Contrast in Phonology: Theory, Perception, Acquisition

Edited by
Peter Avery
B. Elan Dresher
Keren Rice

Mouton de Gruyter

Contrast in Phonology

Phonology and Phonetics
13

Editor
Aditi Lahiri

Mouton de Gruyter
Berlin · New York

Contrast in Phonology

Theory, Perception, Acquisition

edited by

Peter Avery
B. Elan Dresher
Keren Rice

Mouton de Gruyter
Berlin · New York

Mouton de Gruyter (formerly Mouton, The Hague)
is a Division of Walter de Gruyter GmbH & Co. KG, Berlin.

⊗ Printed on acid-free paper which falls within the guidelines
of the ANSI to ensure permanence and durability.

Library of Congress Cataloging-in-Publication Data

Contrast in phonology : theory, perception, acquisition / edited by Peter Avery, B. Elan Dresher, Keren Rice.
p. cm.
Includes bibliographical references and index.
ISBN 978-3-11-019821-8 (hardcover : alk. paper)
1. Grammar, Comparative and general – Phonology, Comparative.
2. Minimal pair (Linguistics) 3. Speech perception. 4. Language acquisition. I. Avery, Peter. II. Dresher, Elan. III. Rice, Keren, 1949–
P217.52.C66 2008
414–dc22
2008031077

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

ISBN 978-3-11-019821-8
ISSN 1861-4191

Cover design: Christopher Schneider, Berlin.
Typesetting: OLD-Media OHG, Neckarsteinach.
Printed in Germany.

Table of contents

Acknowledgements

This volume grows out of a conference on contrast in phonology that was held at the University of Toronto in 2002. The conference was supported through funds from the Social Science and Humanities Research Council of Canada as well as from a grant from the Connaught International Symposium fund at the University of Toronto. We thank those sources for funding the conference. We would also like to take this opportunity to thank those who reviewed abstracts for the conference as well as the participants in the conference for a lively and stimulating event that led to quality papers.

We owe a deep debt of thanks to Chiara Frigeni. Chiara served as editorial assistant for this volume. It is through her persistence and perseverance that the book is now done. Many thanks, Chiara.

Contributors

Peter Avery	York University, Canada
Paul Boersma	University of Amsterdam, The Netherlands
Amanda Boomershine	University of North Carolina Wilmington, USA
Juli Cebrian	Universitat Autònoma de Barcelona, Spain
B. Elan Dresher	University of Toronto, Canada
Paola Escudero	University of Amsterdam, The Netherlands
Paula Fikkert	Radboud University Nijmegen, The Netherlands
Heather Goad	McGill University, Canada
Daniel Currie Hall	University of Toronto, Canada
Kathleen Currie Hall	The Ohio State University, USA
Gunnar Ólafur Hansson	University of British Columbia, Canada
Elizabeth Hume	The Ohio State University, USA
Keith Johnson	University of California at Berkeley, USA
Alexei Kochetov	University of Toronto, Canada
S.-Y. Kuroda	University of California, San Diego, USA
Clara Levelt	Leiden University, Leiden Institute for Brain and Cognition, The Netherlands
Jessica Maye	Northwestern University and the Northwestern Institute on Complex Systems, USA
Jeff Mielke	University of Ottawa, Canada
Keren Rice	University of Toronto, Canada
James M. Scobbie	Queen Margaret University, Scotland, UK
Jane Stuart-Smith	University of Glasgow, Scotland, UK
Daniel J. Weiss	Pennsylvania State University, USA

Introduction

Peter Avery, B. Elan Dresher, and Keren Rice

Since Saussure, it has been recognized that contrast is central to phonological theory. Goldsmith (1996), in his introductory article in *The Handbook of Phonological Theory*, identifies contrast as the question that "lies at the doorstep of phonemic theory." Contrast played an important role in the major phonological schools of the twentieth century, and is again the subject of renewed interest. Despite its centrality, few works have explicitly taken contrast itself as their central theme; this volume puts contrast at the center, so as to make explicit how it works and its importance to phonology.

In particular, we focus on the role that contrast in phonology plays in three areas: phonological theory (grammar), perception, and acquisition.

1. Phonological theory

This section is concerned with the role of contrast in phonological theory and the description of phonological systems. How is contrast determined in a given inventory? To what extent does it play a role in accounting for sound patterns in language? How is it represented? What is the role of noncontrastive features?

Dresher looks at how phonologists decide which feature specifications are contrastive and which are redundant in the phonemes of a given phonological inventory. He argues that phonologists have vacillated between two different and incompatible approaches to this question, one based on minimal pairs, and the other based on a hierarchy of features (Jakobson and Halle 1956). He argues that the former approach is fundamentally inadequate, despite its intuitive appeal, and that the latter approach is superior.

One consequence of adopting an approach to contrast that depends on a feature hierarchy is that the same inventory can be assigned different sets of contrastive feature specifications under different orderings of the features. It follows that the set of contrasts operative in a given inventory are not self-evident, and allow for variability, to the extent that the ordering of features can vary from one language to another.

Hall and Kuroda both address, from very different perspectives, the problematic behaviour of certain phonemes with respect to voicing assimilation.

Hall builds on the general approach to contrastive specification advocated by Dresher and investigates what role redundant features play in phonology. He formulates the strongest version of what he calls the *contrastivist hypothesis* as follows: "redundant features are not present in the phonological computation." He argues that Czech voicing assimilation demonstrates that this strong formulation is not correct. In particular, while an analysis employing minimally contrastive specifications can account well for various subtleties of the Czech voicing assimilation, it also incorrectly predicts that the Czech phoneme *ř* ([r̝]) should become [t] when devoiced; instead, it becomes a voiceless [r̝̊], which is not an underlying phoneme in Czech. To solve this problem Hall proposes a weaker version of the contrastivist hypothesis: "redundant features are not active (but may be present) in the phonological computation." That is, they may play a prophylactic role, preventing mergers that would be expected if only contrastive features were in play.

Kuroda shows that different processes affecting voicing in Japanese do not treat nasals, liquids, and glides in a consistent way: a rule of regressive voicing assimilation is triggered by voiced obstruents, nasals, liquids, and glides; progressive assimilation is triggered only by voiced obstruents and nasals; and rendaku (which causes voicing) is blocked only by voiced obstruents. Kuroda proposes a feature geometry that encodes dependencies that mirror the properties of the vocal tract. In this framework, he proposes that the equivalent of the feature [+voice] is contrastive in obstruents and nasals, but redundant in liquids and glides. In his analysis, progressive assimilation is triggered by contrastive [+voiced], regressive assimilation by any phonetically voiced segment (contrastively or redundantly voiced), and rendaku targets a level of the feature geometry that isolates voiced obstruents, to the exclusion of nasals and other sonorants. Despite the differences in their frameworks, both Hall and Kuroda make crucial use of a distinction between contrastive and redundant feature specifications, and both observe processes that refer to contrastive as well as redundant specifications.

Though making distinctions between contrastive and redundant properties of phonemes, both contributions also illustrate how these distinctions depend on a phonological analysis; they do not simply flow from the phonetics of the inventory. Identifying what the laryngeal contrasts are in Czech and Japanese is a function partly of the general theory and partly of the particular analysis.

Scobbie and Stuart-Smith take the idea of indeterminacy of contrast further, arguing that contrast must be treated as an inherent gradient phenomenon. They argue that ambiguity in deciding whether a surface contrast is phonemic or allophonic, or which properties are contrastive and which are redundant, is not something that the analyst or native speaker language learner can necessarily always resolve. This is particularly so with respect to contrasts that are "marginal" to the system, where marginality is a heterogeneous characteristic that can be due to diverse causes. They propose that "exemplar" approaches to phonological representation (Pierrehumbert 2001, Coleman 2002) might have the flexibility to account for what they call "quasi-phonemic" contrasts in a "fuzzy" inventory.

In the final chapter of this section, **Hansson** considers how contrast affects phonological systems, with special attention to the interplay between vowel harmony and the neutralization of lexical contrast. He observes a striking difference between consonant harmony and vowel harmony with respect to contrast. Consonant harmony often results in the neutralization of an underlying contrast; for example, sibilant harmony in Ineseño Chumash changes underlying /…s…ʃ…/ to […ʃ…ʃ…] as well as /…ʃ…s…/ to […s…s…]. A similar pattern of neutralization, however, is unattested for vowel harmony; there are no known cases, for example, of a language with backness harmony that neutralizes an underlying contrast between a front and back vowel, so that underlying /…æ…+…a…/ becomes […a…+…a…] and /…a…+…æ…/ becomes […æ…+…æ…]. Hansson proposes that the recoverability (Kaye 1974) of a neutralized underlying contrast is much easier in consonant harmony than in vowel harmony, because of the sizes of consonant and vowel inventories and the relative frequency of neutral segments in each type of harmony.

2. Perception

In recent years the effects of contrast on perception have been studied from various points of view. Conversely, perceptual explanations have been suggested for why some contrasts are less likely in certain positions. An important question is the role that non-contrastive features play in the perception of contrasts in first and second languages. Because of the special connection between contrast and perception, it is fitting that this be one of the main themes of this volume. The chapters dealing with perception focus on different aspects of contrast. While Boomershine, Hall, Hume, and Johnson examine surface contrasts that reflect different underlying relationships in different languages, Mielke focuses more on differences in perception that arise due to phonotactic differences between languages. Kochetov addresses perception in a different way, examining the relationship between vowel inventories and the existence of secondary articulations in a language.

Boomershine, Hall, Hume, and Johnson begin their chapter with a discussion of Trubetzkoy (1939), noting that he identifies native language contrasts as having an important influence on perception of speech sounds. They focus on one assumption that he makes, that different degrees of contrast may have different consequences for speech perception. These authors examine the impact of contrast versus allophony on the perception of speech sounds by Spanish-speaking and English-speaking listeners. More particularly, they examine the perception of three sounds, [d], [ð], and [ɾ], that group together differently in English and Spanish in terms of the type of contrasts they participate in. They conclude, supporting findings in the literature, that phonemic contrasts are more perceptually distinct than allophonic contrasts, with English speakers finding [d]/[ð] more perceptually distinct than Spanish speakers, while Spanish speakers found [d]/[ɾ] to be more distinct. Thus, phonemic contrast influences speech perception, and, in addition, surface phonetic detail influences perceptual discrimination judgments. The authors argue that this distinction between contrast and allophony is best accounted for by an exemplar model. Their conclusion is particularly important for an understanding of the role of contrast in perception, showing the important role of phonemic contrast, and, in addition, recognizing that allophony and non-contrastiveness are not treated in the same way.

Mielke, like Boomershine, Hall, Hume, and Johnson, deals with perceptual salience and contrast. He is concerned with the influence of perception on contrast and how contrast influences perception. He focuses specifically on a contrast between /h/ and its absence in four languages. Mielke finds that /h/ deletes in environments where it is perceptually weak cross-linguistically. Nevertheless, differences exist between languages, with /h/ being more perceptible by speakers of some languages than others. Mielke relates this difference to phonotactic restrictions in the different languages. He further argues that, in addition to acoustic factors, functional load has an influence on contrast, with increased functional load associated with contrast maintenance. He thus finds that a variety of factors are important in the preservation or loss of contrast.

Kochetov, too, studies perception, in this case focusing on misperception. He takes as his study the relationship between secondary articulations on consonants and vowel contrasts in phonological inventories. Kochetov argues that interactions between a speaker and listener/learner constrain the relationship between secondary articulations on consonants and vowel inventories, with languages with secondary articulations not having complex vowel systems and languages with complex vowel systems not having secondary articulations. He argues that limitations on production and perception create this tendency to avoid a language having both distinctive secondary articulation contrasts and multiple distinctions in rounding/backness and vice versa. These markedness effects are not part of universal grammar, he argues, but rather result from low-level interactions. He investigates this claim in a simulation between a speaker and a listener, and argues that there is perceptual confusion of vowels and secondary articulations; when both are present they are difficult to replicate, with frequent undershoot. No a priori knowledge of markedness is necessary. Thus, a contrast of the type investigated is very unlikely to develop, as a system of this sort will shift to a stable pattern.

3. Acquisition

The third major focus of this volume is on first and second language acquisition. Much research in child language has looked at the order of acquisition of contrasts; explaining the observed sequence is one of the main goals of this research. At the same time, there have been major advances in the study of the perception of contrasts by infants. Researchers in second language acquisition have devoted much attention to the perception of contrasts, and the extent to which this is disrupted by the different contrastive system of the first language.

3.1. First language (L1) acquisition

Research on infant perception has established that 6–8 month old infants can discriminate contrasts that are not used in the ambient language more easily than adults, and gradually lose this ability in the next few months. **Weiss and Maye** point out that there are also studies that show that some contrasts are difficult for infants to perceive, though adults whose native language uses these contrasts perceive them well. It follows that exposure to these contrasts facilitates their discrimination. Weiss and Maye consider the extent to which statistical learning might facilitate the perception of difficult contrasts. They design an experiment in which continua of synthetically manipulated tokens ranging from prevoiced to short-lag velar stops are presented to infants in two conditions: in one condition more tokens are chosen from the extremes of the continuum, simulating a bimodal distribution; in the other condition more tokens are selected from the middle of the continuum, resulting in a unimodal distribution. Infants exposed to the bimodal distribution indeed do better at discriminating test pairs of prevoiced and short-lag velar stops.

Being able to discriminate phonetic sounds is a prerequisite to acquisition of phonology. But being able to distinguish between two sounds in a phonetic discrimination task does not mean that infants are able to store or represent these contrasts in their developing phonology. Thus, it has been shown that children's ability to discriminate sounds deteriorates significantly when the sounds are presented in the form of contrasting words. **Fikkert and Levelt** propose that there is a fixed order to the development of phonological point of articulation contrasts in words. In considering the patterns exhibited in their database of five Dutch children recorded weekly for about a year, they address some fundamental differences between child language phonology and adult phonology. In particular, child phonology is frequently characterized by an extensive "consonant harmony"; if this kind of harmony reflects universal markedness constraints, it is unexplained why it is unattested in adult phonology. They propose instead that "consonant harmony" in children results from a combination of factors. In early stages of acquisition, children cannot use point of articulation contrastively within a word, resulting in the appearance of harmony. Later, when children begin to make such contrasts, they extrapolate from their developing lexicon to formulate constraints that do not hold, or do not hold as strongly, of adult language. Thus, in their model, children's lexical representations are not adult-like to begin with, as is sometimes assumed, but develop as they are able to manipulate more contrasts independently.

3.2. Second language (L2) acquisition

The final three chapters deal with the acquisition of contrasts in a new language (the target language) and the role that the native language plays in this acquisition. Boersma and Escudero (Dutch learners of Spanish) and Cebrian (Catalan learners of English) look at the acquisition of vowel systems, while Goad (French and English learners of Thai) focuses on laryngeal contrasts. The three chapters all involve perceptual experiments. Taken as a whole, these chapters show that learners do not blindly map from their first language phonetics onto the second language phonetics, though Boersma and Escudero argue that in the initial stages this is the default strategy. Rather, learners dealing with a new phonemic system recalibrate their perception of it in a language-specific way.

Boersma and Escudero ask how learners whose native language has a large number of vowel contrasts (Dutch, in this case) will handle a system (Spanish) with a smaller vowel inventory. They shed light on the mechanisms responsible for the development of a separate "perception grammar" for the second language. They show that while beginning Dutch learners initially will tend to identify Spanish vowel tokens with the auditorily most similar Dutch vowels, over time they tune their perception of Spanish vowels to better align with the Spanish system of contrasts. Thus, proficient learners perceive a token [æ] as the vowel /ɛ/ when they are told it is a Dutch vowel, but as /ɑ/ when listening in "Spanish mode". Boersma and Escudero present an Optimality-Theoretic model of how learners converge on the appropriate perception grammar.

Cebrian looks at how Catalan speakers fare in the perception and production of front vowels in English, where there is a mismatch between the two languages. He shows that native Catalan speakers with little or no knowledge of English readily identify English [i] with Catalan /i/, but have no consistent Catalan mapping of English [ɪ], since Catalan has no such vowel phoneme. Interestingly, native Catalan learners of English do less well in categorizing English [i]. Cebrian finds that whereas native English speakers rely mostly on spectral cues to distinguish /i/ from /ɪ/, Catalan speakers rely more on duration. This study shows that where a new contrast (/i ɪ/) must be acquired, the perception system may have difficulty in reallocating the vowel space, even when one of the vowels (/i/) is an almost perfect fit with one of the vowels (/i/) in the native language system. This result underscores that, as Cebrian writes, "vowels are not acquired individually but as part of a system of contrasts with the consequence that the formation of one vowel category can directly affect the categorization of another vowel."

Goad examines what the acquisition of a new contrast can reveal about the nature of the underlying contrasts in the native language. She focuses on the acquisition of the three-way voicing contrast in Thai (voiced, voiceless unaspirated, and voiceless aspirated) by speakers of French and English, languages with a two-way voicing contrast. As English and French differ in the phonetic implementation of the voicing contrast, it is reasonable to assume that English speakers may perceive the Thai contrasts differently from French speakers. French has a contrast between a plain voiceless stop and a voiced stop; lacking aspiration, it is not surprising that French listeners have difficulty discriminating Thai voiceless unaspirated and voiceless aspirated stops. English voiceless stops are aspirated; nevertheless, English-speaking subjects fare no better than the French-speaking subjects in discriminating the aspirated and unaspirated voiceless stops.

These results can be explained if, as traditional phonological analyses have proposed, English speakers represent only the feature [voice] in their lexical representations, and not aspiration, encoded by the feature [spread glottis]. Goad concludes that lexical representations are abstract, and that a feature that is present in the phonetics, but not in lexical representations, does not necessarily aid in the perception of L2 contrasts that use that feature. Goad goes on to discuss results that appear to point in another direction, arguing that the position that English stops are unspecified for [spread glottis] can be upheld.

4. Summary

The three main areas covered in this volume – theory, perception, and acquisition – are tightly interconnected: research on the acquisition of a contrast may assign a central role to perception; neither of these can be studied in isolation from an account of the place of contrast in phonological theory and description. We hope that this volume will help to illuminate these interconnections and a variety of approaches in contemporary research on contrast, and that it will stimulate further research in these areas.

References

Goldsmith, John A.
1996 Phonological theory. In: John A. Goldsmith (ed.), *Handbook of Phonology*, 1–23. Oxford: Blackwell.

Jakobson, Roman, and Morris Halle
1956 *Fundamentals of Language*. The Hague: Mouton.

Pierrehumbert, Janet B.
2001 Exemplar dynamics: Word frequency, lenition and contrast. In: Joan Bybee and Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*, 137–157. Amsterdam: John Benjamins.

Coleman, John
2002 Phonetic representations in the mental lexicon. In: Jacques Durand and Bernard Lax (eds.), *Phonetics, Phonology, and Cognition*, 96–130. Oxford: Oxford University Press.

Kaye, Jonathan
1974 Opacity and recoverability in phonology. *Canadian Journal of Linguistics* 19: 134–149.

Trubetzkoy, N. S.
1939 *Grundzüge der Phonologie*. Göttingen: Vandenhoeck & Ruprecht.

Theory

The contrastive hierarchy in phonology[1]

B. Elan Dresher

1. Introduction

Since Saussure's famous statement that "dans la langue il n'y a que des différences" (Saussure [1916] 1972: 166),[2] the notion of contrast has been at the heart of linguistic theory. While it is relatively uncomplicated to determine whether or not two sounds are contrastive in a given language (though see Chomsky 1964), it is another matter to determine whether a given feature is contrastive in any particular situation. I will show that from the beginning phonologists have vacillated between two different and incompatible approaches to determining contrastiveness. Further, one of these approaches is provably untenable. The other is more promising, and in the second part of this paper I will look at some applications of it. Given the centrality of the issue, it is remarkable that it has received almost no attention in the literature. Recovering this missing chapter of phonological theory sheds new light on a number of old and new controversies over contrast in phonology.

2. Extraction of contrasts via fully specified minimal pairs

One approach to determining contrastiveness is based on pairwise comparisons of fully specified pairs of phonemes. For example, given segments /p b m/ as in (1a) and the binary features [voiced] and [nasal], /p/ and /b/ contrast with

---

1 I would like to thank the members of the project on Markedness and Contrast in Phonology in the Department of Linguistics at the University of Toronto for many kinds of help over the years, as well as the students in LIN 1221 in Fall 2001. This research was supported in part by grants 410–96–0842, 410–99–1309, and 410–2003–0913 from the Social Sciences and Humanities Research Council of Canada.

2 With reference specifically to speech sounds (Saussure 1972: 163), "Ce qui import dans le mot, ce n'est pas le son lui-même, mais les différences phoniques qui permettent de distinguer ce mot de tous les autres" [What is important in a word is not the sound itself, but the phonetic contrasts that allow us to distinguish this word from all the others].

respect to [voiced], /b/ and /m/ contrast with respect to [nasal], and /p/ and /m/ contrast with respect to both features. In the latter case it is not clear which of these features should be considered contrastive; in the case of /p b/ and /b m/, however, there is clearly only one contrastive feature in each case. Let us define a *minimal pair* as two members of an inventory that are distinguished by a single feature.[3] If we want to determine contrastive features starting from fully specified representations, it makes sense to focus on minimal pairs, because they reveal the contrasting features in the purest way. Pairwise comparison of the minimal pairs in (1a) yields the representations in (1b).

(1) French /p b m/ (Martinet 1964: 64)

a. Full specification

	p	b	m
voiced	–	+	+
nasal	–	–	+

b. Features distinguishing minimal pairs

	p	b	m
voiced	–	+	
nasal		–	+

c. Redundancy rules for (b)

[0 voiced] → [+ voiced]   [0 nasal] → [–nasal]

These are essentially the contrastive specifications proposed by Martinet (1964: 64) in his discussion of how to contrastively specify the consonants of Standard French. The redundancy rules in (1c) then fill in the unspecified features at some point before or during phonetic implementation.

Extraction of contrastive features from fully specified minimal pairs was evidently also used by Trubetzkoy ([1939] 1969), especially in the first part of his book. For example, Trubetzkoy (1969: 68–9) writes that in Standard French, *d* and *n* "are the only voiced dental occlusives". This fact is apparent from the fully specified feature values shown in (2a).[4] He observes further that "neither voicing nor occlusion is distinctive for *n*, as neither voiceless nor spirantal *n* occur as independent phonemes". That is, Trubetzkoy understands a feature to be distinctive in a phoneme if there is another phoneme in the language that

3 This kind of *featural* minimal pair differs from the usual sense of "minimal pair" in linguistics, which is a pair of *words* that differ by a single phoneme: for example, *sit* and *kit*, or *kick* and *kiss*. Determination of word minimal pairs does not require us to identify in what way (i. e., with respect to which features) one phoneme is crucially distinguished from another; it is enough to know that they are different.

4 These features are inferred from Trubetzkoy's discussion. Trubetzkoy assumes that the place feature is multi-valued; in the table, *dnt* = dental, *bil* = bilabial, *alv* = alveolar, and *dor* = dorsal.

is identical except for that feature. This notion of contrastiveness is consistent with extraction of contrastive features from fully specified minimal pairs. Since there is no voiceless *n* to make a minimal pair with *n* based on voicing, and no fricative *n* to make a minimal pair based on occlusion, it follows on this view that voicing and occlusion cannot be distinctive in /n/, as shown in (2b), where only specifications that are contrastive in this sense are retained.

(2) Some French consonants, bilateral oppositions (Trubetzkoy 1969: 68–69)

a. Full specifications

	t	d	n	p	b	m	s	z	k	g
voiced	–	+	+	–	+	+	–	+	–	+
continuant	–	–	–	–	–	–	+	+	–	–
place	dnt	dnt	dnt	bil	bil	bil	alv	alv	dor	dor
nasal	–	–	+	–	–	+	–	–	–	–

b. Contrastive specifications via minimal pairs

	t	d	n	p	b	m	s	z	k	g
voiced	–	+		–	+		–	+	–	+
continuant										
place	dnt	dnt	dnt	bil	bil	bil	alv	alv	dor	dor
nasal		–	+		–	+				

c. Determination of bilateral oppositions

Pair	In common	Shared with	Opposition
t ~ n	[dnt]	d	multilateral
t ~ d	[dnt, –nasal]	–	bilateral
d ~ n	[dnt, +voiced, –cont]	–	bilateral
d ~b	[+voiced, –nasal]	g	multilateral

This approach to determining contrastive features poses problems for one of Trubetzkoy's most fundamental concepts, the classification of oppositions into *bilateral* and *multilateral*. The members of a bilateral opposition are unique with respect to the set of features they share; in a multilateral opposition, the members do not share any set of features not also shared by at least one other member of the inventory. Further, when classifying an opposition as bilateral or multilateral, "Of course, only the phonologically distinctive properties are to be considered" (Trubetzkoy 1969: 68). However, Trubetzkoy cannot maintain this position, given his analysis of French.

Notice in (2b), for example, that with respect to the contrastive features, /d/ and /n/ share only the feature [dental], and this is true also of /t/ and /d/. Thus, /t/ ~ /d/ and /d/ ~ /n/ ought to be classified as multilateral oppositions. Trubetzkoy believes, however, that both /t/ ~ /d/ and /d/ ~ /n/ form bilateral oppositions in French, though he presents no evidence that this is the case. Thus, he concedes that sometimes noncontrastive features must be considered in assessing if an opposition is bilateral, as shown in (2c), where redundant but necessary features are underlined.

To maintain the more principled view that only contrastive features are to be considered in classifying oppositions, Trubetzkoy could either give up the idea that both the /t/ ~ /d/ and /d/ ~ /n/ oppositions are bilateral, and/or adopt a different criterion for determining contrastive features. We will see that there are grounds for doing both of these; in later sections of his book, Trubetzkoy takes quite a different approach to determining whether an opposition is bilateral or multilateral.

Jakobson (1949) apparently took a similar approach to specification of the features of Serbo-Croatian. I say "apparently" because he does not state explicitly how he arrived at his specifications, but we can work backwards to infer what the method was. I present his specifications of oral and nasal stops (only features relevant to this example are included). The shaded squares are those that Jakobson leaves unspecified. They are precisely the specifications that do not distinguish between minimal pairs.[5]

(3) Specifications of oral and nasal stops

	p	b	m	t	d	n	ć	đ	ń	k	g
voicing	–	+		–	+		–	+		–	+
nasality		–	+		–	+		–	+		
saturation	–	–	–	–	–	–	+	+	+	+	+
gravity	+	+	+	–	–	–	–	–		+	+

5 An exception is the specification of /m/ as [–saturation]. Since /m n ń/ are the only [+nasal] segments, the features [saturation] and [gravity] are needed only to distinguish between them. /n/ forms a minimal pair with /ń/ based on [saturation], and with /m/ based on [gravity]. As expected, /n/ is specified for both [saturation] and [gravity], and /ń/ is specified for [saturation] but not for [gravity]. By symmetry, /m/ ought to be specified for [gravity] but not for [saturation]. I suspect the specification of /m/ as [–saturation] is simply an error. I will show below that the minimal pairs method is not able to adequately distinguish all members of an inventory in the general case. Therefore, it is not surprising that Jakobson did not, or was not able to, adhere to it in a strict way.

2.1. An algorithm for extracting contrasts via fully specified minimal pairs

Extraction of contrastive features from fully specified minimal pairs can be implemented by a formal algorithm. Such an algorithm was proposed by Archangeli (1988). I will call this the Pairwise Algorithm, given in (4):

(4) Pairwise Algorithm (Archangeli 1988)
   a. Fully specify all segments.
   b. Isolate all pairs of segments.
   c. Determine which segment pairs differ by a single feature specification.
   d. Designate such feature specifications as "contrastive" on the members of that pair.
   e. Once all pairs have been examined and appropriate feature specifications have been marked "contrastive," delete all unmarked feature specifications on each segment.

An illustration of how this algorithm is supposed to work is given in (5). This is a typical five-vowel system characterized by the features [high], [low], and [back]. According to the Pairwise Algorithm, this five-vowel system, fully specified for these features as in (5a), would be underspecified as in (5b):

(5) Five-vowel system, features [high], [low], [back]
   a. Full specifications

	i	e	a	o	u
high	+	–	–	–	+
low	–	–	+	–	–
back	–	–	+	+	+

   b. Specifications according to the Pairwise Algorithm

	i	e	a	o	u	<u>Minimal pairs</u>
high	+	–		–	+	{i, e}; {o, u}
low			+	–		{a, o}
back	–	–		+	+	{i, u}; {e ,o}

2.2. Problems with extracting contrasts via fully specified minimal pairs

Deriving contrastive features from fully specified minimal pairs is unworkable for several reasons. First, it fails to adequately contrast segments that are not minimal pairs. Consider again example (1), French /p b m/. The contrastive specification in (1b) distinguishes /b/ from /p/ on one side and from /m/ on the other; but what about the contrast between /p/ and /m/? /p/ is [–voiced] and /m/ is [+nasal]; since these are not privative features but truly binary, we cannot conclude that the absence of a specification is necessarily distinct from a specification. Without running through the redundancy rules that tell us how to fill in missing specifications, we cannot decide if /p/ is distinct from /m/ or not. But then we have failed to arrive at a proper contrastive specification. Thus, the Pairwise Algorithm fails the Distinctness Condition proposed by Halle (1959), given in (6). Essentially, it says that 0 is not distinct from a plus or minus value in a binary feature system that is not privative. Examples are shown in (7).

(6) Distinctness of phonemes (Halle 1959: 32)
Segment-type {A} will be said to be *different from* segment-type {B}, if and only if at least one feature which is phonemic in both, has a different value in {A} than in {B}; i. e., plus in the former and minus in the latter, or vice versa.

(7) Examples of distinctness and non-distinctness (Halle 1959: 32)

a. {A} is not "different from" {C}

	{A}	{B}	{C}
Feature 1	+	–	+
Feature 2	0	+	–

b. All three are "different"

	{A}	{B}	{C}
Feature 1	+	–	–
Feature 2	0	+	–

One can argue about whether contrastive specifications ought to meet the Distinctness Condition (I think they do, but Stanley (1967) is one of a number who disagree). However, the minimal pairs method faces much more severe problems of adequacy, in that there are common situations in which it fails by any measure to distinguish the members of an inventory. There are two types of cases in which this occurs.

First, the Pairwise Algorithm will fail when there are too many features relative to the number of phonemes in the inventory. The Pairwise Algorithm succeeds in distinguishing the five vowels in (5) in the three-dimensional feature space defined by the features [high], [low], and [back]. But recall that the Pairwise Algorithm starts from *fully specified* specifications; the limitation of the feature space to three features is arbitrary and unjustified. Full phonetic specification implies that the vowels be specified for *all* vowel features, including [round], [ATR], [nasal], and so on. Even adding just one more feature, say [round], causes the Pairwise Algorithm to fail to differentiate the five-vowel system in (5). The results are shown in (8).

(8) Five-vowel system, features [high], [low], [back], [round]

a. Full specifications

	i	e	a	o	u
high	+	–	–	–	+
low	–	–	+	–	–
back	–	–	+	+	+
round	–	–	–	+	+

b. Specifications according to the Pairwise Algorithm

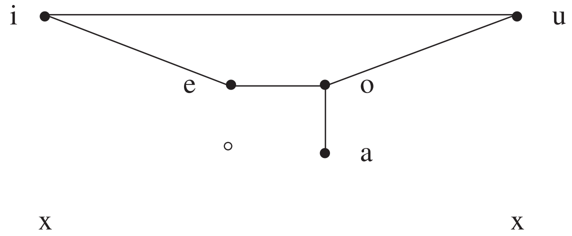
	i	e	a	o	u	Minimal pairs
high	+	–		–	+	{i, e}; {o, u}
low						
back						
round						

The only minimal pairs are {i, e} and {o, u}; the addition of the fourth feature turns what used to be minimal pairs into segments that are distinguished by more than one feature. The features [back] and [round] are each redundant given the other, but one of them has to be retained. In such cases, the Pairwise Algorithm cannot decide which feature to keep and which to discard. It is not clear, then, that an approach to contrast that relies on minimal pairs can handle even the simplest inventories, once all features are taken into account.

In these situations there is a remedy available, and that is to reduce the number of features before employing the Pairwise Algorithm. But then some other mechanism must operate in advance of the Pairwise Algorithm to make the same kinds of decisions it should be making. We shall see that when we spell out what this other mechanism is, the Pairwise Algorithm will be shown to be superfluous.

There is another type of case in which the Pairwise Algorithm fails, and this does not involve extra features, but rather the way in which the members of an inventory are dispersed over the space defined by the feature set. That the Pairwise Algorithm gives a contrastive specification at all, whether correct or not, is due to the connectedness of the paths through the space defined by the set of features. We can model the space corresponding to the inventory in (5) and the minimal pair paths through it with a diagram as in (9). The four nodes in the left half of the diagram are [–back], the four on the right are [+back]; the top four nodes are [–low], the bottom four are [+low]; and the peripheral four nodes are [+high], the inner four are [–high]. An empty circle ◦ represents an unoccupied node, and x represents an impossible combination of [+high, +low]. The members of this inventory are distributed in such a way that every phoneme except /a/ has two neighbours, creating enough minimal pairs to produce a result.

(9) Five-vowel system, features [high], [low], [back]



Archangeli (1988) points out that not every five-vowel system can be assigned a contrastive set of specifications by the Pairwise Algorithm. An example of such an inventory is the vowel system of Maranungku (Tryon 1970), given in (10).

(10) Maranungku, features [high], [low], [back]

a. Full specifications

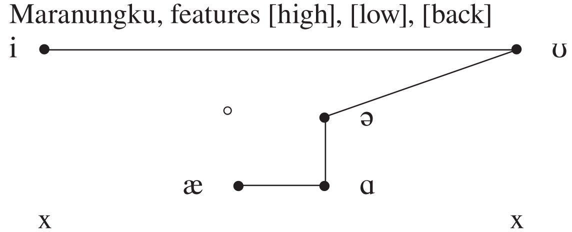
	i	æ	ɑ	ə	ʊ
high	+	–	–	–	+
low	–	+	+	–	–
back	–	–	+	+	+

b. Specifications according to the Pairwise Algorithm

	i	æ	ɑ	ə	ʊ	Contrasts
high				–	+	{ə, ʊ}
low			+	–		{ɑ, ə}
back	–	–	+		+	{i, ʊ}; {æ, ɑ}

In this case, /i/ and /æ/ have the same contrastive specification because they occupy parallel positions in a contrast, as shown graphically in (11), but have no other neighbours that could further differentiate them in terms of this algorithm.

(11) Maranungku, features [high], [low], [back]



Whether or not an inventory has paths that make its members distinguishable by the Pairwise Algorithm is an accidental property, and should not be the basis of a theory of contrast.

3. Specification of contrasts by a hierarchy of features

Another approach to contrast also has roots in the earliest work on contrast in phonology. In his discussion of the Polabian vowel system, Trubetzkoy (1969: 102–103) observes that a "certain hierarchy existed" whereby the back ~ front contrast is higher than the rounded ~ unrounded one, the latter being a subclassification of the front vowels. Trubetzkoy's rationale for this analysis is that the oppositions between back and front vowels are constant, but those between rounded and unrounded vowels of the same height are neutralizable (after *v* and *j* to *i* and *ê*). Also, palatalization in consonants is neutralized before all front vowels, as well as before "the maximally open vowel *a* which stood outside the classes of timbre".

We can understand Trubetzkoy's remarks as suggesting that the feature [back] has wider *scope* than does [rounded]: [back] is relevant to all the vowels in the inventory, apart from *a*, whereas [rounded] has contrastive force only among the front vowels. Scope differences can be equally understood in terms of *ordering*: the feature [back] is ordered ahead of [rounded], notated as [back] > [rounded]. Thus, the vowel inventory is divided on the basis of [back] before a contrast based on [rounded] is made. The statement that *a* "stood outside the classes of timbre" can be understood as implying that the feature that distinguishes *a* from all the other vowels, which we will here call [low], is ordered before all the other vowel features. The diagram in (12) gives a pictorial representation of the feature hierarchy suggested by Trubetzkoy's discussion.[6]

(12) Polabian (Trubetzkoy 1969: 102–3): [low] > [back] > [rounded]

Front		Back	
Unrounded	Rounded		
i	ü	u	
ê	ö	o	Nonlow
e		ɑ	
	a		Low

6 Trubetzkoy (1969: 103) further confirms that he does not consider rounding to be contrastive among the back vowels: "The properties of lip participation were phonologically irrelevant for the back vowels." This, despite the fact that the vowel he represents as *ɑ* "appears to have been pronounced as a back vowel without lip rounding" (Trubetzkoy 1969: 210 n. 21). Presumably, he considered the contrastive distinction between this vowel and the other back vowels to be based on height rather than lip rounding.

Elsewhere, Trubetzkoy (1969: 126) observes that Modern Greek has a bilabial stop /p/ and labiodental fricatives /f v/, and a postdental stop /t/ and interdental fricatives /θ ð/. Is the primary contrast one of occlusion (stop versus fricative) or of place? Trubetzkoy appeals to "parallel" relations between stops and fricatives at different places. In the sibilant and dorsal series (/ts s z/ and /k x ɣ/, respectively), the contrast is unambiguously one of occlusion, since stops and fricatives occur at exactly the same place of articulation. By parallelism, Trubetzkoy proposes that the same contrast should apply to the ambiguous cases, which leads to the conclusion that the minor place splits are phonologically irrelevant. The contrasts in the inventory can be pictured as in (13).

(13) Modern Greek: major place, voicing, occlusion > minor place[7]

	Labial	Apical	Sibilant	Dorsal
voiceless stops	p	t	ts	k
voiceless fricatives	f	θ	s	x
voiced fricatives	v	ð	z	ɣ

In French, however, Trubetzkoy (1969: 126) argues for a split labial series. "For in the entire French consonant system there is not a single phoneme pair in which the relation spirant : occlusive would occur in its pure form". Indeed, Trubetzkoy follows this analysis to its logical conclusion that there is *no* opposition between occlusives and spirants in French, because degree of occlusion cannot be regarded independently of position of articulation. Thus, Greek and French require a different ordering of the continuant feature relative to minor place features.

(14) French obstruents (based on Martinet 1964: 65)[8]

	bilabial	labiodental	apical	alveolar	pre-palatal	dorso-velar
voiceless	p	f	t	s	š	k
voiced	b	v	d	z	ž	g

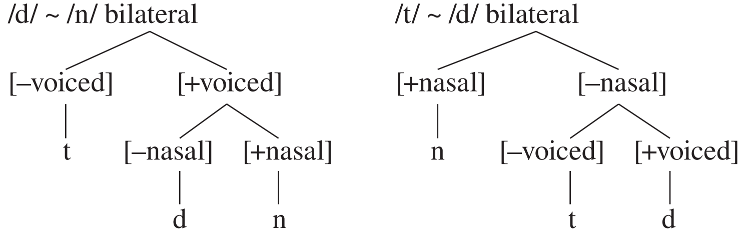
7 I substitute phonetic transcription for Trubetzkoy's Greek letters.
8 As Trubetzkoy does not give a chart, I adapt this one from Martinet (1964), whose analysis is clearly influenced by Trubetzkoy.

These analyses are inconsistent with Trubetzkoy's earlier discussion of bilateral oppositions in French. Whereas earlier he assumed that /t/ and /d/ were contrastively occlusive, according to his later analysis occlusion plays no role at all in the French consonant system. Moreover, in a hierarchical approach to contrastive specification, it is not at all clear that voicing is redundant for /n/, contrary to Trubetzkoy's assertion. For example, if [voiced] is ordered above [nasal], then the voicing contrast will include in its purview the nasal consonants as well, as shown in (15a). In this ordering, /d/ ~ /n/ participate in a bilateral opposition, but /t/ ~ /d/ do not. On the other hand, the features could be ordered as in (15b), in which case nasals are not specified for voicing, /d/ ~ /n/ do not form a bilateral opposition, but /t/ ~ /d/ do.

(15) French dental obstruents and nasals:

a. [voiced] > [nasal]: /d/ ~ /n/ bilateral

b. [nasal] > [voiced]: /t/ ~ /d/ bilateral



The tree diagrams in (15) show one important characteristic of specification by a top-down feature hierarchy: feature values that are logically redundant, such as [+voiced] for /n/, or [-nasal] for /t/, may still be designated as contrastive, if they are high enough on the hierarchy. A further difference from the previous method of determining contrastive features is that changes in the feature hierarchy can result in different contrastive specifications for the same inventory; the method based on fully specified minimal pairs always leads to the same contrastive specifications (where it works at all). Thus, the contrastive feature that distinguishes /p/ from /f/ in French is different from the one that distinguishes these phonemes in Greek; this result is not obtainable from making pairwise comparisons of fully specified segments.

3.1. An algorithm for specifying contrasts by a feature hierarchy

Let us consider a bit more explicitly how contrast is determined using a hierarchy of features. An algorithm corresponding to this idea, which we call the Successive Division Algorithm (Dresher 1998b, 2003), is given in (16).[9] The basic idea is that we start by assuming that all sounds form one phoneme. This primordial allophonic soup is divided into two or more sets by whichever distinctive feature is selected first. We keep dividing up the inventory into sets, applying successive features in turn, until every set has only one member.

(16) Successive Division Algorithm (SDA)
a. In the initial state, all tokens in inventory I are assumed to be variants of a single member. Set I = S, the set of all members.
b. i) If S is found to have more than one member, proceed to (c).
ii) Otherwise, stop. If a member, M, has not been designated contrastive with respect to a feature, G, then G is *redundant* for M.
c. Select a new *n*-ary feature, F, from the set of distinctive features.[10] F splits members of the input set, S, into *n* sets, $F_1 - F_n$, depending on what value of F is true of each member of S.
d. i) If all but one of $F_1 - F_n$ is empty, then loop back to (c).[11]
ii) Otherwise, F is *contrastive* for all members of S.
e. For each set F_i, loop back to (b), replacing S by F_i.

This algorithm solves the problems encountered by the Pairwise Algorithm. First, it adequately contrasts all members of an inventory, not just minimal pairs. Second, it is guaranteed to work in all inventories: it does not require any particular distribution of phonemes in the feature space. Third, it does not have to adopt auxiliary mechanisms for multiple logical redundancies; the ordering of the features in the hierarchy determines which features will be considered contrastive, and which redundant, in every case.

3.2. The rise and fall of the contrastive hierarchy

We have seen that Trubetzkoy's practice in *Grundzüge* does not point to a consistent method for determining contrastive features, but presupposes two

---

9 This algorithm is based on the method proposed by Jakobson and his colleagues in the 1950s (Jakobson, Fant and Halle 1952 and other works discussed in the next section). Dresher, Piggott and Rice (1994) call it the Continuous Dichotomy, echoing the "dichotomous scale" of Jakobson and Halle (1956).

10 I assume that the set of relevant distinctive features for a particular domain is given by some theory of that domain. By "new" feature I mean one that has not already been tried. Thus, the value of F changes every time this step reapplies (I assume some mechanism for keeping track of which features have already been tried, but do not specify it here).

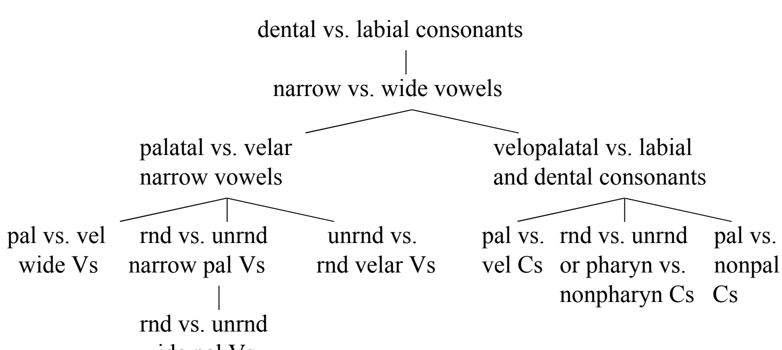
11 That is, if all members of S have the same value of F, then F is not contrastive in this set.

different and incompatible approaches: one based on fully specified minimal pairs, and another based on feature ordering. It is perhaps noteworthy that in the cases where his analysis is accompanied with a clear empirical motivation, it tends to be consistent with the ordering approach, assuming a "certain hierarchy" of features.

We also observed that Jakobson (1949) implicitly relied on fully specified minimal pairs in his analysis of Serbo-Croatian. But like Trubetzkoy, Jakobson was not consistent in this regard, but used a hierarchical approach in other work. Indeed, as far as I know, he was the first person to explicitly argue for a feature hierarchy as a way of determining contrastive specifications. The feature hierarchy was given a prominent place in Jakobson, Fant and Halle (1952) and Jakobson and Halle (1956). The latter refer to this hierarchy as the "dichotomous scale", and adduce "several weighty arguments" in support of this hierarchical approach to feature specification. One argument had to do with information theory, based on work with Colin Cherry (Cherry, Halle and Jakobson 1953). Their second argument involves language acquisition. They suggest that distinctive features are necessarily binary because of the way they are acquired, through a series of "binary fissions". They propose (1956: 41) that the order of these contrastive splits is partially fixed, thereby allowing for certain developmental sequences and ruling out others.

The sequence in (17), for example, concerns oral resonance (primary and secondary place) features. Jakobson and Halle propose that a contrast between dental and labial consonants must be made before a contrast between narrow and wide vowels; following the emergence of this contrast, children may either make a further contrast in the set of narrow vowels, or elaborate contrasts in the consonantal system. As the sequence proceeds, more choices become available.

(17) Predicted acquisition sequences (Jakobson and Halle 1956: 41)



The notion of a feature hierarchy that governs the operative contrasts in a phonological inventory has been fruitfully applied in the field of child language, where it is a natural way of describing developing phonological inventories, along the lines set out by Jakobson and Halle (1956) (Pye, Ingram and List 1987, Ingram 1989, Levelt 1989, Dinnsen et al. 1990, Dinnsen 1992, Fikkert 1994). However, it has had a rockier fate in phonological theory itself.

Despite their arguments for it, the contrastive hierarchy was employed inconsistently by Jakobson and Halle in the late 1950s. Perhaps the inconsistency is due to their failure to arrive at a single universal hierarchy that could apply to all the languages they studied. It appeared in the "branching diagrams" of Halle (1959). The use of "branching diagrams" was challenged on various grounds by Stanley (1967) and subsequently virtually disappeared from the theory of generative phonology. Yet, that was not the end of the story for the contrastive hierarchy.

4. Implicit hierarchies in phonological theory and descriptions

Though he opposed the branching diagrams, Stanley (1967: 408) nevertheless observed that "there is obviously some kind of hierarchical relationship among the features which must somehow be captured in the theory." This intuition has continued to haunt phonological theory, popping up in diverse and sometimes unexpected ways.

The notion of a hierarchy of features is evident in various forms of markedness theory, starting with Chomsky and Halle (1968) and Kean (1980). Here, too, the emphasis has been on finding a single universal hierarchy, though such a quest has not been entirely successful. The same can be said for feature geometry (Clements and Hume 1995, Halle, Vaux and Wolfe 2000) which builds a fixed hierarchy directly into representations. Less obviously, versions of underspecification theory (Kiparsky 1982, 1985, Archangeli 1984, Steriade 1987) also can be shown to assume some notions of a feature hierarchy.

Apart from explicit discussions of phonological theory, feature hierarchies are often implicit in at least a partial way in the common practice of phonologists from a variety of theoretical backgrounds when they are presenting segmental inventories. Tables of segmental inventories are often arranged in descriptive grammars in ways that suggest that certain features have wider or narrower contrastive scope than others, which amounts to a partial feature hierarchy.

Compare, for example, the inventory tables of Siglitun, an Inuit (Eskimo-Aleut) language spoken in the Canadian Arctic, and Kolokuma Ịjọ, an Ijoid (Niger-Congo) language spoken in Nigeria, given in (18) and (19), respectively. I present them as they are given in the sources (with some changes to the phonetic symbols but not to the arrangement). Note in particular the different placements of /l/ and /j/ in these charts. The chart of Ịjọ expresses a hierarchy in which the feature [continuant] has wider scope than such features as [sonorant] and [voiced], and [lateral] has wider scope than [nasal]. The Siglitun chart is not as overtly hierarchical, but it is clear that the feature [lateral] has very narrow scope, confined to making distinctions among apicals, whereas [nasal] is higher in the hierarchy. Apart from the nasals, the other sonorants are not set apart in Siglitun, suggesting that the feature [sonorant] is lower in the hierarchy than in Ịjọ.

(18) Siglitun consonants (Dorais 1990: 70)[12]

	Bilabial	Apical		Velar	Uvular
Stops	p	t		k	q
Voiced fricatives	v	l	j	ɣ	ʀ
Voiceless fricatives		ɬ	s		
Nasals	m	n		ŋ	

(19) Consonant phonemes of Kolokuma Ịjọ (Williamson 1965)[13]

<table>
<tr><td rowspan="4"></td><td colspan="2" rowspan="2">Plosive</td><td colspan="5">Continuant</td></tr>
<tr><td colspan="2">Fricative</td><td colspan="3">Sonorant</td></tr>
<tr><td rowspan="2">Vl.</td><td rowspan="2">Vd.</td><td rowspan="2">Vl.</td><td rowspan="2">Vd.</td><td colspan="2">Non-lateral</td><td rowspan="2">Lateral</td></tr>
<tr><td>Oral</td><td>Nasal</td></tr>
<tr><td>Labial</td><td>p</td><td>b</td><td>f</td><td>v</td><td>w</td><td>m</td><td></td></tr>
<tr><td>Alveolar</td><td>t</td><td>d</td><td>s</td><td>z</td><td>r</td><td>n</td><td>l</td></tr>
<tr><td>Back</td><td>k</td><td>g</td><td>(h)</td><td>(ɣ)</td><td>j</td><td>ŋ</td><td></td></tr>
<tr><td>Labio-velar</td><td>kp</td><td>gb</td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

12 I have simplified Dorais's *j/dj* and *s/ch* to *j* and *s*, respectively. As he makes clear, these are variants of single phonemes. Dorais does not usually indicate variants in his charts, and in related dialects in which /j/ has similar variants he lists only *j*. Therefore, I keep to the usual practice of representing a phoneme by one symbol.

13 I substitute *j* for Williamson's *y*. Williamson notes that Back = palatal, velar or glottal, Vl. = voiceless, and Vd. = voiced. Williamson mentions that some speakers have a marginal phoneme /ɣ/, but she omits it from the table. I have added it because it appears to be no less marginal than /h/, which is included.

So pervasive is the hierarchical approach to inventories that we can find it even in the descriptive practice of those who explicitly argue against it. In *A Manual of Phonology*, C.F. Hockett (1955: 173) reviews the different ways of construing the contrasts in the French obstruent system. He observes that place distinctions can make continuancy redundant (the solution favoured by Trubetzkoy and Martinet, shown in (14)); conversely, continuancy can be used to make minor place distinctions redundant (as in the analysis in (13) of Modern Greek). However, he continues: "Both of these decompositions of the French obstruents have the odor of pure game-playing…" He goes on to suggest that it is simply not possible to ever distinguish between features that are "determining" (that is, contrastive), and those that are "determined" (redundant).

Hockett's conclusion, however, is not consistent with his own practice in the rest of the *Manual*. If we can indeed make no distinctions between "determining" and "determined" features, it would be difficult to assign phonemic symbols to a set of allophones, let alone arrange them into neat schematic diagrams. But this Hockett does in his presentation of types of vowel and consonant systems.

For example, he observes (Hockett 1955: 84) that a 2x2 type of vowel system is widespread. He portrays such a system with the diagram in (20).

(20) A 2x2 vowel system (Hockett 1955: 84)

i	o
e	a

As examples, Hockett cites Rutul (Caucasian), in which the high back vowel is sometimes rounded, sometimes not, depending on environment; Fox and Shawnee (Algonquian), where the low back vowel is usually unrounded, though rounded in certain environments; and a number of other languages. It is particularly telling that the schematic diagram (20), for which he cites no specific language, has /o/ rather than /u/ aligned in the same row with /i/, and /e/ rather than /æ/ in the same row as /a/. He adds, "we class Fox as a two-by-two system despite the fact that the vowel classed as low back, /a/, is typically lower than that classed as low front, /e/". Though he lists no features, the arrangement in (20) can only mean that backness is the contrastive (determining) place/timbre feature, and that roundness is the redundant (determined) feature. The chart further indicates that there are only two phonological height classes, hence a single contrastive (determining) height feature; the phonetic height differences between /i/ ~ /o/ and /e/ ~ /a/ must therefore be considered redundant (determined). Thus, the chart indicates that it is not phonologically (i. e., contrastively) relevant that /o/ and /a/ may be phonetically lower than /i/ and /e/, respectively; the choice of these symbols suggests that /o/ and /e/ might even be at the same height phonetically, though functioning phonemically at different heights. Indeed, the schematization in (20) appears to be specifically chosen to show how the contrastive structure of a vowel system can differ from its surface phonetic appearance.[14]

Hockett (1955) makes decisions like these about which features are contrastive and which redundant throughout his survey of vowel and consonant systems. To take one more example involving vowels, he writes that a 3+1 system "is reported for Amahuaca" (21a), "though the /ɨ/ may be lower than /i u/, placing Amahuaca rather with Ilocano and others" (21b). He observes that in the Filipino (Austronesian) languages represented by (21b), /ə/ has fronted variants, and also higher central or back unrounded variants.

(21) Vowel systems: 3+1 vs. 2+1+1 (Hockett 1955: 84–85)

a. Amahuaca

i	ɨ	u
	a	

b. Ilocano

ɪ		u
	ə	
	a	

It is not important, for the purposes of this discussion, whether Amahuaca (a Panoan language of Peru and Brazil) is as in (21a) or (21b). What is important is that Hockett believes it is meaningful to assign it to one or the other. If there is indeed no way to distinguish between determined and determining features, we could not represent Ilocano as in (21b), since this diagram implies that the determining features of /ə/, for example, are that it is central and mid, even though it has variants that are front and others that are high. Similarly, Amahuaca could not be represented as in (21a) if /ɨ/ is phonetically lower than /i u/ to any extent, because that means making a decision that its centrality and non-lowness are its contrastive features and its lower height relative to the other high vowels is a redundant feature.

14 In this regard, Hockett is following in the tradition of Sapir (1925: 37–51); as Sapir puts it, "And yet it is most important to emphasize the fact, strange but indubitable, that a pattern alignment does not need to correspond exactly to the more obvious phonetic one."

5. Conclusion

I have argued that, despite the often-stated importance of contrast to phonological theory, methods for distinguishing between contrastive and redundant features in any given situation have been little discussed and seldom made explicit. The brief survey above has identified two different and incompatible methods for assigning contrastive features to segments that have been used intermittently in phonology. The first approach, based on fully-specified minimal pairs, has a certain intuitive appeal, but can be shown to be incapable of producing usable contrastive specifications in many situations. The second method, based on setting up a feature hierarchy in which the contrastive scope of features is determined by ordering, is a sounder method that can be applied to any phonological inventory. Thus, the main argument of this paper can be summarized as in (22).

(22) The Contrastive Hierarchy
Contrastive features are determined by establishing a feature hierarchy for a language and applying the Successive Division Algorithm.

It remains an empirical question whether this method of distinguishing between contrastive and redundant features, or any other method, is relevant to the operation of phonological systems. The cases discussed above suggest that contrast is important because contrastive features have a special role to play in phonological patterning. An explicit hypothesis based on this long-standing assumption is formulated as follows by Dresher and Zhang (2005):

(23) Contrast and phonological activity (Dresher and Zhang 2005)[15]
Only contrastive feature values are active in the (lexical) phonology.

The hypotheses in (22) and (23) are the subject of ongoing research in the project on Markedness and Contrast in Phonology at the University of Toronto (http://www.chass. utoronto.ca/~contrast/); see Avery and Rice (1989), Dresher (1998a, b, 2002, 2003), Dresher and Rice (2002), Dresher, Piggott, and Rice (1994), Hall (this volume), Rice (1993, 1997, 2002), and Rice and Avery (1995). Work in this framework also includes the dissertations by Avery (1996) on cross-linguistic voicing contrasts, Causley (1999) on segmental complexity and

15 Hall (2007: 20) calls this the Contrastivist Hypothesis, which he formulates as follows: “The phonological component of a language L operates only on those features which are necessary to distinguish the phonemes of L from one another.”

markedness in Optimality Theory, Dyck (1995) on phonetic and phonological patterning of Spanish and Italian vowels, Ghini (2001) on the phonology of Miogliola, Hall (2007) on phonological and phonetic aspects of the contrastivist hypothesis, with special application to Slavic languages, Walker (1993) on vowel harmony in Altaic, Wu (1994) and Zhou (1999) on Mandarin segmental phonology, and Zhang (1996) on Manchu-Tungusic languages; dedicated issues of the *Toronto Working Papers in Linguistics* (most recently Hall 2003 and Frigeni, Hirayama and Mackenzie 2005); and other references listed on the website.

References

Archangeli, Diana
1984 Underspecification in Yawelmani phonology and morphology. Ph. D. dissertation, MIT.

Archangeli, Diana
1988 Aspects of underspecification theory. *Phonology* 5: 183–207.

Avery, Peter
1996 The representation of voicing contrasts. Ph. D. dissertation, Department of Linguistics, University of Toronto.

Avery, Peter and Keren Rice
1989 Segment structure and coronal underspecification. *Phonology* 6: 179–200.

Causley, Trisha
1999 Complexity and Markedness in Optimality Theory. Ph. D. dissertation, Department of Linguistics, University of Toronto.

Cherry, E. Colin, Morris Halle and Roman Jakobson
1953 Toward the logical description of languages in their phonemic aspect. *Language* 29: 34–46. Reprinted in Roman Jakobson (1962), *Selected Writings I*, 449–463. The Hague: Mouton & Co.

Chomsky, Noam
1964 Current issues in linguistic theory. In Jerry A. Fodor and Jerrold J. Katz (eds.), *The Structure of Language*, 50–118. Englewood Cliffs, NJ: Prentice-Hall.

Chomsky, Noam and Morris Halle
1956 On the logic of phonemic description. Paper presented at the M. I. T. Conference on Speech Communication, June 16, 1956.

Chomsky, Noam and Morris Halle
1968 *The Sound Pattern of English*. New York: Harper & Row.

Clements, G. N. and Elizabeth V. Hume
1995 The internal organization of speech sounds. In John A. Goldsmith (ed.), *Handbook of Phonology*, 245–306. Oxford: Blackwell.

Dinnsen, Daniel A.
1992 Variation in developing and fully developed phonetic inventories. In: Charles A. Ferguson, Lisa Menn and Carol Stoel-Gammon (eds.), *Phonological Development: Models, Research, Implications*, 191–210. Timonium, MD: York Press.

Dinnsen, Daniel A., Steven B. Chin, Mary Elbert and Thomas W. Powell
1990 Some constraints on functionally disordered phonologies: Phonetic inventories and phonotactics. *Journal of Speech and Hearing Research* 33: 28–37.

Dorais, Louis-Jacques
1990 *Inuit Uqausiqatigiit: Inuit Languages and Dialects.* Iqaluit: Arctic College – Nunatta Campus.

Dresher, B. Elan
1998a Child phonology, learnability, and phonological theory. In: Tej Bhatia and William C. Ritchie (eds.), *Handbook of Language Acquisition*, 299–346. New York: Academic Press.

Dresher, B. Elan
1998b On contrast and redundancy. Paper presented at the annual meeting of the Canadian Linguistic Association, Ottawa. Ms., Department of Linguistics, University of Toronto.

Dresher, B. Elan
2002 Determining contrastiveness: A missing chapter in the history of phonology. In Sophie Burelle and Stanca Somesfalean (eds.), *Proceedings of the 2002 annual conference of the Canadian Linguistic Association*, 82–93. Montreal: Département de linguistique et de didactique des langues, Université du Québec à Montréal.

Dresher, B. Elan
2003 Contrast and asymmetries in inventories. In Anna-Maria di Sciullo (ed.), *Asymmetry in Grammar, Volume 2: Morphology, Phonology, Acquisition*, 239–57. Amsterdam: John Benjamins.

Dresher, B. Elan and Keren Rice
2002 Markedness and the contrastive hierarchy in phonology. http://www.chass.utoronto.ca/ ~contrast/.

Dresher, B. Elan and Xi Zhang
2005 Contrast and phonological activity in Manchu vowel systems. *Canadian Journal of Linguistics* 50: 45–82.

Dresher, B. Elan, Glyne Piggott and Keren Rice
1994 Contrast in phonology: Overview. *Toronto Working Papers in Linguistics* 13: iii–xvii.

Dyck, Carrie
1995 Constraining the phonology-phonetics interface, with exemplification from Spanish and Italian dialects. Ph. D. dissertation, Department of Linguistics, University of Toronto.

Fikkert, Paula
1994 *On the Acquisition of Prosodic Structure (HIL Dissertations 6).* Dordrecht: ICG Printing.

Frigeni, Chiara, Manami Hirayama and Sara Mackenzie (eds.)
2005 *Toronto Working Papers in Linguistics (Special Issue on Similarity in Phonology) 24.* Toronto: Department of Linguistics, University of Toronto.

Ghini, Mirco
2001 *Asymmetries in the Phonology of Miogliola.* Berlin: Mouton de Gruyter.

Hall, Daniel Currie (ed.)
2003 *Toronto Working Papers in Linguistics (Special Issue on Contrast in Phonology) 20.* Toronto: Department of Linguistics, University of Toronto.

Hall, Daniel Currie
2007 The role and representation of contrast in phonological theory. Ph. D. dissertation, Department of Linguistics, University of Toronto.

Hall, Daniel Currie
this volume Prophylactic features and implicit contrast.

Halle, Morris
1959 *The Sound Pattern of Russian: A Linguistic and Acoustical Investigation.* The Hague: Mouton. Second printing, 1971.

Halle, Morris, Bert Vaux and Andrew Wolfe
2000 On feature spreading and the representation of place of articulation. *Linguistic Inquiry* 31: 387–444.

Hockett, Charles, F.
1955 *A Manual of Phonology.* Baltimore: Waverly Press.

Ingram, David
1989 *First Language Acquisition: Method, Description and Explanation.* Cambridge: Cambridge University Press.

Jakobson, Roman
1949 On the identification of phonemic entities. *Travaux du Cercle Linguistique de Copenhague* 5: 205–213. Reprinted in Roman Jakobson (1962), *Selected Writings I*, 418–425. The Hague: Mouton & Co.

Jakobson, Roman, C. Gunnar M. Fant and Morris Halle
1952 *Preliminaries to speech analysis.* MIT Acoustics Laboratory, Technical Report, No. 13. Reissued by MIT Press, Cambridge, Mass., Eleventh Printing, 1976.

Jakobson, Roman and Morris Halle
1956 *Fundamentals of Language.* The Hague: Mouton.

Kean, Mary-Louise
1980 The theory of markedness in generative grammar. Ph. D. dissertation, MIT. Reproduced by the Indiana University Linguistics Club, Bloomington, Indiana.

Kiparsky, Paul
1982 From cyclic to Lexical Phonology. In Harry van der Hulst and Norval Smith (eds.), *The Structure of Phonological Representations (Part I)*, 131–176. Foris: Dordrecht.

Kiparsky, Paul
1985 Some consequences of Lexical Phonology. *Phonology Yearbook* 2: 85–138.

Levelt, C. C.
1989 An essay on child phonology. M. A. thesis, Leiden University.

Martinet, André
1964 *Elements of General Linguistics.* With a foreword by L. R. Palmer. Translated by Elisabeth Palmer. Chicago: University of Chicago Press.

