Further experimentation on Beijing Chinese, which likewise permits comparison of stressed and non-final and stressless final syllables, has yielded similar results; for details, see Zhang 2002a.

4.2.2 Standard Thai Consider now the second type of relevant languages, those with both vowel length and coda sonorancy contrasts. We will examine Standard Thai, Cantonese, and Navajo.

Syllables in Standard Thai can be open, closed by an obstruent /p/, /t/, /k/, or /ʔ/, or closed by a nasal /m/, /n/, or /ŋ/. Vowel length is contrastive in closed syllables. I will refer to syllables closed by an obstruent (CVO and CVVO) as checked syllables, and other syllables (CV, CVN, and CVVN) as non-checked syllables. There are five tones in Thai: H, M, L, HL, and LH. On non-checked syllables, all five tones can occur. On CVVO, generally, only HL and L occur, but in rare instances, H can also occur (e.g., *nóot* 'note'; *khwɔ́ɔt* 'quart', both English loanwords). On CVO, generally, only H and L occur, but HL occurs occasionally (e.g., *kɔ̂ʔ* 'then, consequently') (Gandour 1974; Hudak 1987). This tonal distribution is summarised in (14) (adapted from Gandour 1974; parentheses indicate rare occurrence.).

(14) Tonal distribution in Standard Thai (Gandour 1974):

	H	M	L	HL	LH
CV	+	+	+	+	+
CVN	+	+	+	+	+
CVVN	+	+	+	+	+
CVVO	(+)	–	+	+	–
CVO	+	–	+	(+)	–

It can be seen that the distribution of contour tones in Thai is primarily affected by the checked/non-checked distinction, as non-checked syllables can carry both LH and HL whether they have a long or a short vowel. Vowel length is also relevant, since HL can occur on CVVO, but usually not on CVO.[6] However, the crucial distinction for present purposes is that CVVO supports only a subset of the tones supported by CV and CVN, despite its greater phonological vowel length. If the approach taken here is correct, it must be the case that the sonorous rhyme duration of CV and CVN is greater than that of CVVC. In other words, in Thai, the factor *checked vs nonchecked* should outweigh the factor *V vs VV*.

Thai data were collected from two native speakers. For each of the five syllable types – CV, CVVN, CVN, CVVO, CVO – four monosyllabic words

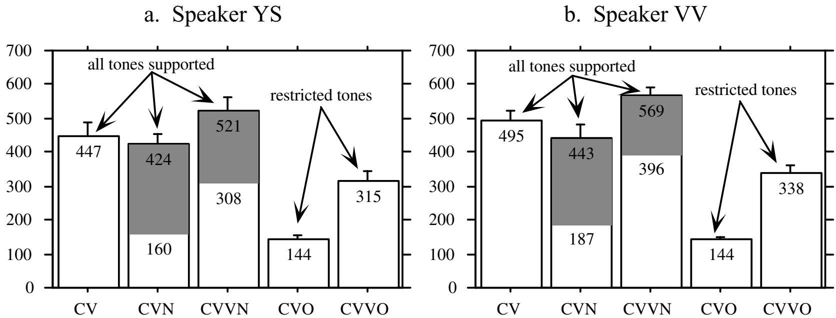


Figure 6.2 Standard Thai sonorous rhyme duration (ms)

were recorded, each with eight repetitions. All words had the nucleus /a/ and had either M or L tone.

The sonorous rhyme duration for the five syllable types are plotted in figure 6.2. The grey portion in the bars indicates sonorous duration contributed by the nasal coda.

For each speaker, a one-way ANOVA with sonorous rhyme duration as the dependent variable and syllable type as the independent variable was carried out. The effect is highly significant for both speakers: for YS, $F(4, 135) = 623.3$, $p < 0.0001$; for VV, $F(4, 135) = 1157.7$, $p < 0.0001$. Fisher's PLSD post-hoc tests show that for both speakers, both CV and CVN have a longer sonorous rhyme duration than CVVO at the significance level of $p < 0.0001$.

The hypothesis CV > CVVO, CVN > CVVO is therefore supported, and the data in Thai are thus consistent with the direct approach. The phonological pattern that more contour tones are allowed on non-checked syllables than checked syllables is in agreement with the phonetic fact that non-checked syllables have longer sonorous rhyme duration.

Gordon (1998) documents a similar pattern for Cantonese. The syllable inventory of Cantonese is the same as Thai: CV, CVN, CVVN, CVO, and CVVO (N = /m, n, ŋ/, O = /p, t, k/). While there is both a vowel length contrast and a checked/non-checked distinction, the distribution of contour tones is only affected by the latter: in CV, CVN, and CVVN, seven different tones, including four contour tones, can occur: 53, 35, 21, 23, 55, 33, 22. But in CVVO and CVO, only the level tones 5, 3, and 2 can occur, even when the syllable contains a long vowel.

Gordon's duration data for different syllable types of Cantonese are shown in figure 6.3. Again, the grey portion in the bars indicates sonorous duration contributed by the nasal coda. As in Thai, Cantonese has a considerably longer sonorous rhyme duration in non-checked syllables than in checked ones, regardless of the phonological length of the nucleus.

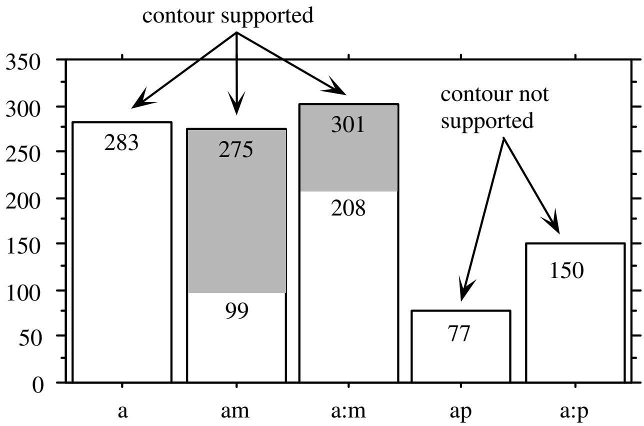


Figure 6.3 Cantonese sonorous rhyme duration (ms)

4.2.4 Navajo Navajo exhibits the opposite pattern to Standard Thai and Cantonese: it restricts contours to long vowels, regardless of the coda. Therefore the crucial question is whether the duration pattern of Navajo is also different from Standard Thai and Cantonese.

Navajo vowel length is contrastive in both open and closed syllables. There are six syllable types (CV, CVO, CVR, CVV, CVVO, and CVVR) and four tones (H, L, HL, LH), with the contour tones restricted to long vowels and diphthongs, that is, CVV, CVVO, and CVVR. Therefore, unlike Thai and Cantonese, the factor that determines the contour distribution in Navajo is vowel length, not coda sonorancy. The tonal distribution of Navajo is summarised in (15).

(15) Tonal distribution in Navajo

	H	L	HL	LH
CV	+	+	−	−
CVO	+	+	−	−
CVR	+	+	−	−
CVV	+	+	+	+
CVVO	+	+	+	+
CVVR	+	+	+	+

The crucial C_{CONTOUR} comparisons are between CVR and CVV and between CVR and CVVO: CVR benefits from having a sonorant coda, while CVV and CVVO benefit from having a long vowel. Given that it is the long vowel that

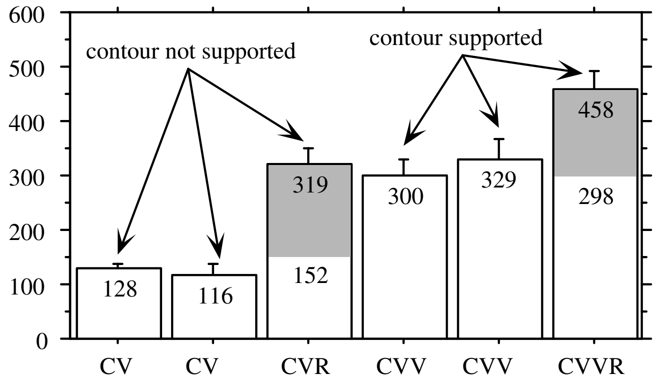


Figure 6.4 Navajo sonorous rhyme duration (ms)

licenses contour tones here, the direct approach leads to a phonetic hypothesis that is crucially different from that for Thai and Cantonese: $\{C_{CONTOUR}(CVV), C_{CONTOUR}(CVVO)\} > C_{CONTOUR}(CVR)$. Since a vowel is phonetically a better tone-bearing segment than a sonorant consonant when they are of comparable duration (see section 2.1.1 and the definition of $C_{CONTOUR}$), we arrive at the following hypothesis: the sonorous rhyme duration of syllables with a long nucleus should be longer than or comparable to that of syllables with a short nucleus. In particular, $\{CVV, CVVO\} \geq CVR$.

The prediction, unfortunately, is not fully testable with the available data. The problem is that we do not know in advance the value of the parameter *a* in the formula (1), which weights the contribution of R vs V in determining the value of $C_{CONTOUR}$. We will see, however, that the data do lend themselves to a plausible interpretation.

Navajo data were collected from one male native speaker. The target syllable is always the second syllable of a disyllabic word. For each syllable type, two words with /i/ and two words with /a/ were used, and all target syllables had a low tone. Eight repetitions were recorded.

The sonorous rhyme duration for each syllable type is plotted in figure 6.4. The grey portion again indicates sonorous duration contributed by the coda consonant.

A one-way ANOVA shows that the syllable type has a significant effect on the sonorous rhyme duration: $F(5, 162) = 596.7$, $p < 0.0001$. CVR has a comparable sonorous duration in the rhyme to CVV and CVVO: it is not significantly different from either CVVO (Fisher's PLSD post-hoc tests, $p > 0.01$) or CVV ($p > 0.01$).

The data here are less conclusive than in the Thai and Cantonese cases. In particular, CVR, which cannot bear contours, has a sonorous duration comparable to the CVVX syllables, which can. However, the data are compatible with the direct hypothesis provided *a* is sufficiently high to give CVV and CVVO substantially higher $C_{CONTOUR}$ values than CVR. There is in fact a crucial comparison that supports this conjecture: in Thai and Cantonese, the sonorous rhyme duration in CVR is considerably longer than that in CVVO; but in Navajo, the two durations are comparable. This difference suggests that a considerable amount of sonorant consonant duration may be needed to balance a given amount of vocalic duration. This suggests that *a* is indeed rather high, and thus that Thai and Cantonese CVR would qualify for contour bearing while Navajo CVR would not. The experimental data thus appear to be compatible with the direct approach.

4.3 Conclusion to the instrumental studies

The fact that all the phonetic cases studied here reveal data patterns consistent with the more restrictive direct approach constitutes significant support for this approach, as this implies that there is no empirical reason for us to adopt the less restrictive structure-only approach. The direct approach is more restrictive because it does not predict situations in which contours are restricted to phonemic long vowels in Thai and Cantonese, or to sonorant-closed syllables in Navajo.

Xhosa and Beijing Chinese illustrate a similar point from the interaction of two different durational factors: stress and final position in a prosodic domain. It turns out that in both languages, stress plays the decisive role in determining the sonorous duration of the rhyme and correspondingly the distribution of contour tones.

5 The basics of a formal analysis

In the previous two sections, I have argued that the distribution of contour tones is best captured by a direct approach, which encodes the phonetic index $C_{CONTOUR}$ of a rhyme. In this section, I sketch out a formalisation of this approach.

5.1 Overview of the theoretical apparatus

The patterns of contour tone distribution that an analysis must capture are the following. First, the distribution of contour tones depends on a phonetic index of the rhyme – $C_{CONTOUR}$; the lower the $C_{CONTOUR}$ value, the more limited distribution the contour tones will have on the rhyme. Second, when a contour tone encounters a syllable with insufficient tone-bearing ability, there are three possible resolutions: increasing the $C_{CONTOUR}$ value of the syllable, flattening out the pitch excursion, or both.[7] For both lengthening and flattening, the change can be either neutralising (merging with some other phonological category) or allophonic.

I posit three families of constraints. Markedness constraints of the family *CONTOUR(*T*)-C$_{\text{CONTOUR}}$(*R*) are violated when contour tones occur on rhymes with certain C$_{\text{CONTOUR}}$ values. Markedness constraints of the family *DUR penalise extra duration on the syllable. Faithfulness constraints of the family PRES(Tone) enforce similarity between tonal input and output. Each of these constraint families has a set of intrinsic rankings, described below.

The interaction of these three constraint families gives rise to the attested patterns of contour tone restriction. In the following sections of this chapter, I formally define these constraint families and discuss their interactions in detail.

5.2 Constraints and their intrinsic rankings

*5.2.1 *CONTOUR(x)-C$_{CONTOUR}$(y)* The Markedness constraints *CONTOUR(*x*)-C$_{\text{CONTOUR}}$(*y*) ban contours when their C$_{\text{CONTOUR}}$ values are too low. Formally, they are defined as follows:

(16) *CONTOUR(x_i)-C$_{\text{CONTOUR}}$(y_j):
no contour tone x_i is allowed on a syllable with the C$_{\text{CONTOUR}}$ value of syllable y_j or smaller.

The *CONTOUR(x_i)-C$_{\text{CONTOUR}}$(y_j) constraints observe two sets of intrinsic rankings, given in (17), which are projected from the phonetics.

(17) a. If C$_{\text{CONTOUR}}$ (y_a) > C$_{\text{CONTOUR}}$(y_b), then *CONTOUR(x_i)-C$_{\text{CONTOUR}}$(y_b) $\gg$ *CONTOUR(x_i)-C$_{\text{CONTOUR}}$(y_a).
b. If contour tone x_m is higher on the *Tonal Complexity Scale* (see (2) and (3)) than contour tone x_n, then *CONTOUR(x_m)-C$_{\text{CONTOUR}}$(y_j) $\gg$ *CONTOUR(x_n)-C$_{\text{CONTOUR}}$(y_j).

These rankings reflect the speaker's knowledge that a structure that is phonetically more demanding should be banned before a structure that is less so; and that a syllable should be able to host a tone with a lower complexity before it can host a tone with a higher complexity.

5.2.2 *DURATION Assuming that each segment in a certain prosodic environment has a minimum duration (Klatt 1973; Allen et al. 1987), I define the *DURATION (abbr. *DUR) constraint family as follows:

(18) *DUR(τ_i): for all segments in the rhyme, their cumulative duration in excess of the minimum duration in the prosodic environment in question cannot be τ_i or more.

These constraints also have an intrinsic ranking, as in (19).

(19) If $\tau_i > \tau_j$, then $*\textsc{Dur}(\tau_i) \gg *\textsc{Dur}(\tau_j)$

For more detailed discussion on this constraint family and how it interacts with the minimal duration requirement for each segment, see Zhang 2002a.

5.2.3 PRESERVE*(Tone)* PRESERVE(Tone) constraints penalise candidates according to their deviation – defined in perceptual terms – from the input. Assume that we can define a function $S(T_I, t)$, which returns the value of perceptual similarity between any pair of tones T_I and *t*. $S(T_I, t)$ can be defined in such a way that if t_1 is perceptually more similar to T_I than t_2, then $S(T_I, t_1) < S(T_I, t_2)$.[8] We can then define the constraint family PRESERVE(Tone) (abbr. PRES(T)) as in (20).

(20) $\forall i$, $1 \leq i \leq n$, $\exists$ constraint PRES(T, *i*), defined as:
an input tone T_I must have an output correspondent T_O which satisfies the condition $S(T_I, T_O) < i$.

This constraint family is internally ranked by the principle that the candidate that deviates the most from the input will be penalised by the highest ranked constraint (cf. the 'P-map' approach of Steriade 2001b). More formally, we have:

(21) PRES(T, *n*) ≫ PRES(T, *n-1*) ≫ . . . ≫ PRES(T, *2*) ≫ PRES(T, *1*).

5.3 Assumptions made in the model

My model relies on the following general assumptions.

Canonicality. I assume that the canonical speaking rate and style are the basis on which the grammar is constructed. Thus, $C_{CONTOUR}$ is calculated from the canonical duration of the sonorous portion of the rhyme. This assumption is necessary because syllable durations and pitch range vary under different speaking rates and styles, and the 'tolerance level' for tone slope varies too. Since the standard mode of speech is what language users are most frequently exposed to and most frequently utilise, it is reasonable to assume that it is this mode that defines the quantitative values that appear in the constraints.

Normalisation. Upon identifying the canonical speaking rate and style, I further assume that speakers are able to normalise duration and pitch across speaking rates and styles (Kirchner 1998; Steriade 1999). Only under this assumption can we discuss the grammatical behaviour of different rates and styles and account for the stability of the phonological system across these rates and styles. For example, if the speaker was not able to normalise, but took the phonetic values in the inputs, outputs, and constraints at face value, then a HL contour on CVO would violate a higher ranked *CONTOUR (x_i)-$C_{CONTOUR}(y_j)$ constraint in the fast speech grammar than in the slow speech grammar, so that the phonological system for the two rates would be different.

This assumption does not preclude the possibility of different phonological behaviour in different speaking rates and styles. It is still possible for particular speech styles to be associated with constraints that are specific to them, for example constraints that refer to the realisation of affective signalling or constraints that refer to absolute duration instead of normalised duration to express physiological limitations, and so on. For discussion, see Kirchner's chapter, as well as Harris 1969 and Ao 1993. But given the overall stability of the phonological system despite the fluctuating speaking rates and styles, I believe that normalisation is a necessary assumption here.

The assumption of normalisation is justified by experimental evidence that speakers can attend to and compensate for fluctuations in speaking rate and style. For example, many perceptual studies show that the speaking rate of the stimuli influences listeners' perceptual boundary between two segments if this boundary is dependent on duration (Port 1979; Miller and Grosjean 1981; Pols 1986). For studies on tone normalisation, see Leather 1983, Moore 1995, and Moore and Jongman 1997.

Contrast constraints. We know that given a phonetic dimension, only a small number of contrasts will emerge in any given language. But if phonetic details such as a minute change of duration or pitch excursion can be included in phonological representations, how can contrasts emerge? Flemming (1995; this volume) and Kirchner (1997) have addressed this problem with proposals to incorporate constraints on the perceptual distance of contrasts (MINDIST constraints by Flemming, POLAR constraints by Kirchner). Here I simply acknowledge that the system in which I operate also needs constraints that achieve such effects, without committing myself to either approach.

5.4 Sample analyses

Consider now the predictions made by this formal apparatus. Suppose that in language *L*, there exists an underlying contour tone *T* whose pitch excursion under the standard speaking rate and style is of Δf. Suppose further that that input form *T* is lodged on a rhyme *R* whose $C_{CONTOUR}$ value is *c* and whose minimum sonorous rhyme duration is *d*. Let us see what predictions (in terms of phonological alternation or allophonic distribution) the apparatus makes.

5.4.1 No change necessary The first possibility is that all members of the PRES(T) and *DUR families outrank *CONTOUR(*T*)-$C_{CONTOUR}(R)$. Under this ranking, the contour faithfully surfaces on the given rhyme without lengthening. This is because any flattening of the contour or lengthening of the sonorous rhyme duration in order to satisfy *CONTOUR(T)-$C_{CONTOUR}$(R) will incur violations of higher ranking PRES(T) or *DUR constraints:

(22) $T_{\Delta f}, R_d \rightarrow \Delta f, d$

	$T_{\Delta f}, R_d$	PRES(T)	*DUR	*CONTOUR(T)-$C_{CONTOUR}$(R)
☞	faithful: $\Delta f, d$			*
	contour reduction: $\Delta f - f_0, d$	*!		
	rhyme lengthening: $\Delta f, d + d_0$		*!	

Languages of this sort are attested. For example, !Xũ (Snyman 1970), ǂKhomani (Doke 1937), and a number of Chinantec languages allow all tones on all syllable types, be they open or checked, long-vowelled or short-vowelled. Although most of the sources I consulted on these languages do not give phonetic details of tone and duration, thus it is possible that the contour tones on shorter syllable types are somewhat flattened, or these syllables are somewhat lengthened, there is some phonetic documentation on Lalana Chinantec (Mugele 1982) which shows that the same contour tone exhibits relative stability of onset and endpoint on different syllable types, and the same syllable type exhibits relatively stable duration when carrying different tones.

The analysis further predicts that on a rhyme R' with a $C_{CONTOUR}$ value greater than c, Δf will also be faithfully realised, since the constraint *CONTOUR(T)-$C_{CONTOUR}$(R') will be even lower ranked than *CONTOUR(T)-$C_{CONTOUR}$(R). This prediction is consistent with the implicational hierarchies established in the survey.

5.4.2 Partial contour reduction Now consider cases in which a particular contour type must appear in partially reduced form (less pitch range) on certain short rhymes. In such cases, *CONTOUR(T)-$C_{CONTOUR}$(R) outranks some but not all PRES(T) constraints, but the *DUR constraint family is still undominated. Under this ranking, the contour is flattened to satisfy the *CONTOUR(T)-$C_{CONTOUR}$(R) constraint, but no extra duration can be added to the sonorous portion of the rhyme, as illustrated in (23).

(23) $T_{\Delta f}, R_d \rightarrow \Delta f\text{-}f_0, d$

$T_{\Delta f}, R_d$	*DUR	*CONTOUR(T)-C$_{\text{CONTOUR}}$(R)	PRES(T)
faithful: $\Delta f, d$		*!	
☞ contour reduction: $\Delta f\text{-}f_0, d$			*
rhyme lengthening: $\Delta f, d + d_0$	*!		

Such flattening occurs in Pingyao Chinese (Hou 1980), where the contour tones 53 and 13, which are fully realised on CV (with a phonetically long vowel) and CVR, have partial realisations 54 and 23 on CVO.

This ranking type also predicts that on a rhyme *R'* with a C$_{\text{CONTOUR}}$ value greater than *c*, Δf will be more faithfully realised, i.e. realised with less or no reduction of the pitch excursion. This is because the relevant *CONTOUR(x)-C$_{\text{CONTOUR}}$(y) constraint *CONTOUR(T)-C$_{\text{CONTOUR}}$(R') will be lower ranked than *CONTOUR(T)-C$_{\text{CONTOUR}}$(R), and this will allow more PRES(T) constraints to exert influence on the output form. This, again, is consistent with the implicational hierarchy established in the survey.

5.4.3 Complete contour reduction The third possibility is to have all *CONTOUR(x)-C$_{\text{CONTOUR}}$(R) and *DUR constraints outrank all the relevant PRES(T) constraints. That is, *CONTOUR(δ)-C$_{\text{CONTOUR}}$(R), where δ represents the smallest perceptible pitch excursion, outranks the PRES(T, i) constraint that penalises changing the tone T to a level tone. This ranking predicts that the tone T will be flattened all the way to a level tone, as illustrated in (24).

(24) $T_{\Delta f}, R_d \rightarrow 0, d$

$T_{\Delta f}, R_d$	*DUR	*CONTOUR(δ)-C$_{\text{CONTOUR}}$(R)	PRES(T, i)
faithful: $\Delta f, d$		*!	
partial contour reduction: $\Delta f\text{-}f_0, d$		*!	
☞ complete contour reduction: $0, d$			*
rhyme lengthening: $\Delta f, d + d_0$	*!		

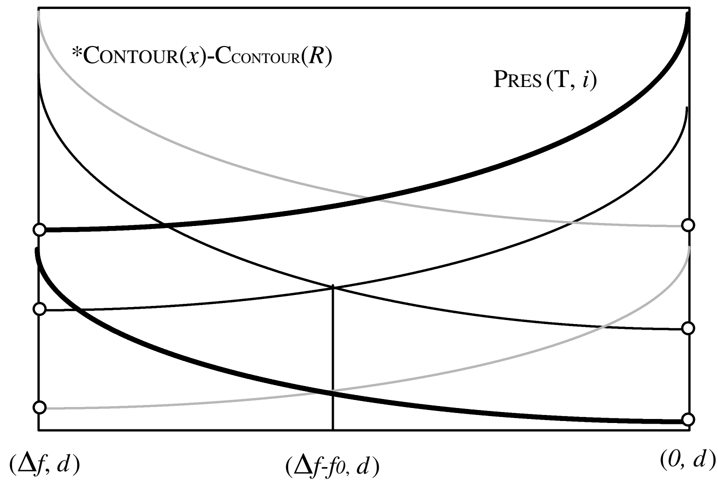


Figure 6.5 Interaction of *CONTOUR(x)-$C_{CONTOUR}$(R) and PRES(T, i) yielding different degrees of contour reduction

This is the most commonly attested pattern of contour tone restrictions in languages, that is, certain contour tones cannot occur on syllables with low $C_{CONTOUR}$ values. We have seen many examples of this sort, for example Xhosa's restriction of contour tones to stressed syllables, Navajo's restriction of contour tones to long vowels, Cantonese's restriction of contour tones to non-checked syllables, and so on.

This ranking further predicts that on a rhyme R' with a $C_{CONTOUR}$ value greater than c, Δf will be more faithfully realised, that is, realised with less or no reduction of the pitch excursion: *CONTOUR(δ)-$C_{CONTOUR}$(R') will be lower ranked than *CONTOUR(δ)-$C_{CONTOUR}$(R), and this will allow more PRES(T) constraints to exert influence on the output form. This is yet again consistent with the implicational hierarchy established in the survey.

5.4.4 Summary The scenarios described in sections 5.4.1–5.4.3 are summarised in the schematic graph in figure 6.5. In the graph, the x-axis represents tonal candidates, arranged from left to right according to the degree of reduction of their pitch range. Thus, the leftmost candidate on the x-axis is the most faithful to the input, with no flattening at all (Δf). The rightmost candidate is the one with complete flattening (to zero). Since all *DUR constraints are always ranked on top in the scenarios described so far, I only consider candidates that respect these constraints, that is, candidates with no lengthening. Thus, in all candidates, d appears as the sonorous rhyme duration.

The *y*-axis represents constraint ranking: the higher the *y* value, the higher the ranking. The curves in the graph represent the highest ranked constraints in the *CONTOUR(*x*)-$\mathrm{C}_{\mathrm{CONTOUR}}$(*R*) and PRES(T, *i*) families that the candidates on the *x*-axis violate.

The thick black lines in the graph indicate the ranking of the two constraint families that ensures the faithful realisation (as described in section 5.4.1) of the pitch excursion Δf, which appears as the leftmost candidate on the *x*-axis. The highest ranked constraint it violates is *CONTOUR(*T*)-$\mathrm{C}_{\mathrm{CONTOUR}}$(*R*). Any other candidate to the right, which deviates from the input, will induce violation of a higher ranked PRES(T, *i*) constraint.

The thin black lines indicate the ranking that produces partial reduction of the contour to Δf-f_0 (cf. section 5.4.2). This is the candidate on the *x*-axis that corresponds to the point of intersection of the two curves. Any candidate further to the left violates a higher ranked *CONTOUR(*x*)-$\mathrm{C}_{\mathrm{CONTOUR}}$(*R*) constraint, and any candidate further to the right violates a higher ranked PRES(T, *i*) constraint.

The grey lines indicate the ranking that forces complete reduction of the contour tone to a level tone, as described in section 5.4.3. Level tone is represented here as the rightmost candidate on the *x*-axis. The highest ranked constraint it violates is the highest ranked PRES(T, *i*) constraint. Any other candidate further to the left which would deviate less from the input will induce the violation of a higher ranked *CONTOUR(*x*)-$\mathrm{C}_{\mathrm{CONTOUR}}$(*R*) constraint.

From these three examples, then, it should be clear that any degree of pitch-range reduction is derivable in this system. Moreover, the same analytic strategy can be straightforwardly extended to cover cases in which duration is altered, either alone or in addition to a pitch range adjustment. The three basic cases are summarised below; for full exemplification of these cases, see Zhang 2002a.

Nonneutralising lengthening. This occurs when *CONTOUR(*T*)-$\mathrm{C}_{\mathrm{CONTOUR}}$(*R*), along with the PRES(T) family, outrank some *DUR constraints. It is found in Mitla Zapotec (Briggs 1961) and Wuyi Chinese (Fu 1984).

Neutralising lengthening. Where the *CONTOUR(*T*)-$\mathrm{C}_{\mathrm{CONTOUR}}(V_d)$ constraint associated with a long vowel outranks *DUR(*d*), and all PRES(T) constraints are ranked on top, the ranking predicts neutralising lengthening when the tone *T* occurs on a short vowel. This is found in Gã (Paster 1999).

Both contour reduction + rhyme lengthening. This outcome, found in Hausa (Newman 1990; Gordon 1998), results when *CONTOUR(*T*)-$\mathrm{C}_{\mathrm{CONTOUR}}$(*R*) outranks some *DUR constraints and some PRES(T) constraints.

5.5 Summary

In this section, I have proposed an explicit formalisation for the direct approach to contour tone distribution and discussed the patterns that are predicted by the model. The model directly encodes phonetic details such as $C_{CONTOUR}$ and the durational properties of the rhyme. I have argued that such a move is necessary (sections 3–4) and makes restrictive predictions.

6 Alternative approaches

In this section, I discuss two alternative approaches to contour tone distribution.

6.1 Moras

Traditionally, the mora is used in phonology to capture the heavy vs light distinction in weight-related phenomena such as stress assignment, compensatory lengthening, metrics, and word minimality (Newman 1972; Hyman 1985; McCarthy and Prince 1986; Hayes 1989; among others). It has also been proposed by Duanmu (1990, 1993, 1994a, 1994b) to be the tone-bearing unit. Observing that Chinese languages with fewer distributional restrictions on contour tones (e.g. Mandarin) have generally longer syllable rhymes than those with more restrictions (e.g. Shanghai), Duanmu argues that a contour tone must be represented as a concatenation of level tones, each of which needs a mora to be licensed. The difference in contour tone restrictions between Mandarin and Shanghai stems from the fact that syllables are bimoraic in Mandarin but monomoraic in Shanghai. I will give three arguments here that the bimoraic status is neither sufficient nor necessary for contour tone licensing.

First, given that the main purpose of the mora is to capture the heavy vs light distinction, the maximum mora count is usually assumed to be two (or at most three, for cases like Estonian where a three-way weight distinction has to be made; Hayes 1989). But the contour tone licensing behaviour sometimes requires more than three levels of distinction. For example, in Mende (Leben 1973; Dwyer 1978; Zhang 2000), long vowels can carry LHL, LH, or HL in monosyllabic words, but only LH or HL in other positions. Short vowels can carry LH or HL in monosyllabic words, HL in the final position of disyllabic or polysyllabic words, but no contour in other positions. This is a four-way distinction, and goes beyond the maximum mora count that any version of moraic theory is willing to accommodate.

Second, sometimes contour tones with the same number of pitch targets have different distributional restrictions. Thus, in many languages, rising tone has a more limited distribution than falling (Mende, Gã, Kɔnni, Kukuya, Tiv). In many Chinese dialects (Pingyao, Shanghai, Fuzhou), contour tones with more pronounced pitch excursion are restricted to non-checked syllables, while contour tones with less pronounced pitch excursions can occur on checked syllables. These asymmetries cannot be captured in a moraic approach, as the same number of pitch targets ought to require the same number of moras to be licensed.

Third, languages do not always favour syllables with clearly higher mora count (such as those with long vowels or sonorant codas) for contour tone bearing. We have seen that many languages allow contour tones more freely in phrase-final syllables and syllables in shorter words. Although the effect of final lengthening may be neutralising in some languages (e.g. Luganda), it is not in many others; and the effect of syllable lengthening in shorter words is not neutralising in any language of which I am aware. Since these effects are purely quantitative, it is implausible to attribute them to the structural property of mora count.

To summarise, the moraic approach attempts to capture the correlation of tonal contour complexity and duration representationally, but the patterns of correlation are too complicated to be accounted for by the limited mechanism of mora counting.

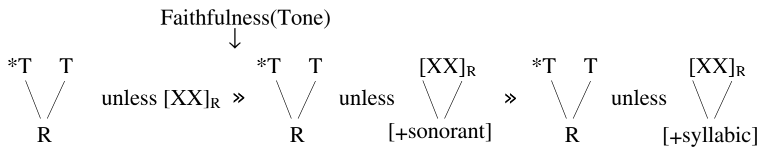
6.2 Gordon's approach

Gordon (1998, 1999, this volume) also recognises that a syllable's contour tone carrying ability is crucially dependent on the duration and sonority of the rhyme. He maintains, however, that the effects of phonetics in phonology are indirect, mediated by phonological structures projected from phonetics. In his system, the constraint that bans contour tones on CV and CVO, for example, takes the form of (25). The constraint does not directly refer to the phonetic measurement of contour tone bearing ability $C_{CONTOUR}$, but to the number of timing slots and the feature [+sonorant].

```
(25)   *T   T              [XX]R          A contour tone is licensed by a rhyme
         \ /    unless      \ /       :
          R             [+sonorant]       containing two timing slots that are [+sonorant].
```

Consequently, unlike the direct approach (in which partial contour reduction, partial rhyme lengthening, and the combination of both are all treated in a uniform fashion as governed by same constraint hierarchy), Gordon must assume that these are phonetic effects, as his analysis of Hausa indicates (26) (Gordon 1998: 247). With the tonal faithfulness constraint flanked between the constraint requiring two timing slots for contour tones and the constraint requiring two [+sonorant] timing slots for contour tones, Gordon's analysis predicts that the falling tone can surface on CVO, but does not predict that the falling tone is partially flattened and the fall-carrying CVO has a lengthened rhyme. Gordon actually does not explicitly specify where the partial flattening and lengthening take place, whether in the phonology proper or in the phonetics.

(26)



With constraint formulation as in (25), Gordon's approach is in fact subject to the same criticisms as the moraic approach. First, representing the contour tone as [T T] concatenation misses any generalisations regarding tonal complexity, particularly those involving tones with the same number of pitch targets. Second, the concept of timing slots in the rhyme is formally identical to moras; then we have the same problems (cf. section 6.1) with the number of distinctions that must be made and whether it is appropriate to add timing slots for phrase-final syllables and syllables in shorter words.

7 Conclusion

I have argued for two points. First, the phonological behaviour of contour tone licensing is determined by the duration and sonority of the syllable rhyme, and the root of this correlation lies in the fact that the production and perception of tonal contours require sufficient sonorous rhyme duration. Second, a formal analysis of contour tone licensing must encode the language-specific phonetic facts of duration and sonority in the constraints; a structure-only approach that only refers to the privileged phonological positions makes erroneous predictions.

In a broader context, the facts of contour tone licensing illustrate the two ways in which phonetics influences the phonological patterning of positional licensing. First, positional licensing is contrast-specific; that is, different phonological contrasts preferentially occur in different positions. This is due to the fact that different contrasts require the support of different phonetic properties. Second, for a particular contrast, its positional licensing behaviour is tuned to language-specific phonetics; that is, the richness of the relevant phonetic property in a given phonological context can differ from language to language, and the positional licensing behaviour of the contrast in question is sensitive to these differences. A valid analysis of positional licensing needs to reflect the relevance of contrast-specific and language-specific phonetics, and the direct approach sketched out above is an example of how such analyses should proceed.

References

Abramson, Arthur S. (1962). The vowels and tones of Standard Thai: acoustical measurements and experiments. *IJAL* 28.2: part III.

Allen, Jonathan, M. Sharon Hunnicutt, and Dennis Klatt, with Robert C. Armstrong and David Pisoni (1987). *From text to speech: The MITalk system*. Cambridge: Cambridge University Press.

Ao, Benjamin (1993). *Phonetics and phonology of Nantong Chinese*. PhD dissertation, Ohio State University.

Bao, Zhiming (1990). *On the nature of tone*. PhD dissertation, MIT.

Beckman, Jill (1997). *Positional faithfulness*. PhD dissertation, University of Massachusetts, Amherst.

Briggs, Elinor (1961). *Mitla Zapotec grammar*. Mexico City: Instituto Linguistico de Verano, Mexico.

Chao, Yuen Ren (1948). *Mandarin primer: An intensive course in spoken Chinese*. Cambridge, MA: Harvard University Press.

(1968). *A grammar of spoken Chinese*. Berkeley and Los Angeles: University of California Press.

Crowhurst, Megan and Mark Hewitt (1997). *Boolean operations and constraint interaction in Optimality Theory*. Rutgers Optimality Archive, 229, http://roa.rutgers.edu.

Doke, Clement. M. (1937). An outline of ‡Khomani Bushman phonetics. In J. D. Rheinallt Jones and C. M. Doke (eds.), *Bushmen of the Southern Kalahari*. Johannesburg: Witwatersrand University Press. 61–88.

Duanmu, San (1990). *A formal study of syllable, tone, stress and domain in Chinese languages*. PhD dissertation, MIT.

(1993). Rime length, stress, and association domains. *Journal of East Asian Linguistics* 2: 1–44.

(1994a). Against contour tones. *LI* 25: 555–608.

(1994b). Syllabic weight and syllabic duration: A correlation between phonology and phonetics. *Phonology* 11: 1–24.

Dwyer, David (1978). What sort of tone language is Mende? *Studies in African Linguistics* 9: 167–208.

Flemming, Edward (1995). *Auditory representations in phonology*. PhD dissertation, University of California, Los Angeles.

Fu, Guotong (1984). Wuyi fangyan de liandu biandiao (Tone Sandhi in the Wuyi dialect). *Fangyan* (Dialects) 1984: 109–27.

Gandour, Jackson T. (1974). The glottal stop in Siamese: Predictability in phonological description. *UCLA Working Papers in Phonetics* 27: 84–91.

(1978). The perception of tone. In Victoria A. Fromkin (ed.), *Tone: A linguistic survey*. New York: Academic Press, 41–76.

Gandour, Jackson T. and Richard A. Harshman (1978). Crosslanguage differences in tone perception: A multi-dimensional scaling investigation. *Language and Speech* 21: 1–33.

Goldsmith, John A. (1976). *Autosegmental phonology*. PhD dissertation, MIT. Reproduced by the Indiana University Linguistics Club.

Gordon, Matthew (1998). The process specific nature of weight: The case of contour tone restrictions. *WCCFL* 17: 235–49.

(1999). *Syllable weight: Phonetics, phonology, and typology*. PhD dissertation, University of California, Los Angeles.

Greenberg, Steven and Eric Zee (1979). On the perception of contour tones. *UCLA Working Papers in Phonetics* 45: 150–64.

Harris, James (1969). *Spanish phonology*. Cambridge, MA: MIT Press.

Hayes, Bruce (1989). Compensatory lengthening in moraic phonology. *LI* 20: 253–306.

He, Weitang (1986). Guangdongsheng Zengcheng fangyan yinxi (The phonemic system of the Zengcheng dialect of Guangdong Province). *Fangyan* (Dialects) 1986: 133–8.

Hoffman, Carl (1963). *A grammar of the Margi language*. London: Oxford University Press.

Hou, Jing-Yi. (1980). Pingyao fangyan de liandu biandiao (Tone sandhi in the Pingyao dialect). *Fangyan* (Dialects) 1980: 1–14.

Hudak, John (1993). *William J. Gedney's 'The Saek Language: Glossary, texts, and translations'. Michigan Papers on South and Southeast Asia*, No. 41. Center for South and Southeast Asian Studies, The University of Michigan.

Hudak, Thomas John (1987). Thai. In Bernard Comrie (ed.), *The world's major languages*. Oxford: Oxford University Press, 757–75.

Hyman, Larry M. (1985). *A theory of phonological weight*. Dordrecht: Foris.

Kao, Diana L. (1971). *Structure of the syllable in Cantonese*. The Hague: Mouton.

Kirchner, Robert (1996). Synchronic chain shifts in Optimality Theory. *LI* 27: 341–50.

(1997). Contrastiveness and faithfulness. *Phonology* 14: 83–111.

(1998). *An effort-based approach to consonant lenition*. PhD dissertation, University of California, Los Angeles. Rutgers Optimality Archive 276, http://roa.rutgers.edu.

Klatt, Dennis H. (1973). Interaction between two factors that influence vowel duration. *Journal of the Acoustical Society of America* 54: 1102–4.

(1975). Vowel lengthening is syntactically determined in connected discourse. *JPh* 3: 129–40.

Lanham, L. W. (1958). The tonemes of Xhosa. *African Studies* 17: 65–81.

Leather, Jonathan (1983). Speaker normalization in perception of lexical tone. *JPh* 11: 373–82.

Leben, William (1973). *Suprasegmental phonology*. PhD dissertation, MIT.

Lehiste, Ilse (1972). Timing of utterances and linguistic boundaries. *JASA* 51: 2108–24.

Li, Lan (1997). Guiyanghua yindang (The sound record of Guiyang). In Jing-Yi Hou (ed.), *Xiandai hanyu fangyan yinku* (The sound archives of modern Chinese dialects). Shanghai: Shanghai Jiaoyu Chubanshe.

Lindblom, Björn, Bertil Lyberg, and Karin Holmgren (1981). *Durational patterns of Swedish phonology: Do they reflect short-term motor memory processes?* Bloomington: Indiana University Linguistics Club.

Lyberg, Bertil (1977). Some observations on the timing of Swedish utterances. *JPh* 5: 49–59.

McCarthy, John and Alan Prince (1986). *Prosodic morphology*. Ms., University of Massachusetts and Brandeis University.

Miller, J. L. and F. Grosjean (1981). How the components of speaking rate influence perception of phonetic segments. *Journal of Experimental Psychology: Human Perception and Performance* 1: 208–15.

Moore, Corinne B. (1995). *Speaker and rate normalization in the perception of lexical tone by Mandarin and English speakers*. PhD dissertation, Cornell University.

Moore, Corinne B. and Allard Jongman (1997). Speaker normalization in the perception of Mandarin Chinese tones. *JASA* 102: 1864–77.

Morev, L.N., A.A. Moskalyov, and Y.Y. Plam (1979). *The Lao language*. Moscow: USSR Academy of Sciences, Institute of Oriental Studies.

Mugele, Robert (1982). *Tone and ballistic syllable in Lalana Chinantec*. PhD dissertation, University of Texas at Austin.

Newman, Paul (1972). Syllable weight as a phonological variable. *Studies in African Linguistics* 3: 301–23.

(1990). Hausa and the Chadic languages. In Bernard Comrie (ed.), *The world's major languages*. Oxford: Oxford University Press, 705–23.

Ohala, John J. (1978). Production of tone. In Victoria. A. Fromkin (ed.), *Tone: A linguistic survey*. New York: Academic Press, 5–39.

Paster, Mary (1999). *Issues in the tonology of Gã*. Senior honours thesis, Ohio State University.

Plomp, R. (1967). Pitch of complex tones. *JASA* 41: 1526–33.

Pols, Louis C. (1986). Variation and interaction in speech. In Joseph S. Perkell and Dennis H. Klatt (eds.), *Invariance and variability in speech processes*. Hillsdale, NJ: Lawrence Erlbaum Associates, 140–62.

Port, Robert F. (1979). The influence of tempo on stop closure duration as a cue for voicing and place. *JPh* 7: 45–56.

Prince, Alan and Paul Smolensky (1993). *Optimality Theory: Constraint interaction in generative grammar*. Rutgers Optimality Archive 537, http://roa.rutgers.edu/.

Ritsma, Roelof J. (1967). Frequencies dominant in the perception of complex sounds. *JASA* 42: 191–8.

Rupp, James E. (1990). The Lealao Chinantec syllable. InWilliam R. Merrifield and-Calvin R. Rensch (eds.), *Syllables, tone and verb paradigms.* (Studies in Chinantec languages) 4: Dallas: Summer Institute of Linguistics, 63–73.

Smolensky, Paul (1995). *On the internal structure of the constraint component of UG*. UCLA colloquium talk, 7 April, 1995. http://hebb.cog.jhu.edu/pdf/on_internal_structure_constraint_component_con_ug-handout.pdf.

Snyman, Jannie Winston (1970). *An introduction to the !Xũ! (!Kung) language*. (Communications from the School of African Studies, University of Cape Town, 34) Cape Town: A. A. Balkema.

Steriade, Donca (1993). *Neutralization and the expression of contrast*. Paper presented at *NELS* 24. University of Massachusetts, Amherst.

(1995). Underspecification and markedness. In John Goldsmith (ed.), *The handbook of phonological theory*. Oxford: Blackwell 114–74.

(1999). Phonetics in phonology: The case of laryngeal neutralization. In Matthew K. Gordon (ed.), *Papers in Phonology* 3, (UCLA Working Papers in Linguistics 2) Los Angeles: Department of Linguistics, UCLA, 25–146. http://www.linguistics.ucla.edu/people/steriade/papers/phoneticsinphonology.pdf

(2001a). Directional asymmetries in place assimilation: A perceptual account. In Elizabeth Hume and Keith Johnson (eds.), *The role of speech perception in phonology*. New York: Academic Press, 219–50.

(2001b). The phonology of perceptibility effects: The P-map and its consequences for constraint organization. Ms. University of California, Los Angeles.

Sundberg, Johan (1979). Maximum speed of pitch changes in singers and untrained subjects. *JPh* 7: 71–9.

Wightman, Collin W., Stefanie Shattuck-Hufnagel, Mari Ostendorf, and Patti J. Price (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *JASA* 91: 1707–17.

Wilson, Colin (2000). *Targeted constraints: An approach to contextual neutralization in Optimality Theory*. PhD dissertation, Johns Hopkins University.

Woo, Nancy (1969). *Prosody and phonology*. PhD dissertation, MIT.

Yip, Moria (1989). Contour tones. *Phonology*. 6: 149–74.

(1995). Tone in East Asian languages. In John. A. Goldsmith (ed.), *Handbook of theoretical phonology.* Oxford, UK: Blackwell, 476–94.

Young, Robert W. and William Morgan Sr (1987). *The Navajo language: A grammar and colloquial dictionary*, rev. ed. Albuquerque: University of New Mexico Press.

Zhang, Jie (1998). Checked tones in Chinese dialects – evidence for phonetically-driven phonology. *CLS* 34: 439–53.

(2000). The phonetic basis for tonal melody mapping in Mende and elsewhere. *WCCFL* 19: 603–16.

(2001). The contrast-specificity of positional prominence – evidence from diphthong distribution. Paper presented at the 75th annual meeting of the Linguistic Society of America, Washington, DC.

(2002a). *The effects of duration and sonority on contour tone distribution: Typological survey and formal analysis*. New York: Routledge.

(2002b). The positional markedness nature of contour tone licensing. Paper presented at the 76th annual meeting of the Linguistic Society of America, San Francisco, CA.

Notes

1. Here and elsewhere, L = low pitch, M = mid, H = high, LH = a contour rising from L to H, and similarly for other contours.
2. Tones here are denoted with the Chao letters (Chao 1948, 1968); ‘5’ and ‘1’ indicate the highest and lowest pitches in a speaker’s regular pitch range.
3. This kind of additive lengthening effect has been documented for English in Klatt 1973 and for Mandarin Chinese in Zhang 2002a. Both works show that a stressed syllable in prosodic-final position is longer than a stressless final syllable or a stressed non-final syllable.
4. In brief, Lealao Chinantec, Margi, and Zengcheng Chinese only have rising tones. This is unexpected given what we know about the differences between rising and falling tones. But they also have contour restrictions that are related to the duration and sonority of the rhyme in the expected direction: Lealao Chinantec limits contours to stressed syllables (Rupp 1990); Margi limits contours to monosyllabic words (Hoffman 1963); and Zengcheng Chinese limits contours to CVV and CVR (He 1986). In Lao (Morev et al. 1979), a rising tone and a high-falling tone can occur on CVO, but not on CVVO; and in Saek (Hudak 1993), complex tone 454 occurs on CVO, but not on CVVO. Without detailed phonetic description and historical knowledge of these languages, I take them as exceptions to the implicational laws and tendencies established in this section.

5. For discussion on other factors that could potentially increase the C_{CONTOUR} value of the rhyme, but do not behave as privileged contour tone licensers in any language in the survey, such as low vowels (as opposed to high vowels), voiced obstruent coda (as opposed to voiceless ones), see Zhang 2002a.
6. The fact that CVVO primarily carries HL and L and CVO primarily carries H and L can be understood from the following historical perspective. In Early Thai (pre-fifteenth century), there was no tonal contrast on checked syllables. Between the fifteenth and seventeenth centuries, a tonal split occurred: on CVVO, the split resulted in HL after a voiced onset and a L after a voiceless onset; on CVO, it resulted in a H after a voiced onset and a L after a voiceless onset. Possibly, the reason why a HL did not result on CVO was that there was not enough duration for the contour to surface.
7. Theoretically, there are various ways to increase the C_{CONTOUR} value of the syllable: increasing the sonorous rhyme duration, changing its sonorant coda into a vowel, making the syllable in question stressed, and so on. The factorial typology with the *CONTOUR(x)-C_{CONTOUR}(y) constraints and IDENT[length], IDENT[vocalic], IDENT[stress] should predict all these patterns. But in reality, I have not seen cases in which the sonorant coda is changed to a vowel or the stress is shifted in order to accommodate a contour tone. This is part of the 'too many solutions' problem, a general issue in Optimality Theory. For proposed solutions, see Steriade 2001b and Wilson 2000.
8. For more discussion on the similarity function and the consequence of using similarity function of this sort in the evaluation of faithfulness constraints, see Zhang 2002a.

7 Vowel reduction*

Katherine M. Crosswhite

1 Background on vowel reduction

Vowel reduction is a well-known phonological phenomenon; the idea that certain vowels might undergo qualitative changes in unstressed positions is likely to be familiar to anyone who has taken an introductory phonology course. Because this phenomenon can be so succinctly described – that is, 'unstressed vowels undergo neutralisation' – it is often assumed that vowel reduction is a unitary phenomenon, with a single formal analysis. In this chapter, I take the contrary position that vowel reduction has two different mechanisms.

Acknowledging the bipartite nature of vowel reduction is key to explaining what I refer to as 'reduction paradoxes' – cases in which vowel reduction patterns indicate that one and the same vowel is both highly marked (i.e. tends to be subject to reduction cross-linguistically) and highly unmarked (i.e. often serves as a reduction vowel, replacing other vowel qualities that are subject to reduction). This sort of paradox can be resolved by recognising two types of constraints that focus on unstressed vowel qualities, but that have separate teleologies. One type of constraint is based on the idea of *prominence*, and is implemented using prominence reduction constraints (Prince and Smolensky 1993). With respect to prominence-reducing vowel reduction, unstressed /a/ is disfavoured, being a highly sonorous vowel. The other is based on the idea of *contrast*, and is implemented using licensing constraints; specifically, licensing constraints focusing on avoiding unstressed noncorner vowels. In this sort of vowel reduction, unstressed /a/ is favoured, since /a/ is one of the three corner vowels /i, u, a/. In what follows, I will lay out the constraints motivating these two types of reduction, their phonetic motivations, and examples of how they work.

An additional point to be made is that the two constraint families alluded to above only identify the vowels to be eliminated by vowel reduction and the contexts in which they are to be eliminated. They do not, however, identify the method for eliminating them. Invoking these constraints within Optimality Theory predicts that languages will vary in the neutralisations used to meet the demands of these constraints, and furthermore that the set of possible neutralisations will correspond to the logical possibilities predicted by a factorial typology combining these reduction constraints with constraints on vowel faithfulness. Indeed, an empirical survey of vowel reduction languages demonstrates a wide variety of vowel reduction patterns, including multiple cases where the same unstressed sub-inventory is achieved via different sets of neutralisations, and furthermore that the observed patterns are a good fit for a factorial typology based on the analysis of vowel reduction provided here (Crosswhite 2001).

1.1 A note on vowel features and vowel faithfulness

One vowel quality that shows up quite commonly in vowel reduction is [ə]. I assume that reduced [ə] is essentially a targetless vowel that is not specified for any vowel features (except perhaps, following Browman and Goldstein 1992, a specification for minimal opening). For example, I assume that in English the weak, reduced vowel [ə] and the full vowel [ʌ] differ in featural representation: whereas [ʌ] is specified mid central, [ə] is not. Assuming this representation, I adopt a DEP/MAX approach to vowel faithfulness (Zoll 1996). For example, the reduction of /i/ to [ə] would require the *deletion* of the feature specifications [+high, +front, −low], in violation of the corresponding MAX[F] constraints. Similarly, reduction to a non-schwa vowel will be treated as the deletion of certain feature specifications, and the insertion of certain others. For example, the reduction of /e/ to [i] would be represented as deletion of [−high] (violating MAX[−high]) and insertion of [+high] (violating DEP[+high]).

2 Contrast-enhancing reduction

The first type of vowel reduction I will consider is contrast-enhancing reduction, in which certain undesirable or perceptually challenging vowel qualities are limited to stressed position. In general, this form of reduction will amount to elimination of noncorner vowels, especially mid vowels. One such reduction pattern is found in Belarusian (Krivitskii and Podluzhnyi 1994), in which the mid vowels /e, o/ both reduce to [a]:

(1) Vowel neutralisations in Belarusian (Krivitskii and Podluzhnyi, 1994)

Vowels under stress		Same vowels unstressed	
ˈn**o**ɣi	‘legs’	n**a**ˈɣa	‘leg’
ˈk**o**l	‘pole’ (nom.)	k**a**ˈla	‘pole’ (gen.)
ˈvʲ**o**snɨ	‘spring’ (gen.)	vʲ**a**ˈsna	‘spring’ (nom.)
ˈmʲ**o**t	‘honey’ (n.)	mʲ**a**ˈdovɨ	‘honey’ (adj.)
ˈʃ**e**pt	‘whisper’	ʃ**a**pˈtatsʲ	‘to whisper’
ˈr**e**ki	‘rivers’	r**a**ˈka	‘river’
ˈspʲ**e**tsʲ	‘to ripen’	paˈspʲ**a**vatsʲ	‘to mature’
ˈklʲ**e**j	‘glue’	klʲ**a**ˈjonka	‘oil-cloth’

This form of reduction produces an end result in which unstressed syllables are limited to the vowel sub-inventory [i, u, a]. As noted by Lindblom (1986), this type of vowel inventory shows maximal dispersion and, therefore, minimal acoustic ambiguity. This and similar types of vowel reduction will be motivated using licensing constraints, which are discussed below.

2.1 Licensing constraints and contrast enhancement

Not all speech sounds are perceived equally well; furthermore, not all speech sounds are equally good in all segmental or prosodic environments. From the speaker's point of view, it may be undesirable for a speech sound to be misperceived – not merely out of charitable concern for the listener, but also out of selfish reasons. If you produce a speech sound that is misperceived, you have expended articulatory effort in an ineffective manner. This approach is the basis for Steriade's (1994a, b) licensing-by-cue approach to phonological neutralisations. If a given contrast is in danger of being missed by the listener, why should the speaker go to the trouble of producing it? In other words, positional neutralisations based on the desire to avoid ineffectual expenditure of articulatory effort can be thought of as the grammatical encoding of the speaker's preference to 'not deploy a feature in positions where its defining [acoustic] cues are necessarily absent or diminished' (Steriade 1994a). There are two logical courses of action for a speaker who wants to avoid ineffectual articulation: (1) Don't say the sound at all, or (2) Say a different, but similar, sound. In this study, I do not consider strategy (1) (deletion), instead I focus on strategy (2) (neutralisation).

To account for contrast-enhancing vowel reduction, I will use licensing constraints – a non-faithfulness-based version of Steriade's (1994a) Implement constraint family. (See also Steriade's (1994b) positional neutralisation constraints, as well as the stress-prominence constraints used by Majors (1998) in the analysis of stress-dependent vowel harmony.) The form of a licensing constraint is as follows:

(2) Lic-Q/β: The vowel quality Q is only licensed in context β,
where Q = any vowel quality or a natural vowel class
β = any context that enhances the accurate perception of Q

Note that a licensing constraint cannot combine just any vowel quality Q with just any context β: licensing is constrained by the requirement that the context β must enhance the accurate perception of Q. It should also be underscored that licensing constraints are not members of the faithfulness constraint family. A constraint such as Lic-Q/β will assign a violation mark for every instance of [Q] that occurs without β, irrespective of whether [Q] is underlying or derived.

In this respect, licensing constraints are similar to the grounding conditions discussed by Archangeli and Pulleyblank (1994). For example, drawing on the fact that tongue root advancement and tongue body raising are articulatorily compatible gestures, while the combination of tongue root advancement and tongue body lowering are articulatorily antagonistic, they posit constraints of the type ATR/Low, which prohibits the [−ATR] feature specification from co-occurring with a [−low] specification (i.e. vowels that are [−ATR] must be low). The licensing constraints used here take a similar approach, applying not to combinations of features (as in ATR/Low), but to combinations of features with positions (such as stressed position). It should also be pointed out that Archangeli and Pulleyblank's motivation for the ATR/Low constraint (and, indeed, for most of their grounding conditions) is based on articulatory considerations: retraction of the tongue root allows easier depression of the tongue body – thus [−ATR] and [+low] are *articulatorily* compatible. Archangeli and Pulleyblank also allow grounding conditions to refer to acoustic compatibility of the type discussed with respect to licensing above, but their emphasis is usually on articulatory compatibility.[1] In the enhancement-based licensing constraints used here, articulatory considerations do not play a role, although acoustic compatibility is required (the licensing context β must enhance correct perception of Q).

The particular licensing constraints considered here are those focused on particular vowel qualities in stressed position. The licensing constraint that motivates most cases of contrast-enhancing vowel reduction is:

(3) Lic-Noncorner/Stress: Noncorner vowels are licensed only in stressed positions.

In order to maintain the phonetic motivation for contrast-enhancing vowel reduction, it is necessary to demonstrate two facts: (1) that noncorner vowels (the quality Q) are subject to misperception, and (2) that stressed position (the context β) enhances their correct perception. These two issues are addressed separately below.

2.2 *Corner vs noncorner vowels*

The corner vowels /i, u, a/ show several special properties: they are the three most common vowels, occurring in almost all languages (Maddieson 1984); and, as a set, they constitute the smallest complete vowel inventory found with any regularity in the world's languages (Maddieson 1984; Lindblom 1989). Although much remains to be learned about these vowels, three characteristics work to single out /i, u, a/ from all remaining vowel qualities: dispersion

(Lindblom 1986), quantal characteristics (Stevens 1989), and focalisation (cf. Stevens 1989; Schwartz et al. 1997).

Dispersion refers to the efficient use of the acoustic space available. Speech sounds should be well dispersed throughout this space, so as to increase the distinctiveness of each of the sounds from the others. It is for this reason that certain vowel inventories are very common across the world's languages: they locate vowels at various points in acoustic space that maximise the acoustic distance between the members of the inventory. This idea, often referred to as Dispersion Theory, has been articulated in a number of articles by Bjorn Lindblom and collaborators (Liljencrants and Lindblom 1972; Lindblom 1986; Lindblom and Maddieson 1988; et al.) Under this theory, the corner vowels /i, u, a/ are special in that they are maximally acoustically distinct: in theory, a vowel system consisting of these three vowels would be easiest in terms of perception because the possibility for confusing an intended vowel quality for an incorrect but adjacent vowel quality is minimised. However, this cannot be the only motivation for treating the corner vowels /i, u, a/ as special. As illustrated in Lindblom 1986, the distance metric alone does not uniquely identify /i, u, a/ as having this quality. The vowel inventory /i, u, ɔ/ also displays maximal dispersion (and, as pointed out by Lindblom, is an attested three-vowel inventory).

The vowels /i, u, a/ are also special in terms of their production: they all show quantal effects (Stevens 1989). A quantal effect occurs when a given change in articulation does not produce a correspondingly large acoustic change. That is, speech sounds that show quantal effects are ones in which the appropriate acoustic quality is more or less consistent with a wide range of articulations. Non-quantal sounds, in contrast, show large changes in acoustic quality for similarly sized articulatory changes. Quantal Theory (Stevens 1989) hypothesises that languages prefer to use speech sounds that show these quantal effects, presumably because they are consistent with a wider range of articulations, and thus easier to produce under a wide range of contexts. In particular, the vowels /i, u, a/ all show this sort of effect. For example, as illustrated by Perkell and Nelson (1982, 1985) and Perkell and Cohen (1986), the articulation of /i, u, a/ does show significant variation in constriction location, but not in terms of constriction degree, supporting the idea that these vowel qualities are fairly stable in the face of certain types of articulatory changes as identified by Quantal Theory. However, it is again the case that quantal effects alone are insufficient to explain why /i, u, a/ constitute a special class: it is not the case that /i, u, a/ are the only vowels that show such effects. Indeed, it has been suggested that central vowels like /ʌ/ and /ə/ also show quantal effects (Ladefoged et al. 1977; Pisoni 1980). However, as pointed out by Diehl (1986), the vowels /i, u, a/ are distinguished by having not only articulatory stability, but also in the fact that they do not share this characteristic with adjacent qualities. That is, the vowels /i, u, a/ not only occupy areas of acoustic stability in the face of articulatory variation, but they are surrounded by areas of instability, where acoustic quality is comparatively sensitive to articulatory variation.

Furthermore, Stevens (1989) also points out that the quantal vowels /i, u, a/ and /y/ are perceptually special by virtue of having spectral prominences caused by convergences: either proximity of two formants (/i/: F3, F4; /a/: F2, F1; /y/: F2, F3), or proximity of the first formant and the fundamental frequency (/u/). Stevens suggests that these proximities are (or tend to be) less than some critical distance necessary for distinguishing both prominences; that is, that the two formants (or F1 and F0) are so close in these vowel qualities as to converge auditorily into a single spectral prominence.

Schwartz et al. refer to the presence of such convergences as giving a vowel a 'focal' property. Based on the finding that these focalisations lead to more stable patterns in discrimination tasks, Schwartz et al. hypothesise that they also lead to increased perceptual salience. For an intuitive explanation of this effect, they quote Lieberman (1971: 57–8) as saying that these vowels 'provide acoustic salience: that is, their formant frequency patterns yield prominent spectral peaks (formed by the convergence of two formant frequencies [. . .]) that make it easier to perceive the sounds, just as, in the domain of colour vision, saturated colours are easier to differentiate than muted ones'. Schwartz et al. then propose a hybrid Dispersion-Focalisation Theory, which seeks to maximise not just distance between vowels (inter-vowel salience, dispersion), but also the intra-vowel salience or 'local focalisation'. Under this approach, the possible three-vowel inventories /i, u, a/ and /i, u, ɔ/ (both of which maximise dispersion) are distinguished, with /i, u, a/ being more optimal since it maximises both dispersion and focalisation.

Given the above, it does seem clear that the vowels /i, u, a/ do have a special status. This status does not seem to be linked to any single characteristic, but perhaps to the amalgamation of several, making this set of vowels particularly auspicious. Their qualities are stable against articulatory variation, they are maximally acoustically distinct from one another, and they are perceptually salient due to convergences of spectral prominences. Not only are they individually 'good' vowel qualities, but as a set they also constitute a particularly desirable vowel inventory. It therefore seems reasonable to suppose that precisely these vowels would be the most desirable to use in cases where correct vowel perception is at risk. Put another way, vowels that are not members of this set might reasonably be limited, in some languages, to those positions where their correct perception is most favourable. As such, this set of vowels seems like a plausible candidate for the quality Q mentioned in a licensing constraint – that is, the set that requires a particular environment to be phonologically licensed. With this in mind, I now turn to consider the other side of the question, namely the position where these vowels are licensed.

2.3 *Motivating licensing under stress*

As mentioned earlier, this section specifically focuses on cases where noncorner vowels are neutralised in unstressed syllables. As just outlined, corner vowels have several special characteristics that make them particularly desirable phonetically. However, is it the case that stressed position is a context that would increase the correct perception of a noncorner vowel? In general, increasing the exposure to any stimulus increases the likelihood that the stimulus will be correctly identified. Assuming that stress engenders increased duration, the condition placed on licensing constraints (β must enhance correct perception of Q) is met, and stressed position is an appropriate context for a licensing effect.

In connection with this, it is interesting to note that the constraint proposed, LIC-NONCORNER/STRESS, appears to be active only in those languages where stress is correlated with increased duration: I have found no examples of stress-induced licensing of vowel features that occur in either languages where stress is strictly intensity-based (i.e. stressed vowels are not longer than unstressed ones, as in contemporary Czech (Palková 1994)), or in languages that use pitch accent. There are, however, interesting examples where some variety of a language that predominantly shows pitch accent or intensity-based stress does possess vowel reduction. In all such cases I have found, the accentual system of the innovating dialect has replaced the prosodic system of the predominating dialects with a duration-based stress system. Examples include Standard Slovene (Bezlaj 1939; Toporišič 1976; Lenček 1982); Horjulj Slovene (Lenček 1982: 145), Botzetierra Basque (Hualde 1991), and dialectal Polish (Urbańczyk 1953: 11). There are also cases that go the other way around, in which a language shows predominantly duration-based stress, but also has dialects that lack vowel reduction. In these cases, the dialects in question lack a strong duration-based stress (cf. Russian with *okan'e*; Kasatkin 1989).

2.4 *Effect of* LIC-NONCORNER/STRESS

The LIC-NONCORNER/STRESS constraint induces elimination of unstressed mid vowels. In the Belarusian example provided earlier, both the unstressed mid vowels /e, o/ were eliminated via lowering to [a]. However, this is not the only possible result for LIC-NONCORNER/STRESS. For example, the exact opposite phenomenon – reduction via raising – is also attested. An example of this reduction pattern is found in the Native American language Luiseño (Munro and Benson 1973):

(4) Vowel neutralisations in Luiseño (Munro and Benson 1973)

	Vowels under stress		Same vowels unstressed	
unstressed	ˈtʃ**o**ka	'to limp'	tʃ**u**ˈkatʃkaʃ	'limping'
/e, o/ raise	ˈh**e**din	'will open'	h**i**ˈdiki-	'to uncover'
	tʃaˈp**o**mkat	'liar'	ˈtʃaʃp**u**mkatum	'liars'
unstressed	ˈm**a**ha	'to stop'	m**a**ˈhamhaʃ	'slow'
/i, u, a/ do	ˈk**u**ˑmit	'smoke'	k**u**ˈmikmiʃ	'smoky colored'
not reduce	ˈʂ**u**ˑkat	'deer'	ˈpaˑʂ**u**kat	'elk'
	taˈk**i**tkiʃ	'straight'	ˈtaˑk**i**ʃ	'stone for smoothing pottery'

In the Luiseño case, unstressed mid vowels are eliminated by raising. Note that the remaining vowels do not undergo reduction, once again creating the maximally dispersed vowel sub-inventory [i, u, a] in unstressed positions. The difference between Belarusian and Luiseño can easily be accounted for by using different rankings for vowel faithfulness constraints: Belarusian eliminates unstressed mid vowels in a manner that preserves their underlying nonhigh nature, while Luiseño eliminates the same vowels in a manner that preserves their underlying colour. The following partial grammars illustrate this state of affairs:

(5) Partial grammars for Belarusian and Luiseño vowel reductions

Belarusian:
LIC-NONCORNER/STRESS, MAX[−high] ≫ MAX[round], MAX[+front]

Luiseño:
LIC-NONCORNER/STRESS, MAX[round], MAX[+front] ≫ MAX[−high]

In both the Belarusian and Luiseño cases, the LIC-NONCORNER/ STRESS constraint is undominated (at least in these partial grammars), thus motivating the elimination of unstressed /e, o/. In addition, in each case at least one, if not two, vocalic faithfulness constraints are also undominated. In the case of Belarusian, the undominated faithfulness constraint is MAX[−high], which derives reduction-via-lowering, as illustrated below for two hypothetical forms:

(6) Reduction via lowering (e.g. Belarusian)

/to'ta/ *'hypothetical word'*	LIC-Noncorner	MAX [−high]	MAX [round]	MAX [+front]
☞ [ta'ta]			*	
[tu'ta]		*!		
[to'ta]	*!			

/teta/ *'hypothetical word'*	LIC-Noncorner	MAX [−high]	MAX [round]	MAX [+front]
☞ [ta'ta]				*
[ti'ta]		*!		
[te'ta]	*!			

As illustrated, the two topmost constraints do not need to be ranked with respect to one another – the fact that the colour-based faithfulness constraints are at the bottom of the constraint hierarchy is adequate to derive reduction via lowering. By reversing the position of the colour-based faithfulness constraints and MAX[−high], reduction via raising is derived:

(7) Reduction via raising (e.g., Luiseño)

/to'ta/ *'hypothetical word'*	LIC-Noncorner	MAX [round]	MAX [+front]	MAX [−high]
☞ [tu'ta]				*
[ta'ta]		*!		
[to'ta]	*!			

/teta/ *'hypothetical word'*	LIC-Noncorner	MAX [round]	MAX [+front]	MAX [−high]
☞ [ti'ta]				*
[ta'ta]			*!	
[te'ta]	*!			

Again, the only ranking that need be specified to derive reduction via raising is that MAX[−high] must be at the bottom of the constraint hierarchy. Other logically possible permutations of these constraints predict yet other neutralisation patterns for eliminating unstressed mid vowels, all of which seem to be attested. For example, by ranking MAX[round] at the bottom of the constraint hierarchy, an asymmetrical reduction pattern (/e/ > [i], /o/ > [a]) is predicted, as shown below. This type of pattern is attested in Contemporary Standard Russian.

(8) An asymmetrical reduction pattern (e.g. Russian)

/to'ta/ *'hypothetical word'*	LIC-Noncorner	MAX [+front]	MAX [−high]	MAX [round]
☞ [ta'ta]				*!
[tu'ta]			*	
[to'ta]	*!			

/teta/ *'hypothetical word'*	LIC-Noncorner	MAX [+front]	MAX [−high]	MAX [round]
☞ [ti'ta]			*	
[ta'ta]		*!		
[te'ta]	*!			

Reversing the position of MAX[+front] and MAX[round] predicts the opposite asymmetrical pattern: /e/ > [a], /o/ > [u]. This pattern is attested in Algueres Catalan (Recasens 1991). Similarly, by adding additional vocalic faithfulness constraints, even more variations on the elimination of unstressed mid vowels are accounted for. Some of these additional patterns include [ATR]-preserving patterns (cf. Bergün Romansch: /ɛ, ɔ/ > [a], /e/ > [i], /o/ > [u], Lutta 1923; Kamprath 1991), and [low]-preserving patterns (cf. Saipanese Chamorro: /æ, ɑ/ > [a]). For a more detailed description of predicted reduction patterns and their attestation, see Crosswhite 2001.

2.5 *Licensing-based reduction creating novel vowel qualities*

Another type of licensing-based reduction creates as its output vowels that do not belong to the basic inventory of vowel phonemes. We consider first a case from Slovene. In Slovene, there are eight phonemic vowel qualities: /i, u, e, o, ɛ, ɔ, a, ə/. Note the phonemic contrast between lax and tense vowels, as illustrated by pairs such as ['tseːsta] 'road' vs ['sɛːstra] 'sister'. However, this opposition is only maintained in long, accented syllables. Since stress is mobile in Slovene, this produces tense ∼ lax alternations, as illustrated below (data from Bidwell 1969):

(9) Vowel neutralisations in Standard Slovene (Bidwell 1969)

	Vowels under stress		Same vowels unstressed	
unstressed ɔ	ˈgɔːra	‘mountain’ nom. sg.	gɔˈreː	‘mountain’ gen. sg.
(no change)	ˈpɔːtɔk	‘stream’ nom. sg.	pɔˈtoːka	‘stream’ gen. sg.
unstressed ɛ	ˈpl**ɛ**ːma	‘tribe’ nom. sg.	pl**ɛ**ˈmeːna	‘tribes’ nom. pl.
(no change)	ˈs**ɛ**ːstra	‘sister’ nom. sg.	s**ɛ**ˈstreː	‘sister’ gen. sg.
unstressed o > ɔ	ˈm**o**ʒ	‘man’ nom. sg.	mɔˈʒjeː	‘men’ nom. pl.
	ˈk**o**ːst	‘bone’ nom. sg.	kɔˈstiː	‘bone’ gen. sg.
unstressed e > ɛ	ˈr**e**ːtʃ	‘word’ nom. sg.	r**ɛ**ˈtʃiː	‘word’ gen. sg.
	ˈts**e**ːsta	‘road’ nom. sg.	ts**ɛ**ˈsteː	‘road’ gen. sg.

The tense ∼ lax distinction is also neutralised in short, accented syllables, as illustrated by forms such as [ˈkm**ɛ**ːta] ∼ [ˈkm**ɛ**t] ‘peasant’ (gen./nom.) As illustrated, the traditional description is that when the tense ∼ lax distinction is neutralised, it is in favour of the lax mid vowels [ɛ, ɔ]. The only other known case of reduction via laxing is from certain north-eastern dialects of Brazilian Portuguese (Brakel 1985; Perrone and Ledford-Miller 1985). The existence of reduction via laxing is something of an anomaly, since the opposite form of reduction – reduction via tensing (see section 3.7) – seems much more common, and moreover fits in with the theoretical approaches to vowel reduction presented here. However, instrumental analysis of the Slovene tense ∼ lax neutralisation by Lehiste (1961) indicates that there is more to the story than would be guessed based on examining traditionally transcribed forms, such as those presented above. She finds that the neutralised vowels usually transcribed as [ɛ] and [ɔ] are actually intermediate between non-neutralised /e/ and /ɛ/ or /o/ and /ɔ/. Lehiste therefore proposes that the Slovene tense ∼ lax neutralisation is actually a case of archiphonemic neutralisation: the neutralised vowels are simply unspecified for laxness or tenseness and might, therefore, be more accurately transcribed as the archiphonemes [E, O]. The graph in figure 7.1, based on the vowel formant measurements reported by Lehiste (1961), illustrates this situation. This is in stark contrast with reduction via tensing. For example, an underlying tense ∼ lax distinction is also neutralised in many dialects of Italian, but this time in favour of the tense vowels. Data illustrating this phenomenon are provided below:

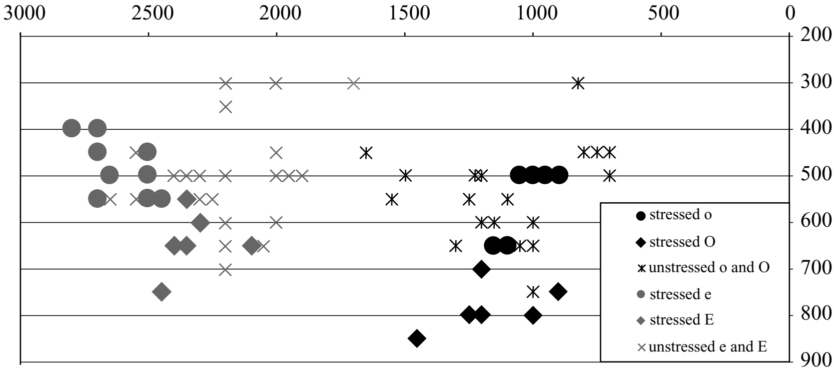


Figure 7.1 Stressed and Unstressed Mid Vowels in Slovene (data from Lehiste 1961)

(10) Vowel neutralisations in Standard Italian

	Vowels under stress		Same vowels unstressed	
unstressed ɔ > o	akˈkɔʎʎere	‘to receive’	akk**o**ʎˈʎɛnte	‘welcoming’
	ˈfɔʎʎo	‘page’	f**o**ʎˈʎetto	‘slip of paper’
unstressed ɛ > e	ˈb**ɛ**lla	‘beautiful’	b**e**lˈlettsa	‘beauty’
	ˈp**ɛ**ddʒo	‘worse’	p**e**dˈdʒore	‘to worsen’
unstressed o	ˈs**o**le	‘sunlight’	s**o**ledˈdʒato	‘sunny’
(no change)	ˈb**o**kka	‘mouth’	b**o**kˈkaʎʎo	‘mouthpiece’
unstressed e	ˈp**e**lo	‘hair’	p**e**ˈloso	‘hairy’
(no change)	ˈr**e**te	‘net’	r**e**ˈticolo	‘network’

Instrumental analysis by Baroni (1996) demonstrates that the neutralised vowels in Italian are not intermediate between the qualities observed for the stressed lax and tense mid vowels. Rather, the neutralised vowels (transcribed as unstressed [e, o] above) are in fact qualitatively indistinguishable from stressed [e, o]. Thus, we have a contrast: a tense ∼ lax distinction is eliminated in unstressed syllables in Italian by replacement (lax vowels are replaced by tense ones), while in Slovene a tense ∼ lax distinction is eliminated in unstressed (and short stressed) syllables by feature elimination (both lax and tense vowels are replaced by an archiphoneme not specified for [ATR]).

A case that is formally similar to the Slovene tense ∼ lax neutralisation is found in Eastern Ojibwa (Bloomfield 1957; Miller 1972). In Eastern Ojibwa, there are basically three monomoraic vowel qualities: /ɪ, ʊ, a/. In unstressed positions (described by Bloomfield (1957) simply as odd syllables counting from the left edge of the word), /ɪ/ and /a/ reduce to [ə], while /ʊ/ reduces to a rounded variant of schwa. Superficially, this pattern resembles reduction of all vowels to [ə] (see section 3.7); however, it differs in that one distinctive feature, namely [round], is preserved under reduction. Recall here that [ə] is analysed as a featureless vowel. Therefore rounded schwa would be represented as a surface vowel with a specification for [round], but no specification for height or advancement.

The licensing effects discussed in the preceding sections can be described as bans on certain vowel qualities in unstressed positions, leading to the elimination of certain vowel specifications. However, both the Slovene and Eastern Ojibwa cases could be modelled by banning certain *vowel distinctions* in a given position, leading to the wholesale absence of some feature in that position. In the case of Slovene, the feature [ATR] cannot occur in monomoraic positions. In the case of Eastern Ojibwa, the features [high] and [front] do not occur in monomoraic unstressed positions. This might suggest licensing constraints such as the following:

(11) Lic-[ATR]/μμ: Feature specifications for [ATR] may only occur in association with bimoraicity.

(12) Lic-[high, front]/Stress: Feature specifications for [high] and/or [front] may only occur in association with stress.

This sort of licensing constraint can be seen as establishing a second vowel inventory for unstressed and/or short positions by redefining the set of features allowable in these contexts. That is, this sort of licensing constraint actually introduces new vowel qualities that occur in unstressed or short positions, but are absent from stressed or long positions. This apparent increase in vocalic complexity dependent on stress/length is seemingly mitigated by the fact that the unstressed/short vowel inventory uses fewer features, and therefore makes fewer vowel distinctions, despite the fact that it introduces additional vowel qualities.

3 Prominence reduction

In the preceding section, I discussed a number of vowel reduction systems based on the idea of avoiding noncorner vowels in unstressed position. In this type of system, the unstressed vowel [a] was considered to be highly desirable. Interestingly, there are also vowel reduction systems that seem to suggest exactly the opposite conclusion: in these systems, [a] is *illegal* in stressless position, and reduces to [ə]. A resolution of this paradox, as we will see, can be found in enriching the constraint system, attributing constraints with conflicting effects to conflicting phonetic teleologies.

One example of a phonological system that reduces stressless [a] to schwa is found in Bulgarian. As in various systems discussed so far, the noncorner vowels /e, o/ are neutralised in unstressed syllables, surfacing as [i, u], respectively, as shown in (13a–c). However, the vowel /a/ is reduced to [ə] (13d–e). Clearly, this reduction pattern is at odds with the phenomenon of contrast enhancement.

(13) Vowel reduction in Bulgarian

	Vowels under stress		Same vowels unstressed	
a.	ˈr**o**guf	'of horn'	r**u**ˈgat	'horned'
b.	ˈ**o**nzi	'that' (masc.)	**u**ˈnazi	'that' (fem.)
c.	ˈs**e**lu	'village'	s**i**ˈla	'villages'
d.	ˈr**a**butə	'work'	r**ə**ˈbotnik	'worker'
e.	ˈgr**a**t	'city'	gr**ə**ˈdets	'town'

A further fact, which fits into this general pattern, is that Bulgarian also has a phonemic /ə/, which does not undergo neutralisation: unstressed /ə/ emerges without change. This only adds to the puzzle. Since the vowel quality [ə] is at least as nonperipheral as the qualities [e, o] (if not more so), why does this vowel escape reduction? To solve this type of vowel reduction paradox, I propose that Bulgarian vowel reduction, and similar reduction types, constitute a formally distinct type of vowel reduction, which I refer to as *prominence reduction*. Prominence-reducing vowel reduction is based on the desire to avoid particularly long or otherwise salient vowel qualities in unstressed positions. A wide variety of vowel reduction patterns seem to fall into this category. This type of vowel reduction will be modelled here using prominence alignment constraints, as discussed by Prince and Smolensky (1993).

3.1 Prominence alignment

The formal mechanism of prominence alignment was first employed by Prince and Smolensky (1993) in an analysis of predictable syllabicity in Imdlawn Tashlhiyt Berber, in which any segment can be syllabic, and the choice of which individual segments will surface as syllable nuclei in any given word is predicted by relative sonority. Specifically, more sonorous segments are chosen to occupy the syllable nuclei positions, while less sonorous segments are chosen to occupy syllable onset positions. Prince and Smolensky analyse this pattern as a case of 'stacking', or aligning, prominent elements. Assuming that syllable nucleus position is a prominent prosodic position, and furthermore that increasing segmental sonority is correlated with increasing prominence, the Imdlawn Tashlhiyt Berber pattern can be accounted for simply by saying that prominence at both the segmental and syllabic levels should co-occur. That is, combinations of prominent syllabic position plus non-prominent segmental material (or vice versa) should be avoided. This is formalised by Prince and Smolensky starting with identification of two 'phonetic scales'. The two scales they used are illustrated below. (The symbol '$_{prom}>$' means 'is more prominent than'.)

(14) Scale 1: Syllabic prominence
peak $_{prom}>$ margin
Scale 2: Segmental prominence (sonority)
a $_{prom}>$ e, o $_{prom}>$ i, u $_{prom}>$ l, r $_{prom}>$ n, m $_{prom}>$ *etc.*

These two 'phonetic scales' are then crossed, as follows. Choose one element of Scale 1 and cross it, in order, with each member of Scale 2. The resulting combinations form the basis of Optimality-Theoretic constraints militating against that particular combination of elements. These constraints are inherently ranked with respect to one another, mirroring the order of elements in Scale 2.

Since Scale 1 has two members, there are two different ways the crossing operation can proceed, producing constraints that either focus on the prominent member of Scale 1, or the non-prominent member of Scale 1. The two constraint families that can be made by crossing the two scales shown in (14) are illustrated below:

(15) *MARGIN/a ≫ *MARGIN/e, o ≫ *MARGIN/i, u ≫ *MARGIN/l, r ≫ *MARGIN/n, m ≫ . . .
*(*MARGIN/X = X is not a syllable margin.)*

(16) . . . ≫ *PEAK/n, m ≫ *PEAK/l, r ≫ *PEAK/i, u ≫ *PEAK/e, o ≫ *PEAK/a
*(*PEAK/X = X is not a syllable peak.)*

Note that the constraints in (15) focus on the non-prominent member of Scale 1, namely, syllable margin position. Following Jian-King (1996), I will refer to this type of constraint family as a prominence reduction family. Note that a prominence reduction family has members ranked in order of decreasing sonority, encoding the idea that reductions of prominence are preferred in non-prominent positions. For example, the fact that *MARGIN/a is the highest-ranked member of this family expresses the idea that the vowel quality [a] is in fact the worst possible syllable margin imaginable – the fact that it is the highest-ranked constraint means that it is the most difficult to violate. Similarly, the constraints in (16) focus on the prominent member of Scale 1, namely, syllable nucleus position. I will refer to this type of family as a prominence alignment family.

Note that its members are ranked in order of increasing sonority, expressing the idea that high prominence is preferred in prominent positions. For example, the fact that *PEAK/a is the lowest-ranked member of this family expresses the idea that [a] is in fact a very good syllable nucleus – the fact that it is the lowest-ranked member means that it is the easiest one to violate.

The idea of crossing two phonetic scales is not limited to sonority and syllabicity. In theory, any two prominence scales can be crossed in a similar way. For example, Kenstowicz (1994) employs this technique to cross sonority with stress (stressed $_{prom}>$ unstressed), explaining why certain high-sonority vowels in some languages attract stress, while certain low-sonority vowels in other languages repel stress. Similarly, Gordon (this volume) uses constraints crossing stress with rhyme prominence to explain patterns of syllable weight, and Crosswhite (2001) crosses sonority with moraicity to explain why certain high-sonority vowels undergo lengthening in stressed position more easily than low-sonority vowels.

3.2 Motivating prominence alignment

Prince and Smolensky (1993) make it clear that the prominence alignment mechanism is intended to be phonetically motivated: the inherent ranking of constraints in these families is intended to mirror some specific physical continuum that could, presumably, be objectively established through measurement. However, the exact physical continuum appropriate for this is far from clear. For example, it has been noted several times that the phonological sonority hierarchy cannot straightforwardly be reduced to any single phonetic characteristic. For example, Fry (1979) attempts to correlate phonological sonority in English with intensity (louder = more sonorous), while Lindblom (1983) suggests a basis for sonority in jaw opening (more jaw depression = greater sonority). However, both of these proposals fail to generate the precise ordering employed by phonologists in analyses of, say, sonority sequencing in syllabification. The major problems with these approaches are laid out by Keating (1983) and Malsh and Fulcher (1989). For example, although segments can be classed by their preferred degree of jaw opening (as in Lindblom 1983), it turns out that most consonants can easily accommodate a wide range of jaw positions, while relatively few segments (such as /s/) have a more demanding, less variable jaw position (Keating 1983). Furthermore, although intensity correlates with sonority fairly well for sonorant segments like vowels, liquids, and nasals, defining sonority in terms of intensity makes incorrect predictions concerning obstruents. Certain obstruents, such as fricatives, have a fairly high intensity due to the presence of noise produced at a constriction site somewhere in the vocal tract, yet they are considered to be low-sonority segments.

In response to these difficulties, some researchers have proposed that there is not a single physical correlate for sonority, but many. For example, Malsh and Fulcher (1989) propose that sonority is correlated to intensity plus jaw opening – under their framework, segments that are both loud and open (such as vowels and sonorous consonants) will be highly sonorous, while segments that are both quiet and close (such as stops) are highly non-sonorous. Segments that are ambiguous, such as /s/, which has a close jaw position but high intensity, will be ordered on a language-specific basis. In a similar tack, Nathan (1989) proposes a number of physical correlates (voicing, openness, 'prolongability') that tend to make a segment more prototypically sonorous. However, I propose that it is too early to give up hope for a straightforward physical correlate for sonority. In connection with this, it is worth noting a similar situation regarding another prominence-based phonological phenomenon, syllable weight. As demonstrated by Gordon (this volume), syllable-weight hierarchies are in some cases straightforwardly correlated with simple duration measurements, but this correlation is imperfect – a slightly different measure, total perceptual energy, provides a better basis for predicting syllable-weight distinctions. With this in mind, it may be the case that some other, more specific, measure may provide a phonetic basis for the phonological sonority hierarchy. For example, in Keating's (1983) discussion of sonority, she points out that jaw opening and the frequency value for the first formant (F1) are both correlated with amplitude at low frequencies: lower jaw position correlates with a higher F1 and higher low-frequency amplitude. In contrast, noise associated with obstruents tends to be broad spectrum or high frequency. Similarly, Stevens (1989: 35–7) notes that sonorant and non-sonorant consonants differ in low-frequency amplitude: producing a constriction somewhere in the vocal tract in turn decreases transglottal pressure and causes a decrease in the amplitude of glottal pulses. According to Stevens, this causes a drop-off in amplitude at low frequencies, specifically in the vicinity of the first harmonic. These observations, coupled with the fact that the human auditory system responds differently for low- and high-frequency sounds (those above or below approximately 3 kHz; Johnson 1980), it seems at least reasonably plausible that sonority should not be equated with general amplitude, but with low-frequency amplitude.

Although this hypothesis has yet to be tested experimentally, it is clear that this hypothesis, or some form of it, satisfactorily accounts for vocalic prominence. Indeed, it may be the case that vocalic prominence and consonantal prominence are not truly the same creature: in determining the phonetic basis for phonological sonority hierarchies, it is generally the ordering of obstruents that is problematic. For example, Wright (this volume) makes a compelling case for the role of perceptual robustness in accounting for consonantal sonority sequencing restrictions. In comparison with obstruents, the relative sonority of vowels and sonorant consonants is relatively straightforward. For example, it has been found for a number of languages that low vowels tend to be longer in duration than mid vowels, which in turn tend to be longer than high vowels (cf. Lehiste (1970) for an overview). Lehiste (1970) proposes that this low > mid > high duration pattern is in fact universal, and rooted in physiological factors. That is, vowels generally require a fairly open vocal tract, and one way of achieving this is through jaw depression. The lower the vowel, the more jaw depression is typically observed, and hence, a longer articulation time. Furthermore, the more open the vocal tract is, the more sound can escape, thus increasing intensity. Although this phonetic explanation does not extend to consonants, it is clear that vocalic prominence is straightforward: the longer and louder a vowel is, the more prominent it is. By virtue of the way in which vowels are articulated, both of the characteristics are generally correlated with jaw depression. This definition of vocalic prominence has the added benefit that it can easily be extended to new vowels. For example, Kenstowicz (1994) proposes extending the sonority hierarchy to include [ə] as the least sonorous of vowels to explain why, in certain languages, this vowel has the special property of repelling stress. Taking the common assumption that [ə] is a mid vowel, this extension is unmotivated. If sonority were correlated simply to tongue height, we would expect [ə] to have a medium sonority, similar to that of [e, o]. However, under the current hypothesis, this extension is expected. For example, Gruzov (1960) explores vowel duration in Mari, a language where [ə] does have the stress-repelling property investigated by Kenstowicz (1994). His data show that this vowel is indeed significantly shorter than the other vowels not showing this behaviour, a characteristic that extends even to cases of stressed [ə], where lengthening under accent is minimal. Similarly, jaw depression measurements for [ə] in Bulgarian indicate that this vowel is very close, showing a jaw depression about the same as for Bulgarian [i, u].[2]

With this interpretation of vocalic prominence in mind, prominence-stacking effects such as low-sonority vowels repelling stress (Kenstowicz 1994), or low-sonority vowels lowering under stress (Crosswhite 1998), can be seen as a grounded effect, in the sense of Archangeli and Pulleyblank (1994). That is, the only characteristic that seems to be held in common between vocalic prominence (sonority) and positional prominence (such as stress or syllabicity) is the presence of either increased intensity, increased duration, or both. By making prominent positions co-occur with sonorous segmental material, entities with the same or similar phonetic characteristics, in this case, duration and intensity, are made to co-occur within a single segmental locus.

This goal can be realised in a number of ways. The specific case of prominence-stacking investigated by Prince and Smolensky was syllabification – in their examples, syllable nuclei were chosen based on their relative sonority, with the most sonorous segments preferentially parsed as syllabic. For example, a low-sonority vowel like [i] would be parsed as an onset or coda if doing otherwise would require a higher sonority vowel like for example, [o] or [a] to be non-syllabic. In other words, the underlying sonority of the segments involved remains static, and prosodic prominence (i.e. syllabicity) is manipulated to provide a good match. However, the opposite is also possible: the sonority of the underlying segments can be increased in order to make them into better syllable nuclei. For example, in the (now-extinct) Native American language Gabrielino (Munro p.c.), short high vowels never surface as syllable nuclei. When such a form would otherwise be expected (i.e. when underlying /iː/ shortens in unaccented position), the vowel lowers to mid. A similar phenomenon occurred historically in most Slavic dialects (the so-called 'fall of the jers', cf., e.g., Vaillant 1958), in which the short high vowels /i, u/ were either deleted, or lowered to a more sonorous vowel quality (usually to [e, o], although the exact vowel qualities vary from dialect to dialect). In both of these cases, the underlying sonority of vowels is manipulated in order to provide a good match for prosodic prominence. It is precisely this sort of case – the reduction of underlying sonority in unstressed positions – that is discussed in more detail below.

3.3 Applying prominence alignment/reduction to vowel reduction

Prominence-reducing vowel reduction results from the application of prominence reduction constraints to vowels in unstressed positions. For example, the following constraint hierarchy is found in several languages with prominence-reducing vowel reduction.

(17) *UNSTRESSED/a ≫ *UNSTRESSED/ɛ, ɔ ≫ *UNSTRESSED/e, o ≫ *UNSTRESSED/i, u ≫ *UNSTRESSED/ə

This constraint hierarchy is calculated by crossing the following two phonetic scales:

(18) Scale 1: Accentual prominence
stressed $_{prom}$> unstressed
Scale 2: Vocalic prominence
a $_{prom}$> ɛ, ɔ $_{prom}$> e, o $_{prom}$> i, u $_{prom}$> ə

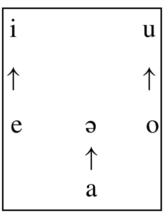
As before, the prominence reduction constraint family that would be based on these scales is produced beginning with the non-prominent member of Scale 1 (unstressed position). This systematically crosses it with the members of Scale 2, starting with the most sonorous member [a]. The highest ranking prominence reduction constraint based on these two scales is therefore *UNSTRESSED/a ('low vowels are not found in unstressed position').

To demonstrate how prominence reduction applies to unstressed vowels in more detail, two examples of prominence-reduction vowel reduction are discussed below: Bulgarian and Sri Lankan Portuguese Creole.

3.4 Prominence reduction in Bulgarian

In Standard Bulgarian (as well as many other Bulgarian dialects), there is a six-vowel inventory: /i, u, e, o, ə, a/. Note that /ə/ is a phonemic vowel in Bulgarian, which can occur both stressed and unstressed. In most of the eastern dialects of Bulgarian, this six-vowel inventory reduces to the three-vowel sub-inventory [i, u, ə] when in unstressed position. This is accomplished via step-wise raising of unstressed nonhigh vowels: unstressed /a/ raises to [ə], while unstressed /e/ and /o/ raise to [i] and [u], respectively. These neutralisations are illustrated below.[3]

(19) Bulgarian vowel neutralisations



Forms illustrating the Bulgarian vowel reduction pattern are provided in (20) (repeated here for the reader's ease from (13)).

(20) Bulgarian vowel reduction

	Vowels under stress		Same vowels unstressed	
a.	ˈr**o**guf	'of horn'	r**u**ˈgat	'horned'
b.	ˈ**o**nzi	'that' (masc.)	**u**ˈnazi	'that' (fem.)
c.	ˈs**e**lu	'village'	s**i**ˈla	'villages'
d.	ˈr**a**butə	'work'	r**ə**ˈbotnik	'worker'
e.	ˈgr**a**t	'city'	gr**ə**ˈdets	'town'

The stepwise nature of Bulgarian vowel reduction has made it difficult to treat as a unified phenomenon in classical generative phonology. In particular, the reduction process cannot be analysed as the elimination of nonhigh vowels, since the reduction of unstressed /a/ produces the nonhigh vowel [ə]. However, under the prominence reduction approach, this conundrum is easily solved: /e, o, a/ are defined as a group not according to distinctive features, but according to sonority. Under the same logic, the vowels /i, u, ə/ group together as a class of low-sonority vowels. Thus, Bulgarian vowel reduction can be formally modelled using the *UNSTRESSED/X constraint family, along with faithfulness constraints for the features [round], [front], and [high]. These constraints are ranked as follows in Bulgarian:

(21) Bulgarian constraint ranking

MAX[round]		MAX[−high]
MAX[+front]	≫	
*UNSTRESSED/a »		*UNSTRESSSED/i, u
*UNSTRESSED/e, o		»*UNSTRESSED/ə

The Faithfulness constraints MAX[round] and MAX[+front] are undominated in this grammar, indicating that underlying colour specifications are always maintained. The constraint *UNSTRESSED/a is also undominated, indicating that unstressed [a] will never occur in Bulgarian. *UNSTRESSED/e, o is dominated only by *UNSTRESSED/a, so unstressed [e] and [o] will likewise never occur. Finally, both *UNSTRESSED/a and *UNSTRESSED/e, o dominate the faithfulness constraint MAX[−high]. This domination allows the underlying [−high] specifications of /e, o/ to be cast off and replaced with [+high] in the neutralisations /e, o/ > [i, u]. It also allows the underlying [−high] specification of /a/ to be deleted without replacement in the neutralisation /a/ > [ə] (as noted earlier, schwa is represented as a targetless vowel, lacking any feature specifications). This is illustrated in (22), demonstrating reduction of unstressed /e, o, a/. Note that we can assume that all DEP[F] constraints are crucially dominated in Bulgarian, and will therefore not be included in the tableaux for sake of brevity.

(22) Reduction of /e, o, a/ in Bulgarian

/ro'gat/ 'horned'	MAX [round]	MAX [front]	*UNSTR-a	*UNSTR-e, o	*UNSTR-i, u	*UNSTR ə	MAX [−high]
☞ [ru'gat]					*		*
[ro'gat]				*!			
[rə'gat]	*!					*	*
[ra'gat]	*!		*				

/se'la/ 'villages'	MAX [round]	MAX [front]	*UNSTR-a	*UNSTR-e, o	*UNSTR-i, u	*UNSTR ə	MAX [−high]
☞ [si'la]					*		*
[se'la]				*!			
[sə'la]		*!				*	*
[sa'la]		*!	*		*		*

/gra'dets/ 'town'	MAX [round]	MAX [front]	*UNSTR-a	*UNSTR-e,o	*UNSTR-i,u	*UNSTR ə	MAX [−high]
☞ [grə'dets]						*	*
[gru'dets]					*!		*
[gro'dets]				*!			
[gra'dets]			*!				

Furthermore, the undominated position of faithfulness constraints for [round] and [front] also prevent the reduction of unstressed /i, u/ to [ə]:

(23) Nonreduction of /i, u/

/ime'na/ 'names' (cf.'*ime* 'name')	MAX round	MAX +front	*UNSTR-a	*UNSTR-e, o	*UNSTR-i, u	*UNSTR ə	MAX −high
☞ [imi'na]					*		
[emi'na]				*!			
[əmi'na]		*!				*	

/buk'var/ 'primer' (cf. '*bukvə* 'letter')	MAX round	MAX +front	*UNSTR-a	*UNSTR-e,o	*UNSTR-i,u	*UNSTR ə	MAX –high
☞ [buk'var]					*		
[bok'var]				*!			
[bək'var]	*!					*	

This account of Bulgarian vowel reduction not only accounts for the stepwise character of the neutralisations, but it also corresponds well with instrumental investigation of Bulgarian vowel reduction by Pettersson and Wood (1987a, 1987b). Their experiments suggest that the vowels targeted by Bulgarian vowel reduction are dispreferred precisely because of their prominence, not because of their featural content. Pettersson and Wood started their investigation by first verifying the existence of acoustically neutralising vowel reduction using spectrographic evidence – the formant frequencies measured for unstressed /e, o, a/ were found to coincide with those measured for the vowels /i, u, ə/, respectively. Since there was no acoustic difference in vowel quality reflecting these underlying vowel contrasts, we can state that Bulgarian vowel reduction is acoustically neutralising. However, using X-ray evidence, they conclude that Bulgarian vowel reduction is not completely neutralising in terms of articulation. That is, they claim based on the X-ray evidence that unstressed /e/, /o/, and /a/ in Bulgarian maintain the same (non-contrastive) tongue postures that are characteristic of their (respective) pronunciation when stressed, but take on a high jaw position when unstressed – that is, a height similar to that seen for the vowels /i/, /u/, and /ə/. For example, they describe the tongue postures used for Bulgarian stressed /i/ as being more bunched and tense, while the tongue postures used for Bulgarian stressed /e/ are more flat and lax – a distinction that was preserved in unstressed syllables, despite the fact that they all had been acoustically neutralised to the quality [i]. On the other hand, they describe the vowels /e, o, a/ as having on average 4 mm more mandible depression when stressed than did the stressed vowels /i, u, ə/. In other words, the change from an underlying *e*-quality to a surface *i*-quality was not brought about by actively changing the configuration or posture of the tongue, but by passively raising it by decreasing the amount of jaw opening, thus creating a high front vowel. Similar changes in jaw position account for the changes /o/ > [u] and /a/ > [ə]. The important observation is that the *articulation* of underlying /e, o, a/ was changed only enough to produce the vowels with the correct *acoustic* quality. In this case, a partial articulatory neutralisation is sufficient to cause a complete acoustic merger. This shows that it is not necessarily the featural content of /e, o, a/ that is disfavoured in unstressed position, but their sonority. That is, a constraint like *UNSTRESSED/a does not make reference to the articulatory or featural qualities of [a], but to its prominence (i.e. duration and amplitude). Therefore, this constraint will be equally non-violated by any vowel that is acoustically [i], regardless of the articulations used to realise that quality. With this in mind, it may be the case that Bulgarian unstressed /e, o/ are, in fact, [−high] on the surface, since they do not adopt a high tongue posture. In this case, it would be necessary to reformulate the analysis for Bulgarian vowel reduction already provided so as to replace articulatory features like [high] with acoustic features like [high F2] or, as suggested by Flemming (1995, this volume), to include both articulatory and acoustic features. For example, the crucially dominated constraint in the analysis already provided may not be MAX[−high], but MAX[−high F2]; whereas faithfulness constraints for tongue body position remain undominated.

This type of prominence reduction analysis is easily extendable to similar phenomena in other languages. One such example, Sri Lankan Portuguese Creole, is discussed in the next section.

3.5 Prominence reduction in Sri Lankan Portuguese Creole

Another interesting case of vowel reduction via prominence reduction comes from Sri Lankan Portuguese Creole (Smith 1978). This language has seven phonemic vowels: /i, u, e, o, æ, a, ɑ/. In this language, all vowels under stress are long. When stress shifts to a different syllable, the vowel shortens. For example, the long stressed mid vowels in [ˈoːj] 'eye' and [ˈtʃeːru] 'fragrance' correspond to short mid vowels in [oˈjaː] 'to see' and [tʃeˈraː] 'to smell pleasant'. However, if the stressed vowel is low, it not only shortens, but also raises to the corresponding mid vowel, as shown in the data below. This presents a case where low and mid vowels are contrastive under stress, but are neutralised (in favour of the mid vowels) when not under stress.

(24) Prominence reduction in Sri Lankan Portuguese Creole

	Low vowels under stress		Same vowels unstressed	
unstressed ɒ > o	ˈ**ɒ**ːbrə	'profession'	**o**bˈreːru	'manual worker'
	ˈn**ɒ**ːmi	'name'	n**o**miˈnaː	'nominate'
unstressed œ> e	ˈp**æ**ːdərə	'stone'	p**e**driˈjaːdu	'ornamented with stones'
	ˈf**æ**ːru	'iron'	f**e**ˈreru	'blacksmith'
unstressed a > ə	ˈb**a**ːjlu	'dance'	b**ə**jlˈdoːr	'dancer'
	ˈb**a**ːrvə	'beard'	b**ə**rˈveːru	'barber'

This vowel reduction pattern is quite similar to the Bulgarian pattern in that high-sonority vowels are eliminated, and in that this is accomplished in a manner that respects underlying colour specifications.[4] The difference between Sri Lankan Portuguese Creole and Bulgarian is that in Bulgarian, the mid vowels /e, o/ are subject to this type of reduction, while in Sri Lankan Portuguese Creole they are immune. To account for the Sri Lankan Portuguese Creole pattern of vowel reduction, the same constraints seen in the Bulgarian analysis can be used, but the ranking must be slightly different. As indicated below, the constraint *UNSTRESSED/e, o is promoted above MAX[−high], although MAX[−high] remains dominated by *UNSTRESSED/a. Furthermore, it should be pointed out that a constraint like *UNSTRESSED/a does not refer only to a specific quality named in the constraint ([a]), but to any other vowel qualities that share its position in the sonority hierarchy.[5] Therefore, [æ] and [ɒ] also fall under the purview of this constraint.

(25) Sri Lankan Portuguese Creole Constraint Ranking

Max[round]
Max[+front] ≫ Max[−high] ≫ *Unstressed/e, o ≫
*Unstressed/a *Unstressed/i, u ≫
*Unstressed/ə

The ranking of *Unstressed/a above Max[−high] ensures that the low vowel /a/ can lose its underlying [−high] specification to emerge as the featureless vowel [ə]. However, the ranking of Max[−high] above *Unstressed/e, o ensures that unstressed /e, o/ will not reduce to [i, u]. (Further, the undominated position of Max[round] and Max[+front] ensures that none of the vowels /e, o, i, u/ will reduce to [ə].) These rankings are illustrated in the following tableaux:

(26) Tableaux for Sri Lankan Portuguese Creole

/nɒmi'na/ 'nominate'	Max [round]	Max [front]	*Unstr a	Max [−high]	*Unstr- e,o	*Unstr- i,u	*Unstr ə
☞ [nomi'naː]					*		
[numi'naː]				*!		*	
[nɒmi'naː]			*!				
[nəmi'naː]	*!			*			*

/bajl'dor/ 'dancer'	Max [round]	Max [front]	*Unstr.- a	Max [−high]	*Unstr.- e,o	*Unstr.- i,u	*Unstr ə
☞ [bəjl'doːr]				*			*
[bajl'doːr]			*!				

/o'ja/ 'to see'	Max [round]	Max [front]	*Unstr.- a	Max [−high]	*Unstr.- e,o	*Unstr.- i,u	*Unstr ə
☞ [o'jaː]					*		
[u'jaː]				*!		*	
[ə'jaː]	*!			*			*

3.6 Prominence reduction and undershoot

As described above, prominence-reducing vowel reduction is a phenomenon in which relatively long vowels are replaced in unstressed position with shorter vowels of a similar quality. As such, prominence-reducing vowel reduction is somewhat similar, at least superficially, to the phonetic phenomenon of vowel undershoot. Undershoot refers to a situation in which a given speech sound is articulated in a manner that does not fully instantiate the canonical realisation of that sound. This usually occurs in contexts where articulation time is short; although the appropriate types of gestures are made (tongue fronting, jaw depression, etc.), the magnitude of the gestures is not adequate to produce a canonical realisation of the given phone. When undershoot applies to vowels, the result is often referred to as 'vowel reduction', as in Lindblom 1963. Here, the term 'vowel reduction' does not refer to the phonemic neutralisation of unstressed vowels, but to gradient changes in vowel articulation that can strike any vowel, stressed or unstressed, under the appropriate conditions, such as under a fast speech tempo (Lindblom 1963; Koopmans-van Beinum 1980). Impressionistically, vowel undershoot has been described as vowel centralisation: an undershot vowel might naively be described as a more [ə]-like realisation of that vowel quality. However, this is not an accurate description of vowel undershoot. Lindblom (1963) demonstrates that vowel undershoot cannot be straightforwardly described as centralisation. Instead, he suggests that undershoot results from the displacement of a given vowel in the direction of the surrounding consonantal environment. In many cases, this sort of coarticulatory effect will, in fact, equate with vowel centralisation. In other cases, it equates to vowel fronting, backing, raising, and so on, dependent on the articulations used in the surrounding consonantal environment. This hypothesis was further investigated in Moon and Lindblom 1994, where changes in vowel duration were seen to affect not only overall vowel quality (shorter vowels = less canonical), but also rate of formant change going into and coming out of the vowel (shorter vowel = more rapid transitions). To sum up, a given vowel phoneme will have a canonical target quality, a quality that may only be instantiated in contexts where the vowel is rather long. In contexts where the vowel is shorter, the articulators must move more quickly to produce the same result in less time. In certain contexts, articulatory effort might not be increased by the amount needed to maintain the canonical target quality, producing a surface quality that is less canonical and coloured to some degree by coarticulation with surrounding segments.

As such, vowel undershoot and prominence-reducing vowel reduction are somewhat similar in terms of end result. Both phenomena result in vowels with shorter articulation times in contexts of low duration such as in unstressed syllables or under fast speech tempos. Some have claimed that prominence-reducing vowel neutralisations like those analysed here are in fact the phonological analogues of undershoot, and that the motivation for prominence-reducing vowel reduction is effort avoidance (Flemming, this volume; Kirchner, this volume). I take the contrary position that these two phenomena are in fact phonetic and phonological inverses of each other. That is, in vowel undershoot, decreased articulation time leads to a change in vowel quality, and can be traced to a pressure to avoid effortful articulations (i.e. ones in which articulator movement must be fast). In contrast, in prominence-reducing vowel reduction, a change in vowel quality leads to a decrease in articulation time. This phenomenon, as presented above, is not motivated by effort avoidance, but by the desire to unite phonological entities with similar phonetic characteristics (enhancement). Although this distinction may at first glance seem overly nice, it makes certain typological predictions that seem to be empirically supported.

The first such prediction concerns the nature of other, similar prominence-reduction phenomena. The enhancement motivation advanced here for prominence-reducing vowel reductions predicts that it is prominence, not articulatory effort, that should be avoided in non-prominent prosodic positions. For example, phonemically bimoraic vowels are longer than their monomoraic counterparts, but not necessarily more effortful. However, many languages shorten bimoraic vowels in unaccented positions (cf. Crosswhite 2000 for one such case). Since, *ceteris paribus*, the gestures used for the monomoraic and bimoraic versions of a given vowel quality are the same, it is hard to see how simply holding the vowel for a longer time makes it inherently more effortful. In fact, it would naively seem that the monomoraic vowel would be more effortful, since the same articulatory gestures must be compressed into a shorter time period, requiring faster articulator movement.

The second prediction concerns the types of phenomena that we would expect to see in prominent positions. Under the effort-avoidance approach, we do not necessarily expect to see any special phenomena affecting prominent positions. Instead, we would expect to see only faithful realisation of underlying qualities, even those underlying qualities that are the most effortful. Instead, we often see prominent positions targeted for augmentation phenomena, which is expected under the enhancement motivation for prominence alignment. For example, many languages will lengthen underlyingly monomoraic vowels in stressed position – even when this eliminates a phonological contrast, namely that of length. Similarly, some languages increase the vocalic prominence of stressed vowels through vowel lowering (two examples of this are discussed in the next section), despite the resulting loss of contrast.

In summary, then, although both prominence-reducing vowel reduction and vowel undershoot produce similar end results, they seem to have different motivations. In particular, the enhancement-based motivation for prominence reduction more easily extends to cases of prominence reduction that cannot be analysed as effort avoidance, as well as to cases of augmentation in prominent prosodic positions.

Having said this, it should be pointed out that effort-avoidance is not wholly unassociated with vowel reduction: I merely claim that effort avoidance is not the motivating power behind it. Effort avoidance does enter into the picture, however, when considering logically possible alternatives to phonemic vowel reduction. Robert Kirchner has suggested (personal communication) that a vowel of any quality can be made to have extremely low sonority by making it extremely short. If this is so, it remains to be explained why prominence-reduction constraints such as *UNSTRESSED/e do not reduce a targeted vowel's sonority simply by making it extremely short while maintaining its quality. Doing so would require either very rapid articulator movement in unstressed syllables or a tolerance for non-canonical, undershot vowel realisations. Both alternatives are possible and attested. Lack of phonemic reduction via use of rapid articulator movement can be seen in grammars where effort-avoidance constraints are low ranked, whereas lack of phonemic reduction via tolerance of undershoot can be seen in grammars where constraints on accuracy of articulation are low ranked. Phonemic vowel reduction of the type considered in this chapter will result when both types of phonetic implementation constraint have a relatively high ranking. For an example of the formal implementation of constraints on effort avoidance to account for gradient sound changes, see Kirchner (this volume).

3.7 Other examples of prominence reduction

In the examples of prominence-reducing vowel reduction discussed above, vowel faithfulness constraints acted in a fairly straightforward manner to determine which subset of vowels underwent reduction, as well as the types of neutralisations used. However, vowel faithfulness constraints can interact with the *UNSTRESSED/X constraint family in other, more complex ways, resulting in additional subtypes of the prominence reduction phenomenon.

Reduction to [ə]. For example, if the entire *UNSTRESSED/X constraint family outranks vowel faithfulness constraints (particularly those demanding preservation of colour features), all unstressed vowels will surface as the extremely low-sonority vowel [ə]. This form of reduction is common, for example, in English and in some non-standard varieties of Russian (Dedova 1988).

If, however, the *UNSTRESSED/X constraint family outranks *all but one* faithfulness constraint, the resulting pattern of reduction is one in which most unstressed vowels reduce to [ə], leaving one or two 'survivors' to surface without reduction. This type of pattern is observed, for example, in some dialects of Italian, in which *all vowels but /a/* undergo reduction to [ə] in pretonic position (Maiden 1995).

And finally, if all of the vowel faithfulness constraints outrank the entire *UNSTRESSED/x constraint family, no reduction will occur. But if just one of the constraints is demoted to the bottom of the hierarchy, a pattern will result in which most unstressed vowels remain unreduced, but one or two vowels are singled out for reduction to [ə]. This type of pattern is observed in the Sadzhava dialect of Ukrainian (Popova 1972), where /e, ɛ/ reduce to [ə] while all other vowels remain unreduced, as well as in the Pavlikianski and Shirokolushki dialects of Bulgarian, where only /i/ and only /i, e/ (respectively) undergo reduction to [ə] (Stojkov 1968). Another well-known example is the reduction of unstressed monomoraic /i, u/ to [ə] in several dialects of Slovene (Bezlaj 1939; Toporišič 1976; Lenček 1982).

Reduction via tensing. As just mentioned, if all vowel faithfulness constraints outrank the entire *UNSTRESSED/X constraint family, they cannot cause any vowel reduction. However, if faithfulness constraints on [ATR] are demoted to a position below *UNSTRESSED/ɛ, ɔ, a form of reduction will occur in which the unstressed mid lax vowels /ɛ, ɔ/ undergo a minimal form of raising, surfacing as the slightly less sonorous vowels [e, o]. This form of reduction is common, occurring for instance in dialects of Italian and Catalan.

Since this is the only form of prominence-reducing vowel reduction that completely respects underlying height and colour features, it is predicted to be quite common. In earlier work (Crosswhite 2001), I have computed the factorial typology of the constraint set proposed here. It predicts that reduction-via-tensing should be able to co-occur with many other patterns of vowel reduction. This type of pattern is in fact attested in languages like Trigrad Bulgarian (Stojkov 1963), where a reduction-via-lowering pattern (/o, ɔ/ > [a], a case of contrast-enhancing reduction, to be discussed shortly) is accompanied by reduction-via-tensing (/ɛ/ > [e]). A similar pattern is found in Majorcan Catalan (Recasens 1991).

It is interesting to note at this juncture that the opposite type of pattern, reduction-via-laxing, has also been reported to occur in two different languages: Standard Slovene and north-eastern dialects of Brazilian Portuguese. However, instrumental analysis of the Slovene case casts doubt on this interpretation of the facts. This type of reduction is discussed as reduction via feature deletion in section 2.5.

3.8 Prominence reduction in additional contexts

Finally, it should be recalled that prominence reduction constraints are based on the combination of two prominence scales – all of the prominence reduction constraints examined in this section deal with (1) an accentual prominence scale, and (2) a vocalic prominence scale (vowel sonority). However, the parameter *accented* vs *unaccented* is not the only type of prominence that can interact with vowel quality in the fashion under consideration. For example, in addition to accentual prominence, there is also weight-based prominence. And indeed, there are cases of vocalic neutralisations that appear to be motivated by this sort of weight-based prominence reduction constraint. One such case, Carniolan Slovene, will be discussed now in order to show how the prominence reduction mechanism can be extended to this type of case.

Many dialects of Slovene exhibit vowel reductions. Although the exact nature of these reductions vary from dialect to dialect, one factor that is shared by most of the Slovene reduction patterns is the conditioning environment: accented bimoraic vowels are immune to reduction, while both unaccented vowels and accented monomoraic vowels are subject to neutralisations. Additionally, surface bimoraic vowels are only allowed in accented syllables in Slovene; if accent moves off an underlyingly long vowel, it concomitantly shortens: cf. [ˈplɛːma] ‘tribe’ vs [plɛˈmeːna] ‘tribes’. Thus it is appropriate to characterise Slovene vowel reduction as targeting monomoraic vowels, whether accented or not. For example, in many Carniolan dialects of Slovene (upon which the literary language is based), monomoraic /i, u/ reduce to [ə]. As mentioned in section 3.7 above, reduction of /i, u/ to [ə] while all other vowel qualities remain unreduced is a special case of reduction to [ə] in which most, but not all, vowel faithfulness constraints outrank the entire family of relevant prominence reduction constraints. In this case, the relevant prominence reduction family is not *Unstressed/X, but *Monomoraic/X. That is, just as *Unstressed/X family requires that non-prominent unstressed positions be filled with low-sonority vowels, the constraint family *Monomoraic/X requires that all monomoraic vowels be low in sonority (while the more prominent bimoraic vowels are left unaffected).[6] The *Monomoraic/X family is illustrated below:

(27) *Monomoraic/a ≫ *Monomoraic/e, o ≫ *Monomoraic/i, u ≫ *Monomoraic/ə

For example, the constraint *Monomoraic/a says that the segment [a] is too sonorous to be only monomoraic: to escape violation, all underlying /a/ must either lengthen or undergo a reduction in sonority. That is, if left unbridled, this constraint family would reduce all monomoraic vowels to [ə]. However, in Slovene, this constraint family is not left unbridled. High-ranking vowel faithfulness constraints block reduction to [ə] for most of the underlying vowel qualities of Slovene:

(28) Constraint ranking for Slovene

		*Monomoraic/a ≫		Max[+high],
Max[+low],	≫	*Monomoraic/e, o ≫	≫	Max[round],
Max[−high]		*Monomoraic/i, u ≫		Max[+front]
		*Monomoraic/ə		

By ranking MAX[+low] and MAX[−high] above all the monomoraic vowel reduction constraints, the nonhigh vowels are completely immune to reduction: any decrease in sonority of a monomoraic mid or low vowel would require violating one of these two undominated constraints. However, the high vowels are not protected in this manner. The ranking of *MONOMORAIC/ə above faithfulness constraints for colour and [+high] leaves the high vowels open to reduction. This analysis correctly predicts that bimoraic /i, u/ are not subject to this form of reduction.