Pye, Clifton, David Ingram and Helen List
1987 A comparison of initial consonant acquisition in English and Quiché. In: Keith E. Nelson and Ann Van Kleeck (eds.), *Children's Language (Vol. 6)*, 175–190. Hillsdale, NJ: Lawrence Erlbaum.

Rice, Keren
1993 A reexamination of the feature [sonorant]: The status of "sonorant obstruents". *Language* 69: 308–344.

Rice, Keren
1997 Japanese NC clusters and the redundancy of postnasal voicing. *Linguistic Inquiry* 28: 541–551.

Rice, Keren
2002 Vowel place contrasts. In Mengistu Amberber and Peter Collins (eds.), *Language Universals and Variation*, 239–270. Westport, CT: Praeger.

Rice, Keren and Peter Avery
1995 Variability in a deterministic model of language acquisition: A theory of segmental elaboration. In: John Archibald (ed.), *Phonological Acquisition and Phonological Theory*, 23–42. Hillsdale, NJ: Lawrence Erlbaum.

Sapir, Edward
1925 Sound patterns in language. *Language* 1: 37–51. Reprinted in Martin Joos (ed.), *Readings in Linguistics I*, 19–25. Chicago: University of Chicago Press, 1957.

Saussure, Ferdinand de
1916/1972 *Cours de linguistique générale*. Publié par Charles Bally et Albert Sechehaye; avec la collaboration de Albert Riedlinger. Éd. critique préparée par Tullio de Mauro. Paris: Payot.

Stanley, Richard
1967 Redundancy rules in phonology. *Language* 43: 393–436.

Steriade, Donca
1987 Redundant values. In: Anna Bosch, Barbara Need and Eric Schiller (eds.), *CLS 23: Papers from the 23rd Annual Regional Meeting of the Chicago Linguistic Society. Part Two: Parasession on Autosegmental and Metrical Phonology*, 339–362. Chicago: Chicago Linguistic Society.

Trubetzkoy, N. S.
1969 *Principles of Phonology.* Berkeley: University of California Press. Translated by C. Baltaxe from *Grundzüge der Phonologie*. Göttingen: Vandenhoek & Ruprecht, 1939.

Tryon, D. T.
1970 *An Introduction to Maranungku (Pacific Linguistics Series B, #14)*. Canberra: Australian National University.

Walker, Rachel
1993 Contrastive specification and vowel harmony mechanisms in the Altaic languages. Forum paper (M. A. thesis), Department of Linguistics, University of Toronto.

Williamson, Kay
1965 *A Grammar of the Kolokuma Dialect of Ijo*. Cambridge: Cambridge University Press.

Wu, Yuwen
1994 Mandarin segmental phonology. Ph. D. dissertation, Department of Linguistics, University of Toronto.

Zhang, Xi
1996 Vowel systems of the Manchu-Tungus languages of China. Ph. D. dissertation. University of Toronto.

Zhou, Hong
1999 Vowel systems in Mandarin languages. Ph. D. dissertation, Department of Linguistics, University of Toronto.

Prophylactic features and implicit contrast[1]

Daniel Currie Hall

1. Background

1.1. The contrastivist hypothesis

The premise behind contrastive specification in phonology is that the phonological rules of any language refer only to those features that are necessary to differentiate the phonemes of that language from one another – that is, distinctive features, *sensu stricto*. It is intuitively obvious that this is the minimum amount of information phonological representations can possibly contain: without at least this much information, there would be no way of assigning different phonetic realizations to different phonemes. At the opposite end of the scale, there is no readily identifiable upper bound; it is conceivable that phonological representations might contain infinitely detailed articulatory and acoustic descriptions of segments (see, e. g., Flemming (1995) and Boersma (1998, 2000) for proposals along these lines). In the investigation of phonological representations, then, it seems methodologically appropriate to take the contrastivist hypothesis as a starting point, and to retreat from it by whatever minimal steps are dictated by empirical evidence. In addition to a clear starting point, this approach provides reliably falsifiable hypotheses, as it is more generally possible to demonstrate empirically that a representation is too impoverished than to show that it is too rich.

This paper investigates the case of Czech (Slavic) voicing assimilation, in which purely contrastive specifications appear to be inadequate, and proposes a minimal retreat from the strongest version of the contrastivist hypothesis, which is stated in (1).

1 I am grateful to Veronika Ambros, Elan Dresher, Keren Rice, Bill Idsardi, Susana Béjar, Elizabeth Cowper, members of the phonology group at the University of Toronto, and audiences at various Montréal-Ottawa-Toronto Phonology Workshops for their helpful comments on earlier versions of this and related work. The research presented here has been supported in part by SSHRC grant #410–99–1309 to Keren Rice and Elan Dresher.

(1) Contrastivist hypothesis, strongest version:
Redundant features are not present in the phonological computation.

The Czech data, I will argue, can be accounted for by augmenting purely contrastive representations with what I will refer to as **prophylactic features** – redundant features that serve to prevent some phonological process from resulting in an unattested neutralization. Although these features are crucially present in segmental representations before and during the phonological computation, no phonological rule need make reference to them in any way. The introduction of prophylactic features thus represents a minimally weaker version of the contrastivist hypothesis, as in (2).

(2) Contrastivist hypothesis, weaker version:
Redundant features are not active (but may be present) in the phonological computation.

1.2 Defining contrast and redundancy

Before the contrastivist hypothesis can be tested, in either version, it must first be more clearly defined. Redundant features are to be excluded from the computation, but what constitutes a redundant feature? There are several different ways in which a piece of information about a segment may be considered to be predictable. For example, consider the universal implicational relations listed in (3):

(3) a. [+glottal, –continuant] → [–voice]
b. [+alveolar] → [+coronal]
c. [+liquid] → [+sonorant]
d. [+low] → [–high]
e. [+high] → [–low]
f. [αF] → not [–αF]

Some of these implications are grounded in articulatory necessity or near-necessity: it is impossible to produce voicing and glottal closure at the same time (3a), and it would be awkward to produce constriction at the alveolar ridge with any articulator other than the tip or blade of the tongue (3b). Others arise from logical necessity: liquids are a subclass of sonorants (3c); the body of the tongue can be low, or high, or neither, but not both (3d,e); and, much more generally, in a binary feature system, the presence of one value for a feature on a given segment precludes the presence of the opposite value of the same feature on the same segment.

Still other instances of redundancy are language-specific, being rooted in the shape of a particular phonemic inventory. For example, for each of the vowels in the inventory in (4a), there is one feature value by which it can be uniquely identified, and from which its values for all other relevant features can be predicted. In the !Xũ (Khoisan) pulmonic egressive stop series shown in (4b), some feature values can be predicted on the basis of the generalizations that (i) stops may be velarized or aspirated, but not both, and (ii) only coronal stops may be velarized.

(4) Language-specific redundancies

a. A common three-vowel inventory

i	u	[–back] → [+high, –low, –round]
		[+round] → [+high, –low, +back]
a		[+low] → [–high, –round, +back]

b. !Xũ pulmonic egressive stop series (Maddieson 1984: 421)

p	t	$t^{ɣ}$	k	[+velarized] → [–spread glottis]
p^h	t^h		k^h	[+spread glottis] → [–velarized]
b	d	$d^{ɣ}$	g	[+velarized] → [+coronal]

The language-specific redundancies in (4) are purely paradigmatic, in that the redundant features are predictable on the basis of other features on the same segment. Standard Bulgarian, (Slavic) in (5), offers an example of syntagmatic predictability. In preconsonantal position, the contrast between plain and palatalized stops is neutralized, and so any stop in this environment is predictably non-palatalized.

(5) Standard Bulgarian: /_ C → [–palatalized] (Kochetov 2002: 30, 34)

C1 \ C2	p	t	k	p’	t’	k’
p	appa	apta	apka	app’a	apt’a	apk’a
t	atpa	atta	atka	atp’a	att’a	atk’a
*p’	*ap’pa	*ap’ta	*ap’ka	*ap’p’a	*ap’t’a	*ap’k’a
*t’	*at’pa	*at’ta	*at’ka	*at’p’a	*at’t’a	*at’k’a

In the phonological literature, various sorts of restrictions on representations have been proposed as a means of eliminating, reducing, or otherwise dealing with redundancy. For example, the use of privative rather than binary features is a means of addressing the implication stated in (3f). Implications that arise from subset relations, such as the ones in (3b) and (3c), can be captured by means of feature-geometric representations (see, e. g., Clements and Hume 1996). Dependency structures like the ones in (6) encode the fact that dental, alveolar, and retroflex sounds are necessarily coronal (6a) and that liquids, nasals, and approximants are all sonorants (6b).

(6) a.
Place
|
Coronal
|
{Dental, Alveolar, Retroflex, …}

b.
Sonorant
|
{Liquid, Nasal, Approx., …}

Restrictions on representations may also limit the identity, configuration, or number of features that may coöccur on a segment. For example, Calabrese (2003) proposes universal markedness filters such as the ones in (7). In any given language, those filters that are active restrict the possible combinations of features on segments and dictate the phonetic interpretation of underspecified representations.

(7)
a. *[+high, –ATR]
b. *[+low, +ATR]
c. *[+high, +low]
d. *[-back, +round]
e. *[+back, –round]
f. *[–stiff vocal folds, +spread glottis]

A more abstract filter such as the one in (8) could, in combination with a theory of coronal unmarkedness (see, e. g., Kiparsky 1985; Avery and Rice 1989; Mester and Itô 1989; Paradis and Prunet 1989, 1991), provide an account of one of the implicational relations in the stop inventory in (4b). The filter in (8) asserts that a consonant cannot have both a marked primary place of articulation and a marked secondary place of articulation; if coronal is taken to be the unmarked value for C-place, then this filter will permit secondary articulations only on coronal consonants.

(8)
*Place
C-Place V-Place
| |
α β

In the framework of Optimality Theory, similarly abstract restrictions on segmental complexity can be imposed through the use of *Structure constraints (McCarthy and Prince 1993; Myers 1997; Causley 1999), which indiscriminately penalize specified structure of any kind. The effects of *Structure constraints are mitigated by faithfulness, as illustrated in (9).

(9)

/F, G/	Max(F)	*Structure	Max(G)
☞ [F]		*	*
[F, G]		**!	
[F, H]		**!	*
[]	*!		*
[G]	*!	*	

Any feature (such as [F] in the example in (9)) whose preservation is mandated by a faithfulness constraint outranking *Structure will be permitted to surface; other features (such as [G]) will be disallowed in the output. Conjoined *Structure constraints (Causley 1999: 194–196) can be used to enforce a complexity ceiling, so that [F] and [G] may not surface together on a single segment, even if each of them is allowed to appear alone, as illustrated in (10).

(10)

/F, G/	*Struc & *Struc	Max(F)	Max(G)	*Structure
☞ [F]			*	*
☞ [G]		*		
[]		*!	*!	*
[F, G]	*!			**
[F, H]	*!		*	**

In this paper, I will assume a version of the contrastivist hypothesis based on the Successive Division Algorithm of Dresher, Piggott, and Rice (1994) and Dresher (1998a, 1998b, 2002, 2003, 2004), a recent instantiation of an idea that has its origins in work by Trubetzkoy (1939), Cherry, Halle, and Jakobson (1953), Jakobson and Halle (1956), and Halle (1959). This algorithm is based on the insight that features are not contrastive or redundant in any absolute sense, but rather that their status is determined by the relative scope of the distinctions they mark. For example, in the /i, a, u/ vowel inventory in (4a), the feature [±high] is predictable if the values for [±low] are known, and the feature [±low] is predictable if the values for [±high] are known, but it does not follow from this that both [±high] and [±low] are redundant. The Successive Division Algorithm says that one of these features may take scope over the other, the feature with wider scope thus being contrastive, and the one with narrower scope redundant, because it cannot serve to make any distinctions within either of the subinventories delimited by the first feature. The algorithm is thus based on the notion of a contrastive hierarchy such as the one proposed by Jakobson and Halle (1956). It allows for cross-linguistic variation, prohibiting the specification of redundant features while allowing for the possibility that a feature that is redundant in one context (language, sub-inventory) may be contrastive in another.

The version of the Successive Division Algorithm assumed here uses privative features, and can be stated as in (11).[2]

(11) Successive Division Algorithm (adapted from Dresher 2003: 56)
  a. The input to the algorithm is an inventory (I) of one or more segments that are not yet featurally distinct from one another.
  b. If I is found to contain more than one phoneme, then it is divided into two subinventories: a marked set M, to which is assigned a feature [F], and its unmarked complement set $\bar{\mathrm{M}}$.
  c. M and $\bar{\mathrm{M}}$ are then treated as the input to the algorithm; the process continues until all phonemes are featurally distinct.

Two consequences of this implementation of the contrastivist hypothesis are of particular relevance to the phenomena to be considered here. First, the output of the algorithm will always contain exactly one segment with minimal feature specifications (i. e., only those features which were present on the input inventory). Considering the inventory as a whole, this means that there will be one entirely unspecified segment. In the initial input to the algorithm, no features have been assigned. After one division has been made, the inventory I has been divided into a marked subset M, all of whose members are specified with some feature F, and an unmarked subset $\bar{\mathrm{M}}$. At this point, there are two possibilities. If $\bar{\mathrm{M}}$ contains only one segment, then that segment will remain fully unspeci-

2 Dresher (1998a) uses privative features and refers to the algorithm as the Successive Binary Algorithm. Dresher (2003) presents a more general version with *n*-ary features, and calls it the Successive Division Algorithm. The Successive Binary Algorithm is simply a special case of the Successive Division Algorithm; it is binary because privative features make binary divisions. An earlier version of the algorithm is presented by Dresher, Piggott, and Rice (1994) as the Continuous Dichotomy Hypothesis.

fied, because it is already distinct from all other phonemes in the inventory. If $\bar{M}$ contains more than one segment, then $\bar{M}$ will become the input I to a new cycle of the algorithm; following the unmarked subset of the unmarked subset through each cycle, we will eventually reach a state in which this set is reduced to a single member.

Secondly, the order in which features are assigned by the Successive Division Algorithm has the potential to enter into an isomorphism with the hierarchical organization of features in the representations of individual segments. Although this possibility has not yet been thoroughly worked out – it is suggested briefly by Béjar (1998) – feature geometry offers a natural representational correlate of the scope relations that arise dynamically in the Successive Division Algorithm. For example, suppose that we have the consonantal (sub)inventory /p, t, k/ as shown in (12a). If the first division (12b) separates /p, k/ from /t/ by means of the feature Peripheral (Rice 1995), any marked feature, such as Labial, that then distinguishes between /p/ and /k/ can be represented geometrically as a dependent of Peripheral (12c). Thus the contrastive hierarchy maps onto a means of organizing features within a segment.

(12) a. {p, t, k}

b. {t} {p, k}
{t}	{p, k}
	|
	Peripheral

c.
{t}	{k}	{p}
	|	|
	Peripheral	Peripheral
		|
		Labial

There is, however, no guarantee that such a tidy mapping will be possible in all cases. For example, there is nothing in the algorithm in (11) to say that the inventory in (12) should not first be divided into {p} and {t, k} by the feature Labial, and the subinventory {t, k} then divided by Peripheral – in which case there would be no feature-geometric correlate to the contrastive scope of the features involved, because Peripheral would depend upon the absence of Labial rather than on the presence of any feature. The mapping from algorithm to feature geometry could in principle be enforced by imposing restrictions on the order of divisions, which could be based on inherent logical superset-subset relations between features (e. g., labial places of articulation are a proper subset of peripheral places of articulation), or on perceptual salience (as in the order of acquisition described by Jakobson and Halle 1956), or simply by a stipulated adherence to an abstract universal feature geometry. (See also Dyck 1995 for further discussion of some potential motivations for restrictions on the order of divisions.) In the Czech example discussed in the following sections, the necessary order of divisions for laryngeal features accords well with the laryngeal feature geometry proposed by Avery (1996); whether this correspondence is more than a happy accident remains to be seen.

The predictive value of the hypothesis embodied in the Successive Division Algorithm depends to a great extent on the assumption that featural representations constrain the range of processes in which segments can participate, and that phonological rules are maximally simple and general. If the phonological computation contains mechanisms that are excessively powerful or formally arbitrary, then it will be able, given even the most parsimonious segmental representations, to produce almost any conceivable pattern – for example, the Czech voicing assimilation facts discussed in the following section might be derived by having a separate rule for every possible sequence of two consecutive obstruents. In the account proposed here, I restrict the power of the phonological computation by assuming that rules are expressed in terms of spreading and delinking of monovalent features, and that the rules that derive voicing assimilation apply to whole classes of segments identified by their voicing features alone. The anomalous voicing behaviour of the two exceptional segments /v/ and /r̝/ must therefore be attributed to their representations, specifically their voicing features, and not to quirks in the rules that apply to them. Other comparably restrictive models of the phonological computation would also be possible, of course, and are predicted to encounter some version of the difficulty discussed in section 3 below.[3]

2. Czech voicing assimilation

In this paper, the empirical testing ground for the contrastivist hypothesis is the inventory of Czech consonants (shown in (13)) and their behaviour with respect to voicing assimilation. Hall (1998, 2003a) demonstrates that the Successive Division Algorithm can assign features based on Avery (1996) that ac-

3 Cf. Hall 2007a: § 5.3 and Hall 2007b for approaches to Czech voicing assimilation using Optimality Theory with privative and binary features, respectively.

count for the five distinct patterns of voicing behaviour in Czech consonants. (The relevant Czech data are drawn from de Bray (1969), Hála (1962), Kučera (1961), Palková (1994), Poldauf et al. (1994), Townsend (1990), and V. Ambros (p. c.).)

(13) The Czech consonant inventory
Orthographic forms are indicated in angle brackets where they differ from IPA.

		bilabial/ labiodental	dental/ alveolar	palatal/ postalveolar	velar/ glottal
stops	voiceless	p	t	c ⟨t'⟩	k
	voiced	b	d	ɟ ⟨d'⟩	g
affricates	voiceless		t͡s ⟨c⟩	t͡ʃ ⟨č⟩	
fricatives	voiceless	f	s	ʃ ⟨š⟩	x ⟨ch⟩
	voiced	v	z	ʒ ⟨ž⟩	ɦ ⟨h⟩
nasals		m	n	ɲ ⟨ň⟩	
trills			r	r̝ ⟨ř⟩	
lateral			l		
glide				j	

At the surface, nearly all clusters of Czech obstruents agree in voicing, as illustrated in (14):

(14) Czech obstruent clusters

a. *hezká*	[ɦeskaː]	'pretty' (fem. nom. sg.)	[*ɦezkaː, *ɦesgaː]
b. *pták*	[ptaːk]	'bird' (nom. sg.)	[*pdaːk, *btaːk]
c. *kde*	[gde]	'where'	[*kde, *gte]
d. *vstal*	[(f)stal]	'he got up'	[*vstal, *(v)ztal, *(v)sdal]
e. *lec+kdo*	[led͡zgdo]	'several people'	[*let͡sgdo, *let͡skdo]

As the data in (15) reveal, this agreement is the result of a process of regressive voicing assimilation, in which all obstruents in a cluster take on the voicing value of the rightmost one. (Word-final obstruent clusters are consistently voiceless.) (15) illustrates the forms of the prepositions *s* /s/ 'with' and *z* /z/ 'from'. Before sonorants, the two prepositions surface with their underlying voicing values (15a); before voiced obstruents, both are voiced (15b); and before voiceless obstruents, both are voiceless (15c).

(15) Regressive voicing assimilation

s /s/ 'with'			*z* /z/ 'from'		
a. sonorant:					
s mužem	[smuʒem]	'with man'	*z muže*	[zmuʒe]	'from man'
s lesem	[slesem]	'with forest'	*z lese*	[zlese]	'from forest'
b. voiced:					
s domem	[zdomem]	'with house'	*z domu*	[zdomu]	'from house'
s hradem	[zɦradem]	'with castle'	*z hradu*	[zɦradu]	'from castle'
c. voiceless:					
s polem	[spolem]	'with field'	*z pole*	[spole]	'from field'
s chybou	[sxibou̯]	'with mistake'	*z chyby*	[sxibi]	'from mistake'

However, there are two exceptions to this pattern. The segment /v/ (historically derived from Common Slavic **w*) undergoes regressive assimilation (and final devoicing), but does not spread its own voicing leftward:

(16) Behaviour of /v/

a. target:	*v tom*	[ftom]	'at that'
b. non-trigger:	*tvořit se*	[tvor̝it se]	'to take shape'
	≠ *dvořit se*	[dvor̝it se]	'to court, woo'

The other exception is /r̝/. Like /v/, /r̝/ undergoes but does not trigger regressive assimilation; unlike /v/, /r̝/ also undergoes a process of progressive assimilatory devoicing:

(17) Behaviour of /r̝/

a. No devoicing:	*řeč*	[r̝et͡ʃ]	'speech'
	břeh	[br̝ex]	'shore'
b. Regressive devoicing:	*tajnůstkářský*	[tajnuːstkaːr̝̊skiː]	'secretive'
c. Progressive devoicing:	*středa*	[str̝̊eda]	'Wednesday'
	před	[pr̝̊et]	'before, ago'

Czech consonants thus exhibit five different patterns of voicing behaviour, which are summarized in the table in (18).[4]

---

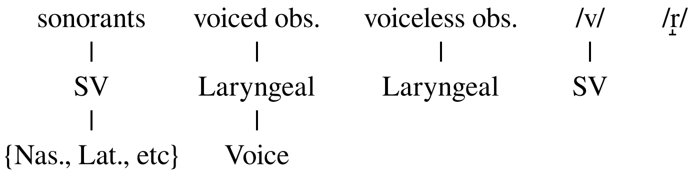
4 In some dialects, /v/ behaves in the same way as /r̝/. The variation is not crucial to the present question.

(18) Summary of voicing patterns

	DEFAULT REALIZATION	REGRESSIVE ASSIM.		PROGRESSIVE ASSIM.	
		TRIGGER	TARGET	TRIGGER	TARGET
sonorants	voiced	no	no	n/a	no
voiced obs.	voiced	yes	yes	n/a	no
voiceless obs.	voiceless	yes	yes	yes	no
/v/	voiced	no	yes	n/a	no
/r̞/	voiced	no	yes	n/a	yes

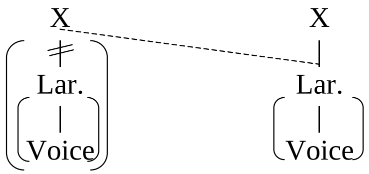
Hall (1998, 2003a) proposes five different sets of laryngeal features, shown in (19), to account for the five different patterns. These specifications represent a combination of Avery's (1996) Laryngeal Voice, Sonorant Voice, and Contextual Voice systems.

(19) Voicing feature specifications



Regressive assimilation is accomplished by leftward spreading of the Laryngeal node, with dependent Voice if present, replacing any existing Laryngeal node on the target.[5]

(20) Regressive voicing assimilation



5 All voicing alternations will be described here in derivational terms. In principle, the same feature specifications could be used in an Optimality Theoretic analysis, with spreading being driven by highly ranked AGREE[VOICE] and MAX[F] constraints, but cf. Hall (2007c) for a discussion of some of the problems with this approach.

Progressive devoicing of /r̝/ is the result of rightward spreading onto a consonant that does not already have any voicing specification, as in (21). (The presence of SV blocks progressive assimilation, but not regressive assimilation. True sonorants are protected from devoicing by the presence of a dependent on SV, but /v/ can be regressively devoiced.)

(21) Progressive voicing assimilation

X X

Lar.

The features in (19) can be assigned by the Successive Division Algorithm as follows. First, the true obstruents, characterized by their ability to spread their voicing properties, are identified by the feature Laryngeal. Within this marked set, the voiced obstruents are distinguished from their voiceless counterparts by the feature Voice. In the unmarked (non-Laryngeal) set, true sonorants and /v/ are distinguished from /r̝/ by their immunity from progressive devoicing, which is encoded by SV; /r̝/ is now fully distinct from all other segments.[6] Within the SV set, the true sonorants are distinguished from /v/ by the specification of features such as Nasal and Lateral, which are dependents of SV; when all the phonemes in the SV class have been differentiated, /v/ is the only segment left with a bare SV node.

This sequence of divisions is represented in (22).

(22) Dividing the inventory

	SV		LARYNGEAL	
	{v}	{r̝}	{p, t, s, …}	
NASAL, LATERAL, ETC.	{m, n, l, …}		{b, d, z, …}	VOICE

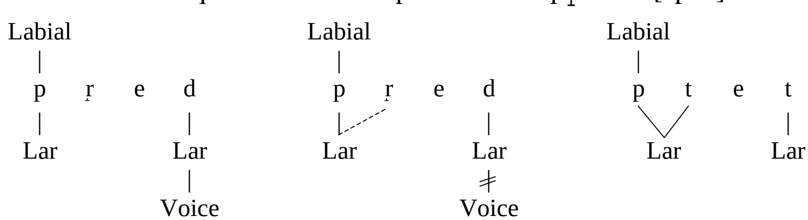
6 The contrast between sonorants and obstruents takes high scope in many languages. Where Czech is unusual is in having two successive divisions (based on SV and Laryngeal), each of which by itself would separate obstruents from sonorants, and which in combination isolate the anomalous /r̝/ from the rest of the inventory. Historically, this situation appears to have arisen from the fact that /r̝/, the Czech reflex of Common Slavic *r^j, has ceased to be a sonorant (unlike its counterpart in Slovak), but without becoming fully absorbed into the class of obstruents (unlike its counterpart in Polish).

3. The problem

The Successive Division Algorithm is thus capable of assigning the five distinct sets of voicing specifications needed to drive the rules in (20) and (21). However, the resulting representations make a prediction about the results of voicing assimilation that is highly counterintuitive, and, more importantly, false.

As expected, there is exactly one minimally specified segment in the inventory, namely /r̝/. Since /r̝/ is fully distinguished from all other segments on the basis of its voicing behaviour alone, there is no need for further features to be assigned to it. Among the set of voiceless obstruents, there will also be exactly one minimally specified segment – that is, one that bears only the feature Laryngeal, and no other features. Exactly which segment this is is difficult to determine; it is quite likely that there are several possible orders of divisions in the obstruent inventory that would produce several different sets of feature specifications that are equally compatible with the phonological behaviour of the segments in question. Crucially, though, there is no phonemic obstruent counterpart to /r̝/. The only phoneme with which /r̝/ alternates is /r/, which is a sonorant; /r/ becomes /r̝/ as part of a morphophonological palatalization process that also turns /t/ to /c/, /n/ to /ɲ/, /s/ to /ʃ/, and so on. Let us suppose, for the sake of argument, that the segment specified only with Laryngeal is /t/, which is, from a typological perspective, a plausible candidate for the status of least marked voiceless obstruent. In that case, when /r̝/ is devoiced, it will become featurally identical to /t/, as shown in the derivation in (23), in which underlying /pr̝ed/ 'before' surfaces as [*ptet].

(23) Unwanted consequences of underspecification: /pr̝ed/ → [*ptet]



Not surprisingly, this is incorrect. Devoicing /r̝/ does not produce [t], nor does it produce any other segment in the phonemic inventory of voiceless obstruents; it produces a voiceless postalveolar fricative trill [r̝̊].

What happens if we try to use the Successive Division Algorithm to specify /r̝/ for place and/or manner, so that it doesn't turn into [t]? Two possible sequences of divisions that would give features to /r̝/ are shown in (24). In (24a), the first feature assigned is Vibrant, which puts /r̝/ into a natural class with the sonorant /r/; in (24b), the first division is based on Palatal, which groups /r̝/ with the other segments that share its place of articulation.

(24) Alternative feature specifications

a. First division:

{r, r̝}		{p, t, v, m, z, …}
|		
Vibrant		

Second division:

{r̝}	{r}	{p, t, v, m, z, …}
|	|	
Vibrant	Vibrant	
|		
Palatal		

b. First division:

{c, ʃ, ʒ, r̝, …}		{p, t, v, m, z, …}
|		
Palatal		

Second division:

{r̝}	{c, ʃ, ʒ, …}	{p, t, v, m, z, …}
|	|	
Palatal	Palatal	
|		
Vibrant		

However, we know that there will always be one minimally specified segment in each inventory. So, in (24a), /r/ will be specified only for Vibrant, and there will be one other segment with no features at all. In (24b), there will be one segment with only Palatal, and one other segment with no features at all. In either of these cases, there will be at least two other segments that, like /r̝/, have no voicing features. There will be no way to explain why these segments do not pattern with /r̝/ with respect to voicing assimilation – we would have to write an arbitrary rule for progressive devoicing that specifically targets a segment with the features Palatal and Vibrant.[7]

4. The solution: Prophylactic features

It appears that we need redundant features to be present in the phonology to prevent /r̝/ from turning into some other phoneme when it is devoiced. Howev-

7 As Mercado (2002) has observed, evidence from phonological processes must take precedence over passive evidence of contrast (minimal pairs) in determining the order of divisions in the Successive Division Algorithm.

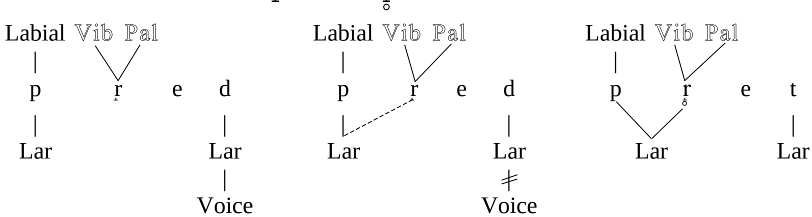
er, it is possible to account for the Czech facts with a minimal retreat from the strongest contrastivist hypothesis to the version in (2). The features Vibrant and Palatal (or any other feature that would suffice to differentiate an assimilated /r̝/ from all underlying obstruents) can be treated as prophylactic specifications on /r̝/. Although they are crucially present on the segment before and during the phonological computation, they are phonologically completely inert. They do not spread; they do not block spreading; they are not targets for spreading or delinking; they are not part of the structural description of any rule. In including them in the underlying representation for /r̝/, we abandon the hypothesis that redundant features are wholly absent from phonological representations, but we are able to maintain the position that the phonological computation does not refer to such features.

In effect, these prophylactic features, although the Successive Division Algorithm characterizes them as redundant, are required to be present precisely for the purpose of maintaining a contrast: they encode the fact that /r̝/ is different from /t/ (or whatever the least specified voiceless obstruent happens to be) in respects other than voicing behaviour. The features that would tell the speaker that a devoiced /r̝/ is different from a /t/ are contrastive elsewhere in the consonant inventory:

– /r̝/ differs in place from /t/ in the same way that /ɟ/ differs from /d/ or /c/ from /t/.
– /r̝/ differs in manner from /t/ in the same way that /r/ differs from /l/ or /n/ or /d/.

Given prophylactically specified Vibrant and Palatal on /r̝/, the derivation of /pr̝ed/ will proceed correctly as in (25). (The use of outlined letters here is intended to indicate the phonological invisibility of the prophylactic features.)

(25) Prophylactic features: /pr̝ed/ → [pr̝̊et]



Prophylactic features may also shed light on a similar phenomenon involving the preservation of contrast through unexpected non-structure-preservation in languages with asymmetric four-vowel inventories. D'Arcy (2002) observes that Yowlumne Yokuts (Penutian), Tagalog (Austronesian), and Tiwi (Australian) all have the vowel inventory shown in (26), and all have various lowering rules that change /i/ into [e] (in some cases also lowering /u/ to [o]).

(26) Asymmetrical four-vowel inventory

i		u
		o
	a	

If we assume that the lowering process is implemented as delinking of the feature High, then the representations in (27) are two plausible sets of feature specifications assignable by the Successive Division Algorithm.

(27) a.

i	u		o	a
High	High	Peripheral	Peripheral	

b.

i	u		o	a
High	High	Peripheral		Low

In (27a), delinking High from /i/ would produce a segment identical to /a/; in (27b), the same process would produce a segment identical to /o/.

An alternative set of specifications, in which Coronal rather than Peripheral is the marked place feature, would predict the correct results for lowering. These specifications are shown in (28).

(28)

u	i		o	a
High	High	Coronal		Low

Given the specifications in (28), lowering of /i/ is correctly predicted to generate a segment not found in the underlying inventory (bearing only the feature Coronal), and lowering of /u/ will produce [o]. However, in Yowlumne there is independent evidence from vowel harmony processes that Peripheral must be the marked place feature. Accordingly, Hall (2003b; 2007a: § 3.1) argues that the Yowlumne facts can be accounted for given the specifications in (27a) together with a prophylactic specification of Low on /a/. The feature Low is never referred to in the phonological computation, but at the end of a derivation it serves to distinguish an underlying /a/ from a lowered /i/, permitting the latter to be realized phonetically as [e]. This differs from the Czech case in that the prophylactic feature is specified not on the segment that undergoes the non-structure-preserving process, but rather on the other phoneme with which it is in danger of being neutralized.

The data from Czech and from Yowlumne thus tell us that purely contrastive specifications are insufficient. The strongest version of the contrastivist hypothesis cannot be maintained. However, prophylactic features offer a minimal retreat from strong contrastivism, by permitting some redundant properties of a segment to be present during, but invisible to, the phonological computation. Under this approach, the role of redundant features is still very narrowly constrained, in contrast to, for example, the approach taken by Nevins (2004), in which all segments are fully specified, but phonological processes vary parametrically as to whether they can 'see' all features or only the contrastive ones. In the system of prophylactic features proposed here, redundant features are never visible to any phonological rule, and not all redundant features are even present in the representation. From the cases discussed in this paper, it is not yet clear under precisely what circumstances a feature may have prophylactic status, but one commonality between the Czech situation and the Yowlumne is that in each inventory, the prophylactic feature has the effect of locating phonetically an underspecified segment that has a broad range of logically possible values and no single counterpart within any of the sets of segments with which it contrasts. In Czech, /r̝/ has no exact phonemic counterpart among either the obstruents or the sonorants; its prophylactic feature or features make it possible to identify this segment as being similar in place to /ʒ/ and similar in manner to /r/. In Yowlumne, the underspecified segment /a/ could, given where the division algorithm places it, just as well be an /e/ (a non-high counterpart to /i/) or an /ɤ/ (an unrounded counterpart to /o/). The prophylactic specification of Low identifies it as not being directly opposite any of the other segments in the inventory, but differing from them all in height. The role of prophylactic features in general, then, seems to be to preserve phonetic information that is irrelevant to the phonology, but otherwise in danger of being rendered irrecoverable by it. Since these features are crucial to the correct phonetic realization of the segments in question, the data from Czech and Yowlumne suggest that sometimes it is appropriate for features to be heard but not seen.

References

Avery, J. Peter

1996 The representation of voicing contrasts. Doctoral dissertation, University of Toronto.

Avery, J. Peter, and Keren D. Rice

1989 Segment structure and coronal underspecification. *Phonology* 6: 179–200.

Béjar, Susana

1998 Segmental inventory and the representation of voicing contrasts. Presented at the Montréal-Ottawa-Toronto Phonology Workshop, University of Ottawa, February 1998.

Boersma, Paul
1998 *Functional Phonology.* The Hague: Holland Academic Graphics.

Boersma, Paul
2000 Learning a grammar in Functional Phonology. In: Joost Dekkers, Frank van der Leeuw, and Jeroen van de Weijer (eds.), *Optimality Theory: Phonology, Syntax, and Acquisition*, 471–523. Oxford: Oxford University Press.

Calabrese, Andrea
2003 On the Evolution of the Short High Vowels of Latin into Romance. In: Ana Teresa Pérez-Leroux and Yves Roberge (eds.), *Romance Linguistics: Theory and Acquisition*, 63–94. Amsterdam: John Benjamins.

Causley, Trisha K.
1999 Complexity and markedness in Optimality Theory. Doctoral dissertation, University of Toronto.

Cherry, E. Colin, Morris Halle, and Roman Jakobson
1953 Toward the logical description of languages in their phonemic aspect. *Language* 29.1: 34–46.

Clements, G. N., and Elizabeth V. Hume
1996 Internal organization of speech sounds. In: John A. Goldsmith (ed.), *The handbook of phonological theory*, 245–306. Oxford: Blackwell.

D'Arcy, Alex
2003 Yawelmani reexamined: A contrastive analysis. Ms., University of Toronto.

de Bray, R. G. A.
1969 *Guide to the Slavonic Languages*. London: J. M. Dent and Sons.

Dresher, B. Elan
1998a Contrast in inventories. Paper presented at the Montréal-Ottawa-Toronto Workshop, University of Ottawa. (http://www.chass.utoronto.ca/~dresher/mot98.html)

Dresher, B. Elan
1998b On contrast and redundancy. Paper presented to the Canadian Linguistic Association, University of Ottawa.

Dresher, B. Elan
2002 Determining contrastiveness: A missing chapter in the history of phonology. In: Sophie Burelle and Stanca Somesfalean (eds.), *Proceedings of the 2002 Annual Conference of the Canadian Linguistic Association*, 82–93. Montréal: Département de linguistique et de didactique des langues, Université du Québec à Montréal.

Dresher, B. Elan
2003 The contrastive hierarchy in phonology. *Toronto Working Papers in Linguistics* 20: 47–62.

Dresher, B. Elan
2004 On the acquisition of phonological contrasts. In: Jacqueline van Kampen and Sergio Baauw (eds.), *Proceedings of GALA 2003, Volume 1 (LOT Occasional Series 3)*, 27–46. Utrecht: LOT Publications.

Dyck, Carrie
1995 Constraining the phonology-phonetics interface, with exemplification from Spanish and Italian dialects. Doctoral dissertation, University of Toronto.

Flemming, Edward
1995 Auditory representations in phonology. Doctoral dissertation, UCLA.

Hála, Bohuslav
1962 *Uvedeni do fonetiky češtiny na obecně fonetickem zakladě* [An introduction to the phonetics of Czech on general phonetic principles]. Prague: Nakladatelství Československé akademie věd.

Hall, Daniel Currie
1998 Contrastive specification for West Slavic voicing assimilation. Paper presented at the Montréal-Ottawa-Toronto Workshop, University of Ottawa.

Hall, Daniel Currie
2003a Laryngeal feature specifications in West Slavic languages. *Toronto Working Papers in Linguistics* 20: 93–114.

Hall, Daniel Currie
2003b Prophylaxis and asymmetry in Yokuts. In: Sophie Burelle and Stanca Somesfalean (eds.), *Proceedings of the 2003 Annual Conference of the Canadian Linguistic Association*, 97–108. Montréal: Département de linguistique et de didactique des langues, Université du Québec à Montréal

Hall, Daniel Currie
2007a The role and representation of contrast in phonological theory. Doctoral dissertation, University of Toronto.

Hall, Daniel Currie
2007b Contrastive specification in Optimality Theory: The revenge of the affricates. Presented to the Canadian Linguistic Association, University of Saskatchewan, Saskatoon, May 2007.

Hall, Daniel Currie
2007c Laryngeal underspecification and Richness of the Base. In: Sylvia Blaho, Patrik Bye, and Martin Krämer (eds.), *Freedom of Analysis?* 11–34. The Hague: Mouton.

Halle, Morris
1959 *The Sound Pattern of Russian: A Linguistic and Acoustical Investigation*. The Hague: Mouton.

Jakobson, Roman, and Morris Halle
1956 *Fundamentals of Language*. The Hague: Mouton.

Kiparsky, Paul
1985 Some consequences of Lexical Phonology. *Phonology Yearbook* 2: 85–138.

Kochetov, Aexei
2002 *Production, Perception, and Emergent Phonotactic Patterns: A Case of Contrastive Palatalization*. New York: Routledge.

Kučera, Henry
1961 *The Phonology of Czech*. The Hague: Mouton.

Maddieson, Ian
1984 *Patterns of Sounds*. Cambridge University Press.

McCarthy, John, and Alan Prince
1993 Prosodic Morphology: Constraint interaction and satisfaction. *Rutgers University Center for Cognitive Science Technical Report* 3. ROA #482.

Mercado, Raphael
2002 Contrastive theory and Québécois oral vowels. Presented at the Montréal-Ottawa-Toronto Phonology Workshop, McGill University, Montréal, February 2002.

Mester, R. Armin, and Junko Itô
1989 Feature predictability and underspecification: Palatal prosody in Japanese mimetics. *Language* 65.2: 258–293.

Myers, Scott
1997 OCP effects in Optimality Theory. *Natural Language and Linguistic Theory* 15: 847–892.

Nevins, Andrew Ira
2004 Conditions on (dis)harmony. Doctoral dissertation, Massachusetts Institute of Technology.

Palková, Zdena
1994 *Fonetika a fonologie češtiny s obecným úvodem do problematiky oboru* [The phonetics and phonology of Czech, with a general introduction to the problems of the field]. Prague: Univerzita Karlova.

Paradis, Carole, and Jean-François Prunet
1989 On coronal transparency. *Phonology* 6: 317–348.

Paradis, Carole, and Jean-François Prunet (eds.)
1991 *The Special Status of Coronals: Internal and External Evidence* (*Phonetics and Phonology* 2). New York: Academic Press.

Poldauf, Ivan, Jan Caha, Alena Kopecká, and Jiří Krámský
1994 *Anglicko-český česko-anglický slovník* [English-Czech Czech-English dictionary]. Prague: Státní pedagogické nakladatelství.

Rice, Keren D.
1995 On vowel place features. *Toronto Working Papers in Linguistics* 14.1: 73–116.

Townsend, Charles
1990 *A Description of Spoken Prague Czech*. Columbus, Ohio: Slavica.

Trubetzkoy, N. S
1939 Grundzüge der Phonologie. *Travaux du cercle linguistique de Prague* 7.

ROA = Rutgers Optimality Archive: http://roa.rutgers.edu/

Contrasts in Japanese: A contribution to feature geometry

S.-Y. Kuroda

The content of this paper was presented in May, 2002, at the Second International Conference on Contrast in Phonology. The paper is printed here in virtually the same form as was submitted to the editors in September of the same year for publication. Since this paper was written, I have published two papers related to the topics discussed below, Kuroda (2003/2004) and Kuroda (2006). In the first paper, I showed that the feature geometry I am proposing in this paper can account for the phenomenon of sonorant assimilation in Korean in a very succinct and revealing way, thus providing further support for the feature geometry I am proposing in this paper. In that account, projection reversal plays a crucial role. The mechanism of projection reversal is introduced in the present paper but it is discussed only tangentially; see section 9 below. As a consequence, important though it is for theoretical reasons, one may feel that more empirical support of the idea is called for. The account of Korean sonorant assimilation fills that gap.

In the second paper I have proposed a rather radical shift in our view of how to relate feature geometry to phonological/phonetic features. Geometry is now not considered as defining hierarchical relations among features directly. It only specifies hierarchical relationships of slots that mediate between abstract phonological structure and acoustic-articulatory/aerodynamic reality by means of phonology/phonetics interface conditions. Ironically, then, the term *feature* geometry may turn out to be justifiable only on historical/etymological grounds. Be that as it may, this shift, on the theoretical side, allows us to get rid of the redundancy rule (37) below from the geometry; the fact that necessitated this rule is now viewed as part of a more general fact that is to be accounted for by the way abstract phonological contrast relates to aerodynamic reality through interface conditions. On the empirical side, voicing and coda nasalization in Japanese can, under the new framework, be accounted for in a much simpler way than below in this paper. Incidentally, projection reversal also plays a crucial role in the newer account of Japanese phonology.

Thus, both theoretically and as a matter of the empirical account of Japanese phonology the present paper represents an earlier stage of my research into aerodynamic feature geometry and I wish to refer the interested reader to the papers cited above for my more recent thoughts on this area. Nonetheless, I think that the present paper still deserves some attention. For one thing, in order to appreciate the consequence of the rather drastic move I have taken for the advancement of the theory in Kuroda (2006), one needs to understand the empirical issues that motivated it. For another, aerodynamic feature geometry is still at an initial exploratory stage, being built on a very limited empirical basis. Whether and how its fundamental ideas are to be executed and implemented successfully in details much depends on expanded empirical studies to be done in this framework. The viability of the change proposed in my later papers for the fundamental conception of the geometry is also yet to be determined on the basis of such studies. From this perspective, the earlier, less compact but rather more transparent analysis done at an initial stage of the theory that is more directly and explicitly connected to the common idea of features should be kept available for future reference and consultation.

1. Introduction

In her paper on the issue of sonorants, Rice (1993: 309) introduces her main theme by comparing Japanese and Kikuyu with respect to the relation between the features [voice] and [sonorant]: "In Japanese as described by Itô & Mester…obstruents and sonorants do not form a natural class with respect to the feature [voice]… In contrast … in Kikuyu both voiced obstruents and sonorants count as voiced …"

(1) Rice (1993)
Japanese {voiced obstruents} ::: {sonorants}
Kikuyu {voiced obstruents, sonorants}

Here, "sonorants" includes "nasals". However, with respect to the problem of the relation between voiced obstruents and sonorants, the situation in Japanese is not as straightforward as Itô and Mester's description might suggest. There are three phenomena in Japanese phonology that relate to this issue:
- Sequential voicing in compound formation known as *rendaku*.
- A progressive voicing assimilation observed in the verb paradigm.
- A regressive process, which at first glance looks like a leftward nasalization triggered by a voiced segment.

This last process is exemplified in certain mimetic adverb constructions, as will be shown below. These three phenomena group sonorants differently with respect to voiced obstruents:

(2)	Rendaku/Lyman's Law	{voiced obstruents} ::: {nasals, liquids, glides}
	The verb paradigm	{voiced obstruents, nasals} ::: {liquids, glides}
	Mimetic adverbs	{voiced obstruents, nasals, glides} [no examples with liquids]

These phenomena were described in the early days of Japanese generative phonology in the 1960s and in more recent works in the 1980s and 90s.[1] However, I have doubts about some aspects of recent treatments of these phenomena, and wish to resurrect the spirit of the earlier treatment, recasting it in the framework of feature geometry, which was not available in the 60s. In this respect, the present study is an attempt to defend an old description and bring it abreast with theoretical advancement in phonology. In doing so, however, I have come to realize that the current conception of feature geometry does not suffice to achieve this goal in an insightful manner and have been led to the idea of a feature geometry more in conformity with the physical reality underlying phonology. This paper is a small beginning of explorations into a feature geometry structurally designed to be homomorphic to the aerodynamic architecture of the articulatory organs.

2. The difference between Itô & Mester's and my account

Itô and Mester maintain that rendaku voicing and progressive voicing observed in the verbal paradigm are manifestations of the same voicing process. However, nasals behave differently in these two phenomena. Let us first consider rendaku. Rendaku is commonly described as the voicing of an initial voiceless obstruent of the second component of a compound word. This is exemplified in (3), with the affected voiced segments boldfaced:[2]

(3) Rendaku voicing

susi	'sushi'	maki-**z**usi	'rolled-sushi'
kami	'paper'	ori-**g**ami	'origami' (folding-paper)
hasi	'chopsticks'	wari-**b**asi	'split-chopsticks'

1 Kuroda (1960, 1965), McCawley (1965, 1968) for the former and Itô's and Itô and Mester's works cited below for the latter.

2 Citation forms of Japanese examples are given largely following the conventions in Martin (1975:15). Phonetic/phonological representations at various levels of derivation are commonly, but not always, given between two slashes. At the phonetic level the manner and the place of articulation are not invariant under the rendaku voicing alternation due to allophonic and systematic phonemic variations. In particular, /b/ alternates with /h/ on the surface. See, for example, Itô & Mester (1986:52 f) for details.

Rendaku voicing, however, is not observed if the second component contains a non-initial voiced obstruent. This constraint is known as Lyman's Law:

(4) Lyman's Law: a constraint on rendaku

kaze	'wind'	kami-kaze (*kami-**g**a**z**e)	'divine wind'
kotoba	'speech'	onna-kotoba (*onna-**g**oto**b**a)	'women's speech'

We must note, however, that liquids and glides, as well as nasals, though they are phonetically voiced, do *not* block rendaku voicing:

(5) Liquids, glides as well as nasals do not block rendaku voicing

kokoro	'heart'	onna-**g**oko**r**o	'women's feeling'
kayu	'rice porridge'	asa-**g**a**y**u	'breakfast porridge'
tanuki	'raccoon dog'	oo-**d**a**n**uki	'big raccoon dog'

Next, let us observe the second phenomenon mentioned above, progressive assimilation in the verb paradigm. A stem-final consonant triggers voicing assimilation of /t/ in three suffixes /ta~da/ 'past/perfect', /te~de/ 'gerund' and /tari~dari/ 'representative'.[3] This process of assimilation is shown in the minimal pair given in (6), although a later process of lenition affects the velars, /k/ and /g/, and makes the effect of the voicing assimilation opaque. In (7), the stem-final /b/ gets voiced and then nasalized due to a general constraint, Coda Nasalization, to which I will return later.

Voicing after verb stems

(6)	kak-u 'write'	kak-ta	(> kai-ta)	'wrote'
	kag-u 'smell'	kag-ta > kag-da	(> kai-da)	'smelled'
(7)	tob-u 'fly, jump'	tob-ta > tob-da	(> ton-da)	'flew, jumped'[4]

In this process, nasals are grouped together with voiced obstruents and voice the following /t/:

(8) yom-u 'read' yom-ta > yom-da (> yon-da) 'read' (past)

3 For the REPRESENTATIVE *tari*, see Martin (1975:566).

4 The nasalization observed in (7) is due to Coda Nasalization, which nasalizes a voiced consonant in syllable coda position; see (29) below. If Coda Nasalization applies before the voicing of the suffix initial /t/, then it would also apply to /kag-ta/ and yield */kan-da/, unless we change the stem-final /g/ to /i/ (or insert /i/ between /g/ and /t/) before Coda Nasalization and complicate the voicing rule considerably, an unwelcome consequence.

However, glides and liquids as well as vowels do not cause this voicing assimilation:

(9)	kar-u	'trim'	kar-ta	(>kat-ta)	'trimmed'
	kaw-u > ka-u	'buy'	kaw-ta	(>kat-ta)	'bought'

(10)	tabe-ru	'eat'	tabe-ta	'ate'
	oki-ru	'wake up'	oki-ta	'woke up'

To sum up, nasals behave differently for the two phenomena we considered, rendaku voicing on the one hand and voicing assimilation in the verb paradigm on the other.[5] However, Itô and Mester take both of these phenomena as manifestations of a general process of assimilation that is triggered by voiced obstruents, excluding liquids, glides and nasals.[6] Itô and Mester deal with the voicing observed after a nasal in the past/perfect form like *yon-da* in (8) by means of another separate process of voicing, Post-Nasal Voicing:

(11) C -> [+voice] / [+nasal]___ (Itô & Mester 1986: 69, (42))

Nasals thus can be taken out of the triggers of voicing assimilation. This is why Japanese, in opposition to Kikuyu, is characterized as in (1) by Rice, based on Itô and Mester.

---

5 Besides the verb paradigm discussed above, we also observe the effect of Progressive Voicing triggered by prefix-final nasals in verbs with the implication of intense action such as the following:

Prefixed intense action verbs

bun-toru > bun-doru	'rob'
hum-sibaru > hun-zibaru	'fasten violently'

Itô & Mester (1996:24) cite *bun-doru* as an example of *verbal root compounding*, analyzing it as derived from *but-toru* 'strike+take'. I follow here the analysis given in *Kojien* of prefixed verbs. My view is that the data cited as examples of VERBAL ROOT COMPOUNDS in the recent literature (or SPECIAL CONSONANT-BASE VERB COMPOUNDS (Martin 1952: 89)) divide into compound verbs and prefixed verbs, though drawing a boundary between them raises delicate questions, the not unfamiliar tension one faces when one has to choose between analysis and etymology. A full-fledged discussion of this topic is beyond the scope of this paper.

6 Itô & Mester (1986:57) assume that "rendaku is essentially a morphological process introducing a linking morpheme in a certain morphological context," i. e., between two components of a compound word. Voicing spreads from this inserted linking morpheme to the initial segment of the second component.

However, Post Nasal Voicing is problematic. It is descriptively equivalent to the constraint Itô and Mester call *NT in later work:

(12) *NT (Itô & Mester 1995)
A nasal may not be followed by a voiceless obstruent.

I agree with Rice (1997) and Vance (2002) that *NT does not hold.[7] Some counterexamples:

(13) intiki 'trickery'; anta 'you'; kenka 'quarrel'; nantomo '(not) at all'

If the constraint *NT is out of place, we cannot have Post-Nasal Voicing. Thus, I conclude that we have to formulate a progressive assimilation rule which includes nasals as triggers.

There is an apparent contradiction in what I have said. On the one hand, I am claiming that we cannot have Post Nasal Voicing. On the other hand, I maintain that nasals trigger Progressive Voicing Assimilation. But there is a crucial difference in these two rules. Post Nasal Voicing, as intended by Itô and Mester, is a general rule with a phonotactic consequence. In contrast, Pro-

---

7 The voicing of the suffix-initial /t/ observed in the verbal morphology we are concerned with is an innovation that took place in Middle Japanese, when the original strictly open syllable structure of Old Japanese started to collapse. Before this innovation, the verb stem took the /i/-ending form (the *renyo*-form, in the traditional terminology) before the relevant suffixes. For example, we have the following historical derivation: *tobi-te* > *ton-de* 'fly'. The issue of *NT does not arise for Old Japanese, as there were no closed syllables in the language. One might be able to identify an intermediate stage between Old and Modern Japanese where arguably *NT held in the Yamato Stratum of the vocabulary. But the invasion of words of Sino-Japanese origin into common usage through time has made it impossible to clearly demarcate the division in Modern Japanese between the native and the Sino-Japanese stratum along with the historical origin. The existence of a Sino-Japanese stratum is arguably real for morphological reasons, but such a stratum can hardly justify a Yamato stratum with the phonological constraint *NT. Besides, violations of open syllable structure and *NT have also arisen within the etymologically native part of the vocabulary.

Itô & Mester (1986:69) originally introduced *NT as a constraint for the Yamato stratum, but they later dissociated it from such a sublexicon stratum in the constraint domain model of lexical organization (Itô & Mester 1995; the constraint domain of *NT contains, but is not limited to, the [Yamato] class. (ibid:823) The domain is specific for *NT and those items that violate it do not count as [Yamato].

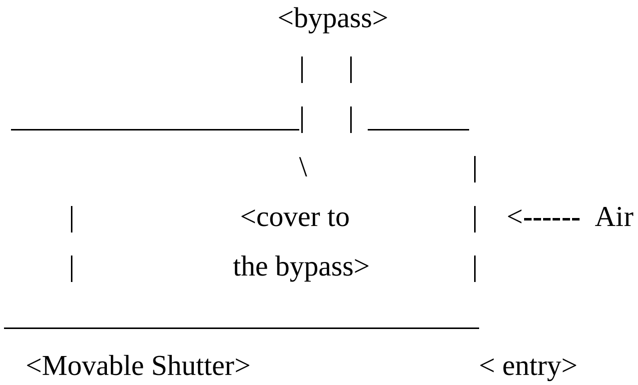
gressive Assimilation in the verbal paradigm is restricted to cross-morphemic context, or, more specifically, between verb stems and affixes.[8]

3. Feature geometry

3.1. Feature trees

At this point, let us shift our attention to feature geometry. My vision is to construct a feature geometry that is faithful to the aerodynamic design of the articulatory organ. The articulatory organ is schematized in the figure in (14).

(14)



The device consists of a main air path (the oral cavity), a bypass (the nasal cavity) and a movable shutter (the lips and tongue). Three parameters in this design are relevant:

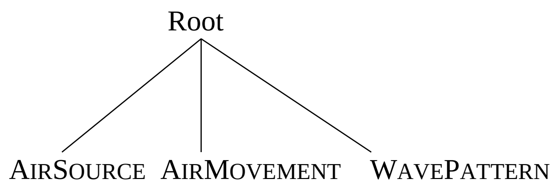
- The states of the entry to the main air path and the cover to the bypass. This parameter determines the quality of the AirSource.
- The degree and manner in which the shutter is opened/closed. This parameter determines the quality of AirMovement

8 Not all morphemes that attach directly to verb stems are affected by Progressive Voicing Assimilation; in fact, only three suffixes are: *ta, te, tari*. For example, the causative verb stem *sase* directly attaches to verb stems, but we do not get */yonzase/ < /yom/+/sase/ for 'make read.' Rather, the initial obstruent /s/ is elided after a stem-final consonant and we have /yom-ase/. It would seem fair to assume that there are some morpho-syntactic reasons why the three suffixes *ta*, *te* and *tari*, but not other suffixes, undergo voicing assimilation. Progressive Voicing Assimilation must specify a proper morpho-syntactic environment for its application, but I leave this matter aside.

– The positioning of the shutter. This last parameter determines the quality of the WavePattern.

The feature geometry I propose, Aerodynamic Geometry (ADG) is structurally homomorphic to this aerodynamic design of the articulatory organ. We have three nodes corresponding to these parameters immediately dominated by Root, as shown in (15):

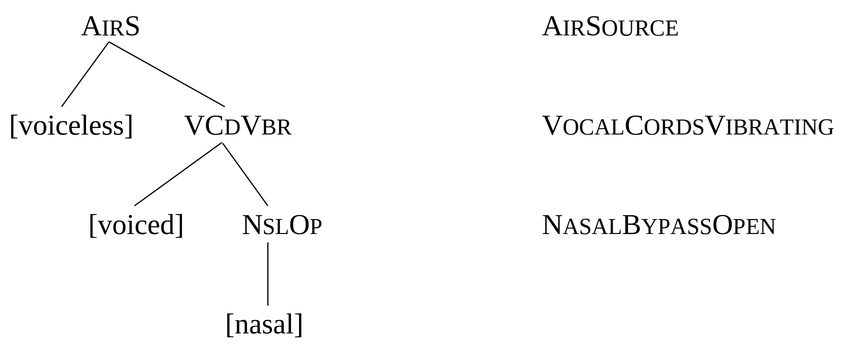
(15) Tree diagram for ADG: the top level



For the topic in Japanese phonology we are now concerned with, voicing and nasalization, what matters is the branch AirSource. AirMovement mostly concerns manner-of-articulation features, and I will return to it later. WavePattern largely concerns place-of-articulation features.

I assume that AirSource has the structure represented by the tree in (16).

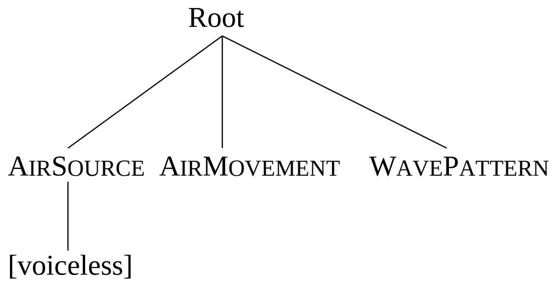
(16) Tree diagram for geometry under AirSource



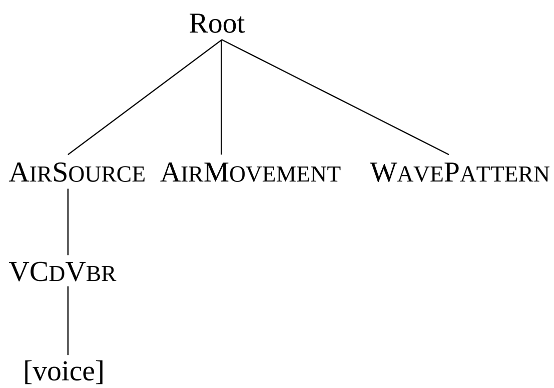
The phonetic/phonological features [voiceless], [voiced] and [nasal] are by definition default values of the three nodes, AirSource, VocalCordsVibrating and NasalBypassOpen, respectively.

All feature trees for phonological segments are, so to speak, embedded in this tree diagram in (16). We can see by inspection that three trees (17a-c) are embedded in (16). They are the relevant part of the feature trees for the segments /t/, /d/ and /n/.

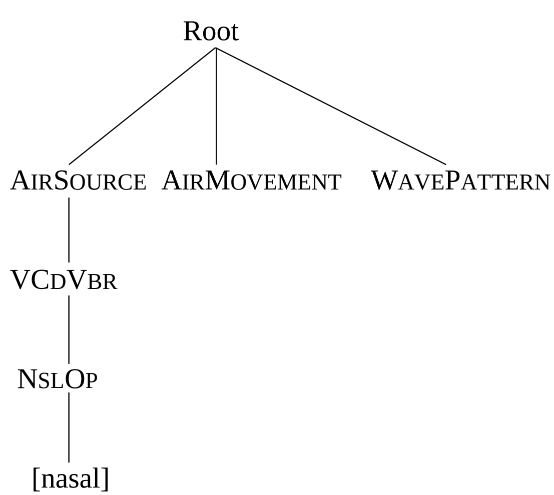
(17) a Feature tree for /t/



b Feature tree for /d/



c Feature tree for /n/



3.2. Redundancy and Underspecification

In this geometry the phonetic dependency of nasality on voicing (that is, the fact that nasal sounds are acoustically voiced) is NOT captured by making [voiced] a redundant feature of nasal sounds. Rather, the significance of the dependency of nasality on voicing is incorporated in the design of the geometry. The node VCdVbr signifies the vibrating vocal cords, an articulatory characteristic shared by non-nasal voiced sounds and nasal sounds. In this geometry, the feature called [voiced] signifies "vibrating vocal cords without the nasal bypass open". This is a characteristic of non-nasal voiced sounds.[9] The phonetic substance of the feature commonly called [voiced] is assigned to the node VCdVbr, rather than to the phonological/phonetic feature here labeled [voiced].

I introduce a familiar type of convention for the application of feature geometry.

(18) Underspecification convention: Default values are left unspecified in the underlying phonological representations.

Under this Convention, the relevant part of the feature trees for /t/, /d/, /n/ are given in (19):

(19) a /t/ = /…[AirS] …/

b /d/ = /…[AirS] …/
|
[VCdVbr]

c /n/ = /…[AirS] …/
|
[VCdVbr]
|
[NslOp]

9 As a matter of fact, the idea to group non-nasal voiced sounds and nasal sounds under one category has already been explored by Nasukawa (1998) in the framework of Element Theory. We see here how a particular design of feature geometry can implement the same idea as Nasukawa's.

4. Feature geometry and Progressive Voicing Assimilation

4.1. Preliminary observation: A linear account

Given this feature geometry and the above conventions, let us consider how the process of Progressive Voicing Assimilation in the verb paradigm can be accounted for. Observe that voicing and nasalization are phonologically redundant predictable features for the liquid /r/ and glide /w/ (as well as /y/) and for the vowels. For these sounds no contrast is relevant under the node AirS, and hence nodes under AirS must be left unspecified. On the other hand, the nasals /n/ and /m/ contrast with the non-nasals /d/ and /b/ under the node VCDVBR; hence, the nasals, in opposition to /r/, /w/ and vowels, share the feature VCDVBR with voiced obstruents in the underlying representations. Then, if we formulate Progressive Voicing Assimilation in terms of VCDVBR, the specified environment of the rule includes voiced obstruents and nasals but excludes /r/ and /w/ as well as the vowels. The initial segment for the suffix /ta~da/ is underlyingly unspecified under AirS. In linear phonology, we can formulate the rule as in (20):

(20) Progressive voicing assimilation in linear phonology

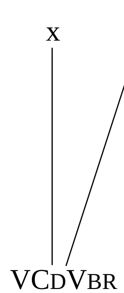
[ ] --> VCDVBR /[VCDVBR]___

By rule (20), the blank segment [ ] gets the node VCDVBR inserted and eventually gets its default value [voiced]. The segment /t/ is converted to /d/.

4.2. A non-linear account with ADG: Horizontal copying

It might seem that we can put (20) directly in the autosegmental formulation as in (21): VCDVBR spreads to the right.

(21) *Progressive Voicing Assimilation in autosegmental phonology (incorrect form)



However, this rule gives the right result only when the stem-final consonant is a voiced obstruent and not when it is a nasal, because in the latter case the suffix initial /t/ would be nasalized. The following examples illustrate this situation.