3.9 Comparison of reduction types

I have proposed that vowel reduction is not a single phenomenon, but two independent, formally distinct phenomena: contrast enhancement and prominence reduction. These two reduction phenomena are sometimes indistinguishable. For example, under both types of reduction, unstressed [e, o] are disfavoured: either because they are noncorner (contrast enhancement), or because they are moderately sonorous (prominence reduction). Furthermore, given the Optimality-Theoretic formulation of these two phenomena, it is entirely possible for either of these categories to give rise to non-canonical vowel reduction patterns that only partially instantiate the ideas of prominence reduction or contrast enhancement. For example, high-ranking for a faithfulness constraint might block a certain vowel from undergoing reduction (e.g. high rank for MAX[+low] could block reduction of unstressed /a/). With this in mind, any case where mid vowels are neutralised via raising to [i, u] is ambiguous: Are they being raised to produce corner vowel qualities, or to produce low-sonority vowel qualities? In some cases, it is possible to distinguish the two. For example, if unstressed /a/ also reduces, for example to [ə], it is a case of prominence reduction: the reduction /a/ > [ə] decreases sonority, but does not produce a corner vowel. Similarly, if any vowel reduces to [a], it is a case of contrast enhancement. For example, if /e/ > [i] but /o/ > [a] (as in Russian), we have a case of contrast enhancement, since one of these reductions, /o/ > [a], produces a corner vowel but does not reduce sonority. However, in cases where both /e, o/ reduce to [i, u], and unstressed /a/ remains unreduced, no definitive categorisation can be made: it could be that /a/ is immune because LIC-NONCORNER/STRESS does not affect unstressed /a/, or it could be the case that high rank of MAX[+low] blocks reduction of /a/.

4 Two-pattern vowel-reduction systems

Finally, it appears that more than one type of vowel reduction can occur in the same language. In this type of vowel-reduction language, there are two sets of neutralisations that occur in unstressed syllables. One, typically a moderate form of reduction, takes place in certain unstressed syllables, while a more extreme form of reduction takes place in the remaining syllables. This is the case, for example, in most dialects of Russian. Here, unstressed /o/ and /a/ neutralise. In the syllable immediately preceding the stress, this generates the surface vowel quality [a], while in other unstressed syllables it generates [ə]. This is demonstrated below with data from Standard Russian:

(29) Russian two-pattern vowel reduction

In stressed σ	In immediately pre-stress σ	In other unstressed σ	gloss
ˈd**o**m (nom. sg.)	d**a**ˈma (nom. pl.)	d**ə**maˈvoj (adj.)	‘house’
ˈg**o**ləvu (acc.)	g**a**ˈlofka (diminutive)	g**ə**laˈva (nom. sg.)	‘head’
ˈk**a**m^{j}in^j (nom. sg.)	k**a**m^{j}ˈn^{j}ej (gen. pl.)	k**ə**m^{j}iˈn^{j}istəj (adj.)	‘stone’
ˈd**a**l^{j}iji (comp.)	d**a**ˈl^{j}ok^{j}ij (adj.)	d**ə**l^{j}iˈko (adverb)	‘far’

A similar pattern is seen in some southern Russian dialects, where unstressed /e/ and /o/ both neutralise to [a] in the syllable immediately preceding the stress, but reduce either to [i] (for underlying /e/ and for underlying /o, a/ preceded by a palatalised consonant) or to [ə] (underlying /o, a/ elsewhere). Other such patterns are found in Rhodope Bulgarian dialects (Miletich 1936), certain Italian dialects (Maiden 1995), and standard Brazilian Portuguese (Dukes 1993; Redenbarger 1981), to name a few. These patterns are briefly described in the table in (30).

(30) Examples of two-pattern vowel reduction

Language	Moderate reduction	Extreme reduction
Southern Russian	Unstressed /e, o/ both neutralise to [a] in the syllable immediately preceding the stress.	In all remaining unstressed syllables, /o, a/ reduce to [ə] (or [i] following a palatalised consonant) and unstressed /e/ reduces to [i].
Contemporary Standard Russian	Unstressed /o/ neutralises to [a] in the syllable immediately preceding the stress.	Unstressed /o/ and /a/ neutralise to [ə] in the remaining unstressed syllables.

Rhodope Bulgarian	Unstressed /e/ and /o/ both neutralise to [a] in any syllable preceding the stress.	Unstressed /e/ and /o/ neutralise to [i] and [u], respectively, in post-tonic positions. In the same contexts, unstressed /a/ neutralises to [ə].
Dialectal Italian	All unstressed vowels *except /a/* neutralise to [ə] in any syllable preceding the stress.	*All* unstressed vowels neutralise to [ə] post-tonically.
Standard Brazilian Portuguese	Unstressed /ɛ, ɔ/ neutralise to [e, o], respectively, in unstressed non-word-final and non-word-initial syllables.	Unstressed /ɛ, e/ and /ɔ, o/ neutralise to [i] and [u], respectively in word-final and word-initial unstressed syllables. Also, unstressed /a/ becomes [ə].

The extreme vowel-reduction processes share some common features. First, the type of neutralisation seen in extreme reduction is always sonority decreasing – in contrast with the moderate reductions, which can be sonority increasing (cf. the Russian and Bulgarian change of unstressed /o/ to [a]). Second, they seem to target those unstressed syllables that are the most durationally impoverished. This fact makes it possible to identify extreme vowel reduction as a case of neutralisation caused by prominence reduction targeting durationally impoverished unstressed syllables, while the changes seen in moderate reduction can be ascribed to a second reduction phenomenon that targets unstressed syllables in general (usually contrast-enhancing reduction, but cf. the dialectal Italian pattern). An interesting parallel that can be drawn at this point is that contrast-enhancing reduction can occur in either stress-timed or syllable-timed languages, but prominence reduction appears to occur only in stress-timed languages (or dialects). Based on evidence from various languages with vocalic prominence reduction phenomena, I conclude that extreme reduction (a form of prominence reduction found in two-pattern systems) occurs in durationally impoverished syllables. In Crosswhite (2001), I argue that such durationally impoverished syllables are in fact nonmoraic.

As noted above, the contexts for extreme vowel reduction comprise the most durationally impoverished syllables found in a given language. Put another way, the moderate vowel reductions are only found in those unstressed syllables that have slightly greater duration than the other unstressed syllables. Two examples – Standard Russian and Brazilian Portuguese – are discussed in the paragraphs below.

The immediately pretonic syllable in Russian has long been recognised as having a special durational status – it is much longer than other unstressed syllables, and can sometimes even be longer than the stressed syllable (this is often the case, for example, with words where the stressed vowel has a low inherent duration ([i, u]), and the immediately pretonic vowel has a high inherent duration ([a]).) This difference in the duration of Russian unstressed syllables is significant enough to have been accurately noted by ear and described by nineteenth-century Russian grammarians. This phenomenon is also easily observed indirectly when listening to Russian speech: at conversational speech tempos, non-immediately pretonic unstressed vowels are often completely or near-completely elided, but immediately pretonic ones are immune to this process. For example, the word /xoroˈʃo/ 'good' has the citation pronunciation [xəraˈʃo], however it is often pronounced [xraˈʃo], but never *[xərˈʃo].

The special durational status of Russian immediately pretonic syllables has also been confirmed experimentally (see, e.g., Zlatoustova 1981). Investigators have also noted a link between this type of durational effect and the presence vs absence of varying degrees of vowel reduction in Russian dialects (Vysotskii 1973; Al'mukhamedova and Kul'sharipova 1980; Kasatkina et al. 1996; et al.). For example, in the Vladimir-Volga Basin dialect group, the immediately pretonic syllable has the same special durational status seen in Contemporary Standard Russian, and in this group the immediately pretonic syllable also has a special status with respect to vowel reduction: all other unstressed syllables are subject to vowel reduction, but the immediately pretonic one is immune (i.e. unstressed /o/ remains [o]). In another case, the immediately pretonic syllable in certain southern Russian dialects predictably displays either moderate reduction or extreme reduction. As shown by Kasatkina et al. (1996), this predictable variation in vowel-reduction pattern is accompanied by changes in prosody – when the immediately pretonic syllable displays moderate reduction, this syllable has relatively longer duration; when it displays extreme reduction, it has a duration about equivalent to that of the other unstressed syllables.

A similar situation is seen in standard Brazilian Portuguese. For example, in instrumental work by Major (1992), it was found that post-tonic syllables in Brazilian Portuguese undergo greater shortening than do pretonic syllables. Based on this and other evidence, Major hypothesises that the post-tonic syllables in Brazilian Portuguese are stress timed, while the pretonic syllables are syllable timed. I will assume here that this is equivalent to saying that post-tonic syllables in Brazilian Portuguese can be nonmoraic, while pretonic ones cannot. Major also suggests that this dichotomy in the durational properties of pre- and post-tonic unstressed syllables is related to the two different vowel reduction patterns seen in this language, although in a way that is slightly different from the relationship hypothesised here. Specifically, Major hypothesises that the sonority-decreasing vowel neutralisations seen post-tonically (ɛ, e > i; ɔ, o > u; a > ə) exist in order to heighten the effect of the post-tonic shortening. Since Major also observed some shortening pretonically (i.e., in faster speech tempos), he suggests that Brazilian Portuguese is in the process of converting from syllable timing (which is typical of several Romance languages, including Spanish and Italian) to stress timing. Judging from the comments made by Brakel (1985) and de Carvalho (1988–92), this process is already at a more advanced stage in the European (Iberian) variant of Portuguese. Both researchers note that one of the differences between Brazilian and European Portuguese is the absence in the European variant of the moderate vowel reductions seen in the Brazilian variant. That is, in European Portuguese, all unstressed syllables are subject to extreme reduction.[7] They further note that European Portuguese differs from Brazilian Portuguese in the type of manipulation that an unstressed vowel endures – in the European variant, unstressed vowels (including pretonic vowels) are subject to extreme shortening, which often results in the devoicing and/or complete deletion of the vowel – a type of pronunciation that is less typical for Brazilian Portuguese.

My general approach to accounting for these observations is that stress-timed languages – with their ultra-short vowels that are prone to devoicing or deletion – are languages that allow some subset of their vowels to surface without associated moras. The exact distribution of these nonmoraic vowels varies from language to language – in Brazilian Portuguese they are post-tonic; in Russian they cannot occur immediately before the stressed syllable; and in Iberian Portuguese they seem to occur in most unstressed syllables.

5 Conclusion

In this chapter, I have considered numerous cases of vowel reduction, all of which fall into one of two broad groups: reduction based on prominence, and reduction based on contrast enhancement. The prominence-based category is motivated by the desire to unite different elements with similar prominence characteristics, and the contrast-enhancing category is motivated by the desire to avoid perceptually challenging vowel qualities in all but the most perceptually favourable positions. These two desiderata are encoded in an Optimality-Theoretic analysis using, respectively, prominence-reduction constraints and licensing constraints.

This dichotomy is especially useful in resolving ‘reduction paradoxes’; for example, the Bulgarian pattern where /i, e, ə, a, o, u/ reduces to [i, u, ə] versus the Belarusian pattern where /i, e, a, o, u/ reduce to [i, u, a]. If these two reduction patterns are superficially compared, we might arrive at the anomalous conclusion that the vowel quality [a] is both marked (it undergoes reduction in Bulgarian) and unmarked (it serves as a reduction vowel in Belarusian). Another example of this sort of 'reduction paradox' is observed when comparing Bulgarian and Sri Lankan Portuguese Creole (/æ, a, ɒ/ > [e, ə, o]) – the mid vowels seem to be both marked (they undergo reduction in Bulgarian) and unmarked (they serve as reduction vowels in Sri Lankan Portuguese Creole). The analysis for vowel reduction advanced here resolves these paradoxes by positing separate families of markedness constraints, based on distinct phonetic motivations. Specifically, the basic claim of this analysis is that vowels are not 'absolutely' marked, and, therefore, vowel-reduction patterns cannot be used as indicators of absolute vocalic markedness.

This appeal to two independent motivations for vowel reduction suggests that, in fact, there is no such thing as a monolithic concept of markedness, at least as far as unstressed vowels are concerned. In effect, the phonological concept of markedness has been replaced by phonetic considerations, which are encoded in phonology using phonetically motivated constraints.

References

Al'mukhamedova, Z. M. and P. È. Kul'sharipova (1980). *Reduktsiia glasnikh i prosodiia slova v okaiushchikh russkikh govorakh* (Vowel reduction and word prosody in Okajuschij Russian dialects). Kazan': Izdatel'stvo Kazan'skogo Universiteta.

Archangeli, Diana and Douglas Pulleyblank (1994). *Grounded phonology*. Cambridge, MA: MIT Press.

Baroni, Marco (1996). An acoustic study of Italian unstressed mid-vowels. Paper presented at Acoustical Society of America, 4 December 1996, Hawaii. http://home.sslmit.unibo.it/~baroni/research.html.

Bezlaj, František (1939). *Oris slovenskega knjižnega izgovora* (A sketch of Slovene literary pronunciation). Ljubljana: Natisnila učiteljska tiskarna.

Bidwell, Charles E. (1969). *Outline of Slovenian Morphology*. Pittsburgh: University Center for International Studies, University of Pittsburgh.

Bloomfield, Leonard (1957). *Eastern Ojibwa: Grammatical sketch, texts, and word list*. Ann Arbor: University of Michigan Press.

Brakel, Arthur (1985). Reflections on the analysis of exceptions to the rule of Iberian Portuguese vowel reduction. *Hispanic Linguistics* 2. 1: 63–85.

Browman, Catherine P. and Louis Goldstein (1992). 'Targetless' schwa: An articulatory analysis. In Gerard J. Docherty and D. Robert Ladd (eds.), *Papers in Laboratory Phonology, II: Gesture, Segment, Prosody*. Cambridge: Cambridge University Press, 26–67.

Casali, Roderic F. (1998). Predicting ATR activity. In M. Catherine Gruber, Derrick Higgins, Kenneth S. Olson, and Tamra Wysocki (eds.). *CLS 34 part 1: Papers from the main session, April 17–19*. Chicago: Chicago Linguistic Society, 56–68.

Crosswhite, Katherine (1998). Segmental vs. prosodic correspondence in Chamorro. *Phonology* 15. 3: 281–316.

(2000). Length neutralization and vowel lengthening in Orlec Chakavian. In Tracy Holloway King and Irina A. Sekerina (eds.), *Annual workshop on formal approaches to Slavic linguistics: the Philadelphia meeting, 1999*. Ann Arbor: Michigan Slavic Publications, 152–67.

(2001). *Vowel reduction in Optimality Theory*. New York: Routledge.

de Carvalho, Joaquim Brandao (1988–92). Reduction vocalique, quantité et accentuation: Pour une explication structurale de la divergence entre portugais lusitanien et portugais brèsilien. *Boletim de Filologia* 32: 5–26.

Dedova, O. V. (1988). Nabl'udeniia nad vokalizmom sovremennogo moskovskogo prostorechiia (Observations on one vocalism of contemporary muscovite *prostorechie*). *Vestnik Moskovskogo Universiteta*, Series 9 'Filologiia', no. 2: 28–36.

Diehl, Randy (1986). Remarks on Stevens' quantal theory of speech. *JPh* 17: 71–8.

Dukes, Beatriz (1993). Vowel reduction and underspecification in Brazilian Portuguese. *UCLA Occasional Papers in Linguistics* 13: *Papers in Phonology*: 21–48.

Flemming, Edward (1995). *Auditory representations in Phonology*. PhD dissertation, University of California, Los Angeles.

Fry, D. B. (1979). *The physics of speech*. Cambridge: Cambridge University Press.

Gruzov, L. P. (1960). *Sovremennyi mariiskii iazyk: fonetika* (Contemporary Mari: Phonetics). IOshkar-Ola: Mariiskoe knizhnoe izd-vo.

Hualde, Jose (1991). *Basque phonology*. London: Routledge.

Jiang-King, Ping (1996). *An Optimality account of tone-vowel interaction in Northern Min*. PhD dissertation, Chinese University of Hong Kong.

Johnson, D. H. (1980). The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *JASA* 68: 1115–22.

Kamprath, Christine (1991). *Toward a restrictive model of unstressed vowel reduction*. Paper presented at the annual meeting of the Linguistic Society of America.

Kasatkin, L. L. (1989). *Russkaia dialektologiia* (Russian dialectology). Moscow: Prosveshcheniie.

Kasatkina, R. F. et al. (1996). Osobennosti prosodii slova v iuzhnorusskikh govorakh (Characteristics of work prosody in South Russian dialects). In T. Nikolaeva (ed.), *Prosodicheskii stroi russkogo iazyka* (The prosodic structure of Russian). Moscow: RAN.

Keating, Patricia (1983). Comments on the jaw and syllable structure. *JPh* 11: 401–6.

Kenstowicz, M. (1994). *Sonority-driven stress*. Ms., MIT. Rutgers Optimality Archive 33, http://rccs.rutgers.edu/roa.html.

Koopmans-van Beinum, F. J. (1980). *Vowel contrast reduction, an acoustic and perceptual study of Dutch vowels in various speech conditions*. PhD dissertation, University of Amsterdam.

Krivitskii, A. A. and A. I. Podluzhnyi (1994). *Uchebnik belorusskogo iazyka dlia samoobrazovaniia* (Textbook of the Belarusian language for self-study). Minsk: Vyshèishaia Shkola.

Ladefoged, Peter, Joseph L. Declerk, and Richard Harshman (1977). Factor analyses of tongue shapes. *UCLA Working Papers in Phonetics* 19. Los Angeles, CA: Department of Linguistics, *UCLA*, 39–46.

Ladefoged, P., R. Harshman, L. Goldstein and L. Rice (1971). Vowel articulation and formant frequencies. *UCLA Working Papers in Phonetics* 38: 16–40.

Lehiste, Ilse (1961). The phonemes of Slovene. *International Journal of Slavic Linguistics and Poetics* 4: 48–66.
(1970). *Suprasegmentals*. Cambridge, MA: MIT.
Lenček, Rado (1982). *The structure and history of the Slovene language*. Columbus, OH: Slavica.
Lieberman, P. (1971). *Uniquely human: The evolution of speech, thought, and selfless behaviour*. Cambridge, MA: Harvard University Press.
Liljencrants, J. and B. Lindblom (1972). Numerical simulations of vowel quality systems: The role of perceptual contrast. *Lg* 48: 839–62.
Lindblom, B. (1963). Spectrographic study of vowel reduction. *JASA* 35: 1773–81.
(1983). Economy of speech gestures. In P. F. MacNeilage (ed.), *The production of speech*. New York: Springer Verlag. 217–45.
(1986). Phonetic universals in vowel systems. In John Ohala and Jeri Jaeger (eds.), *Experimental phonology*. Orlando: Academic Press, 13–44.
Lindblom, B. and I. Maddieson (1988). Phonetic universals in consonant systems. In L. M. Hyman and C. N. Li (eds.), *Language, speech and mind*. London: Routledge, 62–78.
Lutta, C. M. (1923). Der Dialekt von Bergün und seine Stellung innerhalb der raetoromanischen Mundarten Braubeundens. Supplement to *Zeitschrift für Romanische Philologie* 71.
Maddieson, Ian (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.
Maiden, Martin (1995). Evidence from the Italian dialects for the internal structure of prosodic domains. In John Charles Smith and Martin Maiden (eds.), *Linguistic theory and the Romance languages*. Amsterdam: Benjamins, 115–31.
Major, Roy C. (1992). Stress and rhythm in Brazilian Portuguese. In Dale A. Koike and Donald P. Macedo (eds.), *Romance linguistics: The Portuguese context*. Westport: Bergin and Garvey, 3–30.
Majors, Tivoli (1998). *Stress-dependent harmony: Phonetic origins and Phonological analysis*, PhD dissertation, University of Texas, Austin.
Malsch, Derry and Roseanne Fulcher (1989). Categorizing phonological segments: The inadequacy of the sonority hierarchy. In Robert Corrigan, Fred Eckman and Michael Noonan (eds.), *Linguistic categorization*. Oxford: Clarendon, 69–80.
Miletich, L. (1936). 'Akane' i 'polnoglasie' vŭ tsentralniia rodopski govorŭ ('Akane' and 'polnoglasie' in the central Rhodose dialects). *Makedonski pregledŭ* 10. 26–30.
Miller, Patricia (1972). Vowel neutralization and vowel reduction. *CLS* 8: 482–9.
Moon, Seung-jae and Bjorn Lindblom (1994). Interaction between duration, context, and speaking style in English stressed vowels. *JASA* 96: 40–55.
Munro, Pamela and Peter John Benson (1973). Reduplication and rule ordering in Luiseño. *IJAL* 39: 15–21.
Nathan, Geoffrey (1989). Preliminaries to a theory of substance: The substance of sonority. In Robert Corrigan, Fred Eckman, and Michael Noonan (eds.), *Linguistic categorization*. Oxford: Clarendon, 55–67.
Palková, Zdena (1994). *Fonetika a fonologie češtiny* (Phonetics and phonology of Czech). Prague: Universita Karlova.
Perkell, J. S. and M. H. Cohen (1986). An indirect test of the quantal nature of speech in the production of the vowels /i/, /a/, and /u/. *JPh* 17: 123–33.

Perkell, J. S. and W. L. Nelson (1982). Articulatory targets and speech motor control: A study of vowel production. In S. Grillner, A. Persson, B. Lindblom and J. Lubker (eds.), *Speech motor control*. Oxford: Pergamon, 187–204.

(1985). Variability in production of the vowels /i/ and /a/. *JASA* 77. 1889–95.

Perrone, Charles A. and Linda Ledford-Miller (1985). Variation in pretonic /e/ in Brazilian Portuguese: Preliminary studies with popular music of the Northeast and Rio de Janeiro. *Hispania* 68. 1: 154–9.

Peterson, Gordon, and Ilse Lehiste (1960). Duration of syllable nuclei in English. *JASA* 32: 693–703.

Pettersson, Thore and Sidney Wood (1987a). Vowel reduction in Bulgarian and its implications for theories of vowel production: A review of the problem. *Folia Linguistica* 21: 261–79.

(1987b). Vowel reduction in Bulgarian: the phonetic data and model experiments. *Folia Linguistica* 22: 239–62.

Pisoni, D. (1980). Variability of vowel formant frequencies and the quantal theory of speech: A first report. *Phonetica* 37: 285–305.

Popova, T. V. (1972). Paradigmaticheskie konsonantnye riady cheredovanii v iugozapadnykh ukrainskikh dialektakh (na materiale govora s. Sadzhava) (A paradigmatic series of alternations in Southwest Ukrainian dialectal consonantism (based on material from the dialect of the village of Sadzhava)). In G. P. Klepikova (ed.), *Karpartskaia dialektologiia i onomastika* (Carpathian dialectology and onomastics). Moscow: Nauka, 179–239.

Prince, Alan and Paul Smolensky (1993). *Optimality Theory: Constraint interaction in generative grammar*. Rutgers Optimality Archive 537, http://roa.rutgers.edu/.

Recasens, Daniel (1991). *Fonética descriptiva del Catalá*. Barcelona: Institut d'Estudis Catalans.

Redenbarger, Wayne J. (1981). *Articulator features and Portuguese vowel height*. 'Harvard Studies in Romance Languages, 37' Cambridge, MA: Department of Romance Languages and Literatures of Harvard University.

Scatton, E. A. (1984). *A reference grammar of modern Bulgarian*. Columbus, OH: Slavica.

Schwartz, Jean-Luc, Louis-Jean Boë, Nathalie Vallée, and Christian Abry (1997). The dispersion-focalization theory of vowel systems. *JPhon*. 25: 255–86.

Smith, Ian R. (1978). *Sri Lanka Creole Portuguese phonology*. (Dravidian Linguistics Association). Vanchiyoor: Kerala University Cooperative Stores Press.

Steriade, Donca (1994). Licensing by cue. Ms., University of California, Los Angeles.

(1994b). Positional neutralization and the expression of contrast. Ms., University of California, Los Angeles.

Stevens, Kenneth (1989). On the quantal nature of speech. *JPh* 17: 3–45.

Stojkov, Stojko (1963). Akane v govora na s. Trigrad, Devinsko (*Akane* in the dialect of the village of Trigrad, Devinsko). *Bŭlgarski Ezik* 13: 8–21.

(1968). *Bŭlgarska dialektologiia* (Bulgarian dialectology). Sofia: Nauka i isskustvo.

Toporišič, Jože (1976). *Slovenska slovnica* (Slovene grammar). Maribor: Založba Obzorja.

Urbanczyk, Stanislaw (1953). *Zarys dialektologii polskiej* (A sketch of Polish dialectology). Warsaw: PWN.

Vaillant, Andre (1958). *Grammaire comparée des langues slaves*, Vol. 1 *Phonétique*. Lyon: IAC.

Vysotskii, S. S. (1973). O zvukovoi strukture slova v russkikh govorakh (on word sound structure in Russian dialects). In S. V. Bromlei, Iu. S. Azarkh, and L. N. Bulatova (eds.), *Issledovaniia po russkoi dialektologii* (Research on Russian dialectology). Moscow: Nauka, 17–41.

Zlatoustova, L. V. (1981). *Foneticheskie edinitsy russkoi rechi* (The phonetic units of Russian speech). Moscow: Izdatel'stvo Moskovskogo universiteta.

Zoll, Cheryl (1996). *Parsing below the segment in a constraint-based framework*. PhD dissertation, University of California, Berkeley.

Notes

* I would like to thank the following individuals for valuable comments on this chapter, or on work leading to it: Tim Beasley, Bruce Hayes, Patricia Keating, Robert Kirchner, Joyce McDonough, and Donca Steriade.

1. But see Casali 1998 for an argument that the acoustically grounded approach to the relationship between [ATR] and [low] offers a superior account of certain vowel harmony facts.
2. It is also worth noting here that this approach does not necessarily predict that non-reduced mid central vowels, such as English [ʌ], should have the same position in the sonority hierarchy as [ə]. Hence, Peterson and Lehiste's (1960) finding that English [ə] is in fact slightly longer than [ɪ, ʊ, ɛ] is not contrary to the position outlined above, since they classify the vowels in *tuck* and *tug* as [ə].
3. Note that in the pronunciation norm of Sofia and other western areas of Bulgaria, vowel reduction is weaker than in the eastern areas, or even entirely absent. Since Sofia pronunciation defines the standard, many vowel-reducing speakers attempt to suppress vowel neutralisations when speaking in formal registers. In particular, suppression of the reduction /e/>[i] is quite common, and lack of this suppression is rather stigmatised – cf. Scatton 1984.
4. Again, it may be the case that the underlyingly low vowels in unstressed position maintain some articulatory differences, when compared with phonemically mid vowels. Lacking any concrete data on this point, I will continue to use traditional articulatory features in this and subsequent analyses.
5. This being the case, it might be more appropriate to call the constraint *Unstressed/[low]. However, this would imply the parallel *Unstressed/[mid]. Since [ə] is often (though not uncontroversially) classed as a mid vowel, this nomenclature would cause confusion – as laid out in the preceding sections, [ə] is an extremely low sonority vowel while 'full' mid vowels like [e, o] are of medium sonority.
6. It is important to bear in mind that the monomoraic category is in fact intermediate in prominence: it is less prominent than bimoraicity, but more prominent than non-moraicity: $\mu\ \mu_{prom} > \mu_{prom} > \emptyset$. For this reason, the monomoraic category can be the basis for both a prominence reduction constraint family as in the main text (which I designate *Monomoraic/X), as well as a prominence alignment family (*μ/X). The

former places limitations on which vowels are too sonorous to be monomoraic (as opposed to bimoraic), while the latter places limitations on which segments (usually consonants) are not sonorous enough to be moraic. For example, *MONOMORAIC/a says that [a] is too sonorous to be *only monomoraic*, while *μ/t says that [t] is not sonorous enough to contribute phonological weight. These two distinct families differ in their inherent rankings: *MONOMORAIC/a ≫ *MONOMORAIC/t but *μ/t ≫ *μ/a.

7. Other differences between Brazilian and European Portuguese vowel reduction include: (1) in European Portuguese, vowel reduction is obligatory even in slow speech; in Brazilian Portuguese the pretonic 'moderate' reduction is optional in careful (citation) speech; (2) in European Portuguese, unstressed /e/ reduces to [ə], whereas in Brazilian Portuguese it reduces to [i] (post-tonically).

8 Contrast and perceptual distinctiveness

Edward Flemming

1 Introduction

Most ‘phonetically driven’ or functionalist theories of phonology propose that two of the fundamental forces shaping phonology are the need to minimise effort on the part of the speaker and the need to minimise the likelihood of confusion on the part of the listener. The goal of this chapter is to explore the perceptual side of this story, investigating the general character of the constraints imposed on phonology by the need to minimise confusion.

The need to avoid confusion is hypothesised to derive from the communicative function of language. Successful communication depends on listeners being able to recover what a speaker is saying. Therefore it is important to avoid perceptually confusable realisations of distinct categories; in particular, distinct words should not be perceptually confusable. The phonology of a language regulates the differences that can minimally distinguish words, so one of the desiderata for a phonology is that it should not allow these minimal differences, or contrasts, to be too subtle perceptually. In Optimality-Theoretic terms, this means that there are constraints favouring less confusable contrasts over more confusable contrasts.

There is nothing new about the broad outlines of this theory (cf. Lindblom 1986, 1990; Martinet 1955; Zipf 1949; among others), but it has important implications for the nature of phonology. First, it gives a central role to the auditory-perceptual properties of speech sounds, since distinctiveness of contrasts is dependent on perceptual representation of speech sounds. This runs counter to the articulatory bias in phonological feature theory observed in Chomsky and Halle 1968 and its successors. Substantial evidence for the importance of perceptual considerations in phonology has already been accumulated (e.g. Boersma 1998; Flemming 1995; Jun, this volume; Steriade 1995, 1997; Wright, this volume; see also Hume and Johnson 2001: 1–2 and references cited there). This chapter provides further evidence for this position, but the focus is on a second implication of the theory: the existence of constraints on contrasts. Constraints favouring distinct contrasts are constraints on the differences between forms rather than on the individual forms themselves. We will see that paradigmatic constraints of this kind have considerable implications for the architecture of phonology.

The next section discusses why we should expect perceptual markedness to be a property of contrasts rather than individual sounds and previews evidence that this is in fact the case. Then constraints on contrast will be formalised within the context of a theory of phonological contrast. The remainder of the chapter provides evidence for the key prediction of the theory: the markedness of a sound depends on the sounds that it contrasts with.

2 Perceptual markedness is a property of contrasts

The nature of the process of speech perception leads us to expect that any phonological constraints motivated by perceptual factors should be constraints on contrasts, such as the contrast between a back unrounded vowel and a back rounded vowel, not constraints on individual sounds, such as a back unrounded vowel. Speech perception involves segmenting a speech signal and categorising the segments into a predetermined set of categories such as phonetic segments and words. The cues for classification are necessarily cues that a stimulus belongs to one category as opposed to another. So we cannot talk about cues to a category, or how well a category is cued by a particular signal without knowing what the alternatives are. For example, it is not possible to say that a back unrounded vowel presents perceptual difficulties without knowing what it contrasts with. It is relatively difficult to distinguish a back unrounded vowel from a back rounded vowel, so if a language allows this contrast, the back unrounded vowel can be said to present perceptual difficulties, and the same can be said of the back rounded vowel. But if it is known that a back unrounded vowel is the only vowel that can appear in the relevant context, then all the listener needs to do is identify that a vowel is present as opposed to a consonant, which is likely to be unproblematic.

Perceptual difficulty is thus very different from articulatory difficulty. Articulatory difficulty can be regarded as a property of an individual sound in a particular context because it relates to the effort involved in producing that sound. There is no analogous notion of effort involved in perceiving a sound – perceptual difficulties do not arise because particular speech sounds tax the auditory system, the difficulty arises in correctly categorising sounds. Thus it does not seem to be possible to provide a sound basis in perceptual phonetics for constraints on the markedness of sounds independent of the contrasts that they enter into. This point is assumed in Liljencrants and Lindblom's models of how perceptual factors shape vowel inventories (Liljencrants and Lindblom 1972; Lindblom 1986), and similar considerations are discussed in Steriade 1997.

The difference between regarding perceptual markedness as a property of contrasts rather than sounds can be clarified through consideration of alternative approaches to the analysis of correlations between backness and lip-rounding in vowels. Cross-linguistically, front vowels are usually unrounded whereas non-low back vowels are usually rounded. This is true of the common five-vowel inventory in (1), and in the UPSID database as a whole, 94.0 per cent of front vowels are unrounded and 93.5 per cent of back vowels are rounded (Maddieson 1984).

(1)
```
i        u
e        o
     a
```

The perceptual explanation for this pattern is that co-varying backness and rounding in this way maximises the difference in second formant frequency (F2) between front and back vowels, thus making them more distinct. In general, front and back vowels differ primarily in F2, with front vowels having a high F2 and back vowels having a low F2. Lip-rounding lowers F2 so the maximally distinct F2 contrast is between front unrounded and back rounded vowels (Liljencrants and Lindblom 1972; Stevens, Keyser, and Kawasaki 1986). This is illustrated in (2), which shows the approximate positions of front and back rounded and unrounded vowels on the F2 dimension. It can be seen that the distinctiveness of contrasts between front and back rounded vowels, e.g. [y-u], or between front and back unrounded vowels, e.g. [i-ɯ], is sub-optimal.

(2)
```
 i   y   ɯ   u
←────────────
      F2
```

The standard phonological analysis of this pattern of co-variation is to posit feature co-occurrence constraints against front rounded vowels and back unrounded vowels (3).

(3) *[−back, +round]
*[+back, −round]

This analysis does not correspond to the perceptual explanation outlined above. The constraints in (3) imply that front rounded vowels and back unrounded vowels are marked sounds, whereas the perceptual explanation implies that it is the contrasts involving front rounded vowels and back unrounded vowels that are dispreferred because they are less distinct than the contrast between a front unrounded vowel and a back rounded vowel. In Optimality-Theoretic terms, there is a general principle that contrasts are more marked the less distinct they are, which implies a ranking of constraints as in (4), where *X-Y means that words should not be minimally differentiated by the contrast between sounds X and Y. (More general constraints that subsume these highly specific constraints will be formulated below).

(4) *y-ɯ ≫ *i-ɯ, *y-y ≫ *i-u

These two accounts make very different predictions: Constraints on the distinctiveness of contrasts predict that a sound may be marked by virtue of the contrasts it enters into. If there are no constraints on contrasts, then the markedness of contrasts should depend simply on the markedness of the individual sounds, and should be insensitive to the system of contrasts. We will see a range of evidence that markedness of sounds is indeed dependent on the contrasts they enter into – that is, that there are markedness relations over contrasts as well as over sounds – and that the relative markedness of contrasts does correspond to their distinctiveness.

For example, the dispreference for front rounded vowels and back unrounded vowels extends to other vowels with intermediate F2 values, such as central vowels. Most languages contrast front and back vowels, and if they have central vowels, they are in addition to front and back vowels. The same explanation applies here also: since central vowels like [ɨ] fall in the middle of the F2 scale in (2), contrasts like [i-ɨ] and [ɨ-u] are less distinct than [i-u] and consequently dispreferred. But we will see in section 4.1 that in the absence of front-back contrasts, vowels with intermediate F2 values, such as central vowels, are the unmarked case in many contexts. A number of languages, including Kabardian (Kuipers 1960; Choi 1991) and Marshallese (Bender 1968; Choi 1992), have short vowel inventories that lack front-back contrasts. These so-called 'vertical' vowel systems consist of high and mid, or high, mid, and low vowels, whose backness is conditioned by surrounding consonants, resulting in a variety of specific vowel qualities, many of which would be highly marked in a system with front-back contrasts, for example central vowels, back unrounded vowels, and short diphthongs. Crucially there are no vertical vowel inventories containing invariant [i] or [u], vowels that are ubiquitous in non-vertical inventories. That is, there are no vowel inventories such as [i, e, a] or [u, o, a].

This pattern makes perfect sense in terms of constraints on the distinctiveness of contrasts: as already discussed central vowels are not problematic in themselves, it is the contrast between front and central or back and central vowels that is marked (*i-ɨ, *ɨ-u $\gg$ *i-u). In the absence of such F2-based contrasts, distinctiveness in F2 becomes irrelevant, and minimisation of effort becomes the key factor governing vowel backness. Effort minimisation dictates that vowels should accommodate to the articulatory requirements of neighbouring consonants. This analysis is developed in section 4.1.

These generalisations about vertical vowel systems show that the markedness of sounds depends on the contrasts that they enter into, because sounds such as central vowels, which are marked when in contrast with front and back vowels, can be unmarked in the absence of such contrasts. The same pattern is observed in vowel reduction: when all vowel qualities are neutralised in unstressed syllables, as in English, the result is typically a 'schwa' vowel – a vowel type that is not permitted in stressed syllables in the same languages.

This type of contrast-dependent markedness cannot be captured in terms of constraints on individual sounds. As Ní Chiosáin and Padgett (1997) point out, the cross-linguistic preference for front unrounded and back rounded vowels over central vowels suggests a universal ranking of segment Markedness constraints as shown in (5), which implies that any language with [ɨ] will have [i, u] also. But this would imply that if only one of these vowels appears it should be [i] or [u], and certainly not a central vowel. More generally, this approach incorrectly predicts that, if a sound type is unmarked, it should be unmarked regardless of the contrasts it enters into.

(5) *ɨ ≫ *u, *i

Constraints on the distinctiveness of contrasts, and their implications for phonology, are the focus of this chapter. However, it is also essential to consider general constraints, such as effort minimisation, that limit the distinctiveness of contrasts, since actual contrasts are less than maximally distinct. So the first step is to situate constraints on the distinctiveness of contrasts within the context of a theory of phonological contrast. This is the topic of the next section. This model will then be applied to the analysis of particular phenomena, demonstrating the range of effects of distinctiveness constraints, and the difficulties that arise for models that do not include constraints on contrasts.

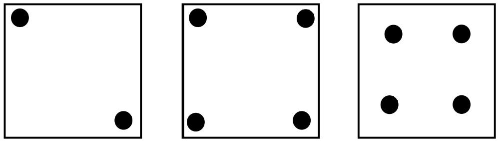
3 The Dispersion Theory of contrast

Constraints on the distinctiveness of contrasts are formalised here as part of a theory of contrast dubbed the 'Dispersion Theory' (Flemming 1995, 1996, 2001) after Lindblom's (1986, 1990) 'Theory of Adaptive Dispersion', which it resembles in many respects. The core of the theory is the claim that the selection of phonological contrasts is subject to three functional goals:

i. Maximise the distinctiveness of contrasts.
ii. Minimise articulatory effort.
iii. Maximise the number of contrasts.

As noted above, a preference to maximise the distinctiveness of contrasts follows from language's function as a means for the transmission of information. This tendency is hypothesised to be moderated by two conflicting goals. The first is a preference to minimise the expenditure of effort in speaking, which appears to be a general principle of human motor behaviour not specific to language. The second is a preference to maximise the number of phonological contrasts that are permitted in any given context in order to enable languages to differentiate a substantial vocabulary of words without words becoming excessively long.

The conflicts between these goals can be illustrated by considering the selection of contrasting sounds from a schematic two-dimensional auditory space, shown in figure 8.1. Figure 8.1a shows an inventory that includes only one



(a)	(b)	(c)
Two categories	Four categories	Four categories
Most separation	Less separation	Least separation
More effort	More effort	Less effort

Figure 8.1 Selection of contrasts from a schematic auditory space

contrast, but the contrast is maximally distinct, that is, the two sound categories are well separated in the auditory space. If we try to fit more sounds into the same auditory space, the sounds will necessarily be closer together, that is, the contrasts will be less distinct (figure 8.1b). Thus the goals of maximising the number of contrasts and maximising the distinctiveness of contrasts inherently conflict. Minimisation of effort also conflicts with maximising distinctiveness. Assuming that not all sounds are equally easy to produce, attempting to minimise effort reduces the area of the auditory space available for selection of contrasts. For example, if we assume that sounds in the periphery of the space involve greater effort than those in the interior, then, to avoid effortful sounds it is necessary to restrict sounds to a reduced area of the space, thus the contrasts will be less distinct, as illustrated in figure 8.1c. Note that while minimisation of effort and maximisation of the number of contrasts both conflict with maximisation of distinctiveness, they do not directly conflict with each other.

3.1 Formulation of the constraints on contrast

Given that the three requirements on contrasts conflict, the selection of an inventory of contrasts involves achieving a balance between them. Optimality Theory (Prince and Smolensky 1993) provides a system for specifying the resolution of conflict between constraints, so this framework is adopted in formalising Dispersion Theory. In this section the functional goals for systems of contrasts posited above are formulated as Optimality-Theoretic constraints.

3.1.1 Maximise the distinctiveness of contrasts Given the considerations outlined in section 2, the measure of distinctiveness that is predicted to be relevant to the markedness of a contrast between two sounds is the probability of confusing the two sounds. Our understanding of the acoustic basis of confusability is limited, so any general model of distinctiveness is necessarily tentative. To allow the precise formulation of analyses, a fairly specific view of distinctiveness will be presented, but many of the details could be modified without affecting the central claims advanced here.

In psychological work on identification and categorisation it is common to conceive of stimuli (such as speech sounds) as being located in a multidimensional similarity space, where the distance between stimuli is systematically related to the confusability of those stimuli – that is, stimuli that are closer together in the space are more similar, and hence more confusable (e.g. Shepard 1957; Nosofsky 1992). This conception is adopted here. The domain in which we have the best understanding of perceptual space is vowel quality. There is good evidence that the main dimensions of the similarity space for vowels correspond well to the frequencies of the first two formants (Delattre, Liberman, Cooper, and Gerstman 1952; Plomp 1975; Shepard 1972), and less clear evidence for a dimension corresponding to the third formant (see Rosner and Pickering 1994: 173ff. for a review).

A coarsely quantised three-dimensional vowel space, adequate for most of the analyses developed here, is shown in (6a–b) (cf. Liljencrants and Lindblom 1972). Sounds are specified by matrices of dimension values, e.g. [F1 *1*, F2 *6*, F3 *3*] for [i]. That is, dimensions are essentially scalar features so standard feature notation is used with the modification that dimensions take integer values rather than +/−. The locations of different vowel qualities are indicated as far as possible using IPA symbols. In some cases there is no IPA symbol for a particular vowel quality (e.g. the unrounded counterpart to [ʊ] which might occupy [F1 *2*, F2 *2*]), while in many cases more than one vowel could occupy a given position in F1-F2 space due to the similar acoustic effects of lip rounding and tongue backing, e.g. central rounded [ʉ] occupies the same position as back unrounded [ɯ]. Also, the IPA low back unrounded vowel symbol [ɑ] is used for a wide range of vowel qualities in transcriptions of English dialects and could have been used to symbolise [F1 *7*, F2 *2*]. Similarly, [y] could also have been used for [F1 *1*, F2 *5*].

(6) a.

F1 \ F2	6	5	4	3	2	1
1	i	i̠	y	ɨ	ɯ	u
2		ɪ	ʏ			ʊ
3		e̝	ø̝		ɤ̝	o̝
4		e	ø	ə	ɤ	o
5			ɛ	ɐ	ʌ	ɔ
6			æ	ɜ		ɑ
7				a	a̠	

b.

F1 \ F3	3	2	1
1	i		y, ɯ, u
2		ɪ	ʏ, ʊ
3		e̝	ø̝, ɤ̝, o̝
4		e	ø, ɤ, o
5		ɛ	ʌ, ɔ
6			ɑ
7			a

The distinctiveness of a pair of vowel qualities should then be calculated from the differences on each of these three dimensions. However, relative distinctiveness on a single dimension can be determined with much greater confidence than distinctiveness involving differences on multiple dimensions, so almost all of the analyses developed here are cases that can be formulated as the selection of a set of contrasting sounds along one perceptual dimension. Consequently we will concentrate on formalising this restricted case. Contrasts on multiple dimensions are discussed in detail in Flemming 2001.

The requirement that the auditory distinctiveness of contrasts should be maximised can be decomposed into a ranked set of constraints requiring a specified minimal auditory distance between contrasting forms (7) (Flemming 1995). The required distance is indicated in the format *Dimension:distance*, for example 'MINDIST = F1:2' is satisfied by contrasting sounds that differ by at least 2 on the F1 dimension.

(7) MINDIST = F1:1 $\gg$ MINDIST = F1:2 $\gg \ldots \gg$ MINDIST = F1:4

To encode the fact that auditory distinctiveness should be maximised, MINDIST = $D{:}n$ is ranked above MINDIST = $D{:}n+1$, that is, the less distinct the contrast, the greater the violation.

3.1.2 Maximise the number of contrasts The requirement that the number of contrasts should be maximised can be implemented in terms of a positive constraint, MAXIMISE CONTRASTS, that counts the number of contrasts in the candidate inventory (Flemming 2001). The largest inventory or inventories are selected by this constraint, all others are eliminated. Of course the largest candidate inventories will usually have been eliminated by higher-ranked constraints, so this constraint actually selects the largest viable inventory.

3.1.3 Balancing the requirements on contrasts The language-specific balance between these first two constraints on contrasts is modelled by specifying the language-specific ranking of the constraint MAXIMISE CONTRASTS in the hierarchy of MINDIST constraints. Effectively, the first MINDIST constraint to outrank MAXIMISE CONTRASTS sets a threshold distance, and the optimal inventory is the one that packs the most contrasting sounds onto the relevant dimension without any pair being closer than this threshold.

The conflict between the two constraints on contrasts is illustrated in the tableau in (8). This tableau shows inventories of contrasting vowel heights and their evaluation by MINDIST and MAXIMISE contrasts constraints. We are considering constraints on contrasts so the candidates evaluated here are sets of contrasting forms rather than outputs for a given input. For simplicity, we are considering only a single perceptual dimension, so the individual vowels are representative of distinctive heights.

MINDIST constraints assign one mark for each pair of contrasting sounds that are not separated by at least the specified minimum distance. For example, candidate (b) violates MINDIST = F1:4 twice because the contrasting pairs [i-e] and [e-a] violate this constraint while [i-a] satisfies it, being separated by a distance of 6 on the F1 dimension. (Note that the number of violations will generally be irrelevant for MINDIST constraints ranked above MAXIMISE CONTRASTS because it will always be possible to satisfy the MINDIST constraint by eliminating contrasts).

MAXIMISE CONTRASTS is a positive scalar constraint, according to which more contrasts are better, so evaluation by this constraint is indicated using one check mark (✓) for each contrasting sound category – more check marks indicate a better candidate according to this constraint. The conflict between the two constraint types is readily apparent in (8): sets of vowel height contrasts that better satisfy MAXIMISE F1 CONTRASTS incur worse violations of the MINDIST constraints.

(8)

		MINDIST = F1:1	MINDIST = F1:2	MINDIST = F1:3	MINDIST = F1:4	MINDIST = F1:5	MAXIMISE CONTRASTS
a.	i-a						✓✓
b.	i-e-a				**	**	✓✓✓
c.	i-e̞-ɛ-a			***	***	*****	✓✓✓✓
d.	i-ɪ-e̞-ɛ-a		**	*****	******	********	✓✓✓✓✓

The effect of ranking MAXIMISE CONTRASTS at different points in the fixed hierarchy of MINDIST constraints is illustrated by the tableaux in (9) and (10). The ranking in (9) yields three distinct vowel heights – that is, the winning candidate is (b). This candidate violates MINDIST = F1:3, but any attempt to satisfy this constraint by improving distinctiveness, as in candidate (a), violates higher-ranked MAXIMISE CONTRASTS by selecting only two contrasting vowel heights. It is not possible to fit three contrasting vowels with a minimum separation of three features on the F1 dimension. Candidate (c) better satisfies MAXIMISE CONTRASTS than (b), maintaining four contrasting vowel heights, but (c) violates higher-ranked MINDIST = F1:2 since [e-ɛ] and [ɛ-a] each differ by only 1 on the F1 dimension.

(9)

		MINDIST = F1:2	MINDIST = F1:3	MAXIMISE CONTRASTS	MINDIST = F1:4	MINDIST = F1:5
a.	i-a			✓✓!		
b.	☞ i-e-a			✓✓✓	**	**
c.	i-e̞-ɛ-a		*!**	✓✓✓✓	***	*****

Thus the particular balance achieved here between maximising the number of contrasts and maximising the distinctiveness of the contrasts yields three contrasting heights. Altering the ranking of MAXIMISE CONTRASTS results in a different balance. For example, if less weight is given to maximising the number of contrasts, ranking MAXIMISE CONTRASTS below MINDIST = F1:3, the winning candidate has just two contrasting vowel heights, differing maximally in F1. It is apparent that the maximally distinct F1 contrast [i-a] is preferred over any sub-maximal contrast, such as [i-æ] (which violates MINDIST = F1:6), although this comparison is not included in the tableau.

(10)

		MINDIST = F1:2	MINDIST = F1:3	MINDIST = F1:4	MAXIMISE CONTRASTS	MINDIST = F1:5
a.	☞ i-a				✓✓	
b.	i-e-a			*!*	✓✓✓	**
c.	i-e̞-ɛ-a		*!**	***	✓✓✓✓	*****

Not all conceivable rankings of MAXIMISE CONTRASTS correspond to possible languages. The balance between maximisation of the number of contrasts and maximisation of the distinctiveness of contrasts is determined by the ranking of MAXIMISE CONTRASTS relative to the MINDIST constraints. Allowing all definable rankings predicts the existence of languages that value the number of contrasts very highly, resulting in a huge number of very fine contrasts, and languages that value distinctiveness very highly, resulting in a handful of maximally distinct contrasts. Neither of these extremes is attested. It seems that there is a lower bound on the distinctiveness required for a contrast to be functional, and that there is an upper bound beyond which additional distinctiveness provides a poor return on the effort expended. This could be implemented by specifying that certain MINDIST constraints, referring to the smallest acceptable contrastive differences, are universally ranked above MAXIMISE CONTRASTS, and that MAXIMISE CONTRASTS is in turn universally ranked above another set of MINDIST constraints that make 'excessive' distinctiveness requirements. However, it would be desirable to derive these bounds from general considerations of perceptibility and communicative efficiency rather than simply stipulating them.

Note that the need to place limits on possible constraint rankings is not novel to the Dispersion Theory. The same issue arises with respect to standard faithfulness constraints: If all faithfulness constraints are at the top of the ranking then all inputs will surface as well-formed outputs, that is, this ranking would yield an unattested language with no restrictions on the form of words. Conversely, if all faithfulness constraints were at the bottom of the ranking, then all inputs would be mapped to a single, maximally well-formed output (presumably the null output, i.e. silence).

3.1.4 Minimisation of effort The analyses above do not include effort-minimisation constraints. No general account of the effort involved in speech production will be proposed here, instead specific constraints such as 'Don't voice obstruents' and 'Don't have short low vowels' will be motivated as they become relevant. If a sound violates an effort constraint that outranks MAXIMISE CONTRASTS, it will not be employed, even if it would allow more contrasts or more distinct contrasts.

3.2 Some effects of MINDIST constraints

3.2.1 Dispersion The most basic consequence of the distinctiveness constraints (MINDIST constraints) is a preference for distinct contrasts. This gives rise to dispersion effects whereby contrasting sounds tend to be evenly distributed over as much auditory space as effort constraints will allow (cf. Liljencrants and Lindblom 1972; Lindblom 1986). This effect has already been demonstrated above in relation to F1 contrasts, and the preference for front unrounded and back rounded vowels discussed in section 2 is another instance of this tendency, applied to contrasts on the F2 dimension. The acoustic effects of lip-rounding mean that the maximal F2 difference is between front unrounded vowels and back rounded vowels (11), so if maximisation of distinctiveness of F2 contrasts outranks maximising the number of contrasts, these are the vowels that will be selected (12). F2 contrasts involving central vowels are necessarily sub-maximal, and thus are dispreferred. Of course, the appearance of non-peripheral vowels may be motivated by the desire to maximise contrasts – that is, if MAXIMISE CONTRASTS is ranked above MINDIST = F2:3.

(11) F2:

6	5	4	3	2	1
i	i̠	y	ɨ	ɯ	u

(12)

		MINDIST = F2:3	MAXIMISE CONTRASTS	MINDIST = F2:4	MINDIST = F2:5
a.	☞ i-u		✓✓		
b.	i-ɯ		✓✓		*!
c.	y-u		✓✓		*!
d.	i-ɨ		✓✓	*!	*
e.	i-ɨ-u	*!	✓✓✓	**	**

This notion of dispersion of contrasting sounds is also closely related to the concept of 'enhancement', a term coined by Stevens, Keyser, and Kawasaki (1986). Stevens et al. observe that 'basic' distinctive features are often accompanied by 'redundant' features which 'strengthen the acoustic representation of distinctive features and contribute additional properties which help the listener to perceive the distinction' (p. 426). They regard the relationship between [back] and [round] in vowels as one of enhancement: [round] enhances distinctive [back]. In terms of the Dispersion Theory, this can be reformulated as the observation that independent articulations often combine to yield more distinct contrasts.

3.2.2 Neutralisation A second basic effect of MINDIST constraints, in interaction with the other Dispersion-Theoretic constraints, is neutralisation of indistinct contrasts. In Dispersion Theory, neutralisation of a contrast results when constraints prevent it from achieving sufficient distinctiveness in some environment. That is, in a ranking of the form shown in (13) where *EFFORT is an effort minimisation constraint penalising some articulation, a contrast will be neutralised in some context if it cannot be realised with a distinctiveness of *d* without violating *EFFORT.

(13) MINDIST $= d$, *EFFORT $\gg$ MAXIMISE CONTRASTS

The distinctiveness that can be achieved for a given degree of effort varies across contexts. Some cues to contrasts are simply unavailable in certain contexts, for example release formant transitions are not available as a cue to consonant place if the consonant is not released into an approximant. In addition, the articulatory effort involved in realising a cue is generally highly context-dependent, for example voicing of an obstruent is more difficult following a voiceless sound than following a voiced sound because it is more difficult to initiate voicing than to sustain it (Westbury and Keating 1986). So the possibility of realising a contrast that satisfies MINDIST $= d$ without violating *EFFORT depends on context, and consequently a given type of contrast may be selected as optimal in some contexts and not in others – that is, the contrast is neutralised in those other contexts. For example, consonant place contrasts may be permitted before sonorants, but neutralised before obstruents, where stop bursts and release transitions are not available. Thus Dispersion Theory provides an account of Steriade's (1995, 1997) generalisation that contrasts are neutralised first in environments where 'the *cues* to the relevant contrast would be diminished or obtainable only at the cost of additional articulatory manoeuvres' (Steriade 1997: 1).

It is important to note that the ranking of other constraints will typically be crucial in making the realisation of a distinct contrast more effortful in a particular context – for example stop bursts will only be absent before obstruents if some constraint requires the stop closure to overlap with the following consonant. In the example we will consider here, metrical constraints on unstressed vowel duration make distinct vowel contrasts more difficult to realise in unstressed syllables.

The analysis of neutralisation will be exemplified with analyses of two common patterns of vowel reduction: reduction from a seven-vowel inventory (14i) in primary stressed syllables to a five-vowel inventory (14ii) in other syllables, as in central Italian dialects (Maiden 1995), and reduction from a five-vowel inventory (14ii) in primary stressed syllables to a three-vowel inventory (14iii) elsewhere, as in southern Italian dialects (Maiden 1995) and Russian (Halle 1959).

(14)	(i)				(ii)				(iii)			
		i		u		i		u		i		u
		e		o		e		o			a	
		ɛ		ɔ			a					
			a									

The central Italian pattern is exemplified in (15) with data from standard Italian (as described in dictionaries). The pairs of words on each line are morphologically related so the parenthesised forms illustrate alternations that arise when stress is shifted off a vowel that cannot appear in an unstressed syllable.

(15)	Stressed vowels		Unstressed vowels	
[i]	v**í**no	'wine'	v**i**nífero	'wine-producing'
[e]	p**é**ska	'fishing'	p**e**skáre	'to fish'
[ɛ]	b**ɛ́**lːo	'beautiful'	(b**e**líno	'pretty')
[a]	m**á**no	'hand'	m**a**nuále	'manual'
[ɔ]	m**ɔ́**lːe	'soft'	(m**o**lːeménte	'softly')
[o]	n**ó**me	'name'	n**o**mináre	'to name, call'
[u]	k**ú**ra	'care'	k**u**ráre	'to treat'

The southern Italian pattern is exemplified by the dialect of Mistretta, Sicily (Mazzola 1976) (16).

(16)	Stressed vowels		Unstressed vowels	
[i]	v**í**nːi	'he sells'	v**i**nːímu	'we sell'
[e]	v**é**ni	'he comes'	(v**i**nímu	'we come')
[a]	**á**vi	'he has'	**a**víti	'he has'
[o]	m**ó**ri	'he dies'	(m**u**rímu	'we die')
[u]	**ú**ɟːi	'he boils'	**u**ɟːímu	'we boil'

These patterns of reduction involve neutralisation of F1 contrasts only. According to the analysis of neutralisation outlined above, this implies that it is more difficult to produce distinct F1 contrasts in unstressed positions. The most likely source of that difference in difficulty is the difference in duration between primary stressed and other vowels in these languages. So the proposed analysis is that producing low vowels is increasingly difficult as vowel duration is reduced, and this motivates raising of short low vowels, leaving a smaller range of the F1 dimension for distinguishing F1 contrasts. This in turn can result in the selection of a smaller number of contrasts.[1]

The most direct evidence for a relationship between vowel duration and the ability to achieve a high F1 comes from Lindblom's (1963) finding that the F1 of Swedish nonhigh vowels decreases exponentially as vowel duration decreases. It is also well established that low vowels are longer than high vowels, other things being equal (Lehiste 1970). These effects are commonly attributed to the greater articulator movement involved in producing a low vowel between consonants: low vowels require an open upper vocal tract to produce a high F1, whereas all consonants (other than pharyngeals and laryngeals) require upper-vocal tract constrictions, so producing a low vowel between consonants requires substantial opening and closing movements. Westbury and Keating (1980, cited in Keating 1985) provide evidence that vowel duration differences are indeed related to distance moved: they found that vowels with lower jaw positions had longer durations in a study of English. Thus, producing a low vowel with the same duration as a higher vowel will typically require faster, and consequently more effortful, movements. Reduction of low /a/ to [ɐ] or [ə] in unstressed syllables is accordingly commonly reported both impressionistically and in experimental studies such as Lindblom 1963.

This correlation between duration and raising of low vowels has been observed in central Italian also: a study of vowels in the speech of five male Italian television news readers (Albano Leoni et al. 1995) found that /a/ in a primary stressed syllable was twice as long as medial unstressed /a/, and the mean F1 of /a/ was 750 Hz in primary stressed syllables, but 553 Hz in medial unstressed syllables (close to the F1 of a stressed lower-mid vowel)[2]. So the inventory in unstressed positions is more accurately transcribed as [i, e, ɐ, o, u], where [ɐ] is a lower-mid central vowel. High and higher-mid vowels had essentially the same F1 in stressed and unstressed positions.

While the direct effect of vowel shortening is to increase the difficulty of producing low vowels, this has obvious consequences for the selection of F1 contrasts: if the lowest vowel in an inventory is lower-mid ([F1 *5*] in the terms used here) this leaves less room for distinguishing F1 contrasts than in stressed syllables where the lowest vowel is truly low ([F1 *7*]), so it is not possible to maintain the same number of height contrasts with the same distinctiveness. Consequently, three vowel heights are selected in unstressed syllables and four in longer, stressed syllables.

This analysis can be formalised in terms of the constraint ranking in (17). The positions of relevant vowels on the F1 dimension are shown in (18).

(17) Unstressed vowels are short, *Short low V, Mindist = F1:2
≫ Maximise contrasts ≫ Mindist = F1:3

(18) F1:

7	6	5	4	3	2	1
a	ɐ	ɛ	e	e̞	ɪ	i
		ɐ	ə			

Unstressed vowels are short is a place-holder for whatever constraints require unstressed vowels to be shorter than stressed vowels. *Short low V is a constraint against expending the effort to produce a short low vowel. This should properly be derived from a general model of articulatory effort (cf. Kirchner, this volume), but for present purposes we will formalise it as a constraint that penalises short vowels with F1 of greater than 5 on the scale in (18).

In stressed syllables, the first two constraints are irrelevant, so the ranking yields four vowel heights, each separated by F1:2, as shown in (19). However, in unstressed syllables, high-ranking Unstressed vowels are short requires short vowels, so *Short low V is applicable also. This effort-minimisation constraint penalises low vowels, so the candidate [i-e̞-ɛ-a] is now ruled out because [a] has [F2 7] (20a). Attempting to maintain four contrasts while avoiding low vowels, as in candidate (b), results in violations of Mindist = F1:2 because [ɛ-ɐ] do not differ in F1. The winning candidate has three vowel heights, and so is evaluated as worse by Maximise contrasts, but satisfies the higher-ranked minimum distance requirement.

(19) Central Italian – vowels in primary stressed syllables

		*Short low V	Mindist = F1:2	Maximise contrasts	Mindist = F1:3
a.	í-á			✓✓!	
b.	í-é-á			✓✓✓!	
c.	☞ í-é̞-ɛ́-á			✓✓✓✓	***

(20) Central Italian – vowels in unstressed syllables

		*Short low V	Mindist = F1:2	Maximise contrasts	Mindist = F1:3
a.	i-e̞-ɛ-a	*!		✓✓✓✓	***
b.	i-e̞-ɛ-ɐ		*!	✓✓✓✓	***
c.	☞ i-e̞-ɐ			✓✓✓	**

A similar ranking (21) derives the southern Italian pattern in which three vowel heights are contrasted in primary stressed syllables, but only two elsewhere. The only difference is that both MINDIST constraints are ranked above MAXIMISE CONTRASTS – that is, distinctiveness requirements are more demanding.

(21) UNSTRESSED VOWELS ARE SHORT, *SHORT LOW V, MINDIST = F1:2 ≫ MINDIST = F1:3 ≫ MAXIMISE F1 CONTRASTS

This is the same ranking of MINDIST constraints used to derive three contrasting vowel heights in (9) above, and the same derivation applies in primary stressed syllables, where the top-ranked constraints are irrelevant. In unstressed syllables, *SHORT LOW V is applicable again, so the lowest vowel possible is [ɐ], and it is not possible to fit a vowel between [i] and [ɐ] while satisfying MINDIST = F1:3 (22b), so only two vowel heights are selected (22c).

(22) Southern Italian – vowels in unstressed syllables

		*SHORT LOW V	MINDIST = F1:2	MINDIST = F1:3	MAXIMISE CONTRASTS
a.	i-e-a	*!			✓✓✓
b.	i-e̞-ɐ			*!*	✓✓✓
c.	☞ i-ɐ				✓✓

This analysis is based on the assumption that the difference between the two patterns of reduction lies in the ranking of MINDIST constraints, but there may also be differences in the characteristic durations of the unstressed vowels. The difficulty of producing a low vowel increases as vowel duration decreases, so if southern Italian unstressed vowels are shorter than central Italian unstressed vowels, then more raising of low vowels may occur, making reduction to a two-height system more desirable. Good evidence that different degrees of shortening can result in different degrees of reduction in this way is provided by Brazilian Portuguese. Brazilian Portuguese combines the two patterns of reduction: the seven-vowel system (14i) is permitted in primary stressed syllables, the five-vowel system (14ii) in syllables preceding the stress, and the three-vowel system (14iii) in unstressed final syllables (stress is generally penultimate) (Mattoso Camara 1972). Since both patterns of reduction occur in the same language, they cannot be accounted for in terms of differences in the ranking of MINDIST constraints, but they can be accounted for in terms of differences in vowel duration. Major (1992) found that final unstressed syllables are substantially shorter than pre-stress syllables (which are in turn shorter than stressed syllables),[3] so the same degree of effort should result in higher vowels in this position. The 'low' vowel is indeed higher in this position, as indicated by the standard impressionistic transcription of the final unstressed vowel system as [i, ə, u] (e.g. Mattoso Camara 1972). Acoustic data reported by Fails and Clegg (1992) show a progressive decrease in F1 of the lowest vowel from primary stressed, to pre-stress, to final unstressed.

We will see that duration-based neutralisation is also central to some of the case studies that provide more direct evidence for distinctiveness constraints (section 4.1), but before turning to those cases, we will consider some additional issues in the formulation and application of Dispersion Theory.

3.2.3 Analysis of words and alternations The analysis of vowel reduction raises two important issues concerning analyses using Dispersion Theory. First, we have only considered the selection of inventories of contrasting sounds, but a phonology must characterise the set of well-formed possible words in a language. The implication of Dispersion Theory is that words must be evaluated with respect to paradigmatic constraints in addition to the familiar syntagmatic Markedness constraints, such as effort minimisation and metrical constraints. That is, words must be sufficiently distinct from other minimally contrasting possible words (MINDIST), and there must be a sufficient number of such contrasting words (MAXIMISE CONTRASTS). Deriving inventories of sounds in particular contexts is an important step towards the analysis of complete words because, for a word to be well formed, each sound in that word must be a member of the optimal inventory for its particular context. We will see in section 5 that developing this basic idea is not simple, but we will postpone that discussion until we have more thoroughly motivated the constraints on contrast.

The second issue raised by the analysis of vowel reduction is how morpheme alternations should be analysed. The analysis in section 3.2.2 derives the distributional fact that [e] is not permitted in short, unstressed syllables in Sicilian Italian, but it says nothing about the fact that [e] alternates with [i] when stress shifts off it, for example [v**é**ni ∼ v**i**nímu] (16). In standard OT, the analysis of alternations centres on faithfulness to the underlying representation of a morpheme, but it is not possible to combine dispersion constraints with the faithfulness-based account of allomorphic similarity because the two are fundamentally incompatible. This is illustrated in (23). This tableau repeats the ranking used in (9) to derive three contrasting vowel heights [i-e-a], with the addition of a top-ranked faithfulness constraint IDENT [F1], which requires that output segments have the same [F1] value as the corresponding input segment – that is, input values of [F1] must be preserved in the output. The problem arises where the input contains a vowel that is not part of the inventory derived by the dispersion constraints, as in (23). In the candidate inventories, the underlined form is the output corresponding to input /ɪ/, whereas the other forms are the set of contrasting vowels required for the evaluation of constraints on contrast.

The inclusion of faithfulness constraints subverts the intended effect of the MINDIST and MAXIMISE CONTRASTS constraints, because it makes the selected inventory of vowel height contrasts dependent on the input under consideration – the same constraints that are supposed to derive three vowel heights, as in (9), yield two [ɪ-a] in (23) because faithfulness to the input F1 forces inclusion of [ɪ] in an output form. The expected three vowel heights are derived if the input is /i/, for example.

(23)

	/ɪ/	IDENT [F1]	MINDIST = F1:3	MAXIMISE CONTRASTS	MINDIST = F1:4
a.	i̲-e-a	*!		✓✓✓	**
b.	ɪ̲-e-a		*!	✓✓✓	**
c.	☞ ɪ̲-a			✓✓	
d.	i̲-a	*!		✓✓	

In Flemming 1995 it is proposed that allomorphy should be analysed in terms of a direct requirement of similarity between the surface forms of a morpheme, that is ‘output-output correspondence’ or ‘paradigm uniformity’ constraints. These constraints have been used to account for cyclicity and related effects (e.g. Benua 1997; Burzio 1998; Kenstowicz 1996; Steriade 1997, 2000), but, as Burzio (1998) observes, they can naturally be extended to account for all similarity relations between realisations of a morpheme including those observed in allomorphy, eliminating any role for an input. So [v**é**ni] alternates with [v**i**n**í**mu] because [i] is the most similar to [e] of the vowels that are permitted in unstressed syllables. However, most of the analyses considered here concern distribution rather than alternations, so we will not pursue this line further here.

4 Evidence for constraints on contrasts

Now we have laid out the basics of a theory that incorporates constraints on the distinctiveness of contrasts, the theory will be applied in the analysis of phenomena that illustrate the effects of these constraints, and that are problematic for other theories. In general terms, the case studies provide evidence for the central prediction that the markedness of a sound depends on the sounds that it contrasts with. Without constraints on contrasts, markedness is predicted to be purely a property of sounds, so the markedness of a sound should be independent of the nature of the sounds that it contrasts with.

4.1 F2 contrasts and vowel dispersion

The first case study concerns the preference for front unrounded and back rounded vowels discussed in section 2. This pattern has already been analysed as a result of the preference for maximally distinct contrasts, that is, the MINDIST constraints (section 3.2.1): the maximal F2 difference is between front unrounded vowels and back rounded vowels, so if maximisation of distinctiveness of F2 contrasts outranks maximising the number of contrasts, these are the vowels that will be selected. F2 contrasts involving non-peripheral vowels (central vowels, front rounded vowels, etc.) are necessarily sub-maximal, and thus are dispreferred.

The novel prediction made by this analysis is that front unrounded and back rounded vowels should only be preferred where there are F2 contrasts. If there are no vowel contrasts that are primarily realised in terms of F2 differences, other constraints are predicted to govern backness and rounding of vowels, the most general of which are effort-minimisation constraints. It is unusual for all vowel F2 contrasts to be neutralised, but there are two circumstances in which this happens: in 'vertical' vowel inventories, and in fully neutralising vowel reduction in unstressed syllables, as in English reduction to 'schwa'. In both cases the predictions of the Dispersion-Theoretic analysis are confirmed: we do not find front unrounded or back rounded vowels in most contexts, rather backness and rounding are governed by minimisation of effort. This means that they are realised as smooth transitions between preceding and following consonants, which frequently results in central or centralised vowel qualities.

4.1.1 Vowel qualities in the absence of F2 contrasts Vowel inventories that lack front-back contrasts are found in Marshallese (Bender 1968; Choi 1992), North-west Caucasian languages (Colarusso 1988) including Kabardian (Kuipers 1960; Colarusso 1992) and Shapsug (Smeets 1984), and some Ndu languages of Papua New Guinea including Iatmul (Laycock 1965; Staalsen 1966). These languages are typically described as having only central vowels; however, this is a claim about the underlying vowel inventory posited as part of a derivational analysis, not an observation about the surface vowels. On the surface, all of these languages distinguish short vowels from longer vowels, with conventional F2 contrasts among the longer vowels, but no F2 contrasts among the short vowels. For example, the north-west Caucasian languages Kabardian and Shapsug have a system of five normal length vowels [i, e, a, o, u] (Kuipers 1960: 23f.; Smeets 1984: 123) and a 'vertical' system of two short vowels, which can be transcribed broadly as [ɨ, ə].[4] However, the precise backness and rounding of these vowels depends on their consonantal context. Colarusso states that in north-west Caucasian languages, 'The sequence C_1ɨC_2 means "go from 1 to 2, letting your tongue follow the shortest path that permits an interval of sonorant voicing." C_1əC_2 means "go from 1 to 2 . . . but at the same time imposing on this trajectory an articulatory gesture which pulls the tongue body down and back"'[5] (1998: 307). Marshallese also has a long vowel inventory with F2 contrasts, but the medial short vowels /ɨ, ə, a/ contrast in height only (Bender 1968). Again the backness and rounding of these vowels is dependent on consonant context: Choi (1992) shows that the F2 trajectory of these vowels is a nearly linear interpolation between F2 values determined by the preceding and following consonants.

These transitional vowel qualities are plausibly analysed as the result of effort minimisation. Although articulatory effort is not well understood, basic considerations imply that higher velocity movements should be more effortful (Kirchner, this volume; Nelson 1983; Perkell 1997), and velocity of movement in a vowel is minimised by adopting a linear trajectory between preceding and following consonants. Some deviation from a linear vowel height trajectory is necessary to achieve a vocalic degree of stricture, and to realise F1 contrasts, but backness and rounding can be interpolated between preceding and following consonants, producing the near-linear F2 movements observed by Choi.

So vertical vowel systems are what we expect given the analysis of F2 dispersion above – where F2 contrasts are neutralised, backness and rounding of vowels are determined by effort minimisation. The resulting vowel qualities are often central, back unrounded, front rounded, or short diphthongs involving these qualities. These are all vowel types that would be highly marked in the presence of F2 contrasts, so the markedness of vowel qualities depends on the contrasts they enter into.

This conclusion holds even more clearly if we follow Choi (1992) in analysing these vowels as being phonetically underspecified for [back] and [round] – that is, these vowels lack specifications for these features in the output of the phonology, and the specific contextual allophones are generated through a process of phonetic interpolation. Such unspecified vowels only occur in the absence of F2 contrasts, so they are not just marked in the presence of F2 contrasts, they are unattested.

The other situation in which we find neutralisation of F2 contrasts is in vowel reduction. In languages such as English (Hayes 1995), southern Italian dialects (Maiden 1995), and Dutch (Booij 1995), all vowel quality distinctions are neutralised in some unstressed syllables. The resulting vowel is usually referred to as 'schwa'. Phonetic studies of schwa in Dutch (Van Bergem 1994) and English (Kondo 1994) indicate that this vowel can also be analysed as the result of effort minimisation predominating where vowel contrasts are neutralised.[6] As in vertical vowels, F2 in schwa is an almost linear interpolation between values for adjacent consonants.[7] Since there are no height contrasts, F1 of schwa is expected to be transitional also. In most consonant contexts an opening movement is required to realise a vocalic stricture, but minimising this opening movement results in a vowel with low F1, comparable to that of high vowels, as observed by Van Bergem (1994) and Kondo (1994).[8] However, these studies did not include schwas adjacent to nonhigh vowels, or separated from them by laryngeals (as in 'saw **a**nother'). Examination of sequences of this kind in English suggests that F1 interpolates from the low vowel to the following consonant, which can result in a relatively high F1 during schwa.

Preliminary investigation of the southern Italian dialect of Bari, based on recordings provided with Valente (1975), suggests that schwa is much the same as in English and Dutch. It is also predicted that the schwa vowels that break up consonant clusters in some Berber and Salishan languages (Dell and Elmedlaoui 1996; Flemming et al. 1994) should be similar transitional vowels since there are no vowel-quality contrasts in these positions. This appears to be correct for Montana Salish.

Schwa is not permitted where there are vowel-quality contrasts (in stressed syllables), but is the unmarked vowel where quality contrasts are neutralised (e.g. in unstressed syllables). So reduction to schwa demonstrates a similar pattern of contrast-dependent markedness to that observed in vertical vowel languages. These patterns cannot be captured in terms of constraints on individual sounds.

The same applies if it is assumed that transitional vowels are simply unspecified for [back] and [round], or [F2]. A constraint against such unspecified vowels would have to be ranked above constraints such as *i, *u to prevent transitional vowels from surfacing in F2 contrasts, but such a ranking implies that unspecified vowels should always be dispreferred, even in neutralisation.

It is often possible to propose a re-analysis of a pattern of contrast-dependent markedness as positional markedness. For example, vertical vowel inventories seem to be restricted to extra-short vowels, and the schwa found in neutralising reduction is very short (see below), so it is possible to formulate a constraint against vowel qualities with non-transitional F2 among extra-short vowels and restrict the Markedness constraints against transitional vowels to apply only to longer vowels (24).

(24) *NORMAL DURATION[ɨ] ≫ *NORMAL DURATION[i], *NORMAL DURATION[u]
*EXTRA-SHORT[i], *EXTRA-SHORT[u] ≫ *EXTRA-SHORT[ɨ]

This strategy runs into difficulties because the full typology of extra-short vowels is more complex. Neutralisation to a single vowel quality results in a schwa vowel in which both F1 and F2 are essentially transitional (although F1 must be above a certain minimum). This implies that schwa should be the least marked extra-short vowel. But if this is the case, schwa should be found in all inventories of extra-short vowels, which is not the case. Vertical vowels have specific F1 targets, and transitional schwa is also excluded from the extra-short vowel inventory [i, ɘ, u] (where [ɘ] is used in the IPA sense of a mid central rounded vowel), found in unstressed final vowels in Brazilian Portuguese (section 3.2.2) and most unstressed syllables in Standard Russian (Crosswhite this volume).

So even among extra-short vowels, markedness depends on the system of contrasts, making it impossible to arrange vowel types in a single markedness hierarchy. Perhaps some basis could be found for differentiating the positions in which we find reduction to [i, ɘ, u], positions in which we find reduction to schwa, and positions in which we find vertical vowels. Then it would be possible to posit distinct hierarchies of vowel markedness for each type of position, but such a proliferation of increasingly specific constraints should prompt us to seek more general organising principles, as we have done here. For example, positing position-specific hierarchies leaves it as a remarkable coincidence that vowels with transitional F2 are unmarked in precisely the hierarchies for positions where F2 contrasts are neutralised, and vowels with transitional F1 are unmarked in positions where F1 contrasts are neutralised.[9]

4.1.2 The motivation for neutralising vowel F2 contrasts In this section we will briefly address the motivation for neutralising vowel F2 contrasts. We have seen that the Dispersion-Theoretic analysis correctly predicts the properties of vowel inventories without F2 contrasts, but we have not yet explained why a language would forgo F2 contrasts in the first place. The argument made above only depends on the outcome of F2 neutralisation, so the motivation for neutralisation is not directly relevant here, but it might be thought that vertical vowel inventories contradict MAXIMISE CONTRASTS by failing to exploit F2 contrasts, so it is useful to show that this is not the case.

The analysis proposed here is that neutralisation of vowel F2 follows the standard pattern described in section 3.2.2: F2 contrasts are neutralised in contexts where it is too difficult to realise them distinctly. A key factor that contributes to this difficulty is very short vowel duration. In section 3.2.2 we saw evidence that short duration makes it difficult to produce high F1 in a vowel because there is little time for the necessary opening and closing movements. Similar considerations apply to the realisation of F2 contrasts in shorter vowels. Lindblom (1963) shows that F2 at the mid-point of a vowel in a CVC where both consonants are the same tends to move closer to an F2 value characteristic of the consonant as the duration of the vowel is reduced. As a vowel becomes shorter, it becomes more effortful to deviate from the least-effort transition between preceding and following consonants by a significant amount, but deviation from the least-effort transition is required to realise distinct F2 values for contrasting vowels. At short durations, the effort of realising a distinct F2 contrast can be sufficient to make neutralisation optimal.

Schwa vowels are typically extremely short – Kondo (2000) reports a mean duration of 34 ms for English – so it is unsurprising that it is difficult to maintain either F1 or F2 contrasts in this context. Vertical vowels are also short, although generally longer than schwa. The Caucasian vowel length opposition is between short vowels and extra-short vowels, rather than between long and short vowels as in Japanese or Finnish (Choi 1991; Kuipers 1960: 24; Smeets 1984: 122; Colarusso 1988: 349), and the Marshallese vertical vowels are comparable to these extra-short vowels (although the low vowel is longer – presumably this is necessary to reach a high F1) (Choi 1992). But in these languages the difficulty presented by short duration is exacerbated by rich inventories of consonant place contrasts. F2 transitions play an important role in realising these contrasts, so it is less possible to facilitate vowel contrasts by co-producing vowels with consonants. Marshallese has an extensive system of palatalisation, velarisation, and labio-velarisation contrasts (e.g. [p^j-$p^ɣ$], [k-k^w]), and sequences such as [p^jup^j] and [$p^ɣ$i$p^ɣ$] obviously require substantial tongue body movement. The Caucasian languages contrast large sets of places of articulation, together with some secondary articulations. So to some extent it appears that vertical vowel inventories are trading vowel F2 contrasts for consonant-centered F2 contrasts. Indeed, analysts have varied between characterising Arrernte as a vertical vowel language with extensive labio-velarisation contrasts, or as a language with vowel F2 contrasts, and a smaller consonant inventory (Ladefoged and Maddieson 1996: 357). However, it is apparent that rich consonant contrasts alone do not give rise to neutralisation, because F2 contrasts are maintained among longer vowels in the same consonant contexts.

The constraint ranking in (25) is a partial formalisation of this analysis of vertical vowels. The constraint *HIGHEFFORT is intended to penalise particularly rapid movements – specifically, with very short vowels, it rules out anything more than small deviations from a smooth transition between tongue body and lip positions for preceding and following consonants. The MINDIST constraint imposes a substantial minimum distance for vowel contrasts in F2, and for contrasts based primarily on F2 during consonant release transitions. This constraint is satisfied by contrasts between fully front and back vowels (e.g. i-u, e-o) or between palatalised and velarised consonants (see sample F2 specifications in (26)).

(25) *HIGHEFFORT, MINDIST = F2:4 ≫ MAXIMISE CONTRASTS

(26) F2:

6	5	4	3	2	1
i	i̠	ɨ̠	ɨ	ɯ	u
	e	ə̟	ə	ɤ	o
Cʲ				Cˠ	

The operation of these constraints is illustrated by the tableau in (27) which shows the selection of an inventory of CVCs with extra-short vowels, considering only secondary articulation contrasts as representatives of consonant contrasts and only F2 contrasts among vowels.

(27)

					*HIGH EFFORT	MINDIST = F2:4	MAXIMISE CONTRASTS
a.	CʲĭCʲ CˠĭCʲ	CʲŭCʲ CˠŭCʲ	Cʲĭ Cˠ Cˠĭ Cˠ	CʲŭCˠ CˠŭCˠ	*!*****		8
b.	CʲĭCʲ Cˠĭ̠Cʲ	Cʲɨ̟̆ Cʲ Cˠɨ̆Cʲ	Cʲĭ̠Cˠ Cˠɨ̟̆ Cˠ	Cʲɨ̆Cˠ Cˠɯ̆Cˠ		*!***	8
c. ☞			Cʲɨ̆Cʲ Cˠɨ̆Cʲ	Cʲɨ̆Cˠ Cˠɨ̆Cˠ			4
d.			CĭC	CŭC			2!

Candidate (a) best satisfies MAXIMISE CONTRASTS since it allows palatalisation-velarisation contrasts on consonants in all positions, and front-back contrasts in vowels; however, CVCs such as [pʲupʲ, pˠipˠ] involve substantial violations of *HIGHEFFORT since they involve movement from two full front-back movements in a short duration. Candidate (b) is intended to include CVCs that barely satisfy *HIGHEFFORT, that is, they represent the maximum allowable effort, while maintaining vowel and consonant contrasts in all positions. However, with such short vowels, the maximum allowable effort results in only small deviations from a smooth transition between the secondary articulations of the consonants (indicated by somewhat arbitrary transcriptions), and consequently indistinct F2 contrasts, so candidate (b) violates the MINDIST constraint. The winning candidate, (c), satisfies *HIGHEFFORT since it involves only transitional vowels, transcribed here with central vowel symbols. There are no vowel F2 contrasts, and the palatalisation-velarisation contrasts satisfy the MINDIST constraint.

Candidate (d) satisfies the *HIGHEFFORT and MINDIST constraints by neutralising consonant contrasts rather than vowel contrasts. This candidate loses out to (c) because it realises fewer contrasts. Probably other considerations contribute to this outcome – for example consonant place contrasts will typically be cued by a release burst or during the consonant constriction itself, as well as by formant transitions, so they may be more distinct than extra-short vowel F2 contrasts – but it seems likely that one advantage of adopting a vertical vowel system is that many consonant contrasts can be differentiated in a relatively short duration (consonant constriction plus transitions), whereas distinct vowel contrasts take longer to realise. So abandoning vowel F2 contrasts may actually be motivated by MAXIMISE CONTRASTS rather than being in conflict with this constraint.

This analysis suggests that vertical vowels are similar to the schwa vowels of Berber and Salish in that they serve primarily to allow the realisation of consonant contrasts, but F1 is not generally implicated in consonant contrasts, so it is possible to simultaneously realise vowel F1 contrasts if vowels are permitted to be somewhat longer than the Berber or Salish schwa.

4.1.3 Related phenomena Dispersion Theory predicts that, where no contrasts are primarily realised on a given dimension, then realisation on that dimension will be governed by minimisation of effort, or other contextual Markedness constraints. Neutralisation of F2 contrasts in vertical vowel inventories and in fully neutralising vowel reduction are examples of this phenomenon. There are probably many other examples of this pattern, but in some cases they can be difficult to detect because the least effort realisation of a sound type is similar to a sound found in contrast. For example, in many contexts, the least-effort laryngeal state for an obstruent will be voicelessness, due to aerodynamic factors discussed in the next section. However, voiceless stops also provide a distinct contrast with voiced stops, so least-effort stops may be similar to contrastively voiceless stops in many contexts. Dispersion theory leads us to expect that non-contrastive voiceless stops should be more prone to partial voicing following a preceding sonorant because effort minimisation disfavours active measures to promote voicelessness, but the differences involved can only be identified by instrumental analysis, so we do not have relevant data for many languages (but see Hsu 1998 for evidence of this pattern in Taiwanese). However, there is good evidence for the related prediction that effortful enhancements of stop voicing should only apply where there are voicing contrasts, as shown in the next section.

Contextual nasalisation of vowels provides another possible example of this type of pattern. It is a slightly more complex case because nasalisation does affect the distinctiveness of vowel-quality contrasts, particularly those involving

F1 (Wright 1986; Beddor 1993), but it obviously has a much greater effect on vowel nasalisation contrasts. So although we expect some general resistance to nasalisation of vowels, it is to be expected that oral vowels should be more tolerant of contextual nasalisation in the absence of nasalisation contrasts.

Again, differences in the magnitude and extent of partial nasalisation can only be determined by instrumental methods. Cohn (1990) shows that contrastive oral vowels in French undergo much less contextual nasalisation than English vowels preceding a nasal, but there is no obvious difference following a nasal. In any case, French may not be the most relevant example since the vowel nasalisation contrasts are generally accompanied by differences in vowel quality. There is evidence that the extreme measure of denasalisation of nasals to avoid contextual vowel nasalisation is only adopted where there are vowel-nasalisation contrasts.

The only way to ensure that a vowel adjacent to a nasal is completely oral is to execute the velum movement during the stop closure, resulting in a brief oral stop. This pattern is observed in a wide variety of languages (Anderson 1976; Herbert 1986), the most striking instance being Kaingang, where nasals are prenasalised preceding an oral vowel (28b), post-nasalised following an oral vowel (28c), and 'medio-nasalised' between oral vowels (28d).

(28)
a. $\tilde{V}m\tilde{V}$
b. $\tilde{V}m^{b}V$
c. $V^{b}m\tilde{V}$
d. $V^{b}m^{b}V$

Herbert (1986) claims that this pattern of realisation is only observed in languages with contrastive nasalisation, as one would expect if partial denasalisation is motivated by the pressure to maximise the distinctiveness of vowel nasalisation contrasts. That is, replacing a nasal by a more marked partially nasalised stop is only justified where it serves to maximise the distinctiveness of a contrast with nasalised vowels, because allophonic partial nasalisation does little damage to the distinctiveness of vowel-quality contrasts.

The schematic ranking in (29) shows the outlines of a Dispersion-Theoretic formulation of this analysis. The dispreference for partially nasalised stops universally outranks constraints against contrasts between partially nasalised vowel qualities (e.g. *ĩ-ẽ, where a single tilde indicates partial nasalisation), so allophonic partial nasalisation of vowels is always preferred to denasalisation of nasal consonants[10]. But the distinctiveness constraint against contrasts between partially and fully nasalised vowels *ṽ-$\tilde{\tilde{v}}$ (where a double tilde [$\tilde{\tilde{v}}$] marks a fully nasalised vowel) can outrank *Partially nasalised stop, so denasalisation can be conditioned by a contrastively oral vowel.

(29) *ṽ-ṽ̃, *Partially nasalised stop ≫ *ĩ-ẽ, *ẽ-ã, etc.

Without constraints on contrast, it is not possible to account for the fact that denasalisation requires vowel nasalisation contrasts, since any constraint that favoured denasalisation adjacent to oral vowels would necessarily apply to all oral vowels, whether or not they contrast with nasalised vowels.

There is one exception to Herbert's generalisation: in some Australian languages, including Gupapuyŋu (Butcher 1999), nasals are optionally pre-stopped postvocalically, although there are no vowel nasalisation contrasts. Butcher suggests that this partial denasalisation serves to ensure that the closure transitions are oral, avoiding the destructive effect of nasalisation on the distinctiveness of formant patterns (Repp and Svastikula 1988; Wright 1986). The distinctiveness of formant transitions is particularly important because the relevant languages distinguish four to six places of articulation among nasals. So this exceptional case also appears to be motivated by distinctiveness constraints.[11]

4.2 Enhancement of stop voicing contrasts

Another example of contrast-dependent markedness is provided by the typology of laryngeal contrasts among stops. A number of languages contrast prenasalised or implosive stops with voiceless unaspirated stops, but do not have plain voiced stops. The preference for prenasalised or implosive stops over plain voiced stops is explained on the grounds that prenasalised and implosive stops are more distinct than voiceless stops (cf. Iverson and Salmons 1996). However, these sounds are also more effortful than plain voiced stops, so most languages forgo these enhancements. Crucially, enhancement of stop voicing does not occur in the absence of contrast – we do not find prenasalisation or implosivisation of intervocalically voiced stops, for example. This is expected if the only reason for exerting the additional effort involved in producing these sounds is to satisfy a constraint on the distinctiveness of contrasts, but, like other contrast-dependent patterns of distribution, it is difficult to account for without constraints on contrast.

Prenasalised and implosive stops are often thought of as more marked than plain voiced stops. While it is true that they are less frequent than plain voiced stops, there is no implicational relationship between these sound types: a substantial number of languages have prenasalised or implosive stops without having plain voiced stops, for example San Juan Colorado Mixtec has prenasalised stops but no plain voiced stops (30) (Campbell, Peterson, and Lorenzo Cruz 1986). This pattern is discussed by Iverson and Salmons (1996) in relation to Mixtec, and by Herbert (1986: 16ff.), who cites a number of other examples, including Fijian, Lobaha, Reef Islands-Santa Cruz languages, and South Gomen. Other examples include Southern Barasano (Smith and Smith 1971) and Guaraní (Gregores and Suárez 1967).

(30) San Juan Colorado Mixtec stops

p	t	tʲ	k
ᵐb	ⁿd	ⁿdʲ	

Languages that contrast voiceless and implosive stops but lack plain voiced stops seem to be less common (Maddieson 1984: 28), but the UPSID database of phonological inventories (Maddieson 1984) includes two examples: Nyangi and Maasai (both Eastern Sudanic). The stops of Nyangi are shown in (32). In addition, Vietnamese voiced stops are often implosive (Nguyen 1970), and Ladefoged and Maddieson report that ‘fully voiced stops in many diverse languages (e.g. Maidu, Thai and Zulu) are often accompanied by downward movements of the larynx that make them slightly implosive’ (1996: 78).

(31) Nyangi stops:

p	t	c	k
ɓ	ɗ	ʄ	ɠ

Voiced stops are distinguished from voiceless stops by a variety of cues. One of the most important is Voice Onset Time (Lisker and Abramson 1964; Lisker 1975), but the presence of voicing during closure (indicated by periodicity and low-frequency energy) is also significant (Stevens and Blumstein 1981). Implosive and prenasalised stops are more strongly voiced than plain voiced stops, and so are better distinguished from voiceless stops in this respect. It is difficult to sustain high intensity of voicing during a stop closure because pressure builds up behind the closure until there is no longer a pressure drop across the glottis. Without a sufficient pressure drop there is no airflow through the glottis, and voicing ceases (Ohala 1983; Westbury and Keating 1986). So voicing tends to decline in intensity through a voiced stop closure. Lowering the velum allows air to be vented from the vocal tract, mitigating the pressure build-up, and thus facilitating the maintenance of high intensity of voicing. In addition, radiation from the nose results in higher intensity of the speech signal than radiation through the neck, which is the only source of sound in an oral stop (Stevens et al. 1986: 439).

Similarly, lowering the larynx during the stop closure, as in implosive stops, expands the oral cavity, reducing the build-up of pressure. Consequently, implosives are characteristically strongly voiced. Lindau (1984) found that the amplitude of voicing actually increases through the course of an implosive closure. Implosives also have very low-intensity release bursts because the intensity of the burst depends on oral pressure at release (Ladefoged and Maddieson 1996: 82). Intensity of the release burst has been shown to be a significant cue to stop-voicing contrasts in English (Repp 1979), so this is also likely to make implosives more distinct from voiceless stops than plain voiced stops.

Given these considerations, it seems likely that languages like Mixtec and Nyangi prefer prenasalised-voiceless and implosive-voiceless stop contrasts over the more common voiced-voiceless contrast because the former are more distinct contrasts (Henton, Ladefoged, and Maddieson 1992; Iverson and Salmons 1996). The conflicting constraint that leads many languages to forgo maximising distinctiveness is probably effort minimisation. Implosives involve more effort than plain voiced stops because they involve an additional larynx-lowering gesture. Prenasalised stops require rapid raising of the velum to produce oral and nasal phases within the same stop.

The analysis can be formalised as follows. We will assume a dimension corresponding to strength of voicing ([voice]) (32), which could be quantified in terms of the intensity of the periodic part of the speech signal. Voiceless unaspirated stops are also distinguished from the voiced stops by VOT, but this difference is not significant here because it is not enhanced by prenasalisation or implosion.

(32) Voice:

0	1	2
t	d	ⁿd, ɗ

For present purposes the fact that prenasalised stops and implosives involved greater effort than plain voiced stops will be implemented as a fixed ranking of constraints against these sound types (33).

(33) *IMPLOSIVE, *PRENASALISED STOP ≫ *VOICED STOP

Then a language like Nyangi, with implosives in place of voiced stops, is derived by the following ranking, as shown in (35).

(34) MINDIST = VOICE:1 ≫ MINDIST = VOICE:2, MAXIMISE CONTRASTS, *PRENASALISED STOP ≫ *IMPLOSIVE ≫ *VOICED STOP

(35)

		MINDIST = VOICE:1	MINDIST = VOICE:2	MAXIMISE CONTRASTS	*PRENASAL-ISED STOP	*IMPLOSIVE	*VOICED STOP
a.	t-d		*!	✓✓			*
b.	☞ t-ɗ			✓✓		*	
c.	t-ⁿd			✓✓	*!		
d.	t			✓!			

We will assume for now that the preference for implosives over prenasalised stops depends purely on the relative ranking of the effort-minimisation constraints against these sound types, so the ranking in (34) derives implosives where *PRENASALISED STOP ≫ *IMPLOSIVE (cf. 35c), while prenasalised stops are derived if this ranking is reversed. The more common voiced-voiceless contrast is derived if MINDIST = VOICE 2 is ranked below both of these effort-minimisation constraints (36).

(36)

		MINDIST = VOICE:1	MAXIMISE CONTRASTS	*PRENASAL-ISED STOP	*IMPLOSIVE	MINDIST = VOICE:2	*VOICED STOP
a.	☞ t-d		✓✓			*	*
b.	t-ɗ		✓✓		*!		
c.	t-ⁿd		✓✓	*!			
d.	t		✓!				

If a voicing contrast is not maintained, the distinctiveness of voicing contrasts is irrelevant, so voicing of stops is determined primarily by effort minimisation. In many contexts, effort minimisation prefers devoicing of stops due to aerodynamic factors reviewed above, but in some contexts, for example following a nasal or in short stops between vowels, voicing appears to be easier to produce and many languages follow effort minimisation, resulting in allophonically voiced stops in these contexts (Westbury and Keating 1986; Kirchner 1998). For example, stops are voiced intervocalically and following nasals in Tümpisa Shoshone (Dayley 1989; Kirchner 1998).[12] Implosives and prenasalised stops, on the other hand, are never preferred by effort-minimisation constraints, so these sounds are only expected in contrast with voiceless stops.

The patterns of distribution analysed here involve a contrast-dependent generalisation: implosives and prenasalised stops can be preferred to voiced stops where they contrast with voiceless stops, but they are never preferred to voiced stops where there is no voicing contrast. That is, there is no post-nasal implosivisation or intervocalic prenasalisation. This situation is difficult to account for without constraints on contrasts because any simple way of deriving implosives/prenasalised stops in place of voiced stops without these constraints is liable to predict that these sounds could also be preferred in the absence of contrast.

In a theory without constraints on contrasts, a preference for implosives over voiced stops implies a ranking of constraints with the effect of that shown in (37). The exact formulation of *VOICED STOP and *IMPLOSIVE is not important, it is only necessary that one favours implosive stops over voiced stops, and the other effectively imposes the reverse preference. We must also assume that faithfulness to the feature that differentiates implosives from plain voiced stops (e.g. [lowered larynx]) is low ranked throughout to explain the absence of contrasts between plain voiced and implosive stops in the relevant languages.

(37) IDENT[VOICE], *VOICED STOP ≫ *IMPLOSIVE

The reverse ranking of *VOICED STOP and *IMPLOSIVE would also have to be allowed to derive the usual voiced-voiceless contrast:

(38) IDENT[VOICE], *IMPLOSIVE ≫ *VOICED STOP

The problem arises when these ranking possibilities are combined with rankings required to analyse allophonic variation in languages without voicing contrasts. The basic ranking for a language without stop voicing contrasts has to place IDENT[VOICE] below the effort-minimisation constraints:

(39) *IMPLOSIVE ≫ *VOICED STOP ≫ IDENT[VOICE]

To derive intervocalic voicing, it is necessary to differentiate the markedness of voiced stops between vowels from their markedness in other contexts. A simple approach is to posit a constraint against intervocalic voiceless stops, *VOICELESS STOP/V__V, ranked above the general constraint against voiced stops (40). But we have already seen that *VOICED STOP must be able to outrank *IMPLOSIVE to account for languages with implosives but no plain voiced stops. So nothing prevents reversing the ranking of these constraints, as in (40), which derives the unattested phenomenon of intervocalic voicing implosivisation, that is, stops are implosive between vowels (41), but voiceless elsewhere (42), because this ranking makes it preferable to replace any voiced stop by an implosive.

(40) *VOICELESS STOP/V_V, *IMPLOSIVE ≫ *VOICED STOP ≫ IDENT[VOICE]

(41)

	/ata/	*VOICELESS STOP/V_V	*VOICED STOP	*IMPLOSIVE	IDENT [VOICE]
a.	ata	*!			
b.	ada		*!		*
c.	☞ aɗa			*	*

(42)

	/ad/	*VOICELESS STOP/V_V	*VOICED STOP	*IMPLOSIVE	IDENT [VOICE]
a.	☞ at				*
b.	ad		*!		
c.	aɗ			*!	

The problem with this approach is that it is not possible to express the fact that implosives and prenasalised stops are only favoured because they yield more distinct contrasts with voiceless stops (or an additional contrast), so nothing favours implosives in the absence of stop-voicing contrasts. Without constraints on contrasts, it is necessary to posit constraints favouring implosives and prenasalised stops independent of contrast, which then predicts that these sounds could be preferred over plain voiced stops in the absence of contrast. The preference for implosives and prenasalised stops must be strictly dependent on the presence of a contrast, which implies constraints on contrasts.

4.3 Allophonic and contrastive nasalisation

The pattern observed in the previous section could be characterised as showing that allophonic stop voicing is not subject to enhancement, whereas contrastive voicing can be enhanced, so it provides evidence that allophonic stop voicing behaves differently from contrastive stop voicing. Comparing the behaviour of allophonic and contrastive instances of a sound type is a good way to investigate the prediction that the markedness of a sound depends on the sounds that it contrasts with, because the same sound types can be observed in different systems of contrasts. For example, an allophonically nasalised vowel generally contrasts with other nasalised vowels, but by definition it does not minimally contrast with its oral counterpart, whereas a contrastively nasalised vowel minimally contrasts both with other nasal vowels and with its oral counterpart. In this section we will see that the markedness of an allophonically nasalised sound can differ from the markedness of the same sound where nasalisation is minimally contrastive. This difference is unexpected if there are no constraints on contrasts because then a constraint against nasalised vowels, for example, applies equally to allophonically and contrastively nasalised vowels.

In Dispersion Theory, a contextually nasalised vowel will generally violate more MINDIST constraints than its oral equivalent because nasalisation reduces the distinctiveness of vowel F1 contrasts (section 4.1.3). The markedness of a contrastively nasalised vowel depends on this factor also, but it is more dependent on the distinctiveness of the contrast with its oral counterpart. Evidence for the significance of this distinction comes from the existence of mismatches between the typologies of nasalisation contrasts and resistance to allophonic nasalisation in nasal harmony (Ní Chiosáin and Padgett 1997).

Blocking of nasal harmony can be illustrated from the Johore dialect of Malay (Onn 1980). In this language, nasality is only contrastive on stops, and spreads rightward from a nasal stop onto a sequence of vowels, glides and laryngeals (43a). All other segment types block the spread of nasalisation (43b).

(43)	a.	mĩnõm	'to drink'	baŋõn	'to rise'
		mãʔ̃ãp	'to pardon'	pənə̃ŋãh̃ãn	'central focus'
		mãj̃ãn	'stalk (palm)'	mə̃nãw̃ãn	'to capture (active)'
	b.	mə̃ratappi	'to cause to cry'	pəŋãw̃ãsan	'supervision'
		mãkan	'to eat'		

Languages with nasal harmony vary as to which segments block the spread of nasality. Building on surveys by Schourup (1972), Piggott (1992) and Cohn (1993), Walker (1998) shows that segments can be arranged into the hierarchy shown in (44) such that if any segment type blocks nasal harmony, all segment types lower on the hierarchy block nasal harmony as well. So in Johore Malay the top two levels of the hierarchy undergo nasalisation – liquids and segments lower on the hierarchy block harmony. Sundanese (Robins 1957; Cohn 1990) is similar to Malay, but glides also block nasalisation.

(44) vowels, laryngeals > glides > liquids > fricatives > obstruent stops

Walker (1998) analyses this generalisation in terms of a corresponding hierarchy of constraints on the 'compatibility' of different segment types with nasality:

(45) *NasObsStop ≫ *NasFricative ≫ *NasLiquid ≫ *NasGlide ≫ *NasVowel ≫ *NasSonStop

Walker does not propose any constraint on nasalised laryngeals, but remarks that they typically pattern with vowels (1998: 50), so *NasLaryngeal should presumably be ranked at the same level as *NasVowel in the hierarchy. Different patterns of blocking are then derived by ranking the constraint that motivates nasal harmony, Spread[+nasal], at different points in this hierarchy. Segment types subject to nasality constraints ranked below Spread[+nasal] undergo nasalisation harmony, while those subject to higher-ranking constraints block harmony. For example, the ranking for Johore Malay is as in (46).

(46) *NasObsStop ≫ *NasFricative ≫ *NasLiquid ≫ **Spread[+nasal]** ≫ *NasGlide ≫ *NasVowel ≫ *NasSonStop

The difficulty faced by this approach is that these same Markedness constraints should also account for the typology of nasalisation contrasts – placing Ident[nasal] at different points in the hierarchy should yield the typology of nasalisation contrasts. As Walker observes (1998: 53), the predicted pattern is broadly correct: the most common nasal sounds are nasal stops, and the next most common are nasalised vowels, while other contrastively nasalised sounds are rare (Maddieson 1984; Cohn 1993). However, Ní Chiosáin and Padgett (1997) note that the predictions concerning nasalised laryngeals are problematic. Nasalisation is never contrastive on glottal stops, suggesting that *NASLARYNGEAL should be high ranked, but on the other hand laryngeals are among the sounds most susceptible to nasal harmony, implying that *NASLARYNGEAL should be low ranked. That is, contrastive nasalisation of laryngeals is very marked, but non-contrastive nasalisation is unproblematic. This situation is predicted by the contrast-based analysis of blocking and contrast: nasalisation contrasts on glottal stops are unsatisfactory because nasalised and oral glottals are acoustically identical (cf. Walker and Pullum 1999), but by the same token, contrasts between glottal stops and other consonants are unaffected by contextual nasalisation of the glottal. The acoustics of [h] are also relatively unaffected by velum lowering (Ohala 1975: 301). The problem with Walker's approach is that it conflates markedness of contrastive and non-contrastive nasalisation.

A faithfulness-based analysis according to which nasal harmony is blocked because it violates faithfulness to the underlying [−nasal] specification of the blocking segment fares no better. Such an account must distinguish IDENT[NASAL] constraints for different segment types, for example IDENT[NASAL]/FRICATIVE, IDENT[NASAL]/LIQUID, and so on. The blocking hierarchy is then derived by imposing a universal ranking on these faithfulness constraints:

(47) IDENT[NAS]/OBSSTOP ≫ IDENT[NAS]/FRICATIVE ≫
IDENT[NAS]/LIQUID ≫ IDENT[NAS]/GLIDE ≫
IDENT[NAS]/VOWEL, IDENT[NAS]/LARYNGEAL

Again the problem is that these faithfulness constraints are required to do double duty: they must also account for the typology of nasalisation contrasts. Nasality contrasts should be derived by the ranking of a general Markedness constraint, presumably *[+NASAL]. However, this makes completely inaccurate predictions concerning the typology of nasal contrasts. For example, the high ranking of IDENT[NAS]/FRICATIVE motivated by the resistance of fricatives to nasalisation implies that nasalisation contrasts on fricatives should be common also. At the other extreme, the susceptibility of vowels to contextual nasalisation should imply that nasalisation contrasts on vowels are more marked than nasalisation contrasts on liquids or glides.[13] Again, the diagnosis is that this approach fails to distinguish markedness of contrastive nasalisation from markedness of non-contrastive, contextual nasalisation.

Dispersion Theory predicts that nasal harmony should be blocked where nasalisation would be articulatorily difficult, or where it would give rise to indistinct contrasts. Ní Chiosáin and Padgett propose a dispersion-based analysis of the behaviour of laryngeals according to which nasalisation contrasts do not arise on laryngeals because they would be indistinct, as above, but laryngeals do not block nasal harmony because a nasalised laryngeal is a low-effort sound, implying an account of blocking in terms of articulatory compatibility with nasalisation. However, there is no obvious articulatory difficulty in lowering the velum during any sound type (although it may result in a change in manner as well as nasalisation). So the blocking hierarchy is more plausibly derived from distinctiveness constraints.[14] Specifically, most of the blocking hierarchy ((47), above) can be derived from the generalisation that nasal harmony is blocked where nasalisation would endanger contrasts with nasal stops. That is, blocking of nasal harmony is a consequence of MINDIST constraints blocking the creation of indistinct contrasts between nasals and nasalised consonants. The ranking of these constraints is shown in (48). Distances are expressed descriptively because it is not clear what dimensions distinguish these sounds.[15]

(48) MINDIST = NAS'FRICATIVE'-NASSTOP ≫ MINDIST = NASLIQUID-NASSTOP ≫ MINDIST = NASGLIDE-NASSTOP ≫ MINDIST = NASVOWEL-NASSTOP ≫ MINDIST = NASLARYNGEAL-NASSTOP

Since laryngeals are largely unaffected by nasalisation, they remain highly distinct from voiced nasals. So MINDIST = NASLARYNGEAL-NASSTOP is at the bottom of the hierarchy, and laryngeals are consequently least likely to block nasal harmony. As for the rest of the hierarchy, nasalising a voiced stop actually results in a nasal, and so would neutralise contrasts. Approximant consonants such as glides, laterals, and rhotics are already similar to nasals in that these are all sonorant consonants, and lowering the velum further reduces this difference. In general, the narrower the oral constriction of a nasalised approximant, the more similar it will be to a nasal stop. In a nasal stop, all airflow is through the nasal cavity, whereas in a nasalised approximant there is airflow through both oral and nasal cavities, but a narrower oral constriction results in less airflow through the oral cavity, and correspondingly more airflow through the nasal cavity, resulting in closer approximation to a nasal stop. Thus nasalised laterals, for example [l̃], and flaps [ɾ̃] are most similar to nasal stops, nasalised glides, [w̃, j̃] somewhat less so, and nasalised vowels least of all. Nasalising a voiced fricative is liable to lead to loss of frication because air is vented through the nose, leaving insufficient oral pressure to generate frication (Ohala and Ohala 1993: 227f.), so the basic result is a narrowly constricted nasalised approximant, which presumably is slightly closer to a nasal than a nasalised liquid (in (48), NAS'FRICATIVE' refers to the sound that results from lowering the velum during a voiced fricative).

So nasalising sounds that are higher on the blocking hierarchy results in less distinct contrasts with nasal stops, that is, violation of higher-ranked MINDIST constraints. The blocking segments in a particular nasal harmony system depend on the position of SPREAD[+NASAL] in the hierarchy of MINDIST constraints. For example, in Johore Malay, MINDIST = NASGLIDE-NASSTOP ≫ SPREAD[+NASAL]. A contrast like [j-ɲ] satisfies this MINDIST constraint, so nasality may spread onto glides (49). On the other hand a contrast involving a nasalised liquid, for example [r̃-n], violates this MINDIST constraint, so liquids block the spread of nasality (50).

(49)

		MAXIMISE CONTRASTS	MINDIST = NASGLIDE-NASSTOP	SPREAD [+NASAL]
a.	majaŋ-mãɲãŋ	✓✓		*****!
b.	mãjaŋ-mãɲãŋ	✓✓		****!
c.	☞ mãj̃ãŋ-mãɲãŋ	✓✓		**
d.	mãɲãŋ	✓!		*

(50)

		MAXIMISE CONTRASTS	MINDIST = NASGLIDE-NASSTOP	SPREAD [+NASAL]
a.	məɾa-mə̃nã	✓✓		***!
b.	☞ mə̃ɾa-mə̃nã	✓✓		**
c.	mə̃ɾ̃ã̤-mə̃nã	✓✓	*!	
d.	mə̃na	✓!		*

Blocking by voiceless consonants has a different basis. Nasalising a voiceless stop results in a voiceless nasal, which is quite distinct from a voiced nasal, but is highly dispreferred for other reasons. A fully devoiced nasal is similar to [h] and also yields indistinct place contrasts since noise is generated mainly at the nostrils, and so is the same regardless of oral place of articulation (Ohala and Ohala 1993: 232). Note that contrastively voiceless nasals are actually voiced during part of the nasal closure (Ladefoged and Maddieson 1996: 113) – Ohala and Ohala suggest that this realisation is adopted precisely to diminish the problems just outlined. Nasalising a voiceless fricative would yield similar results, since frication would be lost, as in nasalisation of a voiced fricative.

The typology of contrastive nasalisation depends partly on the distinctiveness facts just outlined, but many other distinctiveness relations are significant, in particular the distinctiveness of contrasts between corresponding oral and nasal sounds, which is irrelevant where nasalisation is non-contrastive. These additional considerations explain the discrepancies between the typologies of blocking and contrast.

As already noted, the acoustic near-identity between oral and nasalised laryngeals explains why these sounds are not found in contrast.[16] Nasal stops provide good contrasts with oral stops and approximants, and consequently are found in almost all languages (Ferguson 1963; Maddieson 1984). This is probably because they are well differentiated from obstruents by intensity, and from other sonorants by their distinctive formant structure. Nasals have more closely spaced formants than oral sonorants because the nasal-pharyngeal tract is longer than the oral tract. Nasal formant structure is further differentiated from most oral sonorants by the presence of spectral zeroes at the resonant frequencies of the oral cavity behind the closure (Fujimura 1963; Stevens 1999: 487ff.) (laterals also have spectral zeroes, but at higher frequencies than in nasals). Given the desirability of nasal stops, the undesirability of nasalised approximants follows from the fact that nasalised approximants yield poor contrasts with nasal stops for reasons outlined above in the discussion of the MINDIST hierarchy in (48).

5 Conclusion: working with constraints on contrast

We have now seen substantial evidence that phonology includes constraints on contrasts, specifically constraints that favour maximising the distinctiveness of contrasts (MINDIST), and a constraint that favours maximising the number of contrasts (MAXIMISE CONTRASTS). We have also seen that these constraints do not operate independently from more familiar syntagmatic Markedness constraints, for example as a theory of inventories, somehow operating outside of conventional phonological analyses. The interaction between syntagmatic and paradigmatic constraints is central to the derivation of basic phenomena such as neutralisation (section 3.2.2) and blocking in harmony processes (section 4.3). According to Dispersion Theory, the set of well-formed words in a language represents an optimal balance between the number and distinctiveness of the contrasts between words and constraints that define preferred sound sequences, such as effort minimisation and metrical constraints. However, combining paradigmatic and syntagmatic constraints in this way does result in a system with very different properties from an OT grammar based on conventional constraints, because constraints on the distinctiveness of contrasts evaluate relationships between forms. So if we want to determine whether a putative word is well formed, we must consider whether it is sufficiently distinct from neighbouring words. But these words must also be well formed, which implies assessing their distinctiveness from neighbouring words, and so on. Thus it seems that we cannot evaluate the well-formedness of a single word without determining the set of all possible words.

The analyses above avoid this problem by considering only the evaluation of inventories of contrasting sounds (or short strings of sounds) in a particular context rather than evaluating complete words. For example, evaluating vowel inventories effectively involves determining the set of contrasting sounds that are permitted in a syllable nucleus. This makes the evaluation of MINDIST and MAXIMISE CONTRASTS straightforward, since only a small number of contrasting sounds are possible in a given context. This simplification is valid given certain assumptions. First, the context must be well formed. For example, if we are evaluating the set of vowels that can appear before a nasal stop, it must be true that nasal stops are part of an inventory of consonant contrasts that can occur in postvocalic position. Second, nothing outside of the specified context should be relevant – that is, no constraint that is ranked high enough to affect the well-formedness of the inventory should refer to material outside of the specified context.

More generally, the strategy for avoiding the problem of mass comparisons is to derive generalisations about the set of possible words in a language – for example, stressed vowels are all drawn from a certain set – rather than deriving particular words. But this strategy is not actually novel, it is the usual approach to phonological analysis. Even if it is possible to determine whether an individual word is well formed with respect to a constraint ranking, the result of such an exercise is usually not very significant. Showing that a grammar can derive an individual word is not usually the goal of phonological analysis of a language; the goal is to devise a grammar that derives all and only the possible words of that language. The usual intermediate goal is to derive generalisations about all the possible words of the language, exactly as in the analyses here.

For example, in analysing a language it is usual to restrict attention to a single process, for example place assimilation between nasals and stops, ignoring stress assignment, distribution of vowels, and so on. Such an analysis may be illustrated by deriving complete words, for example /kanpa/ → [kampa], but in itself this is uninteresting. The real goal is to derive the generalisation that nasals are always homorganic to following stops. Properly, establishing such a generalisation requires showing that no contrary output is derived if all possible inputs are passed through the grammar (Prince and Smolensky 1993: 91). So with or without paradigmatic constraints, there is an important distinction between deriving individual words using a grammar and reasoning about the properties of the set of words derived by that grammar. Constraints on contrast make complete derivation of individual words difficult, but that does not preclude deriving generalisations about possible words.

To approach the derivation of complete words, it is necessary to derive increasingly comprehensive descriptions of the set of possible words. Such a description need not be a list of possible words, it could be a grammar that generates the possible words. That is, one way to deal with the need to evaluate all words simultaneously could be to evaluate candidate grammars that provide compact characterisations of candidate sets of possible words. Any such solution involves substantial additions to the analytical machinery of phonology, but we have seen that these steps are well-motivated.

References

Albano Leoni, F., M. R. Caputo, L. Cerrato, F. Cutugno, P. Maturi, and R. Savy (1995). Il vocalismo dell'Italiano. Analisi di un campione televisivo. *Studi Italiani di Linguistica Teorica e Applicata* 24: 405–11.

Anderson, Stephen R. (1976). Nasality and the internal structure of segments. *Lg* 52: 326–44.

Beddor, Patrice S. (1993). The perception of nasal vowels. In Marie K. Huffman and Rena A. Krakow (eds.), *Nasals, nasalization, and the velum.* (Phonetics and Phonology 5) Academic Press: San Diego, 171–96.

Bender, Byron W. (1968). Marshallese phonology. *Oceanic Linguistics* 7: 16–35.