(22) Inputs to (21)

```
 tob-ta  'flew'        yom-ta   'read' (past)
  /b/    /t/             /m/     /d/
 AirS   AirS            AirS    AirS
   |                      |
VCdVbr                  VCdVbr
                          |
                        NslOp
```

(23) Outputs by (21)

```
OK                          Undesired
tob-da (> tonda)             *yom-na (>*yonna)
/b/      /t/                  /m/       /n/
AirS     AirS                 AirS      AirS

VCdVbr                        VCdVbr
                                |
                              NslOp
```

The operation we need here is not spreading. Rather, only the symbol VCdVbr must be copied to the right after any nodes dominated by it are delinked. This is the operation called copying introduced in Rice & Avery (1991: 106). Let us then represent this horizontal operation as follows:

(24) Progressive Voicing Assimilation

```
     x          y
     |          |
    AirS       AirS
     |          |
  VCdVbr -> VCdVbr
```

The application of this rule can be illustrated as follows. We have the underlying representations for *tonda* 'flew' and *yonda* 'read':

(25) Inputs to (24)

tob-ta		yom-ta	
/b/	/t/	/m/	/t/
AirS	AirS	AirS	AirS
|		|	
VCdVbr		VCdVbr	
		|	
		NslOp	

(24) applied to these forms yields the following outputs.

(26) Outputs by (24)

tob-da	(> tonda)	yom-da	(>yonda)
/b/	/d/	/m/	/d/
AirS	AirS	AirS	AirS
|	|	|	|
VCdVbr ->	VCdVbr	VCdVbr ->	VCdVbr
		|	
		NslOp	

5. Regressive voicing assimilation

I will now discuss a regressive process. Principal data for this process comes from a particular form of mimetic adverbs, which I call *ri*-extended mimetic adverbs. We can assume that the *ri*-extended forms are derived from two mora mimetic stems $C_1V_1C_2V_2$. These stems form mimetic adverbs either by reduplication, as shown in the first column of (28), or by the following morphological rule that inserts an underspecified consonantal segment C between the two stem moras:

(27) Morphological rule for *ri*-extended mimetic adverbs
$C_1V_1C_2V_2$ -> $C_1V_1CC_2V_2$ ri
where C is an unspecified consonantal segment

The phonetic forms of *ri*-extended mimetic adverbs are given in the second column of the table in (28).

(28) The *ri*-extended mimetic adverbs:

Reduplicated forms	*ri*-extended intensive forms		
$C_1V_1C_2V_2$-$C_1V_1C_2V_2$	$C_1V_1CC_2V_2$-*ri*		
hakihaki	hakkiri		'clearly'
yutayuta	yuttari		'leisurely'
boyaboya	boỹyari	(*boyyari)	'absent-mindedly'
yawayawa	yaw̃wari	(*yawwari)	'softly'
syoboshobo	yombori	(*syobbori)	'discouragedly'
sugasuga	sug̃gari	(*suggari)	'nicely slender'

The point of interest regarding the data given in (28) is this: if C_2 is voiced (including glides /y/ and /w/) the inserted unspecified consonantal segment /C/ gets nasalized. It appears that there is a regressive nasalization triggered by voicing, but such a process cannot be understood as a process of assimilation, and cannot be easily formalized in terms of feature geometry. However, as a matter of fact, this process of nasalization can be factored out into two processes, a regressive voicing assimilation and coda nasalization. In Japanese, voiced codas are necessarily nasalized, that is, we have a rule:

(29) Coda Nasalization

$\text{VCdVbr})_s \text{ -> [nasal]}$

Hence, in order to get the inserted C in (27) nasalized, it suffices to have C become voiced.[10] So, we have the following regressive voicing assimilation:

(30) Regressive Voicing Assimilation[11]

$$\begin{array}{ccc} \text{x} & & \text{y} \\ | & & | \\ \text{AirS} & & \text{AirS} \\ | & & | \\ \text{VCdVbr} & \leftarrow & \text{VCdVbr} \end{array}$$

The derivation of *syombori* is given in (31):

---

10 This analysis is essentially equivalent to that given by McCawley (1968:97) and Itô & Mester (1986:59, n. 14).

11 Regressive Voicing Assimilation is formally a mirror image of Progressive Voicing Assimilation. However, the former is a general rule while the latter, to recall, is restricted to cross-morphemic context; see notes 4 and 7.

(31) syombori

syobo => syoCbo-ri => syobbo-ri => syombo-ri

---C b --- ---C b --- ---C b ---

AirS AirS AirS AirS AirS AirS

| | | | |

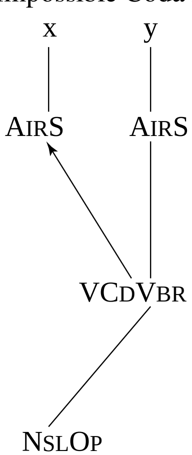
VCdVbr VCdVbr <-- VCdVbr VCdVbr VCdVbr

|

NslOp

We again have Copying, not Spreading. For, if VCdVbr were spread and became a multi-linked node, Coda Nasalization would not be able to nasalize only the left half of this node.

(32) Impossible Coda nasalization



6. The problem of sonorants

Recall, however, that this regressive voicing assimilation must be triggered by liquids and glides. Previously, we saw that Progressive Voicing Assimilation (24) is not triggered by the liquid /r/ or the glide /w/. This condition is satisfied because liquids and glides are underlyingly underspecified and not marked for VCdVbr. In contrast, the glides /y/ and /w/ must trigger Regressive Voicing Assimilation in order for the coda to be nasalized, as *bonyari* and *yanwari* in (28) show, which I repeat in (33):

(33) Part of (28) repeated

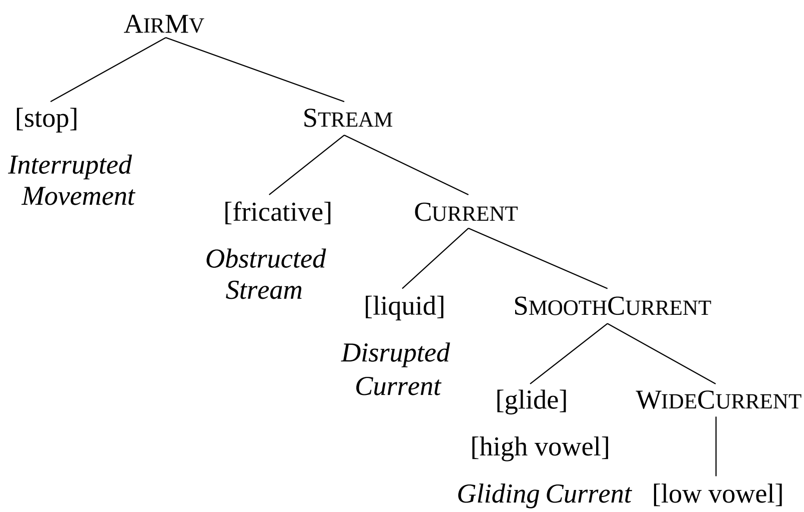
boyaboya	boỹyari	(*boyyari)	‘absent-mindedly’
yawayawa	yaw̃wari	(*yawwari)	‘softly’

For this reason, we need a special default rule for sonorants:

(34) Sonorant Default Voicing Rule
[Sonorant] => [VCdVbr] (sonorant = liquid/glide)

Now, how to deal with the problem of sonorants is another innovation I have in mind for the new design of feature geometry. I do not introduce [sonorant] as a feature, either as a root feature or as an organizing node. Instead, I design the structure under the node AirMovement as one that mirrors the sonority hierarchy of segments. This is shown in the tree diagram in (35).

(35)



The terms in square brackets represent features obtained as default values of the immediately dominating nodes. The terms in italics given under square brackets are “nicknames” for these features suggesting their aerodynamic characteristics in conformity with the design of the geometry.

I assume that the following rule in (36) replaces the informally stated rule (34) in our feature geometry:

(36) Sonorant Default Voicing Rule
Current => VCdVbr

This rule in effect interprets segments dominated by the node CURRENT as sonorant and specifies that sonorant segments are voiced. In the next three sections, I will discuss two general issues that arise in the design of feature geometry exhibited in the diagram (35).

7. Nasals as sonorants

The first issue concerns the generally held view that nasals are sonorants. The intent of the Sonorant Default Voicing Rule (36) is to formally characterize the informal concept of sonorants in our feature geometry by the node CURRENT. Then, the following redundancy rule can be taken as expressing this general view:

(37) Nasal Sonority Rule

NSLOP -> CURRENT

Nasals are usually grouped with oral stops. However, according to Shirô Hattori,

> "… since air cannot flow out [either through the oral or the nasal cavity] during the retention of a stop, the air pressure at the oral cavity and the pharynx increases and the force of closure at the place of articulation is greater for obstruents than for nasals; therefore, exactly speaking, the manner of how the closure is made is not completely the same [for nasals and stops]." (Hattori 1951: 122 [translated from the Japanese by SYK])[12]

Hattori (1951: 124; tr. by SYK) then notes that while nasalized liquids and glides are easy to articulate, fricatives (as well as trills) are difficult or impossible to fully nasalize, since "to articulate usual fricatives, it is necessary for fairly strong breath to flow through the oral cavity." Thus, the sonority of nasals must be greater than STREAM in terms of the tree structure of AIRMOVEMENT (35). The Nasal Sonority Rule (37) conforms to this requirement. Rule (37) together with the structure given in (35) puts nasals like /m/ and /n/ at the same level of sonority as lateral liquids like /l/. This stipulation represents a minimally required condition and suffices for accounting for Japanese phonology.[13]

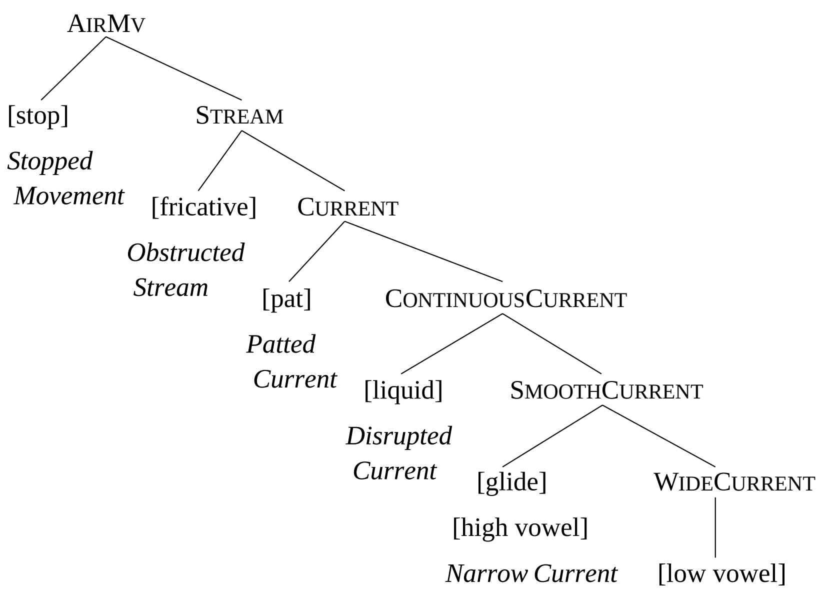
---

12 For experimental evidence for this statement, I quote from Fujimura (1961: 246): "There is a significant difference in the physical mechanism of the motion of the nasal bilabial, compared to that for the stops, because of the overpressure built up behind the closure in the case of stops."

13 (35) and (37) together would mark nasals as [liquid], perhaps a bad choice of the term, but substantive confusion should not arise from it.

However, it has been argued that nasals are to a lesser degree, or less marked, as sonorant, than liquids and that feature geometry must be structured to incorporate this assumption. (McCarthy 1988, Rice & Avery 1991, Rice 1993, Iverson & Sohn 1994.) We can accommodate this position by refining the structure given in (35). For the purpose of separating nasals and liquids for sonority degrees, we insert a new node ContinuousCurrent between Current and SmoothCurrent. The feature [liquid] is now taken as the default of ContinuousCurrent. I label [pat] the default feature of Current in this refined structure to suggest a light closure characteristic of nasals:

(38)



8. Sonorant assimilation in English

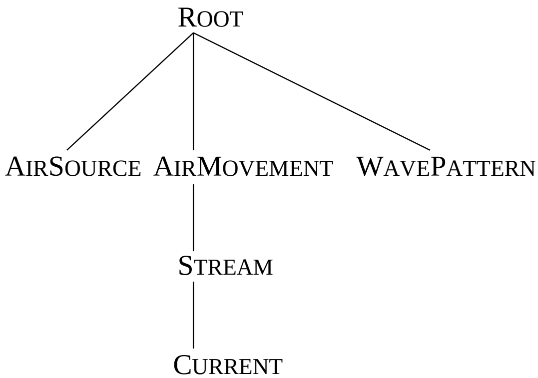
As an example, let us consider the phenomenon of sonorant assimilation in English. I reproduce relevant data from Rice and Avery (1991: 107):

(39)	a	i[m]balance	i[n]dentured	i[ŋ]grown	
		i[m]possible	i[n]tangible	i[ŋ]credible	
	b	i[r]rational	i[l]legible	i[n]numerable	i[m]measurable

The prefix-final segment fully assimilates to a following sonorant, as shown in (b), but not to a following obstruent as seen in (a); in the latter case, the prefix-final segment is realized as a nasal, the assimilation being restricted to the place of articulation. These facts suggest, on the one hand, that the prefix-final segment is a "generic" sonorant, i. e., an unspecified sonorant segment, rather than an unspecified consonantal segment, and, on the other, that the default sonorant in English is an unspecified nasal.

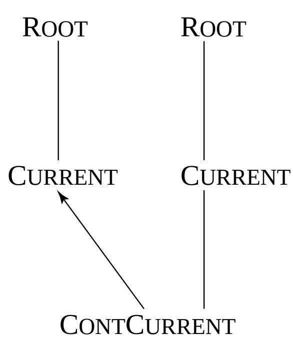
As Current formally characterizes the informal concept of sonorant, the "unspecified" sonorant segment has the following representation in our feature geometry:

(40)



This segment is the final segment of the prefix in question in its underlying representation. We assume that between this prefix-final segment and the initial segment of the stem that follows the prefix, leftward sonorant assimilation applies. Formally, it takes the form of Spread ContinuousCurrent:

(41) Spread ContinuousCurrent



If the stem-initial segment is a stop, a fricative or a nasal, the structural description of Spread ContCurrent is not met: we are left with the sequence unchanged at the underlying level. To illustrate, take, for instance, *insane*. We have the following underlying representation, with irrelevant details omitted:

```
(42)  /…AirS AirMv …/ + /…AirS AirMv…/
                :                  |
             Current             Stream
```

We get the following representation by the default convention:

```
(43)  /…AirS AirMv…/ + /…AirS AirMv …/.
                :          |        |
             Current  [voiceless] Stream
                |                   |
              [pat]            [fricative]
```

The feature [nasal] has yet to be assigned to the prefix-final "default" sonorant segment. But this assignment can be supplied by the Structure Preserving Convention, since, as I assume, English lacks non-nasal "patted" sonorants.[14] Thus, at the surface level, we have the following sequence, a nasal geminate, which is then simplified in a complete assimilation:

```
(44)  /…AirS        AirMv…/   +  /…AirS          AirMv …/.
         |             :             |              |
       VCdVbr       Current     [voiceless]       Stream
         |             |                            |
       NslOp         [pat]                      [fricative]
         |
       [nasal]
```

Next, consider the case where the stem-initial segment is a liquid, as in *irrational*. We have the following underlying representation:

```
(45)  /…AirS …AirMv… / + /… AirS AirMv …/
                 |                 :
              Current          Current…
                                   |
                              ContCurrent
```

14 Arguably alveolar flap /D/ in English as in *writer* and *rider* (Chomsky 1964: (35)) might be taken as a non-nasal [pat]. But this is a phonetic matter, and I assume no non-nasal pats exist in the phoneme inventory of English.

The structural description of Spread ContCurrent is met, and it yields the following representation:

(46) /…AirS AirMv …/ + /… AirS AirMv …/.

Current Current…/

ContCurrent

The Sonorant Default Voicing Rule (36) and the default conventions derive the following representation of a liquid geminate.:

(47) /...AirS AirMv .../ + /... AirMv AirS.../.

VCdVbr Current Current VCdVbr..../

[voiced] [voiced]

ContCurrent

[liquid]

We have thus accounted for the alternation of the prefix *in-* discussed by Rice and Avery.

The attentive reader will, however, have noticed that there is a potential serious flaw in this account in terms of our feature geometry. As it stands now, Spread ContCurrent does not distinguish between liquids and vowels. Take, for example, *inactive*. The relevant part of the underlying representation for this form is:

(48) /…AirS …AirMv… / + /… AirS AirMv …/.

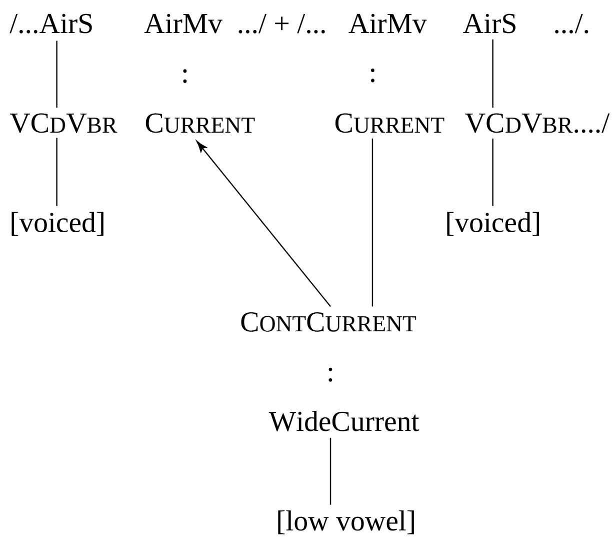
Current Current…

ContCurrent

WideCurrent

The structural description of Spread ContCurrent is met; together with default conventions, it yields the following representation:

(49)



The result would be *iaactive* /ɪææktɪv/ a form with a "geminate" vowel. This undesired consequence leads us to the second general issue I would like to discuss.[15]

---

15 Sonorant/nasal assimilation in Korean also provides support for the refined structure (38). Indeed, our feature geometry in (38) not only can accommodate the account of the sonorant/nasal assimilation given by Iverson and Sohn (1994) but also can account for the spirantization phenomenon as well by one and the same rule: sonorant, nasal and fricative assimilation in Korean can be understood as manifestations of Spread Stream; see Kuroda (2003).

I am fairly confident that with the refined structure (38) our geometry can deal with the kinds of issues involving voicing and sonorant that Avery and Rice (1989) and Rice (1992) tackled with their node SV (Sonorant Voice or Spontaneous Voice). But there is also a problem with (38). The system would impose on us an arbitrary decision: except for the unlikely case where there is an underlying contrast between nasal and nonnasal flaps, we can characterize nasal sounds underlyingly either as [nasal] or as [pat], the other being introduced by a redundancy rule. It would thus seem preferable if we could somehow take (35) as an unmarked situation and devise a separate means for making nasals less marked sonorant than liquids. I leave this issue for future study.

9. The problem of consonants vs. vowels

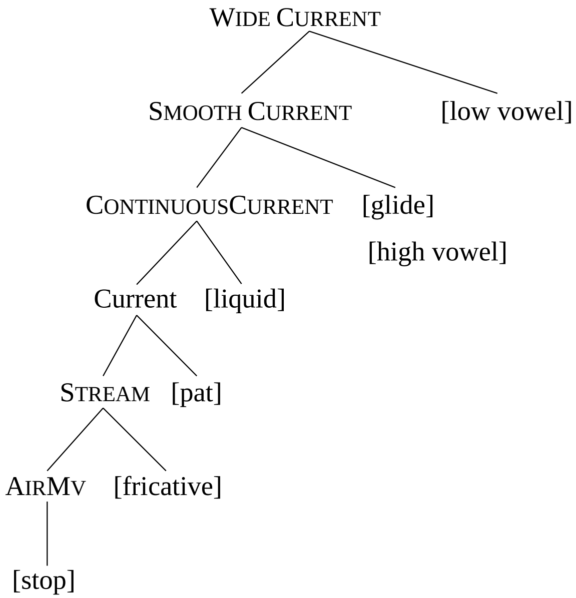
It is common in feature geometry to introduce [consonantal]/[vocalic] as well as [sonorant] as features of the Root node (Clements & Hume 1995: 292; Halle 1995: 2). Such features, however, are deemed arbitrary in the aerodynamic conception of feature geometry. As shown above, the feature [sonorant] is dispensed with in our feature geometry as incongruent with the basic idea of our geometry. It is resolved in the sonority hierarchy, which is structurally mirrored in the node organization under AIRMOVEMENT. Features [consonantal] and [vocalic] are also matters of sonority and must be dispensed with in our feature geometry.

Sounds with a lesser degree of sonority are deemed consonantal, and those with a greater degree are deemed vocalic. Thus, consonants, so to speak, branch off at a higher position in the tree structure, and vowels at a lower position in (35). Stops may well be considered as default consonants, and fricatives as more consonantal than liquids. This fact is reflected by the tree structure under AIRMV, where AIRMV, whose default value is [stop], dominates STREAM, whose default value in turn is [fricative], and STREAM dominates CURRENT, whose default value is [liquid]. This structure also implies that a general rule that affects stops as consonants can be formulated in terms of the node AirMv and must also affect fricatives and liquids, as desired. Thus, it looks as though node AIRMV dispenses with the feature [consonantal], taking over its function.

However, the problem with this line of thought, of course, is that vowels are located down at the bottom of the AIRMV tree and would count, so to speak, as the least consonantal segments. Vowels would then be affected by a rule affecting consonants in general. Likewise, they would also be affected by a rule that affects sonorants, as we have seen at the end of the last section.

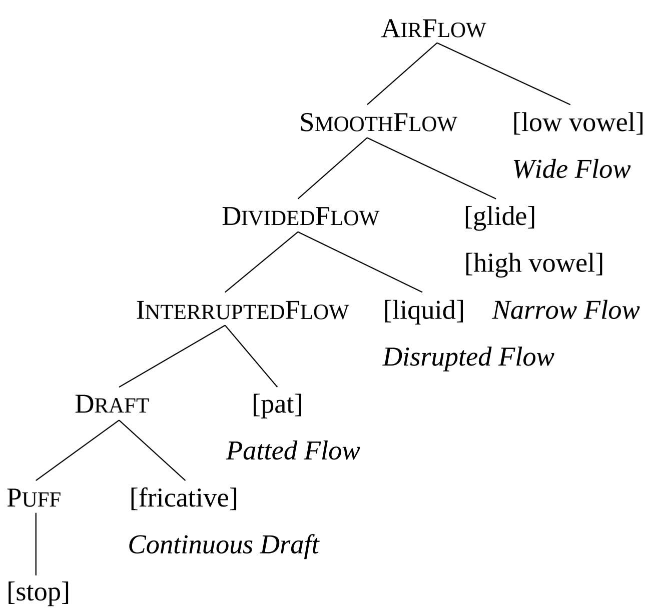
The difficulty we face arises from the fact that we have made an arbitrary choice for a value of a free parameter. Sonority is a scalar measure. When we combine this measure with an entailment relation, there is no intrinsic reason to choose which way the directionality of entailment should take. Let x and y be sonority degree and let $x < y$. If we gloss the sonority scale in terms of "at least as sonorous as" and define E_x as "being at least as sonorous as x," then E_y entails $E_{x.}$ In contrast, if we gloss the sonority scale in terms of "at most as sonorous as" and define E_x as "being at most as sonorous as x," then E_x entails $E_{y.}$ The former perspective gives the geometric structure given in (38). We can envision the geometric structure for the latter perspective if we imagine the tree in (38) as if it were a mobile and if we imagine holding it at the other end. Then we get the following tree:

(50)



However, the connotation of a node changed from "at least as sonorous as" to "at most as sonorous as". It would be better to revise the labels so that they might conform to the reversed entailment relation as suggested below:

(51)



To summarize, we have the geometry of the sonority structure projected in two different perspectives: the consonantal perspective, (38), and the vocalic perspective, (51). What I then propose is that the opposition consonantal vs. vocalic is not one that is determined by properties of segments formalized in terms of features; rather, it is one that inheres in positions (slots) that segments occupy. At a consonantal position, i. e., at a syllable periphery, the subgeometry under AirMv/AirFlow is projected in the consonantal perspective, while at a vocalic position, i. e., at a syllable nucleus, it is projected in the vocalic perspective. The entailment relation determined in one projection does not apply in another projection.

To illustrate how the mechanism of the projected geometry works, let us return to our earlier example *inactive* for the English sonorant assimilation. Since the stem-initial segment is projected in the vocalic perspective, we have the following underlying representation:

<table>
<tr><td>(52)</td><td>/…AirS</td><td>AirMv …/ +</td><td>/… AirS</td><td>AirFlow …/.</td></tr>
<tr><td></td><td></td><td>:</td><td></td><td>|</td></tr>
<tr><td></td><td></td><td>Current</td><td></td><td>[low vowel]</td></tr>
</table>

Spread ContCurrent does not affect this form as its structural description is not met. The relevant default conventions derive the following surface representation, /… -n-a… /, as desired:

<table>
<tr><td>(53)</td><td>/…AirS</td><td>AirMv …/ +</td><td>/… AirS</td><td>AirFlow …/.</td></tr>
<tr><td></td><td>|</td><td>|</td><td>|</td><td>|</td></tr>
<tr><td></td><td>VCdVbr</td><td>Current</td><td>VCdVbr</td><td>[low vowel]</td></tr>
<tr><td></td><td>|</td><td>|</td><td>|</td><td></td></tr>
<tr><td></td><td>NslOp</td><td>[pat]</td><td>[voiced]</td><td></td></tr>
<tr><td></td><td>|</td><td></td><td></td><td></td></tr>
<tr><td></td><td>[nasal]</td><td></td><td></td><td></td></tr>
</table>

10. Rendaku

Let us now return to Japanese phonology and let me add a few more remarks on rendaku. The rendaku phenomenon is commonly described in terms of the voicing of an initial obstruent of the second component of a compound word; see (3). The voicing is subject to the constraint of Lyman's Law: a non-initial voiced obstruent (but not a nasal or a sonorant), if any, in the second component of a compound word blocks rendaku voicing.

The significance of the rendaku phenomenon, I believe, is quite different for Old Japanese and Modern Japanese. The relevant difference between the Old and the Modern Japanese mainstream dialect is that in Old Japanese no word begins with a voiced obstruent and no stem has more than one voiced obstruent.

(54) Old Japanese phonotactic constraints:
   No word-initial voiced obstruent
   No more than one voiced obstruent in a single stem

The rendaku phenomenon in Old Japanese, in my view, is nothing but a simple morpheme structure constraint, that is, nothing but the manifestation of the Obligatory Contour Principle on the tier [voiced], [voiced] in the sense of the feature geometry I presented above, that is, in ordinary terms, non-nasal voiced. The voiced-unvoiced alternation of stem-initial obstruents is the manifestation of the general constraint that delinks the branch VCDVBR dominating [voiced] at word-initial position. This account explains at the same time the rendaku alternation, the absence of multiple voiced obstruents in a single stem, and the absence of word-initial voiced obstruents; it also accounts for the existence of *rendaku-immune* stems, stems that never exhibit the rendaku alternation, even though the voicing would not violate Lyman's Law.

(55) Rendaku-immune stems:

saki	'tip, end'	sio	'tide' (but not 'salt')				
kemuri	'smoke'	kasu	'dregs'	kase	'shackles'	kita	'north'
tuya	'gloss'		tuti	'earth'			

Martin (1987:114) and Vance (1987:69 f)

Looked at this way, the rendaku phenomenon in Old Japanese does not involve a voicing process. Rather, it provides evidence for a devoicing process.

(56) Rendaku in Old Japanese (Kuroda 1963, 2001, 2002)
   OCP on tier [voiced] ([voiced] in the sense defined in (16))
   Delink VCDVBR in env. #___
            --|--
          [voiced]
   Phenomena accounted for by (56)
   The rendaku voiced-unvoiced alternation
   The non-existence of words with an initial voiced obstruent
   The non-existence of words with more than one voiced obstruents
   The existence of *rendaku-immune* words

The matter is quite different with rendaku in later Japanese. In Modern Japanese, words *can* begin with a voiced obstruent, even discounting many such words of Sino-Japanese origin:

(57) Modern Japanese: words with an initial voiced obstruent
dasu ‘bring out’, *dare* ‘who’, *gama* ‘toad’, *gomi* ‘trash’, *barasu* ‘expose’ (not to mention many Sino-Japanese words)

Also, a stem in Modern Japanese can have more than one voiced obstruent:

(58) Modern Japanese: stems with more than one voiced obstruent
goza ‘mat’, *dobu* ‘ditch’.

In Kuroda (2002) I accounted for rendaku in Modern Japanese as an extension of the account given to Old Japanese, in terms of devoicing rather than voicing. The merit of that account could be questioned. Unlike the case of Old Japanese, we cannot relate the rendaku voiced/unvoiced alternation to other general processes or constraints in Modern Japanese phonology. In addition, the rendaku voiced/unvoiced alternation is irregular and arbitrary, as many scholars have noted:

(59) “Lacking any systematic guide, one must learn for each [compound] word whether a non-initial Y[amato] morph group exhibits the alternation or not…” Martin (1952:49)
“I am unable to state the environment in which the ‘voicing rule’ applies. The relevant data are completely bewildering.” McCawley (1968:87)
“There is little doubt that the occurrence of *rendaku* in modern standard Japanese cannot be predicted by any simple principle or set of principles.” Vance (1987:57)

It might be sensible to account for rendaku in Modern Japanese, following traditional lines, in terms of a lexically determined voicing process, with a certain proportion of subregularities. A rule to account for it could be formulated along the lines of Itô & Mester (1986:59), where [voiced] is to be understood as the default feature of VCdVbr, i. e., “non-nasal voiced”:

(60) Spread [voiced]

x x

[voice]

Be that as it may, it is quite questionable to take the voicing observed in the verb paradigm as an aspect of the same process as is responsible for the rendaku voiced/unvoiced alternation at the expense of introducing *NT with an otherwise unmotivated constraint domain; see note 7.

11. Summary of voicing assimilation in Japanese

To summarize, we have the following rules to account for the phenomenon of voicing assimilation, with the given order of application:

(61) PROGRESSIVE VOICING ASSIMILATION

```
        x          y
        |          |
      AIRS       AIRS
        |          |
    VCDVBR -> VCDVBR
```

(62) Sonorant Default Rule
CURRENT => [VCDVBR]

(63) Regressive Voicing Assimilation

```
        x          y
        |          |
      AIRS       AIRS
        |          |
    VCDVBR <- VCDVBR
```

(64) Coda Default
VCDVBR$)_s$ -> [nasal]

Rendaku is a separate lexical mechanism which, regardless of whether it is formulated in terms of voicing or devoicing, affects the tier of feature [voiced].

To return to the opening theme of this paper, the contrast between voiced obstruents and sonorants in Japanese phonology, we have the following three-way taxonomy shown in (2). The first type, manifested in rendaku, is an opposition characterized by the feature [voiced] in the sense of our feature geometry in underlying representation; the second type, manifested by the verb paradigm, is characterized by the node VCDVBR in underlying representation; and the third by the node VCDVBR in phonetic representation:

(65) [Cf: (2)]

Rendaku/Lyman's Law:	[voiced] (underlying)
The verb paradigm:	VCdVbr (underlying)
Mimetic adverbs	VCdVbr (surface)

12. Conclusion

In this paper, I have defended the earlier view of Japanese generative phonology concerning the phenomena of voicing assimilation. For this purpose, it is necessary to have a means to group together nasals with voiced obstruents to the exclusion of liquids and glides. I have justified this grouping on the basis of the idea of a feature geometry homomorphic to the aerodynamic design of the articulatory organs. But the part of the feature geometry to be referred to for this purpose constitutes a small branch of the geometry. Hence, this paper might be equivalent, from one perspective, to using a sledgehammer to crack a nut, and, from another, to building a castle in the air.

From the perspective of Japanese phonology, what concerns us first and foremost is the matter of descriptive adequacy of the competing descriptions, which in particular, hinges on whether Itô & Mester's *NT is viable or not. *NT, in my view, is untenable. We are thus turned back to the classical view of voicing assimilation. But from the perspective of linguistic theory, the issue of descriptive adequacy is not left alone; the claim for a descriptively adequate account is in the end judged by the adequacy of the theory that frames the description or its contribution to the development of explanatory adequacy. In this paper, I wish to claim that a crucial step for an adequate account of Japanese phonology justifies, and is justified by, an aerodynamically motivated feature geometry. Plainly, the issue in Japanese phonology addressed here by itself hardly justifies this geometry. From a theoretical perspective, this paper is mostly conceptually driven, in the hope of suggesting the viability of aerodynamically motivated feature geometry, and is just a small step toward empirical substantiation of this conceptual possibility.

Acknowledgements. I would like to express my gratitude to Eric Bakovic and Sharon Rose, who read various versions of this paper, for valuable comments and to Susan D. Fischer for her help and suggestions for preparing the final draft.

References

Avery, Peter and Keren Rice
1989 Segment structure and coronal underspecification. *Phonology* 6.179–200.

Chomsky, Noam
1964 *Current Issues In Linguistic Theory.* The Hague: Mouton.

Clements, G. N. and Elizabeth V. Hume
1995 The internal organization of speech sounds. In John A. Goldsmith (ed.), *The Handbook of Phonological Theory.* Cambridge: Blackwell. 245–306.

Fujimura, Osamu
1961 Bilabial stop and nasal consonants: A motion picture study and its acoustic implications. *Journal of Speech and Hearing Research*, Vol. 4, 233–247.

Halle, Morris
1995 Feature geometry and feature spreading. *Linguistic Inquiry* 26.1–46.

Hattori, Shirô
1951 *Onseigaku.* Tokyo: Iwanami.

Itô, Junko and R. Armin Mester
1986 The phonology of voicing in Japanese: The theoretical consequences of morphological accessibility. *Linguistic Inquiry* 17.49–73.

Itô, Junko and R. Armin Mester
1995 Japanese phonology. In John A. Goldsmith (ed.), *The Handbook of Phonological Theory.* Cambridge: Blackwell. 817–838.

Itô, Junko and R. Armin Mester
1996 Stem and word in Sino-Japanese. In Takashi Otake and Anne Cutler (eds.), *Phonological Structure and Language Processing: Cross-Linguistic Studies.* Berlin and New York: Mouton de Gruyter. 13–44.

Iverson, Gregory and H.-S. Sohn
1994 Liquid representation in Korean. In Young-Key Kim-Renaud (ed.), *Theoretical Issues in Korean Linguistics.* Stanford: CSLI. 79–100.

Kojien: Sony Electronic Book.
1996 Tokyo: Iwanami Shoten Publishers.

Kurisu, Kazutaka
2000 Richness of the base and root fusion in Sino-Japanese. *Journal of East Asian Linguistics* 9.147–185.

Kuroda, S.-Y.
1960 *Gengo no Kijutsu.* Tokyo: Kenkyusha.

Kuroda, S.-Y.
1963 A historical remark on 'rendaku,' a phenomenon in Japanese morphology. Ms., MIT.

Kuroda, S.-Y.
1965 *Generative Grammatical Studies in the Japanese Language.* Dissertation, MIT. Published by Garland Press, New York. 1979.

Kuroda, S.-Y.
2001 Rendaku and vocabulary stratification. In Bohumil Palek and Osamu Fujimura (eds.), *Proceedings of LP2000, Prague, 2000. Item Order: Its Variety and Linguistic and Phonetic Consequences*. Prague: Charles University Press. 11–34.

Kuroda, S.-Y.
2002 Rendaku. In Noriko Akatsuka and Susan Strauss (eds.), *Japanese/Korean Linguistics 10*. Stanford: CSLI. 337–350.

Kuroda, S.-Y.
2003 Projective feature geometry: A case study in Korean assimilation. *San Diego Linguistics Papers, Issue 1*. Linguistics Department, University of California, San Diego. 83–108. Also printed in *Journal of Cognitive Science*, Institute of Cognitive Science, Seoul National University. 5.1:73–137. 2004.

Kuroda, S.-Y.
2006 Feature geometry and phonetic features: A case study in voicing and coda nasalization in Japanese. *Gengo Kenkyu* No. 129: 91–134.

Martin, Samuel E.
1952 Morphophonemics of Standard Colloquial Japanese. *Language, Supplement. Language Dissertation* No. 47.

Martin, Samuel E.
1975 *A Reference Grammar of Japanese*. New Haven: Yale University Press.

Martin, Samuel E.
1987 *Japanese Language through Time*. New Haven: Yale University Press.

McCarthy, John
1988 Feature geometry and dependency. *Phonetica* 43.84–108.

McCawley, J. D.
1965 The accentual system of Standard Japanese. Dissertation. MIT.

McCawley, J. D.
1968 *The Phonological Component of a Grammar of Japanese*. The Hague: Mouton.

Nasukawa, Kuniya
1998 An integrated approach to nasality and voicing. In Eugeniusz Cyran (ed.), *Structure and Interpretation: Studies in Phonology*. Lublin: Folium. 205–225.

Rice, Keren D.
1992 On deriving sonority: A structural account of sonority relationships. *Phonology* 9.61–99.

Rice, Keren D.
1993 A reexamination of the feature [sonorant]: The status of ‘sonorant obstruents’. *Language* 69.308–344.

Rice, Keren D.
1997 Japanese NC clusters and the redundancy of postnasal voicing. *Linguistic Inquiry* 28.541–551.

Rice, Keren and Peter Avery
1991 On the relationship between laterality and coronality. In Carole Paradis and Jean-François Prunet (eds.), *Phonetics and Phonology 2: The Special Status of Coronals*. San Diegeo, CA: Academic Press. 101–124.

Vance, Timothy J.
1987 *An Introduction to Japanese Phonology*. Albany: State University of New York Press.

Vance, Timothy J.
2002 Another look at vocabulary stratification in Japanese. In Fabrice Cavoto (ed.), *A Collection of Papers in Honour of Alexis Manaster Ramer*. 2 vols. Munich: Lincom Europa.

Quasi-phonemic contrast and the fuzzy inventory: Examples from Scottish English

James M. Scobbie and Jane Stuart-Smith

1. Introduction

In this article we propose that contrast must be treated as a gradient phenomenon at the phonological level, with membership of a phonemic inventory being a matter of degree. This is because, though minimal pairs provide simple and strong evidence of contrast, things are not always so straightforward. Defining "minimal" is one challenge; as is determining which aspects of a contrast are distinctive and which redundant. Non-phonological information is sometimes a necessary consideration. These complications are usually thought to affect the analysis of a phenomenon in a discrete way, tipping the binary balance held by the phonologist towards either one analysis or another. We, on the other hand, see the necessity of evaluating contrastive evidence and of taking other linguistic information into account as being an indication that contrastiveness is a scalar property. We address some patterns in the sound system of Scottish English; ones which provide less than clear evidence of phonemicity – or, as we think, evidence of less than clear phonemicity.

First we review two consonants which are usually regarded as being part of the Scottish inventory, but which are systematically and lexically peripheral and which have been shown in our recent work to be seriously compromised as members of the Scottish Standard English (SSE) consonant inventory. From the vowel system we then present some new data relating to the unpredictability of the distribution of "long" and "short" variants of /ai/. Generally the distribution of these variants (and long/short variants of /i/ and /ʉ/) is predictable from phonological structure, hence allophonic. But part of the pattern involves what we term a "quasi-phonemic" (QP) contrast between such words as *crude* [kʰɹʉd] and *crewed* [kʰɹʉːd] or *side* [sʌɪd] and *sighed* [sɑed].

A number of different near-contrasts from various dialects of English of this general QP type are discussed by Harris (1990, 1994). Under the label of "marginal" contrasts, Harris (1994: 28–31) presents them as key analytic problems. Earlier, Harris had called them "derived", and though this reflects their morphologically complex nature, it uses a derivational metaphor which is better avoided. We have coined the narrower term quasi-phonemic for this class because being marginal in the inventory is a heterogenous characteristic. For example, low type and token frequency, lexically-restricted incidence and phonotactic restrictions make the status of some phonemes marginal, such as Scottish /x/ as we will see, but the *crude* vs. *crewed* contrast is marginal in quite a different way, namely in its systematicity. We will claim that both types of marginality should be reflected directly in phonological theories.

Harris reviews a number of quasi-phonemic (QP) contrasts, of which the Scottish Vowel Length Rule is just one. His suffixing examples fall into two types. One type (including the SVLR) share the general characteristic than an open syllable allophone is conditioned even when it appears before a consonantal suffix C. The QP contrast arises in the context of that consonant between the open allophone (found before suffix/clitic C) and closed allophone (found before tautomorphemic C). One of his examples is *days* vs. *daze* in Northern Ireland English in which *daze* has [ɪə] (like other cases of /e/ in closed syllables) whereas *days* (like *day*) has [ɛː] despite the coda /z/. The other type is when the suffix is syllabic. A word-final coda C (either its mere presence or some aspect of it) conditions an allophone of the previous vowel, e. g. [ɒʷ] in *roll* in London English, which is preserved on suffixation, giving rise to *molar* [aʷ] vs. *roller* [ɒʷ] (cf. also *ladder* vs. *madder* in Belfast or New York English). Perhaps the QP contrast in this case arises through the failure of syllabification of the stem-final word-internal C as an onset (in this example, the /l/ of *roller*). A morpheme-internal C (the /l/ of *molar*) must be ambisyllabic or an onset. The foot structure of *molar* vs. *roller* does not seem to differ: it is the morphological difference which is crucial. A third type, non-suffixing, is where morphosyntax or lexical class directly conditions some variant (*can* vs. *can* in US English).

The SVLR distinction between *side* and *sighed* etc. is quasi-phonemic because while there is a categorical and meaning-bearing difference between the two forms, it is one which is entirely predictable, from morphological structure. Thus the phonetic vowel differences in these Scottish English pairs, if phonologised at all (as length or bimoraicity or headedness or whatever), are in one sense redundant and non-phonemic (Pike, 1947). Since the redundancy is based only on non-phonological structure, we have chosen a terminology which gives precedence to the similarity of this pair to other pairs in which a minimal difference in sound makes a difference in lexical meaning while recognizing that this is not contrast in the strict sense.

Pike's seminal work is an excellent starting point for considering such issues, and much of what he had to say is strongly relevant today, and the sorts of problems we address were well known to him nearly sixty years ago, and so it should perhaps be surprising, then, that such data still seem problematic. As we will show, the more detailed empirical data we gather, the more problematic things seem to get for traditional concepts of phonology, such as a crisp distinction between distinctiveness and redundancy, between contrastive and non-contrastive phenomena.

2. What is phonological and what is not?

Lexical contrast is the defining phenomenon of phonology. As a general concept, contrast is a situation in which phonetic differences (from the obvious to the subtle) reflect and represent categorical differences in meaning. In the canonical case, namely lexical contrast, differences in sound change one word, such as *wood*, into another, such as *burning*. The categoricalness of lexical contrast arises out of semantics, but only sometimes is encoded by utterly clear articulatory or perceptual phonetic categorisation: for example, it is the meanings of *bin* and *bean* which are absolutely disjoint and uninterpolable, not the extensional set of each word's actual phonetic realisations. The categoricalness of lexical contrast demands that in any particular system, such as Scottish English, two words either contrast (such as *love* and *loves*) or do not (like *pull* and *pool*): there is no indeterminacy or intermediacy. Contrasts are relatively easy to establish, and if they form the basis for phonology it follows that it is reasonable to have, as a theoretical goal, a clear-cut, modular, algebraic phonology of words and phrases.

It is obvious, however, that to develop a *theory* of phonology (in order that we can make phonological predictions about typology, acquisition, diachrony and so on), we need to follow in the footsteps of Kenneth Pike and other structuralists and consider much more than unorganised yet categorical meaningful differences in sound. First, we must develop analyses of the systems into which contrasts are organised, a process which demands that we identify the most basic contrastive units, the structures that govern their distribution, and the principles that control their behaviour. A second essential ingredient is to address systematic phenomena which complement lexical contrast, such as morphophonemic alternations, allophonic variation, stress, intonation, and other phrasal phenomena.

These theoretical necessities are intertwined: divining the minimal units of contrast means tracking their distribution in structure *even when they are not actively contrastive*. Bear in mind that since very fine-grained, variable and continuous aspects of phonetics may be language-specific, there is a vast amount of "non-contrastive" information which must be represented mentally by speakers. It is therefore necessary in most theoretical viewpoints to work out which of the myriad of predictable differences in sound actually constitute phonological data, and which are language-specific but phonetic (even if they are conditioned by phonological structure). We see no reason not to use the word "grammar" to encompass the entire cognitive system which we, as language users, have to learn. The crucial debate in phonology is whether such fine phonetic detail is expressed in the same system that is necessary for encoding contrast (usually a symbol-processing formalism) (e. g. Boersma, 1998; Flemming, 2001); whether phonology and phonetics are disjoint (Chomsky and Halle, 1968; Hale and Reiss, 2000); or whether, in mental representations, knowledge of contrast is a fuzzy superimposition on, or abstraction from, knowledge of precise (yet predictably and continuously variable) phonetic targets (Pierrehumbert, 2001; 2002; Hawkins and Smith, 2001; Coleman, 2002; Gafos, 2006; and other aspects of Boersma, 1998).

We might hypothesise, in line with most phonological theory, that (at some level of detail) phonetic and phonological knowledge are distinct. But, as Scobbie's review (2007) of these different approaches points out, adopting a *modular* architecture entails that all sound-systematic data can and must be segregated appropriately. Determining that some set of forms constitutes phonological data relevant to a particular phonological principle – or not – is theoretically crucial. Yet there is no scientific, let alone generally-agreed, basis for making such a decision. This ambiguity about the phonological status of many non-contrastive phenomena is one of the most intractable predicaments hindering advances in phonological theory.

This lack of clarity as to the remit of phonology is due to phonetics and phonology being non-arbitrarily related and to the language-specificity of much phonetic patterning. It might have been hoped that instrumental phonetic analysis (such as laboratory phonology, reviewed by Pierrehumbert, Beckman, and Ladd, 2000) could provide the grounds for an "industry standard" definition of what is, and what is not, phonological data, let alone what is phonemic within phonology. But in practice it is often hard to define exactly which linguistic phenomena are truly phonological deterministically. There are even indications that in many occasions it may be impossible (or misleading) to make a definitive decision about the phonological vs. phonetic status of some phenomena on phonetic, or any other empirical grounds. The uncertainty over a simple binary choice will, we think, increase as more complex phenomena are subjected to empirical analysis, especially when attention is paid to issues of phonological variation and change. The benefits of empiricism may be that we may gain a more realistic impression of the complexities of phonology rather than solving long-standing problems with contemporary theories.

3. Establishing inventories of segments, features, clusters and more

One of the major components of a phonological system is an *inventory* of lexically contrastive units. Such inventories are usually featural or segmental, but in principle can be compiled for any type of linguistic unit. Contrastive inventories are crucial for much cross-linguistic comparison (as in Ladefoged and Maddieson, 1995 for example) but their theoretical status is unclear.

Contrastiveness alone cannot derive an inventory: the fact that *banana* and *bounce* contrast does not take us far. Two mutually dependent initial steps in the establishment of such an inventory are required. These are the identification of: places in structure, such as the syllable onset or first element in a consonant cluster (syntagmatics); and the inventories that pertain at each position (paradigmatics).

If we limit ourselves initially to a lexically contrastive inventory, then the relevant process of identifying the units is the minimal pair test. In such a test (also called a commutation test), pairs (actually n-tuples) of lexically contrastive words must be found which differ from each other in as few potentially phonological characteristics as possible. By definition, these paradigmatic choices will be made in just one syntagmatic position. For example (and putting aside the phonetic naivety which such a phonological statement implies), *bit* and *pit* differ in only the identity of their first segment. If no "smaller" distinction between them can be found, then this establishes two phonemes (let us call them /b/ and /p/) as members of the inventory and a single distinctive feature to encode the minimal difference (let us call it /voice/). In most minimal n-tuples like *pit*, *bit*, *tit*, *kit*, *git*, there will be a gap, in this case /dɪt/, which can be filled with a near-minimal form like *did*, if it is felt that the change in context is irrelevant to the initial consonant. Comparison of a number of such sets offers support to the inventory.

However, there are often ambiguities over the dimension in which a contrast is minimal, making even minimal pairs hard to analyse, let alone near-minimal pairs and partial n-tuples. Indeed it is often unclear whether a contrast *is* minimal. *Beat* and *bead* are usually taken to be a minimal pair, despite the fact that they differ in more than one potentially phonological dimension (this time we are not being quite so phonetically naïve). But in most analyses of English, they are said to differ phonologically (in underlying representations at least) in their final consonant alone. In those varieties of English in which there is a clear systematic vowel duration difference between them, this vowel difference is not relevant to the inventory. If it is phonological at all, it is redundant and appears only in symbolic surface structure, constrained by the grammar. (Phonetically, of course, vowel duration is actually an extremely important correlate of the *beat*/*bead* contrast in many varieties of English; though not Scottish English, as we will see.)

The alternative approach to encoding vowel duration phonologically as vowel length is to call it phonetic. If, like most phonetic allophony, the patterns of vowel duration are subtle, gradient and variable, then they may not be part of surface structure or constrained by symbolic phonological grammars at all. Distinguishing phonological from phonetic allophony is an extremely thorny issue, but is absolutely crucial in surface-oriented phonological theories. A theory of phonology comprising only constraints on surface structure requires a definition of what surface structure is, and what phenomena it represents. Indeed, any theory of phonology needs to define what its "surface" level of representation is, which non-contrastive phonological categories it contains, and state what it is *for* (Scobbie, 2005b; Ladd, 2006).

A final point is that commutative comparisons such as the minimal pair test are limited to paradigmatic substitutions at one place in structure, so cannot be used to establish the inventory across different syntagmatic positions. The concept of a cross-positional phonemic inventory requires further appeals to phonetic similarity and well-formed inventories.

In the face of such indeterminacy, phonological research cannot simply maintain the status quo. More detailed research into these fundamental concepts is clearly required. *Can* the discovery procedures of Pike be amended for today and completed? Or is the indeterminacy of descriptive phonology not a failing, but an indication of a deeper theoretical indeterminacy which should be embraced by theoreticians? We now approach these questions by considering some of the problems relating to the segmental inventory and contrastive content of Scottish English.

4. Scottish Standard English

Native Scots whose grammar and lexis can be classed as Standard (International) English speak with a variety of different accents – of course. For the most part the variation in any geographical location within Scotland is, following Aitken (1984) and Abercrombie (1979), seen as a continuum from local "broad" sound systems with deep roots at one end, to, at the other, varieties influenced in large measure (but usually indirectly and at some considerable historical or social remove) by the standard variety spoken in England. The latter non-vernacular end of the continuum shows, naturally, far less geographical variation within Scotland.[1] Somewhere between a local vernacular variant of Scots and what would be seen as a foreign Anglo-English is Scottish (i. e. Scottish-accented) Standard English, "SSE" (Abercrombie, 1979; Scobbie, Hewlett, and Turk, 1999). It is impossible and undesirable to draw a clean line between such varieties, but our goal here is to probe the problems which arise when considering the structure of any phonological system, in this case SSE, due to system-internal ambiguities over the contrastive phonological status of particular phenomena.

So, Standard English (e. g. as written here) when spoken in Scotland is different from American or Southern Standard British English essentially in its sound system, by definition, with a few minor systematic differences elsewhere, such as the existence of the preposition *outwith* and the grammaticality of *needing washed*. To go from SSE towards Scots, on the other hand, means greatly altering lexis, lexical incidence, morphology, morphosyntax, idiom and to some extent syntax, and (again) the sound system.[2]

When distinguishing the various local versions of Scots from SSE in terms of "accent", i. e. sound system, we think it is not sufficiently clear that few of the aspects characterising the SSE sound system from Scots are phonological on a very narrow interpretation. It is appreciated that SSE and Scots are still remarkably similar, and are clearly closer than SSE and RP. What is not stressed is that the potentially very distinct sound systems of SSE and Scots differ primarily in lexical incidence, the membership of lexical sets, morphophonemics, and even in what phonologists usually call "low level" phonetics, as any sociophonetic study can show. Differences in phonemic inventory and phonotactics are more trivial. Even varieties of broad Scots whose phonologies are most different to SSE, such as Shetlandic (van Leyden, 2002), have segmental inventories which bear closer typological similarities to SSE than SSE does to many other well-known varieties of English. This is not to say that the differences in phonetics and lexical incidence are trivial. As well as being able to cause severe problems for interspeaker intelligibility, they are important characteristics of sound systems with complex geographical, structural and sociolinguistic distributions. For an overview of acquisition, see Scobbie, Gordeeva, and Matthews (2007).

---

1 The effects of population movement and dialect contact are fundamental but additional complications which we cannot address here, as we will attempt to focus as narrowly as possible on phonological issues. For some of the necessary breadth, see Stuart-Smith (2003).

2 For example, see Matthew Fitt's translation into Glasgow Scots: "Zeus, high-heid-yin ae the gods an heid-bummer ae the universe, had a son an he cawed this son Heracles. Heracles was strang as a buhl. He wis built like a hoose-end an had erms like a boxer an legs like cabers. Heracles wis feart at naebody, except his step-maw Hera." (Fitt, Rennie, and Robertson, 2002).

5. The consonant inventory of Scottish Standard English

In this section we concentrate on peripheral items in the consonant inventory of Scottish English and the varying reasons for the dubious status of certain consonant phonemes. For more details and full methodology see Stuart-Smith's various publications based on empirical data gathered from a socially-stratified pool of 32 speakers from Glasgow (Stuart-Smith, Timmins, and Tweedie, 2007; Stuart-Smith, 2003) and references therein (though especially relevant is Macafee, 1983).

5.1. Overview

Generally speaking, the Scottish consonant inventory is familiar from other varieties of English: /p t k b d g tʃ dʒ f θ s ʃ v ð z ʒ m n ŋ h r l w j/. These 24 consonants comprise a relatively simple core, though there are some well-known analytic problems common to many varieties of English: the complementary distribution of /h/ and /ŋ/; the status of /ŋ/ as a segment rather than a sequence; the skewed phonotactics and low functional load of the /θ/-/ð/ contrast (and the ongoing loss of /θ/); the difficult status of post-vocalic /w/ and /j/; the roles of [ʔ] as an allophone of /t/ and as a delimitative marker; and others. The liquids /r/ and /l/ are also of great phonological interest, especially with respect to coda weakening and sandhi, but since there is little argument that SSE at least does have an /r/ and an /l/, we will forego further discussion of these crucial consonants for now.

5.2. The velar fricative *x*

This non-sibilant voiceless fricative phoneme is limited phonotactically to the coda, appears primarily as a singleton and not often in clusters, favours word-final to word-medial contexts and has a highly limited lexical frequency outwith proper names. Informal observation indicates that younger SSE speakers have difficulty thinking of even a handful of words containing /x/, such as *broch* or *loch*. (These words, whether with their /x/ intact or not, have been borrowed into standard English.) The phoneme is more commonly preserved in place names and surnames (and so *Naughty* may have /x/ when a surname even if not when a regular lexeme) and indeed is productively applied to non-English names and words, whether spelt with coda "ch" (*Munich*, *Bruch* and *Bach*), or not (*van Gogh*, *Ahmed* and *Khomeini* with a structurally rare onset /x/).

Despite a limited distribution, the use of a [x] sound in *loch* and the contrast with *lock* are still highly salient for many SSE speakers, and a failure to use [x] may be explicitly brought to the attention of foreigners, including native English speakers. The use of [k] in *loch*, in particular, can cause social offence far beyond any strictly linguistic basis. Even so, /x/ is losing ground among young urban vernacular speakers (Lawson and Stuart-Smith, 1999; Stuart-Smith et al., 2007) and even rural Scots speakers (Marshall, 2004). There are relatively few borrowings into SSE with /x/, and it is far more common in self-evidently Scots lexis (*bourach*, *dicht*, *teuch*, *dreich*, *pech*).[3] SSE speakers will use such Scots lexis only in some contexts (e. g. literary or social ones), and if they are used, it is important they are pronounced "correctly", i. e. with /x/.

In SSE, the high social salience of the phoneme /x/ and the minimal pair *lock*/*loch* seem to provide evidence for the inclusion of /x/ in the inventory, despite its extremely marginal structural status, low functional load, low type and token frequency and propensity for merger with /k/ among many speakers.

Structurally, coda-based [x] and stressed-onset based [h] could be synchronic allophones. They are largely speaking in complementary distribution, and are both non-strident voiceless fricatives. Phonetically, hyperarticulated onset /h/ is sometimes heard to have some [x]-quality, whereas coda [x] is acoustically weak with smooth velar frication. Indeed, heavily weakened /x/ approaches the quality of a devoiced vowel after high or back vowels. (Perhaps we should discount the self-confident handful of speakers who claim to have *Docherty* as ['dɔxɹ̩te] and *Doherty* as ['dɔʰɹ̩te]. It may say more about the similarity between /h/ and /x/ and the potential for mutual substitution than about a potential for contrast, or be another peripheral aspect of the phonology which is spelling-induced.) Finally, whether aspiration on initial /p t k/ is thought to be relevant to the status of /h/ or not, it is interesting that strongly aspirated /k/ may have an affricated release.

5.3. The voiceless labial-velar ʍ

This consonant is limited phonotactically to onset and appears in no clusters. It is of very limited type frequency, but because it appears in "wh" grammatical words, has a fairly high token frequency. There are a number of minimal pairs (*which* vs. *witch*, *whether* vs. *weather*, *whales* vs. *Wales*) which can be seen as strongly supporting the status of /ʍ/ as a member of the inventory. However,

---

3 Scots lexis can be glossed at the *Dictionary of the Scots Language* online: http://www.dsl.ac.uk/dsl/

for the majority of English speakers in the UK these pairs are homophonous, and SSE speakers vary in how aware they are of the contrast if they have it themselves. These factors may explain the persistence of the popular Scottish children's joke: "How do you get two whales in a Mini?" which relies on a [w] in *whales*.[4] Lawson and Stuart-Smith (1999) and Stuart-Smith et al. (2007) present quantitative evidence for the weakening and loss of the requisite phonetic distinction which underpins the contrast among younger speakers who are generally thought to continue Scots in their vernacular, where the contrast is always thought to have been strong (see also Johnston, 1997). Their use of [w] is indicative of a merger, which is echoed by the tendency of highly Anglicised speakers to merge /ʍ/ and /w/. Children may temporarily lexicalise the wrong phoneme developmentally. But on the whole, SSE still contrasts these pairs.

One of the main phonological problems with /ʍ/ is where it goes structurally in the inventory. It seems usually to be regarded as a fricative, yet, inconsistently, to be the voiceless counterpart of the approximant /w/. Alternatively, it may be seen as a cluster /hw/ – in which case /ʍ/ would not be part of the inventory at all. The existence of clear contrast does not solve the analytic problem of phonemicity.

The main argument against the cluster analysis would be that it creates the only cluster in which /h/ would be involved synchronically. And although /w/ appears in several, only /sw/ is well-supported lexically (*sweet*, *swan*, *switch*). Examples of /bw/, /dw/, /gw/, /fw/, /θw/ and /ʃw/ are rare and/or often involve marginal lexemes (*Buenos Aires*, *dwarf*, *Dwight*, *Gwen*, *guano*, *Fuentes*, *foyer*, *thwack*, *Schweppes*) and such argumentation is usually used to establish that a complex segment is not a cluster, but a singleton phoneme. However, /hw/ need not be the only /hC/ cluster in SSE, given other analytic possibilities. Specifically, it may be partnered by the cluster /hj/ e. g. in *huge*, so long as /ju/ is not regarded as a diphthong /iu/, another long-standing indeterminacy of the vowel inventory of English.

These clusters would be phonologically parallel: they are the pair /h/+glide. Additionally, they are phonetically parallel because in production they are very segment-like with little internal sequencing of voice. Generally /hw/ is [ʍ], while /hj/ is [ç]. Finally, note that some SSE speakers who avoid /j/ in clusters have a pattern in which both are reduced to their glide (*which* with [w] and *human* with [j]), whereas the reduction of the cluster /nj/ in *new* is to plain [n].

So even with clear contrasts in those speakers who have not lost it, the status of /ʍ/ is actually in the balance. With its low frequency and without any clear

4 The answer is: "Go down the M6 [a motorway] and turn right."

position in the structure of the consonant system, this "Scottish consonant" has a reasonable claim to be a marginal cluster rather than a marginal phoneme.

6. The Vowel Inventory

We will focus here on one particular phonological vowel system, one commonly discussed in phonological research on SSE. This system is widely found in the fifty-mile span that encompasses Glasgow (the largest city) and Edinburgh (the capital). Several million speakers, the bulk of the Scottish population, live in a number of conurbations in this Central Belt. The starting point for an SSE phonemic inventory are the twelve lexically stressed vowels of Abercrombie's "basic" Scottish vowel system (Abercrombie, 1979). It has five free monophthongs /i e ɔ o ʉ/ (*pea*, *pay*, *paw*, *po*, *pooh*), four checked monophthongs /ɪ ɛ a ʌ/ (*pit*, *pet*, *pat*, *putt*), and three free diphthongs /ai au ɔi/ (*buy*, *bow*, *boy*). SSE lacks a number of tense/lax or monomoraic/bimoraic pairs which are common to other dialects of English. *Pam* and *palm*, *cot* and *caught*, *pool* and *pull* are homophones.

Abercrombie notes that some speakers have additional vowels that can be, in principle, easily established through a minimal pair test. Under the influence of Anglo English, for example, speakers may distinguish *Pam* and *palm*, in which case we would add /ɑ/ to the inventory for *palm*, or, more rarely some other contrasts. The context for our discussion is the readily-established and uncontroversial basic system, but the extent to which these additional contrasts are likely to be evidenced by lexemes with different frequencies or contexts of use is relevant (where we expect patterns in line with Bybee's work, e. g. in Bybee and Hopper, 2002).

6.1. The Scottish Vowel Length Rule

The phenomenon in Scottish English which has received most interest from phonologists is the Scottish Vowel Length Rule (SVLR) (Aitken, 1981; Giegerich, 1992; Scobbie et al., 1999a, 1999b and many others). This is the name given to the complex but mostly *predictable* distribution (hence "rule") of "long" and "short" allophones of vowel phonemes as conditioned by various factors: phonological, phonetic and morphological. To simplify things:[5] in

5 We are going to over-simplify the following characterisation, so that we can move on to considering the facts in the next section which relate to contrast in more detail. The difficulties in characterizing these non-contrastive aspects of the SVLR are no less problematic, and are the focus of on-going research.

word-final stressed syllables, "long" allophones (i. e. those with greater phonetic duration) occur in open syllables and before voiced fricatives and /r/; "short" allophones occur before stops (including voiced ones, crucially for what follows), nasals, voiceless fricatives and /l/. Following McKenna (1988), Scobbie et al. (1999a) and Scobbie (2005a) show that among the monophthongs, /i ʉ/ stand out as having a particularly strong phonetic duration effect, while with /ai/, quality and quantity interact in a particularly revealing way.[6] Establishing exactly which vowels are subject to a phonological SVLR and which vowels are subject to a similar but phonetic pattern remains an absolutely fundamental problem – if, that is, it is thought to be important to separate phonology from phonetics in a sharp modular way.

Many of the phonological discussions of the SVLR focus on the challenge of formalising what "length" means for /ai/, linking that to /i ʉ/, and distinguishing short /i/ from lax /ɪ/ (cf. Escudero and Boersma, 2004 for an empirical study related to the last opposition which indicates it tends to be one cued by quality more than duration). Such issues are important whether the SVLR length distinction is underlying or derived.