Benua, Laura (1997). *Transderivational identity*. PhD dissertation, University of Massachusetts, Amherst.

Boersma, Paul (1998). *Functional phonology: formalizing the interactions between articulatory and perceptual drives.*. The Hague: Holland Academic Graphics.

Booij, Geert (1995). *The phonology of Dutch*. Oxford: Oxford University Press.

Burgess, Eunice and Patricia Ham (1968). Multilevel conditioning of phoneme variants in Apinayé. *Linguistics* 41: 5–18.

Burzio, Luigi (1998). Multiple correspondence. *Lingua* 104: 79–109.

Butcher, Andrew (1999). What speakers of Australian aboriginal languages do with their velums and why: The phonetics of the nasal/oral contrast. *Proceedings of the International Congress of Phonetic Sciences 1999*: 479–82.

Byrd, Dani (1992). Marshallese suffixal reduplication. *WCCFL* 11: 61–77.

Campbell, Sara Stark, Andrea Johnson Peterson, and Filiberto Lorenzo Cruz (1986). Diccionario Mixteco de San Juan Colorado. Mexico City: Instituto Lingüístico de Verano.

Choi, John D. (1991). An acoustic study of Kabardian vowels. *Journal of the International Phonetic Association* 21: 4–12.

(1992). *Phonetic underspecification and target interpolation: An acoustic study of Marshallese vowel allophony.* (UCLA Working Papers in Phonetics 82) PhD dissertation, University of California, Los Angeles.

Chomsky, Noam and Morris Halle (1968). *The sound pattern of English*. New York: Harper and Row.

Cohn, Abigail C. (1990). *Phonetic and phonological rules of nasalization.* (UCLA Working Papers in Phonetics, 76) PhD dissertation, University of California, Los Angeles.

(1993). A survey of the phonology of the feature [nasal]. *Working Papers of the Cornell Phonetics Laboratory* 8: 141–203.

Colarusso, John (1988). *The Northwest Caucasian languages: A phonological survey*. New York: Garland Publishing.

(1992). *A grammar of the Kabardian language*. Calgary: University of Calgary Press.

Crosswhite, Katherine (2001). *Vowel reduction in Optimality Theory*. New York: Routledge.

Dayley, Jon P. (1989). *Tümpisa (Panamint) Shoshone grammar*. Berkeley: University of California Press.

Delattre, Pierre C., Alvin M. Liberman, Franklin S. Cooper, and Louis J. Gerstman (1952). An experimental study of the acoustic determinants of vowel color: Observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word* 8: 195–210.

Dell, François and Mohammed Elmedlaoui (1996). On consonant releases in Imdlawn Tashlhiyt Berber. *Linguistics* 34: 357–95.

Edwards, Jan, Mary E. Beckman, and Janet Fletcher (1991). The articulatory kinematics of final lengthening, *JASA* 89: 369–82.

Fails, Willis C. and J. Halvor Clegg (1992). A spectrographic analysis of Portuguese stressed and unstressed vowels. In Donald P. Macedo and Dale A. Koike (eds.), *Romance linguistics: The Portuguese context*. Westport, CT: Bergin and Garvey, 31–42.

Ferguson, Charles A. (1963). Assumptions about nasals: A sample study in phonological universals. In Joseph Greenberg (ed.), *Universals of language*. Cambridge, MA: MIT Press, 53–60.

Flemming, Edward (1995). *Auditory representations in phonology*. PhD dissertation, University of California, Los Angeles.

(1996). Evidence for constraints on contrast: The dispersion theory of contrast. In Chai-Shune K. Hsu (ed.), *UCLA Working Papers in Phonology* 1: 86–106.

(2001). *Auditory representations in phonology*. New York: Garland Press. [Revised version of Flemming 1995].

Flemming, Edward, Peter Ladefoged, and Sarah G. Thomason (1994). The phonetic structures of Montana Salish. *UCLA Working Papers in Phonetics* 87: 1–34.

Fujimura, Osamu (1962). Analysis of nasal consonants. *JASA* 34: 1865–75.

Gregores, Emma and Jorge A. Suárez (1967). *A description of colloquial Guaraní*. The Hague: Mouton.

Halle, Morris (1959). *The sound pattern of Russian*. The Hague: Mouton.

Hayes, Bruce (1995). *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Henton, Caroline, Peter Ladefoged, and Ian Maddieson (1992). Stops in the world's languages. *Phonetica* 49: 65–101.

Herbert, Robert K. (1986). *Language universals, markedness theory and natural phonetic processes*. Berlin: Mouton de Gruyter.

Hsu, Chai-Shune K. (1998). Voicing underspecification in Taiwanese word-final consonants. *UCLA Working Papers in Phonetics* 96: 90–105.

(1998). Voicing underspecification in Taiwanese word-final consonants. *UCLA Working Papers in Phonetics* 96: 90–105.

Hume, Elizabeth and Keith Johnson (2001). A model of the interplay of speech perception and phonology. In Elizabeth Hume and Keith Johnson (eds.), *The role of speech perception in phonology*. San Diego: Academic Press, 3–26.

Iverson, Gregory K. and Joseph C. Salmons (1996). Mixtec prenasalization as hypervoicing. *IJAL* 62: 165–75.

Keating, Patricia A. (1985). Universal phonetics and the organization of grammars. In Victoria A. Fromkin (ed.) *Phonetic linguistics*. New York: Academic Press, 115–32.

Kenstowicz, Michael (1996). Uniform exponence and base identity. In J. Durand and B. Laks (eds.), *Current trends in phonology*. Paris: CNRS; Salford: University of Salford Publications, 363–93.

Kirchner, Robert (1997). Contrastiveness and faithfulness. *Phonology* 14: 83–111.

(1998). *An effort-based approach to consonant lenition*. PhD dissertation, University of California, Los Angeles. Rutgers Optimality Archive 276, http://roa.rutgers.edu.

Kondo, Yuko (1994). Targetless schwa: Is that how we get the impression of stress timing in English? *Proceedings of the Edinburgh Linguistics Department Conference '94*: 63–76.

(2000). Production of schwa by Japanese speakers of English: An acoustic study of shifts in coarticulatory strategies from L1 to L2. In Michael Broe and Janet Pierrehumbert (eds.), *Papers in laboratory phonology 5: Acquisition and the lexicon*. Cambridge: Cambridge University Press, 29–39.

Kuipers, Aert H. (1960). *Phoneme and morpheme in Kabardian (Eastern Adyghe)*. (Janua Linguarum, series minor, no. 8) The Hague: Mouton.

Ladefoged, Peter and Ian Maddieson (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Laycock, D. C. (1965). *The Ndu language family*. (Linguistics Circle of Canberra, Series C, no.1) Canberra: Australian National University.

Lehiste, Ilse (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Liljencrants, Johan and Björn Lindblom (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Lg* 48: 839–62.

Lindau, Mona (1984). Phonetic differences in glottalic consonants. *JPhon* 12: 147–55.

Lindblom, Björn (1963). Spectrographic study of vowel reduction. *JASA* 35: 1773–81.

(1986). Phonetic universals in vowel systems. J. J. Ohala and J. J. Jaeger (eds.), *Experimental phonology*. Orlando, FL: Academic Press.

(1990). Phonetic content in phonology. *PERILUS* 11: 101–18.

Lisker, Leigh (1975). Is it VOT or a first formant detector? *JASA* 57: 1547–51.

Lisker, Leigh and Arthur S. Abramson (1964). A cross-language study of voicing in initial stops. *Word* 20: 384–422.

Maddieson, Ian (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Maeda, Shinji (1993). Acoustics of vowel nasalization and articulatory shifts in French nasal vowels. In Marie K. Huffman and Rena A. Krakow (eds.), *Nasals, nasalization, and the velum.* (Phonetics and Phonology 5) San Diego: Academic Press, 147–70.

Maiden, Martin (1995). Vowel systems. In Martin Maiden and Mair Parry (eds.), *The dialects of Italy*. London: Routledge, 7–14.

Major, Roy C. (1992). Stress and rhythm in Brazilian Portuguese. In Donald P. Macedo and Dale A. Koike (eds.), *Romance linguistics: The Portuguese context*. Westport, CI: Bergin and Garvey, 3–30.

Martinet, Andre (1952). Function, structure, and sound change. *Word* 8: 1–32.

(1955). *Economie des changements phonétiques*. Berne: Francke.

Mattoso Camara, Joaquim (1972). *The Portuguese language*. Chicago: University of Chicago Press.

Mazzola, Michael L. (1976). *Proto-Romance and Sicilian*. Lisse: Peter de Ridder.

Moraes, João A. de (1998). Brazilian Portuguese. In Daniel Hirst and Albert Di Cristo (eds.), *Intonation systems*. Cambridge: Cambridge University Press.

Nelson, W. L. (1983). Physical principles for economies of skilled movement. *Biological Cybernetics* 46: 135–47.

Nguyen, Dang Liem (1970). *A contrastive phonological analysis of English and Vietnamese*, Vol. 4. (Pacific Linguistics Series C, No. 8) Canberra: Australian National University.

Ní Chiosáin, Máire and Jaye Padgett (1997). Markedness, segment realization, and locality in spreading. Report no. LRC-97-01, Linguistics Research Center, University of California, Santa Cruz.

Nosofsky, Robert M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology* 43: 25–53.

Ohala, John J. (1975). Phonetic explanations for nasal sound patterns. In Charles A. Ferguson, Larry M. Hyman, and John J. Ohala (eds.), *Nasálfest: Papers from a symposium on nasals and nasalization*. Stanford: Stanford University, 289–316.

(1983). The origins of sound patterns in vocal tract constraints. In Peter F. MacNeilage (ed.), *The production of speech*. New York: Springer Verlag, 189–216.

Ohala, John J. and Manjari Ohala (1993). The phonetics of nasal phonology: Theorems and data. In Marie K. Huffman and Rena A. Krakow (eds.), *Nasals, nasalization, and the velum*. San Diego: Academic Press, 225–50.

Onn, Farid M. (1980). *Aspects of Malay phonology and morphology*. PhD dissertation, Universiti Kebangsaan Malaysia, Bangi.

Perkell, Joseph S. (1997). Articulatory processes. In William J. Hardcastle and John Laver (eds.), *The handbook of phonetic sciences*. Oxford: Blackwell, 333–70.

Piggott, Glynne (1992). Variability in feature dependency: The case of nasality. *NLLT* 10: 33–78.

Plomp, Reinier (1975). Auditory analysis and timbre perception. In Gunnar Fant and Michel Tatham (eds.), *Auditory analysis and perception of speech*. New York: Academic Press, 7–22.

Prince, Alan and Paul Smolensky (1993). *Optimality Theory: Constraint interaction in generative grammar*. Rutgers Optimality Archive 537, http://roa.rutgers.edu/.

Repp, Bruno H. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech* 27: 173–89.

Repp, Bruno H. and K. Svastikula (1988). Perception of the [m]-[n] distinction in VC syllables. *JASA* 83: 238–47.

Robins, R. H. (1957). Vowel nasality in Sundanese. In *Studies in Linguistic Analysis*. London: Blackwell, 87–103.

Rosner, B. S. and J. B. Pickering (1994). *Vowel perception and production*. Oxford: Oxford University Press.

Schourup, Lawrence (1972). Characteristics of vowel nasalization. *Papers in Linguistics* 5: 530–48.

Schwartz, Jean-Luc, Denis Beautemps, Christian Abry, and Pierre Escudier (1993). Inter-individual and cross-linguistic strategies for the production of the [i] *vs.* [y] contrast. *JPhon* 21: 411–25.

Shepard, Roger N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika* 22: 325–45.

(1972). Psychological representation of speech sounds. In Edward David and Peter Denes (eds.), *Human communication: A unified view*. New York: McGraw-Hill, 67–113.

Smeets, Rieks (1984). *Studies in West Circassian phonology and morphology*. Leiden: Hakuchi Press.

Smith, Richard and Connie Smith (1971). Southern Barasano phonemics. *Linguistics* 75: 80–5.

Staalsen, Philip (1966). The phonemes of Iatmul. *Papers in New Guinea Linguistics* 5 (Linguistics Circle of Canberra, Series A, No.7): 69–76.

Steriade, Donca (1993). Orality and markedness. *BLS* 19: 334–47.

(1995). Neutralization and the expression of contrast. Ms., University of California, Los Angeles.

(1997). Phonetics in phonology: the case of laryngeal neutralization. Ms., University of California, Los Angeles. [http://www.linguistics.ucla.edu/people/steriade/papers/phoneticsinphonology.pdf]

(2000). Paradigm uniformity and the phonetics-phonology boundary. In Michael Broe and Janet Pierrehumbert (eds.), *Papers in laboratory phonology 5: Acquisition and the lexicon*. Cambridge: Cambridge University Press, 313–34.

Stevens, Kenneth N. (1999). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Stevens, Kenneth N. and Sheila E. Blumstein (1981). The search for invariant acoustic correlates of phonetic features. In Peter D. Eimas and Joanne L. Miller (eds.), *Perspectives on the study of speech*. Hillsdale, NJ: Lawrence Erlbaum, 1–38.

Stevens, Kenneth N., S. Jay Keyser, and Haruko Kawasaki (1986). Toward a phonetic and phonological theory of redundant features. In Joseph S. Perkell and Dennis H. Klatt (eds.), *Invariance and variability in speech processes*. Hillsdale, NJ: Lawrence Erlbaum, 426–49.

Valente, V. (1975). Puglia. In Manlio Cortelazzo (ed.), *Profilo dei dialetti italiani,* Vol. 15. Pisa: Pacini.

Van Bergem, Dick R. (1994). A model of coarticulatory effects on the schwa. *Speech Communication* 14: 143–62.

Walker, Rachel (1998). *Nasalization, neutral segments, and opacity effects*. PhD dissertation, University of California, Santa Cruz.

Walker, Rachel and Geoffrey K. Pullum (1999). Possible and impossible segments. *Lg* 75: 764–80.

Westbury, John and Patricia A. Keating (1980). Central representation of vowel duration. *JASA* 67: S37A.

(1986). On the naturalness of stop consonant voicing. *JL* 22: 145–66.

Wright, James T. (1986). The behavior of nasalised vowels in the perceptual vowel space. In John J. Ohala and Jeri J. Jaeger (eds.), *Experimental phonology*. Orlando: Academic Press, 45–68.

Wright, Richard (1996). Consonant clusters and cue preservation in Tsou. PhD dissertation, University of California, Los Angeles.

Zipf, George K. (1949). *Human behavior and the principle of least effort*. Cambridge: Addison-Wesley.

Notes

I would like to thank the editors for detailed comments on this chapter.

1. Crosswhite (this volume) proposes that vowel raising is desirable in unstressed syllables because it lowers the sonority of the vowel, resulting in a better correspondence between stress and vowel intensity. This predicts that raising of unstressed vowels could occur in languages where stress is not realised by significant differences in vowel duration whereas the present analysis treats shortening as the primary cause of raising. She also suggests that the same vowel quality difference may be less discriminable if vowel duration is shorter, so larger quality differences are required in short, unstressed syllables, which can result in neutralisation in these positions without raising of low vowels. If vowel duration affects distinctiveness in this way, it could be formalised in dispersion theory by ranking MINDIST constraints on shorter vowels higher than comparable constraints on longer vowels.
2. Word-final unstressed syllables were more variable in duration, probably because duration in this position is dependent on phrase-final lengthening effects. F1 of final /a/ was correspondingly more variable. The greater duration of phrase-final vowels does not lead to a larger vowel inventory in this position – this is probably a 'uniformity' effect (Steriade 1997, 2000), i.e. it allows words to have a more consistent realisation across phrasal positions.
3. Moraes (1998) found relatively small differences in duration between pretonic and final unstressed vowels, but he only measured high vowels which tend to be short in any case. Major measured low vowels, which are more relevant here. Moraes (1998) also shows that the duration difference can be eliminated by phrase-final lengthening of final unstressed syllables. As in Italian, it appears to be the phrase-medial characteristics that are relevant to neutralising vowel reduction. It is also interesting to note that Moraes found that final unstressed vowels have much lower intensity than vowels in other positions, and that this remains true even with final-lengthening (this should not be a consequence of vowel raising, since all vowels were high). Intensity should play some role in the perceptibility of vowel contrasts, but this factor is not analysed here.
4. Kuipers actually transcribes the Kabardian high vowel as [ə], the mid-vowel as [a], and the 'long' low vowel as [ā], and Colarusso (1988) follows him in this, but their descriptions, Colarusso's phonetic transcriptions, and acoustic data in Choi (1991) all indicate that the vowels are actually high and mid respectively.
5. The transcription of vowels has been altered in accordance with conventions adopted here.
6. Van Bergem (1994) also concludes that Dutch schwa is the minimum-effort vowel.
7. Consonant F2 is influenced by adjacent full vowels, so the vowel environment also influences schwa quality. There are no data on the influence of vowel environment on vertical vowel quality.
8. The use of IPA [ə] to transcribe this vowel is thus a source of confusion since [ə] is supposed to be a mid central vowel. The fact that schwa is typically a high vowel with transitional F2 helps to explain the use of the [ə] symbol to transcribe high vertical vowels in Caucasian (e.g. Kuipers 1960; Smeets 1984).
9. Crosswhite (2001, this volume) proposes distinct markedness hierarchies for stressed and unstressed vowels where schwa is the most marked stressed vowel, but the least marked unstressed vowel. However, she uses [ə] to refer to both a

mid-central vowel, as found in Brazilian Portuguese vowel reduction, and the transitional vowel found in complete neutralisation, so this analysis fails to account for the distinct contexts in which these two types of vowel arise.

10. It is not clear whether *PARTIALLY NASALISED STOP is properly a constraint against the effort involved in moving the velum with sufficient rapidity and precision to produce a nasalisation contour, or whether partially nasalised stops are dispreferred relative to full nasals because they yield inferior contrasts with some ubiquitous sound category such as voiceless stops. However, it does seem that languages do not have partially nasalised stops unless they also have nasals, either in contrast or in alternation with the partially nasalised stops (Herbert 1986: 16ff.).
11. A form of post-nasalisation can also arise without vowel nasalisation contrasts through a process of 'pre-obstruentisation' discussed by Steriade (1993) (e.g. Diyari, Icelandic). Steriade argues that this process is not denasalisation per se because it is accompanied by pre-stopping of laterals in the same environments (l → dl). Further reason for doubting that it is the orality of vowels that conditions this process of post-nasalisation comes from the fact that they are not conditioned by all oral vowels – it only applies to post-stress or geminate nasals.
12. Not all languages that lack a voicing contrast have fully voiced stops between vowels. This might be due to variation in stop duration – that is, devoicing is only effortful in short stops – or effort minimisation might be opposed by a conflicting preference for voiceless stops because their place cues are more distinct and robust (Wright 1996, this volume).
13. Adopting the set of markedness constraints in (47) rather than undifferentiated *[+NASAL] avoids some of the problematic predictions of a pure faithfulness account at the cost of proliferating constraints, but it still fails to account for the behavior of laryngeals: although IDENT[NAS]/LARYNGEAL is ranked low, so is *NASLARYNGEAL, so it is predicted that contrastively nasalised laryngeals can be derived while excluding all other nasalisation contrasts.
14. Ní Chiosáin and Padgett (1997: fn.25) also consider this possibility.
15. Nasalised vowels and approximants raise potential difficulties for the formant-based approach to spectral quality adopted above because nasal coupling introduces additional formants (resonances of the nasal cavity), so formants of nasalised sounds do not correspond straightforwardly to the formants of oral sounds. In addition, nasal sounds generally include spectral zeroes whose primary effect is to reduce the intensity of nearby formants (Maeda 1993), and formant intensities are not dimensions we have considered so far.
16. Ladefoged and Maddieson (1996: 133) report the existence of contrastively nasalised [h̃] in Kwangali, but nasalisation is also realised on the following vowel, so it is possible to regard this as essentially a vowel nasalisation contrast that is restricted to environments following a consonant that is highly compatible with contextual nasalisation conditioned by a nasalised vowel.

9 Syllable weight*

Matthew Gordon

1 Introduction

The goal of this chapter is to explore the role phonetics plays in shaping the phonology of syllable weight. Standard (moraic and skeletal slot) treatments of weight assume that weight criteria (i.e. what syllables count as 'heavy') may vary from language to language, but that all phonological processes within a given language will employ a uniform weight criterion. An extensive survey of weight-sensitive phonological phenomena, however, shows the opposite: weight criteria are frequently non-uniform within a given language, but particular weight criteria are characteristic of particular classes of weight-sensitive phenomena, with considerable cross-linguistic uniformity for those phenomena. This chapter focuses on weight-sensitive stress and tone. I argue that the divergent weight criteria observed in stress and tone systems largely follow from differences in the phonetic implementation of stress and tone. Phonetics also plays a role in accounting for cross-linguistic variation in weight criteria for a given process: such variation is often attributed to independent phonetic properties of these languages, which are in turn grounded in other phonological properties such as syllable structure.

Despite the importance of phonetics, however, syllable weight is not sensitive only to phonetic considerations. I will claim that languages employ weight distinctions that operate over phonologically symmetrical classes of syllables, even if this means not exploiting the phonetically most effective weight distinction(s).

Finally, I will argue that the ingredients of process-specificity of weight criteria, phonetic effectiveness, and phonological simplicity together play a crucial role in the ranking of a set of Optimality-Theoretic constraints governing weight-sensitive phenomena. The divergent phonetic motivations behind different weight-sensitive processes are reflected in process-specific sets of constraints. Within individual phenomena, constraints referring to the phonetically most effective of the simple weight distinctions are ranked on a language-specific basis above constraints referring to phonetically less effective or more complex distinctions.

The general structure of this chapter is as follows. Following an introduction to the concept of weight in section 2, section 3 summarises results of the cross-linguistic weight survey, arguing against language-internal uniformity of weight criteria and for the process-specificity of weight. Sections 4 and 5 juxtapose weight-sensitive tone and stress, respectively, arguing that differences in their distribution of weight criteria are attributed to differences in the phonetic factors governing the two phenomena. In the context of the discussion of weight-sensitive stress, section 5 also explores two other ingredients in the analysis of weight: the phonetic motivations behind language-specific variation in weight criteria for a single process and the role of phonological simplicity in syllable weight. Finally, section 6 sketches a formal constraint-based analysis of weight which incorporates the phonetic conditioning factors governing weight.

2 Background

2.1 Weight as a phenomenon

Linguists have long observed that many languages display phonological phenomena that treat certain syllable types as heavier than others (e.g. Jakobson 1931; Allen 1973; et al.). For example, stress in Yana falls on the leftmost syllable that is closed or contains a long vowel (Sapir and Swadesh 1960). In words without closed syllables and long vowels, stress falls on the first syllable. Thus in Yana, closed syllables (CVC) and syllables containing long vowels (CVV) are ‘heavier’ than open syllables containing a short vowel (CV).

Other phonological phenomena can also be weight sensitive. For example, in many tone languages, syllables differ in the range of tonal contrasts they may support. Thus, while most languages allow level tones on all syllable types, many restrict contour tones to certain heavy syllables. For example, in Kiowa (Watkins 1984), contour tones are restricted to CVV and syllables closed by a sonorant coda (CV[+son]). Contour tones may not occur on CV or on short-voweled syllables closed by an obstruent (CV[−son]).

Other phenomena in addition to stress and tone have been linked to weight: minimal word requirements (McCarthy and Prince 1986, 1995b), metrical scansion (Hayes 1988), compensatory lengthening (Hayes 1989), reduplication (McCarthy and Prince 1986, 1995b), and syllable templatic restrictions such as prohibitions against long vowels in closed syllables (McCarthy and Prince 1986, 1995b). It is thus clear that syllable weight plays an important role in phonological theory.

2.2 Representations of weight

Two representations of weight that have gained wide acceptance in phonological theory are skeletal slot models, including CV and X slot models (McCarthy

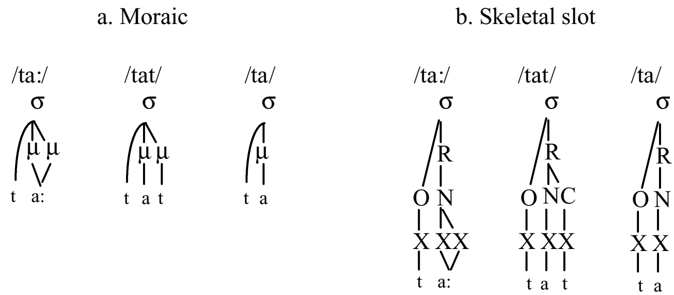


Figure 9.1 Moraic representations of three syllable types in Yana

1979; Clements and Keyser 1983; Levin 1985) and moraic models (Hyman 1985; Hayes 1989). These models assume representations that are projected from properties of the underlying representation, such as segment count and phonemic length. Units of weight, either skeletal slots (in CV and X slot models) or moras, are assigned to segments. In the case of moraic theory, the only segments that are eligible to receive a mora are those in the syllable rhyme, the relevant domain of weight in most cases (Hyman 1985; Hayes 1989). Rhymes with a greater number of segments receive a greater number of weight units. Similarly, contrasts in segmental length are represented by assuming that long segments are associated with two weight units, while short segments are associated with one unit of weight. Weight distinctions are thus reducible to differences in the number of units of weight in a syllable, and, in the case of skeletal slot theory, the affiliation of timing slots. Syllables with a greater number of weight units are 'heavier' than syllables with fewer weight-bearing units. Additionally, in skeletal slot models, it is assumed that weight is calculated over only the nucleus in languages observing either the CVV heavy or the CVV, CV[+son] heavy distinctions (Levin 1985). Sample representations of weight in Yana in Hayes' (1989) moraic and Levin's (1985) skeletal slot models appear in figure 9.1.

Differences in tonal weight can also be captured by moraic and skeletal slot models. It is typically assumed that contour tones result from the combination of two level tones (e.g. Woo 1969; Hyman 1985; Duanmu 1994a, b). Thus, a rising tone reflects the combination of a low tone followed by a high tone, while a falling tone is represented as high tone followed by a low tone. Given the compositionality of contour tones, restrictions against contour tones are usually assumed to arise from a prohibition against associations between more than one tone and a single timing position (either a skeletal slot or mora). Because a contour tone consists of two tones, it requires two timing positions on which to be realised in languages with weight-sensitive tone. For example, the Kiowa restriction against contour tones on CV[−son] and CV follows if one assumes that only sonorants are associated with weight-bearing timing positions in Kiowa.

2.3 *Weight uniformity*

An important claim of both theories of weight is that weight criteria may vary from language to language, but are uniform for different processes within the same language (Hyman 1985; McCarthy and Prince 1986, 1995b; Zec 1994; Hayes 1989). Latin provides an example of uniformity of weight criteria within a single language: the metrical and stress systems, as well as other quantitative phenomena, treat CVV and CVC as heavy and CV as light (Mester 1994). Comparison of Yana stress with Khalkha stress illustrates the parameterisation of weight as a function of language: the stress system in Khalkha treats only CVV as heavy (Bosson 1964; Walker 1995).

Most representations of weight have captured the assumption that weight is a property of languages by parametrising weight representations as a function of language. For example, in Hayes' (1989) moraic theory, some languages (e.g. Yana) assign a mora to syllable-final (coda), while others (e.g. Khalkha) do not. Similarly, in skeletal slot models (e.g. Levin 1985), the syllabic affiliation of sonorant consonants is parameterised on a language specific basis: some languages (e.g. Kiowa) syllabify postvocalic sonorant consonants in the nucleus, while others (e.g. Yana and Khalkha) syllabify them as codas.

Several counterexamples to the moraic uniformity hypothesis have surfaced in recent literature, for example Steriade 1991, Crowhurst 1991, Hyman 1992, and Hayes 1995. For example, Steriade (1991) shows that the stress system, the system of poetic metrics, and the minimal root requirement of Early and Classical Greek are sensitive to different weight criteria from the pitch accent system. At both historical stages of Greek, the stress and metrical systems as well as the minimal root requirement treat both CVV and CVC as heavy. Pitch accent weight criteria are more stringent, however, at both stages. In Early Greek, CVV and CV[+son] are heavy, while in Classical Greek, only CVV is heavy for purposes of pitch accent placement. Another example of conflicted weight criteria comes from Lhasa Tibetan (Dawson 1980). In Lhasa Tibetan, CVV is heavy for stress, while CVV and CV[+son] are heavy for tone; thus, if we conflate weight criteria for both phenomena, we get a three-way hierarchy: CVV > CV[+son] > CV. Crowhurst (1991), Hyman (1992), and Hayes (1995) present additional cases of non-uniformity of weight criteria within a single language.

Cases of conflicted weight criteria are problematic for two reasons. First, there is the issue of representing them formally, which requires reference to at least three levels of weight in a single language. To see this, consider the case of Lhasa Tibetan. If one represents the three-way weight hierarchy in terms of mora count, CVV would need to be trimoraic to be heavier than CV[+son] (which would be bimoraic) and CV (which would be monomoraic). However, CVV should only be bimoraic in moraic theory, which assumes that representations are projected from segment count and phonemic length distinctions. Furthermore, the assumption that tones link to weight-bearing units in one-to-one fashion in languages with weight-sensitive tone would be violated if CVV were trimoraic. The representation of the Lhasa Tibetan facts is also problematic for skeletal slot theories, in which there is no straightforward way to represent the Lhasa distinction between CVV and CV[+son]. The difference between CVV and CV[+son] cannot be captured by assuming that weight for stress is calculated over the nucleus and weight for tone is determined over the rhyme, since CV[–son] also contains a branching rhyme but is nevertheless light for both tone and stress.

A more fundamental challenge presented by cases of conflicted weight criteria concerns the basic conception of weight as a language-driven rather than a process-driven phenomenon. Given the increasing number of cases of conflicted weight criteria reported in the literature, it seems worthwhile to explore systematically the alternative and equally plausible hypothesis that weight is more a function of process rather than of language. Under this view, variation in weight criteria would be attributed principally to differences between weight-based phenomena in the weight distinctions they characteristically employ, rather than to differences between languages. For example, it could turn out that weight-sensitive tone tends to observe different weight criteria than weight-sensitive stress and that this process specificity accounts for many cases of conflicted weight criteria. If this scenario turned out to be true, the focus of the theory of weight should shift from explaining how and why languages differ in terms of their weight criteria to addressing how and why weight criteria differ between weight-sensitive phenomena. Exploring weight as not only a language-driven but also a process-driven property also has the potential to provide insight into cases of weight uniformity. For example, suppose that codas were characteristically weightless for both tone and stress systems. Crucially, if this were true, even if one were to find a language in which coda consonants were weightless for both tone and stress, this convergence of weight criteria would not provide support for the view that weight is uniform as a function of language. The moral of this story is that, when considering the evidence for uniformity of weight, it is as important to pay attention to the cross-linguistic weight patterns displayed by a single process as to convergences of weight criteria within the same language.

3 A cross-linguistic survey of weight

In order to gain a better understanding of the cross-linguistic distribution of weight, a survey of six weight-sensitive phenomena (stress, tone, poetic metrics, compensatory lengthening, minimal word requirements, and syllable template restrictions) in approximately 400 languages was conducted. I briefly summarise the overall results of the survey, which strongly suggest that weight is more process driven than language driven: that is, there are as many, if not more, languages with conflicted weight criteria than there are languages with uniform weight criteria, whereas for any one phenomenon there is a high degree of cross-linguistic convergence for weight criteria. The focus of the discussion in this chapter will be on tone and stress; for more detailed discussion of the other weight-sensitive phenomena, see Gordon 1999.

Before considering overall results of the survey, let us briefly consider how the surveyed phenomena other than tone and stress instantiate weight. *Poetic metrics* is a diagnostic for weight in languages with poetic traditions in which the placement of syllables in the meter is dependent on their weight: heavy syllables show a preference (varying in strength depending on the language) for occurring in strong positions, while light syllables tend to fall in weak positions (Hayes 1988). Languages with *syllable template restrictions* prohibit long vowels in either all closed syllables or syllables closed by a certain type of coda, due to a constraint on the maximum weight of a syllable (Steriade 1991; Hayes 1995). In such languages, coda consonants are heavy, since they contribute to the overall weight of the syllable. Many languages have *minimal word requirements* that are weight-sensitive; in these languages, the smallest content word is a heavy syllable (McCarthy and Prince 1986, 1990, 1995a). Finally, *compensatory lengthening* of a vowel triggered by loss of a coda consonant can be analysed as preservation of the weight of the syllable; whenever loss of a coda (or subset of codas) induces lengthening of the preceding vowel, it is characteristically assumed that the lengthened vowel associates with the weight unit originally linked to the lost consonant (Hock 1986; Hayes 1989).

The first result of the survey is that syllable weight observes a hierarchy, where CVV is heaviest, followed by CV[+son], followed by CV[−son], and finally by CV, as shown in (1) (see also Zec 1988).

(1) CVV > CV[+son] > CV[−son] > CV

For a given phenomenon, languages draw different cut-off points between heavy and light syllables along this hierarchy. For example, Kiowa's tonal system makes the cut-off between CV[+son] and CV[–son], so that CVV and CV[+son] are heavy, while CV[−son] and CV are light. The Yana stress system, on the other hand, makes its cut between CV[−son] and CV, so that only CV is light and all others are heavy. Crucially, a syllable is never heavier than one to its left in the hierarchy.[1]

A second finding of the survey is that processes differ in their distribution of weight criteria. This is shown in table 9.1, which depicts the number of languages displaying a given weight criterion for five of the six phenomena under

Table 9.1. *Weight criteria for different processes*

		Process				
		Stress	Tone	Metrics	Minimal Word	Syllable Template Restriction
Criterion	CVV heavy	35	21	0	17	—
	CVV, CVC heavy	40	3	16	80	53
	CVV, CV[+son] heavy	3	25	0	0	2

discussion (compensatory lengthening is excluded since it is diagnostic for only a single weight criterion, CVV, CVC heavy; see Gordon 1999 for discussion). For reasons of space, table 9.1 is limited to the three most common criteria: CVV heavy; CVV, CVC heavy; CVV, CV[+son] heavy. The distribution of weight criteria for stress, which displays greater cross-linguistic diversity than other phenomena, is discussed in section 5.1. The dash in table 9.1 indicates that syllable template restrictions by their very nature are always sensitive to the presence or type of coda consonant (since it is the presence of the coda that diagnoses the restriction) and are thus unable to diagnose the CVV heavy criterion. First, considering similarities between processes, poetic metrics and syllable template restrictions either exclusively or almost exclusively observe the CVV, CVC heavy criterion. This uniformity is perhaps less striking, however, when one considers that syllable template restrictions are intrinsically not probative in diagnosing the CVV heavy criterion. Minimal word requirements are heavily biased in favour of the CVV, CVC heavy criterion, though there are a substantial minority of languages observing the CVV heavy criterion, unlike in poetic metric systems. Stress systems are almost equally split between the CVV heavy and the CVV, CVC heavy criteria, with a very small number of CVV, CV[+son] heavy languages. Tone, on the other hand, rarely observes the CVV, CVC heavy criterion and is almost equally divided between languages with the CVV heavy criterion and those with the CVV, CV[+son] heavy criterion. Particularly instructive is the comparison of the CVV, CVC heavy and the CVV, CV[+son] heavy criteria for stress and tone: the CVV, CV[+son] criterion is quite common for tone but vanishingly rare for stress, whereas the converse is true for tone. On the other hand, the CVV, CVC heavy criterion is strikingly rare for tone, but is very common for stress. This distributional asymmetry between tone and stress will be attributed (in sections 4 and 5, respectively) to differences in the phonetic factors underlying these phenomena.

Differences in the distribution of weight criteria between different processes can be examined statistically by means of a chi-square test, which assesses differences between pairs of phenomena in the relative proportion of languages displaying a given weight criterion. This test indicates that all pairs of phenomena with the exception of poetic metrics and minimal word requirements differ significantly from each other (at the $p < .01$ level) in their distribution of weight criteria. Thus, comparison of the distribution of weight criteria for different processes argues against the hypothesis that weight criteria are sensitive to individual languages and not to individual processes. If weight were primarily language driven we would not expect processes to differ as much as they do in their cross-linguistic distribution of weight criteria.

It is also possible to test directly the standard assumption that weight criteria are normally uniform for different processes within the same language, by examining languages in the survey with more than one weight-sensitive phenomenon. The most probative languages for testing the weight uniformity hypothesis are those containing multiple weight-sensitive phenomena with differing cross-linguistic distributions in weight criteria, since any convergences in weight criteria between processes with similar weight distributions could be attributed to process-internal rather than language-internal consistency of criteria. Thus, in virtue of the processes involved, we would a priori expect a high degree of convergence in any pairwise comparison of criteria for metrics, syllable template restrictions, and minimal word requirements in a given language possessing at least two of these phenomena. This is in fact what we tend to find: eight of nine languages with both weight-sensitive poetic metrics and a minimal word requirement observe the same criterion, and all four languages with both a weight-sensitive metrical tradition and a syllable template restriction observe the same criterion.

Rather striking is the high degree of conflicting criteria between syllable template restrictions and minimal word requirements: only seven of thirteen languages with both phenomena observe the same criterion, a finding that strongly contradicts the weight uniformity hypothesis. The weight uniformity hypothesis does not fare any better in other pairwise comparisons between phenomena in individual languages. For virtually all other comparisons (stress vs tone, stress vs syllable template restrictions, stress vs minimal word requirements, tone vs metrics, tone vs minimal word requirements), agreement percentages hover at or below 50 per cent (see Gordon 1999 for more discussion). If we compare stress and tone, the focus of this chapter, we see that of the four languages in the survey that allow coda consonants and that have both weight-sensitive tone and stress, in two of them, Krongo (Reh 1985) and Cherokee (Wright 1996), weight criteria for tone and stress agree (CVV heavy in both languages), while in the other two, Classical Greek (Steriade 1991) and Lhasa Tibetan (Dawson 1980), weight criteria disagree. The only pairwise comparison that yields a relatively high level of language-internal consistency is the comparison of stress and metrics, for which six of eight surveyed languages with both phenomena agree in criteria (see Gordon 1999 for analysis of the tendency for stress and metrical systems to observe similar weight criteria).

The upshot of the survey is that weight criteria do not tend to converge any more than one would expect a priori by considering the cross-linguistic distribution of weight criteria for different phenomena. A greater success level in predicting weight criteria is thus achieved through consideration of the process-internal distribution of weight criteria rather than through language-internal comparison of weight criteria for different phenomena.

Given the divergence between different weight-sensitive phenomena in their cross-linguistic distributions of weight criteria, it is natural to seek explanations for weight in terms of factors relevant to the individual phenomena. Throughout the rest of this chapter, we focus on weight-sensitive stress and tone (see Gordon 1999 for discussion of other phenomena), arguing that differences in their phonetic conditioning factors lead to different distributions in phonological weight criteria for the two phenomena, that is, the frequent observance of the CVV, CVC heavy criterion for stress and its corresponding rarity in tone systems, contrasted with the frequent occurrence of the CVV, CV[+son] heavy criterion in tone systems and its rarity in stress systems. The following section focuses on the phonetic underpinnings of weight-sensitive tone; for further discussion of this topic, see Zhang (this volume).

4 The phonetic basis for weight-sensitive tone

The physical correlate of tone is fundamental frequency, which is only present in voiced segments (see Maddieson 1978 for cross-linguistic observations about tone; see Beckman 1986, House 1990, and Moore 1995 for further discussion of its perceptual correlates). In fact, the property that defines a voicing contrast is the fundamental frequency: voiceless segments lack a fundamental, voiced segments have one. Thus, the only type of segment on which tone may be directly realised is a voiced one.[2]

Crucially, the fundamental frequency profile of a segment or syllable (and hence its tonal profile) is cued not only by the fundamental itself but also by the higher harmonics. This is because the harmonics occur at frequencies that are multiples of the fundamental frequency; thus a signal with a fundamental frequency of 200 Hz will have harmonics at 400 Hz, 600 Hz, 800 Hz, 1000 Hz, and at 200 Hz increments thereafter. The presence of harmonics greatly enhances the salience of the fundamental frequency, and can even allow for recovery of the tone when the fundamental itself has been excised from the signal (see House 1990 and Moore 1995 for review of the relevant psychoacoustic literature).

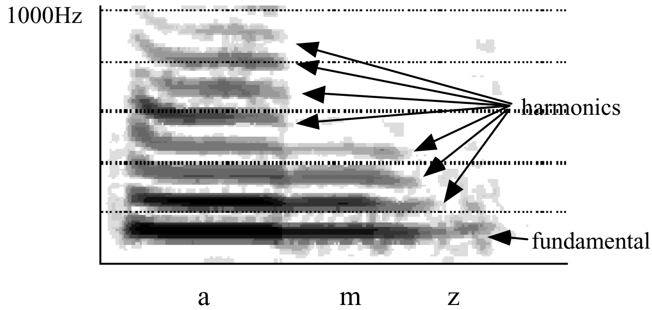


Figure 9.2 Narrowband spectrogram of voiced segments

While the relationship of harmonics to the fundamental in the frequency domain is the same for all segments (harmonics occur at multiples of the fundamental), voiced segments differ in the intensity of their harmonics. Because vowels typically have the most energy at higher frequencies, their higher harmonics have greater intensity than those of consonants. Voiced sonorant consonants also possess a fairly energetic harmonic structure relative to voiced obstruents, but typically do not possess as intense harmonics as vowels. Nevertheless, the more crucial harmonics for the perception of the fundamental, the low frequency harmonics (House 1990), are typically present in sonorants.

In contrast to sonorants, obstruents provide either minimal or no cues to fundamental frequency. Voiceless consonants, including obstruents, do not have a fundamental or harmonics. In voiced obstruents, harmonics above the fundamental typically have very little energy; furthermore, the fundamental itself is typically substantially less intense than in sonorants. The absence of a salient harmonic structure in obstruents and the low intensity of the fundamental are due to the narrow constrictions associated with obstruents. Thus, voiced obstruents are inherently impoverished relative to voiced sonorants in terms of their tonal salience. One would thus expect voiced obstruents to contribute little to the ability of a syllable to carry a contour tone. This fact, taken together with the inability of voiceless obstruents to carry tone, means that the class of obstruents considered as a whole is quite poorly suited to supporting tonal information.[3] A further factor contributing to the characteristic weightless status of obstruents in tonal weight is the relative rarity of languages with voiced coda obstruents.

The relative ability of different segment types to carry tone can be made more vivid by considering a narrowband spectrogram of different types of voiced segments in figure 9.2. Voiceless segments are not included since they lack a fundamental and harmonic structure. In figure 9.2, the vowel has the greatest number of visible harmonics above the fundamental (i.e. those with sufficient intensity to show up in the narrowband spectrogram) and also the most intense ones (as reflected in the darkness of the harmonics). The sonorant consonant also has a relatively rich harmonic structure and relatively intense harmonics, though its harmonics are visibly fewer (again due to decreased intensity at higher frequencies) and less intense than the vowel's. Compared to both the vowel and the sonorant, the voiced obstruent provides very little tonal information: there are no continuous harmonics visible above the fundamental and the fundamental itself is relatively weak in intensity.

The relative salience of tonal information realised on different segment types offers an explanation for the distribution of weight-sensitive contour tone restrictions discussed earlier. Recall the implicational hierarchy of syllable types that may bear contour tones: CVV is heaviest, followed by CV[+son], followed by CV[−son], followed by CV. This hierarchy mirrors the phonetic hierarchy of tonal salience in figure 9.2, under the assumption that contour tones require a longer duration to be realised than level tones (see Zhang's chapter). It is thus crucial that not only the initial portion of the rhyme but also the latter portion possess properties that will allow for recovery of the tonal information. Thus, it is the second half of the long vowel in CVV and the coda consonant in CV[+son] and CV[−son] that serve to differentiate them from each other and from CV in terms of relative ability to carry a contour tone. The hierarchy of syllable types discussed in section 3, CVV > CV[+son] > CV[−son] > CV, thus reduces to a hierarchy characterising the relative ability of different segment types able to support phonetically the latter portion of the contour: V > R > O > Zero, (where the difference between O and zero is not particularly robust; see the discussion earlier).

5 Weight-sensitive stress

Thus far, we have seen that processes differ in their cross-linguistic distributions of various weight criteria and that these differences stem to a large extent from differences in the phonetic conditioning factors that govern these processes. Throughout the rest of the chapter, we will focus on weight-sensitive stress, which provides an instructive contrast to weight-sensitive tone, both in terms of its phonetic underpinnings and its resulting phonological distribution. Examination of weight-sensitive stress also demonstrates the importance of two other elements in a complete analysis of weight. First, it illustrates the important notion (not only for stress, but also for other weight-sensitive phenomena) that weight is guided not only by considerations of phonetic effectiveness but also by principles of phonological simplicity. Second, it will be shown that language-specific differences in weight criteria for a single phenomenon can be attributed to phonetic differences, which are in turn linked to differences between languages with respect to other phonological properties such as syllable

Table 9.2. *Summary of representative weight distinctions*

Weight distinction	No. of languages	Example language
CVV heavy	43	Khalkha
CVV, CVC heavy	43	Yana
CVV, CV[+son] heavy	4	Kwakw'ala
Low V heavy	5	Yimas
Non-high V heavy	3	Komi Jaz'va
Short-central V light	15	Javanese

structure. The relevance of phonological simplicity will be introduced first, in the following section.

5.1 The role of structural simplicity in syllable weight

An interesting feature of syllable weight is the recurrence of relatively simple and symmetrical weight criteria in language after language. To take an example from stress, many languages treat all syllables containing long vowels as heavy (e.g. Khalkha Mongolian), whereas others treat all syllables with branching rhymes, that is, CVV and CVC, as heavy (e.g. Yana). Still others treat all syllables containing a certain vowel quality as heavy (e.g. full vowels in Javanese and low vowels in Yimas). However, we do not find languages with very complex and asymmetrical criteria, even if such criteria might be plausible on purely phonetic grounds. Possible examples would be languages in which long-voweled syllables and those containing low vowels are heavy, or languages in which low vowels followed by a sonorant coda are heavy. An intuitive explanation for the absence of such hypothetical phenomena is that they are phonologically too complex. A major part of the theory of weight proposed here is a limitation on the structural complexity of the available distinctions; the choice of the most phonetically effective distinctions is made only from among the simpler criteria (see Hayes 1999 for similar claims about post-nasal voicing).

Despite its intuitive relevance, defining structural complexity is a very difficult issue. What follows is one proposal that is fully explicit and matches well with my survey data; other possibilities surely exist and remain to be explored.

As a starting point in the discussion of complexity, it is useful to consider in table 9.2 some representative weight distinctions (and the number of languages instantiating them) from the survey of weight-sensitive stress systems in Gordon 1999. The set of weight distinctions observed by phenomena other than stress is a subset of those found in stress systems; thus, a definition of complexity that is adequate for stress will also suffice for other processes. Complex weight hierarchies can be decomposed into a series of binary weight distinctions. For example, the CVV > CVC > CV hierarchy found in Klamath (Barker 1964)

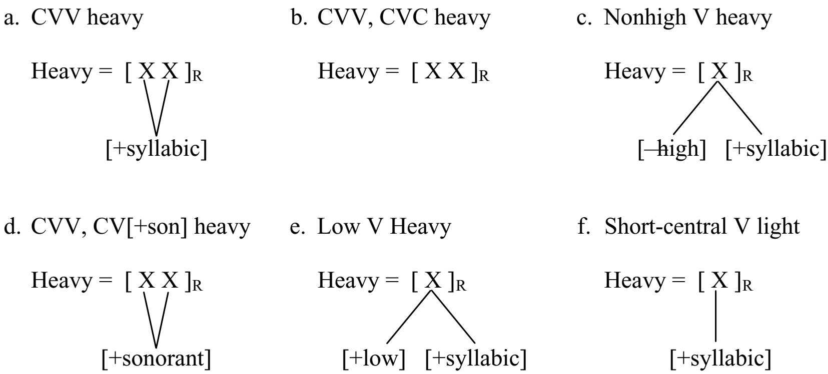


Figure 9.3 Representative weight distinctions

and Chickasaw (Munro and Willmond 1999; Gordon 1999) consists of two weight distinctions: CVV > {CVC, CV} and {CVV, CVC} > CV.[4]

Let us now consider some phonological predicates that define heavy and light syllables for the distinctions in table 9.2. The overall goal of this endeavour will be to provide representations that offer a means of characterising weight distinctions. A theory-neutral notion of weight unit is assumed here, with all segments in the rhyme receiving one weight unit, except for phonologically light central vowels, which are assumed, following Kager (1990), to lack their own weight unit in virtue of their extremely short phonetic duration.[5] Figure 9.3 depicts representations characterising the set of heavy syllables for the weight distinctions in table 9.2. The representations in figure 9.3 define the set of heavy syllables in a language and serve to differentiate them from the light syllables. For example, according to the CVV heavy criterion, all syllables containing at least two syllabic timing positions in the rhyme are heavy. Most of the other representations in figure 9.3 are straightforward, with the possible exception of (f). Adopting the assumption that short-central vowels lack a weight unit of their own, this distinction treats a syllable containing a syllabic timing position as heavy.

Given the representations in figure 9.3, we can hypothesise about the upper limit of formal complexity tolerated by weight distinctions. Most weight distinctions refer only to non-place predicates, that is, timing units and non-place features. The two distinctions that refer to place refer to a place feature linked to a single timing unit. There are no distinctions that refer to a place feature linked to more than one timing unit, as in a hypothetical distinction which treats long low vowels as heavy[6] or one which treats long non-high vowels as heavy. Given the set of distinctions in figure 9.3, I thus offer the following definition of complexity as a working hypothesis: a weight distinction is too complex if it refers to more than one place predicate. I also assume that weight distinctions that require disjoint representations of the heavy syllables are complex, even if they only refer to a single dimension. Thus, for example, a weight distinction that treats long vowels and syllables closed by a lateral as heavy is complex, since there is no single representation of the syllable that encompasses both long vowels and syllables closed by a lateral. This is because long vowels contain no [+lateral] timing positions; there is thus no way for the second timing position in the rhyme to be both simultaneously [+lateral] and [+syllabic].[7] The definition of complexity is formalised in (2).[8]

(2) Definition of complexity: A weight distinction is complex iff
It refers to more than one place predicate.
OR
It makes reference to disjoint representations of the syllable.

This definition of complexity also allows for other sporadically attested weight distinctions not in table 9.2 (see Gordon 1999, 2002 for discussion). As we will see in the phonetic case studies in section 5.8, the notion of phonological simplicity plays an important role in eliminating certain weight distinctions from the set of a priori logically possible weight distinctions, many of which are phonetically very effective.

5.2 Phonetic effectiveness and weight-sensitive stress

In this section, I examine the importance of phonetics in guiding the language-specific choice of weight criteria for weight-sensitive stress. I will argue that languages choose their weight distinctions in order of phonetic effectiveness from among the phonologically simple ones.

Phonetic effectiveness may be defined as the degree to which a particular weight division separates syllables into two maximally distinct groups. In other words, the most effective division of syllables has heavy and light syllables which are most different from each other along some phonetic dimension.

The motivation for this metric of phonetic effectiveness is perceptual. It is hypothesised that languages prefer to rely on weight distinctions based on the largest phonetic differences, since distinctions based on larger phonetic differences are easier to perceive and thus to learn than distinctions based on smaller differences. Furthermore, distinctions relying on relatively large phonetic differences are plausibly easier to deploy, since they harmonise with inherent phonetic prominence. Phonetic and perceptual distinctness (or conversely, lack of distinctness) have been argued to play an important role in phonology in such diverse areas as the construction of segment inventories that maximise the phonetic space (cf. Liljencrants and Lindblom 1972; Lindblom 1986), neutralisation processes eliminating phonetic contrasts which are difficult to implement in a perceptually salient manner (Flemming 1995; Steriade 1999), and phonological processes that strive to preserve or create maximally distinct segments or combinations of segments (Flemming 1995, this volume).

Although the discussion of phonetic effectiveness is couched here within the context of weight-sensitive stress, considerations of phonetic effectiveness may also be assumed to guide choices in weight criteria for tone as well. The only difference between the two phenomena is in the phonetic dimension along which potential weight criteria are evaluated. For tone, I hypothesise that the relevant dimension is the energy found in the sonorant portion of the syllable, whereas, for stress, it is the energy profile of the entire rhyme that is relevant (see Gordon 2001 for discussion and data).

In the present study, phonetic effectiveness was examined along two phonetic dimensions: the duration of the syllable rhyme and the energy of the syllable rhyme. The procedures for measuring duration and energy, and the languages and corpora from which measurements were made, are discussed in sections 5.3–5.7. Results are presented in section 5.8.

5.3 Languages

Six languages displaying various weight distinctions for stress were investigated. Languages were chosen that represented a cross section of attested weight distinctions (Gordon 1999). This chapter will consider results from three languages, one employing the CVV heavy criterion (Khalkha), one with the CVV, CVC heavy criterion (Finnish), and one observing the three-way hierarchy CVV > CVC > CV (Chickasaw). The CVV heavy and the CVV, CVC heavy criteria are the two most common criteria for stress; the hierarchy CVV > CVC > CV is the most common three-way distinction. Results for Khalkha, Finnish, and Chickasaw are similar to those found in other languages, data for which is found in Gordon 1999, 2002.

5.4 Corpora

A corpus of two-syllable words of the form (C)V(ː)C.CV(C) was constructed for each language, varying the rhyme of the first syllable, which was the target syllable, and keeping the vowel in the other syllable constant. Within each language, the first syllable was either phonologically stressed for all words in the corpus (Khalkha and Finnish), or was phonologically unstressed for all words (Chickasaw). The second syllable had the opposite stress level of the first syllable, that is, unstressed in Khalkha and Finnish, and stressed in Chickasaw. By keeping stress uniform for all target syllables, a difference in stress level

Table 9.3. *Vowels and codas measured for Chickasaw, Khalkha, and Finnish*

Language	Vowels and codas measured
Chickasaw	a, i, aː, iː, m, n, l, ʃ, ɬ, b, k
Khalkha	ɑ, u, ɑː, uː, m, n, l, r, s, ʃ, x, k, g
Finnish	ɑ, i, u, ɑː, iː, uː, m, l, r, s, t

between different syllable types was eliminated as a potential confounding factor. The rhymes appearing in the first syllable were varied according to the vowel quality and length (if long vowels occurred in the language) of the syllable nucleus. Three vowel qualities were examined: /i, u/ and a low vowel, either /a/ or /ɑ/. Rhymes containing /i/ were not measured in Khalkha due to confounds created by the vowel harmony system. Rhymes containing /u/ were not measured for Chickasaw due to the absence of this vowel in the inventory. In order to create a more manageable data set for measurement, diphthongs and mid vowels were not examined in any of the languages; thus, the phonetic basis for weight distinctions between long vowels and diphthongs and between mid vowels and other vowel qualities was not examined experimentally. Short vowels were examined in both open syllables and syllables closed by various coda consonants. The set of coda consonants and the vowels examined for each language is listed in table 9.3.

5.5 Measurements

Recall from section 4 that weight-sensitive tone is claimed to be sensitive to the energy profile of the sonorant portion of the syllable rhyme. The phonetic underpinnings behind weight-sensitive stress are not as transparent as those underlying tone, however, as stress is often associated with multiple acoustic properties.

For the present study, duration and energy were targeted as phonetic dimensions for investigation because they are closely linked to the realisation of stress in many languages: both increased energy (along with its perceptual correlate, loudness) and duration are common phonetic correlates of stress. In many languages, stressed syllables are either longer or louder than unstressed syllables, or are *both* longer *and* louder than unstressed syllables. The correlation between stress and increased duration and/or loudness has been experimentally shown for many languages, including English (Fry 1955; Beckman 1986), Polish (Jassem et al. 1968), Mari (Baitschura 1976), Indonesian (Adisasmito-Smith and Cohn 1996), Dutch (Sluijter and van Heuven 1996), and so on and has been impressionistically noted for many other languages.

Six to eight tokens of each word were recorded from one speaker of each language. Words were read in random order and appeared in a carrier phrase. Data were digitised at 16 kHz using the Kay Computerised Speech Lab. Two measurements were made for each rhyme: duration, and a measure that may be termed *total perceptual energy*: the integration of energy over time in the perceptual domain.

A measure of total energy rather than average intensity is most relevant for testing the link between energy and weight-sensitive stress, since psychoacoustic experiments suggest that the ear integrates intensity and time over durations of the magnitude common for syllables in natural speech (see Moore 1995 for a review of the relevant literature).

The procedure for measuring total perceptual energy was as follows. First, in order to control for token-to-token variation in speaking level, average amplitude (RMS) in decibels for each target vowel and the following coda consonant (if any) was calculated relative to a reference vowel. This reference vowel, which was the vowel in the other (non-target) syllable, was kept constant for each set of comparisons. Second, the average amplitude of each segment in the target rhyme was converted to a value representing perceived loudness relative to the vowel in the second syllable. Perceived loudness was computed on the basis of Warren's (1970: 1399) results in experiments designed to measure relative perceived loudness of tones. While Warren's results are based on a different type of stimulus than real speech, they serve as a reasonable and also tractable estimate of the relationship between acoustic energy and perceived loudness. Third, the relative loudness value for each segment was multiplied by the duration of the segment to yield a total energy value for the segment. Finally, if the rhyme contained a coda, total energy values for the vowel nucleus and the coda were added together, yielding a total energy value for the rhyme.

5.6 *Phonetic evaluation of potential weight criteria*

Along the two phonetic parameters of duration and total energy, the syllables measured in a given language were bisected in a number of different ways, with each bisection representing a different formal weight criterion. In determining the phonetic effectiveness of different weight distinctions, the goal was to test all reasonable distinctions against the phonetic data. A total of fifty-five weight distinctions were tested, though not all weight distinctions could be tested in every language due to gaps in the inventory of syllable types in certain languages. Furthermore, in some cases, two different distinctions provided the same division of the data. Thus, for example, distinctions based on the voicing of the coda and those based on the sonorancy of the coda divide the data in the same way for Finnish, as the only sonorants in Finnish are voiced and the only coda obstruents in Finnish are voiceless.

The tested weight distinctions were based on several phonological parameters, including duration (i.e. one vs more than one timing position) and the features ([high]/[low] for vowels, and [coronal], [dorsal], [labial], [voice], [sonorant], and [continuant] for consonants). In addition, the distinction between CVVC (superheavy) and other syllables was tested in Chickasaw.