6.2. Quasi-phonemic contrast involving *i ʉ ai*

As noted, word-final open syllables condition long variants of /i ʉ ai/. When suffixed by /d/ the vowel duration is not short as it is before tautomorphemic /d/ (or /t/) as might be expected under the SVLR.[7] Instead, a long vowel is found, giving rise to something rather like a minimal pair with any word with the same sequence of phonemes (as established up to this point) but in which the final /d/ is tautomorphemic (1–3). Near pairs, which are more common, are in parentheses.

(1) need ≠ kneed, (greed ≠ agreed)

---

6 In unpublished work we show that social factors conditioning /ai/ variation are also crucial to understanding the phonological and phonetic aspects of /ai/ variation.

7 This may be true of some other level 2 suffixes, such as *-ness*, *-ly*, which begin with a shortening consonant, or compounds, but the anecdotal claims in the SVLR literature about this are not supported by actual data and we doubt anything is as simple as it might appear. Bare /d/ as a clitic version of *had* or *would* probably condition long vowels in the words they attach to, but pronoun combinations (*he'd*, *you'd*, *I'd* etc.) typically are short in connected speech, being unstressed.

(2) crude ≠ crewed, brood ≠ brewed, rude ≠ rued, pud ≠ poo'd, mood ≠ moo'd, would ≠ wooed, (Jude ≠ subdued [sʌbdʒʉːd])

(3) side ≠ sighed, tide ≠ tied, (ride ≠ tried)

These differences bear the hall-marks of phonemic contrast, namely a categorical difference in meaning consistently attributable to the presence of a phonetic distinction, but structurally the vowel differences are predictable. The long vowel duration could be attributed to the morphological context directly, or indirectly if a different prosodic structure is proposed. Alternatively, different long/short phonemes could be allocated to different lexemes (albeit on a completely predictable structural basis). The actual analysis does not matter here: the first important point is to note that if the distinctions in (1–3) are not encoded segmentally, then each pair will be phonologically identical in prosody-free underlying representation. Second, if a predictable prosodic distinction were to be introduced then this does not theoretically determine whether the vowel distinctions are or are not encoded in Scottish English surface representations (i. e. as phonological allophones of some kind, such as moraicity or vowel length). Third, a phonological difference at either underlying or surface level in segmental content, including duration, means that there will be six phones corresponding to /i ʉ ai/. (Since prosodic structural differences are segment-independent, it is impossible without further segmental machinery to limit the SVLR to just a subset of all vowels able to appear in open syllables.)

Even if there are six phonological phones, this situation does not mean that all are part of an inventory, partly because derived or redundant structures are not generally accorded this status. However, inventories incorporating redundancy are crucial to understanding phonologisation, are utterly fundamental to surface representations and hence to constraint-based phonology (Scobbie, 2007), and are worthy of theoretical consideration in their own right (Ladd, 2006). We should probably be considering inventories of contrastive dimensions rather than mere segments, because, as Archangeli and Pulleyblank (1994) so clearly point out, segmental vowel inventories are misleadingly large if a basic five vowel system inventory (say) is multiplied 16 times by contrastive binary tone, length, nasality and ATR. In the SSE case, the relevant question therefore might be better asked: does the system have three degrees of length, or both tenseness and length, in *bid*, *bead*, and *freed*?

Support for including length with unarguably contrastive dimensions comes from the strength and categoricalness of the distinctions in (1–3). These differences seem indistinguishable from phonemic contrast from the perspectives of native speaker intuition and phonetic output, and are just as important in characterising the phonology of SSE. Note also that Matthews (2001) shows that the variants of /i/ and /ai/ (as allophones, before voiced and voiceless fricatives) are early-acquired. Unlike true phonemic contrast, however, the categorical meaning differences in (1–3) have a component of predictability in meaning tied to the morphology. Straightforward phonemic contrast does not simultaneously encode a morphological, syntactic or other non-phonological general meaning, nor be conditioned by structure, in addition to a single difference lexical meaning involving one morpheme versus another.

In previous publications we have reviewed the phonetic distributions underpinning a categorical SVLR difference, as well as presented durational and formant analyses of the speakers analysed here. These studies confirm that it is only /i ʉ ai/ that show quasi-phonemic contrast. In other words, the phonetic vowel duration in each of the pairs in (4–8) are no different, despite claims in the literature that they show the same contrast as the pairs in (1–3). We find these claims very interesting, and suspect that a thorough empirical analysis of the native-speaker intuitions on which those claims were based will be an important future addition to the literature. It may be that intuitions about differences are based on morphological / prosodic structure and generalised from /i ʉ ai/ in which they do appear phonetically onto those vowels where, in natural speech at least, there is no distinction.

(4) ode = owed, road = rode = rowed

(5) odd = awed, nod = gnawed

(6) grade = greyed, (afraid = frayed)

(7) aloud = allowed

(8) Boyd = buoyed, (avoid = annoyed)

6.3. Distribution and intuition: *i ʉ ai* in word-final stressed syllables

The few examples of QP contrast for /i ʉ ai/ presented in (1–3) above may have raised some doubt about the generality of the phenomenon. The limited numbers of such pairs may imply this QP contrast is a peripheral or weak phenomenon. However, even a handful of examples of the /x/-/k/ contrast were sufficient to establish the existence of /x/. Additionally, there are numerous near-minimal pairs like *freed* (long). vs. *reed* (short). But as has been mentioned already, other factors support the adoption of a segment in a language's inventory. In this case, because short /i ʉ ai/ are found before voiceless stops, the normal voicing effect on vowel duration is scanty (9). There are therefore also short-long pairs (10) in which the voicing difference (albeit confounded by the morphology) conditions a clear difference in verbs ending in /i ʉ/ and particularly /ai/. Furthermore, all words in the long vowel context are comparable whether the words happen to exist as members of minimal sets or not. The QP contrast is thus thoroughly supported through comparison between various incomplete sets. Finally, suffixed pseudo words, neologisms and nonce verbs (11) seem always to have long vowels, entirely consistent with the pattern.

(9) bleat ≅ bleed, seet ≅ seed, put ≅ pud, newt ≅ nude, bright ≅ bride

(10) skeet < skied, cute < cued, trite < tried, fright < fried

(11) he sky'd the ball, she tree'd the avenue

A rather different argument comes from a phenomenon of particular interest, in which the "wrong" vowel duration shows up. For example, there may be specific lexemes, like *dude*, or *vibes*, in which a long vowel is unexpectedly found for a sizeable minority of speakers. Scobbie (2005b) presents pilot empirical results to clarify the extent and range of such "unpredictable" vowel lengths. For example it seems that final /b ɡ dʒ/ may be more likely to condition a long variant than final /d/ especially in sparse prosodic neighbourhoods, (e. g. the rare coda /ib/), probably indicating that the functional pressures to maintain the quasi-phonemic contrast and to lengthen vowels before voiced stops are greater than the pressure to ensure paradigm uniformity for new or uncommon words. The literature (e. g. Aitken, 1981) is more reliable when reporting strong phonotactic generalisations such as long /ai/ before final /θ/ (*Forsythe*, *Rosyth*, *blythe*) than when reporting the vowel length of individual lexical items. Even so, caution should be exercised until new data is available, on word-internal contexts in particular, as will be clearer when we present the first such results below.

The fact that speakers can have clear intuitions and exceptional lexical specification of long or short variants of /i ʉ ai/ serves to underline the near-phonemic status of the length "contrast". The difference between long and short variants may be structurally allophonic much of the time, but when it is phonologically unpredictable, or when the distribution of long variants becomes highly detailed, the claim that both long and short variants are members of the inventory is strengthened.

In addition, there is some evidence (again largely anecdotal introspection) from level one morphophonology that short /ai/ exists in underlying representation and is not lengthened at level 1, strengthening the case that each variant should be represented in the SSE inventory. For some speakers it appears the irregular plurals of *life*, *wife*, *knife* may be *lives*, *wives*, *knives* with a short /ai/, despite the medial fricative being voiced in the derived environment. On the other hand, *lifes*, *wifes* and *knifes* are also fairly common plurals in otherwise standard speakers (with short /ai/ transparently before /f/), as are the irregular plurals with long vowels. More research is needed on these forms. For /i ʉ/ the evidence is even less clear, but we do not think anyone has ever claimed that irregular *hooves* or *leaves*, for example, may have a short vowel before a voiced fricative.

6.4. Unpredictable lexical incidence of variants of *ai*

It has previously been observed by Aitken that the choice of /ai/ variant in stressed non-final syllables (e. g. in trochees) is even more complex than presented above. For example, he claims that words like *spider* and *cider* have long /ai/ followed by /d/, whereas, if word-internal distribution is the same as word-final distribution, they should have short /ai/ before /d/. Again, such claims are based on introspection and observation rather than on any systematic fieldwork or experimentation and should be taken as a starting point only.

These trochaic patterns are particularly interesting for phonological analysis because, with /ai/ being word-internal, there is no opportunity for quasi-phonemic contrast. It would appear, however, that for many speakers, it is still possible to have a very clear intuition about which variant of /ai/ appears in a given lexeme and for a transcriber to be able to clearly judge very clearly which variant was actually produced. It is thus often possible to draw a SSE informant's attention to the *side*/*sighed* quasi-contrast, and ask which of those two vowels appears in some trochaic word of interest, such as *psycho*, and get a very clear answer that it is one or the other. Note however that there are some speakers who are completely baffled by such a question, or report that the vowel is intermediate or unclear. For them, the variants are presumably either not part of the SSE segmental inventory in the same way as true contrastive vowels, are not part of the inventory in this word-internal context (a polysystemic approach), or are not part of it at all.

We report here some transcription-based findings from Stuart-Smith's large study of Glasgow speech (see references above). Recall that this large-scale study was of a pool of 32 speakers, who were stratified in order to sample SSE and Glaswegian Scottish English. The subjects were either young "Y" (in their mid-teens) or old "O" (40s-60s), male "M" or female "F", and from Bearsden "Bden", a largely middle class suburb of Glasgow, or from Maryhill "Mhill", a largely working class area of the city. As far as we are aware, this is the first empirical investigation of trochaic /ai/.

A number of /ai/ words (where "word" includes high frequency semi-bound morphemes) were incorporated into a wordlist. We focus here only on transcriptional native-speaker judgments of length in these trochaic materials, though we have made extensive (mostly unpublished) transcriptional and duration/formant analysis of /ai/ in monosyllabic (Scobbie et al., 1999b) and trochaic words which back up these judgements. Each speaker's /ai/ in simple and trochaic environments was transcribed on two occasions from digitised tape by the first author, and the rare discrepancies resolved by further speaker-internal comparisons. Unlike /i/ and /ʉ/, the short and long variants of /ai/ have a strong qualitative distinction which makes identification of the variant fairly simple once the transcriber has a model for their acoustic space based on the simpler monosyllabic lexemes.

Table 1. Summary of results for OM, OF and YM subjects. White cell with "s" = short /ai/, empty white cell = long /ai/, and a diagonal line indicates variation.

bible	sidle	libel	micro	nitro	hydro	title	tidal	pylon	crisis	miser
s	s					s	s			

Full results for the older men (OM) and women (OW) and young men (YM) are reported in the appendix, but can be summarized as follows (Table 1).[8] In general, *bible*, *sidle*, *title* and *tidal* are pronounced with short /ai/. *Crisis* is generally short, but may be long among older (especially older male Bearsden) speakers; thus length may be a social variable among older speakers in (some) words in which /ai/ is followed by a voiceless fricative. *Miser* is long, as are *pylon*, *hydro*, *nitro* and *micro*. *Libel* is long among the older speakers, but was largely unfamiliar to and mispronounced by the young males (and of the three who managed it, it was long for two and short for one). Two young males stood out because they had a short vowel in *micro*.

8 OF Speaker 4 from Bearsden has uniformly long /ai/, reflecting her accent generally, which is Anglicised and therefore not really typical of "basic" SSE.

Phonologically, these results exemplify a near contrast (*bible* vs. *libel*) which has often been reported anecdotally, and the preservation of short /ai/ in polymorphemic words based on a closed syllable stem which itself has short /ai/ (*tide* = *tidal*). A completely new result is the interspeaker consensus about short /ai/ before voiced stops in *bible*, *sidle* vs. long /ai/ before voiceless ones *nitro*, *micro*. This shows that the voice and manner of the consonant following /ai/, if it is relevant to the choice of /ai/ variant, is only one aspect of a more complex set of factors. This conditioning system may either be segmentally non-local or possibly prosodic: it seems (from other pilot data) that the nature of the weak syllable, in particular its rhyme, is crucial in conditioning /ai/ variants. For /ai/ plus a voiceless fricative, for example, we suspect a short vowel may be more common in some "long-distance" contexts, e. g. in a trochee terminating in a lateral or rhotic (*rifle*, *cipher*), but a long vowel may be more common in others, such as a trochee terminating in a nasal, obstruent, or vowel (*hyphen*, *Pisces*, *ISA*). Perhaps another way to approach these results is to say that such words are not trochees, but a strong-weak sequence of two monosyllabic feet (like *gymnast*), but it is not clear that shifting the problem onto footing is a revealing step. Rather, we expect gradience.

For example, we suspect that voiced fricatives will generally condition more long vowels than voiceless ones, both in terms of their distribution and in terms of the number of lexemes affected. Further, we suspect that stops and other post-vocalic segments will not pattern identically to fricatives. Overall, these complex conditioning patterns will offer statistical prediction of long and short variants, which is another way of saying the variant of /ai/ is partially unpredictable.

Word-internal /ai/ in obviously non-trochaic contexts may be a little less complex and a little more predictable. A long variant appears foot-finally, even when the post-vocalic consonant is a voiceless fricative (*typhoon*). And footing may determine whether morpheme-final /ai/ is short (*bicycle*) or long (*bisect*).

Turning back to the unpredictability of variants, the behaviour of *libel* suggests an underlying contrast somewhere with *bible*, but the problem is identifying where it is. It may be short vs. long /ai/, the prosodic structure, the syllabification of the /l/, or the presence of a phantom vowel in *libel* (cf. *libellous*). Polysyllabic *tidal*, on the other hand, exemplifies faithfulness to the vowel in *tide*. Aitken suspected an incipient phonemic contrast arising out of these complex distributional generalizations, even though we doubt he perceived just how complex the predictable contexts could be. The very preliminary data in Table 1 offers some support for this view.

Table 2. Results for young female subjects. White cell with "s" = short /ai/, empty white cell = long /ai/, grey cell = no data due to a subject error in reading the word.

			bible	sidle	libel	micro	nitro	hydro	title	tidal	pylon	crisis	miser
YF	**Bden**	**1**	s	s			s		s	s		s	
		3	s	s		s	s		s	s			
		4	s	s		s			s	s		s	
		5	s			s			s			s	
	Mhill	**1**		s					s	s		s	
		2		s					s			s	
		3	s						s	s			
		4	s	s		s	s		s	s		s	

We turn now to individual results from the young women (Table 2), and find a very different pattern – or lack of it. First, there are many examples of /ai/ with a length (short or long) which had not been seen in other speakers above; second, there is a great deal of interspeaker variation; third, phonotactically similar words may have different length vowels. For example, some speakers have an unexpectedly long /ai/ in *bible*, in *sidle*, or in both. Some speakers have an unexpectedly short /ai/ in *micro*, *nitro*, or both. Indeed no two speakers have the same system, and though this may be due to lack of data, we suspect that a larger wordlist would have elicited even more variation in the lexical incidence of short and long /ai/.

These speakers offer support for lexical specification of short and long /ai/, because some of the individual distributions are unlikely to be systematisable on general phonotactics grounds, even complex ones such as were hypothesised above. It is always possible, however, that these 14-year old subjects had not yet learnt the distribution of short and long /ai/, and that there is no language-change aspect to these results. But interpreting Table 2 as a pattern of late acquisition does not solve the problem of the phonemic status of the SVLR variants, and simply underlines the ambiguous, indeterminate and complex nature of the phenomenon in a different way. Phonetic variants of /ai/ are early acquired in simpler environments (Matthews, 2001).

7. Summary, discussion and conclusions

We have considered some of the difficulties in establishing the consonant and vowel inventories of Scottish-accented Standard English (SSE) on fairly narrow phonological grounds. It must not be thought that these difficulties arise due to sociolinguistic or stylistic variation, and that they can be dismissed as just so much "noise" by researchers whose focus is exclusively phonological theory. We think that any variation presented above is relevant to phonology in the narrowest sense. This does not imply that we think sociolinguistic variation is irrelevant to phonology, indeed, quite the opposite. Rather, we think that strictly modular phonology is both based on unrealistic and arbitrary data while at the same time being theoretically limited and unable to deal with phonology's interactions with other modules (Scobbie, 2005a, 2007; Foulkes and Docherty, 2006; Stuart-Smith, 2007).

The contrastive inventory of Scottish Standard English, like any language, offers a number of phonologically uncertain phenomena, and the SVLR is perhaps the most complex of these. In addition to the structurally-conditioned quasi-phonemic contrast in word-final stressed syllables, we examined word-internal /ai/, which has two clear variants. These function as allophones in some contexts, have a QP contrast, and also appear unpredictably when word internal (in the first syllable of a trochee). We presented new data on the lexical incidence of long and short /ai/ from a small empirical study of 32 speakers. In the young female subjects, it is not possible to predict with certainty the lexical incidence of short or long /ai/, whereas the appearance of the variants in other speakers appears to follow statistically certain phonotactic regularities. This unpredictable lexical incidence adds weight to the near-phonemic status of the variants of /ai/, since they seem to have to be specified lexically. Other facts relating to /ai/ may also lend support to the near-phonemic status of both variants, without tipping the balance decisively over. For example, Scots dialect has marginal minimal pairs like *gey* [ɡʌɪ] "very" vs. *guy* [ɡɑe], though speakers with a *gey*/*guy* contrast may have the straightforward QP contrasts described here, a situation which requires further research.

We thus do not offer a solution to the question of whether /ai/ is one member of the inventory of SSE or two. One reason for this is that we hope to leave the reader with the same sense of unease which we feel about the requirement to adopt one ill-fitting and rigid phonological analysis over another. An uncontroversial analysis may be possible given more evidence, but we doubt it. In our experience (and we are adding little here to what was said explicitly by Pike, 1947) *every* language has a rump of potential / actual near-phonemes. These problematic segments are characterized by such factors as low functional load, limited phonotactic distribution, contrast in only a limited phonotactic or grammatical environment, few or no examples of real minimal pairs, speaker intuitions that are variable or at odds with the distributional facts, late acquisition, unpredictable lexical incidence, lexical stratification (so that contrasts may only be found in names, loan words, sub-lexicons etc.), interference from literacy, patterns of variation and change, complex phonetic correlates, abstract cross-positional (e. g. onset to coda) relationships, ambiguity over whether they are singletons or clusters, and low participation in phonological processes.

In SSE, as with every language, the evidence for the contrastive/phonemic status of some segments will always be weaker than it is for others. All contrasts have different functional loads, and some play a very small role in the language. Are subtle differences in phonemicity *outside* or *inside* phonology? From the point of view of phonology, are all phonemes equal? We think the answer is that some contrasts are more contrastive than others, and that this is not merely to say that the functional load of contrasts varies, because while the load on /x/ vs. /k/ may be low, making it peripheral to the inventory, the contrast is clearly phonemic. On the other hand, the SVLR QP contrast is functionally a bit more important, but there are few minimal pairs and the distinction is in part predictable – so the contrastiveness is weak in a quite different way.

Our approach means that phonology should reflect more closely the patterns in the data, or be clearer about how it has abstracted away from them. We think here particularly of "exemplar" approaches (Pierrehumbert, 2001, 2002; Coleman, 2002) which allow a greater flexibility in the way phonological systems interact with phonetics, the lexicon and sociolinguistics. Specific parts of such interactions are explored by Boersma (e. g. Boersma, Escudero, and Hayes, 2003; Boersma, 1998), by Gafos (2006), Foulkes and Docherty (2006), and Scobbie (2006). One thing which we did not mention above which is relevant is that the *phonetic* distinctiveness of /k/ and /x/ on the one hand and /w/ and /ʍ/ on the other is also weakening (Lawson and Stuart-Smith, 1999), tying categorical and phonetic changes together in this case. Other changes (e. g. the derhoticisation of coda /r/) involve shifts in the cues used for a contrast, with resulting systematic re-organisation.

Our position is that it is unsatisfying – and probably misleading – to have to adopt one concrete solution to the "problematic" patterns outlined above. Modular phonologies are by definition ill-structured to capture the ways in which the abstract parts of an individual's grammar can capture and represent the partial, indeterminate and fuzzy nature of concrete phonological phenomena. The best they can do is accept that the phonetic instantiation of phonological categories can be vague and variable "underneath", i. e. in a different module, in a way invisible to the phonology proper. In this regard we disagree absolutely and fundamentally with Pike (and with mainstream generative phonology) that "ultimately, only one accurate analysis can be made of any one set of data" (Pike, 1947: 64). For the "easy" parts of a language, there may well be an obvious and straightforward analysis, but on the periphery, where things get *interesting* because phonology is undergoing change, is hard to acquire, or is highly marked, it is reasonable to posit that the mind of the speaker can entertain alternative or intermediate solutions to the incomplete and ambiguous paradigms that surround them. (Furthermore, intra-speaker variation supports this view.)

Our view is that indeterminate phonological data cannot be explained by models which presuppose that phonology provides unique solutions. An exemplar approach, on the other hand, seems to force messy and ambiguous facts to percolate into higher levels of the analyses, because the basic distributions of exemplars is always present in the grammar. Frequency effects and phonetic detail are not assigned to a different module of the grammar from the phonology, and phonological categories are not merely present or absent. Instead, clear clumps of exemplars in phonetic space self-organise into contextualised categories, and the clarity of such clumping may be a moot point. In a traditional modular approach, category status (and a label) is attributed to phonetic distributions which pass some threshold of phonologization. The phonological module contains constraints or other aspects of the grammar which range over the labels, without *ever* being able to access the underlying distributions, and without any conception that some categories are better-formed or more robust than others. The exemplar view, though as yet very sketchy and lacking in many firm predictions, offers a clear mechanism for expressing gradual phonologisation, gradient contrast, nondeterminism, and fuzzy boundaries, all of which are real and pervasive in any phonology, not just in the case of Scottish English exemplified above.

An alternative is to maintain a modular approach, and to decrease the granularity of the phonological categories, providing labels which are very fine-grained. But we suspect this is merely a notational variant of the exemplar approach. In any case, ultra-fine-grained phonology (incorporating highly-specific phonetic targets which are contextualised in similarly fine-grained fashion) seems to be required in order to deal with learned differences in sound systems. And still a non-deterministic and fuzzy formalism would be required to handle variation, subregularities, gradual phonologisation, and "nearly" phenomena like quasi-phonemic contrast.

Phonological systems often include a fascinating and theoretically contentious body of data, the interpretation of which is equivocal. We do not believe that native speakers arrive at an unequivocal phonological system in such cases as the end point of acquisition. Our view is that a more direct representation of equivocal distributions is required. Thus phonology has to be an *analytic framework* in which core concepts like contrast and categorization could and should be formalised as emergent, flexible, gradient and non-deterministic.

8. Acknowledgments

This research, over a long timescale, has been mainly supported by Leverhulme Trust and ARHC grants to Jane Stuart-Smith and ESRC grants (R000271195, R000237135) to Jim Scobbie, for which we are very grateful. We would like to thank Claire Timmins for all her input in data collection, and Bob Ladd for discussion, and John Harris and the editors for comments on a draft. Errors and omissions remain our responsibility. Thanks too to Chiara Frigeni for all her support.

References

Abercrombie, David
1979 The accents of Standard English in Scotland. In: A.J. Aitken and Tom McArthur (eds.), *Languages of Scotland*, 68–84. Edinburgh: W. & R. Chambers.

Aitken, A.J.
1981 The Scottish Vowel-length Rule. In: Michael Benskin and M.L. Samuels (eds.), *So Meny People, Longages and Tonges*, 131–157. Edinburgh: Middle English Dialect Project.

Aitken, A.J.
1984 Scots and English in Scotland. In: Peter J. Trudgill (ed.), *Language in the British Isles*, 517–532. Cambridge: CUP.

Archangeli, Diana and Pulleyblank, Douglas
1994 *Grounded Phonology*. Cambridge, MA.: MIT Press.

Boersma, Paul
1998 *Functional Phonology*. Amsterdam: HIL.

Boersma, Paul, Escudero, Paula and Hayes, Rachel
2003 Learning abstract phonological from auditory phonetic categories: An integrated model for the acquisition of language-specific sound categories. *Proceedings of the 15th International Congress of Phonetic Sciences*: 1013–1016.

Bybee, Joan and Paul Hopper (eds.)
2002 *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins.

Chomsky, Noam and Halle, Morris
1968 *The Sound Pattern of English*. New York: Harper and Row.
Coleman, John
2002 Phonetic representations in the mental lexicon. In: Jacques Durand and Bernard Lax (eds.), *Phonetics, Phonology, and Cognition*, 96–130. Oxford: Oxford University Press.
Escudero, Paula and Boersma, Paul
2004 Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26: 551–585.
Fitt, Matthew, Rennie, Susan and Robertson, James (eds.)
2002 *The Hoose o Haivers*. Edinburgh: Itchy Coo.
Flemming, Edward
2001 Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18: 7–44.
Foulkes, Paul and Docherty, Gerard J.
2006 The social life of phonetics and phonology. *Journal of Phonetics* 34: 409–438.
Gafos, Adamantios I.
2006 Dynamics in Grammar: Comment on Ladd and Ernestus & Baayen. In: Louis M. Goldstein, D.H. Whalen and Catherine T. Best (eds.), *Papers in Laboratory Phonology 8: Varieies of Phonological Competence*, 51–79. Berlin: Mouton de Gruyter.
Giegerich, Heinz J.
1992 *English Phonology: An Introduction*. Cambridge: CUP.
Hale, Mark and Reiss, Charles
2000 Substance abuse and dysfunctionalism: Current trends in phonology *Linguistic Inquiry* 31: 157–69.
Harris, John
1990 Derived phonological contrasts. In: S. Ramsaran (ed.), *Studies in the Pronunciation of English: A Commemorative Volume in Honour of A. C. Gimson*, 87–105. London: Routledge.
Harris, John
1994 *English Sound Structure*. Oxford: Blackwell Publishers.
Hawkins, Sarah and Smith, Rachel
2001 Polysp: A polysystemic, phonetically-rich approach to speech understanding. *Italian Journal of Linguistics – Rivista di Linguistica* 13: 99–188.
Johnston, Paul A.
1997 Regional variation. In: Charles Jones (ed.), *The Edinburgh History of the Scots Language*, 433–513. Edinburgh: Edinburgh University Press.
Ladd, D. Robert
2006 Distinctive phones in surface representation. In: Louis M. Goldstein, D.H. Whalen and Catherine T. Best (eds.), *Papers in Laboratory Phonology 8: Varieties of Phonological Competence*, 1–26. Berlin: Mouton de Gruyter.

Ladefoged, Peter and Maddieson, Ian
1995 *The Sounds of the World's Languages*. Oxford: Blackwell Publishers.

Lawson, Eleanor and Stuart-Smith, Jane
1999 A sociophonetic investigation of the 'Scottish' consonants /x/ and /hw/ in the speech of Glaswegian children. *Proceedings of the XIVth International Congress of Phonetic Sciences*: 2541–4.

Macafee, Caroline I.
1983 *Varieties of English Around the World: Glasgow*. Amsterdam: Benjamins.

McKenna, Gordon E.
1988 Vowel duration in the Standard English of Scotland. M. Litt dissertation, Department of Linguistics, University of Edinburgh.

Marshall, Jonathan
2004 *Language Change and Sociolinguistics: Rethinking Social Networks*. Basingstoke: Palgrave Macmillan.

Matthews, Benjamin M.
2001 On variability and the acquisition of vowels in normally developing Scottish children (18–36 months). Ph. D. dissertation, Department of Speech and Language Sciences, Queen Margaret University College Edinburgh.

Pierrehumbert, Janet B.
2001 Stochastic phonology. *Glot International* 5: 195–2.

Pierrehumbert, Janet B.
2002 Exemplar dynamics: Word frequency, lenition and contrast. In: Joan Bybee and P. Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*, 137–157. Amsterdam: John Benjamins.

Pierrehumbert, Janet B., Beckman, Mary E. and Ladd, D. Robert
2000 Conceptual foundations of phonology as a laboratory science. In: Noel Burton-Roberts, Philip Carr and Gerard J. Docherty (eds.), *Phonological Knowledge: Conceptual and Empirical Issues*, 273–303. Oxford: Oxford University Press.

Pike, Kenneth
1947 *Phonemics*. Ann Arbor: The University of Michigan Press.

Scobbie, James M.
2005a Interspeaker variation among Shetland Islanders as the long term outcome of dialectally varied input: Speech production evidence for fine-grained linguistic plasticity. *QMU Speech Science Research Centre Working Paper* WP-2. [http://www.qmu.ac.uk/ssrc/]

Scobbie, James M.
2005b The "end" of phonology: The theoretical significance of interface phenomena. Oral paper at the 1st International Conference on the Linguistics of Contemporary English. University of Edinburgh, Scotland, 23–26 June.

Scobbie, James M.
2006 Flexibility in the face of incompatible English VOT systems. In: Louis M. Goldstein, D.H. Whalen and Catherine T. Best (eds.), *Papers in Laboratory Phonology 8: Varieties of Phonological Competence*, 367–392. Berlin: Mouton de Gruyter.

Scobbie, James M.
2007 Interface and overlap in phonetics and phonology. In Gillian Ramchand and Charles Reiss (eds.) *The Oxford Handbook of Linguistic Interfaces*, 17–52. Oxford: Oxford University Press.

Scobbie, James M., Gordeeva, Olga B. and Matthews, Ben
2007 Scottish English. In: Sharynne McLeod (ed.), *The International Guide to Speech Acquisition*, 221–240. Clifton Park, NY: Thomson Delmar Learning.

Scobbie, James M., Hewlett, Nigel and Turk, Alice E.
1999a Standard English in Edinburgh and Glasgow: The Scottish vowel length rule revealed. In: Paul Foulkes and Gerard J. Docherty (eds.), *Urban Voices: Accent Studies in the British Isles*, 230–245. London: Arnold.

Scobbie, James M., Turk, Alice E. and Hewlett, Nigel
1999b Morphemes, phonetics and lexical items: The case of the Scottish vowel length rule. *Proceedings of the XIVth International Congress of Phonetic Sciences* 2: 1617–1620.

Stuart-Smith, Jane
2003 The phonology of modern urban Scots. In: John Corbett, J. Derrick McClure and Jane Stuart-Smith (eds.), *The Edinburgh Companion to Scots*, 110–137. Edinburgh: Edinburgh University Press.

Stuart-Smith, Jane
2007 Empirical evidence for gendered speech production: /s/ in Glaswegian. In Jennifer Cole and Jose Hualde (eds.), *Change in Phonology: Papers in Laboratory Phonology* 9, 65–86. Berlin: Mouton de Gruyter.

Stuart-Smith, Jane, Timmins, Claire and Tweedie, Fiona
2007 "Talkin' Jockney"? : Accent change in Glaswegian. *Journal of Sociolinguistics* 11: 221–260.

Van Leyden, Klaske
2002 The relationship between vowel and consonant duration in Orkney and Shetland dialects. *Phonetica* 59, 1–19.

Appendix

Table A. Results for other subjects. White cell with "s" = short /ai/, empty white cell = long /ai/, grey cell = no data due to a subject error in reading the word.

			bible	sidle	libel	micro	nitro	hydro	title	tidal	pylon	crisis	miser
OM	Bden	**1**	s	s					s	s			
		2	s	s					s	s			
		3	s	s					s				
		4	s	s					s	s			
	Mhill	**1**	s	s					s	s		s	
		2	s	s						s		s	
		3	s	s					s	s		s	
		4	s						s	s		s	
OF	Bden	**1**	s	s					s	s		s	
		2	s						s	s			
		3	s	s					s	s		s	
		4											
	Mhill	**1**	s	s					s	s			
		2	s	s					s	s		s	
		3	s						s	s			
		4	s	s					s	s		s	
YM	Bden	**1**	s						s			s	
		2	s	s					s	s		s	
		3	s	s					s	s		s	
		4	s	s					s			s	
	Mhill	**1**	s	s	s				s	s		s	
		2	s	s					s	s		s	
		3	s	s		s				s			
		4	s	s		s			s	s		s	

Effects of contrast recoverability on the typology of harmony systems

Gunnar Ólafur Hansson

1. Introduction

Harmony, like all other types of assimilation, can be viewed as an instance of contextual neutralization: in a given environment one member of a [+F] : [–F] opposition is allowed, while the other is prohibited (see, e. g., Steriade 2001).[1] This straightforward fact is summarized schematically in (1).[2]

(1) Harmony as neutralization (example with root-to-affix directionality)
 a. If root contains [+F], then…
  affix segments are neutralized to [+F] (that is, [–F] is "not licensed")
 b. If root contains [–F], then…
  affix segments are neutralized to [–F] (that is, [+F] is "not licensed")

In the literature on phonological harmony systems it has often been assumed that the elimination of a [+F] : [–F] contrast in the targeted positions – and the ensuing predictability of [±F] values in those positions – is the very "goal", or main function, that underlies harmony itself. This interpretation has typically been motivated with respect to speech perception (Suomi 1983; Kaun 1995) or general processing/ parsing considerations (Kaye 1989; the idea goes back to Trubetzkoy 1939).

Relating harmony to neutralization in this manner brings up an important question which, somewhat surprisingly, is rarely asked in the literature on harmony systems and their formal analysis. To what extent does harmony result in true neutralization in the narrowest possible sense: the obliteration of existing

1 The research reflected here was partly supported by SSHRC Standard Research Grant 410–2004–0710, and by an Early Career Scholar Award from the Peter Wall Institute for Advanced Studies.

2 Here and throughout, all featural contrasts will be rendered formally as binary [+F] : [–F] oppositions, rather than as the presence vs. absence of a privative feature, [F] : Ø, or as mutually incompatible privative features, [F] : [G] (e. g., [ATR] vs. [RTR]). This is solely for simplicity of exposition, and questions of feature valency, and of the formal representation of specific featural contrasts, are entirely orthogonal to the discussion and argumentation throughout this work (see § 4 for elaboration on this point).

lexical contrasts? In this context it is useful to make a terminological distinction between what I will henceforth refer to as *actual* and *virtual* neutralization, respectively, shown schematically in (2). (Note that my choice of representing the disfavoured feature value in the neutralization environment as "[–F]", rather than "[+F]", is entirely arbitrary and not in any way significant.)

(2) a. Actual neutralization (eliminates attested lexical contrast):
UR: /…+F…/ /…–F…/
↓ ↙
SR: […+F…] *[…–F…]
b. Virtual neutralization (no attested lexical contrast to eliminate)
UR: /…+F…/ (*/…–F…/)
↓
SR: […+F…] *[…–F…]

As an example of actual neutralization, (2a), consider the suspension of the /m/ : /n/ contrast in word-final position in Finnish. For example, the two NOM.SG forms [avain] 'key' and [jæsen] 'member' both have word-final [n].[3] Labials like [m] simply do not occur in this position in Finnish words, though they are allowed in other positions (cf. [maː] 'land', [silmæ] 'eye'). Crucially, we can see how the neutralization obliterates an existing lexical /m/ : /n/ contrast by looking at other word forms that are morphologically related to [avain] and [jæsen], such as the NOM.PL forms [avaimet] 'keys', [jæsenet] 'members'.

Compare this to a similar neutralization of the /m/ : /n/ contrast in final position in Mandarin Chinese. Here again, just as in Finnish, we find only word-final [n] in surface forms ([sān] 'three', [hə̌n] 'very'), even though [m] does occur in other environments ([mə́n] 'door'). But as pervasive and systematic as this neutralization pattern may be, it cannot be shown to result in the obliteration of any actual lexical contrasts. There simply do not exist any individual words or morphemes which could be argued to contain final /m/ in their lexical representation. Hence this is a case of virtual neutralization, as in (2b).

It should be noted at this point that in Optimality Theory (Kager 1999; McCarthy 2002, 2003; Prince and Smolensky [1993] 2004), virtual and actual neutralization are in effect equated, by way of the *Richness of the Base* tenet (see McCarthy 2002: 68–82). In a virtual-neutralization environment like (2b), [–F] is consistently absent from output strings. The very systematicity of this

3 The morphological abbreviations occurring in glosses in this paper are as follows: NOM = nominative; SG = singular; DU = dual; PL = plural; POT = potential; EMPH = emphatic; CTFG = centrifugal; 1, 2, 3 = first, second, third person.

gap entails that the phonological grammar of the language in question must have the capacity to repair any potential input representations which contain [–F] in that environment – no matter how hypothetical these may be – by rendering them unfaithfully in the output. In Optimality Theory, all conceivable inputs, real and hypothetical alike, must map onto well-formed outputs, and all neutralization is therefore "actual" in the sense of (2a). From this perspective, what gives Mandarin the appearance of being different from Finnish is simply a consequence of the morphological structure of the former. If a morpheme-final segment never alternates between word-final (coda) and word-medial (onset) position, for example as a consequence of morpheme concatenation, a hypothetical morpheme-final /m/ gets no chance to show its true colours in any surface forms containing the morpheme in question. Consequently, a learner of Mandarin will never see a reason to posit a final /m/ in that morpheme in the first place. Similarly, if all Finnish [n]-final words happened to behave like [jæsen] (with [n] in all related forms), thus making Finnish a case of virtual rather than actual neutralization, then this would not reflect any difference in the phonology of Finnish as such. Instead, the lack of existing lexical entries with stem-final /m/ would have to be an accidental gap in the lexicon, of no particular significance for the phonological analysis of the language.[4]

Positional neutralization phenomena in the world's languages are usually of the Finnish type, where alternations among morphologically related forms provide evidence of lexical contrasts which the neutralization is (partially) obliterating. It should be noted that the locus of such contrasts need not be in root morphemes, as in the Finnish example, but may be in affixes. For example, the underlying value for [±voice] in the English *-th* and *-s* suffixes, though neutralized after voiceless obstruents (*eighth* [eɪtθ] vs. *eights* [eɪts]), emerges intact after sonorants (*ninth* [naɪnθ] vs. *nines* [naɪnz]).

If harmony is merely a particular instantiation of contextual neutralization, our expectation is that it should pattern in ways similar to other types of neutralization. In particular, we ought to expect to see cases where *actual neutralization* results from the assimilation processes of harmony, especially in light of the fact that from an Optimality Theory perspective, all neutralization is strictly speaking "actual", as explained above. The central goal of this paper

4 This is a slight oversimplification, albeit one which is of no consequence for the ensuing discussion. It would in fact be perfectly possible to force neutralization to [n] to extend to related forms as well (where the nasal is non-final), by invoking some mechanism ensuring paradigmatic identity, such as Uniform Exponence (Kenstowicz 1997), Output-Output Correspondence (Benua 2000), Paradigm Uniformity (Steriade 2000), or Optimal Paradigms (McCarthy 2005).

is to demonstrate that things are not quite so simple. In its purest form, actually-neutralizing assimilation is robustly attested only for consonant harmony, while it is conspicuously absent from the typology of vowel harmony systems. This curious asymmetry among harmony systems, which thus far appears to have gone unnoticed, needs to be explained in some principled way.

I propose an explanation in terms of the relative *recoverability* of the lexical contrasts in question: their "discoverability" on the basis of available surface evidence (see Kaye 1974). I argue that in positions targeted by harmony, lexical contrasts are far more easily recoverable – and hence more easily and securely acquired – under consonant harmony than under vowel harmony. The ultimate source of the bias is a trivial yet substantial asymmetry between consonants and vowels with respect to inventory size, inventory structure, and general phonotactic distribution. Finally, I briefly address how the crucial type of actually-neutralizing harmony, which is attested among consonant harmony systems but not in vowel harmony, has important implications for the analysis of directionality effects in output-oriented frameworks like Optimality Theory.

2. Neutralization patterns in harmony systems

In any harmony system, the segmental inventory of the language in question can be partitioned into three classes of segments with respect to their participation, or lack thereof, in the harmony pattern. (Note that this classification is intended as purely taxonomic, with no particular implications as regards theoretical assumptions.)

(3) Classes of segments in a given harmony system:
  a. all non-neutral [+F] segments
  b. all non-neutral [–F] segments
  c. all neutral segments (may be an empty set)

For example, in a typical ATR harmony system like that of Akan, (3a) consists of [+ATR] /i, u, e, o/, (3b) of [–ATR] /ɪ, ʊ, ɛ, ɔ/, and (3c) of the low vowel /a/. In Turkish palatal harmony, (3c) is an empty set, as there are no neutral vowels, whereas (3a) consists of [+back] /ɯ, u, ɑ, o/, and (3b) of [–back] /i, y, e, ø/.

Individual morphemes in the lexicon may of course be similarly classified with respect to the kinds of segments they contain, just as the English suffixes *-th* and *-s* can be classified as containing an underlyingly [–voi] and [+voi] obstruent, respectively. This yields the following typology of harmony processes, based on whether or not the harmony gives rise to actual neutralization, and if so, under what circumstances such neutralization takes place.[5] My decision to represent the non-harmony-triggering feature value in (4c) as [–F], rather than [+F], is entirely arbitrary. The idea is simply that in the (4c) case, one of the two [F]-values is inert, failing to trigger assimilation; for a particular feature in a particular language, that value might well be [+F] rather than [–F].

(4) Four-way typology with respect to (actual) neutralization in affixes:
  a. lexical [±F] contrast not maintained with any root (contrast unrecoverable)
  b. lexical [±F] contrast maintained with neutral roots only
  c. lexical [±F] contrast maintained with [–F] roots or neutral roots
  d. lexical [±F] contrast maintained with [+F], [–F] or neutral roots (no harmony)

Type (4d) can obviously be ignored, as it consists of languages which do not exhibit any harmony whatsoever. The remaining three are attested cross-linguistically to varying degrees. Type (4a) is by far the most common, and seems to be equally well attested for vowel harmony and consonant harmony. As for type (4c), it is well attested for consonant harmony, but perhaps somewhat less so for vowel harmony. The most interesting type by far is (4b) which, though robustly attested among consonant harmony systems, appears to be entirely unattested for other kinds of harmony.

The following sections illustrate this typology ranging over (4a-c), not only with real examples but also, in the case of typological gaps, with made-up examples, so as to show what such a system would look like if it did exist. As vowel harmony is a more widespread phenomenon than consonant harmony, it will provide our point of departure, in § 2.1, followed by a corresponding survey of consonant harmony in § 2.2. Each is divided into subsections corresponding to the three neutralization patterns in (4a-c).

---

5 The following discussion is restricted to lexical contrasts in affixes, ignoring roots as harmony undergoers (e. g., in dominant-recessive harmony, umlaut, metaphony, etc.). Surface evidence for lexical [+F] : [–F] contrasts is usually much more readily available for root morphemes, for the following reasons: (i) unlike affixes, a root may frequently occur on its own as an independent word; (ii) a root forms the central “hub” of an entire paradigm of morphologically related forms, in ways that an affix does not; and (iii) relevant information about a root’s lexical representation may be distributed across several forms in that paradigm. In very rare cases, affixes may themselves act as independent roots in certain constructions, in which case their underlying contrastive [±F] value may become apparent. This appears to be the case in Hungarian, which is otherwise much like Finnish in the relevant respects (see Ringen and Vago 1998).

2.1. Neutralization patterns in vowel harmony

2.1.1. Vowel harmony with complete neutralization

In the typical case, corresponding to (4a), harmony is manifested solely as "virtual" neutralization, in that there is no evidence of an underlying lexical contrast in affixes which is being obliterated by the harmony. In systems of this kind, the surface [±F] value in the affix is completely predictable given the root. That is, a hypothetical [±F] contrast among affixes, were it to exist, would get no chance to surface intact. An example is Finnish palatal harmony, illustrated in (5); /i, e/ are neutral vowels.

(5) Finnish [±back] harmony (adessive suffix /-llA/):

a. /katu-llA/	[kadulla]	'on the street'	([+back] root)
b. /pøytæ-llA/	[pøydællæ]	'on the table'	([–back] root)
c. /vete-llA/	[vedellæ]	'on the water'	(neutral root)

Since a lexical [±back] specification for the suffix vowel cannot be determined conclusively, that vowel is here represented archiphonemically as /A/ for convenience. Finnish does not have contrasting pairs of affixes with inherently back vs. front vowels, such as a pair /-lla/ vs. /-llæ/ (or even, say, /-lla/ vs. /-tæ/) with separate meanings or functions. Such a contrast would be perfectly conceivable in principle, and would presumably surface intact after a neutral vowel. After all, this is precisely what happens root-internally, as shown in (6).

(6) Root-internal [±back] contrast after neutral vowels in Finnish:

a. /nenæ/	[nenæ]	'nose'
b. /mela/	[mela]	'oar, paddle'

Another example of a vowel harmony system of this type is tongue root harmony in Akan (Archangeli and Pulleyblank 1994), which affects prefixes and suffixes alike. Here /a/ is neutral, cooccurring with both vowel sets (/bisa/ 'to ask', /pɪra/ 'to sweep').

(7) Akan [±ATR] harmony in prefixes and suffixes:

a. /O-susu-I/	[o-susu-i]	's/he measured (it)'	([+ATR] root)
b. /O-fʊrʊ-I/	[ɔ-fʊrʊ-ɪ]	's/he went up'	([–ATR] root)
c. /O-kasa-I/	[ɔ-kasa-ɪ]]	's/he spoke'	(neutral root)

Just as in the Finnish case, Akan affix vowels do not show any evidence of contrasting lexically for the harmonizing feature. Instead, their surface [±F] specification is completely predictable from context.

2.1.2. Vowel harmony with asymmetric neutralization

In a number of cases, only one [F]-value appears to be active (or "dominant"), inducing harmony on nearby vowels. For example, [+ATR] might spread but not [–ATR], [+round] but not [–round], and so forth. Most cases of harmony processes which target vowels and consonants alike fall in this category as well. For example, nasal harmony typically involves nasalization only, to the exclusion of denasalization (see Walker 2000a for a typology). As a result, in the pattern corresponding to (4b), lexical contrasts will be neutralized only in the vicinity of a segment with the active value (represented here as [+F], regardless of what the actual + or – designations might be in practice). Segments with the inactive value (here [–F]) do not trigger harmony, and thus do not condition neutralization; nor do neutral segments, to the extent that the system in question contains any. Taking an example from nasal harmony, a contrast like /n/ : /l/ in suffixes might be neutralized (to [n]) after [+nasal] roots while remaining intact after [–nasal] roots.

Having only one feature value be active is the essential ingredient in so-called dominant-recessive vowel harmony. The prototypical system of this kind also involves bidirectional spreading (affix-to-root as well as root-to-affix; e. g., in Kalenjin, Turkana, Nez Perce, etc.), which creates additional complications. However, this is not always the case, as unidirectional dominant-recessive harmony also exists (*pace* Baković 2000). An example of this is tongue root harmony in Karajá, a Macro-Jê language of Brazil (Ribeiro 2001, 2002). Just as in most dominant-recessive tongue root systems, [+ATR] is the dominant value, triggering assimilation in nearby [–ATR] vowels. With regard to directionality, the harmony is strictly regressive/anticipatory: recessive vowels which follow a dominant one are unaffected, as shown in (8).

(8) Right-to-left [+ATR] harmony in Karajá
a. Permitted vowel sequences:
[+ATR]…[+ATR]
[–ATR]…[–ATR]
[+ATR]…[–ATR] (no progressive harmony)
b. Prohibited vowel sequence:
*[–ATR]…[+ATR] (→ [+ATR]…[+ATR] by regressive harmony)

As shown below, vowel harmony in Karajá holds both morpheme-internally (9a) and between morphemes (9b). Vowels whose surface [±ATR] value is entirely determined by harmony are underlined in the examples. These are in a position of neutralization, namely preceding a [+ATR] vowel, where any and all [+ATR] : [–ATR] contrasts are neutralized to [+ATR].

(9) Examples of Karajá tongue-root harmony (data from Ribeiro 2001, 2002)

a. Root-internally:

/kube/	[ku̲be]	‘palm’	([+ATR]…[+ATR])
/dɔrɛ/	[dɔrɛ]	‘parrot’	([–ATR]…[–ATR])
/tʃuʃɔ/	[tʃuʃɔ]	‘quati’	([+ATR]…[–ATR])

b. Between morphemes:

/r-ɪ-ɗɔ=r-e/	[ri̲ɗo̲re]	‘s/he ate (it)’
/r-ɔ-tʃuhɔ=rɛrɪ/	[ro̲tʃuhɔrɛrɪ]	‘he is cursing’
/r-ɔ-tʃuhɔ=r-e/	[ro̲tʃu̲ho̲re]	‘he cursed’

To make a more direct comparison with Finnish or Akan, it is important to note that unlike in the latter two systems, a lexical [±ATR] contrast is attested in suffixes and clitics in Karajá. On the one hand, there are underlyingly [+ATR] clitics (such as /=le/ EMPHATIC), which always surface with their [ATR] specification intact. On the other hand, there are also underlyingly [–ATR] clitics (such as /=kɛ/ POTENTIAL), the vowels of which are realized as [–ATR] or [+ATR] depending on context, as shown in (10).

(10) Harmony alternation in enclitic /=kɛ/ (Ribeiro 2002):

[rɛlɛk̲ɛ̲	relek̲e̲le]
/r-ɛlɛ=kɛ	r-ɛlɛ=kɛ=le/
CTFG-become=POT	CTFG-become=POT=EMPH

‘He was in the process of becoming [a dolphin]’

In sum, Karajá vowel harmony does neutralize actual [+ATR] : [–ATR] contrasts in clitics and affixes (e. g., /=le/ vs. /=kɛ/), but only before vowels with the dominant feature value, [+ATR]. Elsewhere the underlying contrast is upheld.

Other examples of vowel harmony systems with these properties are surprisingly hard to come by, though they most certainly do exist. Even Karajá is far from being an ideal case; for example, lexical [±ATR] contrasts are conspicuously absent from prefixes.[6] It is also worth noting that Karajá involves tongue-root harmony, as do all reported cases of bidirectional dominant-recessive harmony (Baković 2000); the reasons for this typological limitation are not known. Another case of tongue-root harmony with the same kind of asymmetric neutralization pattern is the Mọba dialect of Yoruba (Perkins 2005). Here

6 For this reason, the prefixes in (9b) should perhaps preferably be rendered as /I-/, /O-/ rather than /ɪ-/, /ɔ-/. If such a prefix contrast did exist, underlyingly [+ATR] prefixes would be expected to surface consistently as [+ATR], whereas [–ATR] ones would alternate between [+ATR] and [–ATR], just as the /=kɛ/ clitic does in (10).

proclitics with mid vowels contrast lexically in [±ATR], but harmony neutralizes this contrast before vowels with the dominant feature value, [–ATR] (or, alternatively, privative [RTR]): in that context, all mid-vowel proclitics surface as [–ATR], even ones which are underlyingly [+ATR].

In the realm of height harmony, a case of asymmetrically neutralizing harmony involving vowel height does appear to be found in C'Lela (Dettweiler 2000; Pulleyblank 2002), where harmony is triggered only by [–high] vowels, including /a/. A lexical [±high] contrast in suffixes and clitics (e. g., 2.SG /vu/ vs. 2.PL /no/) is neutralized after [–high] roots (11a), but emerges intact after [+high] roots (11b).

(11) Asymmetrically neutralizing height harmony in C'Lela (Pulleyblank 2002)

a. Neutralization after [–high] roots:

/batk vu/	→	[batkᵊ vo̠]	'released you-SG'
/batk no/	→	[batkᵊ no]	'released you-PL'

b. Contrast after [+high] roots:

/buzk vu/	→	[buzᵊkᵊ vu]	'chased you-SG'
/buzk no/	→	[buzᵊkᵊ no]	'chased you-PL'

In sum, relatively few vowel harmony systems with the property of asymmetric neutralization appear to be attested. For example, I have yet to find any solid cases involving rounding (despite the fact that [+round] is typically "dominant" in rounding harmony systems) or the back/front dimension. The interim conclusion is that this type of partially neutralizing vowel harmony is fairly rare.[7]

2.1.3. Vowel harmony with symmetric neutralization

This brings us to the last pattern, corresponding to (4c), which appears to be entirely unattested in the cross-linguistic typology of vowel harmony systems.

7 It is important not to confuse the pattern in (11) with another phenomenon, quite commonplace in vowel harmony systems, whereby certain individual morphemes fail to undergo harmony (though the two phenomena are sometimes hard to distinguish in practice). In such cases, the disharmonic behavior of an affix vowel is idiosyncratic, not an automatic consequence of it being specified lexically as [+F] rather than [–F] (or vice versa). Indeed, one typically finds that disharmonic affixes with both [F]-values exist in a given language. In Turkish palatal harmony, for example, the [–back] suffix /-gen/ '(poly)-gon' and the [+back] suffix /-(i)jor/ PRESENT are equally disharmonic in their own idiosyncratic way (cf. [ɑltɯgen] 'hexagon', [gelijor] 's/he is coming').

In this case, neutralization is symmetric in feature-value terms. That is, affixal vowels are neutralized toward [+F] or [–F] depending on the root, just as they are in the Finnish and Akan cases in (5)–(7) above. What is crucial about this (nonexistent) type, however, is that an underlying lexical contrast does emerge intact when no harmony trigger is present – that is, when the root happens to contain only neutral vowels.

Since vowel harmony systems of this kind do not appear to be attested, a hypothetical example will have to suffice as illustration. The one shown in (12) is modelled on Finnish palatal harmony, as laid out in (5) above.

(12) Pseudo-Finnish: [±back] harmony with marginal contrast preservation

a. Lexical [±back] contrast in (certain) suffixes:

/-llæ/	ADESSIVE ('on X; with X')	
/-ssa/	INESSIVE ('in X')	

b. Neutralization to [+back] after non-neutral [+back] roots:

/katu-llæ/	[kadu-lla]	'on the street'
/katu-ssa/	[kadu-ssa]	'in the street'

c. Neutralization to [–back] after non-neutral [–back] roots:

/pøytæ-llæ/	[pøydæ-llæ]	'on the table'
/pøytæ-ssa/	[pøydæ-ssæ]	'in the table'

d. Contrast preserved after neutral roots:

/vete-llæ/	[vede-llæ]	'on the water'
/vete-ssa/	[vede-ssa]	'in the water'

The way in which Pseudo-Finnish differs from real Finnish is twofold. Firstly, vowels of individual suffixes carry their own contrastive [±back] specification. Secondly, this specification comes to light whenever there is no harmony trigger in the root, as in (12d). In real Finnish the relevant forms in (12d) both have [æ]: [vede-llæ, vede-ssæ]. That (12d) is perfectly conceivable in principle is evident from real Finnish forms such as the ones in (13), where a [+back] vowel may occasionally be found after a neutral root:

(13) Genuine Finnish: spurious "contrast" in affixes after neutral roots

a. /vete-llA/	[vede-llæ]	'on/with water'
b. /vere-llA/	[vere-lla]	'on/with blood'

However, what is happening in (13) is rather a matter of lexical contrast among *roots* (typically analyzed in terms of absence vs. presence of a floating [+back] autosegment), which simply happens to be realized on the affix vowel in the surface representation.

2.2. Consonant harmony

Turning now to consonant harmony and its cross-linguistic typology, a few fundamental differences are worth noting which sometimes render direct comparison with vowel harmony difficult. Generally, consonant harmony only operates between segments that are highly similar to one another (Walker 2000b; Hansson 2001; Rose and Walker 2004). It is typically also limited to segments which are contrastively specified for the feature in question. That feature may well be redundant or irrelevant for most segments in the inventory (e. g., a coronal-specific feature such as [±distributed] in the case of vowels and non-coronals, and perhaps some coronals as well). As a result, the class of neutral segments is generally much larger in consonant harmony systems. In a sibilant harmony system, for example, all non-sibilant consonants (as well as all vowels) can be considered neutral, a point to which I shall return in § 3. Most of the individual systems illustrated below are discussed at greater length in Hansson (2001).

2.2.1. Consonant harmony with complete neutralization

The pattern corresponding to (4a), so ubiquitous in vowel harmony, is not extremely common among consonant harmony systems, though it is nevertheless fairly well attested. As an example, consider the sibilant harmony found in several Omotic languages of Ethiopia, such as Koyra (Hayward 1982), shown in (14). Here affix sibilants agree with root sibilants in [±anterior] (or its analogue in alternative feature systems).

(14) Sibilant harmony in Koyra (Hayward 1982)

a. Neutralization to [–ant] after root with [–ant] sibilant

/goːtʃ-uS-/	[goːtʃ-uʃ-]	‘cause to pull’
/paʃ-uS-/	[paʃ-uʃ-]	‘cause to cover up’

b. Neutralization to [+ant] after root with [+ant] sibilant

/kes-uS-/	[kes-us-]	‘cause to go out’
/suːz-uS- /	[suːz-us-]	‘cause to bless’

c. Neutralization to [+ant] after neutral (sibilant-free) root

/tup-uS-/	[tup-us-]	‘cause to tie’
/ʔuːʔ-uS- /	[ʔuːʔ-us-]	‘cause to sip’

This is completely parallel to the Finnish case in (5) above. In Koyra, an affixal sibilant is realized as [–anterior] after roots containing a [–anterior] sibilant, otherwise as [+anterior]. In Finnish, an affixal vowel is realized as [+back] after roots containing a [+back] vowel, otherwise as [–back].

2.2.2. Consonant harmony with asymmetric neutralization

This pattern, corresponding to (4b), is even more robustly attested for consonant harmony than the complete-neutralization pattern just presented, and seems far more common than what we saw for vowel harmony in § 2.1.2. An example of a system with these properties is the long-distance [±nasal] agreement found in many Bantu languages. For example, in Yaka (Hyman 1995), a nasal anywhere in the word forces all subsequent voiced consonants to surface as nasals as well. Thus sequences like *[m…d] or *[n…b] are prohibited, and are repaired to [m…<u>n</u>], etc., whenever they arise through morpheme concatenation (within the appropriate morphological domain). Only [+nasal] is active, not [–nasal]: affix segments can be nasalized through harmony but never denasalized. The examples in (15) are drawn from Hyman (1995), and also directly from Ruttenberg (1968) via the on-line CBOLD database (http://www.cbold.ddl.ish-lyon.cnrs.fr/).[8]

(15) Neutralization patterns in Yaka nasal consonant harmony:

a. Lexical [±nasal] contrast among suffixes:

/-idi/	PERFECTIVE
/-an-/	RECIPROCAL

b. Neutralization after roots with [+nasal] voiced C:

/-tsúm-idi/	[-tsúm-i<u>n</u>i]	'sewed'
/-tsúm-an-/	[-tsúm-a<u>n</u>-]	'sew each other' (contrived form)

c. Contrast preserved after roots with [–nasal] voiced C:

/-kúd-idi/	[-kúd-i<u>d</u>i]	'chased'
/-kúd-an-/	[-kúl-a<u>n</u>-]	'chase each other'

d. Contrast preserved after neutral roots (no voiced C):

/-kík-idi/	[-kík-i<u>d</u>i]	'connected'
/-kík-an-/	[-kík-a<u>n</u>-]	'connect each other'

Just as in the Karajá and C'Lela vowel harmony systems discussed in § 2.1.2, affix segments which contain the active/dominant [F]-value (here [+nasal]) surface intact regardless of context. Affix segments containing the inert/reces-

---

8 Note that [d] and [l] are allophones in complementary distribution in Yaka, [d] occurring before [i], and [l] occurring elsewhere; the phoneme is referred to here as /d/.

sive value [–nasal], on the other hand, alternate depending on the harmonic context.

2.2.3. Consonant harmony with symmetric neutralization

We are now left with the neutralization pattern which appeared to be missing from the typology of vowel harmony systems, namely (4c), wherein a lexical contrast in affixes emerges only with neutral roots. Despite the fact that consonant harmony is comparatively much rarer than vowel harmony, there is no corresponding gap in the typology of consonant harmony. On the contrary, the symmetric-neutralization pattern in (4c) is robustly attested for consonant harmony. A case in point is the sibilant harmony found in Navajo, as well as in many other Athabaskan languages, illustrated in (16).

(16) Sibilant harmony in Navajo (data from Sapir and Hoijer 1967)

a. Lexical contrast in prefixes:

/si-/	ASPECT (usually perfective, though not in these examples)
/ʃi-/	1.SG (possessive)

b. Neutralization to [–ant] before roots with [–ant] sibilant:

/si-ɣiʃ/	[ʃi-ɣiʃ]	'it is bent, curved'
/ʃi-tʃ'íːʔ/	[ʃi-tʃ'íːʔ]	'my intestines'

c. Neutralization to [+ant] before roots with [+ant] sibilant:

/si-sį́/	[si-zį́]	'it (long object) lies'
/ʃi-tseʔ/	[si-tseʔ]	'my rock'

d. Contrast preserved before neutral (sibilant-free) roots:

/si-ʔą́/	[si-ʔa]	'it (round object) lies'
/ʃi-taːʔ/	[ʃi-taːʔ]	'my father'

Another well-known case is the sibilant harmony found in many Chumashan languages, such as Ineseño (Applegate 1972; Poser 1982; Lieber 1987: 145–150), which has prefixal contrasts like 3.SUBJ /s-/ vs. DU.SUBJ /iʃ-/. Each surfaces intact before a neutral root, preserving the underlying /s/ : /ʃ/ contrast. Before a root containing /s/, /ts/, etc., both prefixes surface with [s]; before a root with /ʃ/, /tʃ/, etc., both surface with [ʃ].

The neutralization patterns displayed by the sibilant harmony systems of such languages as Navajo and Ineseño Chumash are entirely analogous to those of the hypothetical Pseudo-Finnish vowel harmony system outlined in (12) earlier. In all three, affix segments are contrastively specified for [+F] vs. [–F], and this underlying contrast emerges only in contexts where the affixes in question attach to neutral roots, whereas it is neutralized in all other circumstances, to [+F] or to [–F] depending on the root.

3. Explaining the typological gap: the role of recoverability

The brief survey in the preceding section raises an important question. Given the fact that vowel harmony is such a common phenomenon, in contrast to the comparative rarity of consonant harmony, why is it that actual neutralization – the obliteration of real lexical contrasts – is attested in the typology of consonant harmony systems but not (or only marginally so) in that of vowel harmony systems?

To my knowledge, the only work which comes close to addressing this problem is Lieber (1987: 145–150). To be exact, the question she raises is a slightly different but closely related one: why is *feature-changing* harmony so remarkably rare? (See § 4 for discussion of the feature-changing vs. feature-filling distinction in this context.) In fact, the only case of feature-changing harmony of which Lieber is aware is the Chumash sibilant harmony system just mentioned. Her suggested explanation for the rarity of such harmonies invokes the relative markedness of different rule types. In her analysis of Chumash sibilant harmony, sibilant harmony is decomposed into an ordered sequence of two rules (following Poser 1982). First, an unbounded delinking rule removes contrastive [±distributed] (or perhaps [±anterior]) specifications from sibilants whenever these are followed by another sibilant somewhere later in the word. Next, a feature-filling rule spreads [±distr] (or [±ant]) specifications to these same sibilants. From this Lieber conjectures that "if […] feature-changing harmonies require unbounded Delinking rules […], and if this sort of rule is highly marked and therefore very costly to a grammar, then we would expect feature-changing harmonies to be rare, perhaps virtually nonexistent" (Lieber 1987: 149).

The most obvious problem with this explanation is that it is utterly circular. From the observed rarity of unbounded delinking rules we infer that these must be "costly" elements of grammar (never mind the vagueness of the "cost" notion itself), and because feature-changing harmony employs a costly kind of operation, it is consequently rare. Why not stipulate instead that it is simply feature-changing harmony as such which is "highly marked and therefore very costly" (especially considering the fact that the very notion of unbounded delinking operations is only needed as a component of such harmonies in the first place)? Secondly, the interpretation of feature-changing harmony as delinking plus spreading rests entirely on a serialist conception of phonological grammars, and becomes utterly meaningless in a parallelist constraint-based perspective such as that of Optimality Theory. Finally, even if we were to accept Lieber's explanation for the rarity of feature-changing (and thus actually-neutralizing) harmony, we are still left with a bigger conundrum: why is it only attested in consonant harmony, not vowel harmony? Given that the latter kind of harmony is so vastly more common in the world's languages, we ought to expect the exact opposite to be the case.

I suggest that the answer to the question instead ultimately lies in the relative *recoverability* of lexical contrasts under these different kinds of harmony: vowel harmony on the one hand and consonant harmony on the other. The term "recoverability" here refers simply to the relative amount of surface evidence available to language learners, on the basis of which they can establish whether such a lexical contrast exists in the first place (cf. Kaye 1974).[9]

In order for a lexical contrast to exist in affix vowels (or consonants), it must of course be learnable. That is to say, generations of learners need to be able to reliably discover the existence of that contrast from surface evidence available in the ambient stimulus data. In order for this to be possible, there need to exist at least some contexts in which the contrast is manifested as such on the surface rather than neutralized. Consider now the fact that every root typically contains at least one vowel and at least one consonant. In order for an affix contrast to surface intact – such that a learner might be expected to notice its existence – the nearest relevant root vowel (in vowel harmony) or consonant (in consonant harmony) must not be a harmony trigger, but rather a neutral or non-harmony-inducing segment of some kind. What needs to be determined, then, is the following: what are the odds that this will indeed be the case, and are these odds any different in vowel harmony than in consonant harmony?

There are several fundamental asymmetries between vowels and consonants that bear on this matter, most of them deriving from some rather mundane facts of life. Firstly, there is a striking difference in the nature of the vowel space and the "consonant space", which is in turn directly reflected in inventory structure. The features which form the basis of vowel harmony systems tend to cross-cut the entire vowel space: every vowel is either front or back, either rounded or unrounded, and so forth. The features involved in consonant harmony, on the other hand, tend to be relevant only for segments occupying small subregions of the consonant space. For example, it is only dorsals that can be either velar or uvular (the basis of harmony in a small handful of languages; see Hansson 2001); similarly, only coronal segments can be either [+anterior] or [–anterior], [+distributed] or [–distributed], and so forth (at least in most versions of distinctive feature theory). For this reason, a great number of segments in any given consonant harmony system are ones for which the [+F] vs. [–F] categorization simply does not apply, and which are therefore, by definition, neutral segments.

9 Note that this is a sense of the term "recoverability" which is different from that used in works concerned with perceptual cues and their role in phonology (e.g., Silverman 1997).

Secondly, as mentioned earlier, a definitive hallmark of consonant harmony processes is that relative trigger/target similarity plays an extremely important role (Walker 2000b; Hansson 2001; Rose and Walker 2004), which further shrinks the set of segments participating in the harmony. For example, non-sibilant coronals like /t/ or /n/ appear to be neutral in all sibilant harmony systems, laryngeal harmony is frequently limited to obstruents which are homorganic, and so forth. In effect, then, consonant harmony is nearly always *parasitic* on features other than the harmonizing one, whereas this seems somewhat less typical of vowel harmony systems.[10] (This is perhaps in part an illusion; it might be that pairs vowels which are highly distinct, like [y] vs. [a], should nevertheless count as being far more similar to one another than a consonant pair like, say, [kʰ] vs. [r], merely by virtue of both being vowels.)

Finally, neutral vowels, when they are present at all in a system, tend to be the odd man out: the lone exception among all the vowels. Common examples are /a/ in height harmony or tongue root harmony, and /i/ in palatal or rounding harmony. Compare this with consonant harmony where, for the reasons just mentioned, neutral consonants (those which neither trigger nor undergo assimilation) are usually in an overwhelming majority in the inventory, greatly outnumbering their non-neutral counterparts.

Recall that the missing vowel harmony type is one where an underlying [±F] contrast in suffixes does exist, but is maintained only in forms containing a neutral-vowel root. Furthermore, those forms constitute the sole potential source of evidence available to the learner that such a lexical contrast exists in the first place. As it turns out, the consonant/vowel asymmetries just outlined lead to a severe reduction in the extent to which such crucial evidence is readily available in vowel harmony systems as compared to consonant harmony systems. To see why this is so, I ask the reader to consider, as a thought experiment, a hypothetical language displaying both [±ATR] vowel harmony and [±anterior] sibilant harmony, where the facts in (17) hold true.

10 Strictly speaking, the proper comparison should therefore be between consonant harmony and *parasitic* vowel harmony in particular. Obviously, the cross-linguistic absence of the neutralization pattern in § 2.1.3 holds true *a fortiori* for that particular subset of vowel harmony systems, and one may ask why this should be so. This might suggest that it is the general applicability of [±F], rather than its redundancy, that is the crucial factor. Neutral /i, e/ in Finnish palatal vowel harmony are phonetically [–back], and do require a [–back] suffix vowel as in (5c), while in a sibilant harmony system, neutral non-coronals such as /k/ or /m/ are simply neither [+anterior] ([or [+distributed]) nor [–anterior] (or [–distributed]) and hence cannot impose either feature value on other segments.

(17) Contrast recoverability under harmony: a thought experiment
a. The segment inventory of language *L* consists of 7 vowels and 28 consonants.
b. 6 of the 7 vowels form three [±ATR] pairs (e. g., /u/ : /ʊ/); unpaired /a/ is neutral and cooccurs freely with either kind of vowel.
c. 4 of the 28 consonants are sibilant coronals, forming two [±anterior] pairs (e. g., /s/ : /ʃ/); the rest are neutral (non-coronals and non-sibilant coronals) and cooccur freely with either kind of sibilant.
d. All vowels have the exact same frequency of occurrence, as do all consonants; each root is a CV syllable (exactly one C and one V).

Given these facts (admittedly somewhat unrealistic in their simplicity), the probability that a given affix will find itself in a *neutralizing environment* – that is, cooccurring with a non-neutral root – is 6 to 1 (86 %) for the ATR harmony, whereas it is only 1 to 6 (4 to 24, i. e. 14 %) for the sibilant harmony. In other words, an affixal [±ATR] vowel contrast will manifest itself as such only very rarely in surface forms (14 % of the time, to be precise), whereas an affixal [±ant] sibilant contrast will surface intact in the vast majority of surface forms (86 % of the time). Even if we make assumption (17d) much less artificial and drastically expand the template of possible root shapes to C(C)V((C)C) – such that any one of up to four consonants could potentially be a harmony-inducing sibilant – it is still the case that around 50 % of all conceivable roots will be neutral (sibilant-free). In other words, an affix sibilant would still find itself in a non-neutralizing environment about half the time, whereas for affix vowels the same is true only about 14 % of the time.

Lexical contrasts in vowel harmony systems are thus far less easily recovered, and hence harder for successive generations of learners to discover and internalize, than are corresponding contrasts in consonant harmony systems. Due to the paucity of surface evidence that some affix vowels are underlyingly [+F] whereas others are [–F], one would expect such contrasts to show a very strong tendency to disappear over time. In consonant harmony systems with the same properties, lexical contrasts in affix segments will be much more easily recoverable (and learnable), and these are therefore predicted to be far less vulnerable to loss over time. In sum, a vowel harmony system of the relevant (4c) type, were it to exist, would be expected to be diachronically unstable, showing a strong tendency to shift toward the ubiquitous (4a) type.

From this I suggest the following conjecture. The observed (synchronic) asymmetry in the cross-linguistic typology of harmony systems is nothing more than a reflection of this *diachronic* asymmetry between consonant and vowel harmony. The explanation for the absence of vowel harmony systems of the relevant type should thus not be sought in the synchronic design principles of grammar, for example by modifying our theory of Universal Grammar so as to circumscribe the range of possible languages to exclude systems of this kind. Instead, the typological gap is better seen as a product of the diachronic trajectories of language change. These trajectories are in turn defined and shaped by the (admittedly synchronic) learnability factors which influence language transmission across generations – or, rather, which influence individual learners' success in replicating the grammars of the speakers providing the ambient input data.

4. The importance of the missing neutralization type

The typological asymmetry discussed in § 2 and § 3 may seem like a rather trivial issue of no particular consequence for the phonological analysis of harmony. As it turns out, however, the "missing" harmony type, which is attested for consonant harmony systems but not for vowel harmony, has serious implications for questions of considerable theoretical importance.

First of all, it should be re-emphasized that in systems like those described for Navajo and Ineseño Chumash in § 2.2.3 (as well as the non-existing Pseudo-Finnish system laid out in § 2.1.3), harmony must be viewed as a genuinely *feature-changing* process. In Navajo, for example, the affixal sibilants which are targeted by harmony demonstrably contrast underlyingly for [±anterior]. It is therefore absolutely clear that the harmony has the power to change not only input [+ant] to output [–ant] but also to change input [–ant] to output [+ant] (cf. Navajo /si-ɣiʃ/ → [ʃi-ɣiʃ] 'it is bent, curved' and /ʃi-tseʔ/ → [si-tseʔ] 'my rock'). Note that this fact is entirely independent of how one chooses to construe the /s/ : /ʃ/ contrast representationally. For the sake of the argument, let us assume that rather than binary [–ant] vs. [+ant], we instead view /ʃ/ as being distinguished from /s/ by the presence of some monovalent feature [F] (e. g., [posterior]). Navajo sibilant harmony must then have the power not only to add (or spread) this feature [F] to the sibilant of the possessive /si-/ prefix (whenever the following root contains an [F]-carrying sibilant), but also to remove that same feature from the sibilant of the /ʃi-/ aspect prefix (whenever the following root contains a sibilant not carrying [F]).[11] The same reasoning applies if

11 As noted above, Poser (1982) and Lieber (1987: 145–150) capture the feature-changing character of sibilant harmony processes like those of Chumash and Navajo by decomposing them into a delinking rule and a (feature-filling) harmony rule. Avery and Rice (1989: 194), who represent [±anterior] contrasts with privative [posterior],

/ʃ/ and /s/ are considered to be distinguished by two monovalent and mutually incompatible features [F] and [G] (roughly corresponding to [–ant] vs. [+ant], similar to the common use of privative [ATR] and [RTR] in the analysis of tongue-root vowel harmony systems). Before roots containing sibilants specified as [F], harmony has the effect of removing [G] from (and adding/spreading [F] to) the sibilant of the /si-/ prefix. And before roots containing sibilants specified as [G], harmony must likewise be capable of removing [F] from (and adding/spreading [G] to) the sibilant of the /ʃi-/ prefix. Before roots containing no sibilant at all, the underlying featural specifications of the prefix sibilants, be they [+F]/[–F], [F]/Ø or [F]/[G], surface intact and unchanged.

For this reason, it is absolutely impossible in principle to recast the type of harmony found in Navajo and Chumash as being in any way strictly *feature-filling*, as has frequently been done in autosegmental analyses of vowel harmony. By contrast, the sibilant harmony in Koyra (see § 2.2.1), just like the analogous and ubiquitous vowel harmony systems described in § 2.1.1, can easily be interpreted in feature-filling terms. Since there is no evidence for an underlying /ʃ/ : /s/ contrast among affix sibilants in Koyra, it is quite possible to view these as being underlyingly unspecified for the relevant feature(s). Koyra sibilant harmony might then be interpreted as involving only [–ant] (or "[F]" in either of the alternative privative analyses outlined in the previous paragraph). In other words, after roots containing a [–ant] sibilant, a suffix sibilant in Koyra takes on the [–ant] (or [F]) specification of this root sibilant. In all other contexts, including after roots which happen to contain a [+ant] sibilant, that same suffix sibilant simply gets specified by default as [+ant] (or [G], or simply left unspecified for privative [F]).

---

view Chumash sibilant harmony as fusion of [coronal] nodes, where "fusion is right-headed, so the features of the rightmost sibilant remain". However, their claim that on this analysis "sibilant harmony is not feature-changing" is puzzling (perhaps reflecting an excessively narrow technical sense of the term "feature-changing"). In sequences like /ʃ…s/ → [s…s], the (right-headed) fusion operation must somehow involve delinking or deletion of the first sibilant's [posterior] specification. In their more detailed treatment of Ponapean velarization agreement along the same lines, Avery and Rice are more explicit in suggesting that deletion/delinking is indeed implicated: "[t]he result of the fusion is that *only secondary features of the righthand segment, the head, are maintained* [emphasis added]" (Avery and Rice 1989: 182). For this reason, it is hard to see how the term "feature-changing" is any less descriptive of their node-fusion analysis of sibilant harmony in Chumash (or Navajo) than it is of the delinking-plus-feature-filling analyses proposed by Poser (1982) and Lieber (1987).

The inherently feature-changing character of the consonant harmony systems of languages like Navajo, Tahltan (Shaw 1991) and the Chumashan languages has profound and devastating consequences for unification-based approaches like Declarative Phonology (Scobbie 1991; Russell 1993; Bird 1995; Coleman 1998). In such models, phonology is construed as *monotonic*, such that any kind of destructive effects that remove or alter lexically specified information are disallowed in principle. Not surprisingly, proponents of declarative approaches to phonology have attempted to explain away Chumash sibilant harmony as a mere "phonetic process" outside the realm of the phonological grammar (Russell 1993; Bird 1995). See Poser (to appear) for a host of counterarguments against such an interpretation, most of which apply at least as strongly to Navajo and Tahltan as well.

A second and more subtle problem concerns absolute directionality and its analysis in output-oriented constraint-based frameworks like Optimality Theory. As it turns out, a subset of the languages with actually-neutralizing consonant harmony also obey fixed regressive directionality (see Hansson 2001 for a survey of directionality patterns in consonant harmony systems). For example, sibilant harmony in Ineseño Chumash (Applegate 1972; Poser 1982; Lieber 1987: 145–150) proceeds from right to left, with no regard whatsoever for morphological constituency or prosodic structure. The sibilant which happens to be the rightmost one in the word simply determines the [±ant] value of any and all preceding sibilants, as illustrated in (18).

(18) Right-to-left sibilant harmony in Ineseño (Applegate 1972)

a. /s-apitʃ^{h}o-it/	[ʃapitʃ^{h}olit]	'I have a stroke of good luck'
b. /s-apitʃ^{h}o-us/	[sapits^{h}olus]	'he has a stroke of good luck'
c. /s-apitʃ^{h}o-us-waʃ/	[ʃapitʃ^{h}oluʃwaʃ]	'he had a stroke of good luck'

Note that here, just as in Navajo, harmony is symmetrically feature-changing (triggering both [+ant] → [–ant] and [–ant] → [+ant] as unfaithful input-output mappings), as well as being actually-neutralizing. As I demonstrate elsewhere (Hansson 2001, in prep.), the specific combination of symmetric neutralization with absolute directionality of assimilation creates severe and unexpected problems for output-oriented approaches to phonology, and is in fact impossible to handle in standard versions of Optimality Theory.

For reasons of space this complex issue can only be touched on briefly here. The core of the problem is that the kinds of output well-formedness constraints which are ultimately responsible for driving harmony – whether these be construed as Agree[F], Align[F], Spread[F], or something else entirely – cannot in and of themselves guarantee that the manner in which harmony is achieved will adhere to a particular directionality of assimilation. This is illustrated by the tableau in (19). Here the intended derivation is /Cɛ-CuC/ → [Ce-CuC], with regressive [+ATR] harmony (similar to the Karajá pattern in § 2.1.2); in the constraint labels, "[+A]" stands for [+ATR] (or, equivalently, privative [ATR]). A constraint like ALIGN-L[+ATR], for example, is defined as requiring that any [+ATR] autosegment occurring in the output be aligned with the left edge of the word.

(19)

UR: /Cɛ-CuC/	ALIGN-L[+A]	SPREAD-L[+A]	AGREE[±A]	IDENT[±A]
a. C ɛ C u C | | [-A] [+A]	*!	*!	*!	
☞ b. C e C u C \ / [+A]				*
☞ c. C ɛ C ʊ C \ / [-A]				*

Note that even though a right-to-left orientation has essentially been built into the ALIGN-L[+ATR] and SPREAD-L[+ATR] constraints, this does nothing to help rule out the left-to-right spreading alternative in (19c). That candidate satisfies such constraints *vacuously*, by not containing any output [+ATR] element at all. The responsibility for preferring (19b) over (19c) must obviously fall to other constraints. Baković (2000) suggests that these may be of two kinds, each giving rise to its own distinctive pattern. Output-output correspondence to the stem of affixation (e. g., IDENT[±ATR]-SA), when ranked sufficiently high, will result in *stem control* or "cyclic" harmony, an extremely common pattern (see Ringen and Vago 1998 for a variation on this idea, using positional input-output faithfulness to root vowels). Alternatively, a Markedness or Faithfulness constraint favouring one [F]-value over the other (*[–ATR] or IDENT[+ATR]-IO, or a local conjunction of the two) will guarantee that the directionality goes from vowels with the favoured value to vowels with the disfavoured one. The resulting pattern is a typical *dominant-recessive* harmony. Either strategy would suffice to select (19b) over (19c), assuming for simplicity that, in our hypothetical example, /CuC/ is the stem and /Cɛ-/ a prefix.

However, both strategies break down when combined simultaneously with both (i) absolute directionality and (ii) actually-neutralizing harmony. This is exactly what we find in the sibilant harmony of Ineseño Chumash in (18) above.

Here we need to ensure not only that /…s…ʃ…/ → […ʃ…ʃ…], but also that /…ʃ…s…/ → […s…s…]. Stem control can obviously not be appealed to, since harmony may go from an affix (suffix) sibilant to a root sibilant just as easily as from a root sibilant to an affix (prefix) sibilant. However, a dominant-recessive analysis fails as well, since harmony alternately favours [+ant] over [–ant] and [–ant] over [+ant], depending simply on which type of sibilant happens to follow the other in the linear sequence. Neither feature value can be designated as the dominant or "active" one in the operation of this harmony system.

In fact, the problem of enforcing absolute directionality of this kind appears to be intractable in standard Optimality Theory. I have argued elsewhere (Hansson 2001, in prep.) that the only viable solution within an Optimality Theory architecture appears to be to formalize the harmony-driving constraint as a *targeted constraint* (Wilson 2001). Such constraints differ from conventional Markedness constraints in that they circumscribe the range of possible repairs for the offending structure. Most importantly, a targeted constraint of the type *[–αF] / __ [αF], while seemingly equivalent to a standard agreement constraint like *[–αF][αF] or Agree[F], differs from the latter in that it favours only those candidates which have repaired the targeted marked element as such (here, the [–αF] segment on the left), not ones which involve modification of the surrounding context (here, the [αF] segment on the right). In other words, such a constraint will prefer the regressive-assimilation candidate [αF]…[αF] over unassimilated *[–αF]…[αF], without simultaneously (and equally) preferring the progressive-assimilation alternative [–αF]…[–αF] as a conventional (non-targeted) agreement constraint would. Directionality ties like that shown in (19) are thus broken in a consistent manner that is independent of the feature values involved, the morphological or prosodic affiliation of the interacting segments, or any other conceivable factors beyond the linear precedence relation itself. In the above example, regressive assimilation will be ensured both for cases of the [–αF]…[αF] type (→ [αF]…[αF]) and for ones of the [αF]…[–αF] type (→ [–αF]…[–αF]).

This is shown in tableaux (20)–(21), which render schematically the regressive [±anterior] sibilant harmony observed in Inseseño Chumash. Because targeted constraints do not impose a total ordering on the entire candidate set, but rather a partial ordering – involving only those candidate pairs which differ in terms of the specified repair to the targeted marked structure – the format of tableaux is necessarily slightly unorthodox. In place of asterisks, each tableau cell lists which (other) candidates, if any, a constraint deems to be more harmonic than the candidate under consideration. Parentheses indicate harmonic orderings of this kind (i.e. preferences) which are cancelled out by conflicting harmonic orderings assigned by a higher-ranked constraint. The bottom row displays how a total ordering over the full candidate set is gradually built up, going from higher-ranked to lower-ranked constraints, until one candidate emerges as most harmonic. Note in particular that no individual constraint directly prefers the regressive-assimilation candidate over its progressive-assimilation competitor. Rather, that preference emerges by transitivity: regressive assimilation (20b)/(21c) beats no assimilation (20a)/(21a) on *[–αant] / __ [αant], the targeted constraint, whereas the latter beats progressive assimilation (20c)/(21b) on simple Faithfulness to input [±ant] values.

(20)

/ʃ…s/	*[–αant] / __ [αant]	IDENT[±ant]-IO
a. ʃ…s	s…s ≻ ʃ…s!	
☞ b. s…s		(ʃ…s ≻ s…s)
c. ʃ…ʃ		ʃ…s ≻ ʃ…ʃ!
cumulative ordering	s…s ≻ ʃ…s	☞s…s ≻ ʃ…s ≻ ʃ…ʃ

(21)

/s…ʃ/	*[–αant]/__[αant]	IDENT[±ant]-IO
a. s…ʃ	ʃ…ʃ ≻ s…ʃ!	
b. s…s		s…ʃ ≻ s…s!
☞ c. ʃ…ʃ		(s…ʃ ≻ ʃ…ʃ)
cumulative ordering	ʃ…ʃ ≻ s…ʃ	☞ʃ…ʃ ≻ s…ʃ ≻ s…s

That the fundamental problem of accounting for absolute directionality in output-oriented frameworks has not previously been noted is hardly surprising. The problem can arise only when the harmony system in question displays precisely the kind of marginal contrast preservation (symmetric neutralization) defined in (4c). As we have seen, harmony of this type is entirely unattested among vowel harmony systems. The specific combination of marginal contrast preservation with absolute directionality of assimilation is found only in a small subset of consonant harmony systems.

5. Summary

We have seen how a close examination of certain aspects of the cross-linguistic typology of harmony systems reveals an asymmetry with respect to the neutralization patterns caused by harmony. Neutralization of actual lexical contrasts (in affixes) appears to be unattested in vowel harmony systems – or at least in those systems where both feature values are active in the harmony – whereas that same kind of neutralization does occur in a number of consonant harmony systems.

The central claim made here has been that this asymmetry falls out from considerations of contrast recoverability. It was demonstrated how the surface evidence needed for reliably establishing the existence of lexical contrasts (in positions targeted by harmony) is necessarily quite limited in a typical vowel harmony system, far more so than in a typical consonant harmony system. Owing to these learnability factors, such contrasts therefore have a very high likelihood of disappearing over time in vowel harmony systems, while that likelihood is much smaller for consonant harmony systems.

Finally, the existence of actually-neutralizing harmony of this kind, attested in consonant but not vowel harmony, was shown to have profound implications for the analysis of harmony within output-oriented models like Optimality Theory, as well as for unification-based approaches to phonology.

References

Applegate, Richard B.
1972 Ineseño Chumash grammar. Doctoral dissertation, University of California, Berkeley.
Archangeli, Diana and Douglas Pulleyblank
1994 *Grounded Phonology.* Cambridge, MA: MIT Press.
Avery, Peter and Keren Rice
1989 Segment structure and coronal underspecification. *Phonology* 6: 179–200.
Baković, Eric
2000 Harmony, dominance and control. Doctoral dissertation, Rutgers University.
Benua, Laura
2000 *Phonological Relations Between Words.* New York: Garland.
Bird, Steven
1995 *Computational Phonology: A Constraint-Based Approach.* Cambridge: Cambridge University Press.
Coleman, John
1998 *Phonological Representations: Their Names, Forms and Powers.* Cambridge: Cambridge University Press.
Dettweiler, Stephen H.
2000 Vowel harmony and neutral vowels in C'Lela. *Journal of West African Languages* 28: 3–18.
Hansson, Gunnar Ólafur
2001 Theoretical and typological issues in consonant harmony. Doctoral dissertation, University of California, Berkeley.

Hansson, Gunnar Ólafur
In prep. Absolute directionality in output-oriented phonology. Ms., University of British Columbia.

Hayward, Richard J.
1982 Notes on the Koyra language. *Afrika und Übersee* 65: 211–268.

Hyman, Larry M.
1995 Nasal consonant harmony at a distance: The case of Yaka. *Studies in African Linguistics* 24: 5–30.

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

Kaun, Abigail R.
1995 The typology of rounding harmony: An optimality theoretic approach. Doctoral dissertation, University of California, Los Angeles.

Kaye, Jonathan D.
1974 Opacity and recoverability in phonology. *Canadian Journal of Linguistics* 19: 134–149.

Kaye, Jonathan D.
1989 *Phonology: A Cognitive View*. Hillsdale, NJ: Lawrence Erlbaum.

Kenstowicz, Michael
1997 Base identity and uniform exponence: Alternatives to cyclicity. In: Jacques Durand and Bernard Laks (eds.), *Current Trends in Phonology: Models and Methods*, 363–394. Salford: University of Salford.

Lieber, Rochelle
1987 *An Integrated Theory of Autosegmental Processes*. Albany, NY: State University of New York Press.

McCarthy, John J.
2002 *A Thematic Guide to Optimality Theory*. Cambridge: Cambridge University Press.

McCarthy, John J. (ed.)
2003 *Optimality Theory in Phonology: A Reader*. Oxford: Blackwell.

McCarthy, John J.
2005 Optimal paradigms. In: Laura J. Downing, T. Alan Hall and Renate Raffelsiefen (eds.), *Paradigms in Phonological Theory*, 170–210. Oxford: Oxford University Press.

Perkins, Jeremy
2005 The RTR harmonic domain in two dialects of Yoruba. M. A. thesis, University of British Columbia.

Poser, William
1982 Phonological representations and action-at-a-distance. In: Harry van der Hulst and Norval Smith (eds.), *The Structure of Phonological Representations*, Vol. 2, 121–158. Dordrecht: Foris.

Poser, William
to appear On the status of Chumash sibilant harmony. *Southwest Journal of Linguistics*.

Prince, Alan and Paul Smolensky
2004 *Optimality Theory: Constraint Interaction in Generative Grammar.* Oxford: Blackwell. First appeared in 1993 as Technical Report RuCCS-TR-2, New Brunswick, NJ: Rutgers University Center for Cognitive Science.

Pulleyblank, Douglas
2002 Harmony drivers: No disagreement allowed. *Proceedings of BLS* 28: 249–267.

Ribeiro, Eduardo Rivail
2001 [ATR] vowel harmony and palatalization in Karajá. Ms., University of Chicago.

Ribeiro, Eduardo Rivail
2002 Directionality in vowel harmony: The case of Karajá (Macro-Jê). *Proceedings of BLS* 28: 475–485.

Ringen, Catherine O. and Robert M. Vago
1998 Hungarian vowel harmony in Optimality Theory. *Phonology* 15: 393–416.

Rose, Sharon and Rachel Walker
2004 A typology of consonant agreement as correspondence. *Language* 80: 475–531.

Russell, Kevin
1993 A constraint-based approach to phonology and morphology. Doctoral dissertation, University of Southern California.

Ruttenberg, Piet
1968 *Lexique yaka-français, français-yaka*. Kinshasa.

Sapir, Edward and Harry Hoijer
1967 *The Phonology and Morphology of the Navajo Language*. (University of California Publications in Linguistics 40.) Berkeley, CA: Univ. of California Press.

Scobbie, James M.
1991 Attribute-value phonology. Doctoral dissertation, University of Edinburgh.

Shaw, Patricia A.
1991 Consonant harmony systems: The special status of coronal harmony. In: Carole Paradis and Jean-François Prunet (eds.), *The Special Status of Coronals: Internal and External Evidence*, 125–157. (Phonetics and Phonology 2.) San Diego: Academic Press.

Silverman, Daniel
1997 *Phasing and Recoverability*. New York: Garland.

Steriade, Donca
2000 Paradigm uniformity and the phonetics-phonology boundary. In: Michael B. Broe and Janet B. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, 313–335. Cambridge: Cambridge University Press.

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*, 219–250. New York: Academic Press.

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

Trubetzkoy, N. S.
 1939 *Grundzüge der Phonologie*. (Travaux du Cercle Linguistique de Prague 7.) Prague.

Walker, Rachel
 2000a *Nasalization, Neutral Segments, and Opacity Effects*. New York: Garland.

Walker, Rachel
 2000b Long-distance consonantal identity effects. *Proceedings of WCCFL* 19: 532–545.

Wilson, Colin
 2001 Consonant cluster neutralisation and targeted constraints. *Phonology* 18: 147–197.

Perception

The impact of allophony versus contrast on speech perception[1]

Amanda Boomershine, Kathleen Currie Hall, Elizabeth Hume, and Keith Johnson

1. Introduction

The perceptual consequences of phonological contrast have long been of interest to phonologists and phoneticians. In Trubetzkoy's well-known *Grundzüge der Phonologie* (1939: 78), for example, he speculates that an opposition between speech sounds that is always contrastive in a given language will be perceived more clearly than an opposition that is neutralizable in some context. Furthermore, even within the category of neutralizable oppositions, he predicts that perception will fluctuate depending on factors such as context. There are three important assumptions that underlie Trubetzkoy's speculations. First, that one's native language experience influences the ability to perceive speech sounds. Second, that the phonological relation holding between sounds in a language has an impact on a listener's perception of those sounds. And third, that it is not simply the presence versus the absence of phonological contrast that is relevant to perceiving a sound. Rather, Trubetzkoy pinpoints different categories, or degrees, of contrast and suggests that each may have a particular consequence for speech perception.

Trubetzkoy's first and second assumptions, that one's native language experience – particularly the phonological relations between sounds – influences the ability to perceive speech sounds, are now well established in the literature. For example, studies in second language learning have found that listeners are more adept at perceiving sounds of their native language than those of a second language acquired later in life, e. g., Polka and Werker (1994), Strange (1995), Dupoux et al. (1997), Best et al (1998), Francis and Nusbaum (2002). Familiar illustrations include the perception of English /l/ and /r/ by Japanese listeners

1 The authors would like to acknowledge the assistance of Mary Beckman, Lauren Collister, Jim Harmon, and Terrell Morgan; members of The Ohio State University Phonies group and Department of Spanish and Portuguese; audiences at the 2004 Mid-Continental Workshop on Phonology, the 2005 Montreal-Ottawa-Toronto Phonology Workshop, and the 2005 OSU Hispanic Linguistics Colloquium; funding from the OSU Department of Linguistics and NIH grant number R01 DC004421; and of course all of our participants.

and that of Hindi dental and retroflex stops by American English listeners. Since the liquids /l/ and /r/ are non-contrastive in Japanese, Japanese listeners have difficulty distinguishing between them, even though they are fully contrastive in English (Goto 1971; MacKain et al. 1981). For similar reasons, perceiving a distinction between the Hindi stops is more challenging for English speakers than it is for Hindi speakers (Werker et al. 1981; Pruitt et al. 1998). The conclusion that can be drawn from these and other studies is that while listeners have little difficulty distinguishing between contrastive native sounds, they are less successful when it comes to non-native sounds that do not serve a contrastive function in their own language.

Less is known, however, concerning Trubetzkoy's third assumption, especially as it relates to the potential impact of phonological relations other than contrast on speech perception. It is this last point that we are especially concerned with in this paper.

As noted above, it is well established that while listeners have no difficulty distinguishing between native sounds that are contrastive, they are less successful when it comes to sounds that do not occur in their own language. Furthermore, there is evidence suggesting that it is not simply the presence versus the absence of phonemic contrast that is relevant to perceiving a sound. Partial contrast, where an otherwise contrastive pair of elements is neutralized in some context, has also been shown to influence perception, as Trubetzkoy predicted. For example, drawing on perception data on Mandarin tone (Huang 2001), Hume and Johnson (2003) conclude that not only is perceptual distinctiveness a function of phonological contrast, but that *partial* contrast reduces perceptual distinctiveness for native listeners. Thus, contrast seems to be more nuanced than is often assumed in the speech perception literature.

This finding then raises the question as to whether other phonological relations also shape perception. Consider non-contrastiveness. As noted above, listeners typically have greater difficulty distinguishing between sounds that do not occur in their own language, and are thus non-contrastive, than they do with native sounds that are contrastive. In addition to this typical notion of non-contrastiveness, sounds that do in fact co-occur within a single language can also be in a non-contrastive relation, such as when they are allophones of the same phoneme. While two sounds with an allophonic distribution both occur in a speaker's phonetic inventory, they never effect a change in meaning. In English, for example, the phones [d] and [ɾ] can be considered allophones of a single phoneme, /d/, with [ɾ] occurring intervocalically when the first vowel is stressed, e. g. [ráyɾiŋ] "riding," and [d] occurring elsewhere, e. g. [rayd] "ride." Crucially, however, substituting [d] for [ɾ] in "riding" has no effect on the meaning of the word.

Given the lack of contrast between a pair of allophones, we would expect them to be perceived as less distinct than a pair of contrastive sounds, all else being equal. Theories of speech perception generally predict this result (see, e. g. Lahiri 1999; Gaskell and Marslen-Wilson 2001), although the means by which they do so vary in their predictions for the perception of other pairs of sounds in the language. There is also some experimental support for the idea that allophony plays a role in speech perception (e. g., Dupoux et al. 1997; Harnsberger 2001; Johnson 2004), though its precise influence on perception has not been directly tested. For example, Harnsberger's (2001) results from an AXB classification task point to a near merger in the perception, by Malayalam listeners, of allophonically-related dental and alveolar nasal consonants. These coronal nasals are in complementary distribution in the language, with the dental occurring morpheme-initially and the alveolar occurring both morpheme-finally and intervocalically (Mohanan and Mohanan 1984). Contrastive nasals such as bilabial [m] versus velar [ŋ], on the other hand, showed greater perceptual separation in Harnsberger's study. Findings such as these suggest that the simple presence of a sound in an inventory is not the only source of information concerning the relative perception of that sound. The sound's phonological relatedness to other sounds in the inventory must also be taken into consideration.

This paper explores the impact of contrast versus allophony on the perception of speech sounds in a series of four experiments contrasting the behavior of Spanish-speaking and English-speaking listeners, and considers how these empirical results should be integrated into a theory of speech perception. In addition to the basic finding that models of speech perception are in fact correct in their prediction that phonemic contrasts are more perceptually distinct than allophonic contrasts, the results of experiments like the ones presented here can be used to differentiate models of speech perception based on the mechanisms by which this more basic finding is predicted in the different models, as will be discussed. In section 6, we consider the effectiveness of two different models in accounting for the results: a phonological inferencing model (e. g. Gaskell and Marslen-Wilson 1998) and an exemplar model (e. g. Goldinger 1992, 1996; Palmeri et al. 1993; Johnson 1997a, b, 2004; Coleman 2002; Pierrehumbert 2003; Hawkins 2003). To anticipate our conclusion, both models are successful in predicting our findings relating to allophony versus phonemic contrast. Only the exemplar model, however, is able to account for the full range of results obtained in this study.

The experiments presented in this paper make use of the fact that English and Spanish place similar sounds, namely [d], [ð], and [ɾ], in very different positions in the linguistic system of contrasts. As illustrated in (1), the phones

[d] and [ɾ] are allophones of a single phoneme in English while [d] and [ð] are contrastive ([do] *dough* versus [ðo] *though*). Conversely in Spanish, [d] and [ð] are allophones of a single phoneme (*de* [ð]*onde* 'from where', [d]*onde* 'where'), while [d] and [ɾ] are separate phonemes. Note, however, that [d] and [ɾ] are never lexically contrastive in Spanish since the sounds do not appear in the same context: [ɾ] occurs in medial position and [d] in initial position.

(1) Phonological grouping of [ð], [d], and [ɾ] in English and Spanish. Sounds within parentheses pattern as allophones of a single phoneme, and are contrastive with sounds outside parentheses.

English	[ð]	([d]	[ɾ])
Spanish	([ð]	[d])	[ɾ]

In general, we expect that when sounds are contrastive in a language, listeners will be more attuned to the phonetic contrast between these sounds and thus judge them to be more different from each other than sounds that are in a non-contrastive relationship within a given language.

While the pairs [ð]/[d] and [d]/[ɾ] display different phonological relations in Spanish and English, the pair [ɾ]/[ð] patterns similarly in terms of phonological representation. In each language, these sounds are associated with different phonemes, but one sound of the pair is in an allophonic relationship with a different sound that is also present in the inventory of the language, as shown in (2).

(2) Surface and phonemic correspondences of [ð] and [ɾ] in Spanish and English
(a) Spanish: surface contrast [ð] – [ɾ] corresponds to phonemic contrast /d/ – /ɾ/
(b) English: surface contrast [ð] – [ɾ] corresponds to phonemic contrast /ð/ – /d/

The patterning of the two sounds [ɾ]/[ð] is also similar in that in both languages the distinction between the phones signals lexical, or surface, distinctions, as (3) illustrates.

(3) Surface contrast of [ɾ] and [ð]
(a)English

	[lɛðr̩] leather	[lɛɾr̩] letter
	[mʌðr̩] mother	[mʌɾr̩] mutter
(b)	Spanish	
	[kaða] cada 'each'	[kaɾa] cara 'face'

To summarize, the phonological relations of each of the three pairs of sounds are given in (4). The first pair, [d]/[ɾ], is contrastive in Spanish and allophonic in English. In neither language does this pair display a surface contrast. The pair [d]/[ð], on the other hand, displays contrast at the phonemic and surface levels in English, while in Spanish it is allophonic and thus contrasts on neither level. Finally, the phonological relations of the pair [ɾ]/[ð] are the same in both languages, being contrastive both at surface and phonemic levels.

(4) Summary of phonological relations among [d], [ð], and [ɾ] in English and Spanish

Pair:	[d] – [ɾ]		[d] – [ð]		[ɾ] – [ð]	
Language:	English	Spanish	English	Spanish	English	Spanish
phonemic (underlying) contrast	-	+	+	-	+	+
surface contrast	-	-	+	-	+	+

Given the similar patterning of the latter pair across the two languages, we would expect the perceived difference between intervocalic [ð] and [ɾ] to be about the same for both Spanish and English listeners. On the other hand, given the allophonic/contrastive differences with the remaining two pairs, we would expect the pairs to pattern differently in the two languages. Specifically, contrastive pairs should show greater perceptual separation than the allophonic pairs.

To explore the perception of the contrastive and allophonic relations among [d, ɾ, ð] in the two languages, we used two experimental paradigms, intending to differentiate processing that might emphasize surface contrast from processing at a more phonemic level. To capture phonological processing, listeners were asked to rate the perceived difference between the sounds, forcing them to categorize each sound and then compare it to a second categorized sound. To capture surface phonetic processing, listeners were asked to make speeded AX discrimination judgments; such tasks are generally assumed in the literature to access a more purely auditory level of discriminability (see, e. g., Fox 1984; Strange and Dittman 1984; Werker and Logan 1985). Because the pattern of contrasts at surface and phonemic levels differs for the [d]/[ɾ] comparison, we expected that if one task taps surface contrast effects while the other taps phonemic contrast then we might see differing patterns of response with the two paradigms. It will be seen in the following sections, however, that these predictions regarding paradigm differences were not borne out.

1.1. Structure of the paper

Section 2 describes an experiment in which Spanish-speaking and English-speaking listeners were asked to rate the perceived similarities of pairs of non-identical stimuli: [d]/[ɾ], [d]/[ð], and [ɾ]/[ð]. Because the phonologies of Spanish and English group these sounds differently (see (1) above) and because the rating task is an off-line judgment task, we expected to see a strong effect of native language background on the listeners' similarity ratings. Section 3 presents results from a speeded discrimination study using the same stimuli that were used in experiment 1. We expected to find in this experiment a much smaller effect of native language on perceptual distance because the speeded discrimination task is a much more on-line task which may tap earlier "phonetic" processing (Werker and Logan 1985). Surprisingly, Spanish-speaking and English-speaking listeners differed in this experiment just as they differed in the rating task using these stimuli. Sections 4 and 5 present rating and speeded discrimination experiments that are identical to experiments 1 and 2 in every regard, except that in these experiments the stimuli were produced by speakers of Greek, who in their native language make all of the contrasts tested in the experiments (whereas the speakers for experiments 1 and 2 were English-speaking linguists). Finally, the differences and similarities between the two sets of experiments, as well as the implications of the experiments for theories of speech perception, are presented in section 6.

2. Experiment 1: Rating [d], [ɾ], [ð] pairs

2.1. Methods

2.1.1. Stimuli

Materials consisted of two tokens of each of the following VCV sequences: [ada], [aɾa], [aða], [idi], [iɾi], [iði], [udu], [uɾu], and [uðu]. The tokens were produced by two American English speaking trained phoneticians, one male and one female. The speakers recorded multiple examples of the stimuli using a head-mounted microphone in a soundproof booth. The speakers attempted to produce equal stress on the first and second syllables. In order to control the amplitude across tokens and speakers, the peak amplitude was equated for each of the tokens. The two best recordings for each VCV sequence were used as stimuli in the studies. These materials were used as stimuli in both experiment 1 and experiment 2.

2.1.2. Participants

One group of native Spanish speakers and one group of native American English speakers participated in the experiment. The native Spanish speakers (N = 10, 3 men, 7 women) were students or friends of students at The Ohio State University, and were from a variety of Spanish-speaking countries, including Mexico, Colombia, Spain, Argentina, Puerto Rico, and Peru. They were paid a small sum for participating in the experiment. The native English speakers (N = 18, 8 men, 10 women) were undergraduate students at The Ohio State University enrolled in introductory linguistics courses who participated in the experiment for partial fulfillment of a course requirement. They were screened in a post-test questionnaire and only subjects who had no Spanish speaking experience were included in this experiment. The native English-speaking participants thus had a mean self-rating of their Spanish ability of 0 on a scale from 0–7, where a score of 7 is equivalent to native competency, and a score of 0 is equivalent to no experience in that language. The native Spanish-speaking participants had a mean self-rating of their English ability of 5 on a scale from 0–7. None of the speakers reported any history of speech or hearing disorders.

It should be noted that all of the native Spanish-speaking participants in the experiments reported here had an advanced level of English (i. e. they were bilingual). They were, however, run in a Spanish setting (the experimenter spoke to them in Spanish and the post-experiment questionnaire was presented in Spanish), so we believe that their English abilities had a minimal influence on their perception (see e. g. Marian and Spivey 2003 for a discussion of how the language of the experimental setting affects participant performance). We are currently running experiments on monolingual native Spanish speakers, and we expect to find that the monolingual Spanish speakers pattern very similarly to the Spanish speakers with a high degree of English. If anything, we expect that the inclusion of Spanish speakers with some knowledge of English in our experiments would bias the results against finding a difference between the perception of phonemic and allophonic pairs across languages; foreshadowing the results, the fact that such a difference was found is further indication that these Spanish speakers were operating in a Spanish mode.

Furthermore, while some of the native English-speaking participants did have knowledge of another foreign language (e. g. French, German, Japanese, etc.), none had familiarity with any language where the phones [d], [ð], and [ɾ] are in a fully contrastive relationship, such as Greek. Also, their mean self-rated ability in any foreign language was at a very low level, and such a superficial acquaintance with a second language does not seem to affect perception to any significant degree (see Boomershine et al. 2004).

2.1.3. Procedure

In this similarity rating task, participants were told that they would hear a pair of sounds and be asked to rate how similar those sounds were on a scale of 1–5, where 1 was 'very similar' and 5 was 'very different.' The participants were each seated at a computer that was connected to a 5-button response box, with up to four participants taking part in the study at a time. The participants listened to the stimuli through headphones, and then judged the similarity of the sounds using the button box. The pairs were presented in a different random order for each participant, using E-Prime software (v. 1.1; Psychological Software Tools, Pittsburgh, PA). The listeners heard pairs of stimuli, separated by one second of silence, such as [ada] <1 sec silence> [aɾa]. The talker and vowel context were the same for every pair so that the only difference in each pair was the consonant. The stimuli presented in each pair were always physically different tokens, even when they were both examples of a single sound (e. g. [ada] … [ada]). The participants were given four practice trials, and then the opportunity to ask questions before proceeding to the four test blocks (360 test trials total). They received no feedback in this experiment.

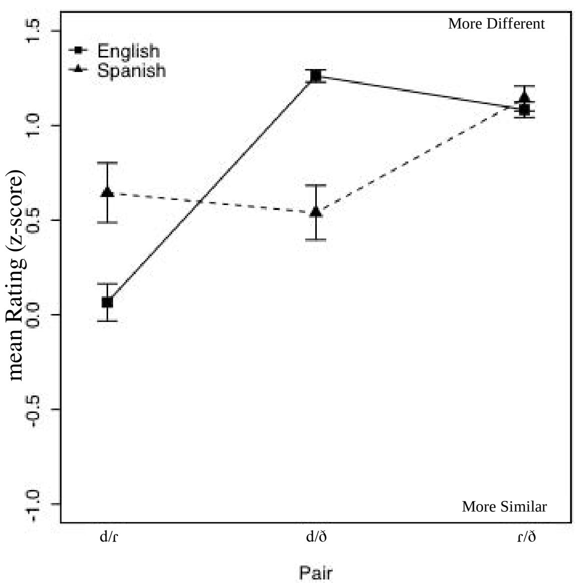


Figure 1. Results of experiment 1. Normalized similarity rating of [d], [ð], and [ɾ] by Spanish-speaking and English-speaking listeners.

2.2. Results

To analyze the rating task results, the rating scores for each speaker were normalized to compensate for differences in use of the 5-point scale (e. g. avoiding use of the endpoints, etc.). The scores were normalized using a standard z-score transformation, such that each participant's scores were centered around 0, with scores above zero indicating "more different" and scores below zero indicating "more similar." The normalized results with their 95 % confidence intervals are shown in Figure 1.

A repeated measures analysis of variance showed that there was a main effect of pair ($F[2, 52] = 31.621$, $p < 0.05$). That is, regardless of native language, the pairs were not all rated the same. There was also a significant pair by group interaction effect ($F[2,52] = 22.174$, $p < 0.05$), meaning that a participant's response to a given pair was dependent on the language group he was in. As shown in the figure, Spanish speakers found the pair [d]/[ɾ] (which is phonemically contrastive in Spanish but allophonic in English) more different than did the English speakers. Subsequent planned comparison independent samples t-tests showed that this difference was significant ($t(26) = 3.29$, $p < 0.05$). Furthermore, English speakers found the pair [d]/[ð] (which is phonemically contrastive in English but allophonic in Spanish) more different than did the Spanish speakers ($t(26) = 4.902$, $p < 0.05$). The pair [ɾ]/[ð], however, was rated the same by both Spanish and English speakers ($t < 1$); this pair is composed of allophones of different phonemes in each language.

2.3. Discussion

The results from experiment 1 provide strong evidence that allophonic relationships influence the perceived distance between sounds at a phonological level of processing. As expected from the fact that [d] and [ð] are allophones of the same phoneme in Spanish, but are separate phonemes in English, Spanish-speaking listeners rated pairs of stimuli contrasting [ð] and [d] as being much more similar sounding than did the American English listeners. Parallel to this, as expected from the fact that [d] and [ɾ] are in an allophonic relationship in English while phonemic in Spanish, English-speaking listeners rated [d]/[ɾ] pairs as being more similar than did Spanish-speaking listeners. There was no significant difference in the ratings by both groups of listeners of the pair [ɾ]/[ð], which are allophones of different phonemes, an expected result given the similarity in the phonological relations of the pair in the two languages.

The results also indicate that on average, listeners rated [d]/[ɾ] pairs as more similar to each other than the [d]/[ð] pairs, and we hypothesize that this is due to the raw auditory discriminability of these particular tokens. Experiment 3 returns to this question, but first we turn to experiment 2 which uses a "phonetic" listening task that might be sensitive to patterns of surface contrast.

3. Experiment 2: Discriminating [d], [ɾ], [ð] pairs

In experiment 1, it was found that the native language of a listener had a strong impact on the listener's judgments of phonetic sound similarity. Given that the similarity rating task invites the listener to use metalinguistic knowledge and ponder the sounds during each trial, it is perhaps not surprising that the language difference was observed. Experiment 2 tests the same contrasts, with speakers of Spanish and English again, but this time using a discrimination task that is intended to require much more "phonetic" or "psychoacoustic" listening, as Werker and Logan (1985) found. Because the patterns of contrast among [d], [ð], and [ɾ] in Spanish and English differ depending on whether we are focusing on surface phonetic contrast or on phonemic category-level contrast, we sought to test in this experiment whether the surface pattern of contrast would influence listeners' responses in a lower-level listening task.

It should be noted that there is some evidence that even in a speeded discrimination task, which should tap a much lower level of processing than similarity rating, listeners' responses are influenced by linguistic experience. Huang (2001, 2004) observed that Mandarin listeners responded with relatively longer reaction times in a speeded discrimination task (as compared with English-speaking listeners) when the sounds they were asked to discriminate were lexically related to each other. Specifically, the phonological neutralization of the dipping and rising tones of Mandarin resulted in longer reaction times for discriminations pairing these tones. English listeners did not show any effect of the Mandarin tone neutralization pattern. Interestingly, Huang found this effect of lexical/phonological contrast in a speeded discrimination task, which is generally assumed to be less prone to such language-specific effects. What Huang did not show is whether the linguistic experience reflected in her experiments relates to surface contrast or phonemic contrast. This experiment addresses this issue.

3.1. Methods

3.1.1. Stimuli

The stimuli that were used for experiment 1 were also used in this experiment.

3.1.2. Participants

The participants in this experiment were drawn from the same pool as those in experiment 1. The native Spanish speakers (N = 13, 3 men, 10 women) self-rated their ability in English at a mean value of 5.7; the native English speakers (N = 17, 3 men, 14 women) had no reported knowledge of Spanish. None of the speakers reported any history of speech or hearing disorders.

3.1.3. Procedure

In this discrimination task, the participants were told that they would hear a pair of sounds and be asked to judge whether the sounds were identical or different. "Identical" meant physically the same token (e. g. the same token of [ada] twice), while "different" meant either a different token of the same stimulus or two completely different stimuli (e. g. [ada] – [ada] where the two were not the same production, or [ada] – [aɾa], etc.). As with the rating task of experiment 1, the participants were seated at a computer connected to a 5-button response box, and the experiment was run using E-Prime software. The participants were asked to indicate whether each pair of sounds they heard was physically identical or different by pressing button 1 on the response box if they were the same tokens and button 5 if they were different tokens. Within each stimulus pair, the stimuli were separated by 100 ms of silence (a shorter interval than in the rating task, used to induce "phonetic" listening). Participants were given four practice trials before completing the three randomized test blocks (288 test trials). After responding to each stimulus pair, the participants were given feedback as to the accuracy of their response, their average percent correct overall, and their response time (ms). This feedback was used to encourage both heightened accuracy and shorter response times.

3.2. Results

The average results and 95 % confidence intervals for the "different" pairs from the discrimination task, shown in Figure 2, are very similar to those from the rating task. This figure shows normalized reaction times. Reaction time for these "different" pairs is taken to be a measure of perceptual distance, where slower reaction times indicate a smaller distance (see for example, Takane and Sergent, 1983); hence "more different" is at the bottom of the graph and "more similar" is at the top. As with the rating scores of experiment 1, we normalized the data in this experiment using a z-score transformation to correct for individual differences in overall reaction time. Consistent with the results from experiment 1, there was a main effect of pair ($F[2,56] = 22.162$, $p < 0.05$), indicating that some pairs were harder to discriminate than others, regardless of the native language of the listener.

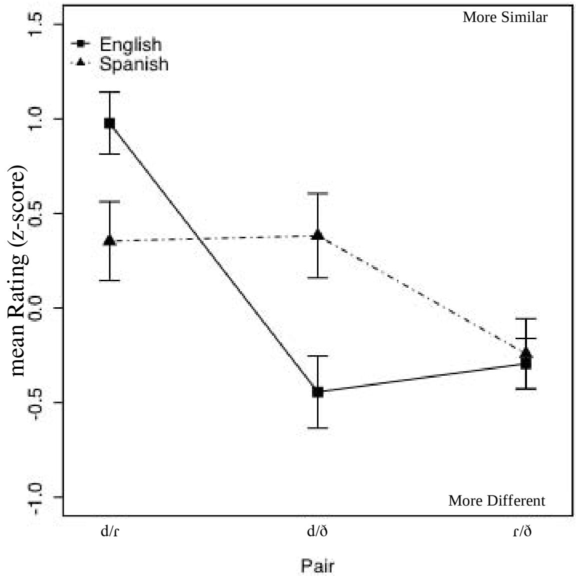


Figure 2. Results of experiment 2. Normalized reaction times for speeded discrimination [d], [ð], and [ɾ] by Spanish-speaking and English-speaking listeners.

There was also a significant pair by group interaction effect ($F[2, 56] = 3.876$, $p < 0.05$), indicating again that the pattern of pair reaction times differed depending on which group the listener was in – i. e., that native language influenced discrimination reaction time. Recall that slower reaction times are associated with more difficult discrimination and therefore with higher similarity. As predicted by the rating task, Spanish listeners were faster at discriminating the pair [d]/[ɾ], which is phonemic in Spanish, than were English listeners for whom [d]/[ɾ] are allophonically related. In subsequent planned comparison independent samples t-tests, this difference was found to be significant ($t(28) = 2.373$, $p < 0.05$), indicating that [d]/[ɾ] is perceived as less similar by the Spanish listeners. Not surprisingly, the English listeners were faster than the Spanish listeners at discriminating the pair [d]/[ð] ($t(28) = 2.823$, $p < 0.05$), given that these sounds have a phonemic relation in English but an allophonic one in Spanish. Finally, for the pair [ɾ]/[ð], the difference in reaction times of the two groups was not statistically significant ($t < 1$).

3.3. Discussion

The results from the discrimination task in experiment 2 are strikingly similar to those from the rating task in experiment 1. Again, there is strong evidence that allophony influences the perceived distance between sounds. As we found in the first experiment, a pair of sounds that is phonemic in one language (e. g. Spanish [d]/[ɾ]; English [d]/[ð]) was judged to be less similar than in the language where it is allophonic. Further, the native language of the listener did not impact the judgment of [ɾ]/[ð]. The pair [d]/[ɾ] is of particular interest here because it does not contrast on the surface in Spanish, just as it doesn't in English. We expected that this lack of surface contrast might make it pattern more like the English [d]/[ɾ] pair. However, even in this discrimination task, Spanish listeners found [d]/[ɾ] to be more different than English listeners did. For Spanish listeners, these two sounds are allophones of different phonemes, so evidently this more abstract level of contrast influences perception even in this on-line reaction-time experiment.

With respect to task, the results from experiment 2 support the findings of Huang (2001, 2004) where cross-linguistic speech perception differences were found using a discrimination task. As noted above, it is commonly assumed in the L2 perception literature that "phonetic" listening tasks, such as discrimination, may obscure cross-linguistic speech perception differences (Werker and Logan, 1985; Huang, 2001 and 2004). The observation that the phonological relations of the pairs in each language impacted the discrimination of the sounds in experiment 2 thus provides further evidence that language-specific influences that emerge in an off-line task can also be observed in an on-line task.

One concern regarding experiments 1 and 2 is that the stimuli were produced by English speakers (linguists trained to be able to produce IPA symbols, but native English speakers nonetheless), and we were comparing responses of English-speaking listeners with those of Spanish-speaking listeners. In a post-test questionnaire the majority of the Spanish-speaking listeners identified the stimuli as having been produced by English speakers, presumably because the coronals were pronounced with an alveolar place of articulation rather than with the dental place of articulation used in the pronunciation of coronals in Spanish. As a result, the stimuli may have been less natural for Spanish listeners than they were for English listeners. (Interestingly though, the majority of the native English speakers did not identify the stimuli as English.) To address this concern, we conducted two further experiments identical to the first two, except that the stimuli were produced by Greek speakers, as opposed to American English speakers. Discussion of these experiments follows.

4. Experiment 3: Rating Greek [d], [ɾ], [ð] pairs

Experiments 3 and 4 replicate experiments 1 and 2 in almost every detail. The listeners were drawn from the same populations and the tasks were the same as in the first two experiments. The only difference was that new speech tokens were used in experiments 3 and 4. We were interested to know whether the evidence for a role of phonemic contrast in speech perception could be replicated in an experiment with new stimuli.

An additional test inherent in these last two experiments has to do with two separable factors in speech perception. Experiments 3 and 4 manipulate one of these factors and hold the other constant, allowing us to examine the former's effect on speech perception. The first factor is the raw auditory/phonetic contrast between sounds. Thus, although [l] and [m], for example, are just as phonemically different from each other as are [p] and [m], we expect that listeners would rate the [p]/[m] contrast as more different than they would the [l]/[m] contrast because the auditory contrast between [p] and [m] is greater than that between [l] and [m]. The second factor is a language-specific mechanism of some sort that responds to speech in a way that is appropriate for, or trained by, the speech sounds and phonological patterns of a particular language. This factor operates the same way across the four experiments; that is, there are no changes in the linguistic identities of the stimuli (still intervocalic [ɾ], [ð], and [d]), and there are no changes in the characteristics of the populations of listeners being tested (though the actual participants were different in all four experiments). By rerunning experiments 1 and 2 with a new set of stimuli produced by speakers of a different language, we expect that the first factor, raw phonetic/auditory discriminability, of the stimuli may change. Comparing the results of experiments 1 and 2 with those of experiments 3 and 4 may thus help us identify aspects of the listeners' response patterns that are affected by the linguistic system of contrast, and pull these apart from aspects of the data that may be due solely to phonetic properties of the particular stimuli used in the test.

4.1. Methods

4.1.1. Stimuli

New stimuli were prepared for this experiment and for experiment 4. Materials consisted of two tokens of the same VCV sequences that were used in experiments 1 and 2: [ada], [aɾa], [aða], [idi], [iɾi], [iði], [udu], [uɾu], and [uðu]. Multiple tokens of these were produced and recorded by two native speakers of Greek, one male and one female, using a head-mounted microphone in a soundproof booth. Greek speakers were chosen because all three of the test phones, [d], [ɾ], and [ð], are contrastive in Greek and are produced naturally in intervocalic position. The speakers attempted to produce equal stress on the first and second syllables. In order to control the amplitude across tokens and speakers, the peak amplitude was equated for each of the tokens. The two best recordings for each VCV sequence were used as stimuli in the studies. These materials were used as stimuli in both experiment 3 and experiment 4.

4.1.2. Participants

Again, participants were drawn from the same pools as experiments 1 and 2. The native Spanish speakers (N = 7, 2 men, 5 women) had a mean self-rating of their English ability of 5.5. The native English speakers (N = 10, 3 men, 7 women) had a mean self-rating of their Spanish ability of 1.6.[2] None of the participants reported any history of speech or hearing disorders.

4.1.3. Procedure

The similarity rating procedure that was used in experiment 1 was also used in this experiment. Participants heard pairs of physically different stimuli and responded with a rating score from 1 (very similar) to 5 (very different).

4.2. Results

The results of experiment 3 are shown in figure 3; as in the graph from experiment 1, "more similar" is at the bottom of the graph and "more different" is at the top, and the means are plotted along with their 95 % confidence inter-

2 Note that in this experiment, some of the English-speaking subjects in this experiment did in fact have some exposure to Spanish, unlike those in experiments 1 and 2. We included these participants because, in an experiment not reported on here (see Boomershine et al. 2004, 2005), we found no significant difference in responses to these stimuli by native English speakers who had anywhere from no Spanish experience to an intermediate level with a self-rating of 4.5 on a scale from 0–7. As is reported in that study, only native English speakers who are advanced Spanish speakers (with a self rating greater than 5) begin to approach the perceptual characteristics of the native Spanish speakers; the native English speakers with an advanced level of Spanish patterned almost identically to the native Spanish speakers in the discrimination task and in between the native English speakers with little or no experience in Spanish and native Spanish speakers in the rating task.

vals. These results were analyzed in the same way as those of experiment 1, reported in section 2.2, using a repeated measures analysis of variance on z-score normalized rating scores. There was not a significant main effect of pair ($F[2,30] = 2.389$, $p > 0.05$). However, as in experiments 1 and 2, there was a significant pair by group interaction effect ($F[2,30] = 20.289$, $p < 0.05$). Subsequent planned comparison independent samples t-tests show that the English listeners rated the pair [d]/[ɾ] (which is allophonic in English) as more similar than did the Spanish speakers (for whom the pair is phonemic) ($t(15) = 4.652$, $p < 0.05$). Similarly, Spanish listeners rated the pair [d]/[ð] (which is allophonic in Spanish) as more similar than did English listeners (for whom the pair is phonemic) ($t(15) = 5.162$, $p < 0.05$). Finally, there was no significant difference between the two language groups in the rating of [ð]/[ɾ] ($t < 1$).

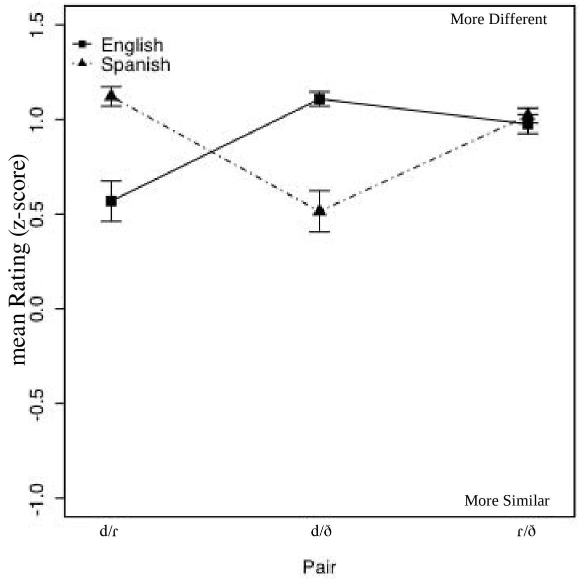


Figure 3. Results of experiment 3. Normalized similarity rating of [d], [ð], and [ɾ] by Spanish-speaking and English-speaking listeners. Stimuli produced by Greek speakers.

4.3. Discussion

The results from experiment 3 also provide evidence that allophonic relationships influence the perceived distance between sounds in phonological processing. The allophonic pairs for English listeners ([d] and [ɾ]) and for Spanish listeners ([d] and [ð]) were both rated as being more similar than the non-allophonic pairs. These results are very similar to those for experiment 1, which used the same task but involved different stimuli. One interesting difference between experiments 1 and 3 is that in experiment 1, the native Spanish speakers thought that the [ɾ]/[ð] distinction was the most salient, while in experiment 3, they found the [d]/[ɾ] distinction most salient. This is most likely due to the change in the raw perceptibility of the stimuli; in experiment 1, the stimuli were produced by native English speakers who perhaps did not make a particularly clear distinction between [d] and [ɾ], which are allophonic in English, but in experiment 3, the stimuli were produced by native Greek speakers, who do make a distinction between [d] and [ɾ] in production. The Spanish listeners found these stimuli, therefore, more perceptually distinct than those produced by English speakers.

5. Experiment 4: Discriminating Greek [d], [ɾ], [ð] pairs

5.1. Methods

5.1.1. Stimuli

The stimuli that were used for experiment 3 were also used in this experiment.

5.1.2. Participants

The participants in this experiment were drawn from the same pools as the other three experiments. The native Spanish speakers (N = 7, 4 men, 3 women) had a mean self-rating of their English ability of 6.1; the native English speakers (N = 11, 5 men, 6 women) had a mean self-rating of their Spanish ability of 1.18. None of the speakers reported any history of speech or hearing disorders.