5.7 *A quantitative metric of phonetic effectiveness*

For the parameters of duration and total energy, distinctions were compared in a three-step process. First, a one-factor analysis of variance was performed, treating rhyme type (e.g. /an/, /am/, /is/, /uk/, etc.) as the independent variable, and duration and energy as the dependent variables. The purpose of this initial analysis was merely to determine whether syllable type had an effect on duration and energy values.

The second step was to compare the mean values for heavy and light syllables for each weight distinction. Weight distinctions for which the means for heavy and light syllables were most divergent were deemed to be the most effective weight distinctions. Because they are not dependent on number of tokens, differences in means were used to determine the relative effectiveness among the weight distinctions. The metric of phonetic effectiveness adopted as a differentiator of weight criteria is summarised in (3).

(3) Definition of phonetic effectiveness
Weight distinction x is more effective than weight distinction y if the difference between the mean energy of heavy syllables and the mean energy of light syllables for distinction x is greater than the difference between the mean energy of heavy syllables and the mean energy of light syllables for distinction y.

The final step in evaluating the phonetic effectiveness of different weight distinctions was to perform a discriminant analysis for each distinction to determine how well it sorted syllables into heavy and light groups. Each weight distinction was treated as a categorical variable with two values: one for light syllables and another for heavy syllables. Significance levels and Wilkes' lambda values for each weight distinction were examined to determine how reliable various weight distinctions were in differentiating heavy and light syllables. Lower

Wilkes' lambda values generally indicate greater robustness in the statistical difference between heavy and light syllables.[9]

5.8 *The link between energy and language-specific weight distinctions*

In this section, the results of the phonetic study of the link between duration and energy and phonological weight are presented. Strikingly, there was a very close overall association between weight criteria and energy. Phonological weight distinctions chosen by languages were the ones that were phonetically most effective along the energy dimension. Duration, on the other hand, was a less effective predictor of certain weight criteria than energy. For this reason, the results of the energy study are presented first in sections 5.8.1–5.8.3; discussion of duration is deferred to section 5.9.

5.8.1 Chickasaw: CVV > CVC > CV Let us begin with Chickasaw (Munro and Willmond 1999; Gordon 1999); a language that makes a ternary weight distinction of the CVV > CVC > CV type, the most common three-way weight hierarchy.

Because it possesses more than a binary weight distinction, Chickasaw provides a relatively tough testing ground for establishing an association between energy and syllable weight. As a first step, an analysis of variance was conducted to determine whether rhyme type had a significant effect on energy values. This ANOVA indicated a highly significant effect of rhyme type on energy values: $F(21, 153) = 15.215$, $p < .0001$. In table 9.4, the relative effectiveness of different weight distinctions in Chickasaw is compared. Distinctions are ordered by phonetic effectiveness, with the more effective distinctions (as described in section 5.7) on top. Mean values are normalised as a ratio relative to the top-ranked distinction, which is assigned an arbitrary value of 100. For example, a weight distinction with a value of 50 in table 9.4 has a 50 per cent smaller difference in energy between heavy and light syllables than the top-ranked distinction. Table 9.4 also includes Wilkes' lambda values and significance levels according to the discriminant analyses.

Note that all of the ties in table 9.4 between two weight distinctions are the result of two weight divisions completely overlapping. For example, the first two distinctions in the column of complex distinctions, the distinction between long low vowels and other rhymes (i.e. /aː/ heavy) and the distinction between long back and long low vowels and other rhymes (i.e. /aː/ and /uː/ heavy), are equivalent for the data set examined, since Chickasaw does not have a long /uː/. Equivalent weight distinctions of this sort are surrounded by brackets. The boldfaced distinctions are the ones actually employed in Chickasaw. Due to space constraints, only those complex distinctions that are superior to

Table 9.4. *The most effective weight distinctions in Chickasaw*

Simple				Complex			
Distinction	Diff	W-λ	p-val.	Distinction	Diff	W-λ	p-val.
				{ /aː/ heavy	100	.657	.0000 }
				/aː, uː/ heavy	100	.657	.0000
{ **VV heavy**	**80.6**	**.603**	**.0000**	/aː, iː/ heavy	80.6	.603	.0000 }
				VV, a[+son] heavy	73.3	.581	.0000
				VV, a[+nas] heavy	72.5	.613	.0000
				V, hiV[+dor] light	71.6	.796	.0000
VV, VC heavy	**71.5**	**.862**	**.0000**				
VVC heavy	67.8	.799	.0000				
VV, V[+son] heavy	64.8	.661	.0000				
VV, V[+voi] heavy	56.3	.760	.0000				
VV, V[+cont] heavy	55.9	.747	.0000				
VV, V[−nas] heavy	31.7	.934	.0006				
+low V heavy	17.7	.975	.0351				

at least one of the actual phonological distinctions are listed in table 9.4, and in subsequent tables for other languages. All of the simple distinctions after the phonological ones are listed in order of relative phonetic effectiveness according to differences between means. Note the following abbreviations in table 9.4 and subsequent tables: [+voi] represents a voiced consonant, [−voi] a voiceless consonant, [+cont] a continuant, [−cont] a non-continuant, [+nas] a nasal, [−nas] a non-nasal, [+lab] a labial, [−lab] a non-labial, [+cor] a coronal, [−cor] a non-coronal, [+dor] a velar and [−dor] a non-velar. Strikingly, the two phonetically most effective weight distinctions among the simple distinctions are precisely the ones exploited by the phonology of Chickasaw. The optimal simple distinction singles out the heaviest member of the weight hierarchy, CVV, while the second best simple distinction designates the two heaviest syllable types in Chickasaw (CVV, CVC). The three-way phonological hierarchy thus results from the combination of the top two phonetic distinctions. The Chickasaw data thus provide strong evidence for a match between syllable weight and the phonetic property of energy.

The Chickasaw data also provide corroboration for the importance of phonological simplicity in syllable weight. Six of the top eight weight distinctions are ruled out only by virtue of their complexity. If simplicity did not play a role in the phonology of weight, one would incorrectly expect the phonology to

Table 9.5. *The most effective weight distinctions in Khalkha*

Simple				Complex			
Distinction	Diff	W-λ	p-val.	Distinction	Diff	W-λ	p-val.
				VV, a[+nas] heavy	100	.635	.0000
				VV, a[+lab] heavy	99.1	.708	.0000
VV heavy	**89.7**	**.832**	**.0000**				
VV, VC heavy	48.1	.949	.0047				
VV, V[+son] heavy	43.9	.879	.0000				
VV, V[+voi] heavy	38.8	.906	.0001				
VV, V[+cont] heavy	13.9	.988	.1769				
VV, V[−nas] heavy	13.5	.991	.2398				
+low V heavy	11.9	.991	.2441				
VV, V[−son] heavy	2.9	.999	.7770				

observe the Caː heavy criterion. One would also incorrectly predict several other complex weight distinctions to surface before the CVV, CVC heavy criterion.

5.8.2 Khalkha: CVV heavy Unlike Chickasaw, Khalkha Mongolian observes a simple binary weight distinction for stress, that is, CVV is heavy (Bosson 1964; Walker 1995). An analysis of variance indicated a highly significant effect of syllable type on energy: $F(21, 132) = 5.857$, $p < .0001$. Individual distinctions are compared in table 9.5. Table 9.5 shows that the phonological weight distinction between CVV and other rhymes is the phonetically most effective distinction among the structurally simple weight distinctions. There are only two distinctions that are superior phonetically to the actual phonological distinction: CVV, a[+nas] heavy and CVV, a[+lab] heavy; however, both of these distinctions are structurally complex, since the heavy syllables in both distinctions cannot be united in a single representation. Thus, Khalkha provides evidence both for a link between syllable weight and total energy, and also the importance of phonological simplicity in the determination of syllable weight.

5.8.3 Finnish: CVV, CVC heavy The Finnish stress system treats both CVV and CVC as heavy (Sadeniemi 1949) for purposes of determining secondary stress. An analysis of variance found a highly significant effect of syllable type on energy values: $F(21, 149) = 34.300$, $p < .0001$. Table 9.6 lists the relative phonetic effectiveness of different weight distinctions in Finnish.

As table 9.6 shows, the link between energy and phonology is quite strong, as in other languages: the phonological weight distinction is also the most effective distinction phonetically, tied only with a number of complex distinctions.

Table 9.6. *The most effective weight distinctions in Finnish*

Simple				Complex			
Heavy	Diff	W-λ	p-val.	Heavy	Diff	W-λ	p-val.
VV, VC heavy	**100**	**.431**	**.0000**	V, hiV[+dor] light	100	.431	.0000
				V, V[+dor] light	100	.431	.0000
				VV, V[−dor] heavy	100	.431	.0000
				hiV in open σ light	99.5	.604	.0000
VV, V[+son] heavy	62.6	.555	.0000				
VV, V[+voi] heavy	62.6	.555	.0000				
VV, V[+cont] heavy	57.2	.628	.0000				
VV heavy	53.8	.836	.0000				
VV, V[−nas] heavy	52.8	.736	.0000				
−back V heavy	12.0	.985	.1161				
+low V heavy	1.1	1.00	.8872				

5.8.4 The phonetics of syllable weight: a summary The experimental data in the preceding sections indicate a number of important facts. First, in all of the languages with weight-sensitive stress, the phonological weight distinction(s) are also the phonetically most sensible of the simple distinctions. This is true not only of languages with binary weight distinctions like Finnish and Khalkha, but also Chickasaw, which observes a three-way weight hierarchy. These data suggest a strong link between the phonology of weight-sensitive stress and a measure of total energy.

Although space constraints do not permit showing all the data here, a similarly good fit between phonetic effectiveness and phonological weight obtained for other languages examined in the phonetic study (see Gordon 1999, 2002 for details). In addition, differences in the language-specific choices in weight criteria for phenomena other than stress turn out to be correlated with language-specific phonetic properties, along different phonetic dimensions than the energy dimension. For example, language-specific differences in weight criteria for tone are associated with differences in sonorous duration, whereas differences in weight criteria for syllable template restrictions appear to be correlated with differences in rhyme duration, that is, syllable template restrictions reflect an upper limit on the total duration of the rhyme. The interested reader is referred to Gordon 1999 for discussion of the phonetic motivations behind these and other weight-sensitive phenomena.

An equally important fact emerging from the data is that syllable weight is not sensitive only to phonetic properties. Rather, phonological simplicity plays an important role in Chickasaw and Khalkha in filtering out weight distinctions that

Table 9.7. *Effectiveness of simple weight distinctions in terms of duration*

Weight distinction	Chickasaw		Khalkha		Finnish	
Heavy	Diff.	Rank	Diff.	Rank	Diff.	Rank
CVV heavy	**14.4**	**7**	**2.3**	**6**	22.3	6
CVV, CVC heavy	**101.2**	**1**	76.3	1	**95.5**	**1**
CVV, V[+son] heavy	23.5	6	−6.1	5	39.7	4t
+low V heavy	8.6	8	−1.3	7	10.0	7
−back V heavy	NA	NA	NA	NA	1.3	8
CVVC, CVCC heavy	30.3	4	NA	NA	NA	NA
CVV, V[+voi] heavy	28.8	5	−9.2	4	39.7	4t
CVV, V[+cont] heavy	32.6	3	26.3	3	44.9	3
CVV, V[−nas] heavy	35.6	2	33.4	2	46.1	2

may be phonetically quite effective but nevertheless are too complex. Thus, the overall picture that emerges is that syllable weight is the result of compromise between achieving the often conflicting goals of constructing a phonetically sensible grammar that also manipulates a relatively simple set of phonological predicates.

5.9 The relationship between duration and syllable weight

Section 5.8 demonstrated that a measure of integrated energy lines up quite closely with phonological weight in a number of languages with different weight distinctions. Duration was also examined in the languages studied, but unlike energy, it often failed to match with phonological weight.

In table 9.7, the effectiveness of the simple weight distinctions in the duration domain is considered. Complex weight distinctions are not considered in table 9.7, as they are eliminated from consideration independently of phonetic effectiveness. The difference (in milliseconds) between the means for heavy and light rhymes according to each distinction appears along with a ranking indicating the effectiveness of a given weight distinction relative to others. Thus, the first-ranked weight distinction is the most effective one, followed by the second-ranked one, and so on. The weight distinction(s) actually employed by the phonology of a given language is indicated by boldface. Ties in ranking are indicated by a lowercase ‘t’ in the rank column.

The CVV, CVC heavy distinction is the most effective distinction in all languages examined. This includes languages in which this distinction is actually employed (Finnish and Chickasaw), as well as languages that employ different weight distinctions (Khalkha). The CVV, CVC heavy is thus phonetically quite an effective distinction from a durational standpoint.

However, other phonological weight distinctions do not provide as good a phonetic fit in terms of duration. For example, the distinction according to which CVV but not CVC is heavy is ranked far behind the CVV distinction in all languages, including one that exploits the CVV distinction but not the CVV, CVC heavy distinction: Khalkha. In Khalkha, the CVV vs CVC, CV distinction divides syllables into two groups whose means differ from each other by only 2.3 ms. Furthermore, in another language that exploits both the CVV heavy and the CVV, CVC heavy distinctions (Chickasaw), the CVV heavy distinction is surpassed by several other weight distinctions that do not emerge in the phonology. In summary, the duration data fits well with the CVV, CVC heavy criterion, but does not provide a good fit with the CVV heavy criterion.

5.10 The influence of phonological structure on phonetic variation

In section 5.8, evidence for a close match between phonological weight for stress and a phonetic measure of energy was presented. What has not yet been explored in the context of weight-sensitive stress is the directionality of the relationship between phonetics and phonology. Thus, one may ask whether languages tailor their stress systems to fit their phonetic characteristics or whether languages adapt certain phonetic patterns to maximise the phonetic effectiveness of their weight systems. A third and intermediate possibility is that both systems influence each other.

Weight-sensitive stress provides arguments for an integrated model of the phonology/phonetics relationship whereby one aspect of phonological structure influences phonetic properties which in turn influence another phonological phenomenon. One of these arguments will be considered in detail in section 5.10.1. Further evidence for this phonology/phonetics relationship is considered in Gordon 1999, 2002.

5.10.1 The function of syllable structure in language specificity of weight criteria Results of the present study suggest that a basic phonological property of a language, its syllable structure, can trigger phonetic differences between languages which in turn lead to variation in weight criteria. Some of the relevant data demonstrating the effect of syllable structure on the phonetic dimension underlying weight-sensitive stress come from comparison of languages observing the CVV heavy criterion for stress and those displaying the CVV, CVC heavy criterion for stress. The exemplar language in this chapter for the CVV heavy criterion is Khalkha, while Finnish employs the CVV, CVC heavy criterion. As has been shown, for both languages, the phonological weight criterion is also the phonetically most effective of the simple weight distinctions.

Interestingly, Khalkha differs from Finnish not only in terms of weight criterion, but also in terms of its inventory of coda consonants. Let us consider the difference along two dimensions: the number of permissible voiceless codas relative to the number of permissible voiced codas, and the number of obstruent codas relative to the number of sonorant codas. The reasons for examining voicing and sonorancy will become apparent shortly.

Both the voiceless-to-voiced ratio and the obstruent-to-sonorant ratio for codas is much larger in Khalkha than in Finnish. This structural difference can be seen just by comparing the set of attested voiced codas and voiceless codas in the target languages, without weighing their relative lexical frequencies; for example, if one assumes that all codas (excluding recent loans) are weighted equivalently whether they occur in 10 words or 100 words. Thus, what is claimed to be relevant here is the type frequency, and not necessarily the token frequency. According to Poppe (1951), Khalkha has the following inventory of coda consonants, including codas that are clearly phonemic and those that are allophonic: [p, t, ts, tʃ, kʲ, k, s, ʃ, x, m, n, ŋ, l, r, b, g]. If one splits this inventory along the voicing dimension, there are slightly more voiceless codas than voiced codas: nine voiceless codas, [p, t, ts, tʃ, kʲ, k, s, ʃ, x], as compared to seven voiced codas, [m, n, ŋ, l, r, b, g]. Divided along the sonorancy dimension, there are eleven obstruent codas, [p, t, ts, tʃ, kʲ, k, s, ʃ, x, b, g], and five sonorant codas, [m, n, ŋ, l, r]. Finnish has five sonorant codas, all of them voiced [m, n, ŋ, r, l], and four obstruent codas, all of them voiceless [s, p, t, k].[10]

The reason sonorancy and voicing are relevant to the present discussion is that differences between segments along these dimensions are reliably associated with differences in energy. Sonorants characteristically have greater energy than obstruents and voiced sounds typically have greater energy than voiceless sounds, all else being equal. Although these generalisations are not without exception, sonorancy and voicing are two of the best, if not the best, features for predicting energy values. If one considers the energy of CVC syllables as a whole, CVC will, all else being equal, have greater energy if a larger set of the coda consonants are voiced rather than voiceless. Similarly, CVC will have greater energy if a larger set of the coda consonants are sonorants rather than obstruents. This argument of course adopts the assumption made above that all coda consonants are weighted equally in the calculation of energy for CVC as a whole.

Following this line of reasoning, CVC in Khalkha would be expected to have less energy than in Finnish, since Khalkha has both a greater obstruent-to-sonorant ratio and a greater voiceless-to-voiced ratio of coda consonants than Finnish. This hypothesis can, in fact, be tested by examining the energy of CVC relative to both CV and CVV in Khalkha and Finnish. The Khalkha data includes three sonorant codas (m, r, l) and five obstruent codas (s, ʃ, x, k, g). Considered along the voicing dimension, four voiced consonants (m, r, l, g) and

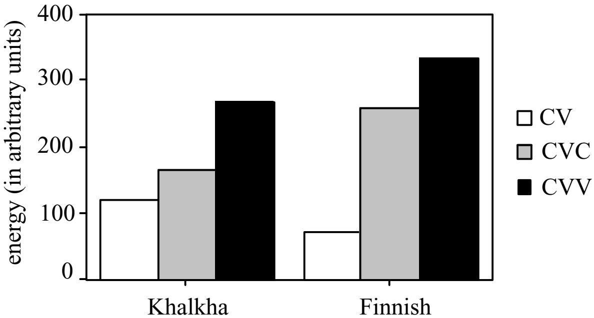


Figure 9.4 Energy values for CV, CVC, and CVC in Khalkha and Finnish

four voiceless consonants (s, ʃ, x, k) were included. The Finnish data includes three sonorant codas, all of them voiced (m, r, l), and two obstruent codas, both of them voiceless (s, t). The corpus for the two languages thus roughly reflects differences between the two languages in the type frequency of voiced relative to voiceless consonants and in the type frequency of sonorants relative to obstruents.

Given the differences in the set of codas examined for Finnish and Khalkha, one would also expect differences in the energy of CVC relative to CVV and CV between the two languages. In particular, CVC should be closer to CVV in energy in Finnish than in Khalkha. Conversely, CVC should be closer to CV in Khalkha than in Finnish. This hypothesis is tested in figure 9.4, which contains energy values for CVV, CVC, and CV in Khalkha and Finnish.

As predicted, CVC is closer in energy to CV than CVV in Khalkha, whereas CVC is closer to CVV than to CV in Finnish. This result corresponds to the difference in the weight of CVC in the two languages. In Khalkha, CVC is light, whereas in Finnish, CVC is heavy. The overall picture that thus emerges is that one language-specific aspect of the phonological system, syllable structure, leads to phonetic differences between languages, which in turn are responsible for differences in weight criteria. Positing this link between structural properties and syllable weight via the intermediary of phonetics makes the interesting prediction that weight distinctions are at least partially predictable if one considers the syllable structure of a language.

This prediction can be tested by examining the inventory of coda consonants in other languages employing either the CVV heavy or the CVV, CVC heavy distinctions for stress. The account given here would predict that languages with the CVV heavy criterion should have a greater obstruent-to-sonorant coda ratio and/or a greater voiceless-to-voiced coda ratio than languages employing the CVV, CVC heavy criterion.

This hypothesis was tested by examining the set of coda consonants for languages in Gordon's (1999) survey which observe either the CVV heavy or the CVV, CVC heavy criteria for stress and which possess both closed syllables and either long vowels or diphthongs. The inventory of codas was examined for a total of sixty-two languages. The results, which are presented in greater detail in Gordon 1999 are summarised here. Of these sixty-two languages, in twenty-three, both the sonorant-to-obstruent ratio and the voiced-to-voiceless ratio are less than one, and in twenty-four, both the sonorant-to-obstruent ratio and the voiced-to-voiceless ratio are at least one.

Strikingly, of the twenty-three languages in which both the sonorant-to-obstruent ratio and the voiced-to-voiceless ratio are less than one, twenty-two employ the CVV heavy criterion, just as predicted by the hypothesis that weight is ultimately determined in large part by coda inventory. The only exceptional language is Yana (Sapir and Swadesh 1960), which has the CVV, CVC heavy criterion yet has sonorant-to-obstruent and voiced-to-voiceless ratios of less than one. Conversely, of the twenty-four languages in which both the sonorant-to-obstruent ratio and the voiced-to-voiceless ratio are at least one, all but four observe the CVV, CVC heavy criterion, again as predicted. A chi-square test, in which languages were coded categorically as either containing sonorant-to-obstruent and voiced-to-voiceless ratios of less than one or containing sonorant-to-obstruent and voiced-to-voiceless ratios of at least one, confirmed that the close link between coda inventory and weight is not due to chance: $\chi^2 = 29.644$, $p < .0001$.

Future research should investigate the extent to which phonological weight criteria are the phonetically most effective in languages that are exceptional in either their voiced-to-voiceless coda ratio or their sonorant-to-obstruent coda ratio, or, even more importantly, languages that are exceptional along both dimensions (e.g. Yana). As far as the present research is concerned, though, it is striking that coda inventories serve as an excellent predictor of weight criteria, as predicted by the proposed account in which syllable weight is ultimately dependent on syllable structure.

In summary, data presented in this section suggest that coda inventory plays an important role in establishing phonetic patterns which in turn are responsible for language-specific choices in weight criteria for stress.

6 A constraint set for weight-sensitive stress

Thus far, we have seen that the phonological weight distinctions in languages are those that are both phonetically effective and structurally simple. Discussion of the formal representation of weight-based stress has been kept to a minimum thus far, used only in the discussion of structural complexity in section 5.1. In this section, I explore the way in which phonetic conditioning factors can be incorporated into a formal theory of weight. The discussion here will focus on stress, though a similar relationship between phonetic effectiveness and the formal analysis of weight obtains for other weight-sensitive phenomena, with the dimensions along which phonetic effectiveness is calculated differing between phenomena (Gordon 1999). Before proceeding, a caveat is necessary. The present proposal is not intended to be a comprehensive metrical theory, which although an integral part of a complete account of syllable weight, goes well beyond the scope of this chapter.

The model I briefly sketch here as a formalism for syllable weight is couched within an Optimality-Theoretic framework and follows work by Prince and Smolensky (1993), Kenstowicz (1994) and others in which much of the burden of phonology is shifted from representations to constraints. In their accounts of weight-sensitive stress, Prince and Smolensky (1993) and Kenstowicz (1994) posit constraints referring to different syllable types involved in a hierarchy of prominence. These constraints capture what Prince and Smolensky (1993: 38) term ‘prominential enhancement that calls directly on contrasts in the intrinsic prominence of syllables’.

Following this work, I assume that the structurally simple weight distinctions mentioned are reflected in constraints referring to weight-sensitive stress. We may also speculate that constraints referring to complex weight distinctions also exist in the grammar but are destined to be mired at the bottom of the constraint hierarchy, precisely because they are complex and thus are unlikely to be entertained by the language learner evaluating simple criteria before complex ones. Under this view, the learner first tests simple weight criteria against a map of phonetic experience (cf. Hayes 1999) and only proceeds to more complex criteria after the simple ones have proven themselves to be poorly suited to the language. Thus, what is innate is not the set of constraints, but the learning algorithm which tests simple weight distinctions before complex ones.[11]

The constraints discussed here refer to stress. All of the representations in section 5.1 that refer to heavy syllables appear as positively stated constraints requiring that the given syllable be stressed. For example, the CVV, CVC heavy distinction is reflected in the high ranking of the constraint in (4).

(4) STRESS $[XX]_R$: CVV and CVC syllables are stressed.

One violation of the constraint is incurred for each instance of an unstressed syllable containing $[XX]_R$.

The structurally simple constraints are ranked on a language-specific basis according to how well different weight distinctions fit the phonetic map. More effective weight distinctions in a given language are ranked ahead of less effective distinctions in the family of weight-sensitive constraints. For example, if the CVV, CVC heavy distinction is the optimal phonetic distinction in a language, that language will rank STRESS $[XX]_R$ above all the other stress constraints. Under this view, the default ranking of constraints in a given language is determined on the basis of phonetic effectiveness. There is limited opportunity for purely inductive learning of the rankings, as, for example, in a hypothetical language in which the observed weight criterion is not the phonetically most effective one. The constraints on stress are interleaved with other constraints; for example, constraints against more than one stress per word, constraints requiring stress in a word, constraints against stress clashes, and so on.

In complex weight hierarchies, more than one constraint is ranked highly enough in the grammar to be active; for example, in Chickasaw both STRESS$[XX]_R$ and the constraint requiring that CVV be stressed are highly ranked.

In the next section, a sample analysis illustrating the interaction between the STRESS constraints and other constraints is presented.

6.1 Yana stress and the STRESS *constraints*

Recall that in Yana (Sapir and Swadesh 1960), stress falls on the first syllable in a word that is either closed (5a) or contains a long vowel or diphthong (5b). If there are no such syllables in the word, stress falls on the first syllable (5c).

(5) Yana stress
   a. sibúmk'ai 'sandstone'
   b. suk'óːniyaː 'name of Indian tribe', ʒáuxauyaː 'Hat Creek Indians', tsiniyáː 'no'
   c. p'údiwi 'women'

Only a few constraints are necessary to account for Yana. First, the STRESS constraint in (4), requiring that all CVV and CVC syllables be stressed, is highly ranked. A second constraint, inviolable in Yana and perhaps in all languages, requires that there be a single syllable in every stress domain that has greater stress than others, following Prince's (1983) Culminativity condition on metrical representations. This constraint is formulated in (6).

(6) ONE STRESS: A word has one and only one stressed syllable.

The final constraint needed is one that requires that stress fall as far to the left as possible in the event of a weight tie. This constraint is a member of the ALIGN family of constraints (McCarthy and Prince 1993).

(7) ALIGN (σ́, L, PrWd): Stresses are aligned with the left edge of a prosodic word; i.e. stresses must fall on the first syllable; one violation is incurred for every syllable separating a stress from left edge of word.

ALIGN (σ́, L, PrWd) competes with its antithesis, ALIGN (σ́, R, PrWd), which requires that stresses be aligned with the right edge of a prosodic word. The crucial ranking of our three constraints is as in (8).

(8) ONE STRESS ≫ STRESS $[XX]_R$ ≫ ALIGN (σ́, L, PrWd)

With these rankings, we are in a position to consider a few sample tableaux for Yana. First, we consider in (9) a four syllable form with a long vowel in the second and final syllables. This form demonstrates that the relevant STRESS constraint is ranked above ALIGN (σ́, L, PrWd).

(9)

Input suk'oːniyaː	ONE STRESS	STRESS $[XX]_R$	ALIGN (σ́, L, PrWd)
☞ (a) suk'óːniyaː		*	*
(b) suk'oːníyaː		**!	**
(c) suk'oːniyaː	*!	**	
(d) suk'óːniyáː	*!		****
(e) suk'oːniyáː		*	***!
(f) súk'oːniyaː		**!	

Candidates (b) and (f) each incur two violations of the STRESS constraint and are destined not to surface. Candidate (f), with stress on the initial syllable rather than the first long vowel, shows that the STRESS constraint is ranked above ALIGN (σ́, L, PrWd). If the opposite ranking obtained, candidate (f) would incorrectly emerge as the victor. Candidates (c) and (d) are also eliminated from consideration, since they have no stresses, thereby incurring a fatal violation of ONE STRESS. Candidate (e) stresses a syllable with a branching rhyme, but it does not have stress on the branching rhyme closest to the left edge of the word; hence it commits two more violations of the ALIGN constraint than the winning candidate (a).

In tableau (10), a form containing a closed syllable and a diphthong is considered. The surface form contains stress on the first syllable that is closed or contains a long vowel or diphthong.

(10)

Input sibumk'ai	ONE STRESS	STRESS $[XX]_R$	ALIGN (σ́, L, PrWd)
☞ (a) sibúmk'ai		*	*
(b) sibumk'ái		*	**!
(c) síbumk'ai		**!	

Finally, we consider in (11) a form with only light syllables. The initial stress of the winning candidate reflects the ranking of ALIGN (σ́, L, PrWd) over its lower-ranked sister ALIGN (σ́, R, PrWd).

(11)

Input: p'udiwi	ONE STRESS	STRESS $[XX]_R$	ALIGN (σ́, L, PrWd)	ALIGN (σ́, R, PrWd)
☞ (a) p'údiwi				**
(b) p'udíwi			*	*
(c) p'udiwí			*!*	

7 Conclusions

In summary, this chapter has presented evidence from tone and stress systems showing that syllable weight has an important process-specific component that is not modelled in standard theories of weight. The basis for the process-specificity of weight is phonetic: weight-sensitive tone and stress differ in terms of the phonetic dimensions along which they operate. The portion of the rhyme characterised by sonorant energy is the relevant domain for determining weight for tone, whereas weight-sensitive stress is sensitive to the energy profile of the entire rhyme. Because weight-sensitive tone and weight-sensitive stress differ in their phonetic underpinnings, they display different phonological distributions in their weight criteria.

On the level of specific languages, phonetic considerations were also shown to account for cross-linguistic variation in weight criteria for a single process. The language-specific choice in weight criteria for a given phenomenon is linked to language-specific phonetic properties. Many, but not all, of these language-specific phonetic differences can in turn be attributed to differences between languages in other aspects of the phonological system. The model that thus emerges is one in which phonetics and phonology are interleaved. Certain language-specific aspects of the phonology, such as segment inventories and syllable structure, shape the phonetic map against which the phonetic goodness of potential phonological weight criteria is evaluated.

The selection of weight criteria is not completely driven by considerations of phonetic goodness, however. Rather, the set of weight criteria being evaluated is constrained by a notion of phonological simplicity, which filters out more complex weight criteria. This reliance on a combination of phonetic effectiveness and phonological simplicity results in adoption of weight criteria that are both phonetically and phonologically sensible. Finally, the combination of phonetic effectiveness and phonological simplicity shapes the language-specific ranking of a series of constraint sets, each referring to different weight-sensitive phenomena.

References

Adisasmito-Smith, Niken and Abigail C. Cohn (1996). Phonetic correlates of primary and secondary stress in Indonesian: A preliminary study. *Working Papers of the Cornell Phonetics Laboratory* 11: 1–16.

Allen, W. Sidney (1973). *Accent and rhythm*. Cambridge: Cambridge University Press.

Baitschura, Uzbek (1976). Instrumental phonetische Beiträge zur Untersuchung der Sprachmelodie und des Wortakzentes im Tscheremissischen. *Etudes Finno-Ougriennes* 13: 109–22.

Barker, Muhammad Abd-al-Rahman (1964). *Klamath grammar*. (University of California Publications in Linguistics, 32) Berkeley: University of California Press.

Bauer, Winifred (1993). *Maori*. New York: Routledge.

Beckman, Mary (1986). *Stress and non-stress accent*. Dordrecht: Foris.

Boas, Franz (1947). *Kwakiutl grammar with a glossary of the suffixes*, ed. Helene Boas Yampolsky and Zelig Harris. *Transactions of the American Philosophical Society* 37.3: 201–377.

Bosson, James E. (1964). *Modern Mongolian* (Uralic and Altaic Series, 38). Bloomington: Indiana University.

Broselow, Ellen, Susan Chen, and Marie Huffman (1997). Syllable weight: Convergence of phonology and phonetics. *Phonology* 14: 47–82.

Clements, George N. and Samuel Jay Keyser (1983). *CV phonology: A generative theory of the syllable*. Cambridge, MA: MIT Press.

Crowhurst, Megan (1991). Demorification in Tübatulabal: Evidence from initial reduplication and stress. *NELS* 21: 49–63.

Dawson, Willa (1980). *Tibetan phonology*. PhD dissertation, University of Washington.

De Lacy, Paul (1997). Prosodic categorization. Ms Rutgers Optimality Archive, ROA-236, http://roa.rutgers.edu.

Duanmu, San (1994a). Against contour tone units. *LI* 25: 555–608.

Syllabic weight and syllabic duration: A correlation between phonology and phonetics. *Phonology* 11: 1–24.

Flemming, Edward (1995). *Auditory representations in Phonology*. PhD dissertation, University of California, Los Angeles.

Fry, D. B. 1955. Duration and intensity as physical correlates of linguistic stress. *JASA* 27: 765–68.

Gordon, Matthew (1999). *Syllable weight: Phonetics, phonology, and typology*. PhD dissertation, University of California, Los Angeles.

(2001). A typology of contour tone restrictions. *Studies in Language* 25: 405–44

(2002). A phonetically-driven account of syllable weight. *Lg* 78: 51–80.

Hayes, Bruce (1988). Metrics and phonological theory. In F. Newmeyer (ed.), *Linguistics: The Cambridge survey*, vol. 2, *Linguistic theory: extensions and implications*. Cambridge: Cambridge University Press.

(1989). Compensatory lengthening in moraic phonology. *LI* 20: 253–306.

(1995). *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

(1999). Phonetically-driven phonology: The role of Optimality Theory and inductive grounding. In Michael Daswell, Edith Moravcsik, Michael Noonan, Frederick Newmeyer, and Kathleen Wheatley (eds.), *Functionalism and formalism in linguistics, volume I, general papers.* Amsterdam: John Benjamins, 243–85.

Hock, Hans (1986). Compensatory lengthening: In defence of the concept 'Mora'. *Folia Linguistica* 20: 431–60.

House, David (1990). *Tonal perception in speech*. Lund: Lund University Press.

Hyman, Larry M. (1985). *A theory of phonological weight*. Dordrecht: Foris.

(1992). Moraic mismatches in Bantu. *Phonology* 9: 255–66.

Jakobson, Roman (1931). Die Betonung und ihre Rolle in der Wort- und Syntagma-Phonologie. In *Roman Jakobson, Selected Writings*, vol. 1., The Hague: Mouton, 117–36.

Jassem, W., J. Morton, and M. Steffen-Batóg (1968). The perception of stress in synthetic speech-like stimuli by Polish listeners. *Speech Analysis and Synthesis* 1: 289–308.

Kager, René (1990). Dutch schwa in moraic phonology. *CLS* 26 *Parasession on the syllable in phonetics and phonology*, 241–56.

Kao, Diana (1971). *Structure of the syllable in Cantonese*. The Hague: Mouton.

Kenstowicz, Michael (1997). Quality sensitive stress. *Rivista di Linguistica* 9: 157–88.

Kirchner, Robert (1998). *An effort-based approach to consonant lenition*. PhD dissertation, University of California, Los Angeles. Rutgers Optimality Archive 276, http://roa.rutgers.edu.

Levin, Juliette (1985). *A metrical theory of syllabicity*. PhD dissertation, MIT.

Liljencrants, Johan and Björn Lindblom (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Lg* 48: 839–62.

Lindblom, Björn (1986). Phonetic universals in vowel systems. In J. Ohala and J. Jaeger (eds.), *Experimental phonology.* New York: Acedemic Press, 13–44.

Maddieson, Ian (1978). Universals of tone. In Joseph Greenberg, Charles Ferguson, and Edith Moravcsik (eds.), *Universals of human language*, vol. 2, *Phonology*, Stanford: Stanford University Press, 335–65.

McCarthy, John (1979). *Formal problems in Semitic phonology and morphology*. PhD dissertation, MIT.

McCarthy, John and Alan Prince (1986). *Prosodic morphology*. Ms., University of Massachusetts and Brandeis University.

(1990). Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language and Linguistic Theory* 8: 209–83.

(1993). *Generalized alignment*. Ms., University of Massachusetts and Rutgers University. Rutgers Optimality Archive 7, http://ruccs.rutgers.edu/roa.html.

(1995). Prosodic morphology, In J. Goldsmith (eds.) *Handbook of phonological theory.* Cambridge, MA: Blackwell, 318–66.

Mester, Armin (1994). The quantitative trochee in Latin. *NLLT* 12: 1–61.

Meyer-Eppler, W. (1957). Realisation of prosodic features in whispered speech. *JASA* 29: 104–6.

Moore, Brian (1995). *An introduction to the psychology of hearing*. San Diego: Academic Press.

Munro, Pamela and Catherine Willmond (1999). *Chikashshanompa' Kilanompoli'* (Let's speak Chickasaw). Los Angeles: UCLA Department of Linguistics.

Ohala, John (1983). The origin of sound patterns in vocal tract constraints. In Peter MacNeilage (ed.), *The production of speech*, New York: Springer Verlag, 189–216.

Payne, J. (1990). Asheninca stress patterns. In Doris Payne (ed.), *Amazonian linguistics*. Austin: University of Texas Press, 185–212.

Poppe, Nicholas (1951). *Khalkha-Mongolische Grammatik, mit Bibliographie, Sprachproben und Glossar*. Wiesbaden: F. Steiner.

Prince, Alan (1983). Relating to the grid. *LI* 14: 19–100.

Prince, Alan and Paul Smolensky (1993). *Optimality Theory: Constraint interaction in generative grammar*. Rutgers Optimality Archive 537, http://roa.rutgers.edu/.

Reh, Mechthild (1985). *Die Krongo-Sprache (Nìinò-mó-dì): Beschreibung, Texte, Wörterverzeichnis*. Berlin: D. Reimer.

Sadeniemi, Matti (1949). *Metriikkamme perusteet (Bases of our metrics)*. Helsinki: SKST.

Sapir, Edward and Morris Swadesh (1960). *Yana dictionary* (University of California Publications in Linguistics, 22). Berkeley: University of California Press.

Sluijter, A. M. C. and V. J. van Heuven (1996). Spectral balance as an acoustic correlate of linguistic stress. *JASA* 100: 2471–85.

Steriade, Donca (1991). Moras and other slots. *Proceedings of the Formal Linguistics Society of the Midamerica* 1: 254–80.

(1999). Licensing laryngeal features. *UCLA Working Papers in Phonology* 3: 25–146.

Walker, Rachel (1995). Mongolian stress: Typological implications for nonfinality in unbounded systems. *Phonology at Santa Cruz* 4: 85–102.

Warren, R. M. (1970). Elimination of biases in loudness judgements for pure tones. *JASA* 48: 1397–413.

Watkins, Laurel (1984). *A grammar of Kiowa*. Lincoln: University of Nebraska Press.

Wightman, C. W., S. Shattuck-Hufnagel, M. Ostendorf, and P. J. Price (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *JASA* 92: 1707–17.

Wilson, Stephen (1986). Metrical structure in Wakashan phonology. *BLS* 12: 283–91.

Woo, Nancy (1969). *Prosody and phonology*. PhD dissertation, MIT.

Wright, Richard (1996). Tone and accent in Oklahoma Cherokee. In Pam Munro (ed.), *Cherokee papers from UCLA*. Los Angles: UCLA Department of Linguistics, 11–22.

Zec, Draga (1994). *Sonority constraints on prosodic structure*. New York: Garland.

Zhang, Jie (2001). *The effects of duration and sonority on contour tone distribution: Typological survey and formal analysis*. PhD dissertation, University of California, Los Angeles.

Notes

* The author wishes to thank the following people who contributed to the ideas and data presented in this chapter, though they should not be held responsible for any shortcomings that remain: Bruce Hayes, Donca Steriade, Adam Albright, Arto Anttila, Diana Archangeli, Marco Baroni, Taehong Cho, Abby Cohn, Katherine

Crosswhite, Gorka Elordieta, Edward Flemming, Cecile Fougeron, Ed Garrett, Carol Genetti, Steve Greenberg, Mike Hammond, Chai-Shune Hsu, Kathleen Hubbard, Larry Hyman, Sun-Ah Jun, Pat Keating, Michael Kenstowicz, Paul Kiparsky, Robert Kirchner, Jody Kreiman, Paul De Lacy, Peter Ladefoged, Margaret MacEachern, Ian Maddieson, Armin Mester, Pam Munro, David Perlmutter, Janet Pierrehumbert, Doug Pulleyblank, Sharon Rose, Dan Silverman, Sumiko Takayanagi, Motoko Ueyama, Draga Zec, Jie Zhang, Cheryl Zoll, and last but not least, the speakers who provided the phonetic data. Thanks to Bruce Hayes and Robert Kirchner for their many insightful comments on an earlier draft of this chapter.

1. Cantonese tone presents an apparent exception to this hierarchy of weight. In Cantonese, contour tones may occur on CV[+son], CVV[+son], and CV, but not on CV[−son] and CVV[−son]. As it turns out, however, the vowel in phonemic CV is actually phonetically long, while the phonemic long vowel in CVV[−son] is quite short phonetically (Kao 1971; Gordon 1999). These phonetic observations account for the heavy status of CV and the light status of CVV[−son], the two apparent counterexamples to the hierarchy in the text.
2. A voiceless segment does not possess a fundamental frequency itself, although its laryngeal settings may influence the fundamental frequency of neighbouring voiced segments, as demonstrated by the link between voiceless consonants and high tone in many languages. Influences of this sort are not strictly weight sensitive and are thus not discussed here. I further abstract away from differences in perceived pitch attributed to manipulation of acoustic properties other than fundamental frequency (e.g. intensity, spectral shape) in whispered speech (cf. Meyer-Eppler 1957).
3. In fact, there is evidence that, in the three languages in the survey (Hausa, Musey, and Luganda), which superficially appear to treat CV[−son] as heavier than CV, the coda obstruent in CV[−son] does not actually support tonal contrasts on the surface (see Gordon 2001 and Zhang this volume, 2001).
4. Note that, for languages with greater than a binary weight distinction for stress, all distinctions are included in table 9.2; for example, the Chickasaw stress system, which observes the hierarchy CVV > CVC > CV, contributes one token to the CVV heavy row and one to the CVV, CVC heavy row.
5. The representation of short-central vowels is a difficult issue, and one that is logically tied into the matter of how central vowels should be represented. In languages in which centralised vowels are light, they are characteristically quite short (see discussion in Gordon 1999), and thus can plausibly be assumed to lack a timing position. Given this approach, the only cross-linguistic variation in the number of timing positions assigned to segments is between languages in which central vowels are very short in duration and those in which they are not. In languages in which central vowels are very short, they do not carry a timing position. In languages in which central vowels are not particularly short, they are associated with a timing position. In this way, the representation of central vowels is predictable based on duration.
6. Kara (De Lacy 1997), in fact, makes a weight distinction between long low vowels and other rhymes. Because the only long vowel in Kara, however, is /a:/, this weight distinction can be expressed as a simple distinction like that in Khalkha between CVV and other syllables; it is thus phonologically simple in Kara.

7. There is one attested weight distinction that is predicted to be complex by the complexity metric developed thus far: Asheninca, which observes the weight hierarchy CVV > Ca(C), Ce(C), Co(C), CiC > Ci > Cɨ in its stress system (Payne 1990). What is problematic is the uniform weight of CiC and all syllables containing a nonhigh vowel to the exclusion of Ci, a distinction that appears to require disjoint representations. Following Kager's (1990) analysis of non-moraic central vowels (see text), I assume that /i/ in open syllables in Asheninca lacks a weight unit. The weight distinction between heavy Ca(C), Ce(C), Co(C), CiN and light Ci is thus represented as in (f) in figure 9.3, where all heavy syllables have at least one weight unit. There is evidence that /i/ is phonetically very short in open syllables in Asheninca and therefore plausibly does not carry its own weight unit. First, /i/ in open syllables is centralised after coronal stridents, suggesting that /i/ is very short and that this durational reduction prevents the tongue from reaching the peripheral high front target required for the canonical realisation of [i]. The result is a hypoarticulated central vowel, which is so short that it is acoustically deleted between voiceless consonants, e.g. ʃit͡sa → ʃt͡sa 'intestinal worm' (Payne 1990: 190). /i/ is not centralised in closed syllables, suggesting that, unlike /i/ in open syllables, it is long enough for the tongue to reach a more forward articulation. Furthermore, a fast speech process deleting secondary stressed /i/ (and not other vowels) in open syllables adjacent to a heavier syllable also suggests that Ci is extremely short (Payne 1990: 201).
8. One might ask why weight distinctions referring to place features are discriminated against from a simplicity standpoint. It is plausible that the reason why place features are penalised more than other types of predicates stems from considerations related to the size of the hypothesis space being tested by learners of a weight system. Place features are distinctive for all segments in a syllable rhyme, including both vowels and consonants. Non-place features (e.g. manner and voicing features), on the other hand, are typically redundant for at least the vowel in a syllable. Thus, the set of possible contrasts in place of articulation for the entire rhyme is larger than the set of possible contrasts in manner or voicing. For this reason, it is plausible that the discrimination against place features by the complexity metric merely reflects an attempt to reduce the hypothesis space of the learner.
9. The lower the Wilkes' lambda value, the greater the amount of variance in the data attributed to differences between the heavy and light syllables and the less the amount of variance attributed to differences among members of the heavy group or the light group. Because the Wilkes' lambda values are affected by factors such as sample size which are not claimed to be relevant to the hypothesis examined here, they were not used as the definitive criterion for ranking weight distinctions in order of phonetic effectiveness; rather, as pointed out in the text, mean values were used to rank the relative phonetic effectiveness of distinctions.
10. /h/ also appears in coda position in Finnish; it is unclear whether it should be treated as a sonorant or an obstruent.
11. An alternative view, not inconsistent with the data presented in this chapter, is that evaluation of complexity is innate and that constraints referring to complex distinctions do not even exist in the grammar.

10 Consonant lenition

Robert Kirchner

1 Introduction

The term 'lenition' (< L. *lenis* 'weak') refers to synchronic alternations, as well as diachronic sound changes, whereby a sound becomes 'weaker', or where a 'weaker' sound bears an allophonic relation to a 'stronger' sound. An explicit, unified characterisation of this 'weakening' has been a vexed question of phonological theory (Bauer 1988); but the core idea, as applied to consonants, is some *reduction in constriction degree or duration*. The term thus uncontroversially includes:

- *degemination*, i.e. reduction of a long (geminate) to a short (singleton) consonant (e.g. tː → t);
- *flapping*, i.e. reduction of a stop to a flap (e.g. t → ɾ);
- *spirantisation*, i.e. reduction from a stop (or affricate) to a fricative or approximant continuant (e.g. t → {θ, θ̞});
- reduction of other consonants to approximants (e.g. r → ɹ, s → s̞);
- *debuccalisation*, i.e. reduction to a laryngeal consonant (e.g. t → ʔ, s → h);
- and, at its most extreme, complete *elision* (e.g. t → Ø).

Voicing (e.g. t → d), although ostensibly involving an adjustment in laryngeal specification rather than reduction of constriction, is also standardly included in this typology. This traditional classification is, I believe, justified, for at least two reasons: (a) the pattern of voicing is similar to that of other lenition processes in terms of its contexts and conditions; and (b) voicing does in fact conform to the constriction reduction characterisation above, upon a closer examination of the articulatory implementation of voiced vs voiceless consonants.

Despite the pervasiveness of lenition in natural language sound systems, this class of processes has received relatively little attention in the theoretical literature. In particular, previous treatments have failed to offer an empirically adequate, unified formal characterisation of lenition, or to account for the contexts in which lenition typically occurs. Let us briefly consider the two most standard approaches.

First, *autosegmental feature spreading* treatments have been proposed (e.g. Harris 1984, handling Spanish spirantisation as [+continuant] spreading; cf.

Mascaró 1983; Jacobs and Wetzels 1988; see Selkirk 1980; Mascaró 1987; Cho 1990; Lombardi 1991 for similar treatments of voicing). But feature spreading cannot be extended to lenition generally, for degemination, debuccalisation, and elision can only be expressed in autosegmental theory as deletion or delinking of phonological material. Moreover, this approach fails to give a natural account of the most typical lenition context, viz. intervocalic position: it suffices to spread the relevant feature from *either* adjacent vowel, and so the role of the other vowel in conditioning the lenition is unexplained.

An alternative approach, often tentatively suggested (e.g. Foley 1977; Churma 1988; Clements 1990; Hock 1991; Ní Chiosáin 1991; Elmedlaoui 1993; Lavoie 1996), but rarely fleshed out in explicit analyses, is the notion of lenition as *sonority promotion.* But if we take the sonority scale (e.g. stops > voiceless fricatives > voiced fricatives > nasals > liquids > high vowels/glides > low vowels (Dell and Elmedlaoui 1985)) seriously as a characterisation of lenition, we incorrectly predict, for example, that fricatives can lenite to nasals. Moreover, the class of vowel reduction processes (see Flemming and Crosswhite's chapters), which would appear to be the vocalic counterpart of consonant lenition, typically involves *raising* (and centralisation), for example a → ə; but the higher the vowel, the *less* sonorous it is. Finally, the sonority promotion proposal says nothing per se about the contexts and conditions under which lenition naturally occurs.

Following Stampe (1972), Lindblom (1983, 1990), Hock (1991), Boersma (1998), and others, I proceed from the intuition that lenition is driven by an imperative to minimise articulatory effort. Unlike standard approaches, however, I argue that lenition patterns arise *directly* from this effort minimisation constraint (which I style Lazy), interacting with preservation of perceptual distinctions, within an Optimality-Theoretic grammar (see Jun 1995, this volume, Boersma 1998, and Flemming 2002 for deployment of versions of this effort-minimisation constraint in analyses of other sorts of phonological and phonetic patterns). Formally, I assume that for each candidate provided by GEN (the Optimality-Theoretic candidate generating function) the *effort cost* (a mental estimate of the biomechanical energy[1] required for articulatory production of the candidate) is computed; and Lazy violations are assessed for the candidate based on this effort cost. Spirantisation, for example, is analysed in terms of rankings where Lazy dominates preservation of an input [–continuant] specification[2] (1a); under the opposite ranking (b), spirantisation is blocked:

(1) a.

	/d/	Lazy	Pres(cont)
	d	**!	
☞	ð	*	*

b.

	/d/	Pres(cont)	Lazy
	d		**
☞	ð	*!	*

This assumes that stops, ceteris paribus, are more effortful than continuants, due to the greater distance that the articulator must travel in the former. More generally, lenition receives a unified formal treatment under this approach, in terms of the ranking schema LAZY ≫ *lenition-blocking constraint*: the type of structural change occurring in a given language depends upon which of the lenition-blocking constraints, if any, are ranked below LAZY.

The common restriction of lenition to intervocalic position likewise receives a straightforward effort-based treatment: the more open the flanking segments, the greater the displacement (hence effort) required to achieve a given degree of consonantal constriction (cf. Grammont 1933; de Jong et al. 1992; Bybee 2001). The primacy of intervocalic position as a context for lenition thus falls out from the natural assumption that the impetus to lenite more effortful gestures is stronger than the impetus to lenite easier gestures. This result can be captured formally by decomposing the LAZY constraint into a series of effort thresholds (i.e. LAZY_n = 'do not expend effort $\geq n$') and interleaving lenition-blocking constraints within this series. Assuming, for example, that the effort required for a [b] in intervocalic position is at least x, and the effort required in post-consonantal position is at most y, ceteris paribus $x > y$, hence LAZY(x) ≫ LAZY(y), and spirantisation of /b/ in intervocalic position can be obtained by the following ranking:

(2)

		LAZY(x)	PRES(cont)	LAZY(y)
	b /V__V	*!		*
☞	β /V__V		*	
☞	b /C__V			*
	β /C__V		*!	

Note, however, that for cases of *complementary distribution*, for example no word-initial fricatives and no non-initial stops, the use of faithfulness as the lenition-blocking constraint (in this case, positional faithfulness to continuancy, for word-initial position) is insufficient:

(3)

		PRES(cont/#__)	LAZY	PRES(cont)
☞	a. #ka → #ka		**	
	#ka → #xa	*!	*	*
	b. #aka → #aka		**!	
☞	#aka → #axa		*	*
	c. #xa → #ka	***!**	******	*****
☞	**#xa → #xa**		*****	
	d. #axa → #aka		**!	*
☞	#axa → #axa		*	

If, as in (3c), some word-initial obstruent is underlyingly [+cont] (and the OT tenet of Richness of the Base (Prince and Smolensky 1993, ch. 9) prevents us from excluding such an input), both faithfulness and LAZY favour the fricative candidate; thus it is impossible to rule out word-initial fricatives. An additional class of lenition-blocking constraints is therefore required: these must not only block lenition, but also actively induce fortition, for example requiring word-initial obstruents to be realised as stops (*[+cont,−son]/#__). It seems plausible that these fortition constraints are, like the positional faithfulness constraints (Steriade 1999), grounded in perceptual considerations. For example, the release burst of a stop contains salient place of articulation cues (e.g. Wright, this volume); thus, by militating in favour of consonants with a release burst, this constraint can be viewed as enhancing the perceptibility of the consonant; and the allocation of more robust cues to word-initial position may be viewed as reflecting the greater importance of word-initial consonants in lexical access. The precise formulation of the fortition constraints, however, and their relation to broader perceptual considerations, is not central to the thrust of this chapter; for a more thorough treatment of perceptual enhancement in phonology, see Flemming, this volume.

In the remainder of this chapter, I further develop and illustrate this approach, with an analysis of Florentine Italian lenition, traditionally referred to as 'Gorgia Toscana'. The analysis must be regarded as tentative, in that it relies upon certain assumptions concerning the relative effort cost of particular articulatory gestures: ultimate confirmation or falsification thus awaits a programme of articulatory experimentation and modelling to establish a more objective quantification of effort. Nevertheless, it will become obvious, as we examine these data, that lenition indeed constitutes a unified phenomenon, not an arbitrary collection of unrelated processes, thus motivating a unified treatment such as the effort-based approach affords. I further show that the contexts for lenition (both obligatory and stylistically conditioned) are elegantly characterised in terms of articulatory effort thresholds. Finally, the Florentine data exemplify a number of generalisations concerning lenition typology (documented in Kirchner 1998, 2000), based on a survey of 272 languages containing lenition patterns:

(4) (i) All else being equal, lenition occurs more readily the greater the openness of the flanking segments (the widely attested pattern of intervocalic lenition being a special case).
(ii) All else being equal, lenition occurs more readily the faster or more casual the speech.
(iii) Unaffricated stops never synchronically spirantise to strident (sibilant or labiodental) fricatives.

(iv) Geminate stops never undergo voicing or reduction of oral constriction unless they concomitantly degeminate.

As I will show, all these generalisations fall out from the proposed effort-based approach, coupled with some plausible phonetic assumptions concerning the relative effort required for particular types of gestures.

2 Data

The facts are principally drawn from Giannelli and Savoia's (1979) study of the Florentine consonant system and its relation to neighbouring dialects (henceforth 'G&S').

2.1 Obligatory processes

In intervocalic position (with an optional intervening liquid or glide, henceforth 'quasi-intervocalic position'),[3] the Florentine voiceless stops (/p, t, k/) obligatorily lenite:

(5)			Orthography	Gloss
	a.	kaɸo	*capo*	'head'
		plastiha / la ɸlastiha	*(la) plastica*	'(the) plastic'
		pentola / la ɸentola	*(la) pentola*	'(the) pot'
		pɾeso / el anno ɸɾeso	*(l'hanno) preso*	'(they have) taken'
		pjɛna / ll ɛɾa ɸjɛna	*(era) piena*	'(s/he was) full'
	b.	pɾaθo	*prato*	'meadow'
		pjɛθɾa	*pietra*	'stone'
		tjɛne / e lo θjɛne	*(lo) tiene*	'(s/he) has (it)'
		tavola / la θavola	*(la) tavola*	'(the) table'
		tɾave / la θɾave	*(la) trave*	'beam'
	c.	amixo	*amico*	'friend'
		pɔxo	*poco*	'little'
		biʃixletta	*bicicletta*	'bicycle'
		kasa / la xasa	*(la) casa*	'(the) house'
		koltɛlli / i xoltɛlli	*(i) coltelli*	'(the) knives'
		kwattɾini / i xwattɾini	*i quattrini*	'(the) money'

(Indeed, G&S report that many Florentine speakers have difficulty producing voiceless stops in this position, when attempting to imitate Standard Italian.)

The *voiced* velar stop obligatorily lenites in this position as well:

(6)	gamba / la ɣamba	*(la) gamba*	'(the) leg'
	gjanda / la ɣjanda	*(la) ghianda*	'(the) acorn'
	gɾatːa / e si ɣɾatta	*(si) gratta*	'(it) scratches'
	e seɣa	*sega*	's/he mows'
	i ssuɣo	*il sugo*	'the juice'

Likewise, the affricates, /tʃ/ and /dʒ/ obligatorily spirantise in this position:

(7)	a.	peʃe	*pece*	'pitch'
		baʃo	*bacio*	'kiss'
		paʃe	*pace*	'peace'
		tʃena / la ʃena	*(la) cena*	'dinner'
	b.	e ɾiʒetta	*rigetta*	'reject'
		faʒano	*fagiano*	'pheasant'
		dʒoɾni / i ʒoɾni	*(i) giorni*	'(the) days'
		dʒɔxa / e ʒɔxa	*gioca*	's/he plays'

Note that this spirantisation, as it applies to the affricates, is neutralising: for example, *la cena* ('the dinner') and *la scena* ('the scene') are both realised as [la ʃena].