5.1.3. Procedure

The speeded discrimination procedure that was used in experiment 2 was also used in this experiment. Participants heard pairs of stimuli and responded "same" if the stimuli were the same tokens of the same type of stimuli (e. g. the same token of [ada] twice) and "different" if they were not (either different types, e. g. [ada] – [aɾa], or different tokens, e. g. [ada] – [ada] where the two were not the same production). Participants were told after each pair whether they were correct or incorrect and were given their response time and their overall average percent correct, in order to encourage fast, accurate responses.

5.2. Results

The results and 95 % confidence intervals for experiment 4 are shown in figure 4; as in figure 2, "more similar" is at the top of the graph and "more different" at the bottom. As with the reaction time data of experiment 2, reaction time is taken as a measure of perceptual distance, and each listener's reaction times are z-score normalized to remove individual differences in overall speed of responding.

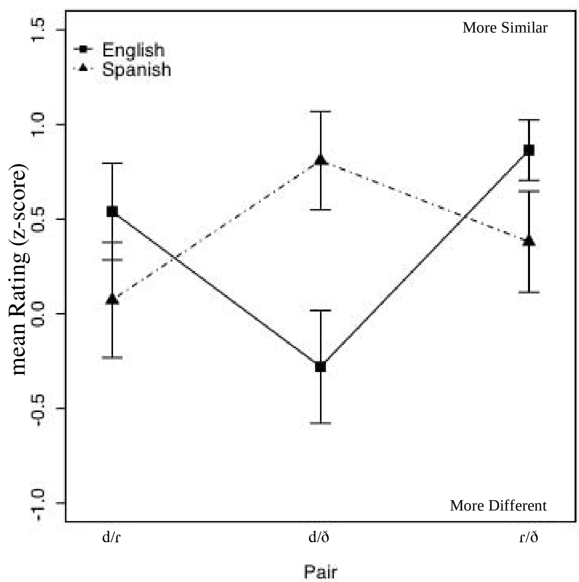


Figure 4. Results of experiment 4. Normalized reaction times for speeded discrimination [d], [ð], and [ɾ] by Spanish-speaking and English-speaking listeners. Stimuli produced by Greek speakers.

In this experiment, there was not a significant effect of pair (F = 1.122). There was, however, a significant pair by group interaction ($F[2,32] = 5.939$, $p < 0.05$). Subsequent planned comparison independent samples t-tests showed that as with experiment 3, the English listeners found [d]/[ɾ] more similar than the Spanish listeners did, though this difference was not quite significant in this particular experiment (t = 1.156). As in the previous experiments, too, the Spanish listeners found [d]/[ð] to be significantly more similar than the English listeners did ($t(16) = 2.538$, $p < 0.05$). Interestingly, there was also a trend in this experiment toward a difference between the two groups for the pair [ð]/[ɾ]; unlike all three of the other experiments, where the two groups had responded to this pair in the same way, in this experiment, the English listeners found [ð] and [ɾ] to be much more similar than the Spanish listeners did, though as with [d] and [ɾ], this difference was not quite significant ($t(16) = 1.664$).

5.3. Discussion

The results from experiment 4 again confirm our hypotheses about the role of allophony as opposed to phonemic contrast in perception: each pair was found to be less perceptually distinct by listeners for whom the pair is allophonic than by listeners for whom it is phonemic. The lack of significance between the two groups in the discrimination of [d]/[ɾ] may again be due to the raw auditory discriminability of the stimuli in this experiment as opposed to experiment 2, which used the same task but English-produced stimuli. That is, in experiment 4, perhaps the native English listeners found the Greek [d]/[ɾ] to be more distinct than the English [d]/[ɾ] of experiment 2 because the Greek [d] and Greek [ɾ] are inherently more different. The difference between the English and Greek stimuli might also explain why there was a (non-significant) tendency for Spanish speakers to find [ɾ]/[ð] more distinct than the English speakers did in experiment 4; if the Greek stimuli are acoustically more like Spanish phones, then perhaps the Spanish listeners simply had an easier time perceiving the difference than did the English listeners. Further experimentation on the raw phonetic discriminability of all of these sounds needs to be carried out to confirm these conjectures. Importantly, however, the Spanish speakers still found the [d]/[ɾ] pair to be more distinct than did the English speakers, while the English speakers found the [d]/[ð] pair to be more distinct than did the Spanish speakers.

6. General Discussion and Conclusion

6.1. Discussion

In summary, all four experiments showed a similar pattern. Across languages, speakers of a language in which a particular pair of sounds is contrastive at a phonemic level perceive that pair as being more perceptually distinct than do speakers of a language in which the pair is not phonemically contrastive. In each of the experiments, the English speakers found [d]/[ð], which is a phonemically contrastive pair in English but allophonic in Spanish, to be more perceptually distinct than the Spanish speakers did. Similarly, the Spanish speakers found [d]/[ɾ], which is phonemically contrastive in Spanish but allophonic in English, to be more perceptually distinct than the English speakers did. The pair [ð]/[ɾ] had about the same level of perceptual distinctiveness in the two languages; recall that in each language one sound of the pair is in an allophonic relationship with a different sound that is also present in the inventory of the other language.

This pattern of results is interesting because while [d] and [ð] are both phonemically and surface contrastive in English, [d] and [ɾ] are only phonemically contrastive in Spanish since they do not contrast in any surface minimal pair (see (4)). It is not too surprising to find that the phonemic level of contrast was related to listeners' ratings of sound similarity in experiments 1 and 3, given that the rating task used in these two experiments encourages a degree of off-line contemplation of the sounds. However, the AX discrimination experiments reported here (experiments 2 and 4) used a "phonetic" listening task that was designed to tap an earlier level of processing in order to see possible effects of the presence or absence of surface contrast. In fact the AX speeded discrimination task is a common psychoacoustic task that is generally assumed to show phonetic responses, but here it apparently does not: the results of experiments 2 and 4 closely matched those of 1 and 3. This leads us to wonder if it would be possible in any listening task to see "phonetic" responding independent of phonological structure.[3]

Of course, this is not to say that phonetic characteristics do not matter. One of the most noticeable differences between experiments 1 and 2 on the one hand and experiments 3 and 4 on the other was that the contrast between [d] and [ɾ] seemed to be much more salient in the second set of experiments, for both English and Spanish listeners. Because the only thing that changed between the two sets of experiments was the specific acoustic stimuli being used, we assume that this change in experimental materials created the difference in results; that is, the differences between the two sets of experiments (1 and 2 on the one hand, and 3 and 4 on the other) was due to the raw phonetic differences between the stimuli, *not* to differences in phonological patterning. The similarities between the two sets of results, on the other hand, are strongly tied to the phonological

3 It should be noted that there have also been claims that shorter inter-stimulus intervals and a lower degree of uncertainty in the task may reduce language-specific effects (Polka 1991; Fox 1984). It would be interesting to see if these effects, however, can ever actually be eliminated from processing.

systems of the native languages of the listeners. Evidently the Greek [d] and [ɾ] tokens were more distinct from each other than were the American English [d] and [ɾ] tokens. Given the lack of a [d]/[ɾ] contrast in English and the presence of such a contrast in Greek, it makes sense to believe that the Greek speakers would be better at keeping them separate in production, which would then transfer over to a better ability by listeners to differentiate them.

It is also interesting to note that in all of the experiments, there was a tendency for English listeners to perceive [d]/[ð] as more distinct than [ɾ]/[ð] despite the observation that there is no apparent representational difference in English between the two pairs; the sounds in each pair are contrastive at phonemic as well as surface levels, as shown in (5) (repeated from (4)).

(5) English [d]/[ð] versus [ɾ]/[ð]

	d/ð	ɾ/ð
Phonemic contrast	/d/ – /ð/	/d/ – /ð/
Surface contrast	[d] – [ð]	[ɾ] – [ð]

It may be that this tendency is simply a result of the raw overall auditory qualities of the sounds in question, an issue that must be explored by further research. It is also possible, however, that the difference is due to the fact that /d/ and /ð/ are each phonemes of English in a traditional analysis, while [ɾ] is simply an allophone of /d/. Although this difference is not indicated by the representations given in (4) or (5), perhaps the notion of contrast is even more finely nuanced than we have shown here. Again, we leave this question to later research.

In sum, the data presented in this paper suggest that phonemic contrast strongly influences speech perception, and that surface phonetic detail influences perceptual discrimination judgments. These results are important in that any model of speech perception must account for them, making sure that the phonemic level of representation is kept distinct from the allophonic level, with the phonemic level resulting in more distinct perceptual contrasts than the allophonic level. There are multiple perceptual models that achieve or could achieve this result; we outline two of them below: a phonological inferencing model and a lexical processing model.

6.2. Modeling the role of allophony and contrast in speech perception

In Gaskell and Marslen-Wilson's (1998) phonological inferencing model of speech perception, the acoustic signal is perceived in terms of the phonological representation that produced it. For instance, suppose that in Spanish there is a lenition rule that changes an underlying stop /d/ into the fricative [ð], and that the word *donde* 'where' has the abstract lexical representation /donde/. With these assumptions, this type of model predicts that the lenition rule is "undone" during the perception of *de* [ð]*onde* 'from where', producing a formal (phonetic/phonological) representation that matches the lexical representation. Thus, the prediction is that in a language with an allophonic relation between [d] and [ð], the acoustic signal of *de* [ð]*onde* is perceived exactly like that of [d]*onde*. The difference between Spanish and English is that English has no such rule, so that the perception of [ð]*onde* would not be subject to the undoing of such a rule, and there would be a distinction between the signals [d]*onde* and [ð]*onde*. Hence, English speakers are correctly predicted to find [d] and [ð] more distinct than Spanish speakers. Note that the same argument can be made, with the role of the languages reversed, for the relation between [d] and [ɾ] – English listeners undo a flapping rule and perceive [d] and [ɾ] as the same, while Spanish listeners have no such rule and perceive [d] and [ɾ] as distinct. In either case, the distinction between [ð] and [ɾ] would be correctly predicted to pattern similarly in the two languages, because in each language, each of these phones is mapped to a different phonological representation.

Interestingly, this model also predicts that the difference between different realizations of [d] will be indistinct from [ð] in Spanish or [ɾ] in English, as each sound is immediately linked to its underlying phonological representation. This prediction was indirectly tested in the rating experiments (experiments 1 and 3) by the comparison of ratings of [d]/[ð] to [d]/[d] and [ð]/[ð] pairs in Spanish and the comparison of ratings of [d]/[ɾ] to [d]/[d] and [ɾ]/[ɾ] pairs in English. In both sets of comparisons, it was found that the pair containing two different articulatory realizations of the same phoneme (e.g. [d]/[ð] or [d]/[ɾ]) was rated as significantly more different than the pairs containing the same articulatory realizations (e.g. [d]/[d], [ð]/[ð], or [ɾ]/[ɾ]). This result was found in experiment 1 with the English-produced stimuli, in which planned comparison paired samples t-tests showed that for Spanish listeners, the difference between [d]/[ð] and [d]/[d] was significant [$t(9) = -7.45$, $p < 0.05$], as was the difference between [d]/[ð] and [ð]/[ð] [$t(9) = 9.403$, $p < 0.05$]. Similarly, for English listeners the difference between [d]/[ɾ] and [d]/[d] was significant [$t(17) = -7.324$, $p < 0.05$], as was the difference between [d]/[ɾ] and [ɾ]/[ɾ] [$t(17) = 7.558$, $p < 0.05$]. The same pattern was found in experiment 3 with the Greek-produced stimuli, where for Spanish listeners, [d]/[ð] versus [d]/[d] was significantly different [$t(6) = 12.304$, $p < 0.05$], as was [d]/[ð] versus [ð]/[ð] [$t(6) = 11.072$, $p < 0.05$]. For the English listeners in experiment 3, the comparison of [d]/[ɾ] versus [d]/[d] was significantly different [$t(9) = 12.613$, $p < 0.05$], as was [d]/[ɾ] versus [ɾ]/[ɾ] [$t(9) = 12.260$, p

< 0.05]. While the inferencing model can thus account for the differences found across languages in the comparison of allophonic versus phonemic pairs, it is not powerful enough to correctly predict perceptual differences for the different types of "allophones of the same phoneme" found within a single language.[4]

In a lexical processing model, on the other hand, both types of results are predicted. In this approach, differences between phonological representations come at the lexical level, once listeners have tried to access words themselves, rather than being a property of the signal-to-representation mapping.

One type of lexical processing model is an exemplar model (see, e. g., Goldinger 1992, 1996; Palmeri et al. 1993; Johnson 1997a, b, 2004; Coleman 2002; Pierrehumbert 2003; Hawkins 2003). In an exemplar model, grammar is an emergent property over stored exemplars of all utterances heard. Word recognition is achieved by matching an incoming acoustic signal with the most similar stored representation of the signal, as defined by the amount of activation of the various stored representations. Hence an incoming [d] will activate stored examples of [d] more than it will activate stored examples of, say, [z], and so it will be recognized as [d]. Allophonic relations in this kind of model are represented by high co-activation (Johnson, 2006). For example, in Spanish, an incoming [d] will activate both [d] and [ð] because there are words that variably contain each different pronunciation. In English, on the other hand, [d] will activate [d] and [ɾ], but not [ð]. High rates of co-activation will make two sounds less perceptually distinct; the results of the experiments here would therefore be correctly predicted. Further, as with the phonological inferencing model, [ð] and [ɾ] are correctly predicted to pattern similarly across the two languages; in this case, because they are not activated by the same incoming signals. That is, in English, [ð] is activated only by an incoming [ð], and [ɾ] is activated by an incoming [d], [t], or [ɾ]. Thus, the signals that activate [ð] and [ɾ] do not overlap. Similarly for Spanish, [ð] is activated by an incoming [d] or [ð], while [ɾ] is activated by an incoming [ɾ]; the activation signals are again non-overlapping.

An exemplar model also predicts that even though an incoming [d] in Spanish will activate both [d] and [ð], as will an incoming [ð], the perception of a [d]/[d] pair will differ from that of a [d]/[ð] pair. This result comes about for a

4 One might reasonably suggest that listeners' ability to detect the phonetic differences that separate allophones of the same phoneme is based on purely auditory processing abilities that are quite separate from speech perception. We are sympathetic with such an explanation of listeners' performance. Unfortunately however, the phonological inferencing model, as it has been presented in the literature, denies this possibility by suggesting that all allophones lead to the "perception" of the underlying phoneme.

number of reasons. First, the acoustic representation of an incoming signal is not completely removed; every utterance is stored with its acoustic representation intact, and similarity between signals is calculated over these acoustic representations. Second, the words that are activated by an incoming signal will depend on this similarity matching. Consequently, the words activated by an incoming [d] might be somewhat different than those activated by an incoming [ð], and words that are activated by both signals may be activated to a greater or lesser extent by one than the other. This use of the acoustic representation of the signal in activating words in the lexicon allows such a model to predict both the difference in the perception of phonemic and allophonic pairs across languages as well as the difference in the perception of pairs of allophones of the same phoneme within a language.

In summary, while both phonological inferencing and exemplar models are able to correctly predict the differing influences of allophony versus phonemic contrast on perception (a phonemic relationship is perceptually more distinct than an allophonic relationship, regardless of the actual identity of the sounds in question), only the exemplar model is successful in accounting for the differences in perception that one finds within a language between pairs of the "same" allophone of one phoneme (e. g. [d]/[d]) and "different" allophones of one phoneme (e. g. [d]/[ð]).

Returning to the speculations of Trubetzkoy, we see this study as providing further evidence for his claim that the particular phonological relation holding between sounds in a language has an impact on a listener's perception of those sounds. Our direct test of allophony versus contrast points to the need to include both in the inventory of phonological relations shown to influence perception. The inventory thus includes phonemic contrast, partial contrast due to phonological neutralization, the non-contrastive relation of allophony, as well as non-contrastiveness due to the absence of one or more of the sounds in a language's sound system. The extent to which the two types of non-contrastiveness differ with regards to their impact on speech perception remains an open question and one that must be addressed in future research.

References

Best, Catherine
1994 The emergence of native-language phonological influences in infants: A perceptual assimilation model. In Judith Goodman and Howard Nusbaum (eds.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words* (pp. 167–224). Cambridge, MA: MIT Press.

Best, Catherine, McRoberts, Gerald, and Sithole, Nomathemba
1988 Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 345–360.

Boomershine, Amanda, Hall, Kathleen Currie, Hume, Elizabeth, and Johnson, Keith
2004 The influence of contrast vs. allophony on perception: The case of Spanish and English. Paper presented at the Mid-Continental Workshop on Phonology, Northwestern University, October 30, 2004.

Boomershine, Amanda, Hall, Kathleen Currie, Hume, Elizabeth, and Johnson, Keith
2005 The influence of language experience on speech perception: The case of Spanish and English. Paper presented at The Hispanic Linguistics Symposium, The Pennsylvania State University, November 10, 2005.

Coleman, John
2002 Phonetic representations in the mental lexicon. In Jacques Durand and Bernard Laks (eds.), *Phonetics, Phonology, and Cognition* (pp. 96–130). Oxford: Oxford University Press.

Dupoux, Emmanuel, Pallier, Christophe, Sebastian, Nuria, and Mehler, Jacques
1997 A destressing 'deafness' in French? *Journal of Memory and Language, 36*, 406–421.

Fox, Robert
1984 Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance, 10*(2), 526–540.

Francis, Alexander, and Nusbaum, Howard
2002 Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 349–366.

Gaskell, M. Gareth, and Marslen-Wilson, William
1998 Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 24*(2), 380–396.

Gaskell, M. Gareth, and Marslen-Wilson, William
2001 Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language, 44*(3), 325–349.

Goldinger, Stephen
1996 Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(5), 1166–1183.

Goto, Hiromu
1971 Auditory perception by normal Japanese adults of the sounds "l" and "r." *Neuropsychologia, 9*, 317–323.

Harnsberger, James
2001 The perception of Malayalam nasal consonants by Marathi, Punjabi, Tamil, Oriya, Bengali, and American English listeners: A multidimensional scaling analysis. *Journal of Phonetics, 29*, 303–327.

Hawkins, Sarah
2003 Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics, 31*(3–4), 373–405.

Huang, Tsan
2001 The interplay of perception and phonology in tone 3 sandhi in Chinese Putonghua. In Elizabeth Hume and Keith Johnson (eds.), *Studies on the Interplay of Speech Perception and Phonology* (Vol. 55, pp. 23–42). Columbus, OH: Ohio State University Working Papers in Linguistics.

Huang, Tsan
2004 Language-specificity in auditory perception of Chinese tones. Unpublished PhD dissertation, The Ohio State University, Columbus, OH.

Hume, Elizabeth, and Johnson, Keith
2001 A model of the interplay of speech perception and phonology. In E. Hume and K. Johnson (eds.), *The Role of Speech Perception in Phonology* (pp. 3–26). San Diego: Academic Press.

Hume, Elizabeth, and Johnson, Keith
2003 The impact of partial phonological contrast on speech perception. *Proceedings of the Fifteenth International Congress of Phonetic Sciences* (pp. 2385–2388).

Johnson, Keith
1997a Speech perception without speaker normalization. In Keith Johnson and John Mullennix (eds.), *Talker Variability in Speech Processing* (pp. 145–165). San Diego: Academic Press.

Johnson, Keith
1997b The auditory/perceptual basis for speech segmentation. In Kim Ainsworth-Darnell and Mariapaola D'Imperio (eds.), *Papers from the Linguistics Laboratory* (Vol. 50, pp. 101–113). Columbus, OH: Ohio State University Working Papers in Linguistics.

Johnson, Keith
2004 Cross-linguistic perceptual differences emerge from the lexicon. In Augustine Agwuele, Willis Warren, and Sang-Hoon Park (eds.), *Proceedings of the 2003 Texas Linguistics Society Conference: Coarticulation in Speech Production and Perception* (pp. 26–41). Sommerville, MA: Cascadilla Press.

Johnson, Keith
2006 Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics, 34*, 485–499.

Lahiri, Aditi
1999 Speech recognition with phonological features. In *Proceedings of the XIVth International Congress of Phonetic Sciences* (pp. 715–718). San Francisco.

MacKain, Kristine, Best, Catherine, and Strange, Winifred
1981 Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics, 2*, 369–390.

Marian, Viorica, and Spivey, Michael
2003 Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition, 6*(2), 97–115.

Mohanan, Karuvannur, and Mohanan, Tara
1984 Lexical phonology of the consonant system in Malayalam. *Linguistic Inquiry, 15*(4), 575–602.

Palmeri, Thomas, Goldinger, Stephen, and Pisoni, David
1993 Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(2), 309–328.

Pierrehumbert, Janet
2003 Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46*, 115–154.

Polka, Linda
1991 Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America, 89*, 2961–2977.

Polka, Linda, and Werker, Janet
1994 Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 421–435.

Pruitt, John, Akahane-Yamada, Reiko, and Strange, Winifred
1998 Perceptual assimilation of Hindi dental and retroflex consonants by native speakers of Japanese and English. *Proceedings of the Joint Meeting of ASA/16th ICA*.

Strange, Winifred
1995 Cross-language studies of speech perception: A historical review. In W. Strange (ed.), *Speech Perception and Linguistic Experience* (pp. 3–45). Baltimore: York Press.

Strange, Winifred, and Dittman, Sibylla
1984 Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception and Psychophysics, 36*(2), 131–145.

Takane, Yoshio, and Sergent, Justine
1983 Multidimensional scaling models for reaction times and same-different judgments. *Psychometrika, 48*, 393–423.

Werker, Janet, Gilbert, John, Humphrey, Keith, and Tees, Richard
1981 Developmental aspects of cross-language speech perception. *Child Development, 52*, 349–353.

Werker, Janet, and Logan, John
1985 Cross-language evidence for three factors in speech perception. *Perception and Psychophysics, 37*, 35–44.

Interplay between perceptual salience and contrast: /h/ perceptibility in Turkish, Arabic, English, and French

Jeff Mielke

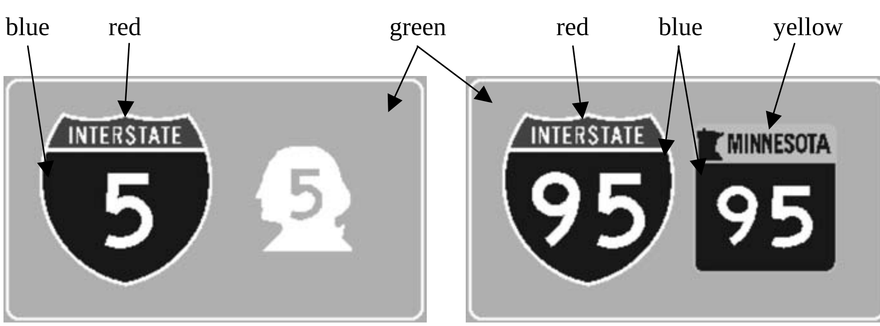
1. Introduction

Phonology is complicated. Systems of phonological contrast are intertwined with external factors such as perceptual distinctness, articulatory ease, functional load, and frequency, and an understanding of how contrasts emerge and dissolve requires attention to these factors. It is well known that the mechanisms of language use and language change favor contrasts which are perceptually distinct over those which are indistinguishable, and contrasts involving sounds which are easy to articulate over contrasts involving sounds which are difficult to produce (see e. g., Steriade 1997, 2001, 2004; Flemming 2002 [1995], 2004, Silverman 1997 [1995]; Wright 1996, 2001; Chang, Plauché, and Ohala 2001; Bybee 2001). Contrasts with greater functional load (i. e. contrasts which distinguish more words or more frequent words) may be less likely to be neutralized than those which seldom distinguish words (see e. g., Martinet 1955), and functional load is intimately related to frequency. While Martinet's claims are controversial, the results reported here support a connection between functional load and contrast maintenance.

The goal of this paper is to illustrate how perception influences contrast in phonological systems, how contrast impacts perception, and how the functional load and perceptual distance of a contrast can interact to resist or induce neutralization. To this end, data will be drawn from the status of phonological contrast between /h/ and its absence in Turkish, Arabic, English, and French, and the perceptibility of the contrast for speakers of these languages. This phoneme and these languages are selected because /h/ is a perceptually weak sound, vulnerable both to misperception and to deletion, because these four languages allow the contrast between /h/ and ∅ in different environments, and because speech perception data for /h/ in different environments for speakers of these four languages is readily available.

Section 2 summarizes the methods and results of a perception experiment performed by Mielke (2003). The focus of this earlier article is the interaction between perception and phonology, while the focus of this one is the implications of the experimental results for the relationships between contrast and other factors. These overlap as follows. Section 3 deals very briefly with the influence of perceptual distance on the maintenance and neutralization of contrast, and as such is basically a summary of discussion in Mielke (2003). Section 4 addresses the influence of phonological contrast on speech perception, in much more depth than the earlier article. Section 5 deals with the interaction between perceptual distance and functional load in their effects on contrast, something which is not addressed at all in Mielke (2003).

The process of identifying phonemes from auditory cues is much like the process of identifying highway signs from visual cues. The visual image of a sign or the auditory signal of an allophone is composed of cues, which are useful for identification, as well as noise, which does not benefit identification and is largely ignored by perceivers. A driver looking for the interstate highway will likely not take all available visual information into account when reading a sign. Signs with different meanings are distinguished visually, but not all meaningful contrasts are equally distinct. For example, in figure 1, the markers on the sign on the left represent a contrast with greater perceptual distance than the one on the right, according to some visual cues (**size**, **main color**, **secondary color**, and **shape**).



large	vs.	small	**size**		*both large*	
blue	vs.	white	**main color**		*both blue*	
red	vs.	∅	**secondary color**	red	vs.	yellow
shield	vs.	head	**shape**	shield	vs.	square

Figure 1. Maintenance of a perceptually robust contrast (Washington, left) and a perceptually weaker one (Minnesota, right)

Size and **main color** are salient cues, and in order to correctly distinguish an interstate highway marker from a Washington state highway marker it may not be necessary for a driver to notice the exact radius of the curve on the shield or the shape of George Washington's nose, or even to look for **secondary color** or **shape** differences. A driver familiar with the Washington state/interstate contrast is likely to initially identify any large, predominantly blue marker with a warm color on top as an interstate highway marker. This is all fine and good until the driver is presented with the Minnesota state/interstate contrast, and both signs look like interstate highway markers. Although the two signs are in different categories in the Minnesota system, they are not distinguished by any cues necessary to discriminate the Washington state/interstate contrast. But a driver who is familiar with the Minnesota state/interstate contrast will most likely attend to **secondary color** and **shape** in order to make the distinction. Visual images (and allophones) can be parceled out into cues and noise, but what counts as a cue and what counts as noise is not necessarily universal. Rather, different drivers (and different speakers) who are familiar with different contrasts attend to different cues. Washington drivers do not need to notice **secondary color** and **shape** cues because these cues are redundant. The functional load placed on these cues in the Minnesota system (because at least two crucially different highway markers are distinguished by them) allows Minnesota drivers to maintain what on a larger scale is a relatively weak contrast. But the lack of perceptual distance leads to confusion for non-native drivers, and may leave the Minnesota marker more vulnerable to sign-changing legislation.

This analysis of visual contrast may also be applied to phonological contrast. Like road signs, phonemes have cues which may or may not be exploited, and different listeners may attend to different cues, while treating others as noise. The remainder of this paper will be concerned with how native phonology can determine the perceptual strategies employed by listeners in phoneme recognition, and how factors such as perceptual salience and functional load can determine the presence or absence of contrast.

2. /h/ deletion and perceptibility

2.1. Turkish /h/ deletion

Mielke (2003) showed that the pattern of /h/ deletion in Turkish can be explained on the basis of the perceptual salience of /h/ in different segmental contexts. /h/ is optionally deleted in fast speech in Turkish, but only in certain segmental contexts (Lewis 1967, Sezer 1986). Sezer (1986) reports that /h/ is optionally deleted *before* sonorant consonants (1a), but not after them (1b). When /h/ is deleted from preconsonantal or final position, compensatory lengthening of the preceding vowel occurs, as in (1a). /h/ is optionally deleted *after* voiceless stops and affricates (2b), but not before them (2a). /h/ is optionally deleted before *and* after voiceless fricatives (3a and 3b), as well as intervocalically (4a), but /h/ is not deleted word-initially (4b). Sezer reports that /h/ does not delete word-finally (4c), but informal native speaker judgments indicate that it deletes in this environment as well, and production experiments show that word-final /h/ deletion is conditioned by the initial segment of the following word, occurring under conditions identical to those under which word-internal /h/ is deleted (Mielke 2002a). Turkish deletes /h/ in segmental contexts that are not obviously related in a formal phonological sense (i. e., the environments conditioning /h/ deletion do not form what is traditionally considered to be a natural class in terms of widely-accepted phonological features), so its phonology is fertile testing ground for the hypothesis that perception influences contrast, because /h/ deletion is the loss of contrast between /h/ and ∅.

(1) /h/ is only deleted *before* sonorant consonants.

a. fihrist	~	fi:rist[1]	'index'
köhne	~	kö:ne	'old'
kahya	~	ka:ya	'steward'
b. merhum		*merum	'the late'
imha		*ima	'destruction'

(2) /h/ is only deleted *after* voiceless stops and voiceless affricates.

a. kahpe		*ka:pe	'harlot'
ahtʃi		*a:tʃi	'cook'
b. ʃüphe	~	ʃüpe	'suspicion'
metʃhul	~	metʃul	'unknown'

(3) /h/ is deleted before *and* after voiceless fricatives.

a. mahsus	~	ma:sus	'special to'
b. safha	~	safa	'step'

(4) /h/ is deleted intervocalically and word-finally, but not word-initially.

a. tohum	~	toum	'seed'
b. hava		*ava	'air'
c. timsah		?timsa:	'crocodile'

1 Compensatory lengthening occurs when /h/ is deleted from coda position (Sezer 1986), but see also Barnes (2001) for an account of this phenomenon based on syllabic isochrony rather than moraic isochrony.

Mielke's (2002b) production study of native Turkish speakers in Columbus, Ohio, found deletion in fewer environments than Sezer reports. This paper is concerned with the deletion pattern reported by Sezer.

A perception experiment was designed to test the relative salience of /h/ in various phonetic environments by speakers of various languages: Turkish, which allows /h/ in many environments, Arabic, which also allows /h/ in many environments, English, which allows /h/ only in prevocalic environments, and French, which has no /h/ sound at all.

2.2. Perception experiment methods

2.2.1. Stimuli

320 nonword stimuli were produced in isolation by a male native speaker of Turkish and recorded in mono using a Shure SM10A head-mounted microphone through a Symetrix SX202 dual mic preamp into a Teac V-427C stereo cassette deck. The stimuli were then digitized at 22050 Hz using a Marantz PMD222 portable cassette recorder and SciCon R&D Inc.'s PCQuirer signal analysis software, and amplitude normalized using Syntrillium's CoolEdit audio editing software.

All stimuli were disyllabic and produced with final stress. 68 stimuli contained intervocalic consonant clusters consisting of /h/ preceded by one of nine different types of consonant (voiceless stop, voiceless affricate, voiceless fricative, voiced stop, voiced affricate, voiced fricative, nasal, liquid, glide). Another 68 stimuli contained intervocalic consonant clusters consisting of /h/ followed by a consonant. 68 foil stimuli contained a single consonant between vowels and no /h/. 24 stimuli contained /h/ in one of three vowel environments (initial, intervocalic, and final), and 12 corresponding foil stimuli contained no /h/. Half of the consonant foil stimuli contained a long vowel before the consonant and all of the word-final foil stimuli contained a long final vowel, to simulate the compensatory lengthening that occurs in Turkish when /h/ is deleted from preconsonantal or word-final position. An additional 80 nontarget stimuli without /h/ were also recorded.

2.2.2. Subjects

The subjects consisted of five female and 14 male native speakers of Turkish in Columbus, Ohio, aged 19–33, 14 female and seven male Ohio State University undergraduates, all native speakers of American English, one male and twenty female native speakers of French in Paris, aged 18–28, and two female and ten male native speakers of Arabic in Paris, aged 20–36, including seven Moroccans, three Algerians, one Mauritanian, and one Jordanian. While the Arabic-speaking subjects clearly have been exposed to diverse varieties of Arabic, the varieties represented in the study are similar with respect to /h/ (cf. Maltese Arabic) (Zawadowski 1978), and variation is not expected to impact the results.

2.2.3. Procedures

The stimuli were randomized and played to subjects over Sennheiser HD 420 headphones from a CTX EzBook 700 laptop computer in a sound-attenuated booth or room. As subjects heard each nonword they were presented on a computer screen with all the segments in the word other than /h/, and instructed to click on the point in the nonword where they heard /h/, or to click on a button representing no /h/ if they heard no /h/. An ‘h’ appeared on the screen at the point where the subject clicked.

2.2.4. Data Analysis

Sensitivity (d’) (Green and Swets 1966; Winer 1971; MacMillan and Creelman 1991) was computed for each subject for each of the 21 environments; d’ is a measure of sensitivity based on z scores (the number of standard deviations from the mean of a standard normal distribution) of hit and false alarm rates: d’ = z(H) – z(F). d’ is positive when the hit rate exceeds the false alarm rate (i. e., subjects report hearing /h/ more often when it is present than when it is not). A d’ of zero indicates that hit and false alarm rates are the same, that subjects have no sensitivity to the presence or absence of /h/. For example, given a hit rate of 75 % and a false alarm rate of 30 %, d’ = z(0.75) – z(0.3) = 0.674 – (–0.524) = 1.199. See Mielke (2003) for further details on procedures and data analysis.

2.3. Results

The experiment yielded a d’ value for each of the 21 environments in each of the four languages (larger d’ values mean greater perceptual distance). These results are given in table 1.

Table 1. Results of the perception experiment: The average d' value is reported for each phonetic context and each language tested (T = voiceless stops, C = voiceless affricates, S = voiceless fricatives, D = voiced stops, J = voiced affricates, Z = voiced fricatives, N = nasals, L = liquids, Y = glides, V = vowels, # = word boundary).

	Turkish	Arabic	English	French		Turkish	Arabic	English	French
V_T	2.900	2.813	0.559	0.699	T_V	2.408	2.764	0.907	1.272
V_C	3.002	3.224	0.642	0.563	C_V	2.633	2.396	0.669	0.321
V_S	2.663	2.846	0.356	0.700	S_V	2.300	3.033	1.189	0.767
V_D	3.299	2.975	0.713	0.910	D_V	3.078	3.481	1.839	1.965
V_J	3.345	3.146	1.030	1.152	J_V	3.415	2.947	1.348	0.740
V_Z	3.340	3.329	1.204	0.876	Z_V	2.700	3.243	1.834	1.384
V_N	3.303	3.121	0.829	0.994	N_V	3.550	3.604	2.353	2.116
V_L	3.340	3.482	1.377	1.122	L_V	3.636	3.488	2.167	1.840
V_Y	3.462	3.542	1.823	1.295	Y_V	2.757	3.403	1.410	1.038
V_V	2.897	3.004	2.152	1.512	V_V	2.897	3.004	2.152	1.512
V_#	1.123	0.951	0.259	0.153	#_V	3.194	3.018	2.490	1.799

The results show that all else being equal, Turkish /h/ deletion occurs in environments where /h/ is perceptually weak *crosslinguistically* (see Mielke (2003) for a more detailed discussion of this aspect of the results). In essence, contrast between /h/ and ∅ is maintained more consistently in environments where the perceptual distance between /h/ and ∅ is large, and contrast is more likely to be neutralized in environments where perceptual distance is small. Figure 2 shows the perceptual distance (d') between /h/ and ∅ in eight environments for Turkish listeners, including four environments where deletion occurs, indicated on the chart by the word "deletes".

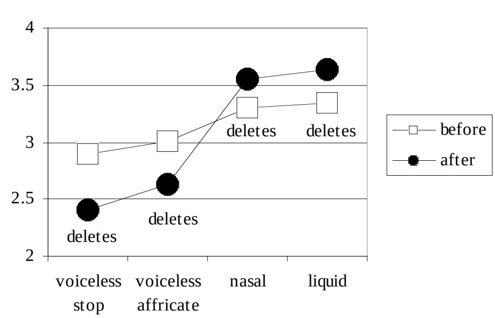


Figure 2. Turkish /h/ deletion occurs in perceptually weak environments.

Deletion occurs predominantly in environments where d' is low. Some additional factors related to the pattern shown here are discussed below in section 5.

3. Perception influences contrast

Because auditory nerve fibers exhibit a greater response at the onset of a stimulus signal (such as a vowel) than at the offset (Bladon 1986; Pickles 1988; Wright 1996), and CV transitions provide better cues than VC transitions (Fujimura, Macchi, and Streeter 1978; Ohala 1992), /h/ is more salient before a vowel, and therefore less perceptible before a sonorant consonant than after, because /h/ is always prevocalic when it follows a sonorant. Further, preconsonantal allophones of /h/ in Turkish may also provide weaker cues than prevocalic allophones of /h/ (e. g., because they are produced with less aspiration). Both of these claims are consistent with Turkish /h/ deletion patterns.

The fact that the opposite deletion pattern exists for voiceless stops and affricates can be explained on the basis of the fact that /h/ is immediately adjacent to aspiration or frication when it follows a voiceless stop or affricate, whereas when /h/ precedes a voiceless stop or affricate, it is separated from the noise by the stop closure. /h/ should be less perceptible after these sounds than before them. This is also consistent with Turkish /h/ deletion patterns.[2]

The Turkish results in figure 2 show that /h/ is more perceptible after nasals and liquids than before them, and as predicted, the pattern is reversed for voiceless stops and affricates. For each pair of environments in figure 2 (before and after a consonant), deletion occurs in the environment with lowest perceptibility in the pair.[3] The same underlying perceptual tendencies exist for all four groups of subjects, even if the ordering of d' values varies from language to language. Most notably, the tendency for prevocalic /h/ to be most salient is stronger for speakers of English and French, neither of which language allows /h/ in preconsonantal or word-final position. See Mielke (2003) for further discussion. In short, perceptual distance (which results from robust acoustic cues) leads to the

2 The perception and deletion of /h/ in the intervocalic context are believed to differ substantially from the perception and deletion of /h/ in other contexts. Intervocalic /h/, like postvocalic /h/ (Kavitskaya 2001), is more vowel-like, and may have more articulatory motivations for deletion, and may also be produced very differently in natural speech than in the experimental stimuli. Further, the discussion of contrast and perceptibility in this chapter focuses on pairs of contexts (preconsonantal vs. postconsonantal), and intervocalic /h/ does not present such a pair. For all of these reasons, it is not discussed further here, but see Mielke (2003) for a discussion.

3 Cross-type comparisons are discussed below in section 5.

presence or maintenance of contrast. Likewise, the lack of robust acoustic cues leads to perceptual similarity, which leads to the lack or loss of contrast.

4. Phonological contrast influences perception

4.1. Language-specific perceptual distance

While perceptual distance is predictable in part from the acoustic cues present in a given environment, listeners with more experience with the contrast "/h/ vs. ∅" are better at perceiving the contrast. Turkish and Arabic allow /h/ (and contrast it with ∅) in many environments, and Turkish and Arabic speakers are more sensitive to the contrast than speakers of English, which allows the contrast "/h/ vs. ∅" in fewer environments, and French, which does not contrast /h/ with ∅ at all.

One hypothesis is that listeners are most sensitive to /h/ in specific environments where it occurs in a language they are familiar with, and less sensitive in other environments. But as shown in figure 3, the presence or absence of contrast in a particular environment is insufficient to predict perceptibility, as shown by perceptibility of the contrast "/h/ vs. ∅" before and after affricates. Figure 3 shows the perceptibility of /h/ and whether or not /h/ is contrastive in four different environments for speakers of all four languages.

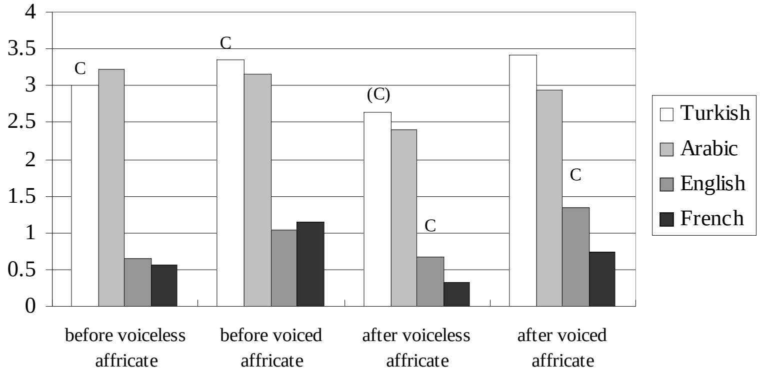


Figure 3. Contrast in a specific environment is not a good predictor of perceptibility. (C indicates presence of contrast between /h/ and ∅. Contrast is neutralized after voiceless affricates in Turkish casual speech.)

Neither Arabic nor French contrasts /h/ with ∅ before or after affricates, but the contrast is far more perceptible for Arabic listeners in these environments.

Further, among the four languages, only English permits /h/ after voiced affricates (e. g. *sagehen* and various words formed with +*hood* and +*house*), but the /h/ vs. ∅ contrast after voiced affricates is more perceptible for Arabic and Turkish listeners than for English listeners.

Arabic and Turkish listeners may perceive the contrast between /h/ and ∅ in unfamiliar contexts because the environments are acoustically similar to those where contrast exists. Turkish lacks /h/ after voiced affricates, but Turkish speakers are exposed to the contrast between /h/ and ∅ after voiced fricatives and voiceless affricates (consonants which are acoustically similar to voiced affricates, especially at the right edge). Arabic has /h/ before and after stops and fricatives, environments where the cues to the presence or absence of /h/ are similar to the cues to the presence or absence of /h/ before and after affricates. English has the contrast in two of the environments in figure 3, but not in any of the non-prevocalic environments where Turkish and Arabic permit /h/, so English speakers may rely on cues which are not found in these environments, thus failing to make the distinction in non-prevocalic environments, and also failing to benefit from additional cues which could increase their ability to perceive /h/ in prevocalic environments too. French lacks the contrast not only in the environments in figure 3, but in all environments, and so the contrast is relatively imperceptible.

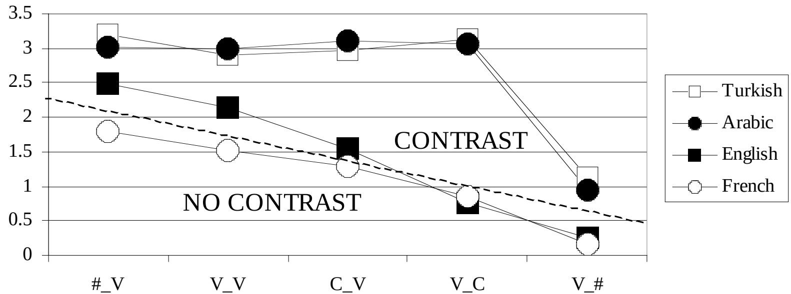


Figure 4. Contrast is a better predictor of perceptibility for general classes of environments.

When results for preconsonantal and postconsonantal environments are pooled, the experimental results correspond more clearly to the phonotactic restrictions in the four languages. Figure 4 shows the perceptibility and contrastiveness of /h/ in four environments for each of the four languages in the study. English allows the contrast only in prevocalic environments: #_V, V_V, C_V, where the contrast is most perceptible. French lacks the contrast "/h/ vs. ∅" in all contexts, and the contrast is less perceptible for French listeners than for other listeners. Turkish allows the contrast "/h/ vs. ∅" in all five environments shown: #_V, V_V, C_V, V_C, and V_# (but lacks the contrast after voiced affricates, and in many environments in casual speech (due to /h/ deletion): after voiceless obstruents, intervocalically, and before fricatives and sonorant consonants). Arabic allows the contrast "/h/ vs. ∅" in all five environments as well (but lacks the contrast before and after affricates, and word-finally in casual speech due to /h/ deletion).

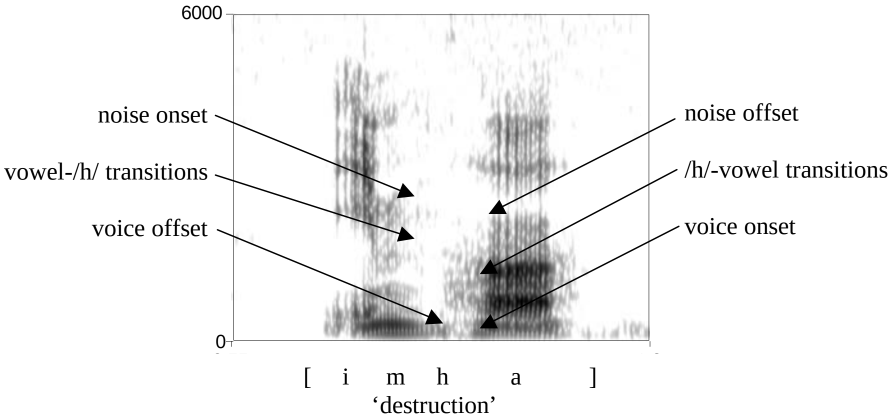


Figure 5. How to tell if there is an /h/[4]

Figure 5 shows some of the acoustic cues to the presence of /h/. In the best case scenario, /h/ is marked on the left edge by the offset of voicing, the onset of noise, and transitions from vowel formants to breathy /h/ formants, and on the right edge by the onset of voicing, the offset of noise, and transitions from /h/

4 "Noise onset" is present when /h/ is not preceded by a noisy sound (stop, fricative, or affricate), "noise offset" is present when /h/ is not followed by a noisy sound (fricative or affricate), "V-/h/ transitions" is present when /h/ is preceded by a vowel, "/h/-V transitions" is present when /h/ is followed by a vowel, "voice offset" is present when /h/ is voiceless and preceded by a voiced sound, and "voice onset" is present when /h/ is voiceless and followed by a voiced sound. The cue "vowel-/h/ transitions" would not be present for the /h/ in the figure 5. Many more language-specific cues exist, such as allophonic variation of other segments in words/stimuli. For example, Turkish /r/ is devoiced and fricated when it is followed by a voiceless obstruent (such as /h/), but not when it is followed a vowel. This should serve as a cue to the presence of /h/ only for the Turkish listeners. That cues of this type are less relevant to listeners who do not know Turkish and the fact that the Arabic and Turkish results are similar might allow these cues to be ignored for now.

formants to vowel formants. Not all of the available cues are necessary to perceive contrast in a given environment, but not all of the cues are always present, either. When the number of cues is small, the cues which are present are more important, and listeners who are not attending to certain cues are unlikely to detect /h/ in environments where they are the only cues available, just as Washington drivers are less likely to be sensitive to the Minnesota state/interstate highway sign contrast, because they are less likely to attend to the visual cues **secondary color** and **shape**, which are necessary to make the contrast.

4.2. Predicting language-specific perception

To explore which cues to /h/ different listeners are attending to, the experiment stimuli were coded with respect to the presence or absence of the six acoustic cues shown in figure 5, and a stepwise linear regression was performed for each language with d' as a dependent variable and the six cues as independent variables.

Table 2. Stepwise linear regression

	English (R^2 = .738)	French (R^2 = .540)
1. noise onset	B = 1.118, t = 4.601, p <.001	B = .800, t = 3.159, p = .006
2. /h/-V transitions	B = .934, t = 4.188, p = .001	B = .601, t = 2.585, p = .019
3. voice onset	B = .753, t = 3.376, p = .004	B = .534, t = 2.299, p = .034
(constant)	B = -.389, t = -1.352, p = .194	B = .060, t = -.201, p = .843

Significant factors were found only for English and French listeners (table 2). English listeners attend to **/h/-vowel transitions**, **noise onset**, and **voice onset**, cues that are present for nearly all instances of /h/ in English. When they occur, the other less salient cues are always redundant. French listeners also attend to **/h/-vowel transitions**, **noise onset**, and **voice onset**. CV transitions and the onset of noise and voicing are more salient than VC transitions and the offset of the same stimuli, even though they occur at different temporal locations.

While English and French listeners attend only to the most salient cues, Turkish and Arabic listeners must also attend to cues which are less salient (such as **noise offset**, **vowel-/h/ transitions**, and **voice offset**) because they are non-redundant cues for perceiving /h/ in preconsonantal position. Thus, Turkish and Arabic listeners perceive /h/ consistently well even in the absence of many of the more salient cues which are crucially important to English and French listeners. Because all the stimuli in the experiment contain at least some cues that these speakers attend to, the linear regression finds no cues to be particularly important.

The results of the linear regression show that the presence or maintenance of contrast (here the maintenance of the contrast between /h/ and ∅ in particular environments) leads to attention to acoustic cues, which in turn leads to perceptual distance. To perceive /h/ in environments where there are few cues to its presence, speakers of Turkish and Arabic must pay attention to less salient cues, but because they attend to these cues, they are more sensitive to /h/ in environments where there are fewer cues to its presence. Because of the lack of contrast in certain environments, the same cues can be ignored by English listeners because they are always redundant, and can be ignored by French listeners because they never mark a meaningful contrast. The absence (or loss) of a contrast leads to a lack of attention to acoustic cues, and therefore a loss of perceptual distance. In this way, listening to the "/h/ vs. ∅" contrast in Turkish or Arabic is like driving in Minnesota, and listening to the same contrast in English is like driving in Washington.

5. Functional load interacts with perceptual distance to influence phonological neutralization

Acoustic factors alone cannot explain the pattern of deletion. As seen above in figure 2, /h/ can be deleted when it is the first element in a consonant cluster or when it is the second element in a consonant cluster, but it is not necessarily deleted in environments where it is least perceptible, because /h/ is less perceptible before voiceless stops and affricates (where it does not delete) than before nasals and liquids (where it does delete). This last fact may be surprising if perceptual salience is the only factor which can explain the deletion pattern, but it is not the only factor. Articulatory ease, functional load, and frequency, for example, are reasonable factors to draw on as well. In this section, functional load will be incorporated into the discussion explaining the deletion pattern in Turkish.

Functional load achieved popularity among linguists in the 1950s (e. g. Hockett 1955, Martinet 1955), and more recently there has been a resurgence in interest in the concept (e. g. Surendran and Niyogi 2003, Hume 2006). In this chapter, a very simple interpretation of the concept will suffice: contrasts which distinguish more words (or which distinguish more frequent words) have greater functional load than contrasts which distinguish fewer (or less frequent) words.

/h/ always deletes on the side of a consonant where it is perceptually weakest, but regardless of the initial position of /h/, the result of deletion is an intervocalic consonant (figure 6).

$$\text{VChV} \xrightarrow{\text{/h/ deletion}} \text{VCV} \xleftarrow{\text{/h/ deletion}} \text{VhCV}$$

Figure 6. Two deletion rules with the same output

Table 3. Contrasts in a hypothetical language. The functional load of VtV vs. VhtV is increased when /h/ deletion merges VthV with VtV.

VthV	vs.	VtV	vs.	VhtV		VtV	vs.	VhtV
		ata				ata		
				ahto				ahto
athu	→	atu				atu×2		
etha						eta		
ethe	→	ete				ete×2		
etho						eto		
		eti				eti		
		etu	!	ehtu		etu	!	ehtu
				ohta				ohta
othe				ohte		ote	!	ohte
		oto	!	ohto		oto	!	ohto
				ohti				ohti
				ohtu				ohtu
ithe	→	ite				ite×2		
		iti				iti		
ithu				ihtu		itu	!	ihtu
utha						uta		
		ute				ute		
		uto				uto		
uthi				uhti		uti	!	uhti
utu				uhtu		utu	!	uhtu
	3		2		Minimal pairs		6	

If the functional load of a contrast is defined (following e.g., Hockett 1955) in terms of the number or frequency of pairs of words differing only in the contrasting elements (here "/h/ vs. ∅" in a particular environment) then a neutralization in one environment can increase the functional load of a contrast in another environment.[5] For example, suppose that in a hypothetical language (figure 7) there are ten words of the form VtV, ten words of the form VthV, and ten words of the form VhtV, and that there are some minimal pairs distinguished only by the presence or absence of /h/ before or after /t/. If /h/ deletion eliminates the contrast "/h/ vs. ∅" after /t/, turning VthV words into VtV words, then there will be twenty words of the form VtV, and an increased number of minimal pairs (six instead of two) distinguished by the presence or absence of /h/ before /t/. This means that the functional load of the contrast "/h/ vs. ∅" before /t/ is greater after the introduction of /h/ deletion after /t/ than it was before. If functional load helps to preserve contrasts, then the contrast "/h/ vs. ∅" before /t/ is less likely to be eliminated now that /h/ deletion is permitted on the opposite side of the consonant.

The data in figure 7 show that for five types of consonants, /h/ deletion occurs either before each consonant (the case with nasals and liquids) or after each consonant (the case with voiceless stops and affricates), depending on where the perceptual distance between /h/ and ∅ is smallest, except in the case of voiceless fricatives. Deletion occurs before *and* after voiceless fricatives.

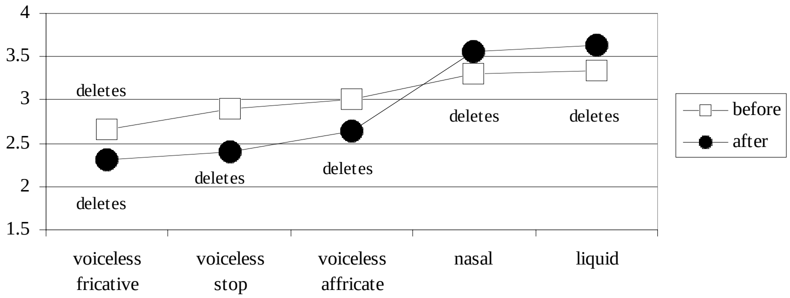


Figure 7. Functional load explains the imperfect mapping from perceptibility to deletion.

The deletion of /h/ before nasals and liquids is difficult to reconcile with the lack of deletion before voiceless stops and affricates (where it is less perceptible), if changes in functional load are not taken into consideration. Deletion that occurs on one side of a consonant inhibits deletion on the other side of that consonant by increasing the functional load of the contrast that would be neutralized. As in the example in table 3, deletion of /h/ after voiceless stops in Turkish increases the frequency of intervocalic stops, and as a result, the

5 A simpler definition of functional load which counts contrasts rather than lexical items leads to the same result in this case.

functional load of the contrast between intervocalic stops and /h/ + voiceless stop clusters (the contrast "/h/ vs. ∅" before voiceless stops) is increased.[6] If functional load and perceptual distance are two factors that cause a contrast to resist deletion, it follows that deletion does not occur when functional load has increased as the result of another deletion unless the perceptual distance of the contrast is sufficiently small. This is the case with voiceless fricatives, the consonant type next to which /h/ is least perceptible, and the only consonant type which conditions /h/ deletion before and after it.

In this section it has been shown how functional load can influence contrast. Increased functional load is associated with the maintenance of contrast, and likewise, the lack of functional load is associated with the neutralization of contrast, which in turn leads to an increase in the functional load of another contrast which involves the output of the neutralization, and can potentially block the neutralization of the second contrast.

6. Conclusion

This paper has shown how contrast in phonological systems shapes language-specific perception through acoustic cues, and how the functional load and perceptual distance of a contrast interact to determine whether the contrast will be maintained or neutralized. The aspects of the relationship between perceptual salience and contrast that have been illustrated are shown in figure 8.

The results of Mielke's (2003) perception experiment show that perceptual distance (the product of robust acoustic cues) leads to the presence or maintenance of contrast. Likewise, the lack of robust acoustic cues leads to perceptual similarity, which leads to the lack or loss of contrast. The linear regression of sensitivity values in terms of acoustic cues shows that phonological contrast leads to attention to acoustic cues, which in turn leads to perceptual distance. Finally, the mismatch between raw d' values and Turkish /h/ deletion illustrate how another language-specific consideration, functional load, can influence

6 /h/ deletion before sonorant consonants is not truly neutralizing, since non-prevocalic /h/ deletion results in compensatory vowel lengthening. For this reason, it might be expected to have less of an impact on the functional load of the contrast "/h/ vs. ∅" after sonorant consonants, and less of an inhibitory effect on /h/ deletion in this context. Since /h/ is so salient after sonorant consonants, deletion would not be expected anyway, regardless of the impact of compensatory lengthening on the functional load of the contrast.

contrast, and allows deletion in one environment to block deletion in another, by increasing its functional load as deletion occurs. This sketch of the interplay between functional considerations has only scratched the surface of the body of system-external language-specific and universal factors which can be drawn upon for explanation of phonological phenomena.

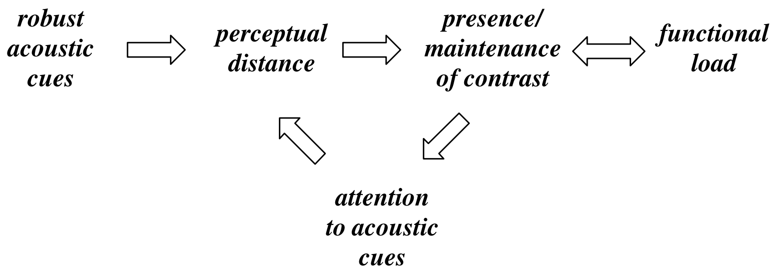


Figure 8. Interplay between perceptual salience and contrast

Acknowledgements

This research has been supported in part by a pre-dissertation research travel grant from The Ohio State and has benefited from comments from Elizabeth Hume, Keith Johnson, Mary Beckman, Dave Odden, Tsan Huang, Misun Seo, Giorgos Tserdanelis, Pauline Welby, Steve Winters, Matt Makashay, Peggy Wong, Robin Dodsworth, Shelome Gooden, Lena Ovcharova, Andrea Sims, and Donca Steriade. The French and Arabic parts of the experiment were made possible by the help of Nick Clements and Annie Rialland.

References

Bladon, A.
1986 Phonetics for hearers. In G. McGregor (ed.), *Language for Hearers*, 1–24. Oxford: Pergamon.

Bybee, J.
2001 *Phonology and Language Use*. Cambridge University Press.

Chang, S., M. Plauché, and J. Ohala
2001 Markedness and consonant confusion asymmetries. In E. Hume and K. Johnson (eds.), *The Role of Speech Perception in Phonology*, 79–101. New York: Academic Press.

Flemming, E.
2002 [1995]. *Auditory Representations in Phonology*. New York: Routledge.

Flemming, E.
2004 Contrast and perceptual distinctiveness. In B. Hayes, R. Kirchner, and D. Steriade (eds.), *The Phonetic Bases of Markedness*, 232–276. Cambridge University Press.

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: Part 4. 337–46.

Green, D. M., and J. A. Swets
1966 *Signal Detection Theory and Psychophysics.* New York: Wiley.

Harrell, R. S.
1966 *A Dictionary of Moroccan Arabic: Arabic-English*. Washington, D. C.: Georgetown University Press.

Hockett, C. F.
1955 *A Manual of Phonology.* Baltimore: Waverly Press, Inc.

Hume, E.
2006 Language specific and universal markedness: An information-theoretic approach. LSA Annual Meeting. Colloquium on Information Theory and Phonology.

Hume, E., and K. Johnson
2001 A model of the interplay of speech perception and phonology. In E. Hume and K. Johnson (eds.), *The Role of Speech Perception in Phonology*, 3–26. New York: Academic Press.

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

Kavitskaya, D.
2002 *Compensatory Lengthening: Phonetics, Phonology, Diachrony.* New York: Routledge.

Kohler, K.
1990 Segmental reduction in connected speech: Phonological facts and phonetic explanations. In W. J. Hardcastle and A. Marchal (eds.), *Speech Production and Speech Modeling*, 69–92. Dordrecht: Kluwer Academic Publishers.

Kornrumpf, H.-J.
1979 *Langenscheidt's Universal Dictionary: Turkish-English, English-Turkish*. Berlin and München: Kangenscheidt.

Lewis, G. L.
1967 *Turkish Grammar.* Oxford: Clarendon Press.

MacMillan, N. A., and D. Creelman
1991 *Detection Theory: A User's Guide.* New York: Cambridge.

Martinet, André
1955 *Économie des Changements Phonétiques.* Bern: Francke.

Mielke, J.
2002a Formalizing the perception-phonology interaction: Looking for evidence of the P-map in speech style variation. Paper presented at the Montreal-Ottawa-Toronto Phonology Workshop, Montréal.

Mielke, J.
2002b The diachronic influence of perception: Experimental evidence from Turkish. *Proceedings of BLS 29.*

Mielke, J.
2003 The interplay of speech perception and phonology: Experimental evidence from Turkish. *Phonetica* 60.3: 208–229.

Oflazer, K.
1994 *PC-Kimmo Turkish Lexicon*. European Corpus Initiative.

Ohala, J.
1992 Alternatives to the sonority hierarchy for explaining segmental sequential constraints. *Chicago Linguistics Society: Papers from the Parasession on the Syllable*, 319–338. Chicago: CLS.

Pickles, J. O.
1988 *An Introduction to the Physiology of Hearing*. San Diego: Academic Press.

Sezer, E.
1986 An autosegmental analysis of compensatory lengthening in Turkish. In L. Wetzels and E. Sezer (eds.), *Studies in Compensatory Lengthening*, 227–250. Dordrecht: Foris Publications.

Silverman, D.
1997 [1995]. *Phasing and Recoverability*. New York: Garland Press.

Steriade, D.
1997 Phonetics in phonology: The case of laryngeal neutralization. Ms., UCLA.

Steriade, D.
2001 Directional asymmetries in place assimilation: A perceptual account. In E. Hume and K. Johnson (eds.) *The Role of Speech Perception in Phonology*, 219–250. New York: Academic Press.

Steriade, D.
2004 The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In K. Hanson and S. Inkelas (eds.), *The Nature of the Word: Essays in Honor of Paul Kiparsky*. Cambridge, Mass.: MIT Press.

Surendran, D., and P. Niyogi
2003 Measuring the functional load of phonological contrasts. Tech. Report TR-2003–12, Univ. of Chicago Comp. Sci. Dept.

Valdman, A.
1976 *Introduction to French Phonology and Morphology*. Rowley, MA: Newbury House Publishers.

Winer, B. J.
1971 *Statistical Principles in Experimental Design*. New York: McGraw-Hill.

Wright, R.
1996 Consonant clusters and cue preservation in Tsou. PhD dissertation, UCLA.

Wright, R.
2001 Perceptual cues in contrast maintenance. In E. Hume and K. Johnson (eds.), *The Role of Speech Perception in Phonology*, 251–277. New York: Academic Press.

Zawadowski, Y. N.
1978 *The Maghrib Arabic Dialects*. Moscow: Central Department of Oriental Literature.

Self-organization through misperception: Secondary articulation and vowel contrasts in language inventories[1]

Alexei Kochetov

1. Introduction

Languages that maintain distinctive secondary articulation contrasts tend to avoid multiple vowel contrasts, particularly rounding contrasts in front and back vowels. At the same time, languages with complex vowel inventories very rarely show distinctions in secondary consonant articulations, for example, in palatalization or labialization. These observations are based both on an analysis of the UPSID Database (Maddieson & Precoda 1992) and on an examination of inventories of a number of languages of Europe that exhibit at least one of the above mentioned contrasts. In this paper I provide an explanatory account of these co-occurrence restrictions on seemingly unrelated segments and derive the two mutually exclusive patterns through a learning simulation. I demonstrate that these markedness effects emerge naturally from low-level interactions between a speaker and a listener/learner as a result of limits on what can be successfully transmitted through the speech communication channel. The key factor in the process is the failure on the part of the listener to correctly process overlapped gestures that happen to share the same articulator. The results suggest that physical limitations on production and perception of speech sounds play an important role in the emergence of common systems of phonological contrasts.

1 This research is supported by a Post-doctoral Fellowship from the Social Sciences and Humanities Research Council of Canada (SSHRC), #756–2001–0145, held at Haskins Laboratories, New Haven, CT. The initial survey of sound inventories was conducted as a part of the Project on Contrast in Phonology supported by SSHRC grants #410–99–1309 and #410–96–0842 to Elan Dresher and Keren Rice, University of Toronto. I am grateful to Louis Goldstein for insightful comments on general theoretical issues that have influenced this work. I am also thankful to Jaye Padgett, anonymous reviewers, and the editors of the volume for helpful comments and suggestions on earlier versions of the paper. *All errors are my own.*

2. Observations

In this paper I focus on two types of contrasts: the high vowel contrasts that involve front/back and rounded/unrounded dimensions (e. g., /i/ vs. /y/ and /ɯ/ vs. /u/), and contrasts in secondary articulations in consonants, "plain" versus palatalized (Cʲ), velarized (Cˠ), or labialized (Cʷ). Both types of contrasts are known to be marked. For instance, the vowels /y/ and /ɯ/ are less common in world languages than /i/ and /u/; so are consonants with distinctive secondary articulation (Maddieson 1984: 124–125; 38). The presence of the marked segments (e. g., front rounded /y/ and palatalized dental/alveolar /tʲ/) often implies the presence of the unmarked ones (front unrounded /i/ and non-palatalized dental/alveolar /t/). What is interesting, however, is that the two types of contrasts very rarely co-occur in language inventories.

2.1. UPSID

An analysis of the UPSID Database (UCLA Phonological Segment Inventory Database: Maddieson & Precoda 1992; 451 languages) shows that languages tend to maintain either distinctive secondary articulation contrasts in stops or rounding and backness contrasts in high vowels. At the same time, languages that contrast unrounded and rounded vowels of the same tongue position (e. g., /i/ vs. /y/ and /y/ and /ɯ/) very rarely show distinctions in secondary consonant articulations.

The database contains 134 languages that have at least one of the following segments: high vowels /y/ or /ɯ/ contrastive with their unmarked counterparts /i/ and /u/; palatalized, velarized, and labialized stops of any place of articulation. These languages are listed in (I) in the Appendix. Note that Maddieson & Precoda (1992) list only "true palatalized" segments, that is, those characterized by a simple addition of a secondary palatal approximant-like constriction and no modification of the primary place, such as dental/alveolar /tʲ dʲ/ (Maddieson 1984: 166–167). Thus the palatal stops /c ɟ/ and post-alveolar affricates /tʃ dʒ tɕ dʑ/ are not listed there, even though in a given language they may pattern together phonologically with other palatalized consonants (e. g. /pʲ/ or /kʲ/).

Out of 134 languages, 81 (60 %) have consonants with secondary articulation at least at one place of articulation, but allow neither /y/ nor /ɯ/ (in addition to /i/ and /u/). Another 47 languages (35 %) have the vowels of interest but no secondary articulation distinction in the consonants. Only 6 languages (4 %) have both types of marked contrasts. Thus Mari and Selkup are listed in

UPSID as having palatalized stops (dentals/alveolars and/or labials) together with the high front rounded /y/. The other 4 languages (Highland Chinantec, Lue, Mbabaram, and Kawaiisu) have labialized stops of only one place of articulation together with high back unrounded /ɯ/. It should be noted, however, that the status of palatalized consonants in Mari and Selkup is not completely clear, since other sources do not mention these segments in inventories of these languages (Vinogradov 1966b, Vinogradov 2001).

Overall, the set of languages with contrastive secondary articulations hardly "overlaps" with the set of those that distinguish backness and rounding contrasts. Exceptions seem to be limited to the languages that exhibit less robust, marginal contrasts in secondary articulation.

2.2. Languages of Europe

In order to further test these observations I turn to languages of Europe since many of these are known to have a greater than average number of vowels (Maddieson 1984: 128) as well as complex consonant inventories. Many of these languages have plain-palatalized contrasts in stops or/and front rounded vowels. The survey reported here is based primarily on the following sources: Ball & Fife (1993), Comrie & Corbett (1993), Harris & Vincent (1988), Iartseva (1993), Iartseva (1997), König & Van der Auwera (1994), MacAulay (1992), Vinogradov (1966 a, b), and Vinogradov (2001). A list of 46 languages that exhibit the above mentioned contrasts is given in (II) in the Appendix.[2] Where the status of palatalized labials and velars in a language is disputed, these consonants are listed in parentheses. In a number of languages, palatalized counterparts of plain dentals/alveolars are realized as alveolo-palatal /tɕ dʑ/ or palato-alveolar affricates /tʃ dʒ/ and, in some cases, as palatal stops /c ɟ/. These are also given in parentheses.

The results are very similar to our findings based on UPSID.[3] 46 % of the languages (22 languages) exhibit secondary palatalization but have no front rounded vowels. In almost half of these languages the contrast between plain and palatalized consonants is fairly robust, extending to two or three places of articulation: labial, coronal, and velar. 46 % of the sample languages (22

2 15 of these languages are also listed in the UPSID database.

3 While it is clear that many of the language characteristics described below are results of contact-induced changes (see, for example, Jakobson 1971 on the "Eurasian *Sprachbund*"), certain linguistic constraints are apparently at play and determine what types of contrasts can freely co-exist in a language inventory.

languages) exhibit the opposite pattern: front rounded vowels /y/, /ø/, or /œ/ occurring "at the expense of" secondary palatal articulation.[4] Only in 8 % of the languages (4 languages) do palatalized consonants and front rounded vowels co-occur: Estonian, Karelian, Veps, and Chuvash. The first three of these languages belong to the Baltic group of the Finno-Ugric family, and they all exhibit palatalized coronal stops in addition to front rounded vowels /y/ and /œ/. Veps and some dialects of Karelian are also reported to have palatalized labials and velars (Iartseva 1993, Vinogradov 1966b); however, it is not clear from the sources whether these are phonemically contrastive (at least in Veps). There are also distributional restrictions on both palatalized coronals and front rounded vowels: for instance, in Estonian /tʲ dʲ/ do not contrast word-initially, and /y/ and /œ/ occur only in initial syllables. It should be noted that in many of the languages with front rounded vowels these segments often participate in the processes of palatal vowel harmony (e. g., Finnish or Tatar) or umlaut (e. g., German or Icelandic). Interestingly, vowel harmony in the languages with palatalized segments, Estonian and Veps, is no longer productive (Iartseva 1993).

Two Turkic languages, Karaim and Gagauz, are listed both with "front rounded vowel languages" and with "palatalization languages." This is done because different dialects of these languages exhibit one of the two types of contrasts: either complex vowel contrasts or secondary palatal articulation contrasts (Comrie 1981: 63–64, Iartseva 1997, Vinogradov 2001). It is interesting that palatal vowel harmony in some dialects corresponds to consonant secondary articulation harmony (e. g., Karaim *køzymde* ~ *kʲozʲumʲdʲa* 'in my eye': Comrie 1981: 63). Yiddish presents another interesting case: the lack of front rounded vowels sets it apart from other Germanic languages. The absence of these segments correlates with the phonemic status of palatalized sonorants /nʲ/ and /lʲ/, and in some dialects with palatalized dentals /tʲ dʲ/ (König & Van der Auwera 1994; Vinogradov 1966a).

To summarize, there is a strong tendency for languages to avoid having both distinctive secondary articulation contrasts and multiple distinctions in rounding/backness, and for languages with multiple vowel contrasts to avoid distinctions in secondary consonant articulations. The main question is: why are these two seemingly unrelated contrasts incompatible? In the rest of this paper I will provide an explanatory account of this phenomenon.

4 These languages seem to have a higher number of basic vowels and diphthongs in general.

3. Sources of explanation

One approach to phonological universals assumes that all markedness effects are innate, pre-specified in Universal Grammar. Thus the facts of incompatibility of the two marked contrasts have to be built into either harmonic rankings of constraints (Optimality Theory; Prince & Smolensky 1993) or universal phonological representations (e. g., Clements 1985). In this paper I consider an alternative approach that argues that these markedness effects arise due to lower-level factors – primarily due to the limitations of speech production and perception.

This view builds in part on existing work investigating the role of low-level articulatory and perceptual factors in shaping phonological structure (Ohala 1981, Kawasaki 1982, Browman & Goldstein 1986, 1999, Hume & Johnson 2001, Pierrehumbert et al. 2001, among others). It also crucially relies on the concept of *self-organization*, or spontaneous emergence of order that is characteristic of various natural dynamic systems (see, for example, Kauffman 1995). Some recent work in the fields of Artificial Intelligence and Artificial Life has demonstrated that complex structures and high-level ontologies can emerge due to low-level sensory-motor interactions of simple autonomous entities – robots or simulated agents. Significantly, this is achieved without any prior specification for this higher-level knowledge (see Pfeifer & Scheier 2001 for a review; see also Brooks 1991, Langton 1995, and Steels 1995).

The self-organization approach, extended to phonology, can be stated as follows: high-level phonological structure – phonological markedness effects – can result from low-level speaker-listener interactions without being directly specified in Universal Grammar. A simplified version of these interactions can be seen as production and perception of lexical items (sensory-motor coordination) and certain kinds of higher-level processing of the perceived input (categorization and generalization). In this approach markedness effects take on a different meaning. "Phonologically unmarked" can be understood as stable with respect to either production, or perception, or higher-level processing, or, in dynamic terms, an equilibrium position. "Phonologically marked" would mean unstable with respect to either production, or perception, or higher-level processing, or a non-equilibrium position. Note that the notion of marked or unmarked may reflect a combined effect or interaction of these kinds of factors. Over time languages tend to retain stable, unmarked, phonological structures and discard the structures that are less stable, or marked.

Returning to the problem in question, how can we explain the apparent incompatibility of complex vowel contrasts and secondary articulation contrasts? It is hypothesized here that a grammar that allows both types of contrasts is highly unstable with respect to production and perception. That is, either the speaker's articulation of these contrasts or the listener's perception of them, or both these activities have an error rate high enough to affect the transmission of this grammar from the speaker to the listener/learner. Given these natural limitations, the system will easily give way to more stable patterns with either of the two marked types of contrasts.

It is crucial for our analysis that high vowels and secondary articulations in consonants are phonetically related. Both segments involve the same articulators: *tongue body*, which is either fronted (as, e. g., for /i/ and /pʲ/) or backed (as, e. g., for /ɯ/ and /pˠ/); and *lips*, which are rounded (as, e. g., for /y/ and /tʷ/) or unrounded (as, e. g., for /ɯ/ and /tˠ/). As a consequence of this articulatory similarity the corresponding high vowels and secondary articulations are also similar acoustically and perceptually. These factors are built into the simulation discussed below.

4. Simulation

The hypothesis outlined above can be investigated using a computer simulation of speaker-listener/learner interactions where the "speaker" and the "listener" are "agents" or simple autonomous entities. Agent-based programming has been used recently to investigate various emergent phonological phenomena (e. g., gestural phasing: Browman & Goldstein 1999; vowel inventories: de Boer 2000; word pronunciations: Liberman 2002; vowel harmony: Harrison et al. 2002). The simulation presented in this paper is far less elaborate than some of those in the works mentioned above; however, it appears to be adequate to handle the problem at hand.

4.1. A hypothetical language

In order to test whether unmarked patterns can emerge through speaker-listener interactions I intentionally chose a hypothetical language with excessively marked inventories of consonants and vowels. This language employs four consonants that share their primary place and differ in their secondary articulation: palatalized, labio-palatalized, velarized, and labialized (1a). It has four high vowels that are differentiated along the front-back and rounded-unrounded dimensions (1b), thus corresponding to the secondary articulations. Lexical items in this language are limited to the shape C_1VC_2 (where $C_1 = C_2$), giving the total of 16 items (2). Each of the items has a distinct meaning; however, the details of phonological-semantic mapping are not important here. Note that in this lexicon all four consonants and all four vowels are fully contrastive, that is, they occur in all logically possible environments.

(1) a. Consonant inventory: {C^j $C^{\text{ɥ}}$ $C^{\text{ɣ}}$ C^w}
b. Vowel inventory: {i y ɯ u}

(2)

C^jiC^j	C^jyC^j	C^jɯC^j	C^juC^j
$C^{\text{ɥ}}$i$C^{\text{ɥ}}$	$C^{\text{ɥ}}$y$C^{\text{ɥ}}$	$C^{\text{ɥ}}$ɯ$C^{\text{ɥ}}$	$C^{\text{ɥ}}$u$C^{\text{ɥ}}$
$C^{\text{ɣ}}$i$C^{\text{ɣ}}$	$C^{\text{ɣ}}$y$C^{\text{ɣ}}$	$C^{\text{ɣ}}$ɯ$C^{\text{ɣ}}$	$C^{\text{ɣ}}$u$C^{\text{ɣ}}$
C^wiC^w	C^wyC^w	C^wɯC^w	C^wuC^w

Note that "C" in our analysis represents a stop of any place of articulation, since our focus is on the properties (phonetic and phonological) of the secondary articulations rather than differences in the primary place. In addition, omitting the primary place of articulation substantially simplifies modeling of the articulation and perception of the consonants of interest.

4.2. Agent interactions

The interactions involve two agents: an adult agent, Agent A, and a learning agent, Agent B (Figure 1). Each of the agents consists of the following components, or modules: production, perception, and lexicon/grammar.

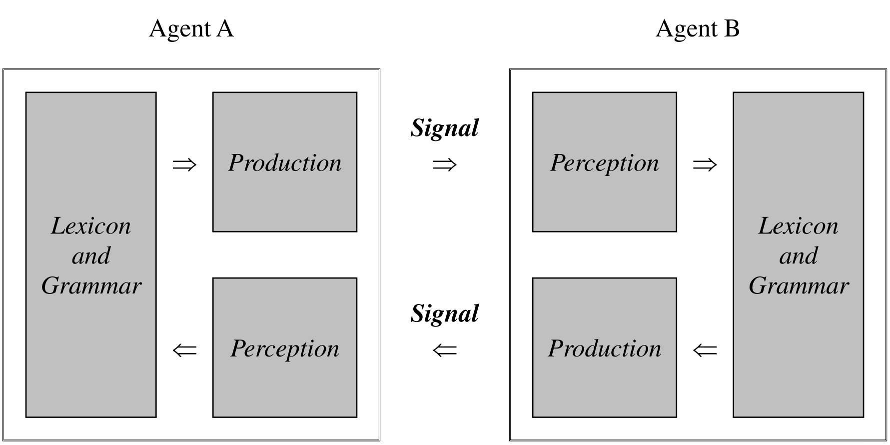


Figure 1. Speaker-listener interactions

In brief, an interaction between the agents proceeds as follows: Agent A picks up a lexical item from the lexicon and produces it as sequences of overlapping articulatory targets (as described below). The acoustic signal resulting from the production is presented to the listener/learner, Agent B. Whether correctly recovered or misidentified, the items are stored in the learner's "lexicon". Further generalizations across the recovered items and the inventory in general are also assumed, but are not implemented in the current simulation. Subsequently, Agent B produces an item from his/her lexicon and adjusts the item's representation based on the communicative success (see de Boer 2000) and additional tokens of this item. The initial part of the interaction, Agent A's production and Agent B's perception, is of particular interest to us, since it is here where most errors are likely to occur. (This part of the interaction is currently implemented using *Matlab*.)

The details of production and perception modules draw heavily from de Boer's agent-based simulation of emergent vowel inventories (2000). These modules, together with the lexicon/grammar component, are described in more detail below.

4.3. Production module

The production module models targets of vowels and secondary articulations – backness, height, and rounding – based on the articulatory model of Maeda (1989). Each articulatory target is assumed to be [0] or [1], where [1] denotes the targets "front", "high", or "rounded", and [0] denotes the opposite specifications: "back," "low," and "unrounded" (3). An articulation of each segment is modeled as a vector of these numbers. The segments of interest have the same height but contrast in backness and rounding. Vowels and the corresponding secondary articulations are specified the same way.

(3) Vowels and Consonants (secondary articulation)

			Backness	Height	Rounding	
/i/, /C^{j}/	=	[	0	1	0	]
/y/, /$C^{\text{ɥ}}$/	=	[	0	1	1	]
/ɯ/, /$C^{\text{ɣ}}$/	=	[	1	1	0	]
/u/, /C^{w}/	=	[	1	1	1	]

Words are modeled as vectors of articulatory targets for each segment (cf., Liberman 2002), that is, as matrices of the digits 0 and 1. Thus the lexical item /C^{j}uC^{j}/ is represented as in (4).

(4)

$$\begin{array}{lccc} & C^j & u & C^j \\ \text{Backness} & \left[\begin{matrix}0\\1\\0\end{matrix}\right. & \begin{matrix}1\\1\\1\end{matrix} & \left.\begin{matrix}0\\1\\0\end{matrix}\right] \\ \text{Height} & & & \\ \text{Rounding} & & & \end{array}$$

	C^j	u	C^j
Backness	0	1	0
Height	1	1	1
Rounding	0	1	0

It is important to note that representing articulatory targets with discrete values (0 or 1) does not mean that their realization is also discrete. First, achievement of articulatory targets in humans is never perfect, and this fact is captured in the simulation by adding "articulatory noise," random fluctuations within the range of ±0.25. In other words, the backness target for /u/, which is specified for [1], can be realized during the production as any value between [0.75] and [1]; similarly, the same parameter for /i/, which is specified for [0], can be realized as any value between [0] and [0.25]. Second, articulatory gestures involved in the production of lexical items are subject to overlap, or co-production (Browman & Goldstein 1986), which tends to result in an "undershoot" of their targets (Lindblom 1963, 1989). This is particularly true when two almost simultaneously activated gestures have conflicting targets, such as tongue body backing for /u/ ([1]) and tongue body fronting ([0]) for /C^j/ in C^juC^j. Thus the main cause of undershoot is purely dynamic: there are physical limits on how well articulators can attain their targets.

This view of gestural overlap is consistent with phonetic accounts of languages with secondary articulations: an achievement of the secondary articulation targets leads to a remarkably different quality of adjacent vowels (e.g., Russian: Bolla 1981, Kochetov 2002; Marshallese: Choi 1992; Irish: Ó Dochartaigh 1992; cf. Ladefoged & Maddieson 1996: 354–366). Thus, /u/ is substantially fronted after palatalized consonants and /i/ is backed after velarized segments. The reverse is often observed in languages with multiple front-back vowel contrasts: an attainment of vowel targets results in allophonic velarization or palatalization of adjacent consonants (e.g., Turkic languages: Comrie 1981: 63).

In this model it is assumed that when two gestures with conflicting articulatory targets overlap, only one of them achieves the target completely (an undershoot rate u = 0), while the target of the second gesture is always undershot.[5] The degree of undershoot is set up to be 0, 0.25, and 0.50. The first one (u = 0) is

5 Although this "either … or" interpretation of undershoot involves a certain oversimplification, it seems to be a plausible approximation given the language data reported in the previous paragraph. A more realistic model should allow for degrees of undershoot of both targets, or "blending" of gestures (as in GEST, a computational gestural model: Browman & Goldstein 1990). It is still an empirical question, however, which targets are undershot more than others in a given language.

highly unlikely to be observed in natural speech; it is used in the simulation as a starting point. The other two degrees of overlap are likely to be more typical, at least of casual and fast speech.[6]

A 25 % undershoot of the vowel target of /C^juC^j/ is shown in (5a). Thus, backness and rounding parameters for /u/ are reduced from [1] to [0.75], since the near-simultaneous secondary articulation targets are specified for [0] (the first and third rows). There is no reduction in height, since all three targets are specified for [1]. The same degree of undershoot of the consonant secondary articulation targets is shown in (5b), where we can see a 25 % shift to a more back and rounded articulation of /C^j/.[7]

(5) a.

$$\begin{array}{ll} & \begin{array}{ccc} C^j & u & C^j \end{array} \\ \begin{array}{l} \text{Backness} \\ \text{Height} \\ \text{Rounding} \end{array} & \begin{bmatrix} 0 & 0.75 & 0 \\ 1 & 1 & 1 \\ 0 & 0.75 & 0 \end{bmatrix} \end{array}$$

b.

$$\begin{array}{ll} & \begin{array}{ccc} C^j & u & C^j \end{array} \\ \begin{array}{l} \text{Backness} \\ \text{Height} \\ \text{Rounding} \end{array} & \begin{bmatrix} 0.25 & 1 & 0.25 \\ 1 & 1 & 1 \\ 0.25 & 1 & 0.25 \end{bmatrix} \end{array}$$

4.4. Signal

An acoustic signal resulting from the production is calculated based on Vallée 1994 (as reported in de Boer 2000). Only first and second formants are used in the analysis (F1 and F2, in Hertz). These formants for our vowels and secondary articulations are shown in (6). In order to ensure acceptable perceptual quality all lexical items were synthesized using *Synthworks*, an acoustic synthesizer.

(6) Formants of vowels and secondary articulations (Hertz)

			F1	F2	
/i/, /C^j/	=	[	252	2202	]
/y/, /$C^{ɥ}$/	=	[	250	1878	]
/ɯ/, /$C^{ɣ}$/	=	[	305	1099	]
/u/, /C^w/	=	[	276	740	]

---

6 It is well established that gestures tend to show more overlap and thus more undershoot in fast speech (e. g., Byrd 1992, Lindblom 1989, Perrier et al. 1996).

7 Note that the numbers actually generated by the production module will not be the same due to the addition of articulatory noise.

Acoustic noise, as random fluctuations of formants within certain ranges (±100 Hz for F1 and ±200Hz for F2), is added to the signal. Adding noise is expected to make the learning situation closer to real-life acquisition, where lexical items are hardly ever acquired in complete silence.

4.5. Perception module

The resulting signal is presented to Agent B, the listener/learner. The listener's recovery of items from the signal involves extracting formants at 3 points in time (the onset, midpoint, and offset of the vowel), converting them to an auditory scale (in Barks; see de Boer 2000 for details), and matching the output to the available vowel and consonant categories, shown in (7). This is achieved by calculating a Euclidean distance from each of the categories. For example, if a part of the signal is identified as having the values F1 = 3.08 Barks and F2 = 9.98 Barks, it is labeled as /ɯ/ (or /$C^{\text{ɣ}}$/), since this category is the closest to the recovered signal (a distance of 0.48 Barks; compared to 1.54 Barks for /y/ or /$C^{\text{ɥ}}$/, 1.76 Barks for /u/ or /C^{w}/, and 2.09 Barks for /i/ or /C^{j}/).

(7) Formants of vowels and secondary articulations (Barks)

			F1	F2	
/i/, /C^{j}/	=	[	2.52	13.65	]
/y/, /$C^{\text{ɥ}}$/	=	[	2.50	12.59	]
/ɯ/, /$C^{\text{ɣ}}$/	=	[	3.08	9.10	]
/u/, /C^{w}/	=	[	2.78	6.82	]

Obviously, the categories are not perceptually equidistant. While /i/ and /ɯ/ are fairly close to their corresponding rounded vowels, /y/ (0.58 Barks) and /u/ (1.29 Barks), the distance between, for example, /y/ and /ɯ/ or /u/ is substantially higher (2.00 and 3.17 Barks, respectively). For simplicity, the shape of lexical items CVC, where C1 = C2, is assumed to be known to the learner.

4.6. Lexicon and grammar

The limitations of articulation – overlap of gestures, with additional articulatory and acoustic noise – have important consequences for perception and, ultimately, for the lexicon and the grammar. Lexical items produced by the speaker with undershot targets may not always be perfectly perceived by the listener/learner. As a result of perceptual confusion, some items will end up being represented in the lexicon of Agent B differently from those of Agent A. As discussed in the next section, an instance of $C^{j}uC^{j}$ with vowel undershoot can be interpreted as /$C^{j}uC^{j}$/, /C^{j}ɯC^{j}/, etc.; while the same item with a consonant undershoot is likely to be perceived as /$C^{j}uC^{j}$/, /$C^{ɥ}uC^{ɥ}$/, etc.

All tokens of a particular lexical item are temporarily stored in the lexicon and are used in the calculation of its abstract representation. This representation consists of the segments most frequently occurring in the stored tokens. Suppose that tokens for a particular lexical item vary in the quality of the vowel: 65 out of 100 tokens have /u/, 25 have /y/, and 10 have /ɯ/. The most common vowel among these, /u/, will be the one used in the lexical representation. It is assumed here that the agent's grammar, or rankings of constraints, is constructed based on the acquired lexical items. The mechanism of this ranking is not explored in the simulation (see Kochetov 2002).

5. Results

In this section I describe the results of the simulation by examining the results under the conditions of three different degrees of undershoot. The first case involves no overlap of vocalic targets and thus no undershoot. As already mentioned, this is an unrealistic situation, but it serves as a baseline for the other cases. The second case involves a certain degree of overlap of targets, and as a consequence, an undershoot of one of them by 25 %. The third case presents a substantial overlap of the targets which leads to a 50 % undershoot of one of them. This degree of undershoot is likely to be typical of fast casual speech. A sample run, perception of the item /$C^{j}uC^{j}$/ based on 100 produced tokens, is presented in (III) in the Appendix.

In each case, undershoot of vowels and secondary articulations is considered separately. Recall that in each case the goal of the learner, Agent B, is to build a lexicon based on perceived tokens. This lexicon may or may not turn out to be identical to the lexicon of Agent A.

5.1. No undershoot

Running the simulation with no undershoot of targets results in a very high degree of success on the part of Agent B in replicating the target lexicon (see (2)). There is a very high probability that all the lexical items are perceived and stored correctly. We can see this from the sample run for /$C^{j}uC^{j}$/ in (III) in the Appendix. In a few instances, the listener confuses perceptually similar vowels

(/i/ and /y/; /ɯ/ and /u/) and similar secondary articulations (/C^{j}/ and /C^{ɥ}/; /C^{ɣ}/ and /C^{w}/). Yet given a high number of presented tokens per each word (100), the errors are unlikely to influence the learner's choice of the underlying form. Given these perceptual results, Agent B will posit the underlying form /C^{j}uC^{j}/, which is identical to that of Agent A.[8]

Overall, the "perfect" production ensures the near-perfect transmission of the lexicon from Agent A to Agent B. I now turn to a more realistic production that involves overlap of gestures and undershoot of targets.

5.2. Undershoot of 25 %

The results show that a 25 % undershoot of all vowel targets has important consequences for perception. I first consider the situation when the vowel target is undershot, while the consonant secondary articulation target is fully achieved. Under these circumstances, the back rounded /u/ between secondary front articulations, /C^{j}/ and /C^{ɥ}/, is perceived by Agent B more often as the back unrounded /ɯ/, rather than /u/ (see (III) in the Appendix). This is shown by the rightmost arrows in (8). In other words, the original lexical items /C^{j}uC^{j}/ and /C^{ɥ}uC^{ɥ}/ (shaded) are identified as homophonous to the original items /C^{j}ɯC^{j}/ and /C^{ɥ}ɯC^{ɥ}/. Similarly the front unrounded /i/ between secondary back articulations, /C^{ɣ}/ and /C^{w}/, is perceived most often as the front rounded /y/, rather than as /i/. This is shown by the leftmost arrows in (8). Given this tendency, the most likely lexicon of Agent B will fail to distinguish between /i/ and /y/ and between /ɯ/ and /u/ in certain environments, leading to the virtual reduction of the vowel contrasts from 4 to 3. At the same time the contrast in secondary articulations remains intact.

(8)

C^{j}iC^{j}	C^{j}yC^{j}	C^{j}ɯC^{j} ←	*C^{j}uC^{j}
C^{ɥ}iC^{ɥ}	C^{ɥ}yC^{ɥ}	C^{ɥ}ɯC^{ɥ} ←	*C^{ɥ}uC^{ɥ}
*C^{ɣ}iC^{ɣ} →	C^{ɣ}yC^{ɣ}	C^{ɣ}ɯC^{ɣ}	C^{ɣ}uC^{ɣ}
*C^{w}iC^{w} →	C^{w}yC^{w}	C^{w}ɯC^{w}	C^{w}uC^{w}

The second situation involves a full achievement of vowel target while the consonant secondary articulation target is undershot by 25 %. In this case Agent B fails to correctly identify secondary palatal articulations in the environment of back vowels: /C^{j}/ is commonly perceived as /C^{ɥ}/ (as shown by the rightmost

8 Note that different outcomes (e. g., /C^{j}ɯC^{j}/, /C^{ɥ}uC^{ɥ}/, or /C^{ɥ}ɯC^{ɥ}/) are also technically possible but only when the lexical form is based on a very small number of tokens.

arrows in (9)). The same applies to the secondary labial articulation in the environment of front vowels (as shown by the leftmost arrows). The resulting lexicon (9) will distinguish between 4 vowels and will fail to distinguish between /C^{j}/ and /C^{ɥ}/ and between /C^{w}/ and /C^{ɣ}/ in certain contexts, thus leading to the reduction of consonant contrasts from 4 to 3.

(9)

C^{j}iC^{j}	C^{j}yC^{j}	*C^{j}ɯC^{j}	*C^{j}uC^{j}
C^{ɥ}iC^{ɥ}	C^{ɥ}yC^{ɥ}	C^{ɥ}ɯC^{ɥ}	C^{ɥ}uC^{ɥ}
C^{ɣ}iC^{ɣ}	C^{ɣ}yC^{ɣ}	C^{ɣ}ɯC^{ɣ}	C^{ɣ}uC^{ɣ}
*C^{w}iC^{w}	*C^{w}yC^{w}	C^{w}ɯC^{w}	C^{w}uC^{w}

5.3. Undershoot of 50 %

Now we will see how increasing the degree of overlap and the degree of undershoot of targets may further affect perception and the resulting lexicon.

The results show that a 50 % undershoot of the vowel targets leads to a higher perceptual error rate than in the previous case and thus to a lexicon dramatically different from the original one. The most likely outcome is shown in (10). Directions of mis-identifications are shown by arrows; mis-identified items are shaded. Note that the front vowels /i/ and /y/ in (10) are in complementary distribution, with /i/ occurring only between palatalized consonants. The back vowels /ɯ/ and /u/ are also in complementary distribution, with /u/ restricted to the environment between labialized consonants. Interestingly, /u/ between palatalized consonants is often considered perceptually similar to /y/ (see (III) in the Appendix).[9] Similarly, we find frequent perception of /i/ between labialized or velarized consonants as /ɯ/.

Overall, the contrasts in vowels are reduced to the distinction between [front] and [back]. All four consonants are found in the lexicon, although with certain positional restrictions.

(10)

C^{j}iC^{j}	*C^{j}yC^{j}	C^{j}ɯC^{j}	*C^{j}uC^{j}
*C^{ɥ}iC^{ɥ}	C^{ɥ}yC^{ɥ}	C^{ɥ}ɯC^{ɥ}	*C^{ɥ}uC^{ɥ}
*C^{ɣ}iC^{ɣ}	C^{ɣ}yC^{ɣ}	C^{ɣ}ɯC^{ɣ}	*C^{ɣ}uC^{ɣ}
*C^{w}iC^{w}	C^{w}yC^{w}	*C^{w}ɯC^{w}	C^{w}uC^{w}

9 This perceptual similarity explains the fact that French and German sequences /C/+/y/ are often adapted in Russian as the sequence /C^{j}/+/u/ (Avanesov 1972). See also the example from Karaim in Section 1.2.

Ultimately, the grammar based on these lexical forms would maintain the contrast between multiple secondary articulations {C^{j} $C^{ɥ}$ $C^{ɣ}$ C^{w}} (although limited positionally) and differentiate vowels only based on the front/back dimension, {y ɯ} or {i u}.

The same degree of undershoot of secondary articulation targets results in the contrast in vowels being fairly well maintained, while the contrast between the consonants becomes highly restricted (11). There is no contrast between secondary rounded and unrounded articulations for both front (/C^{j}/ vs. /$C^{ɥ}$/) and back (/$C^{ɣ}$/ vs. /C^{w}/) tongue positions. The quality in terms of rounding/unrounding of a consonant is predictable from the neighboring vowel environment: /C^{j}/ occurs only in the context of /i/ and /$C^{ɥ}$/ is found elsewhere. Similarly, /C^{w}/ occurs in the environment of /u/ and /$C^{ɣ}$/ is found in all the other vowel environments.

(11)

C^{j}iC^{j}	*C^{j}yC^{j}	*C^{j}ɯC^{j}	*C^{j}uC^{j}
*$C^{ɥ}$i$C^{ɥ}$	$C^{ɥ}$y$C^{ɥ}$	$C^{ɥ}$ɯ$C^{ɥ}$	$C^{ɥ}$u$C^{ɥ}$
$C^{ɣ}$i$C^{ɣ}$	$C^{ɣ}$y$C^{ɣ}$	$C^{ɣ}$ɯ$C^{ɣ}$	*$C^{ɣ}$u$C^{ɣ}$
*C^{w}iC^{w}	*C^{w}yC^{w}	*C^{w}ɯC^{w}	C^{w}uC^{w}

Thus the grammar constructed based on this lexicon would differentiate a full range of vowel contrasts {i y ɯ u} (although restricted positionally), and distinguish consonants by their front or back secondary articulation, {$C^{ɥ}$ $C^{ɣ}$} or {C^{j} C^{w}}.

5.4. Summary

The main result of the simulation is that a grammar such as the target grammar in (12a), that allows multiple contrasts in backness and rounding both in vowels and secondary articulations, is highly unstable because it cannot be well replicated by the learner. Recall that perceptual confusion of vowels and secondary articulations distinguished solely by lip rounding is not uncommon even when their targets are fully achieved. This confusion increases substantially in more natural speech, when gestures overlap in time and their targets are undershot. As we saw, there are certain attractors, default states, at which the grammar naturally arrives. The first one, default grammar 1 (12b), allows multiple secondary articulation contrasts at the expense of vowel distinctions. The second grammar, default grammar 2 (12c), limits secondary articulation contrasts, while maintaining multiple vowel distinctions. These more stable grammars are exhibited by the majority of the languages in our typological survey: languages tend to have either contrastive secondary articulations or front/back rounded/unrounded contrasts in vowels. Note that a grammar that limits both secondary articulations (e. g., only "plain" consonants) and vowel contrasts (e. g., only front vs. back distinction) (12d) is likely to be even more stable in terms of production and perception. As we know, this is the state of affairs characteristic for most of the world's languages: 70 % of the UPSID languages have neither (surface) secondary articulation contrasts, nor rounding contrasts in front or back vowels.

(12) a. Target grammar
- Multiple secondary articulation contrasts
- Multiple vowel contrasts

b. Default grammar 1
- Multiple secondary articulation contrasts
- Limited vowel contrasts

c. Default grammar 2
- Limited secondary articulation contrasts
- Multiple vowel contrasts

d. Default grammar 3
- Limited secondary articulation contrasts
- Limited vowel contrasts

This simulation has allowed me to explain one of many phonological markedness phenomena observed in language. A number of questions related to the results require further consideration. First, the lexicon discussed here is a result of initial processing, based on 100 tokens. The learner is likely to restructure this lexicon based on subsequent communication as well as by making certain generalizations over segments and environments. The pressure to avoid homophony, which is as high as 50 % in our case, is also likely to affect the process. Second, it is likely that the choice of segments in variable cases (e. g., /i/ or /y/ and /ɯ/ or /u/) is influenced by other factors, namely, the general preference for /i/ and /u/ over /y/ and /ɯ/, which appear to result from more complex long-term interactions (see de Boer 2000) and possibly from other factors. Third, positing an a priori set of phonemic categories made the simulation more manageable by restricting the choices of the learner. More realistically, it would not be surprising if the vowel and secondary articulation categories constructed by the learner based on the highly variable input were not identical to those of the speaker [10] (see Liberman 2002 on modeling of word pronunciation in populations of agents). Finally, further work should aim to rely on more complex interactions and a more realistic model of human speech production and perception. It should use a wider range of lexical items and give more attention to higher-level processing of the perceived input.

6. Conclusion

In this paper I have attempted to demonstrate that an investigation of low-level speaker-listener interactions provides insight into the causes of phonological markedness (cf. Ohala 1981, Kawasaki 1982, de Boer 2000, among others). Apparent restrictions on co-occurrence of certain vowel and secondary articulation contrasts in language inventories can be generated in a simulated environment with no *a priori* knowledge of markedness. No "innate" restrictions against having both types of contrasts in inventories need to be assumed, since such a system is highly unstable due to limitations on articulation and perception. A language having this system will inevitably "self-organize" by shifting to a more stable pattern: with either rounding contrasts in the vowels, or secondary articulation contrasts in the consonants, or none of these marked contrasts.

References

Avanesov, R. I.
1972 *Russkoe literaturnoe proiznoshenie* [Russian literary pronunciation]. Moscow: Prosveshchenie.

Ball, Martin J. and James Fife (eds.)
1993 *The Celtic Languages*. New York, N. Y.: Routledge.

Bolla, K.
1981 *A Conspectus of Russian Speech Sounds*. Cologne: Bölau.

Brooks, Rodney A.
1991 Intelligence without representation. *Artificial Intelligence Journal* (47): 139–159.

Browman, Catherine and Louis Goldstein
1986 Towards an articulatory phonology. *Phonology Yearbook* 3: 219–252.

10 Thus, a vowel of /C^{j}uC^j/ can be represented not only as invariable /u/, /ɯ/, or /y/, but also as any intermediate values, such as high central vowels /ʉ/ or /ɨ/, or even mid central /ə/ or /ɵ/. The exact quality of the secondary articulation of /C^j/ can also vary, with a "plain" consonant a possible outcome.

Browman, Catherine and Louis Goldstein
1990 Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics* 18: 299–320.
Browman, Catherine and Louis Goldstein
1999 Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée, Grenoble*: 25–34.
Byrd, Dani
1996 Influences on articulatory timing in consonant sequences. *Journal of Phonetics* 24(2): 209–244.
Choi, J. D.
1992 Phonetic underspecification and target-interpolation: An acoustic study of Marshallese vowel allophony. *UCLA Working Papers in Phonetics* 82.
Clements, G. N.
1985 The geometry of phonological features. *Phonology Yearbook* 2: 223–252.
Comrie, Bernard
1981 *The Languages of the Soviet Union*. Cambridge: Cambridge University Press.
Comrie, Bernard and Greville G. Corbett (eds.)
1993 *The Slavonic Languages*. London/New York: Routledge.
de Boer, Bart
2000 Self-organization in vowel systems. *Journal of Phonetics* 28: 441–465.
Harris, Martin and Nigel Vincent (eds.)
1988 *The Romance Languages*. London: Croom Helm.
Harrison, K. David, Mark Dras, and Berk Kapicioglu
2002 Agent-based modeling of the evolution of vowel harmony. In: M. Hirotani (ed.), *Proceedings of the Northeastern Linguistics Society 32*, 217–236. Amherst, MA: GLSA.
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*. New York: Academic Press: 3–26.
Iartseva, V. N.
1993 *Iazyki mira: Ural'skie iazyki* [Languages of the world: Uralic languages]. Moscow: Nauka.
Iartseva, V. N. (ed.)
1997 *Iazyki Rossiiskoi federatsii i sosednix gosudarstv. Enciklopediia* [Languages of the Russian Federation and neighbouring states, Encyclopedia]. Vol. 1. eds. Moscow: Nauka.
Jakobson, Roman
1971 K kharakteristike evraziiskogo yazykovogo soyuza. [Characterizing the Eurasian *Sprachbund*] In: *Roman Jakobson: Selected Writings*, vol. 1: Phonological Studies. The Hague: Mouton: 144–201.

Kauffman, Stuart
1995 *At Home in the Universe: The Search for the Laws of Self-organization and Complexity*. Oxford: Oxford University Press.

Kawasaki, Haruko
1982 An acoustic basis for universal constraints on sound sequences. Doctoral dissertation, University of California, Berkeley.

Kochetov, Alexei
2002 *Production, Perception, and Emergent Phonotactic Patterns: A Case of Contrastive Palatalization*. New York/London: Routledge.

König, Ekkehard and Johan Van der Auwera (eds.)
1994 *The Germanic languages*. New York/London: Routledge.

Ladefoged, Peter and Ian Maddieson
1996 *The sounds of the World's Languages*. Cambridge, MA: Blackwell.

Langton, Christopher G.
1995 *Artificial Life: An Overview*. Cambridge, MA: MIT Press.

Liberman, Mark
2002 Simple models for emergence of a shared vocabulary. Paper presented at the Eighth Conference on Laboratory Phonology (LabPhon VIII), June 27–30, 2002, Yale University and Haskins Laboratories.

Lindblom, Björn
1963 Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35: 1773–1781.

Lindblom, Björn
1989 Phonetic invariance and the adaptive nature of speech. In: B. A. G. Elsendoorn and H. Bouma (eds.), *Working Models of Human Perception*. London: Jovanovich Publishers: 139–173.

MacAulay, Donald (ed.)
1992 *The Celtic Languages*. New York: Cambridge University Press.

Maddieson, Ian
1984 *Patterns of Sounds*. Cambridge: Cambridge University Press.

Maddieson, Ian and Kristin Precoda
1992 UPSID and PHONEME (version 1.1). University of California at Los Angeles.

Maeda, Shinji
1989 Compensatory articulation during speech: Evidence from the analysis and synthesis of tract shapes using an articulatory model. In: W. J. Hardcastle and A. Marchal (eds.), *Speech production and speech modeling*. Dordrecht: Kluwer: 131–149.

Ó Dochartaigh, Cathair.
1992 Irish. In: Donald MacAulay (ed.), *The Celtic Languages*. New York: Cambridge University Press: 11–99.

Ohala, John J.
1981 The listener as a source of sound change. In: C. S. Masek, R. A. Hendrick, and M. F. Miller (eds.), *Papers from Chicago Linguistic Society Parasession on Language and Behavior*. Chicago, IL: CLS: 178–203.

Perrier, P., H. Loevenbruck, and Y. Payan
1996 Control of tongue movements in speech: The Equilibrium Point Hypothesis perspective. *Journal of Phonetics* 24: 53–75.
Pfeifer, Rolf and Christian Scheier
2001 *Understanding Intelligence*. Cambridge, MA: MIT Press.
Pierrehumbert, Janet, Mary E. Beckman, and D. Robert Ladd
2001 Conceptual foundations of phonology as a laboratory science. In: N. Burton-Roberts, P. Carr, and G. Docherty (eds.), *Phonological Knowledge*. Oxford: Oxford University Press: 273–304.
Prince, Alan and Paul Smolensky
1993 Optimality Theory: Constraint interaction in generative grammar. Ms., Rutgers University.
Steels, Luc
1995 Intelligence – dynamics and representations. In: Luc Steels (ed.), *The Biology and Technology of Intelligent Autonomous Agents*. Berlin: Springer-Verlag.
Vallée, N.
1994 Systèmes vocaliques: de la typologie aux prédictions. Thèse préparée au sein de l'Institute Communication Parlée (Grenoble-URA C.N.R.S. no. 368).
Vinogradov, V. A. (ed.)
2001 *Iazyki Rossiiskoi federatsii i sosednix gosudarstv. Enciklopediia* [Languages of the Russian Federation and neighbouring states, Encyclopedia]. Vol. 2. eds. Moscow: Nauka.
Vinogradov, V. V. (ed.)
1966a *Iazyki narodov SSSR*. Vol. 1. *Indoevropeiskie iazyki* [Languages of the USSR, Indo-European languages]. Moscow: Nauka.
Vinogradov, V. V. (ed.)
1966b *Iazyki narodov SSSR*. Vol. 3. *Finno-ugorskie i samodiiskie iazyki* [Languages of the USSR, Finno-Ugric and Samoyed languages]. Moscow: Nauka.

Appendix

(I) UPSID (Maddieson & Precoda 1992) languages that have either of the following: secondary articulation contrasts in stops (e. g. labialized vs. plain or labialized vs. palatalized), rounding contrasts in high vowels (e. g. /y/ and /ɯ/ vs. /i/ and /u/), or both.[11]

In each case, only marked consonant and vowel counterparts are mentioned.

11 Note that it was not possible to determine whether the listed languages had corresponding stop + glide sequences, since the database (Maddieson & Precoda 1990) did not contain information on language phonotactics (see also Maddieson 1994: 166–167). This question has to be addressed in future work.

a. Languages having secondary articulation contrasts in stops but no rounding contrast in high vowels (81):
   – palatalized and labialized/velarized stops: Irish, Lakkia, Kam, Lai, Kabardian, Igbo, Hausa, Tera, Amuzgo, Tsimshian, Nambakaengo;
   – velarized and labialized stops: Chipewyan*;
   – palatalized stops: Lithuanian, Russian, Bulgarian, Saami, Nenets, Resigaro*, Ocaina*;
   – labialized stops: Sui, Lenakel, Pohnpeian, Kwaio, Taishan, Lak, Rutul, Archi, Kpelle, Kohumono, Konyagi, Kolokuma Ijo, Amharic, Awiya, Iraqw, Beja, Ngizim, Dahalo, Hadza, Haida, Tlingit, Navajo, Huave*, Mixtec, Tseshaht, Kwakw'ala, Quileute, Lushootseed, Luiseno, Hopi, Picuris, Diegueno, Zuni, Tonkawa, Wiyot, Wichita, Nahuatl, Bella Coola, Upper Chehalis, Caddo, Huasteco, Shuswap, Southern Nambiquara, Yupik, Kwoma, Guarani, Ticuna, Siona, Iranxe, Tarascan, Warao, Paya, Cuna, Movima, Saliba, Guambiano, Yupik, Kwoma, Dani, Wantoat, Yessan Mayo.

   * languages with /ɯ/ instead of /u/ (i. e., no contrast in rounding).
b. Languages having a rounding contrast in high vowels but no secondary articulation contrast in stops (47):
   – /y/ and /ɯ/ (contrastive with /i/ and /u/): Turkish, Chuvash, Yakut, Korean, Naxi;
   – /y/ (or /ʏ/): Breton, German, Norwegian, French, Albanian, Finnish, Hungarian, Nganasan, Azerbaijani, Kirghiz, Bashkir, Tuva, Dagur, Iai, Mandarin, Changzhow, Fuzhow, Ejagham, Tzeltal, Huari;
   – /ɯ/ (or /ɯ̞/): Khanty, Komi, Vietnamese, Khmer, Parauk, Sre, Nikobarese, Nyah Kur, Bruu, Yay, Lungchow, Bai, Dafla, Ao, Tulu, Aizi, Fe?Fe?, Karib, Apinaye, Jivaro, Araucanian, Panare.
c. Languages having both a secondary articulation contrast in stops and a rounding contrast in high vowels (6):
   – labialized stops, /y/ and /ɯ/: Highland Chinantec (/k^w/);
   – labialized stops and /ɯ/: Lue (/k^w/), Kawaiisu (/k^w/), Mbabaram (/g^w $^ng^w$ /);
   – palatalized stops and /y/: Mari (/p^j b^j t^j/), Selkup (/t^j/).

(II) Table 1. Languages of Europe having either of the following: secondary palatal articulation contrasts in stops (palatalized vs. plain; languages 1–22), rounding contrasts in high or mid vowels (/y/ or /ø/ (/œ/) vs. /i/ and /e/ (/ɛ/); languages 23–44), or both (languages 45–48). In each case, only marked consonant and vowel counterparts are listed. *Notes:* sounds in parentheses = status disputed, marginal, or not realized as "true" palatalized consonants"; the short/lax vs. long/tense distinction is ignored.

	Language	Group, family	Palatalized consonants			Front vowels	
			labial	coronal	velar	high	mid
1	Belorussian	Slavic, IE	p^{j} b^{j}	(ts^{j} dz^{j})	k^{j} g^{j}		
2	Bulgarian	Slavic, IE	p^{j} b^{j}	t^{j} d^{j}	k^{j} g^{j}		
3	Irish	Celtic, IE	p^{j} b^{j}	(tɕ dʑ/tʃ dʒ)	k^{j} g^{j}		
4	Lithuanian	Baltic, IE	p^{j} b^{j}	t^{j} d^{j} (or tʃ dʒ)	k^{j} g^{j}		
5	Roma	Indo-Aryan, IE	p^{j} b^{j}	t^{j} d^{j} t^{hj}	k^{j} g^{j} k^{hj}		
6	Saami (Eastern)	Finno-Ugric, Uralic	p^{j} b^{j}	t^{j} d^{j} t^{jj} d^{jj}	k^{j} g^{j}		
7	Russian	Slavic, IE	p^{j} b^{j}	t^{j} d^{j}	k^{j} g^{j}		
8	Scots G.	Celtic, IE	(p^{j} b^{j})	(tɕ dʑ/tʃ dʒ)	(c ɟ)		
9	Manx	Celtic, IE		(tʃ dʒ)	k^{j} g^{j}		
10	Nenets	Samoyed, Uralic	p^{j} b^{j}	t^{j} d^{j}			
11	Polish	Slavic, IE	(p^{j} b^{j})	(tɕ dʑ)	(k^{j} g^{j})		
12	Upper Sorbian	Slavic, IE	p^{j} b^{j}	(tɕ dʑ)			
13	Lower Sorbian	Slavic, IE	p^{j} b^{j}	(ɕ ʑ)			
14	Liv	Finno-Ugric, Uralic		t^{j} d^{j}			
15	Erzya Mordva	Finno-Ugric, Uralic		t^{j} d^{j}			
16	Moksha Mordva	Finno-Ugric, Uralic		t^{j} d^{j}			
17	Ukrainian	Slavic, IE		t^{j} d^{j}			
18	Yiddish	Germanic, IE		(t^{j} d^{j})			
19	Czech	Slavic, IE		(c ɟ)			
20	Slovak	Slavic, IE		(c ɟ)			
21	Karaim I	Turkic, Altaic	p^{j} b^{j}	t^{j} d^{j}	k^{j}		
22	Gagauz I	Turkic, Altaic	p^{j} b^{j}	t^{j} d^{j}	k^{j} g^{j}		
23	Karaim II	Turkic, Altaic				y	ø
24	Gagauz II	Turkic, Altaic				y	ø
25	Albanian	Albanian, IE				y	
26	Occitan	Romance, IE				y	
27	Bashkir	Turkic, Altaic				y	ø
28	Danish	Germanic, IE				y	ø hi ø lo
29	Faroese	Germanic, IE				y	ø (œ)
30	Finnish	Finno-Ugric, Uralic				y	ø
31	Frisian	Germanic, IE				y	ø
33	Gorno-Mari	Finno-Ugric, Uralic				y	œ
33	Hungarian	Finno-Ugric, Uralic				y	ø
34	Icelandic	Germanic, IE				ʏ	œ
35	Izhora	Finno-Ugric, Uralic				y	œ
36	Mari	Finno-Ugric, Uralic				y	œ
37	Norwegian	Germanic, IE				y	ø

38	Swedish	Germanic, IE		y_1 (y_2)	ø
39	Tatar	Turkic, Altaic		y	ø
40	Vod’	Finno-Ugric, Uralic		y	œ
41	French	Romance, IE		y	œ ø
42	Breton	Celtic, IE		y	œ ø
43	Dutch	Germanic, IE		y	œ ø
44	German	Germanic, IE		y ʏ	œ ø
45	Karelian	Finno-Ugric, Uralic	$(p^j\ b^j)$ $t^j\ d^j$ $(k^j\ g^j)$	y	œ
46	Veps	Finno-Ugric, Uralic	(p^j) $t^j\ d^j$ $(k^j\ g^j)$	y	œ
47	Estonian	Finno-Ugric, Uralic	$t^j\ d^j$	y	œ
48	Chuvash	Turkic, Altaic	t^j	y	ø

(III) Perception of the item /C^juC^j/, a sample run based on 100 tokens; the numbers indicate Agent B’s “responses” separately for vowels and consonants (C1 = C2); the highest numbers are given in bold.

Table 2. No vowel undershoot; no consonant undershoot

i	y	ɯ	**u**
0	0	24	**76**
C^j	$C^ɥ$	$C^ɣ$	C^w
73	27	0	0

Table 3. Vowel undershoot of 25 %; no consonant undershoot

i	y	ɯ	u
0	0	**87**	13
C^j	$C^ɥ$	$C^ɣ$	C^w
77	29	0	0

Table 4. Vowel undershoot of 50 %; no consonant undershoot

i	y	ɯ	u
0	39	**61**	0
C^j	$C^ɥ$	$C^ɣ$	C^w
81	19	0	0

Table 5. Consonant undershoot of 25 %; no vowel undershoot

i	y	ɯ	**u**
0	0	26	**74**
C^{j}	**$C^{ɥ}$**	$C^{ɣ}$	C^{w}
7	**89**	4	0

Table 6. Consonant undershoot of 50 %; no vowel undershoot

i	y	ɯ	**u**
0	0	29	**71**
C^{j}	$C^{ɥ}$	**$C^{ɣ}$**	C^{w}
7	46	**54**	0

Acquisition

First language (L1) acquisition

The role of contrast in the acquisition of phonetic systems

Daniel J. Weiss and Jessica Maye

1. Introduction

An abundance of research has focused on the process by which the infant's ability to discriminate phonetic contrasts is pruned over the course of the first year, such that only native language contrasts remain discriminable. In contrast, our study focuses on the process by which contrasts that are initially poorly discriminated are facilitated via exposure to a language in which those contrasts are phonemic. Our goal is to determine whether the distribution of sounds in a language can facilitate the discrimination of difficult phonetic contrasts. The experiment presented here represents the start of a larger research project whose goal is to better understand how statistical information in the speech stream guides speech perception within a given phonetic system. In future research we hope to use these studies to address larger issues regarding mechanisms constraining language acquisition in humans.

2. Background and current study

It has been well documented that infants are often better than adults at discriminating phonetic contrasts that are not phonemic in the native language (e. g. Trehub, 1976; Werker et al., 1981). Further, there is an abundance of evidence that as infants develop they begin to discriminate only those phonetic categories that are phonemically contrastive with each other in the native language (e. g. Werker & Tees, 1984; Werker & Lalonde, 1988; Werker & Polka, 1993). This shift from a language-general system of speech perception to a language-specific system occurs relatively early in life. Werker and Tees (1984) found that 6–8 month old infants from English-speaking households could discriminate contrasts found in Hindi and Salish languages that are difficult for adult English speakers to discriminate. At ages 8–10 months, the proportion of infants able to make these discriminations was significantly reduced. By the age of 10–12 months, infants were as poor at making these discriminations as English-speaking adults. These results suggest that although most phonetic contrasts are discriminated by young infants, over the course of development the set of discriminable contrasts is pared down such that it matches the set of contrasts found in the native language.

While there has been a wealth of research supporting the notion that infant speech perception undergoes this paring process, a limited number of studies have indicated that not all phonetic contrasts are well discriminated in early infancy. For these difficult contrasts, exposure to a language in which the contrast is phonemic appears to have a facilitory effect. Aslin and colleagues (1981) found that sensitivity to the contrast between prevoiced and short-lag stop consonants is weak in young infants, and exposure to a language that utilizes such contrasts phonemically is required in order to achieve adult-like competence in discrimination. Polka, Colantonio, and Sundara (2001) investigated the discrimination of /d/-/ð/ in English and French infants and adults. They found that this contrast is discriminated poorly by infants from both language communities, as well as by French-speaking adults; whereas English-speaking adults discriminate it well. This finding demonstrates that exposure to a language in which this contrast in phonemic (i. e. English) facilitates discrimination, while discrimination remains poor if exposed to a language that does not utilize this contrast (i. e. French). Studies such as these indicate that there are some phonetic contrasts that are initially difficult to discriminate; but that exposure to a language that utilizes the contrast phonemically facilitates their discrimination. In addition, some contrasts may be intermediate in their initial difficulty, such that development can take the form of either facilitation or loss, depending on the phonemic status of the contrast in the native language. For example, although both English and Japanese infants discriminate English [r]-[l] at 6 months of age, by 12 months English infants show increased discrimination, while Japanese infants show decreased sensitivity to the contrast.

The goal of the current study is to investigate the mechanism by which such facilitation occurs. Specifically, we hypothesize that the statistical distribution of speech sounds in an infant's input may be a driving factor in the process of facilitation of difficult contrasts. We believe that the shape of the distribution may indicate to the infant which phonetic categories are contrastive in the native language.

3. Phonetic Learning by mode detection

A recent study by Maye, Werker, and Gerken (2002) demonstrated that the discrimination of phonetic contrasts is affected by the distribution of sounds in a speech stream. This study tested the hypothesis that information within the speech stream itself might indicate to infants which sounds are used contrastively in the language. Despite a large degree of phonetic variation, tokens of one speech sound category (e. g. [p^h]) will tend to be more acoustically similar to members of their own category than to members of other, contrasting categories (e. g. [b]). Because of this, the distribution of two categories that are used contrastively within a language should approximate a bimodal distribution of acoustic features; whereas sounds that are not used contrastively should approximate a unimodal distribution.[1]

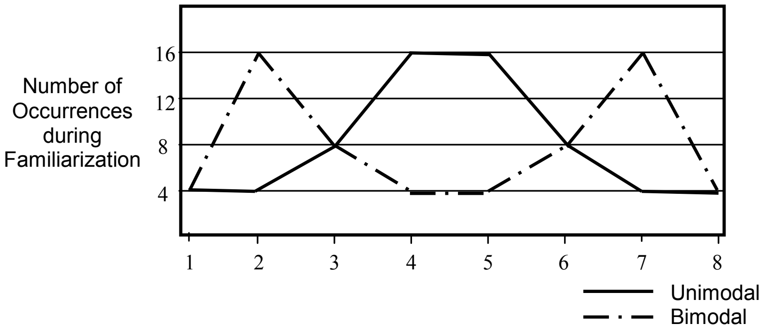


Figure 1. Bimodal vs. unimodal distributions on a phonetic continuum. Maye et al. (2002) utilized both bimodal and unimodal distributions; the current study compares exposure to a bimodal distribution with no prior exposure to phonetic stimuli.

Maye and colleagues tested 6–8 month old infants on their discrimination of voiced [da] vs. voiceless unaspirated [ta].[2] While these sounds are not used contrastively in English, they are discriminable to 6–8 month old infants as

1 Studies demonstrating that speech sound categories are reflected in this sort of distributional evidence include Lisker & Abramson (1964), Magloire & Green (1999), and Sundberg & Lacerda (1999).

2 The unaspirated [ta] stimulus utilized by Maye et al. (2002) was excised from the syllable [sta], and thus the [t] contained coarticulatory effects from the preceding [s]. As a result, the place of articulation for [t] differed slightly from [d]. The continuum between [da] and [ta] was created by altering the formant transitions into the following vowel. In addition, prevoicing was present on the first three tokens of the continuum. Thus, although it included some manipulation of prevoicing, it was not strictly a voicing contrast *per se*. See Pegg & Werker (1997) for discussion of this contrast and Maye et al. (2002) for a complete description of these stimuli. While the contrast used in the present study was similar to that of Maye et al. (2002), we manipulated only one phonetic parameter (i. e., voicing) and tested discrimination of non-endpoint stimuli in order to ensure that discrimination would be difficult.

well as English speaking adults (Pegg & Werker, 1997). Infants were familiarized to either a unimodal or a bimodal distribution of sounds along an 8-point continuum from [da] to [ta] (see Figure 1) for 2.3 minutes, and then tested on their discrimination of the endpoints of the continuum. Infants familiarized to a bimodal distribution of the continuum discriminated the endpoints at test, while infants familiarized to a unimodal distribution did not. The fact that this contrast has been shown to be discriminable to English-learning infants at 6–8 months suggests that familiarization to a unimodal distribution of the sounds suppressed infants' discrimination.

4. Experiment

The results of the Maye et al. (2002) study demonstrate that during the age range when infants are honing in on native language contrasts, they are sensitive to statistical cues in speech that reflect native language categories. Furthermore, the distribution of sounds in the input affects whether or not infants discriminate a contrast. It is therefore likely that this statistical learning mechanism contributes to the development of speech perception. The current study asks whether this statistical learning mechanism can also facilitate the discrimination of difficult phonetic contrasts that an infant might encounter in the native language. In particular, while Maye et al. (2002) showed that a unimodal distribution can suppress discrimination of a previously discriminable contrast, our study asks whether exposure to a bimodal distribution can *facilitate* the discrimination of a previously non-discriminable contrast. In order to test this hypothesis, we tested infants' discrimination of a phonetic contrast that has been reported to be poorly discriminated in early infancy: namely, the contrast between prevoiced and short-lag stops (Aslin et al., 1981). We compared the discrimination of infants familiarized to a bimodal distribution of the sounds with that of infants given no relevant preexposure, with the prediction that infants exposed to a bimodal distribution will demonstrate better discrimination of the contrast.

5. Methods

Subjects. Thirty-two infants from English-speaking homes were included in the study. Subjects ranged in age from 7 months, 12 days to 8 months, 25 days (mean = 8 months, 11 days). An additional 13 subjects were run but excluded from analysis for the following reasons: crying (n = 6), failure to habituate (n = 2), failure to dishabituate to post-test (n = 1), no usable test trials (n = 1), parental interference (n = 1), experimenter error (n = 1), equipment failure (n =

1). Infants were randomly assigned to either the Control or Bimodal condition of the experiment.

Stimuli. We recorded several tokens of the syllables /ga/ and /ka/ as produced by a male speaker of Hindi, a language in which the voiced-voiceless contrast is one of prevoiced vs. short-lag voice onset time (VOT). We created a continuum of prevoiced to short-lag stimuli by synthetically manipulating these naturally produced syllables. Four tokens of [ka], differing slightly in length and intonation contour, were chosen as exemplars, from which four experimental continua were created. The exemplars had relatively long voicing lags, which were then edited by removing portions of voicing lag (using SoundEdit 16 v2.0), to create tokens at four VOT values: 0 ms, 7 ms, 14 ms, and 21 ms voicing lag. To create the prevoiced end of the continuum, prevoicing from naturally produced tokens of [ga] was spliced onto the beginning of the 0-msec lag [ka] tokens, to create tokens at four prevoicing values: 100 ms, 75 ms, 50 ms, and 25 ms voicing lead. Thus, the experimental stimuli consisted of four 8-point [ga]-[ka] continua (based on the four [ka] exemplars with differing intonation), each ranging from –100 ms to 21 ms VOT (see Figure 2).

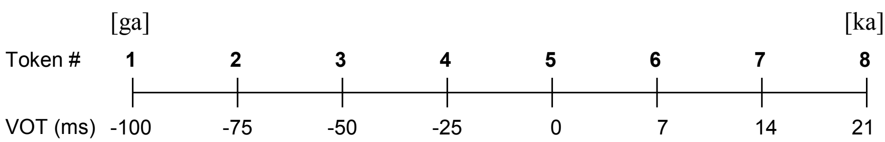


Figure 2. Experimental continuum.

During the familiarization phase of the experiment, infants in the Bimodal condition heard all four of the 8-point [ga]-[ka] continua, presented in random order with an inter-stimulus interval (ISI) of 1 second. The frequency of presentation for these stimuli during familiarization exemplified a bimodal distribution (see Figure 1). That is, tokens near the endpoints of the continua were presented more frequently than tokens from the center. For the Control condition, the familiarization stimuli consisted of a random sequence of tones. Tokens 3 and 6 from the [ga]-[ka] continua were used during the test phase for infants in both conditions (see below).

Procedure. During the experiment, infants were seated on their parent's lap inside a soundproof chamber (Industrial Acoustics, Inc.). Infants faced a television monitor (Hitachi Vm-905AU), which was positioned above a hidden speaker (Boston Acoustics: located behind a curtain), and below a video camera (Sony Hyper HAD: also partially occluded from view). The parent wore sound-canceling headphones (Peltor Workstyle) and listened to music throughout the entire procedure. The experimenter sat outside the chamber and monitored the experiment via closed-circuit television connected to the camera in the chamber.