2.2 *Variable processes in quasi-intervocalic position*

However, the foregoing pattern is inextricable from a much broader pattern of *variable* lenition. That is, although /p, t, k, ɡ, tʃ, dʒ/ obligatorily lenite, as described above, the *degree* of lenition varies along a scale, from close fricatives all the way to Ø, depending on speech rate and register. Moreover, most of the other consonants in the inventory likewise undergo some kind of lenition in quasi-intervocalic position, again depending on rate and register. The larger picture of obligatory and variable lenition in Florentine is conveyed by table 10.1 where levels A through K represent a conflation of speech rate and register factors: level A corresponds to the slowest, most careful speech style; B is somewhat faster or more casual (or a modicum of both); and so on, up to K, the fastest, most careless level.[4] In a nutshell, the lower the register, or the higher the rate, the more extreme the lenition. Note that table 10.1 describes the patterns among older speakers: among younger speakers /k/ lenites directly to [h] in natural speech (say at level C). Further note that, in all the foregoing

Table 10.1. *Lenition variation in quasi-intervocalic position, according to rate/register*

	A	B	C	D	E	F	G	H	I	J	K
/k/	x̝	x	x̞			h				ɰ	Ø
/t/	θ̝	θ	θ̞						h	ð̞	Ø
/p/	ɸ̝	ɸ	ɸ̞							β̝	Ø
/g/	ɣ̝	ɣ	ɰ					h			Ø
/b, d/	b,d		β̝, ð̝	β̞, ð̞					ɦ		Ø
/tʃ,dʒ/	ʃ,ʒ				ʃ̞ , ʒ̞					ʒ̞	
/v/	v				ʋ						Ø
/ʃ,s,f/	ʃ,s,f				ʃ̞,s̞,f̞					ʒ̞,z̞,ʋ	
/m, n/	m,n		β,ĩ								Ø
/ɾ,rr,l/	ɾ,rr,l				ɹ, ɹɹ, l̞						

examples of spirantisation, the output is never strident, except in the case of the affricates (/tʃ, dʒ/ → [ʃ, ʒ]), where the inputs are already strident. This restriction on spirantisation outputs appears to be universal, as noted in generalisation (4iii) above (for documentation of this generalisation, see Kirchner 1998: ch. 4).

2.3 Lenition outside of quasi-intervocalic position

Furthermore, in fast/casual speech styles, the environment for lenition expands beyond quasi-intervocalic position:

(8)		Slow/Careful	Fast/Casual		
	a.	[ɪ kuʃilo	[ɪ x̞uʃilo	*cucilo*	'sew it'
	b.	[ɪ tjɛni	[ɪ θ̞ç̞ɛni	*tieni*	'you have it'
		la ʃesta	la ʃesθa	*la cesta*	'the basket'
	c.	[ɪ pɔɾtalo	ɸɔɾtalo	*portalo*	'carry it'
		e lo spɛɾo	e lo sɸ̞ɛɾo	*lo spero*	'I hope'

Affricates likewise frequently spirantise outside of quasi-intervocalic position in natural speech styles; voiced stops and nasals reduce somewhat more rarely. In sum, the lenitions that are obligatory (or which apply in all but the most careful speech styles) in quasi-intervocalic position can apply in other positions as well, but only in fast or casual speech; while the lenitions that are restricted to fast/casual speech in quasi-intervocalic position fail to apply outside of this position. It is as if, outside of quasi-intervocalic position, the set of reductions shifts, across the board, several rate/register levels to the right, relative to their position in table 10.1.

2.4 The behaviour of geminates

A final wrinkle concerns the (near-)immunity of geminate consonants to these lenition processes. Geminate stops can spirantise (to very close fricatives), but only at the fastest rate of speech, in intervocalic position.

(9)	Slow/Careful – Fast/Casual (levels A – J)	Extremely Fast/ Careless (level K)	
/bɾutto/	[bɾutto]	[bɾuθ̞θ̞o]	‘ugly’
/fɾeddo/	[fɾeddo]	[fɾeð̞̆ð̞̆o]	‘cold’

Unfortunately, G&S give no data on the duration of these spirantised geminates. But, as this is a very-fast-speech phenomenon, their duration would almost certainly be much shorter than in slow or moderate rates: that is, they are probably no longer phonetically ‘long’.[5] G&S do transcribe these spirantised segments as geminates; but this is quite compatible with the phonetic shortening claim, because cues other than consonant duration remain to distinguish the geminates from the singletons: (a) the shortened duration and reduced quality of the vowel that precedes the geminate (cf. Bertinetto 1981; Smith 1992); (b) in the case of the voiceless geminates, a somewhat aspirated release; and (c) an interval of more reduced acoustic energy compared to corresponding singletons, due to the more fortis constriction in the geminates (i.e. the lenited geminates are near stops whereas the lenited singletons are weak approximants or Ø). The Florentine lenition pattern thus exemplifies another typological generalisation (4iv): geminates never spirantise, and obstruent geminates never undergo voicing, unless they concomitantly degeminate, as is documented and more fully discussed in Kirchner 2000; 2001: ch. 5.

3 Effort

In order to develop a concrete effort-based analysis of the foregoing lenition patterns, it is necessary to make explicit assumptions concerning the relative effort costs of the relevant consonants, in the relevant contexts. I do this in terms of a set of inferences, discussed in the remainder of this section, concerning relative effort cost of particular consonant types, which follow largely from the equation of effort with biomechanical energy (e.g. Lindblom 1983). These inferences are expressed as Optimality-Theoretic universal ranking conditions of the form, LAZY(*C*, *K*, *R*) ≫ LAZY(*C'*, *K'*, *R'*), where *C* refers to some class of consonants, *K* to some context, and *R* to some rate of speech. These inferences can only be viewed as a down-payment on a major programme of articulatory experimentation and modelling, to establish a more comprehensive and precise quantification of articulatory energy expenditure (see also Boersma 1998: chs. 2, 7). Nevertheless, this tentative analysis serves to illustrate how a set of phonetically plausible effort relations can be incorporated into a formal phonological account of the lenition patterns of a particular language, and an account of certain typological generalisations concerning lenition. We now proceed to an examination of the particular inferences concerning relative effort cost upon which the analysis is based.

3.1 Effects of constriction degree

The closer (more constricted) the consonant, the greater the effort cost, because of the greater distance that the articulator must travel to reach its constriction target (assuming a more open rest position of the articulator): greater velocity, hence energy, is required to move a mass (in this case, an active articulator) a greater distance in a given amount of time. This observation concerning degree of constriction gives rise to the ranking conditions in (10). For example, the greater degree of constriction for affricates relative to stridents implies that LAZY(vcl_strid_affric, *K*, *R*) must outrank its counterpart LAZY(vcl_strid_fric, *K*, *R*).

(10) LAZY(vcl_strid_affric, *K*, *R*) ≫ LAZY(vcl_strid_fric, *K*, *R*)
LAZY(vcd_strid_affric, *K*, *R*) ≫ LAZY(vcd_strid_fric, *K*, *R*)
LAZY(vcl_stop, *K*, *R*) ≫ LAZY(vcl_clos_fric, *K*, *R*)
LAZY(vcd_stop, *K*, *R*) ≫ LAZY(vcd_clos_fric, *K*, *R*)
LAZY(vcl_fric, *K*, *R*) ≫ LAZY(vcl_approx, *K*, *R*)
LAZY(vcd_fric, *K*, *R*) ≫ LAZY(vcd_approx, *K*, *R*)
LAZY(vcl_clos_fric, *K*, *R*) ≫ LAZY(vcl_fric, *K*, *R*)
LAZY(vcd_clos_fric, *K*, *R*) ≫ LAZY(vcd_fric, *K*, *R*)
LAZY(trill, *K*, *R*) ≫ LAZY(long_vcd_approx, *K*, *R*)
LAZY(flap, *K*, *R*) ≫ LAZY(vcd_approx, *K*, *R*)
LAZY(nas, *K*, *R*) ≫ LAZY(vcd_approx, *K*, *R*)
LAZY(lat, *K*, *R*) ≫ LAZY(vcd_approx, *K*, *R*)
LAZY(vcl_approx, *K*, *R*) ≫ LAZY(vcl_glot_fric, *K*, *R*)
LAZY(vcd_approx, *K*, *R*) ≫ LAZY(vcd_glot_fric, *K*, *R*)

3.2 *Effects of speech rate and register*

The faster the speech rate, the greater the effort cost of a given gesture. Greater velocity, hence energy, is required to achieve a given constriction target in a shorter amount of time.

Lowering of speech register likewise induces an across-the-board shift towards hypoarticulation, that is, more drastic lenition, as we have seen in table 10.1. Register sensitivity might be modelled in terms of register-specific lenition-blocking constraints, with ranking conditions of the form LENITION-BLOCKING-CONSTRAINT/*Reg* ≫ LENITION-BLOCKING-CONSTRAINT/*Reg'*, where *Reg* refers to a higher (more formal) register than *Reg'*.

(11)

		PRES(cont)/*Reg*	LAZY$_x$	PRES(cont)/*Reg'*
☞	a. aka → aka (*Reg*)		*	
	aka → axa (*Reg*)	*!		*
	b. aka → aka (*Reg'*)		*!	
☞	aka → axa (*Reg'*)			*

As shown in (11), a ranking schema where LAZY is interleaved with these register-specific lenition-blocking constraints yields the desired result that a given lenition process may apply in a lower register, but be blocked in a higher register. This treatment captures the intuition that in formal registers of speech, the speaker assigns greater importance to the perceptual needs of the hearer, maintaining contrasts and fortifying segments for the sake of greater clarity. Note, however, that *lower* ranking of low-register lenition-blocking constraints (relative to LAZY) has the same effect as *higher* ranking of lower register-specific LAZY constraints (relative to lenition-blocking constraints).

(12)

		LAZY*x*	PRES(cont)	LAZY*y*
☞	a. aka → aka (*Reg*)	*		*
	aka → axa (*Reg*)		*	
	b. aka → aka (*Reg'*)	*!		*!
☞	aka → axa (*Reg'*)		*	

Register sensitivity of the lenition pattern can thus be modelled in the same way as rate sensitivity, by treating articulatory gestures in lower registers *as though* they had a higher effort cost. Thus, a gesture that counts as sufficiently cost-effective in slow/formal speech may be evaluated as too costly in fast/casual

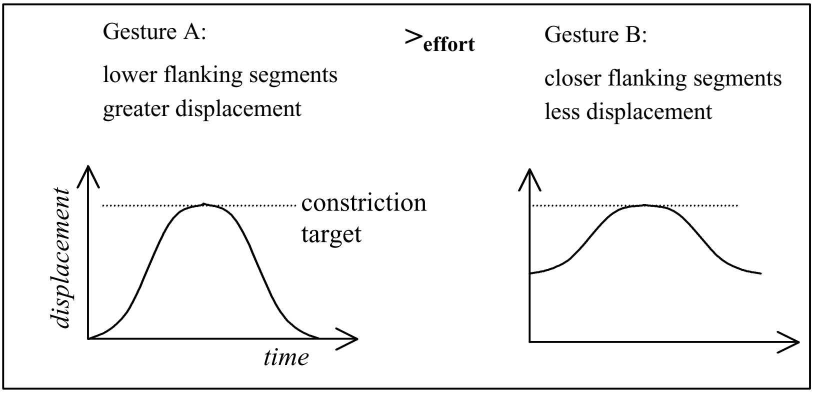


Figure 10.1 Displacement and openness of flanking segments

speech, resulting in lenition. For reasons of expository simplicity, I adopt the latter strategy. This allows us to dispense with a separate hierarchy of register-specific lenition-blocking constraints, instead conflating the treatment of rate- and register-sensitivity in terms of the following ranking conditions; in other words, every constraint must outrank an otherwise identical constraint that governs the next most rapid/casual style. In (13), the letters A–K refer to the conflated rate/register levels identified in table 10.1.

(13) Lazy(*C*, *K*, A) ≫ Lazy(*C*, *K*, B)
Lazy(*C*, *K*, B) ≫ Lazy(*C*, *K*, C)
. . .
Lazy(*C*, *K*, J) ≫ Lazy(C, *K*, K)

3.3 Effects of flanking segments

Ceteris paribus, the more open the segments that flank the consonant, the greater the effort cost. This is because of the greater distance that the articulator must travel to reach a given constriction target, if it must begin from and/or return to a lower position, as schematised in figure 10.1. Simplifying this contextual continuum into G&S's categorical distinction between quasi-intervocalic (which I abbreviate as V_V) vs non-quasi-intervocalic contexts (C_V), I formalise this observation in terms of the ranking condition

(14) Lazy(*C*, V_V, *R*) ≫ Lazy(*C*, C_V, *R*).

Of course, true biomechanically based effort values would distinguish among consonants in the context /a__a/ (maximum displacement, hence maximum propensity to lenite) vs /i__i/ (less displacement, hence less propensity to lenite), vs /i__r/ (even less displacement), and so on, rather than a categorical distinction between quasi-intervocalic position and other contexts. There is in fact some evidence that the actual conditioning of Florentine lenition is closer to the gradient situation than to the categorical distinction. First, G&S present data suggesting that elision of /k/ is more likely to be blocked before high vowels than before nonhigh vowels.[6]

		Frequent	Rarer	Orthography	Gloss
(15)					
	(a)	baa	bax̞a	*baca*	'worm'
		ɸ̞ɔo	ɸ̞ɔx̞o	*poco*	'little'
		ʃɛe	ʃɛhe	*cieche*	'blind-f.pl.'
		bua	bux̞a	*buca*	'hole'
		le ɔɾna	le hɔɾna	*le corna*	'the horns'
		la assetta	la x̞assetta	*la cassetta*	'the box'
		i oltɛlli	i x̞oltɛlli	*i coltelli*	'the knives'
	(b)	i bahini	i βaini	*i bachini*	'the little worms'
		e sɔno x̞inaθ̞i	e sɔno inaθ̞i	*sono chinati*	'they are leaning'

That is, lenition of /k/ is sensitive to the height of the following vowel, as predicted by a true displacement-based account. Secondly, G&S observe that lenition tends to be blocked, or to occur to lesser degree, in the phrase-final foot:

(16) ˈlː ɛɾa ˈʃɛx̞a   *but frequently*   ˈlː ɛɾa ʃɛha ða un ˈɔcːo
era cieca   *era cieca da un occhio*
'she was blind'   'she was blind in one eye'

This phrase-final inhibition of lenition can readily be understood as an effect of the common phonetic process of phrase-final lengthening. That is, a given speech style may be just fast and casual enough to trigger lenition in the 'body' of the phrase; but in the final foot, the speech rate slows down to the point that lenition is blocked. Again, this suggests that the conditioning of lenition is more complex than a categorical distinction between quasi-intervocalic position and other contexts. Rather, it appears that, to the extent G&S discuss these factors, any factor that would raise the effort cost in a given context boosts the probability of lenition occurring in that context. I therefore assume that G&S's categorical distinction between quasi-intervocalic position and other contexts is an idealisation. However, since G&S frame their description of Florentine in

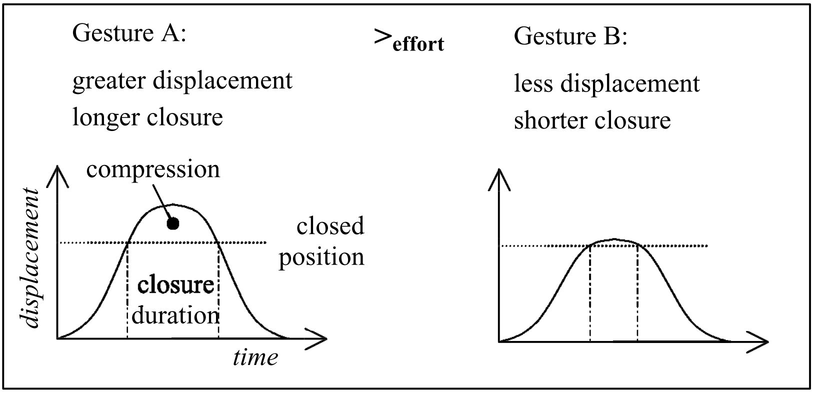


Figure 10.2 Closure duration as displacement

terms of this idealisation, I follow suit in this analysis of their data. The point is, though, that there is reason to believe that if we instead adopted a more biomechanically plausible gradient set of contextual distinctions in effort cost, the accuracy of the analysis would actually improve.

3.4 *Effects of closure duration*

Among non-continuants, the longer the closure, the greater the effort cost (assuming that the longer closure is due to a gesture with higher displacement: displacement above the 'closed' position is interpreted as compression of the active articulator against the passive articulator, as schematised in figure 10.2). This inference gives rise to the following ranking conditions:

(17) Lazy(gem_vcl_stop, *K*, *R*) ≫ Lazy(vcl_stop, *K*, *R*)
Lazy(gem_vcl_affric, *K*, *R*) ≫ Lazy(vcl_affric, *K*, *R*)
Lazy(vcd_stop, *K*, *R*) ≫ Lazy(flap, *K*, *R*)
Lazy(vcl_stop, *K*, *R*) ≫ Lazy(vcd_stop, *K*, *R*)

The last of these conditions is based on the longer (more fortis) closure that typically characterises a voiceless stop vs the shorter (more lenis) closure typical of a voiced stop (see also section 3.5).

3.5 *Aerodynamic effects*

Westbury and Keating (1986) demonstrate that, in utterance-medial position when preceded by a voiced sonorant, singleton stops undergo *passive voicing*, unless they are devoiced by active abduction (or constriction) of the glottis, assuming an adducted rest position of the glottis.[7] Moreover, my own simulations, using the computational model of vocal tract aerodynamics described in Westbury and Keating 1986, show that medial singleton continuant consonants likewise undergo passive voicing, unless actively devoiced. These aerodynamic considerations give rise to the following ranking conditions:

(18) LAZY(vcl_strid_fric, V_V, *R*) ≫ LAZY(vcd_strid_fric, V_V, *R*)
LAZY(vcl_stop, V_V, *R*) ≫ LAZY(vcd_stop, V_V, *R*)
LAZY(vcl_affric, V_V, *R*) ≫ LAZY(vcd_affric, V_V, *R*)
LAZY(vcl_clos_fric, V_V, *R*) ≫ LAZY(vcd_clos_fric, V_V, *R*)
LAZY(vcl_fric, V_V, *R*) ≫ LAZY(vcd_fric, V_V, *R*)
LAZY(vcl_approx, V_V, *R*) ≫ LAZY(vcd_approx, V_V, *R*)
LAZY(vcl_glot_fric, V_V, *R*) ≫ LAZY(vcd_glot_fric, V_V, *R*)

On the other hand, in geminates, the longer closure (or partial constriction in the case of fricatives) results in a build-up of oral pressure, causing passive *devoicing* (see Kirchner 1998/2001: 124–5; cf. Ohala 1983), in any context. Voicing can be extended during a geminate by various cavity expansion gestures, for example pharynx expansion and larynx lowering (Rothenberg 1969)[8]; but such cavity expansion gestures necessarily involve additional effort:

(19) LAZY(gem_vcd_fric *K*, *R*) ≫ LAZY(gem_vcl_fric, *K*, *R*)
LAZY(gem_vcd_affric, *K*, *R*) ≫ LAZY(gem_vcl_affric, *K*, *R*)
LAZY(gem_vcd_stop, *K*, *R*) ≫ LAZY(gem_vcl_stop, *K*, *R*)

3.6 *Effects of precise constriction*

Notwithstanding the general inference in section 3.1 above, that consonants with reduced constriction are less effortful, *strident* (strongly fricated) fricatives and affricates are more effortful than corresponding stops, due to the additional antagonist muscle activation involved in the former. Unlike weakly fricated (or unfricated) continuants, in which the active articulator can approach the passive articulator more or less ballistically (figure 10.3b), *strident* fricatives involve a sustained interval of precise, close constriction (figure 10.3c). Plausibly, achievement of such a constriction requires antagonist muscle activation, in opposition to the (agonist) closure activation, to bring and keep the articulator in close constriction, without allowing it to go all the way to closure, as schematised in figure 10.4. In sibilants in particular, it is known that such antagonism is required: namely, the tongue-blade constriction is partially opposed by a stiffening and bracing of the sides of the tongue against the molar gumline, to produce a grooved channel for the airflow (Ladefoged and Maddieson

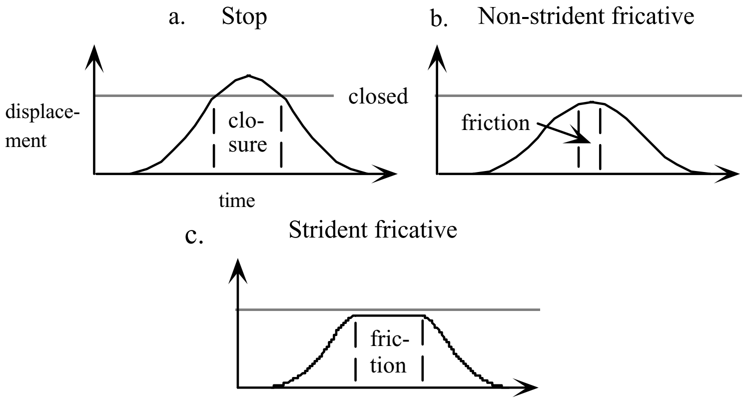


Figure 10.3 Constriction in stop, non-strident, and strident fricative

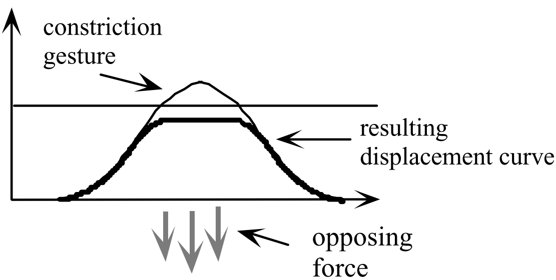


Figure 10.4 Sustained partial constriction achieved by agonist force

1996: 146–7). The effort cost of the sustained partial constriction schematised in figure 10.4, with both agonist and antagonist force, is greater than that required for the corresponding stop (figure 10.3a) (see Kirchner 1998/2001: chs. 2, 4 for further discussion, including a computational simulation of a simplified mass-spring model of consonant constriction that supports this effort claim). Geminate fricatives, a fortiori, are assumed to be more effortful than geminate stops, due to the even longer sustained partial constriction in the former. Geminate fricatives are also assumed to be more effortful than geminate affricates, in that geminate affricates are like singleton affricates, plus lengthened closure. These considerations give rise to the following ranking conditions:

(20) LAZY(vcl_strid_fric, *K*, *R*) $\gg$ LAZY(vcl_stop, *K*, *R*)
LAZY(vcd_strid_fric, *K*, *R*) $\gg$ LAZY(vcd_stop, *K*, *R*)
LAZY(trill, *K*, *R*) $\gg$ LAZY(flap, *K*, *R*)
LAZY(long_vcd_approx, *K*, *R*) $\gg$ LAZY(vcd_approx, *K*, *R*)

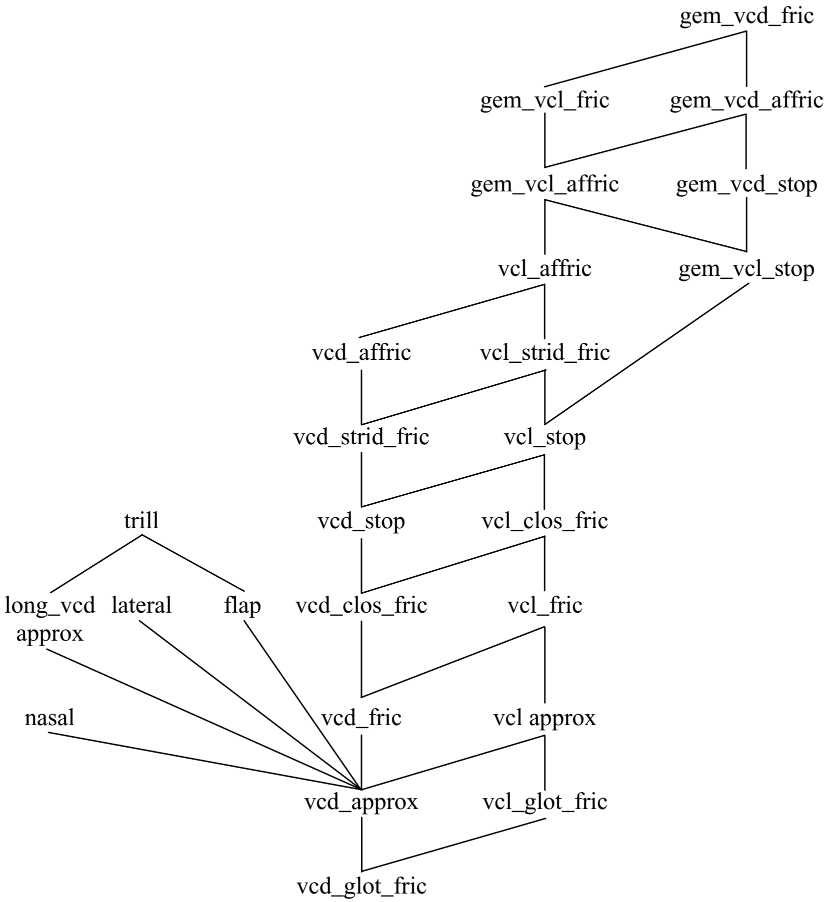


Figure 10.5 Partial Hasse diagram of ranking conditions on Lazy constraints, inferred from considerations of relative articulatory effort, abstracting away from context and rate/register

Lazy(gem_vcd_fric, *K*, *R*) ≫ Lazy(gem_vcd_affric, *K*, *R*)
Lazy(gem_vcl_fric, *K*, *R*) ≫ Lazy(gem_vcl_affric, *K*, *R*)
Lazy(gem_vcl_affric, *K*, *R*) ≫ Lazy(gem_vcl_stop, *K*, *R*)
Lazy(gem_vcd_affric, *K*, *R*) ≫ Lazy(gem_vcd_stop, *K*, *R*)

3.7 Summary

The ranking conditions motivated above can be summarised in terms of the Hasse diagram in figure 10.5 (ignoring context and rate). The complete

Hasse diagram, including contexts and rates, cannot legibly fit on a page, but can be viewed (in a number of formats) at http://www.linguistics.ucla.edu/people/hayes/PBP/.

4 Analysis

Bearing in mind the tentative status of the effort relations inferred above, we can now proceed to an analysis of the lenition patterns described in section 2, in terms of a ranking of particular levels of LAZY relative to faithfulness and fortition constraints, subject to the effort-based ranking conditions motivated in the previous section.

4.1 The stable spirantisations

Voiceless stops. Spirantisation of stops in quasi-intervocalic position at level A is captured in terms of the following ranking:

(21)

		LAZY(vcl_stop, V_V, A)	*−strid, +cont, +cons	PRES(cont)
	p, t, k → p, t, k	*!		
☞	p, t, k → ɸ, θ̞, x̞		*	*

The subordination of PRES(cont) to the fortition constraint, *[−strid, +cont, +cons], ensures that the spirantisation is allophonic, that is, that when considerations of effort do not demand a continuant, namely in the earlier levels in non-quasi-intervocalic position, no nonstrident continuant can surface; rather it must strengthen to a stop, even if underlyingly [+cont] (and, a fortiori, if already [−cont] underlyingly).

(22)

		LAZY(vcl_stop, C_V, A)	*−strid,+cont, +cons	PRES(cont)
☞	ɸ, θ̞, x̞ → p, t, k			*
	ɸ, θ̞, x̞ → ɸ, θ̞, x̞		*!	

The reduction is to a close fricative, rather than an approximant, at this level. Assume a feature [close], which distinguishes the close fricatives [ɸ, θ̞, x̞, β, ð̞, ɣ̞] from the open ones [ɸ, θ, x, β, ð, ɣ]. To block further reduction, I posit another fortition constraint banning [–close] continuant consonants, ranked above the effort cost of the fricatives at this level.

(23)

		*[−clos,+cons, +cont]	LAZY(vcl_clos_fric, V_V, A)	LAZY(vcl_fric, V_V, A)
☞	ɸ, θ̞, x̞		*	*
	ɸ, θ, x	*!		

Voiced stops. The ranking thus far predicts, however, that all the voiced stops should spirantise as well, whereas only /g/ does so. This behaviour might be attributed to a difference in release properties of the various stops. Velar stops are characterised by a noisy release (seen on a spectrogram as multiple bursts), whereas labials and coronal stops typically have a crisp release. Thus, velar stops are somewhat less acoustically distinct from continuants than coronal or labial stops are. I posit the feature [crisp release]: labial and coronal stops are [+crisp rel]; continuants and velar stops are [−crisp rel]. Spirantisation of the labials and coronals, then, is blocked by ranking PRES(crisp rel) above LAZY(vcd_stop, V_V, A).

(24)

		PRES(crisp rel)	LAZY(vcd_stop,V_V, A)	PRES(cont)
	g → g		*!	
☞	g → ɣ̞			*
☞	b, d → b, d		*	
	b, d → β̞, ð̞	*!		*

Assuming LAZY(vcl_stop, V_V, A) ≫ PRES(crisp rel), however, there is no equivalent blocking in the voiceless stops.

Affricates. The low ranking of PRES(cont) further entails spirantisation of the affricates at this level.

(25)

		LAZY(vcl_affric, V_V, A)	LAZY(vcd_affric, V_V, A)	PRES(cont)
	tʃ → tʃ	*!		
☞	tʃ → ʃ			*
	dʒ → dʒ		*!	
☞	dʒ → ʒ			*

Blocking of further reduction is attributable to PRES(strid):

Table 10.2. *Lenition variation in non-quasi-intervocalic position.*

	A	B	C	D	E	F	G	H	I	J	K
/k/	k					x̞	x	x̝			
/t/	t					θ̞	θ	θ			
/p/	p					ɸ̞	ɸ	ɸ̝			
/g/	g					ɣ̞	ɣ	ɰ			
/b,d/	b, d					b,d		β̞,ð̞	β̝, ð̝		
/tʃ,dʒ/t	ʃ,dʒ					ʃ,ʒ				ʃ̝, ʒ̝	
/v/	v					v				ʋ	
/ʃ,s,f/	ʃ,s,f					ʃ,s,f				ʃ̝, s̝, f̝	
/m,n/	m, n					m,n		β,ɹ̃			
/ɾ,rr,l/	ɾ,rr,l					ɾ,rr,l				ɹ,ɹɹ, l̝	

(26)

		Pres(strid)	Lazy(vcl_ strid_fric,V_V,A)
☞	tʃ → ʃ		*
	tʃ → ʃ̝	*!	

Stability of these spirantisations. Since the Lazy constraints for levels B–K are necessarily ranked higher than the corresponding Lazy constraints for level A, the lenition-blocking constraints that are subordinated to Lazy constraints for level A are, by transitivity, also subordinated to Lazy constraints for levels B–K. The foregoing stops and affricates, which spirantise in quasi-intervocalic position at level A, must, a fortiori, spirantise (or lenite even further) in this position at all faster rates and lower registers of speech. Therefore, the spirantisation (or further lenition) of these stops and affricates is obligatory at all levels in quasi-intervocalic position.

4.2 The variable reductions, quasi-intervocalic position

We will now consider the reductions in quasi-intervocalic position, for each of the rate/register levels above A, as summarised in table 10.2.

Level B. At this level, the stops reduce to open rather than close fricatives. Assume a feature [close], which distinguishes the close fricatives [ɸ̝, θ̝, x̝, β̝, ð̝, ɣ̝] from the open ones [ɸ, θ, x, β, ð, ɣ]. The further reduction is obtained by subordinating *[−clos,+cons,+cont] to the effort cost of the close fricatives at this level:

(27)

		Lazy(vcl_clos_fric, V_V, B)	Lazy(vcd_clos_fric, V_V, B)	*[−clos, +cons, +cont]	Lazy(vcl_clos_fric, V_V, A)	Lazy(vcd_clos_fric, V_V, A)
	ɸ, θ̞ , x̞	*!			*	
☞	ɸ, θ, x			*		
	ɣ̞		*!			*
☞	ɣ			*		

The voiced velar in fact reduces all the way to an approximant ([ɰ]), as shown in (28). In the voiceless approximants however, this is blocked at this level, by a fortition constraint requiring the voiceless continuants to be [−son]:

(28)

		*[−voi, +cont, −son]	Lazy(vcl_ fric, V_V, B)	Lazy(vcd_ fric, V_V, B)
☞	ɸ, θ, x		*	
	ɸ̝, θ̝, x̝	*!		
	ɣ			*!
☞	ɰ			

Level C. At this level, the voiceless stops likewise reduce to approximants, /b/ and /d/ spirantise, and the nasals spirantise to nasalised approximants (and /k/ debuccalises in the younger sociolect). Reduction of the voiceless stops to approximants follows from the ranking shown in (29):

(29)

		Lazy(vcl_fric, V_V, C)	*[−voi, +cont, −son]	Lazy(vcl_ fric, V_V, B)
	ɸ, θ, x	*!		*!
☞	ɸ̝, θ̝, x̝		*	

(For the idiolects in which the close approximants are somewhat more prevalent (let us say they occur at level C as well as B, and reduce to open approximants at level D), the ranking would be Lazy(vcl_fric, V_V, D) ≫ *[−voi,+cont,−son] ≫ Lazy(vcl_fric, V_V, C).) Spirantisation of /b, d/ at this level follows from the ranking in (30):

(30)

	LAZY(vcd_stop, V_V, C)	PRES(crisp rel)	LAZY(vcd_stop, V_V, B)
b, d → b, d	*!		*
☞ b, d → β̞, ð̞		*	*

The outcome is a close rather than an open fricative, attributable to disjunctive combination of faithfulness constraints (cf. Smolensky 1995; Kirchner 1996), PRES(crisp rel) ∨ PRES(clos):

(31)

	PRES(crisp rel) ∨ PRES(clos)	LAZY(vcd_ clos_fric,V_V, C)
☞ b, d → β̝, ð̝		*
b, d → β̞ ð̞	*!	

That is, violating LAZY(vcd_ clos_fric, V_V, C) is ranked as worse than violating either PRES(crisp rel) or PRES(clos) individually; but violating both in one step, as in the second candidate above, is worse than violating LAZY(vcd_ clos_fric, V_V, C).

Spirantisation of the nasals at this level follows from the ranking shown in (32):

(32)

	LAZY(nas, V_V, C)	*[+nas,+cont]	LAZY(nas, V_V, B)
m, n → m, n	*!		*
☞ m, n → β̝, ɹ̃		*	

The ranking of *[+nas,+cont] above LAZY(nas, V_V, B) ensures that the spirantisation of the nasals is blocked at earlier levels. Finally, debuccalisation of /k/ in younger idiolects follows from the ranking shown in (33):

(33)

	PRES(cor,lab)	LAZY(vcl_approx, V_V, C)	PRES(dors)
k → x̞		*!	
☞ k → h			*
☞ p, t → ɸ, θ̝		*	
p, t → h	*!		*!

Level D. At this level, the voiced stops spirantise to approximants rather than fricatives. This result follows from the following ranking:

(34)

	Lazy(vcd_fric, V_V, D)	Pres(son)	Lazy(vcd_fric, V_V, D)
b, d → β, ð	*!		*
☞ b, d → β̞, ð̞		*	

Level E. At this level, the strident consonants reduce to nonstrident approximants, and the liquids reduce to approximants. Reduction to nonstridents follows from the ranking shown in (35):

(35)

	Lazy(vcl_strid_fric,V_V, E)	Lazy(vcd_strid_fric,V_V, E)	Pres(strid)
tʃ, s, ʃ, f, → ʃ, s, f	*!		
☞ tʃ, s, ʃ, f → ʃ̞, s̞, f̞			*
dʒ, v → ʃ, v		*!	
☞ dʒ, v → ʒ̞, ʋ			*

Ranking of Pres(strid) above Lazy(vcl_strid_fric, V_V, D) ensures that stridency is maintained at earlier levels. Reduction of the liquids follows from the rankings given below (it is assumed that the reduced /l/ is [+lateral, +continuant], and that reduced /rr/ is [+rho,+long,–trill]).

(36)

	Lazy(lat,V_V, E)	*[+lat,+cont]	Lazy(lat,V_V, D)
l → l	*!		*
☞ l → l̞		*	

(37)

	Pres(long)	Lazy(trill,V_V, E)	*[+rho,+long,−trill]
rr → rr		*!	
☞ rr → ɹɹ			*
rr → ɹ, ɾ	*!		

(38)		Lazy(flap,V_V, E)	*[+rho,−long,−flap]
	ɾ → ɾ	*!	
☞	ɾ → ɹ		*

The fortition constraints, *[+lat,+cont], *[+rho,+long,−trill], and *[+rho, −long,−flap], here serve to block lenition of the liquids at levels D and earlier.

Level F. At this level, /k/ debuccalises in the older sociolect. This follows from the ranking shown in (39):

(39)		Lazy(vcl_approx, V_V, F)	Pres(dors)	Lazy(vcl_approx, V_V, E)
	k → x̞	*!		*
☞	k → h		*	

Level G. For the quasi-intervocalic context under discussion, there are no further reductions at this level; see section 4.3 for other contexts.

Level H. At this level, the voiceless labial debuccalises:

(40)		Lazy(vcl_approx,V_V, H)	Pres(lab)
	p → ɸ̞	*!	
☞	p → h		*

Level I. At this level, /g/ debuccalises to [ʰ] (in the older sociolect). This is captured by the following ranking Lazy(vcd_approx, V_V,I) ≫ Pres(dors) ≫ Lazy(vcd_approx,V_V,H):

(41)		Lazy(vcd_approx, V_V, I)	Pres(dors)	Lazy(vcd_approx, V_V, H)
	g → ɰ	*!		*
☞	g → ɦ		*	

Moreover, /t/ debuccalises:

(42)

		LAZY(vcl_approx, V_V, I)	PRES(cor)	LAZY(vcl_approx, V_V, H)
	t → θ̞	*!		*
☞	t → h		*	

On the other hand, we must prevent debuccalisation of the voiceless approximants derived from stridents [ʃ̞, s̞, f̞], despite the violability, at this level, of all the PRES(place) constraints. This requires another disjunctively combined faithfulness constraint, PRES(strid) ∨ PRES(place):

(43)

		PRES(strid)∨ PRES(place)	LAZY(vcl_approx, V_V, I)	PRES(cor)	PRES(lab)
☞	tʃ, ʃ, s → ʃ̞, s̞		*	*	
	tʃ, ʃ, s → h	*!		*	
☞	f → f̞		*		*
	f → h	*!			

Level J. At this level, all voiceless approximants become voiced:

(44)

		LAZY(vcl_glot_fric,V_V,J)	PRES(voi)
	p, t, k → h	*!	
☞	p, t, k → β̞ ð̞, ɰ		*
	tʃ, ʃ, s, f → ʃ̞, s̞, f̞	*!	
☞	tʃ, ʃ, s, f → ʒ̞, z̞, ʋ		*

Further reduction to [ɦ] (debuccalisation + voicing) is blocked by a disjunctively combined faithfulness constraint, PRES(voi)∨ PRES(place).

(45)

		PRES(voi)∨PRES(place)	LAZY(vcd_approx, V_V,J)
	p, t, k → ɦ	*!	
☞	p, t, k → β̞, ð̞, ɰ		*

Level K. At this final level, the singleton stops and nasals delete.

← +strid | −strid →

{s, ʃ} > {z, ʒ, f} > v > {f, θ, x} > {β, ð, ʋ}

← +more strid | −more strid →

Figure 10.6 Stridency scale.

(46)

		LAZY(vcd_ approx, V_V, K)	LAZY(vcd_glot_ fric, V_V, K)	PRES(seg)
	p, t, k → β̞, ð̞, ɰ	*!		
	p, t, k → ɦ		*!	
☞	p, t, k → Ø			*
	b, d, g → β̞, ð̞, ɰ	*!		
	b, d, g → ɦ		*!	
☞	b, d, g → Ø			*
	m, n → β̞, ɪ̃	*!		
	m, n → ɦ		*	
☞	m, n → Ø			*

For the remaining consonants, elision must be blocked. For the liquids, this result can be obtained by having PRES(lateral) and PRES(rhotic) in undominated position:

(47)

		PRES(lat)	PRES(rho)	LAZY(long_vcd_ approx, V_V, K)	LAZY(vcd_ approx, V_V, K)
☞	l → l̞				*
	l → Ø	*!			
☞	rr → ɹɹ			*	
	rr → Ø		*!		
☞	ɾ → ɹ				*
	ɾ → Ø		*!		

For the stridents, the results are slightly more complicated: /v/ does delete, whereas the other stridents do not. Acoustic stridency, that is, noise intensity, however, is not binary: we can distinguish among degrees of noisiness (cf. Flemming 1995) (see figure 10.6).
Thus, in addition to the standard feature [strident], which draws the line between [v] and [ɸ, θ, x], I posit a feature [more strident], which distinguishes the

sibilants and [f] from [v] and the non-strident fricatives. Blocking of deletion of the [+more strident] fricatives can now be attributed to a disjunctive faithfulness constraint, PRES(+more strident) ∨ PRES(place).

(48)

		PRES(+more strid)∨ PRES(place)	LAZY(vcd_approx, V_V, K)
☞	tʃ, ʃ, ʒ, s, f → ʒ̞, z̞, ʋ		*
	tʃ, ʃ, ʒ, s, f → Ø	*!	
	v → ʋ		*!
☞	v → Ø		

Finally, the geminate stops and affricates, which resist lenition at all earlier levels, succumb to spirantisation (and, I assume, degemination) at this level, reducing to close fricatives. Resistance to spirantisation at level J (and all previous levels) is attributable to high ranking of PRES(long):

(49)

		LAZY(gem_vcl_fric, V_V, J)	PRES (long)	LAZY(gem_vcl_stop, V_V, J)	LAZY(vcl_clos_fric, V_V, J)
☞	pp, tt, kk → pp, tt, kk			*	
	pp, tt, kk → ɸ̞ɸ̞, θ̞θ̞, x̞x̞	*!			*
	pp, tt, kk → ɸ̞, θ̞, x̞		*!		

(Recall from section 3 that spirantisation of a geminate without concomitant degemination *increases* the effort cost.) Spirantisation at level K, then, follows from subordination of PRES(long) to LAZY(gem_vcl_stop, V_V, K):

(50)

		LAZY(gem_vcl_stop, V_V,K)	PRES(long)
	pp, tt, kk → pp, tt, kk	*!	
☞	pp, tt, kk → ɸ̞, θ̞, x̞		*

The ranking LAZY(gem_vcd_stop, V_V, K) ≫ PRES(long) ensures the same result for the geminate voiced stops. Further reduction of the geminates, beyond a close fricative, is blocked by an (undominated) disjunctively combined faithfulness constraint:

(51)

		PRES(long)∨ PRES(clos)	LAZY(vcd_clos_fric, V_V, K)
☞	bb, dd, gg → β̞, ð̞, ɣ̞		*
	bb, dd, gg → Ø	*!	

Summary: quasi-intervocalic position. In the foregoing analysis, the rankings of effort minimisation, faithfulness, and fortition constraints (figure 10.7), in conformity with the effort-based ranking conditions of section 3, yields the pattern above of rate- and register-sensitive reduction in quasi-intervocalic position.

4.3 Lenition outside of quasi-intervocalic position

The behaviour of consonants in non-quasi-intervocalic position follows from the same approach of interleaving lenition-blocking constraints among the LAZY hierarchy imposed by the ranking conditions of section 3. Since the effort cost of consonants in non-quasi-intervocalic position is lower than in quasi-intervocalic position, lenition processes that apply at level A in quasi-intervocalic position can apply only at later levels in non-quasi-intervocalic position; while lenition processes that apply at later levels in quasi-intervocalic position do not apply at all in non-quasi-intervocalic position.

Let us assume that in non-quasi-intervocalic position, voiceless stops spirantise at level F (the mid-point in our rate/register scale): that is, spirantisation is blocked at level E and earlier. This can be captured by ranking LAZY(vcl_stop, C_V, F) ≫ *[−strid,+cont,+cons] ≫ LAZY(vcl_stop, C_V, E).

(52)

		LAZY(vcl_stop, C_V, F)	*−strid, +cont,+cons	LAZY(vcl_stop, C_V, E)
☞	p, t, k *(level E)*			*
	ɸ̞, θ̞, x̞ *(level E)*		*!	
	p, t, k *(level F)*	*!		*
☞	ɸ̝, θ̝, x̝ *(level F)*		*	

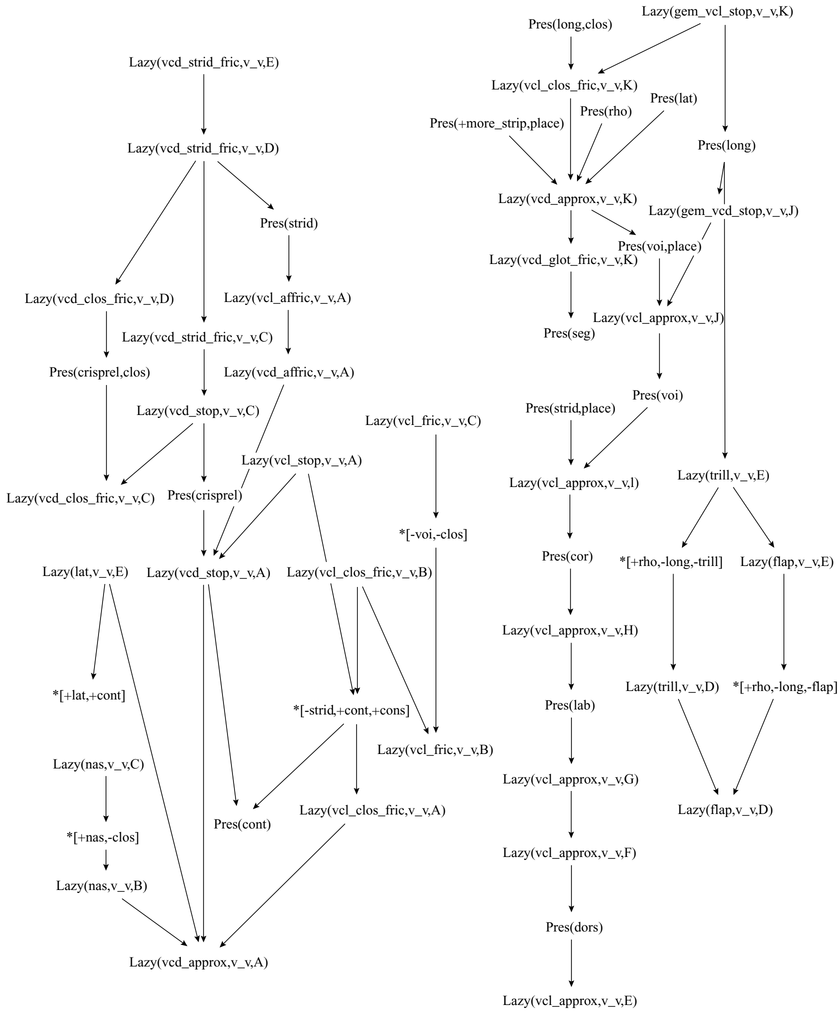


Figure 10.7 Hasse diagram of Lazy relative to Pres and fortition constraints, for Florentine Italian

More generally, substituting Lazy(*x*, C_V, F) for Lazy(*x*, V_V, A), Lazy(*x*, C_V, G) for Lazy(*x*, V_V, B), and so on, in the hierarchy in figure 10.7, we obtain the result that in non-quasi-intervocalic position, the set of processes shift, across the board, several rate/register levels to the right, relative to their behaviour in quasi-intervocalic position (compare tables 10.1 and 10.2). That is, we correctly account for the rate/register-sensitive spirantisation of the stops, affricates, and nasals observed in non-quasi-intervocalic position.

4.4 Capture of typological generalisations

Having seen how this constraint system can be deployed for a single, in-depth analysis, we now turn to typological issues. It is obvious that generalisations (4i) and (4ii), concerning the greater tendency towards lenition in the context of open flanking segments, and in fast/casual speech, are captured under this proposal: as the effort cost of a given consonant constriction gesture rises, due to phonetic context, due to speech rate, or (extrinsically) due to speech register, the impetus to lenite accordingly becomes stronger. We have further noted that the Florentine pattern exemplifies two additional generalisations, (4iii) and (4iv): specifically, the non-stridency of spirantisation outputs; and the immunity of geminates to spirantisation and voicing lenition. Let us now consider how these generalisations follow, automatically, from the effort-based approach, along with certain plausible phonetic inferences outlined above. First, because strident fricatives have a higher effort cost than corresponding stops (section 3.6), no ranking of LAZY with other constraints permits an input stop to map to a strident fricative: the strident incurs a worse violation of LAZY than the input stop.

(53)

	Input: d	LAZY	PRES(cont)
	d	**	
	ð	*	*
👎	z	***	*

(The '👎' indicates that this candidate loses under either ranking.) This result holds true even if we introduce an active fortition constraint, which militates against non-strident fricatives (motivated by the cross-linguistic markedness of weak fricatives such as [ɸ, ð, ɣ]). While such a constraint, if ranked above LAZY, is capable of blocking spirantisation, it cannot cause the strident candidate to emerge as the winner, again because the strident incurs a worse violation of LAZY than the input stop:

(54)

	Input: d	*[+cont,−son,−strid]	LAZY	PRES(cont)
☞	d		**	
	ð	*!	*	*
	z		**!*	*

The strident candidate can only emerge as the winner if the input is already affricated:

(55)

	Input: la tʃena	Pres(strid)	Lazy	Pres(cont)
	la tʃena		****!	
☞	la ʃena		***	*
	la tena	*!	**	
	la jena	*!	*	*

Similarly, given the effort relation *geminate stops > geminate fricatives* (section 3.7), and *voiced geminate obstruents > voiceless geminate obstruents* (section 2.8), geminates cannot spirantise or voice.

(56)

	appa	Lazy	Pres(cont)	Pres(voi)
☞	appa	*		
👎	aɸɸa	**	*	
👎	abba	**		*

Spirantisation (or voicing) can only apply to a geminate if the output concomitantly degeminates:

(57)

	appa	Lazy	Pres(cont)	Pres(length)
	appa	**!		
	aɸɸa	**!*	*	
☞	aɸa	*	*	*

5 Conclusion

We have seen, through the analysis of Florentine lenition, that both stable lenition processes and rate/register-sensitive consonant reduction processes, within and beyond quasi-intervocalic position, can be accounted for in a unified manner, in terms of a consistent ranking of Lazy thresholds versus faithfulness and fortition constraints (again, provisionally assuming the effort values posited in section 3 to be roughly reflective of the relative biomechanical energy costs of the various consonant types). The analysis, moreover, captures the typological generalisations in (4), concerning the greater tendency towards lenition in the context of open flanking segments, and in fast/casual speech, as well as the non-stridency of spirantisation outputs, and the immunity of geminates to spirantisation and voicing lenition. In contrast, within standard approaches, the Florentine lenition pattern cannot be given a unified characterisation: for this set of processes does not correspond to the spreading, addition, or deletion of a particular feature or feature-geometric node; nor can the across-the-board shift towards hypoarticulation at faster rates and lower registers be captured within standard frameworks. Nor do standard approaches afford a natural treatment of lenition in quasi-intervocalic position, nor the expansion of lenition beyond this context at faster rates and lower registers.

References

Bauer, L. (1988). What is lenition? *Journal of Linguistics* 24: 381–392.

Bertinetto, P. (1981). *Strutture prosodiche dell'italiano: Accento, quantità, sillaba, giuntura, fondamenti metrici*. Florence: Presso l'Accademia della Crusca.

Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Cho, Young-mee Yu. (1990). *Parameters of consonantal assimilation*. PhD dissertation, Stanford University.

Churma, D. (1988). On 'On geminates': An alternative approach to inalterability. Ms., SUNY-Buffalo.

Clements, G. (1990). The role of the sonority cycle in core syllabification. In J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the grammar and physics of speech*. Cambridge: Cambridge University Press, 283–333.

Dell, F. and M. Elmedlaoui (1985). Syllabic consonants and syllabification in Imdlawn Tashlhiyt Berber. *Journal of African Languages and Linguistics* 7: 105–30.

Elmedlaoui, M. (1993): Gemination and spirantisation in Hebrew, Berber, and Tigrinya: A fortis-lenis module analysis. *Linguistica Communicatio* 5.1: 121–76.

Flemming, E. (1995). *Auditory representations in phonology*. PhD dissertation, University of California, Los Angeles.

(2002). *Auditory representations in phonology*. New York: Routledge.

Foley, J. (1977). *Foundations of theoretical phonology*. Cambridge: Cambridge University Press.

Giannelli, L. and L. Savoia (1979). Indebolimento consonantico in Toscana. *Revista Italiana di Diallettologia* 2: 23–58.

Grammont, M. (1933). *Traité de phonétique*. Paris: Delgrave.

Harris, J. (1984). La espirantización en castellano y su representación fonológica autosegmental. In Bartra et al. (eds.), *Estudis grammaticals*. Bellaterra: Universitat Autònoma de Barcelona.

Hock, H. (1991). *Principles of historical linguistics*, 2nd edn. Berlin: Mouton de Gruyter.

Jacobs, H. and L. Wetzels (1988). Early French lenition: A formal account of an integrated sound change. In H. van der Hulst and N. Smith (eds.), *Features, segmental structures and harmony processes*. Dordrecht: Foris, 105–29.

Jong, K. de, M. Beckman, and J. Edwards (1993). The interplay between prosodic structure and coarticulation. *Language and Speech* 35: 45–58.

Jun, Jongho (1995). *Perceptual and articulatory factors in place assimilation: An optimality theoretic approach*. PhD dissertation, University of California, Los Angeles. [http://yu.ac.kr/∼jhjun/].

Kirchner, R. (1996). Synchronic chain shifts in Optimality Theory. *LI* 27: 341–50.

(1998). *An effort-based approach to consonant lenition*. PhD dissertation, University of California, Los Angeles. Rutgers Optimality Archive 276, http://roa.rutgers.edu.

(2000). Geminate inalterability and lenition. *Lg* 76: 509–45.

(2001). *An effort approach to consonant lenition*. New York: Routledge.

Ladefoged, P. and I. Maddieson (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Lavoie, L. (1996). Consonant strength: Results of a data base development project. *Working Papers of the Cornell Phonetics Laboratory* 11: 269–316.

Lindblom, B. (1983). Economy of speech gestures. In P. MacNeilage (ed.), *The production of speech*. New York: Springer Verlag, 217–45.

(1990). Explaining phonetic variation: A sketch of the H&H theory. In W. Hardcastle and A. Marchal (eds.), *Speech production and speech modeling*. Dordrecht: Kluwer, 403–39.

Lombardi, L. (1991). *Laryngeal features and laryngeal neutralization*. PhD dissertation, University of Massachusetts, Amherst.

(1998). Evidence for Max-Feature. Ms., University of Maryland, College Park.

Mascaró, J. (1983). Continuant spreading in Basque, Catalan, and Spanish. In M. Aronoff and R. Oehrle (eds.), *Language sound structure*. Cambridge, MIT Press, 287–98.

(1987). A reduction and spreading theory of voicing and other sound effects. Ms., Universitat Autònoma de Barcelona.

Nespor, M. and I. Vogel (1982). Prosodic domains of external sandhi rules. In H. van der Hulst and N. Smith (eds.), *The structure of phonological representations (part I)*. Dordrecht: Foris, 225–55.

Ní Chiosáin, M. (1991). *Topics in the phonology of Irish*. PhD dissertation, University of Massachusetts, Amherst.

Ohala, J. (1983). The origin of sound patterns in vocal tract constraints. In P. MacNeilage (ed.), *The production of speech*. New York: Springer Verlag, 189–216.

Ojamaa, K. (1976). *Temporal aspects of phonological quantity in Estonian*. PhD dissertation, University of Connecticut.

Pind, J. (1999). Speech segment durations and quantity in Icelandic. *JASA* 106: 1045–53.

Prince, Alan and Paul Smolensky (1993). *Optimality Theory: Constraint interaction in Generative Grammar*. Rutgers Optimality Archive 537, http://roa.rutgers.edu/.

Selkirk, E. (1980). Prosodic domains in phonology: Sanskrit revisited. In M. Aronoff and R. Oehrle (eds.), *Language sound structure*. Cambridge: MIT Press, 107–29.

Smith, Caroline (1992). *The temporal organization of vowel and consonant gestures*. PhD dissertation, Yale University.

Smolensky, Paul (1995). On the internal structure of the constraint component *CON* of UG. Rutgers Optimality Archive 86, http://roa.rutgers.edu.

Steriade, D. (1999). Alternatives to the syllabic interpretation of consonantal phonotactics. In O. Fujimura, B. Joseph, and B. Palek (eds.), *Proceedings of the 1998 Linguistics and Phonetics Conference*. Prague: Charles University Press (Karolinum), 205–242.

Stampe, D. (1972). *How I spent my summer vacation*. PhD dissertation, University of Chicago.
Westbury, J. and P. Keating (1986). On the naturalness of stop consonant voicing. *JLg* 22: 145–66.

Notes

1. In section 3.2 below, this definition is elaborated: the effort cost assessed by LAZY reflects biomechanical energy augmented by a numerical index inversely proportional to speech register.
2. The PRES(feature) constraints are formally equivalent to the MAX(feature) constraints of, for example Lombardi 1998.
3. Although quasi-intervocalic position spans word boundaries, Nespor and Vogel (1982) have noted that it does not span the boundary of an intonational phrase, i.e. 'the domain over which an intonation contour is spread'. For example: [$_I$ le **x**ase **x**aɾine] [$_I$ **k**ostano molto **x**aɾe in ameɾi**x**a] 'Cute houses are very expensive in America'.
4. Note that table 10.1, including reference to 'levels A–K', is not taken directly from G&S, but is rather my summarisation of their description of these processes.
5. Cf. Pind 1999, showing, for Icelandic, that the mean duration of fast-speech geminates is even less than that of corresponding slow-speech singletons. Ojamaa (1976) shows the same for Estonian.
6. G&S characterise these data in terms of a generalisation that the velar stop is more likely to elide when flanked by vowels that are identical with respect to height, frontness, and peripherality (presumably equatable with tenseness), or when followed by a [+back] vowel; however, the examples they give do not conform to this generalisation. Thus, [SEe] preferably undergoes elision in spite of non-matching [tense] specifications.
7. But see Boersma 1998, questioning this aspect of Westbury and Keating's model.
8. The other principal strategy of avoiding passive devoicing, 'nasal leak' (allowing air to leak out of the nasal passages), carries a perceptual cost: risking confusion of the stop with a nasal consonant.

11 Language processing and segmental OCP effects*

Stefan A. Frisch

1 Introduction

The Arabic verbal roots are subject to a long-distance phonotactic constraint that is well known for its implications for autosegmental representation (McCarthy 1986, 1988, 1994). In this constraint, originally proposed as an instantiation of the Obligatory Contour Principle (Goldsmith 1979), repeated place of articulation features are not allowed within a root. Subsequent research has shown that the details of consonant occurrence in the Arabic roots are complex, with the strength of the phonotactic restriction gradiently dependent on the similarity of the consonants involved, the presence of intervening segments, and the contrasts available in the segmental inventory of Arabic (Pierrehumbert 1993; Frisch, Pierrehumbert, and Broe 2004).

The gradience of the phonotactic patterns in the Arabic lexicon provide strong evidence for a functional phonetic motivation for the constraint. The similarity avoidance constraint in Arabic is quantitatively dependent on similarity, distance between segments, segment frequency, and segmental position in the word. No formal model that prohibits feature co-occurrence like the autosegmental OCP can capture the richness of the patterning. However, a wide range of evidence from psycholinguistics suggests that processing a sequence of similar items is more difficult than processing a sequence of dissimilar items. Thus, we can account for the presence of similarity avoidance constraints in the phonotactics of Arabic as a consequence of functional pressure to make language processing as easy as possible. I claim that the richness of phonotactic patterns directly (quantitatively) reflects the functional explanation. In this way, statistical analysis of the lexicon provides a novel type of evidence for functionally motivated constraints and rules out alternative formal explanations (see Hawkins 1994 for similar arguments at the syntactic level). Statistical patterns in the synchronic lexicon arise as the result of a diachronic influence of processing difficulty on language change. Over time, functional pressures on the language have shaped the lexicon by influencing borrowing, the creation of nonce forms, and the loss of lexical items.

Despite the diachronic origin of the similarity avoidance patterns, native speakers are aware of these patterns, so they must be considered a part of the synchronic linguistic knowledge of the speakers. The co-occurrence constraint on homorganic consonants influences metalinguistic judgements of acceptability for novel roots (Berent and Shimron 1997; Frisch and Zawaydeh 2001) and the accommodation of borrowed lexical items (Frisch, Pierrehumbert, and Broe 2004).

Moreover, a phonotactic constraint based on processing difficulty should be universal. In fact, similarity avoidance constraints for homorganic consonants like those in Arabic have been found in a wide range of languages, such as English, Javanese, and Ngbaka. Analogous constraints that apply to repeated laryngeal features rather than repeated place features are also attested across unrelated languages such as Sanskrit, Hausa, and Souletin Basque (MacEachern 1999). Further, in cases where lexical patterns have been analysed statistically, the co-occurrence patterns are gradient and quantitatively depend on similarity (Berkley 1995, 2000; Buckley 1997; Frisch 1996; Pierrehumbert 1993).

In this chapter, a functional account of segmental OCP effects is given. Languages avoid sequences of repeated similar segments because they are difficult to serialise. An explanation for the difficulty of repetition can be found in models of language processing. Current language processing models use activation and competition in a neural network of linguistic units to account for similarity-dependent error rates in perception and production. In an activation/competition model of the encoding of a serial sequence, the units (e.g. segments) to be serialised must be activated and deactivated in the proper sequence (Dell, Burger, and Svec 1997). For a segment that has already been encoded, the node in the network corresponding to that segment has fired, and that node must be inhibited so that it does not continue to fire. Simultaneously, for a segment that is soon to be encoded, the node corresponding to that segment must be excited so that it is ready to fire at the proper time. If a sequence involves a repeated segment, the periods of inhibition and excitation may overlap and disrupt encoding of the correct sequence.

The chapter is organised as follows. Section 2 provides an overview of current theories of language processing, focusing on models of segmental processing. Section 3 reviews the segmental OCP patterns in Arabic and other languages, and shows these patterns are gradient and similarity dependent. Section 4 presents the processing account of similarity avoidance constraints and reviews some outstanding problems with this account. Section 5 concludes the chapter with a brief summary.

2 Current theories of language processing

Phonetic and psycholinguistic research on speech perception and production has found that language processing is highly interactive. The perception of structures at the level of speech sounds, spoken words, and larger syntactic constituents involves three general stages: the activation of compatible structures, competition between structures that share common input, and selection of the winning structure (e.g. Swinney 1979; Zwitserlood 1989). In addition to activation and competition within a structural level, there is activation between levels, such that one can influence another (e.g. Ganong 1980). Similarly, the production of speech sounds, words, and phrases appears to involve interactive activation and competition (Dell 1986; Stemberger 1983). Errors in speech production reveal competition between articulatory plans. The processing-based account of similarity avoidance constraints relies on the notion of activation and competition between segments in phonological encoding, so evidence for activation models of phonological processing is presented below.

Theories of phonological processing come from two sources of data. Research on speech perception has sought to determine the organisation and access of phonological information in the mental lexicon. Research on speech production has examined how lexical information is accessed and encoded as a sequence of gestures. In both cases, much of what is learned about language processing comes from studying the systematic errors that are made. Errors tell us something about items that are partially activated during processing. Additional evidence for activation/competition can be gleaned from differences in ease or speed of processing. Tasks involving greater competition between segments or words generally take longer to perform than tasks with no competition.

In this section, I first review evidence in favour of activation/competition from speech perception and speech production. I then introduce connectionism as a formal model of activation/competition in language processing. Finally, psycholinguistic evidence for the processing difficulty involved in repetition is reviewed and connectionist models of serial encoding that explain the processing difficulty are presented.

2.1 Spoken-word recognition and lexical neighbourhood effects

There is a great deal of evidence from speech perception and spoken-word recognition that the phonological lexicon is organised as a multidimensional acoustic-phonetic space, and the recognition of a word takes place when a lexical item becomes sufficiently activated in comparison to its competitors. The organisation of phonological words has been described in terms of groups of words that are phonetically similar, called *lexical neighbourhoods*. For example, the lexical neighbourhood for *cat* would include words such as *bat*, *fat*, *cut*, *kit*, *cap*, *can*, *scat*, and *cattle*. Substantial evidence for lexical neighbourhoods has been found by comparing the performance of experimental participants in processing words that differ in their neighbourhood characteristics (Luce and Pisoni 1998).

Three factors related to the organisation and activation of words in lexical neighbourhoods have been shown to influence the processing of phonological words. It has long been known that more frequent words are easier to process when compared to less frequent words (e.g. Miller, Heise, and Lichten 1951). Presumably, high-frequency words like *buy* are easier to activate than low-frequency words like *bough*. In addition, the *density* of a lexical neighbourhood influences performance. Words that have many lexical neighbours (e.g. *cat*, *lick*) are more difficult to process than words that have fewer lexical neighbours (e.g. *quiz*, *purge*). Finally, the frequency of the neighbours of a target word influences processing. Words with many high-frequency neighbours are more difficult to process than words with mostly low-frequency neighbours. The density of its neighbourhood and frequency of its neighbours determines how much competition the target word receives from other lexical items (Goldinger, Luce, and Pisoni 1989). Words that are high in frequency with only a few low-frequency neighbours are easy to process, while low-frequency words with many high-frequency neighbours are hard to process.

The competition between words and their lexical neighbours influences processing in a variety of tasks (Luce 1986). In an *identification task*, the participant is asked to determine what word has been presented when the word's identity is masked by noise. In this task, ease of processing is reflected in the accuracy in identification, and easy words are identified more accurately than hard words. In a *repetition naming task*, the participant is asked to produce the word as soon as possible after it is presented (either auditorily or visually). In the repetition naming task, ease of processing is reflected in the latency between the presentation of the stimulus and the start of the participant's production of it. Easy words are produced after a shorter delay than hard words. In the *lexical decision task*, the participant is asked to identify whether the stimulus is a word or a nonword. In lexical decision, the influence of lexical neighbourhoods depends on whether the stimulus is a word or nonword. For a response of 'word' for a word (e.g. *kite*), participants are more quickly able to answer if the word is from a dense neighbourhood than a sparse neighbourhood. For a response of 'nonword' for a nonword (e.g. *gite*), participants are more quickly able to answer if the nonword is from a sparse neighbourhood. Both patterns make sense if words are organised into lexical neighbourhoods. A word in a dense neighbourhood will activate a relatively large number of other words, prompting the participant to respond 'word'. For a nonword in a sparse neighbourhood, there will be relatively few words that are activated, so the 'nonword' response will not receive much competition.[1]

Further evidence for the activation of lexical items in processing comes from research on the time course of spoken-word recognition using eye-tracking equipment. In these experiments, participants are given verbal instructions to manipulate objects represented by icons on a computer screen. Before an object can be manipulated, the participant must fixate their gaze upon the object to co-ordinate the movement of the computer cursor to the object. By examining when participants fixate on an object relative to when the name of that object is spoken, it can be demonstrated that participants begin looking at objects before the entire name has been spoken, and further that they are much more likely to fixate on the target object or a lexical neighbour of that object than on a phonologically unrelated object (Allopenna, Magnuson, and Tanenhaus 1998). Assuming that the probability of fixation on an object is a monotonic function of the activation strength of the object's name, the patterns of fixation compare well to a model of phonological processing based on activation and competition. Allopenna et al. (1998) present simulations using the TRACE model (discussed in section 2.3) that provide a very good fit to their eye-tracking data. Overall, we find consistent evidence that the auditory input in speech perception and spoken-word recognition triggers the activation of word and segment units that compete with one another to be recognised as the percept.

2.2 Phonological encoding in speech production

The study of speech production has primarily focused on speech errors that occur in spontaneous speech or in error-inducing experiments. In a speech error, some phonological unit is misproduced. Such errors primarily occur from misordering elements in the speech plan, though in some cases no source for the error is apparent. Errors can occur at the level of the phrase, word, morpheme, or segment, with segmental errors being the most common (Fromkin 1971). Examples of segmental speech errors are given in (1).

(1) *frish gotto* for *fish grotto*
blake fruid for *brake fluid*
spicky point for *sticky point*

In speech errors between segments, errors occur most commonly between segments that share many distinctive features. For example, pairs like p/t, r/l, and s/z have a high error rate, while pairs like p/r and s/m have a low error rate. In other words, errors occur more commonly between segments that are similar to one another. Frisch (1996: ch. 4) analysed consonant-error data for English from portions of two naturally occurring speech-error corpora, the Arizona-MIT corpus (Shattuck-Hufnagel 1979) and the Stemberger corpus (Stemberger 1991). Table 11.1 shows the number of errors in each corpus for each level of similarity in the 'observed' column. Similarity was computed using a metric based on shared features and natural classes (Frisch, Pierrehumbert, and Broe 2004). The number of errors that would be expected if errors were randomly distributed is given in the 'expected' column. The relative error rate (the observed

Table 11.1. *Consonant segment errors aggregated by similarity for two naturally occurring error corpora*

	MIT-Arizona corpus			Stemberger corpus		
Similarity	Observed	Expected	O/E	Observed	Expected	O/E
0–0.1	72	519.3	0.14	26	197.1	0.13
0.1–0.2	246	416.5	0.59	98	178.0	0.55
0.2–0.3	234	113.5	2.06	100	57.1	1.75
0.3–0.4	195	88.2	2.21	82	40.6	2.02
0.4–0.5	288	93.1	3.09	131	31.8	4.12
0.5–0.6	238	42.4	5.61	82	14.4	5.68

divided by the expected) is given in the 'O/E' column. Table 11.1 shows that error rate between consonants is clearly highly dependent on similarity.

At the word level, errors also tend to occur between words that are similar. For example, in error-inducing experiments, words with the same vowel are more likely to interact in a consonant error (e.g. *bad back* for *mad back*) than words that do not share a vowel (e.g. *bade back* for *made back*, Dell 1984). In segment errors within polysyllabic words, prosodic similarity is also relevant. Errors are more likely to occur between segments in similar word and stress positions (Frisch 2000; Shattuck-Hufnagel 1992). In errors between whole words, similarity also plays a role. Errors of whole-word substitution are more likely for words that share several segments. Examples include the substitution of *recession* for *reception*, *liberal* for *liveable*, and the blending of *correlated* and *corroborated* to produce *corrobolated* (Fromkin 1971).

As in the case of speech perception, the common patterns of speech-error production described above can be accounted for by a processing model based on activation and competition between phonological units. Segments that share features will activate one another via those shared features, increasing the likelihood that a similar, but incorrect, segment will be selected for encoding instead of the intended segment. Analogously, words that share segments will activate one another based on those shared segments, increasing the likelihood that a similar word or part of a word will be selected for encoding instead of the intended item.

2.3 Models

In both perception and production research, connectionist models that employ spreading activation have been most successful at accounting for the effects of activation and competition in language processing in both error-free and error performance (see Elman, Bates, Johnson, Karmiloff-Smith, Parisi, and Plunkett 1996 for an introduction). These models simulate parallel processing over a

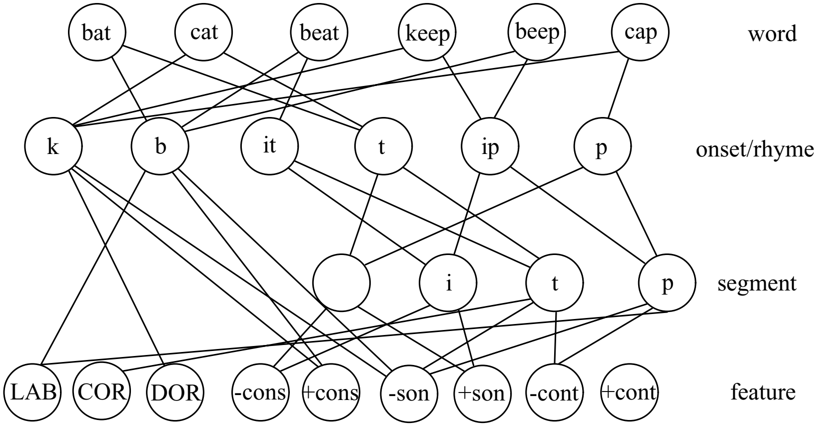


Figure 11.1 Model of phonological encoding, following Dell 1986. Phonological words activate intermediate phonological levels, which in turn activate phonological features to be encoded for articulation.

large set of simple units. Each unit, or node, has an activation level. Nodes are grouped into layers as in figure 11.1. The layers commonly reflect hierarchical layers of linguistic units, such as words, syllables, segments, and features. Each model has an input layer and an output layer that represent the interface between the processing model and semantic, auditory, or articulatory systems. In a model of speech production, words are the input, and the output is a set of segments, features, or articulatory gestures. In a model of speech perception, the input would be acoustic data, distinctive features, or segments, and the output would be a set of units that represent word meanings. In between the input and output are one or more layers of *hidden units*. These layers represent intermediate levels of representation or intermediate stages of processing. In the case of feature-to-word models or word-to-feature models, intermediate layers could include morphemes, syllables, onsets, rhymes, and segments. For example, the TRACE model of spoken-word recognition (McClelland and Elman 1986) has an input feature layer that roughly corresponds to the acoustic features of Jakobson, Fant, and Halle (1952), a word layer for output, and a (hidden) phoneme layer in between.

Processing in a connectionist model is simple. Activation is spread between nodes via weighted connections. In most models, connections are bidirectional, so the units on either end will excite one another when activated, or inhibit one another when not activated. In each time unit of processing, each node's activation level is adjusted according to the amount of excitation and inhibition it is receiving (which is a function of the activation level of nodes connected to it and their connection strengths).

The Dell (1986) model of phonological encoding for speech production, shown in figure 11.1, uses word, syllable, onset, rhyme, segment, and feature nodes. Phonological encoding in this model is simulated by first activating a word that is to be produced (e.g. *keep*). This word spreads activation to corresponding onset and rhyme nodes (e.g. /k/ and /ip/), which in turn activate their corresponding segments (e.g. /i/, and /p/) and features (e.g. [+DOR], [−son]). The onset and rhyme units also activate other words (e.g. *cat* and *beep*) that will introduce competition for the production of the correct segments. The model is run for a number of cycles of spreading activation, after which the most active onset, nucleus, and rhyme segments are selected to be the segments that are encoded. Since the Dell model contains a feature level, it can account for many of the basic facts of segmental speech errors. Similar segments will be activated by shared features, and so will be more likely to be selected for one another, producing an error. In addition, the shared segment nodes between words will cause words with similar segments to be more likely to interfere with one another and result in a speech error.

In some models, the activation of one node inhibits the activation of another node. In the TRACE model of word recognition, words that share segments inhibit one another. Effectively, words in TRACE compete for the segments that are activated by the acoustic feature input. This competition will produce many of the lexical neighbourhood effects discussed above. Words with many neighbours will receive a great deal of inhibition from similar words, and so will be harder to activate over their threshold level. Frequency effects are incorporated in these models by giving units different thresholds of activation for encoding. High-frequency units require less activation to fire than low-frequency units, so high-frequency units are more reliably encoded. Also, high-frequency neighbours to a word will provide more competition than low-frequency neighbours to a word as high-frequency neighbours will be more likely to fire when only partially activated.

2.4 *The problem of serial encoding*

In spoken-word recognition and speech production, language processing unfolds over time. In processing language over time, it is important not only to recognise or produce the appropriate linguistic units as discussed above, but to also recognise the correct order of elements. The words *pat*, *tap*, and *apt* all involve the same segments, and it is their ordering that differentiates the words. A wide range of psycholinguistic studies have shown that sequences with repeated items are more difficult to process than those that do not contain repetition. Presumably, the problem is that repetition introduces a potential confusion in the linear ordering. For example, when producing *Julius Caesar's seizures*, there is high phonological similarity between *Caesar's* and *seizures*.

All segments but one are repeated, and the two distinct segments /z/ and /ʒ/ are highly similar. The likelihood of a speech error is high, as after planning the sequence /si/, it is potentially unclear whether this is the first /si/ in *Caesar's* or the second /si/ in *seizures*, and the sequences /siz/ and /siʒ/ must both be planned. Consistent with this example, it has been found that sequences containing repeated segments have higher speech-error rates, in both laboratory experiments and naturally occurring errors (e.g. Dell 1984; Stemberger 1983).

It might at first appear that the repetition problem would only exist for production, as there is no serial planning that must take place in perception. However, there is the possibility for perceptual confusion as well in cases where items are repeated. In the case of *Caesar's seizures* the perceptual system must determine that there were two distinct productions of the repeated segments. For example, whatever perceptual mechanism indicates that an /s/ is being heard at the beginning of the first word must be reset before the next word arrives (see MacKay 1970 for a neurologically based account of this phenomenon). If /s/ is activated at the beginning of the next word, is it because there was an /s/ in the previous word, or because a new /s/ has come along? The problem is most acute in cases of immediate repetition of a segment (Boersma 1998). For example, the segmental difference between *heavy oak* and *heavy yoke* is the presence of a /j/ onset on the second word (/i#o/ versus /i#jo/). But, phonetically /i/ and /j/ are basically the same. Unless there is a pause between the words, it is difficult to determine whether /j/-like transitions into the /o/ are the result of coarticulation from a previous /i/ or from a distinct /j/ segment. There is no point during the juncture where the perceptual system can determine that the /i/ has ended and a /j/ begun. This does not imply that it is impossible to perceive a contrast between *heavy oak* and *heavy yoke*. Additional phonetic cues are present, such as the overall duration of the [i/j] and the presence or absence of glottalisation at the word juncture. However, these cues rely on suspending judgement on the identity of a portion of speech until sufficient evidence is accumulated to resolve the uncertainty. So, as in the case of production, there is a certain amount of encoding and sequencing that takes place during perception.

There is substantial evidence that a certain amount of time is required for the perceptual system to detect an item and reset itself to detect the same item again. In visual perception of words, there is a well-known phenomenon of *repetition blindness* (Ericksen and Shutze 1978; Kanwisher 1987). In experiments using very rapid presentation of word sequences, repeated items are sometimes perceptually fused and reported only once, even though there is a brief period of time in between stimulus presentations in which a blank screen is presented. This effect has even been shown for repeated letters within words. For example, in a very rapid presentation of the visual sequence *tell*, *shell*, *oe*, the repeated letter sequence *ell* is often perceived only once, such that participants report seeing *tell*, *shoe* (Morris & Harris 1997).

Analogous perceptual errors on repetition in rapidly presented auditory stimuli have also been reported. Participants listening to rapid sequences of words (that must be artificially sped up to produce the effect) can miss a repeated item, reporting it only once (Miller and MacKay 1994). The same effect has also been found at a more abstract level. MacKay and Miller (1994) played rapid sequences of words in which each phonological word was distinct, with no repetition, but the sequence did contain a sequence of near synonyms. In this case, repetition at the semantic level also produced an increase in misreporting what was heard. Thus, we might conclude that the repetition effect is not entirely auditory or articulatory, but instead results from the need to parse a sequence into discrete units at several levels of abstraction.

Processing difficulties that result from repetition can also be found at the morphological level. Stemberger and MacWhinney (1990) examined morphological errors where speakers fail to mark a past tense verb correctly. For example, the past tense of *kid* is *kidded*, producing a sequence containing repeated /d/ segments. Speech errors in which the past tense was intended but not marked (e.g. *kid* for *kidded* or *walk* for *walked*) were more frequent in the case where the word final segment was phonologically similar to the past tense morpheme. In other words, they found a much higher error rate in cases where past tense marking would produce a repeated sequence of similar segments.

2.5 Models of phonological encoding

Many connectionist models of language processing have been adapted from models of reading isolated words, and so do not address the process of serialisation. Serialisation is an integral part of spoken language processing, even for isolated words. As activation-based models of language processing have evolved they have begun to address the problem of serial encoding (e.g. Hartley and Houghton 1996). One solution to the problem is to add a set of *sequencing nodes* that are interconnected with excitatory and/or inhibitory links that would cause them to activate and deactivate in sequence (e.g. Dell, Burger, and Svec 1997). The representation of each word is tied to this cascade of sequencing nodes so that the words' segments (and their corresponding features) would become activated in the correct order. A simplified example of a model of this type is shown in figure 11.2. This model is one example of a model to produce English syllables that is a modification of the Dell 1986 model. The sequencing nodes correspond to the abstract structural units of onset, nucleus, and coda. Activation begins with an abstract word node, which in turn activates the segments of the word and the onset node. Over time, activation of the onset node builds. Activation also begins to build in the nucleus. As activation builds in

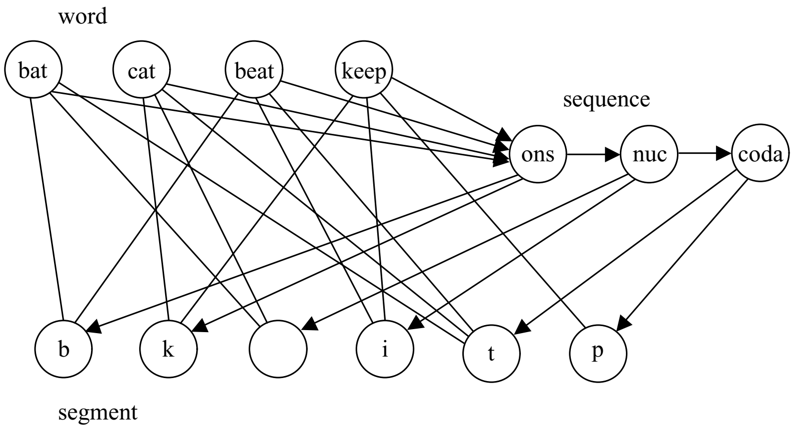


Figure 11.2 Structural model of sequential processing. Independent onset, nucleus, and coda nodes are linked to syllables and their corresponding segments. Activation spreads so that the onset fires first, the nucleus fires second, and the coda fires third.

the nucleus, the activation level of the coda also begins to build. After the onset fires, its activation is lost, and the nucleus is the next most activated element. Once the nucleus fires, activation will continue to build in the coda, until the coda fires. As each sequencing node fires, the most activated segment node is selected to be encoded in that position.[2]

3 The segmental OCP

The phonological phenomena to be accounted for by these language processing models are long-distance similarity avoidance constraints for segments within the lexical patterns of a language. The most well-known of these, the consonant co-occurrence restrictions in the triconsonantal root morphemes of Arabic, were formalised by McCarthy (1986, 1988) using the Obligatory Contour Principle (Goldsmith 1979). The OCP prohibits the repetition of elements on an autosegmental tier. McCarthy proposed, in the case of Arabic, that the OCP is parameterised to operate on privative place of articulation tiers, blocking the generation of a root that contains repeated homorganic segments as C_1C_2, C_2C_3, or C_1C_3. McCarthy also proposed that each place of articulation tier could be selectively sensitive to manner features. These technical refinements to the OCP-Place constraint in Arabic were necessary to account for the fact that roots with C_1C_2 such as those in (2) are not found while those in (3) are relatively rare and those in (4) are quite frequent.

(2) */b m h/
 */t d h/
(3) ?/d s h/
 ?/t ʃ h/
(4) /d n h/
 /t r h/

Segmental OCP effects are found in a wide variety of languages. Place-based consonant co-occurrence constraints are prevalent in the Semitic languages (Bender and Fulass 1978; Buckley 1997; Greenberg 1950; Hayward and Hayward 1989; Koskinen 1964). They have also been found in unrelated languages, including Hawaiian and Serbo-Croatian (MacKay 1970), Ngbaka (Broe 1995), English (Berkley 1995; Frisch 1996), French (Plenat 1996), Italian (Frisch, Pierrehumbert, and Broe 2004), Javanese (Mester 1986), Russian (Padgett 1992), and other languages (Yip 1989). Across languages, there appear to be differences in the details of the co-occurrence restrictions for homorganic consonants. For example, in Ngbaka the co-occurrence restrictions are limited to a very few homorganic consonant pairs, and most homorganic consonant combinations are allowed. In Arabic the co-occurrence constraints forbid nearly all homorganic consonant pairs.

When the segmental OCP constraints within a particular language are examined more closely it becomes apparent that homorganic segments are constrained to different degrees, suggesting that the all-or-nothing autosegmental OCP cannot properly explain the patterns. Pierrehumbert (1993) observed that the Arabic consonant co-occurrence constraints depend on the similarity of the homorganic consonant pairs involved. For example, the alveolar consonants divide into distinct series of obstruents and sonorants, where co-occurrence between the series is frequent, while co-occurrence within the series is uncommon (Greenberg 1950; McCarthy 1994). Frisch, Pierrehumbert, and Broe (2004) showed that many additional sub-patterns of stronger and weaker co-occurrence restrictions can be identified for consonant pairs in Arabic. Table 11.2 gives rates of consonant co-occurrence for different levels of consonant similarity for adjacent and nonadjacent Arabic consonant pairs. Similarity is computed using features and natural classes as described in Frisch, Pierrehumbert, and Broe (2004). Nonhomorganic consonant pairs have a similarity value 0, and identical pairs have a value of 1. Observed and expected totals are taken from an on-line version of Wehr's dictionary (Cowan 1979) that contains native and assimilated Arabic triconsonantal roots (Pierrehumbert 1993). Relative occurrence of the combinations (observed divided by expected) is also given, and shows a consistent decrease in co-occurrence rate as similarity increases. There is also a higher rate of co-occurrence for similar but nonadjacent consonant pairs, suggesting that distance between segments also plays a role in the constraint.

Table 11.2. *Consonant co-occurrence between Arabic consonants that are adjacent (C_1C_2 or C_2C_3) and nonadjacent (C_1C_3) in the root by homorganic consonant pair similarity, along with mean wordlikeness ratings on a 1–7 scale from a group of 24 native Arabic speakers*

	Adjacent			Nonadjacent			Wordlikeness
Sim	Obs	Exp	O/E	Obs	Exp	O/E	Rating
0	2978	2349.3	1.27	1411	1248.4	1.13	3.3
0–0.1	451	365.2	1.23	219	203.2	1.08	–
0.1–0.2	492	550.6	0.89	308	283.7	1.09	3.0
0.2–0.3	151	260.2	0.58	96	124.2	0.77	2.9
0.3–0.4	29	131.2	0.22	50	67.0	0.75	2.7
0.4–0.5	14	180.2	0.08	75	103.2	0.73	2.7
0.5–0.6	3	40.8	0.07	8	25.0	0.32	–
0.6–1	0	90.2	0.00	13	40.4	0.32	1.8
1	1	199.6	0.01	16	103.9	0.15	2.5

Despite the gradience and lexical nature of the similarity avoidance constraint, these patterns are learned and generalised by native speakers. Frisch and Zawaydeh (2001) conducted an experiment in which novel verb forms were given to native Jordanian Arabic speakers for well-formedness judgements. Frisch and Zawaydeh found that verbs from novel roots with OCP violations were judged less acceptable than verbs without violations. This was the case even when the consonant pairs in the nonviolating root were unattested in the lexicon (i.e. an accidental gap). The rightmost column in table 11.2, adapted from Frisch and Zawaydeh, gives average wordlikeness judgements for novel forms across a variety of similarity levels. These data show that influence of similarity on consonant co-occurrence in the lexicon is part of the productive phonological knowledge of Arabic speakers.

In the other languages that have been examined statistically for OCP-Place effects, gradient patterns of co-occurrence have also been found (Berkley 1995, 2000; Buckley 1997; Frisch, Pierrehumbert, and Broe 2004). In each of these cases, similar homorganic consonant pairs are found to be more restricted than dissimilar homorganic consonant pairs, and distance between the consonants reduces the influence of similarity. In some cases, there are no categorical consonant co-occurrence restrictions, but OCP-Place effects are still found as a statistical influence on lexical patterns.

OCP-Place effects in English provide an interesting example of the gradient OCP. There are some gaps in the English lexicon that have traditionally been attributed to the OCP. Davis (1984) discusses a gap of the form sC_iVC_i, where the C_i are identical noncoronals (e.g. **spop*, **skik* and longer words that contain analogous sequences). Coronal forms are robustly attested as in *state*, *stoat*, *stat(istic)*, *stet*, and *astute*. Davis analysed these patterns as the result of a categorical OCP constraint that is active over *sCVC* sequences, where coronals have special status. However, a quantitative analysis of the English lexicon finds that sequences of C_iVC_i are disfavoured more generally (Berkley 1995, 2000; Frisch 1996: Pierrehumbert 1994). It appears that Davis' categorical gap is just an extreme case of the more general pattern of similarity avoidance in English. Frisch (1996; ch. 10) analyses the gap in noncoronal sC_iVC_i sequences as a result of the combined effects of the gradient OCP and a low expected probability of combination of noncoronal sC_iVC_i.

Similar segmental OCP constraints have also been found for laryngeal features (Ito and Mester 1986; MacEachern 1999; Steriade 1982). For example, MacEachern (1999) examined eleven languages that have constraints prohibiting repeated or similar laryngeal specifications within morphemes. As in the case of OCP-Place, the laryngeal co-occurrence constraints across different languages constrain co-occurrence to different degrees. MacEachern showed that the co-occurrence constraints follow a cross-linguistic implicational hierarchy based on similarity. If a language allows two segments with relatively similar laryngeal specifications, such as an ejective and a glottal stop, then the language will also allow two segments with relatively dissimilar laryngeal specifications, such as an ejective and /h/. In earlier work, MacEachern (1997) analysed these languages with a *SIMILARITY constraint hierarchy, directly encoding the similarity avoidance constraint. This was later replaced with an analysis using conjoined constraints that ban repeated feature specifications, but the implication that the constraint depends on segmental similarity remains the same.

In the case of Arabic, repeated identical consonants are restricted to the strongest degree of all, and they generally do not occur. For a functional similarity avoidance constraint, identity is the worst possible violation. However, in many languages that have segmental OCP effects, identical segments are treated differently than segments that are highly similar but nonidentical. For example, in Peruvian Aymara, combinations of nonhomorganic ejectives (e.g. /t'/ and /k'/) or homorganic ejectives and aspirates (e.g. /t'/ and /t^h/) are prohibited, but two identical ejectives or two identical aspirates are allowed (MacEachern 1999). In Ngbaka, repeated similar homorganic segments are not allowed (e.g. /p/ and /b/, /b/ and /m/) but repeated identical segments are allowed (Broe 1995). The greatest challenge to the similarity avoidance account is how, in some languages, identity provides a special case that is not considered a violation. The architecture of the language processing module provides a potential solution to this dilemma which is discussed below.

The fact that segmental similarity avoidance constraints appear to be gradient provides important evidence that they are functionally motivated. While many phonological processes are categorical, phonotactic constraints over the lexicon need not be. The lexicon is a large database of phonological patterns. These patterns are potentially influenced by external forces and change over time. It is therefore reasonable to conclude that functional constraints will be directly reflected in phonotactic patterns. In other words, functionally poor patterns will be scarce to the extent that the patterns are poor, and functionally good patterns will be frequent to the extent that the patterns are good. The distribution of patterns in the lexicon can capture the graded nature of phonetic and cognitive functional constraints. These patterns are generally not categorical in their influence on linguistic performance and their impact on the lexicon need not be categorical either. However, this is not meant to imply that the functional patterns are irrelevant to formal linguistic theory. There are a number of cases where the OCP has been used to account for categorical patterns in morphophonological processes (e.g. McCarthy 1988). It is unsatisfactory to consider gradient cases of segmental similarity avoidance as completely unrelated to categorical cases of segmental similarity avoidance. Rather, it is likely that the morphophonological OCP effects are the result of the same functional similarity-avoidance constraint. In these cases the functional constraint has been grammaticised and regularised into a categorical pattern.

A second place where gradient phonological processes can be observed is in sociolinguistic variation. In fact, there is a particularly well-studied case of a variable dissimilation constraint in dialects of English that also sheds some light on the issue of grammaticalisation of functional constraints (Labov 1971; Guy 1991). In a pattern that is analogous to the errors in nonmarking of past tense studied by Stemberger and MacWhinney (1990), varieties of African American Vernacular English have systematic, but variable, nonmarking of past tense in cases where repetition of similar segments would result, as in (5).

(5) [spat] – [spatəd] ‘spotted’
    [ɡa$^{\text{I}}$d] – [ɡa$^{\text{I}}$dəd] ‘guided’

Guy (1991) refers to this process as *variable t-deletion*. Patterns of t-deletion are not limited exclusively to past tense morphology or to cases of repetition and the process appears to be phonological rather than morphological. However, contexts where the morphology would introduce repetition show the highest rates of t-deletion, suggesting that the origin of the deletion pattern was in that context and subsequently the pattern has spread more generally through the lexicon. This change in progress could play out in any number of ways, resulting in categorical patterns that only partially reflect the original functional motivation for the change.

There is another phonological phenomenon that, in general, should share common properties with the segmental OCP constraints. These are harmony constraints. Harmony promotes assimilation, rather than dissimilation. As with the similarity avoidance constraints, I am primarily concerned here with non-local harmony. Cases involving assimilation of adjacent segments can be analysed as the result of articulatory simplification, and in these cases, repetition is avoided by reducing a sequence of gestures into a single one (see Jun, this volume). Harmony patterns provide a challenge to a functionally grounded account of similarity avoidance. In addition, there are cases of long-distance consonant assimilation. Any theory that predicts that similarity is problematic for language processing would predict that harmony would be a functionally poor pattern. However, there are many other cases where distinct functional tensions push in opposite directions. The most well known is the tension between economy of articulatory effort and maximisation of acoustic distinctiveness (Lindblom 1990; Jun, this volume; Kirchner, this volume). In the Optimality-Theoretic approach to functional explanation of synchronic patterns, the resolution of these tensions is resolved one way or another through a language-particular constraint ranking. Thus, it may merely be the case that some languages place the functional value of repetition avoidance above the value of harmony, while others do the reverse. This is clearly a fruitful topic for further research (see Kaun, this volume, for further discussion).

4 The language-processing account

The primary conclusion of this chapter, that segmental OCP constraints can be attributed to processing difficulty during serialisation of the segments of a word, has been independently proposed by Berg (1998: 88–98). Berg points out that serialisation is particularly problematic in the case of Arabic, where consonant sequences in root morphemes could easily be confused. Berg and Abd-El-Jawad (1996) present evidence that Arabic consonants are unusually susceptible to speech errors that involve consonant segment reordering, as shown in (6).

(6) /takriib/ for /takbiir/ ‘glorification’
/maraaʕiʃ/ for /maʃaaʕir/ ‘feelings’

Berg claims this error pattern can be attributed to the nonconcatenative morphology of Arabic. Prunet, Béland, and Idrissi (1999) present additional evidence that this is the case. They examined the productions of a bilingual Arabic and French speaker with aphasia. They found he made many misordering errors in Arabic, while those errors were systematically absent in French.

Arabic root structure provides a case where the language is particularly vulnerable to the problem of serialisation during production. Berg argues that segmental OCP effects in Arabic are present because the nonconcatenative morphology makes Arabic particularly vulnerable to the processing difficulty caused by repetition. This vulnerability can be explained by the model of serial encoding presented above. In Arabic, the connections between the root consonants and the sequencing nodes would be weak because the root consonants appear in different syllable positions in the verb form, adjacent to a variety of vowels. Examples for the root /ʃ r b/ are shown in (7).

(7) ʃariba 'he drank'
ʃurba 'a drink'
maʃrabun 'tavern'

Since segmental OCP effects are also found in languages that do not have nonconcatenative morphology, it must be the case that repetition difficulty is a universal of language processing, though the effects of the constraint may be weaker or stronger depending on the other aspects of a language's phonology and morphology. Similarity avoidance in Arabic is particularly strong due to the special problems of sequencing segments in a nonconcatenative morphological system.

Boersma (1998: 415–40) also takes a phonetic approach to the OCP. Boersma primarily focuses on a perceptual difficulty with repetition, in particular in the case of parsing a sequence of segments where there is immediate repetition. In order to aid recoverability, a language might use epenthesis between repeated consonants. Alternatively, a language might block vowel deletion in cases where repeated identical segments would then result. Boersma considers the perceptual motivation for the OCP to be valid only for adjacent segments. For distant segments, he proposes an articulatory constraint on repetition similar to that of Berg (1998). However, a perceptual motivation for long-distance dissimilation should not be so readily dismissed. In speech perception, a listener must take a rapid stream of air pressure fluctuations and turn it into a sequence of segments, words, phrases, and clauses. Given the high rate of information conveyed in speech, the identification of segments must be very rapid, and consequently may be considered cognitively difficult. A robust speech-perception module has to deal with many cases of missing information or ambiguity (see Wright, Frisch, and Pisoni 1999 for an overview). One possible aid is higher order information, such as the knowledge of the identity of a word being used to supply information about the segments in that word. Thus, the perceptual system does not want to immediately commit to any particular percept of a segment, as additional information might cause a change in the identity of missing or ambiguous segments. If segmental decisions are delayed, or stretched out over time somewhat (i.e. they last longer than the fleeting acoustic pattern corresponding to the segments of the word), then repetition of similar segments within a word may result in blending of perceptual traces and consequently a misperception. For example, the mechanism that recognises segments might still be gathering information about the identity of the first segment in a CVC word when the third segment is spoken. In that case, the new and old consonants would interact and compete, possibly interfering with proper encoding.

Given the preceding discussion, the account of long-distance similarity avoidance constraints is a relatively simple one. Psycholinguistic research and models of language processing suggest that processing repeated items is difficult. In the cases of segmental OCP constraints for place features or laryngeal features discussed above, it appears that repeated gestures using the same articulator are avoided.[3] This is a processing-based explanation, and from a purely phonological standpoint this level of detail is sufficient as a tentative explanation of the prevalence of similarity avoidance constraints in the world's languages.

4.1 Further predictions

Greater insight into the nature of segmental OCP effects can be gained by considering whether additional phonological predictions can be made by a processing-based account. Activation-based models of lexical access and phonological encoding provide a good basis for an account of the repetition constraint. In particular, activation and competition in these models is sensitive to the similarity of phonological units, so segments will compete with one another to the extent that they are similar. Whether the processing model uses some type of activation cascade or is based on a recurrent network, repeating similar items will create competition between the repeated segments, thus disrupting serialisation.

There is a second aspect of these models of serialisation that has not, so far, been discussed. As the serial processing of a word progresses, elements that have been successfully encoded have a greater and greater influence on the encoding of the remaining segments. For example, in the *cohort model* of spoken-word recognition, the candidate set of competing words narrows as phonetic input is processed, until finally a unique candidate is activated (Marslen-Wilson and Welsch 1978). For example, upon hearing [k^h], many words are possible (e.g. *cat*, *cut*, *cute*, *kite*, *kit*). Once most of the segments in a word have been perceived, it is very likely that lexical access will converge on the correct word, and very unlikely that any segment will become highly activated unless it is a possible completion of the word. For example, after [k^hep] there are only a few possibilities (*capes*, *caped*, *caper*, *capon*). A similar context effect has been demonstrated in speech production. Sevald and Dell (1994) showed that it is more difficult to say sequences of words in which initial segments are repeated (e.g. *cat cab*) than it is to say words in which final segments are repeated and initial segments are not repeated (e.g. *cub tub*). This is another example of the repetition effect. For *cat cab*, the repeated word-initial segments increase the amount of competition between words, making it more difficult to encode the correct final segment (/t/ or /b/). However, in cases where the repeated segments are word final (*cub tub*) there is no evidence of

Table 11.3. *Onset and coda consonant co-occurrence in English monomorphemes as a function of homorganic consonant similarity. Data are given separately for each of the first three syllables*

	Syllable 1			Syllable 2			Syllable 3		
Sim	Obs	Exp	O/E	Obs	Exp	O/E	Obs	Exp	O/E
0	1258	1099.6	1.14	558	491.2	1.14	101	99.4	1.02
0–0.1	145	143.2	1.01	64	66.2	0.97	13	17.2	0.76
0.1–0.2	317	275.1	1.15	227	213.7	1.06	67	49.4	1.36
0.2–0.3	162	150.5	1.08	146	117.1	1.25	39	30.3	1.29
0.3–0.4	98	112.6	0.87	61	71.9	0.85	16	17.0	0.94
0.4–0.5	42	73.6	0.57	29	44.4	0.65	14	20.2	0.69
0.5–0.6	56	92.8	0.60	65	79.9	0.81	20	26.3	0.76
0.6–1	14	35.7	0.39	24	31.4	0.76	11	8.2	1.34

lexical competition, and the activation of repeated segments actually facilitates encoding.

Since preceding context helps to ensure correct encoding of later segments in the word, it should be the case that the influence of the segmental OCP constraint is sensitive to preceding context. In other words, the negative effects of repetition on processing should be ameliorated somewhat for repeated segments that occur nearer to the end of lexical items. This hypothesis has been investigated for English and Arabic OCP-Place constraints (Frisch 1996, 2000). A comparison of the strength of OCP-Place constraints for Arabic C_1C_2 and C_2C_3 consonant pairs found some evidence that the C_2C_3 constraints are weaker (Frisch 2000). In English, where the overall OCP-Place constraint is not as strong as in Arabic, there is also evidence that the constraint is weaker nearer to the end of lexical items. Table 11.3 shows relative rates of co-occurrence (O/E) for similar homorganic onset and coda consonant pairs in CVC syllables of English monomorphemes for each of the first three syllable positions in the word. Similarity of the consonant pairs is determined using features and natural classes (Frisch, Pierrehumbert, and Broe 2004). Identical consonant pairs are not included, as the details of identical consonant co-occurrence in English are complex (see Frisch 1996 for discussion). While the data for later syllables in the words are somewhat sparse, there is a clear trend towards greater co-occurrence. There is higher O/E for OCP-Place violations away from the beginning of the word, especially in the third syllable. In the third syllable, there is no clear evidence of a segmental OCP constraint.

There is one further sub-regularity in the OCP-Place constraints of Arabic, and it too provides evidence that quantitative patterns in the lexicon reflect the subtle influence of functional processing constraints. Note that the analysis of the segmental OCP as a similarity avoidance constraint predicts that

Table 11.4. *Influence of consonant probability on Arabic onset and coda consonant co-occurrence*

	Consonant probability	
Sim	Low	High
0	1.30	1.25
0–0.1	1.17	1.26
0.1–0.2	0.56	1.08
0.2–0.3	0.26	0.72
0.3–0.4	0.14	0.25
0.4–0.5	0.05	0.08
0.5–0.6	0.00	0.09
0.6–1	0.00	0.00
1	0.00	0.01

all consonant combinations of relatively equal similarity should be relatively equally represented as violations of the constraint. In fact, however, the distribution of violations of the OCP within a similarity group is unequal, and appears to be influenced by consonant frequency. Table 11.4 shows consonant co-occurrence for violations of the OCP-Place constraint in Arabic divided into high-frequency and low-frequency consonant groups. Relative to their expected frequency (O/E), high-frequency consonant pairs are more likely to appear as OCP-Place violations than low-frequency consonant pairs. Since the O/E measure takes expected frequency into account, it appears that high-frequency consonant pairs are less constrained by the OCP than low-frequency consonant pairs. This sub-regularity can be accounted for by lexical processing models that use activation/competition.

Recall that lexical neighbourhoods are an important influence on lexical processing. Also note that high-frequency consonants are found in words in dense lexical neighbourhoods, and low-frequency consonants are found in words in sparse neighbourhoods. In this case, the frequency-based difference in O/E for equivalent OCP-Place violations can be seen as a lexical neighbourhood effect, or equivalently as the influence of particular word exemplars on the constraint. If a particular consonant pair is attested as a violation of OCP-Place, it is likely that the violation can serve as a template for other words to violate the constraint using the same consonant pair. Since high-frequency consonant pairs are more frequent than low-frequency consonant pairs, it is more likely that a high-frequency violation will be found in the lexicon. If this violation then serves as an analogical model, the effect will be one where the 'rich get richer', as the violation is supported by its lexical neighbours. In other words, already high-frequency consonant pairs gain an additional advantage due to neighbourhood density that will make them more robust violators of OCP-Place than low-frequency consonant pairs.

As in the case of the gradient nature of the constraint itself, the further gradient effects of position-in-word and consonant frequency on the segmental OCP provide additional evidence that the constraint is functionally motivated. If the OCP constraint were a purely formal feature co-occurrence restriction, these processing factors should have no effect on the constraint, and gradient interactions such as these should not be found.

4.2 The reduplication problem

There is one other unresolved challenge to a functionally based segmental OCP constraint. Languages sometimes employ reduplication as a morphological process. Reduplication creates sequences of repeated segments. It could be that reduplication shares some of the benefits of vowel harmony, as discussed by Kaun (this volume). However, reduplication has been a long-standing problem for connectionist models of language processing, so it would be worth considering more carefully how reduplication and processing interact. Using standard models and training procedures, neural networks cannot learn to generalise the copying process in reduplication to words outside of the training set. This contrasts with other morphological processes like affixation and nonconcatenative morphology, where generalisation to novel words is automatic (Gasser 1998). The lack of automatic generalisation for reduplication in connectionist models appears to be a substantial failure in their ability to provide a general processing model for reduplication. One solution to processing reduplication is to implement a special module in the model that performs the copying when it is needed. This module is similar to the recurrent layer of a recurrent network, providing a memory for the items to be reduplicated. Thus, in a language with reduplication, it may be that special processes are developed to circumvent the repetition problem for reduplicative morphemes.[4]

Though this solution might appear ad hoc, it actually sheds light upon an outstanding problem in the cross-linguistic pattern of similarity avoidance constraints. Recall that some languages appear to have an exception to the laryngeal or place-based OCP constraint for identical segments. The functional difficulty in processing identity, and only identity, can be avoided via a recurrent repetition structure of the same sort that is used for reduplication. While repeated identical segments can benefit from a repetition node, similar but nonidentical segments can never be aided in this manner, as they are not true repetitions. Whether the dissimilarity constraint is based on gestural or perceptual difficulty, there is no way to implement a special copying process for nonidentical segments, so highly similar but nonidentical segments will always be functionally bad. Since maximally similar but nonidentical segments are always the most restricted, the special needs of the processing model offers an account of a cross-linguistically true generalisation about long-distance segmental OCP effects that has not previously been explained (cf. MacEachern 1999). Treating certain OCP problems in parallel with reduplication has also been proposed in formal analyses of the OCP (see Gafos 1998 and Rose 2000 for discussion).

5 Conclusion

Segmental similarity avoidance constraints, like the OCP-Place constraint in Arabic, can be functionally motivated using current theories of language processing that predict difficulty in processing repetition. These constraints have been shown to be gradient, such that lexical co-occurrence patterns reflect degrees of well-formedness with respect to the functional constraint. The gradient nature of these constraints provides strong evidence for their functional motivation. For similarity avoidance constraints, the degree of co-occurrence is a function of similarity, both within individual languages and in typological patterns across languages (Frisch, Pierrehumbert, and Broe 2004; MacEachern 1999). Similarity plays a key role in the functional motivation of the constraint. To the extent that repeated items are similar, the repeated items will be difficult to individually activate and encode in the proper serial sequence in perception and production. Similarity leads to mutual activation and competition that interferes with correct identification and serialisation. To the extent that activation and serialisation are universal properties of language processing, segmental OCP constraints can be explained through universal forces that shape all languages.

References

Allopenna, P. D., J. S. Magnuson, and M. K. Tanenhaus (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language* 38: 419–39.

Bender, M. L. and H. Fulass (1978). *Amharic verb morphology*. East Lansing, MI: Michigan State University African Studies Center.

Berent, I. and J. Shimron (1997). The representation of Hebrew words: Evidence from the obligatory contour principle. *Cognition* 64: 39–72.

Berg, T. (1998). *Linguistic structure and change: An explanation from language processing*. Oxford: Clarendon Press.

Berg, T. and H. Abd-El-Jawad (1996). The unfolding of suprasegmental representations: A cross-linguistic perspective. *JL* 32: 291–324.

Berkley, D. M. (1995). Variability in obligatory contour principle effects. *CLS* 30: 1–12.

(2000). *Gradient OCP effects*. PhD dissertation, Northwestern University.

Boersma, P. (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Broe, M. (1995). Specification theory and Ngbaka cooccurrence constraints. Paper presented at the Mid-Continental Workshop on Phonology, Columbus, OH.

Buckley, E. (1997). Tigrinya root consonants and the OCP. *Penn Working Papers in Linguistics* 4: 19–51.

Cowan, J. (1979). *Hans Wehr: A dictionary of modern written Arabic*. Wiesbaden: Otto Harrasowitz.

Davis, S. (1984). *Topics in syllable geometry*. New York: Garland.

Dell, G. S. (1984). Representation of serial order in speech: Evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10: 222–33.

(1986). A spreading activation theory of retrieval in sentence production. *Psychological Review* 93: 283–321.

Dell, G. S., L. K. Burger, and W. R. Svec (1997). Language production and serial order: A functional analysis and a model. *Psychological Review* 104: 123–47.

Dell, G. S., C. Juliano, and A. Govindjee (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science* 17: 149–95.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science* 14: 179–211.

Elman, J. L., E. A. Bates, M. H. Johnson, A. Karmiloff-Smith, D. Parisi, and K. Plunkett (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Eriksen, C. and D. Schutze (1978). Temporal factors in visual information processing: A tutorial review. In J. Requin (ed.), *Attention and performance VII*. Hillsdale, NJ: Lawrence Erlbaum.

Frisch, S. A. (1996). *Similarity and frequency in phonology*. PhD dissertation, Northwestern University.

(2000). Temporally organized representations as phonological units. In M. Broe and J. B. Pierrehumbert (eds.), *Papers in laboratory phonology V: Language acquisition and the lexicon*. Cambridge: Cambridge University Press, 283–98.

Frisch, S. A., J. B. Pierrehumbert, and M. Broe (2004). Similarity avoidance and the OCP. *NLLT* 22: 179–228.

Frisch, S. A. and B. A. Zawaydeh (2001). The psychological reality of OCP-Place in Arabic. *Lg* 77: 91–106.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Lg* 47: 27–52.

Gafos, D. (1998). Eliminating long-distance consonant spreading. *NLLT* 16: 223–78.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology* 6: 110–25.

Gasser, M. (1998). Representation, generalization, and modularity in connectionist models of language. Paper presented at the Program in Linguistics Colloquium, University of Michigan.

Goldinger, S. D., P. A. Luce, and D. B. Pisoni (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language* 28: 501–18.

Goldsmith, J. (1979). *Autosegmental phonology*. New York: Garland.

Greenberg, J. (1950). The patterning of root morphemes in Semitic. *Word* 5: 162–81.

Guy, G. R. (1991). Explanation in variable phonology: An exponential model of morphological constraints. *Language Variation and Change* 3: 1–22.

Hartley, T. and G. Houghton (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language* 35: 1–31.

Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Hayward, K. and R. Hayward (1989). 'Guttural': Arguments for a new distinctive feature. *Transactions of the Philological Society* 87: 179–93.

Ito, J. and R. A. Mester (1986). The phonology of voicing in Japanese. *LI* 17: 49–73.

Jakobson, R., G. Fant, and M. Halle (1952). *Preliminaries to speech analysis*. Cambridge, MA: MIT Press.

Kanwisher, N. (1987). Repetition blindness: Type recognition without token individuation. *Cognition* 27: 117–43.

Koskinen, K. (1964). Kompatibilität in den dreikonsonantigen hebräischen Wurzeln. *Zeitschrift der Deutschen Morgenlandischen Gesellschaft* 144: 16–58.

Labov, W. (1971). Variation in language. In C. Reed (ed.), *The learning of language*. New York: Appleton-Century Crofts, 187–221.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H and H theory. In W. Hardcastle and A. Marchal (eds.), *Speech production and speech modeling*. Dordrecht: Kluwer, 403–39.

Luce, P. A. (1986). *Neighborhoods of spoken words in the mental lexicon*. (Research in spoken language processing technical report 6) Bloomington, IN: Speech Research Laboratory, Indiana University.

Luce, P. A. and D. B. Pisoni (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19: 1–36.

MacEachern, M. R. (1997). Laryngeal similarity effects in Quechua and Aymara. *WCCFL* 15: 337–51.

(1999). *Laryngeal cooccurrence restrictions*. New York: Garland.

MacKay, D. G. (1970). Phoneme repetition in the structure of languages. *Language and Speech* 13: 199–213.

MacKay, D. G. and M. D. Miller (1994). Semantic blindness: Repeated concepts are difficult to encode and recall under time pressure. *Psychological Science* 5: 52–5.

Marslen-Wilson, W. and A. Welsch (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* 10: 29–63.

McCarthy, J. J. (1986). OCP effects: Gemination and antigemination. *LI* 17: 207–63.

(1988). Feature geometry and dependency: A review. *Phonetica* 43: 84–108.

(1994). The phonetics and phonology of Semitic pharyngeals. In P. Keating (ed.), *Papers in laboratory phonology III: Phonological structure and phonetic form*. Cambridge: Cambridge University Press, 191–283.

McClelland, J. L and J. L. Elman (1986). The TRACE model of speech perception. *Cognitive Psychology* 18: 1–86.

Mester, R. A. (1986). *Studies in tier structure*. PhD dissertation, University of Massachusetts, Amherst.

Miller, G. A., G. A. Heise, and W. Lichten (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology* 16: 329–55.

Miller, M. D. and D. G. MacKay (1994). Repetition deafness: Repeated words in computer-compressed speech are difficult to encode and recall. *Psychological Science* 5: 47–51.

Morris, A. L. and C. L. Harris (1997). A sublexical locus for repetition blindness: Evidence from illusory words. In M. G. Shafto and P. Langley (eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum, 519–24.

Padgett, J. (1992). OCP subsidiary features. *NELS* 22, 335–46.

Pierrehumbert, J. B. (1993). Dissimilarity in the Arabic verbal roots. *NELS* 23: 367–81.

(1994). Syllable structure and word structure. In P. Keating (ed.), *Papers in laboratory phonology III: Phonological structure and phonetic form*. Cambridge: Cambridge University Press, 168–90.

Plenat, M. (1996). De l'interaction des contraints: Une étude de case. In J. Durand and B. Laks (eds.), *Current trends in phonology: Models and methods*. Salford, UK: University of Salford Press, 585–617.

Prunet, J.-F., R. Béland, and A. Idrissi (1999). The mental representation of Semitic words. Ms., University of Quebec, Montreal.

Rose, S. (2000). Rethinking geminates, long-distance geminates and the OCP. *LI* 31: 85–122.

Sevald, C. and G. S. Dell (1994). The sequential cueing effect in speech production. *Cognition* 53: 91–127.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial order mechanism in sentence production. In W. Cooper and E. Walker (eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Lawrence Erlbaum.

(1992). The role of word structure in segmental serial ordering. *Cognition* 42: 213–59.

Stemberger, J. P. (1983). *Speech errors and theoretical phonology: A review*. Bloomington, IN: Indiana University Linguistics Club.

(1991). Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory and Language* 20: 161–85.

Stemberger, J. P. and B. MacWhinney (1990). Form-oriented inflectional errors in language processing. *Cognitive Psychology* 18: 329–54.

Steriade, D. (1982). *Greek prosodies and the nature of syllabification.* PhD dissertation, MIT.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)considerations of context effects. *Journal of Verbal Learning and Verbal Behavior* 18: 645–59.

Wright, R., S. Frisch, and D. B. Pisoni (1999). Speech perception. In J. G. Webster (ed.), *Wiley encyclopedia of electrical and electronics engineering*. New York: John Wiley & Sons, 175–95.

Yip, M. (1989). Feature geometry and cooccurrence restrictions. *Phonology* 6: 349–74.

Zawaydeh, B. A. (1999). *The phonetics and phonology of gutturals in Arabic.* PhD dissertation, Indiana University.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition* 32: 25–64.

Notes

* I would like to thank Gary Dell, Bruce Hayes, Janet Pierrehumbert, Dan Silverman, and especially Robert Kirchner for their comments on this chapter. Parts of this work were presented at Johns Hopkins University, the University of North Carolina, and the 75th meeting of the Linguistic Society of America and I would like to thank those audiences for their comments. This work was supported in part by the *Language Learning* Research Assistant Professorship at the University of Michigan.

1. Note that in this case the presentation of a word from a dense neighborhood prompts a faster response than the presentation of a word from a sparse neighborhood, so it is not always the case that greater lexical density makes the psycholinguistic task more difficult. Rather, greater lexical density always leads to greater activation/competition between words, and the effect of that activation depends on the task.
2. Another solution to processing activation patterns over time is to add a special layer of nodes that represent a memory for the context preceding the current state of the network. This layer is called a *recurrent layer*, and is usually implemented as a layer that is identical in structure to the hidden layer. The recurrent layer records the activation level of each node of the hidden layer from the previous time step. This allows the model to have some knowledge of its previous state which can then be used to guide the activation pattern in the hidden layer to the next state (e.g. Elman 1990). A recurrent network such as this could be used in the task of encoding words (e.g. Dell, Juliano, and Govindjee 1993). Informally speaking, the processor knows what sequence of segments is trying to be produced from the input word, and the recurrent layer reflects how far in the sequence of internal steps of processing it has proceeded. This knowledge is then used to change to the appropriate internal state and produce the next segment in the sequence.
3. The gestural account of Arabic co-occurrence is not entirely suitable (McCarthy 1994) and perceptual factors may also be involved (Zawaydeh 1999).
4. In cases where reduplication does not preserve identity, the reduplication process results in a less grievous violation of similarity avoidance. In these cases, the functional difficulty is at least partially circumvented. To my knowledge, connectionist models of language processing have yet to be applied to cases of non-identity in reduplication.

Index