The experiment began with a 2.5 minute familiarization phase. During familiarization, infants were presented with a silent cartoon clip while hearing either the control (tones) or experimental ([ga]-[ka]) auditory stimuli. Upon completion of the familiarization stimulus presentation, the screen went blank and the test phase began, which was identical for infants in both conditions. Infants' discrimination was tested using a habituation-dishabituation procedure. On each test trial, a multi-colored bullseye appeared on the screen. When the infant looked at the screen (as determined by the experimenter outside the booth), presentation of a sound file was initiated. During habituation, the sound file consisted of the four exemplars of token 6 ([ka], 7 ms VOT) from the experimental continua, presented in random order for a maximum of 60 seconds (ISI = 1 sec). Infant looking times were monitored by the experimenter and recorded using a G4 Macintosh computer. When infants looked away from the target for 2 seconds, the trial would terminate (i.e. the bullseye disappeared from the screen and the auditory stimuli terminated). The next trial began when the infant re-oriented towards the screen.

The threshold for habituation was calculated on the basis of the first 3 trials whose summed looking time was at least 18 seconds total. Habituation was defined as any 3 trials subsequent to these initial 3 whose looking times were half or less than half of the sum of the initial 3 trials. Any trials with looking times less than 2 seconds were excluded. The maximum number of trials to habituation was 20. After habituation occurred (or 20 trials elapsed with no habituation), two change trials were presented. Infants who failed to meet the habituation criterion within 20 trials went on to the change trials but were excluded from analysis (n = 2). During the change trials, the same multi-colored bullseye appeared on the screen, but the sound file consisted of the four exemplars of token 3 ([ga], -50 ms VOT) from the experimental continua, presented in random order for a maximum of 60 seconds (ISI = 1 sec). Following the change trials was a single post-test trial, in which the same visual stimulus was presented along with an acoustically very different sound (the nonce word *bupoki*, produced by a synthetic female voice), repeated for a maximum of 60 seconds (ISI = 1 sec). This post-test trial served to ensure that if infants did transfer habituation to the test stimuli, it was not due to overall habituation to the test apparatus. One infant failed to dishabituate to either the change trials or the post-test trial, and was excluded from analysis.

Results. The looking time data for the last two habituation trials and the two test trials are presented in Figure 3. Due to large individual differences in overall looking times, we used z-scores to normalize the data. A 2x2 mixed-design ANOVA revealed no main effect of Condition (F[1,30] < 1, ns) or Trial Type (habituation vs. change; $F[1,30] < 1$, ns), but a significant interaction effect ($F[1,30] = 5.114$, $p<.05$). In particular, as revealed by planned pairwise comparisons, infants in the Bimodal condition showed a significant increase in looking time on the change trials as compared with the last two habituation trials ($t[15] = 1.886$, $p<.05$); whereas infants in the Control condition showed a non-significant *decrease* in looking time on change trials ($t[15] = 1.250$, $p = .115$).

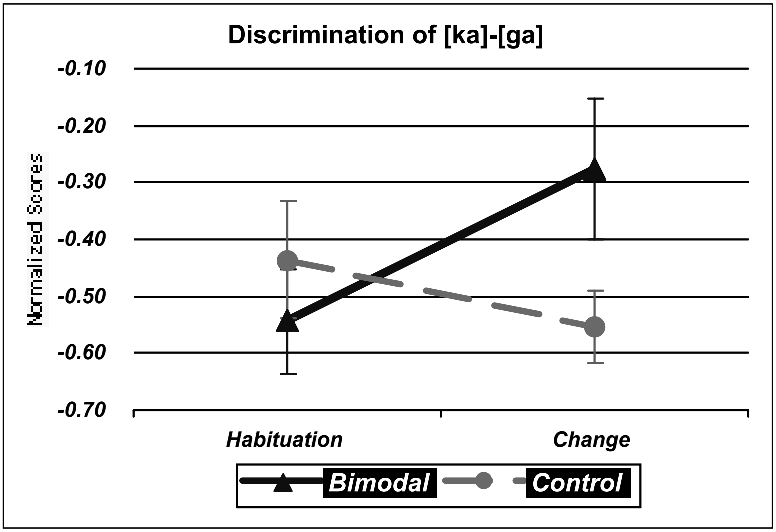


Figure 3. Normalized looking time scores for habituation trials vs. change trials for infants in each condition. Habituation scores represent the average of the last two habituation trials ([ka], VOT 7 ms); change scores represent the average of the two change trials ([ga], VOT –50 ms).

In addition, in the Bimodal condition 13 out of 16 infants showed an increase in looking time for change trials as compared with the last 2 habituation trials. This represents a significant difference in the proportion of respondents between conditions as only 8 out of 16 infants in the Control condition showed increased looking time for change trials (binomial test, test proportion = .50, $p < .022$).

6. General discussion

As predicted, infants exposed to a bimodal distribution during familiarization discriminated the contrast between prevoiced vs. short-lag velar stops, whereas infants without relevant pre-exposure did not discriminate the contrast. These results support the hypothesis that statistical cues regarding the contrastiveness of two sounds in a language can facilitate the discrimination of a difficult phonetic contrast. More generally, our results provide further evidence that during the period of development when infants are honing in on native language contrasts, they are sensitive to distributional cues within the speech stream.

One additional control condition is necessary before we can conclude that it is the bimodal distribution in particular that is responsible for the experimental group's discrimination of the contrast. It could be that infants in the experimental condition are able to discriminate the habituation and change stimuli based solely on the additional exposure to the Hindi sounds (rather than to a specifically bimodal distribution of the sounds). To control for this, we are currently running a condition in which infants are exposed to a unimodal distribution of the same continua (see Figure 1). Based on the findings of the Maye et al. (2002) study, we predict that infants in the unimodal condition will not discriminate between the habituation and change stimuli. In fact, the transfer of habituation may be even more robust for the unimodal group than the control group, since the aforementioned study found that a unimodal distribution suppressed discrimination.

In future research, our goal is to determine how the learning of one contrast in a language affects the acquisition of additional contrasts. When infants learn to discriminate two phonetic categories, it is possible that the learning is restricted to those particular sounds. However, an alternate possibility is that infants' initial acquisition of speech sound categories occurs at a more abstract level, such as the level of the phonetic feature. If this is the case then we might find that exposure to a bimodal distribution of sounds at one place of articulation actually facilitates discrimination of the same featural contrast at an untrained place of articulation. In contrast, if learning is specific to the familiarization stimuli, then familiarization should have no effect on the discrimination of untrained stimuli.

Maye (2000) addressed this question with adult subjects, by presenting native English speakers with voiced vs. voiceless unaspirated stop consonants at either the alveolar or velar place of articulation. Subjects were familiarized with one place of articulation, and then tested on their discrimination of the contrast, first at the trained place of articulation, and subsequently at the untrained place of articulation. Subjects were randomly assigned to one of three familiarization conditions: Bimodal, Unimodal, and No Familiarization. The results showed that while there was a significant effect of familiarization on discrimination of the trained contrast (with the Bimodal group showing greater discrimination than the Unimodal group), there was no generalization to the untrained contrast. In other words, neither the Bimodal or Unimodal group differed from the No Familiarization group on their discrimination of the contrast at the untrained place of articulation.

Although the Maye (2000) study found no generalization to an untrained place of articulation, there are two reasons that infants might perform differently. First, the adult study assessed discrimination using a metalinguistic task. Adult subjects were asked to imagine that the syllables they heard were words in a foreign language. They were then presented with pairs of syllables and asked to indicate whether they thought that the two syllables were the "same word" or "different words" in the language (although there were no meanings associated with the "words"). The fact that subjects were asked to make a metalinguistic interpretation of the stimuli, rather than simply to indicate whether they heard *any difference* between the two syllables, may have interfered with potential generalization. Because infant studies do not introduce metalinguistic factors, there should be no such drawback. In addition, we plan to run a second adult study using methods that provide a more direct measure of discrimination.

A second possibility is that infants may learn phonetic categories in a fundamentally different way than adults do. Infants are in the process of learning a first phonetic system, while adults already have well-established phonetic systems that they must add to or alter in order to incorporate new or different contrasts. Thus, it would not be surprising if the constraints on infant phonetic category learning were different from constraints that operate over adult phonetic retuning.

If we do find that infants learn phonetic contrasts at the level of the feature, it will provide an opportunity to investigate whether phonetic learning is guided by markedness principles. Markedness principles reflect cross-linguistic regularities regarding the relative likelihood of occurrence for different linguistic elements. For example, velar sounds are relatively more rare (more marked) than coronal sounds. In addition, all languages that utilize velar sounds also utilize coronal sounds. The reverse is not true: not all languages with coronals also have velars. It is this sort of statistical regularity in the environment that is likely to be encoded into a learning mechanism adapted by natural selection. Mechanisms that are able to adapt to such environmental regularities should confer an advantage to the user by confining the learning space and increasing the speed of acquisition. Thus, markedness implications are a prime candidate for principles that might be innately encoded in the human mechanism for language acquisition. If this is the case, then we may find asymmetries in the generalization of newly learned phonetic contrasts that reflect markedness implications. Specifically, we would predict generalization from marked to unmarked places of articulation, but not vice versa. That is, an infant trained on a velar contrast should be able to discriminate the same featural contrast at a coronal place of articulation; but infants trained on coronals should not generalize to velars. This prediction is predicated on the fact that it would be beneficial for a language learning mechanism to be aware of the fact that rare phonemic contrasts are predictive of the inclusion of more common phonemic contrasts.

Finally, although in this study we only familiarized and tested infants' discrimination in a single phonological context (namely, word-initial position), it is likely that phonetic learning of this nature is context-specific. That is, two sounds that occur in complementary distribution (two different phonological contexts; e. g., one sound occurs only in word-initial position, the other only in foot-medial position), they do not count towards the same distribution. This makes it possible for a set of phonetic exemplars to form a bimodal distribution in one context (e. g. syllable-initial), and a unimodal distribution in another context (e. g. syllable-final). This phenomenon is known to linguists as "neutralization" of a contrast, and is common cross-linguistically (e. g. German word-final voicing neutralization, English foot-medial flapping of alveolar stops, Korean stop neutralization in coda position).

Evidence for context-specific phonetic learning comes from the fact that English-speaking adults show poor discrimination between English voiced [d] and voiceless unaspirated [t] (the latter occurring only immediately after /s/, while the former never occurs in this position). Pegg and Werker (1997) found that when the initial /s/ was excised from the syllable /sta/, English-speaking adults had trouble differentiating the remaining [ta] from a token of /da/. The coarticulatory influence of the preceding /s/ causes unaspirated [t] in English to differ slightly from [d] in place of articulation, causing consistent differences between these two sounds both in burst properties and formant transitions (Pegg & Werker, 1997). Thus, the two sounds occur frequently in English and have consistently different acoustic properties; yet, English-speaking adults discriminate them poorly. A similar phenomenon is evident in English-speakers' poor discrimination of the [d] and [ð] allophones of /d/ (Boomershine et al., this volume; Maye, 2005).

References

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., and Perey, A. J.
1981 Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. *Child Development, 52*, 1135–1145.

Boomershine, A., Currie Hall, K., Hume, E., and Johnson, K.
this volume The impact of allophony versus contrast on speech perception.

Lisker, L., and Abramson, A. S.
1964 A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*, 384–482.

Magloire, J., and Green, K. P.
1999 A cross-language comparison of speaking rate effects on the production of voice onset time in English and Spanish. *Phonetica, 56*, 158–185.

Maye, J.
2000 Learning speech sound categories on the basis of distributional information. Unpublished doctoral dissertation, University of Arizona.

Maye, J.
2005 Development of phonotactic constraints on phonetic discrimination in infancy. 29th Annual Boston University Conference on Language Development.

Maye, J., Werker, J. F., and Gerken, L. A.
2002 Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82 (3)*, B101-B111.

Pegg, J. E., and Werker, J. F.
1997 Adult and infant perception of two English phones. *Journal of the Acoustical Society of America, 102*, 3742–3753.

Polka, L., Colantonio, C., and Sundara, M.
2001 A cross-language comparison of /d/~/ð/ discrimination: Evidence for a new developmental pattern. *Journal of the Acoustical Society of America*, 109, 2190–2201.

Sundberg, U., and Lacerda, F.
1999 Voice onset time in speech to infants and adults. *Phonetica, 56*, 186–199.

Trehub, S. E. (1976) The discrimination of foreign speech contrasts by infants and adults. *Child Development, 47*, 466–72.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., and Tees, R. C.
1981 Developmental aspects of cross-language speech perception. *Child Development, 52*, 349–355.

Werker, J. F., and Lalonde, C. E.
1988 Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology, 24*, 672–83.

Werker, J. F., and Polka, L.
1993 Developmental changes in speech perception: New challenges and new directions. *Journal of Phonetics, 21*, 83–101.

Werker, J. F., and Tees, R. C.
1984 Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49–63.

How does Place fall into place?

The lexicon and emergent constraints in children's developing phonological grammar[1]

Paula Fikkert and Clara Levelt

In this paper we address the acquisition of place of articulation (PoA) features in words by Dutch children. We show that there is a particular developmental pattern, repeated across children. This pattern can be accounted for by (a) assuming that the child's underlying phonological representation in the lexicon becomes gradually more specified, (b) the emergence of segmental markedness constraints, and (c) referring to the distribution of PoA patterns in the target language. Consonant harmony is an epiphenomenon of this general developmental pattern of PoA organization in words. Generalizations that the child makes over his or her own productive lexicon are grammaticalized as high-ranking markedness constraints, which force PoA features to be linked to certain positions in the word.

1. Introduction

There are two salient aspects to Consonant Harmony (CH) that previous analyses have not accounted for in a satisfactory way. First, harmony between non-adjacent consonants remains a rather peculiar phenomenon, which is specific to child phonologies. In accounts of CH within the framework of Optimality Theory (Prince and Smolensky [1993] 2004), CH forms are treated as unmarked forms. These forms are triggered by some high-ranked markedness constraint, which at some point is either demoted to regions where its presence can no longer be felt (Levelt 1994, 1995; Goad 1998, 2001, 2003), or the constraint undergoes a change in the domain of application (Bernhardt and Stemberger 1998; Pater and Werle 2001, 2003; Pater 2002). These measures

1 We would like to thank the audiences of the Second International Conference on Contrast in Phonology, Toronto 2002, GLOW, Utrecht, 2002, and the Child Phonology Conference in Vancouver 2003 for valuable comments on our presentations, which have found their way in the present paper, and the editors of this volume and an anonymous reviewer for their detailed and helpful comments. Paula Fikkert was supported for this research by her NWO grant "Changing Lexical Representations in the Mental Lexicon".

have to be taken since CH of primary PoA features does not appear in adult language at all. In contrast, other unmarked aspects of children's initial productions, like a CV syllable structure or a minimal Prosodic Word shape, never disappear from the language, and can also emerge under certain circumstances as the optimal output from some more marked input (*The Emergence of The Unmarked*, McCarthy and Prince 1994). The child-language-specificity, either of the constraint or of the domain of application of the constraint, is a problem if we want the substance of grammars, including child grammars, to be stable, and if we want child grammars to mirror cross-linguistic adult grammars (see Pater (2002) for similar reflections).

The second salient fact is that CH is an emerging phenomenon in children's productions. In initial vocabularies there are no CH forms, and children are in fact surprisingly faithful to the PoA structure of the adult target words they are attempting. In the case of CH, target forms that at later stages lead to CH productions are simply not attempted in the early stages. This initial selection of target words that can be produced faithfully cannot be easily accounted for by any grammar, but it is certainly not expected in a grammar where markedness constraints initially outrank faithfulness constraints – the accepted view of an initial developmental grammar in Optimality Theory today (see Boersma and Levelt 2003 and references therein; Gnanadesikan [1995] 2004). Furthermore, it would be expected that subsequent demotion of Markedness constraints in the grammar would give rise to more faithful productions, rather than less faithful ones. What we find is that children initially aim for productions that are both faithful and unmarked, and later drop the concern for faithfulness.[2]

A neglected issue concerning CH data is how CH forms relate to other forms in the vocabulary. CH forms have been treated as an isolated set of data in most accounts. However, here we will show that they are an epiphenomenon of the way children handle Place of Articulation in their vocabulary as a whole.

In the remainder of this paper we will elaborate on the above facts and issues, and show how they can be dealt with. We propose, specifically, that the nature of the initial phonological system is different from the system in more advanced stages of development because it is closely tied to the developing lexicon. In the initial stages, the development of lexical representations and the acquisition of a phonological system go hand in hand. We argue that constraints can emerge in the grammar, as grammaticalized generalizations over the child's early pro-

2 The underlying reason for this development could be that a vocabulary of just unmarked and faithful words becomes too limited to express the things the child wants to express.

ductive lexicon. It remains to be seen whether these constraints are transient, or form a more permanent part of the grammar. At least, traces of the effect of these constraints can be found in the adult target language (Fikkert et al. 2004). Furthermore, we present evidence from production for initial "holistic" and un(der)specified phonological representations. In the course of development these representations become segmentalized and more specified.

A constructionist or emergentist view of the child's grammar and of the child's lexical forms is of course not new (e. g. Ferguson and Farwell 1975; Macken 1978; Menn 1983; Moskowitz 1973; Vihman 1996, Vihman and Velleman 2000; Waterson 1971). Our aim here is to reconcile this view with a generative approach (in casu OT-based research) on acquisition (see also Pater 2002) by pointing out where, when and how a developing grammar is supplied with "constructionist" elements. However, our primary goal here is to determine the exact nature of the data.

2. Materials and Methods

Since our claim is that CH is a consequence of emerging constraints, which are built on the structure of the initial lexicon, our interest lies in the development of the distribution of PoA features over words. With that objective we studied the PoA structure of every word in the corpora of five children acquiring Dutch as their first language. In addition, we studied the PoA structure of words in both the language intake of the children, and in child directed language input (the Van de Weijer (1998) corpus). By *intake* we mean the adult target words that children attempt to produce, which is a selection from the adult input.

First we studied longitudinal, developmental data from 5 children acquiring Dutch as their first language:[3] Tom (1;0–2;22), Jarmo (1;4.18–2;4.1), Robin (1;4.14–2;4.28), Eva (1;4.12–1;11.8) and Noortje (1;7.14–2;11). We recorded data of these children every other week for a period of about one year. These children were selected out of the original group of 12 children from the CLPF (*C*lara *L*evelt and *P*aula *F*ikkert) database (Fikkert 1994, Levelt 1994) because they were recorded from the earliest stages of meaningful speech production. A total of 8407 spontaneous utterances were analyzed (onomatopoeic forms and immediate repetitions were excluded from the analysis). All the words in these utterances were coded for their PoA structure in the following way: labial consonants were represented by P, coronal consonants by T and dorsal

3 These data can be found in CHILDES (MacWhinney 2000).

consonants by K. Round (labial and dorsal) vowels were represented by O, coronal (front) vowels by I and low vowels by A (see Pater and Werle [2001] for a similar method). In addition, front rounded vowels were coded as IO. However, as these vowels occurred infrequently and only at more advanced stages of development, they did not influence the main pattern. In words of more than one syllable only the stressed syllable was coded. Thus, a CVCV form with stress on the initial syllable (where V stands for either a long or a short vowel) was coded CVC-. A word like *baby*, for example, was coded as PIP-. As there was no difference in the developmental patterns of PIP versus PIP- words[4] we collapsed both types in our further analyses. In the case of consonant clusters, the PoA feature of the least sonorant consonant in obstruent-sonorant clusters was taken as the basis for coding, as in most instances this is the consonant that survives in children's cluster reduction patterns (Fikkert 1994, Barlow 1997, Jongstra 2003). For similar reasons, in the case of /sC/-clusters the PoA feature of the /C/ was coded. /h/ was coded as placeless H. In (1) we provide some examples of our coding of children's utterances:

(1) Child Utterance Coding

Target	Child Production	Coding	Result
brood *bread*	[bop]	b= P	POP
/brot/		o= O	
		p = P	
snoep *candy*	[fup]	f = P	POP
/snup/		u = O	
		p = P	
paard *horse*	[pat]	p= P	PAT
/part/		a = A	
		t = T	
trein *train*	[tɛin]	t = T	TIT
/trɛin/		ɛi = I	
		n = T	
lachen *laugh*	[lɑχə]	l =T	TAK-
/lɑχə /		ɑ = A	
		χ = K	

We coded the adult target words in a similar way, as exemplified in (2):

4 This is itself an interesting finding, as it strengthens the claim that there is a word pattern, rather than a syllable-based pattern. Codas and onsets of second unstressed syllables behave similarly with respect to PoA.

(2) Adult Target Coding

Target	Coding	Result
brood /brot/	br = P	POT
	o = O	
	t = T	
snoep /snup/	sn = T	TOP
	u = O	
	p = P	
paard /part/	p = P	PAT
	a = A	
	rt = T	
trein /trɛin/	tr = T	TIT
	ɛi = I	
	n = T	

To return to data sources, in addition to the adult targets we also coded 914 words from a list of words that 6-year olds are supposed to know and use. This list (Schaerlaekens, Kohnstamm and Lejaegere 1999; Zink 2001) is comparable to the MacArthur-Bates Communicative Development Inventories, parent report forms for assessing language skills in young children (www.sci.sdsu.edu/cdi/). We call these words the expected words. Finally, the utterances in the child directed speech database (containing 173,752 words) of Joost van de Weijer (1998) were coded.[5] These were the sources we used to gain information about the PoA structure of the input and intake of language learners.

While we coded all words in both the children's utterances and their corresponding targets, here we limit the discussion to those words that have at least two consonants, i. e. words with a CVC(-) coding, since we are particularly interested in how PoA in words with two consonants develops in children's outputs. We will only illustrate the very first stage with the PoA structure of CV and VC(-) words.

In order to see whether a developmental pattern could be found for the distribution of PoA features over words, the PoA patterns of the child utterances and those of the adult targets were aligned on separate Guttman scales, as will be shown below in section 3. Guttman scaling is a procedure for obtaining an order in data, and for checking to what extent an order is followed (Torgerson 1963). We assumed a pattern to have been acquired if it occurred at least three times during one session, even if the child produced three similar PoA forms of

5 We want to express our gratitude to Joost van de Weijer for generously sharing his data with us.

one target word. As the data could be aligned quite nicely on the scale, we concluded that the PoA structures were acquired in a particular order over time.

For every child the data were aligned on three scales, one showing the order of appearance of the different PoA patterns in the child's production data, one showing the order of appearance of attempted target PoA structures, and one showing the order of faithful productions of attempted adult targets. Finally, we calculated the distribution of the different PoA patterns in the list of words children are supposed to know and use, as well as in the set of attempted adult targets and in the child directed input database. We did this in order to check whether frequency in the input, or intake, influences the order of development in production.

3. Results: PoA patterns in children's utterances, intake and input

In the following sections, we first discuss the PoA patterns found in the production data of the children, and those found in the attempted targets (the intake). Subsequently, we investigate how faithful productions of adult targets develop, and when unfaithful productions appear. Finally we present the distribution of PoA patterns in the language intake and input.

Table I. PoA patterns in production (Jarmo)

Stage	**Produced forms by Jarmo**								
	PVP	TVT	KVK	PVT	PVK	TVK	KVT	TVP	KVP
I 1;4.18–1;5.28	PA PO	TA TI	KA KO						
II 1;5.28–1;6.14	PI	TO	KI						
II/III 1;6.14–1;10.23	PAP POP PIP	TIT TAT TOT		POT PAT					
III/IV 1;10.23– 1;11.21			KIK KOK	PIT	POK	TIK			
IV/V 1;11.21–2;2.6			KAK		PIK	TOK TAK	KIT KAT KOT	TAP TIP TOP	
V 2;2.6									KOP KIP KAP

3.1. Development of PoA patterns in children's production data

In this section we provide the developmental patterns of PoA structures from two children, Jarmo and Robin. The patterns of the three other children are remarkably similar. In table (I) and (II), we have summarized the results from the Guttman scaling procedure for Jarmo and Robin (see Appendix A for data from the other three children). Since the Guttman scale is thought to reflect an order, and since here we are talking about a developmental order, every step in the scale is taken to reflect a developmental stage.

Table II. PoA patterns in production (Robin)

Stage	**Produced forms by Robin**										
			PVP	TVT	PVT	TVK	PVK	KVK	TVP	KVT	KVP
I 1;5.11– 1;6.22	PA PO	TA TI	PAP	TAT TIT							
II/III 1;6.22– 1;8.10	PI		POP		POT			KAK*			
II/III 1;8.10– 1;9.22		TO		TOT	PAT PIT						
IV 1;9.22– 2;2.27			PIP			TIK TAK TOK	PIK PAK POK	KIK KOK			
V 2;2.27– 2;3.17									TIP TAP TOP	KIT KAT KOT	KOP KAP KIP

* The only early forms of the shape KAK are onomatopoeic forms like *kwak* 'quack'.

From these data a general, five-step developmental pattern arises. For every stage we highlight the most salient development:

Stage I

At the first stage, both consonants (C_1 and C_2) in the words are labial (P), Coronal (T) or Dorsal (K). In other words, C_1 equals C_2 with respect to PoA features. In addition, the vowel either carries the same PoA feature as the consonants

(V = C, that is POP, TIT and KOK) or it is A, a low vowel.[6] Not all children have dorsal initial words, though. For instance, one child, Tom, only has POP, TIT, PAP and TAT words in the earliest stages of production (1;2.14–1;3.24). It seems that all words are harmonic; i. e. words are completely coronal (TIT), labial (POP) or dorsal (KOK). Low vowels in Dutch seem to be neither front nor back, nor round. In short, they seem to lack primary PoA features, and are solely distinguished by features under the Tongue Height Node (Lahiri and Evers 1991, Levelt 1994; see also section 4.2.1). Therefore they do not interfere with the PoA structure of the word.

Stage II

At the second stage, both consonants still share their place of articulation. However, the vowel can now be different from the consonant(s): we find PI(P), TO(T) and KI(K) patterns in addition to the forms discussed above. Between 1;5.27 and 1;6.13 Jarmo, for instance, starts producing PI, TO and KI word forms. From the next stage on we represent vowels as "v", as their nature is no longer restricted and can freely combine with all consonant patterns that are allowed in the child's system.

Stage III

At stage III, C_1 and C_2 can carry different PoA features for the first time, but in a very restricted way. At first, the only pattern with two consonants that differ in PoA is PvT: C_1 is labial, C_2 is coronal.

Stage IV

Here PvK and TvK appear in the children's data. In other words, C_2 can be realized as dorsal.

Stage V

Finally, at stage V, we find P-final and K-initial combinations: TvP, KvT and KvP.

In reality, the developments do not always strictly succeed each other. Developments can overlap in time, as can be seen above in the data of Jarmo and

6 Often, words in the initial stage are mostly of the structure CV. In that case, we also find that V = C (PO, TI or KO) or V= A (PA, TA, KA).

Robin. However, taking the patterns of all children together, the five proposed stages stand out. The data of each individual child may not necessarily show evidence for all five stages, but, importantly, they are never in conflict with the proposed stages either.

3.2. Development of PoA patterns in children's intake

The PoA patterns in the intake, i. e. the targets that the child aims to produce, are quite similar to the PoA patterns in production, as can be seen in Table III for Robin (see Appendix B for data from the other four children):

Table III. Attempted targets by Robin

Period	**ATTEMPTED targets by Robin**								
1;5.13–1;5.27	POP PAP PIP	TIT	TIK* TAK*						
1;5.27–1;6.22				PIT					
1;6.22–1;7.29		TAT TOT		PAT POT	KOT KAT				
1;7.29–1;8.10							TAP	KIK	
1;8.10–1;9.22						POK PIK			KOP
1;9.22–1;11.9						PAK	TOP TIP		KIP
From 1;11.9			TOK		KIT				KAP

* TIK and TAK only occur in the onomatopoeic expression *tik tak* 'tick tock'.

The general pattern is that:

I There is an initial preference for $C_1 = C_2 = V$ (or V = A) structures also found in the language intake.

II The first combination of different consonants in the intake is PvT, like in production.

III P-final adult targets are attempted relatively late.

The development of target selection thus resembles the development of produced forms. However, there is more variation between children. Naturally, the set of attempted target forms should be larger than the set of produced forms to guarantee a learning effect.

Table IV. Faithful vs. unfaithful use of PoA pattern

PoA pattern in production	First Faithful use	First Unfaithful use for Targets	
TIT	1;5.13	1;5.27	PIT
		1;7.15	TOT
		1;8.26	KAT
		2;0.20	KIT
PAP	1;5.13	1;6.22	PAT
		1;8.26	TAP
		1;11.9	KAP
TI	1;5.13	1;7.15	TIK
		1;7.15	PIK
		1;7.29	KIK
TA	1;5.13	1;5.13	TAK
POP	1;5.27	1;7.15	TOT
		1;7.15	KOT
		1;8.10	KOP
		1;9.22	TOP
POT	1;7.15	1;8.10	TOT
		1;9.8	KOT
TAT	1;7.15	1;7.15	KAT
PAT	1;7.29	1;7.29	TAP
TOT	1;8.26	1;9.8	KOT
TIK	1;9.22	1;10.9	KIK
PIP	1;10.9	1;10.9	TIP
		1;10.9	KIP
TOK	1;11.9*	1.10.23	KOK

3.3. Development of faithfully produced adult targets

An interesting result comes from the development of faithful productions. By a faithful production we mean a word-production that has the same PoA structure as the target adult word. What we find is that at the early stages, four of the five children produced all words faithfully with unfaithful productions appearing only later.[7] This is best illustrated with the Guttman scale for the child

7 From the one exception, Eva, we do not have recordings from her earliest attempts to speak. We might thus have simply missed the fully faithful stage in production.

	1;5.13	1;5.25	1;6.10	1;6.22	1;7.15	1;7.27	1;8.10	1;8.24	1;9.8	1;9.22	1;10.5	1;10.21	1;11.7	1;11.21	2;0.7	2;0.21	2;1.6	2;1.25	2;2.27	2;3.22
TIT	TIT-	TIT	TIT	TIT	TIT	TIT	TIT TAT	TIT	TIT	TIT	TIT	TIT	TIT	TIT	TIT	TIT	TIT	TIT	TIT	TIT
PAP	PAP	PAP	PAP	PAP	PAP	PAP	PAP	PAP	PAP	PAP	PAP	PAP	PAP	PAP	PAP	PAP	PAP	PAP	PAP	PAP
POP	*PO*	POP	POP	POP	POP	POP	POP	POP	POP	POP	POP	POP	POP	POP	POP	POP	POP	POP	POP	POP
PIP	PIP-		PIP	PIP	PIP	PIP	PIP	PIP	PIP	PIP	PIP	PIP	PIP	PIT-	PIP	PIP PIT	PIP	PIP	PIP	PIP
TAT					TAT	TIT	TAT	TAT	TAT	TAT TIT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT TAK	TAT
KAK					KAK	KAK	KAK		KAK					TAK PAK						
POT					POT	PAT TIT	PAT PIT	POT	POT	POT	POT POP	POT	POT	POT POK	POT	POT POP	POT	POT	POT	POT
PIT		*TIT*				PIT	PIT	PIT	PIT	PIT	PIT	PIT	PIT	PIT	PIT	PIT	PIT	PIT	PIT	PIT
PAT				PAP	PAP	PAT	PAT PAP					PAT PAP	PAT PAP		PAT PAP	PAT	PAT	PAT	PAT	PAT
TOT					POP TIT	POT TIT	POT TIT	TOT	TOT	TOT	TOT	TOT	TOT	TOT	TOT	TOT	TOT	TOT	TOT	TOT
PIK					TI			TI		PIK	PIK	PIK TI	PIK TI	PAK		PIK PAK	PIK	PIK	PIK PAK	PIK PAK
TIK	TI TA	TI	TI	TI	TI	TI	TI	TI	TI	TIK TAK	TIK	TIK	TIK	TIK	TIK	TIK	TIK	TIK	TIK	TIK
TAK	TA	TA	TA	TA	TA	TA	TA	TA	TA	TA	TAK	TAK	TAK	TAK	TAK	TAK	TAK	TAK	TAK	TAK
POK							POK		PO		POK POP	---	POK	POK	POK	POK	POK	POK	POK	POK
PAK										PAK		PAK	PAK	PAK	PAK	PAK	PAK	PAK	PAK	PAK
TOK													TOK	TOK	TOK	TOK	TOK	TOK	TOK	TOK
KIK						*TI* KIK				KIK	TIK	TIK	TIK	TIK KIK	TIK KIK	*TI* TIK	KIK	KIK *TIK*	KIK	KIK
KOK							TOT					TOK	TOK	TOK		TOK	TOK	KOK	KOK	KOK
KIT																TIT	TI		KIT	KIT
KOT					POP	OT	POT		TOT		TOT	TOT	TOT	TOT	TOT	TOT	TOT POT	TOT	KOT	KOT
KAT					TAT			TIT	TIT	TIT	TIT TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT TAK	KAT	KAT
TAP						PAT	PAT	PAP	PAP	PAP	PAP	PAP	PAP	PAP	PAP	PAP	*PAP* TAP	PAP	TAP	TAP
TIP											PIP	PIP	TIP	TIP PIP	PIP	PIP	PIP TIP	PIP TIP	TIP	TIP
KAP													PAP	PAP		PAP	PAP	PAP	KAP	KAP
KOP							POP		POP	POP	POP	POP	POP	POP KOP	POP	POP	POP		KOP	KOP
TOP										POP	POP	POP	POP	POP	POP TOP	POP	POP	POP		TOP
KIP											PIP	PIP	PI				PIP	PIP		KIP

Figure 1. Guttman scale of PoA patterns produced by Robin

Robin, in Figure 1. From left to right are the ages at which the patterns are produced. The first column shows the PoA pattern of the attempted adult target words. Faithful productions are shaded. In the first set of recordings there are no productions outside of the shaded area, i. e., all the productions are faithful.[8] The forms without shading are the unfaithful productions.

An examination of the unfaithful productions reveals that every pattern used as an unfaithful substitute has previously (or simultaneously) been used faithfully.[9] This is shown in Table IV (see Appendix C for data from the other 4 children). For every PoA pattern that is used as a substitute at some point, the date of its first faithful use is in the "First Faithful use" column, and the date of its first unfaithful use is in the next column. It is also indicated for which target pattern the production pattern is used as a substitute. This same pattern is found for the other children in our study: faithful productions of a specific PoA pattern appear before this pattern is used unfaithfully – or in some cases they appear simultaneously – in the recorded data.

3.4. Distribution in intake and input

Our final results come from the distribution of the different PoA patterns in the intake and input. The distribution was calculated in the three sets of data discussed earlier.

(3) Distribution of the different PoA patterns in intake

a. "Expected" vocabulary		b. Attempts at adult targets		c. Child Directed Speech	
KK	2.74 %	KP	3.13 %	PP	1.1 %
PP	5.14 %	KK	3.95 %	KK	2.5 %
KP	5.14 %	PP	5.69 %	TP	2.6 %
TP	10.50 %	TP	6.00 %	PK	4.1 %
KT	10.83 %	PK	8.52 %	KP	6.2 %
PK	11.27 %	KT	9.45 %	TK	10.7 %
TK	13.24 %	TK	9.78 %	KT	12.5 %
TT	15.65 %	TT	25.72 %	PT	27.2 %
PT	25.49 %	PT	27.76 %	TT	33.1 %

8 Except for the target word *tiktak* 'tick-tock', which is always produced [tita], leading to the TI and TA entries in Figure 1.

9 The only exception is TOK (marked with *), which appears very late.

The low-to-high order of frequencies is quite similar in the three lists: KK, PP, TP and KP have the lowest frequencies, KT, PK and TK are in the middle-range, and TT and PT occur in the data most frequently.

3.5. Summary of results

From the longitudinal data of language learners of Dutch, a clear developmental pattern emerges in the PoA structure of their productions. Generalizing over the entire set of data, we find the following stages:

(4) Stages in the development of PoA structures in production

Stage	*Development*	*Production patterns (cumulative)*
I	$C_1 = C_2 = V$ (or V=A)	POP, PAP, TIT, TAT, KOK, KAK
II	$C_1 = C_2$	PIP, TOT, KIK
III	$C_1 = P$, $C_2 = T$	PVT
IV	$C_2 = K$	PVK, TVK
V	$C_2 = P$, $C_1 = K$	TVP, KVT, KVP

The development of selected adult targets for production shows a similar pattern. These developmental patterns are related because of the salient finding that a PoA pattern is produced faithfully before it is used as a substitute.

4. Discussion

4.1. Generalizations over developmental patterns

For a satisfactory and comprehensive explanation of the results presented in section 3 we need to take into account at least the following factors: (a) the developing representation of phonological units, (b) the developing lexicon, and (c) the specific input. Below we formulate five generalizations over the developmental patterns that relate to these factors:

1. *Whole Word Stage:* PoA contrast is initially defined on the entire word, and in this sense the word is not analyzed into separately targetable segments yet (Waterson 1987, Menn 1983, Levelt 1994).
2. *Staged Segmentalization:* After the whole-word stage, words become segmentalized. First, consonants become separate from vowels. Subsequent-

ly, PoA contrast becomes defined over different consonant positions: C_1 is the dedicated position for labial, C_2 is the dedicated position for dorsal (Moskowitz 1973; Vihman et al. 1993).
3. *Emerging Constraints:* Language learners are constrained by their own lexicon: lexical patterns are overgeneralized, i. e. the structure of the lexicon builds constraints into the grammar. This accounts for the pattern of initial faithfulness and emerging unfaithfulness in the data (Ferguson and Farwell 1975, Menn 1983).
4. *Unspecified Coronals.* Compared to Labial and Dorsal, the position of Coronal segments is not restricted to a specific position. We hypothesize that this is because Coronal is unspecified in the lexical representation (Paradis and Prunet 1991).
5. *Input Frequency Effect:* There is a correlation between input-frequency and order of development as soon as segments in words are separately specifiable. Inter- or intra-language input-specific distribution of PoA features can thus lead to different orders of development (Moskowitz 1973).

These points will feature in the remainder of the discussion, where we will elaborate on the different developmental stages.

4.2. Stage I

4.2.1. One word, one feature

As a generalized initial stage it was found that in production words have very restricted PoA patterns, namely TIT, TAT, POP, PAP and for some children also KOK, KAK.[10] Translating these patterns back to features, the patterns TIT, POP and KOK represent structures that can be captured by referring to a single PoA feature, Coronal (i. e., unspecified), Labial, or Dorsal, respectively. The A stands for a low vowel /a/ or /ɑ/, and we assume that a low vowel has no PoA specification, only the tongue height specification Low (Lahiri and Evers 1991). This is why the low vowels can appear together with either Coronal, Labial or Dorsal consonants in the TAT, PAP and KAK patterns.

In the initial stage, then, every produced word contains a single PoA characterization. It thus appears that with respect to PoA, the entire unsegmentalized word forms the representational unit of specification (Waterson 1987, de Boysson-Bardies and Vihman 1991). If the specification is Labial, and the

10 For children with only CV-syllables the patterns are TI, TA, PO, PA and KO, KA.

vowel is non-low, the result is POP; if the vowel is low the result is PAP. No PoA specification (i. e., Coronal) leads to TIT, or TAT in case the vowel is low, and a Dorsal specification leads to KOK, or KAK in case the vowel is low.

We will illustrate this stage with the initial recorded vocabularies of two children, Robin and Eva. These children differ in one respect: Robin is almost entirely faithful to the PoA structure of the adult target, except for some syllable-structure induced dissimilarities (5e, g, i), while Eva's productions can have a PoA structure that is fairly unfaithful to the adult target structure.

(5) Initial vocabulary of Robin (1;5.11)

	adult target	*gloss*	*child's production*	*target structure*	*production structure*
a.	*die*	that one	ti	TI	TI
b.	*huis*	house	hœys	HIT	HIT
c.	*thuis*	home	tœs	TIT	TIT
d.	*zes*	six	sɛs	TIT	TIT
e.	*tik tak*	tick-tock	tita	TIK TAK	TIT
f.	*aan*	on	an	AT	AT
g.	*daar*	there	da	TA	TA
h.	*niet*	not	nt	TIT	TT
i.	*pop*	doll	pɔ	POP	PO
j.	*mamma*	mommy	mɑma	PAP	PAP
k.	*aap*	monkey	ap	AP	AP

(6) Initial vocabulary of Eva (1;4.12)

	adult target	*gloss*	*child's production*	*target structure*	*production structure*
a.	*dicht*	closed	dɪə	TIT	TI
b.	*eend*	duck	ein	IT	IT
c.	*eten*	eat	eitɪ	IT	IT
d.	*trein*	train	tæin	TIT	TIT
e.	*neus*	nose	nɛs	TIT	TIT
f.	*konijn*	rabbit	tɛin	TIT	TIT
g.	*teen*	toe	ten	TIT	TIT
h.	*patat*	french fries	tɑt	TAT	TAT
i.	*staart*	tail	tat	TAT	TAT

	adult target	*gloss*	*child's production*	*target structure*	*production structure*
j.	*daar*	there	da	TA	TA
k.	*bed*	bed	dɛt	**PIT**	**TIT**
l.	*prik*	injection	tɪt	**PIK**	**TIT**
m.	*kijk*	look	tɛit	**KIK**	**TIT**
n.	*beer*	bear	dɛ	**PI**	**TI**
o.	*oma*	granny	oma	OP-	OP-
p.	*op*	on	ɔp	OP	OP
q.	*open*	open	opə	OP	OP
r.	*aap*	monkey	ap	AP	AP
s.	*buik*	tummy	bœyp	**PO/IK**	**POP**
t.	*brood*	bread	mop	**POT**	**POP**
u.	*sloffen*	slippers	pɔfə	**TOP**	**POP**
v.	*poes*	cat	puf	**POT**	*POP*
w.	*schoenen*	shoes	umə	**KOT**	**OP**

As can be judged from the types of words that are attempted, the initial *recorded* vocabulary of Robin reflects the actual initial set of words in his active vocabulary, while Eva's initial recorded vocabulary reflects a more advanced stage of lexical development: she is clearly past the fully faithful stage of production. However, her unfaithful productions still all fit the initial stage of "one word, one PoA feature".

The data in (6k, l) and in (6s, t, u, v) could easily be mistaken for cases of CH in the classic sense of one consonant assimilating in PoA with another non-adjacent consonant. However, two aspects suggest that this is not the appropriate analysis.[11] First, there would be both labial harmony and coronal harmony but no dorsal harmony. This is somewhat unexpected, especially given the underspecification of coronal. Second, both types of harmony would apply to the same sequence of consonants, namely labial-coronal (6k, t, v). In (6l), an apparent case of coronal harmony – given markedness, coronal harmony is curious in itself – there is actually no coronal consonant in the target adult word that could trigger harmony. However, there is a coronal (front) vowel. The data in (6n) and (6w) confirm that it is the vowel that determines the PoA structure for the entire word: in *beer* (6n) the vowel is coronal, there are no coronal con-

11 In § 4.3.1 it becomes clear that the harmony analysis is in fact never the appropriate analysis.

sonants in the target, and labial /b/ is substituted with coronal /d/. In *schoenen* (6w) we find the opposite: the vowel is labial, there are no labial consonants present in the target, and coronal /n/ is substituted with labial /m/. Since the vowel is a salient segment in perception, it is not surprising that the PoA value of this segment should attract the highest amount of attention and feature as the PoA specification for the entire word in production.

Faithfulness to the underlying PoA specification of a vowel outranks faithfulness to the underlying PoA specification of a consonant, and apparently only one specification (or no specification at all) is possible. The surface form therefore carries only the PoA feature of the target adult vowel.

4.2.2. Origin of one word, one feature stage

What is the origin of this initial PoA pattern? It is unlikely that the pattern results directly from a high-ranking markedness constraint in the grammar. Harmonic forms, i. e. forms with consonant harmony, are usually dispreferred in the languages of the world (Frisch et al. 2004). An account in terms of an innate and universal markedness constraint requiring such harmonic forms is therefore not the most obvious solution.

MacNeilage and Davis (2000) give a biomechanical explanation for a similar pattern in babbling and early words. They found the following fixed patterns of CV productions: Coronal C + front V (i. e., TI), Labial C +central V (i. e. PA) and Dorsal C + back V (i. e. KO). According to MacNeilage and Davis these patterns result from mandibular oscillation – an opening-closing movement of the jaw which forms the CV *frame* – in combination with a tongue that remains fixed in either front, central or back position during that oscillation, the *content*.

Waterson (1971) states that the child initially has difficulty planning and producing rapid articulatory movements. Limiting the number of PoA features to one per word leads to a reduction of the processing and production load.

These explanations could form the phonetic, or psycholinguistic, grounding for a grammatical constraint such as "one word, one PoA feature" that is active in the grammar at this particular developmental stage. However, since both the biomechanical restrictions and the planning and production difficulties of the early stages will disappear over time with experience and maturation, it is very unlikely that this particular constraint is a universal constraint of the grammar. This could thus very well be a transient, maturational aspect of the grammar. Since the biomechanical or processing difficulties are highly unlikely to reappear later in life, no adult language will have this constraint actively participating in the grammar. We return to the issue of transient constraints in 4.3.

4.2.3. Perception

An alternative to a "one word, one PoA feature" constraint in the grammar arises from considering the role of perception at this particular stage of development. From numerous studies it has become firmly established that infants are able to discriminate speech sounds at high levels of accuracy (for an overview see Jusczyk 1997). The fact that children specifically select words for production that conform to a certain pattern illustrates this ability. In contrast, it turns out that as soon as children start to learn word meanings, they are no longer such accurate perceivers (Stager and Werker 1997; Werker et al. 2002; Pater, Stager, and Werker 2004; Fikkert, Levelt, and Zamuner 2005). Sound sequences like /bɪ/ and /dɪ/, which young infants can discriminate, cannot be discriminated by older infants – 14 months old – when word meanings are involved. Stager and Werker (1997) found that it is not until the age of 17 months that infants can discriminate minimal pairs like /bɪ/ and /dɪ/ that have semantic referents.

In the initial stage in our study, the children are between 14 and 17 months old, i. e., precisely the period during which children cannot discriminate /bɪ/ and /dɪ/ if word meaning is involved, and where they have just set out to build a lexicon. Non-accurate perception, or rather, an incomplete storage in the lexicon of what is perceived, can thus be expected, leading to incompletely specified lexical representations. We expect vowels to be perceived quite accurately as they are the perceptually salient segments (see Kuhl 2000 for an overview). The perceived PoA characteristic of the vowel is thus mapped successfully onto the lexical representation. The consonants, however, are less accurately identified, and leave gaps in their lexical representation. The word *prik*, for example, could be lexically represented as in (7). As in Stager and Werker's /bɪ/ versus /dɪ/ case, the child is not sure about the PoA feature of the consonants, and their PoA is therefore left unspecified.

(7) Incomplete lexical representation
prik (injection)

Adult output:	[pr ɪ k]
Child's Lexical representation:	C ɪ C
	Cor[12]

12 We will assume that coronal is underspecified and therefore not present in the underlying representation. The strongest evidence for this claim comes from the fact that coronal often appears when other sounds are disallowed, as in the case of the U-shaped pattern of development to be discussed below in (12).

In production, the PoA feature that is available from the lexical representation is used to fill out the unspecified segments. The child is therefore faithful to the underlying representation, and the discrepancy between adult target and child production results from an incomplete representation, which in turn results from the incomplete storage of perceptual features in the phonological representation.

How can we decide between the two accounts, a grammatical constraint versus incomplete storage in the lexical representation, for the initial stage? We opt for the incomplete storage account for the following reasons. First, it is indicative that the period in which the children in Stager and Werker's study had problems with linguistic perception coincides exactly with the period in which children produce the completely harmonic forms. Recent experimental studies on the early perception of TIT and POP forms has also confirmed the hypothesis that initial representations are holistic and underspecified (Fikkert et al. 2005, Fikkert 2006). Furthermore, the harmonic data are cross-linguistically uncommon, and very different from the data from subsequent developmental stages. Assuming a detailed phonological representation for the initial stage renders the developments in the next stage unexpected and hard to account for. Below we discuss how subsequent developments follow from the growing phonological awareness of the learner: segmentalization of the word-unit and the discovery of segmental patterns in both the surrounding language and the child's own lexicon.

4.3. Stages II-IV: Staged segmentation

Segmentation of the unit "word" can be seen as an instance of developing phonological awareness: the ability to deal explicitly with phonological elements (Ferguson and Farwell 1975). In the data of some of the children we saw that as a first step in word segmentation, the category vowel is separated from the category consonant. Where in the initial stage we found predominantly POP and TIT and some KOK patterns, in the second stage we also find TOT, PIP, and for some children also KIK, patterns. From now on we will focus on the consonants.

4.3.1. Labial Left

As soon as consonants in a word can be separately specified, severe limitations on features in combination with certain positions in the word become apparent. The first non-identical, in terms of PoA, combination of consonants is, for every child in our study, PT. The same observation has been made in the early words of children from five different language communities (MacNeilage and Davis 2000), and has been referred to as fronting: a sequencing of consonants proceeding from more forward to more backward places of articulation across the word (Ingram 1974).

According to MacNeilage and Davis (2000) this pattern is basic because it reflects the young child's tendency to start a word in an easy way – a labial consonant only requires a jaw movement, without the additional tongue movement required at the other places of articulation.

Another likely reason for this specific distribution of PoA features within a word to emerge early is the frequency with which it occurs in words from the target language, the input. Words with an initial Labial consonant are highly frequent in Child Directed Speech: Joost van de Weijer (p. c.) reports that 26.19 % of all CVC(V) words directed to a child have a labial segment at C_1, and 19.8 % of all CVC(V) words have an initial labial consonant and a coronal segment at C_2. PT is the most frequent pattern in the input after TT in his database. In (3) above, we saw that the PT intake of children is of a similar magnitude: PT words form 25.49 % of the words in the required vocabulary, and 27.76 % of the attempted adult targets. Among others, Jusczyk, Luce, and Charles-Luce (1994) demonstrated that infants are aware of the relative frequency of occurrence of different phonotactic patterns: infants prefer to listen to words with frequently occurring phonotactic patterns. Zamuner, Gerken, and Hammond (2004) present similar results for older children.

Learners start adding words to their lexicon that have this PT pattern, like *bad* 'bath', *bed* 'bed', *pet* 'cap' and *poes* 'cat', and these targets are faithfully produced by the child, i. e. with a PT pattern. Subsequently, learners analyze their vocabulary and deduce a pattern: labial is connected to C_1. This generalization over the learner's production lexicon gives rise to a preference: Labial should be at the left edge. At this point the lexical pattern "Labial Left" becomes part of the grammar, as a constraint [Labial, and can be overgeneralized. This is illustrated in the tableaux in (8). In order to show the interaction of [Labial with Faithfulness, we have supplied the grammar in the tableaux in (8) with the Faithfulness constraints Max(Lab), Dep(Lab) and Linearity.

(8) OT grammar I: [Labial

a. *poes* 'cat' /pus/

/pus/	[Labial	Max(Lab)	Linearity	Dep(Lab)
☞pus				
puf				*
sup	*!		*	

b. *soep* ‘soup’ /sup/

/sup/	[LABIAL	MAX(LAB)	LINEARITY	DEP(LAB)
sup	*!			
pus			*!	
☞fup				*
sus		*!		

c. klimmen ‘climb’ /klɪmə/

/klɪmə/	[LABIAL	MAX(LAB)	LINEARITY	DEP(LAB)
kɪmə	*!			
mɪkə			*!	
☞ pɪmə				*
kɪ k ə		*!		

While targets like *poes* can be faithfully produced without violating [LABIAL, the faithful candidates for targets like *soep* and *klimmen* are not the optimal candidates. The optimal candidate for *soep* (8b) is [fup], and the optimal candidate for *klimmen* (8c) is [pɪmə] in this particular grammar: they satisfy [LABIAL, and in addition they satisfy the higher ranked faithfulness constraints MAX(LAB) and LINEARITY.

Again, forms like [fup] and [pɪmə] used to be analyzed as resulting from a harmony process between two consonants. In our analysis, however, they result from the interaction of the requirement that Labial be linked to C_1, and the faithfulness constraints MAX(LAB), LINEARITY and DEP(LAB). To illustrate this with chronology, around the age of 1;7.15 Robin starts to attempt more and more adult target words with a PvT structure. As discussed earlier, these targets are produced faithfully. One month later, in the recording at 1;8.12, the first cases of Labial CH appear. Except for TvT and PvP and a quickly disappearing KvK, no other patterns are produced, or even attempted with any frequency. According to our analysis, then, there is no pressure in the grammar for two consonants to share a PoA feature, but this apparent pattern is a consequence of the introduction of the constraint [LABIAL. Additional support for a non-harmonic approach comes from metathesis in child language and from cases where the to-be-aligned feature Labial does not come from an input labial consonant, but from a vowel. In the literature we find the observation that some children metathesize T/KvP forms to PvT/K forms (Menn 1983, Velleman 1995). It is clear that these metathesized forms result from the same [LABIAL constraint in combination with a slightly different ordering of the faithfulness constraints. In the grammar of metathesizing children, LINEARITY, which controls the sequence of segments, is ordered below DEP(LAB), as in (9):

(9) Metathesis of *kip* 'chicken' /kɪp/

/kɪp/	[LABIAL	MAX(LABIAL)	DEP(LAB)	LINEARITY
kɪp	*!			
☞pɪk				*
pɪp			*!	

In (10), data from Robin and Eva show a [LABIAL effect originating with a target labial vowel:

(10) VC "harmony" resulting from [LABIAL

a. *doen* 'do'	/dun/	[bun]	Eva (1;7.15)
b. *schoenen* 'shoes'	/sχunə/	[bunə]	
c. *schoen* 'shoe'	/sxun/	[pun]	Robin (1;8.10)
d. *goed* 'good'	/xut/	[fut]	

As shown in (11) these data result from the same (partial) grammar as the CH and metathesis data above:

(11) VC Harmony of *doen* 'do' /dun/

/dun/	[LABIAL	FAITH(LABIAL)
dun	*!	
☞bun		
din		*!

The question is whether this emergent constraint [LABIAL is a transient constraint, or whether it establishes itself firmly in the grammar as an I-language constraint. If it is part of the I-language grammar, we should find evidence for [LABIAL cross-linguistically. It certainly leaves a trace: the cross-linguistic high frequency of labial initial words (Davis, McNeilage, and Matyear 2002). Grammaticalization of the "Labial Left" lexical pattern in the learner's grammar could re-establish the high frequency of PT/K words in vocabularies. We need to be on the lookout for cases of "The Emergence of The Unmarked" (McCarthy and Prince 1994) that possibly refer to [LABIAL. A first attempt to experimentally test this claim was a rhyming experiment in which subjects were found to supply rhyme-words more often with an initial labial consonant than with other places of articulation. This suggests that both older children and adults have a preference for initial labials (Fikkert et al. 2004).

*4.3.2. *[Dorsal*

Labial and dorsal are considered to be marked PoA features. During development it appears that labial and dorsal segments are in complementary distribution for a while: labial becomes specifically linked with C_1, while dorsal is banned from this position. In the data of some children, like Eva, we find evidence for both a general ban on dorsal and, later, the more specific constraint banning dorsal from initial position. In the data of most children, however, target words containing dorsal in C_2 position, like *dragen* 'carry', *drinken* 'drink' and *pakken* 'catch', appear, and are faithfully produced. Moreover, target words containing Dorsal in C_1 position, which were produced faithfully in the initial "holistic" stage, e. g., *koek* /kuk/ 'cookie', all of a sudden are produced unfaithfully, faithful [kuk] becomes unfaithful [tuk]. This U-shaped developmental pattern is especially salient in the data of Noortje.

(12) U-shaped development (data from Noortje)

a. Stage I					
	KOK	*koek* 'cookie'	→	[kuk]	(2;3.7)
		klok 'clock'	→	[kɔk]	(2;5.23)
	KIK	*kikker* 'frog'	→	[kɪk]	(2;2.21)
		kijk 'look'	→	[kɛik]	(2;5.23)
b. Stage III					
	KOK	*koek* 'cookie'	→	[touk]	(2;8.17)
		klok 'clock'	→	[tɔk]	(2;8.17)
	KIK	*kijk* 'look'	→	[tɛik]	(2;8.17)
		kikker 'frog'	→	[tika]	(2;9.1)
c. Later stage					
	KOK	*kruk* 'stool'	→	[kyk]	(2;9.29)
		kuiken 'chicken'	→	[kœyk]	(2;10.12)

In (12a) we see that dorsal-initial target words are faithfully produced in the early stages in which the word is not, or hardly, segmentalized. The data in (12b) show a sudden dislike for dorsal: the exact same target words from the earlier stage are no longer produced faithfully. U-shaped patterns are not uncommon in acquisition data (Stemberger, Bernhardt, and Johnson 1999), but are hard to account for in a traditional OT account where the initial state is Markedness >> Faithfulness. We cannot account for the development from (12a) to (12b) by referring to changes in the ranking between an innate markedness constraint of the type *[DORSAL "No initial Dorsals" and FAITH(DORS), since the demotion of *[DORSAL would give rise to more faithful productions, rather than less faithful ones. We assume, then, that *[DORSAL has emerged in the grammar, in a high-ranked position. In (12c), finally, the constraint against dorsal in initial position has lost its force, and target initial dorsals can be produced faithfully again. In (13), we provide examples of other dorsal-initial targets that appear simultaneously with the data in (12b) and have no initial dorsal in the child's production.

(13) Initial K > T elsewhere (data from Noortje)

a. KIT > TIT

kleine 'little'	→	[tɛinə]	(2;8.17)
kind 'child'	→	[tɪnts]	(2;9.15)
kers 'cherry'	→	[tɛs]	(2;9.29)

b. KAT > TAT

koud 'cold'	→	[tɑuts]	(2;7.2)
kan 'can'	→	[tɑnə]	(2;9.1)

c. KOT > TOT

grote 'big'	→	[dotə]	(2;9.1)
kousen 'stockings'	→	[tɑusa]	(2;10.26)

d. KIP > PIP > TIP

kip 'chicken'	→	[pɪp]	(2;6.5)
		[tɪp]	(2;10.12)

The example in (13d) merits some additional information. Because of the two constraints [LABIAL and *[DORSAL, discussed above, neither TP nor KP targets can be produced faithfully for some time. As long as [LABIAL is high-ranked, both TP and KP targets will be produced PP, hence [pɪp] for KP target *kip* 'chicken' and [fup] for TP target *soep* 'soup'. The force of [LABIAL is the first to wane. This results in faithful productions of TP targets. KP targets are still problematic because of *[DORSAL. The production of KP targets does, however, evolve, namely from PP to TP: target *kip* is now produced [tɪp].

What happens to *[DORSAL in the grammar? Again we could say that it leaves a frequency trace in the language: Dorsal-initial words have a relatively low frequency compared to labial-initial and coronal-initial words. In the Child Directed Speech database of Van de Weijer (1998), the distribution of the PoA of initial consonants is the following (Van de Weijer, p. c.): Coronal

51 %, Labial 25 %, Dorsal 11 %, and "other" (/h/ and orthographic "r" which is hard to classify) 13 %. Its effect can also be seen in the nasal stop series: the dorsal nasal is banned from C_1 position in Dutch, as well as in many other languages.

4.3.3. Input/intake and order of development

Why do the Dutch children have the particular order of development of PoA patterns that is observed and not some different order? Why are PvT words so early and TvP words so late? It turns out that as soon as consonants with different PoA features can be combined in production, at Stage III, the order of acquisition correlates very well with the distribution of the different PoA patterns in the surrounding language, in this case Dutch. This is shown in (14) for the list of expected words:

(14) Correlation intake-development I

List of required words			*Development*	
PvT	233	25.49 %	Stage III	PvT
TvT	143	15.65 %		
TvK	**121**	**13.24 %**	Stage IV	TvK PvK
PvK	**103**	**11.27 %**		
KvT	**99**	**10.83 %**	Stage V	KvT TvP KvP
TvP	**96**	**10.5 %**		
PvP	47	5.14 %		
KvP	**47**	**5.14 %**		
KvK	25	2.74 %		

As can be seen, PvT is the most frequent PoA pattern in the set of 914 Dutch input words, and PvT is also the first pattern that is produced after the initial two "whole word" stages. The K-final patterns TvK and PvK have the next highest frequencies, and also occur next in production. This is followed by the K-initial pattern KvT, both in frequency and in appearance. The P-final patterns TvP and KvP have the lowest frequencies and are also produced last. If attempted adult targets are also an indication of adult input (de Boysson-Bardies and Vihman 1991), then almost the same correlation is found, with a slight discrepancy between the KvT and PvK (italicized):

(15) Correlation intake-development II

Attempted adult targets		*Development*	
PvT	27.8 %	Stage III	PvT
TvT	25.7 %		
TvK	**9.8 %**	Stage IV	TvK *PvK*
KvT	***9.5 %***		
PvK	***8.5 %***	Stage V	*KvT* TvP KvP
TvP	**6 %**		
PvP	5.7 %		
KvK	4 %		
KvP	**3.1 %**		

The distribution of PoA patterns in the Child Directed Speech data correlates less well for the KvT and PvK patterns (italicized in (16)): KvT has a relatively high frequency in this set of data, and PvK a relatively low frequency. However, it does correlate for the PvT pattern and the P-final patterns.

(16) Correlation input-development

Child Directed Speech		*Development*	
TvT	331 %		
PvT	**27.2 %**	Stage III	PvT
KvT	*12.5 %*		
TvK	**10.7 %**	Stage IV	TvK *PvK*
KvP	**6.2 %**	Stage V	*KvT* TvP KvP
PvK	*4.1 %*		
TvP	**2.6 %**		
KvK	2.5 %		
PvP	1.1 %		

As can be seen in (14), (15) and (16), intake frequency does not correlate particularly well with PoA development in the earliest stages, Stage I and Stage II;

specifically, KK and PP are of very low frequency, yet are produced very early. This is consistent with the claim that in the initial two stages the language learner has a different, less detailed, lexical representation. However, input frequency does correlate with PoA development as soon as consonants can receive separate PoA feature specifications. Given that differences in frequency can be very minor, a perfect correlation between input/intake frequency and developmental order can hardly be expected. On linguistic grounds we expect to find generalizations over these patterns, in terms of Dorsal-initial patterns, Labial-initial patterns, etc. This is worked out in detail in Fikkert, Levelt, and Van de Weijer (2002).

We proposed that the constraint underlying the apparent cases of Labial consonant harmony, [LABIAL, emerged in the grammar based on the child's lexicon. The word patterns in the early lexicon correlate with high-frequency word patterns in the input/intake. Indirectly, then, Labial harmony in Dutch child language can be traced back to the high frequency of Labial-initial words in the adult input.

5. Summary and conclusions

In this paper we provided evidence that there is a fixed order of development of PoA contrast in words, both in production and in the selected targets. At the first stage, the word is an unanalyzed whole and we find only a single PoA in a word. We thus find words with the POP, TIT, and for some children KOK patterns, in addition to PAP, TAT and sometimes KAK patterns. Although these words are harmonic, they are clearly not the result of an assimilatory process between consonants, but show that PoA is not yet contrastively used within words

At stage two, segmentalization of words starts. For most children the vowel becomes separately specifiable from the rest of the word, i. e. the consonants, and we often find a separate PoA for the vowel and consonants; i. e. vowels and consonants can now contrast in PoA within a word. Here we find patterns such as TOT, PIP and KIK but combinations of different consonants have yet to appear.

At stage three, further segmentalization takes place and this development follows a strict pattern; first, we find that labial consonants are preferred at the left edge of the word, and subsequently we find that dorsal consonants are preferred at the right edge but banned from the left edge. At this third stage, the lexicon is first rapidly expanded with words from the target language that have the PoA structure PvT. Words with this structure are highly frequent in the target language. These labial-initial words are faithfully produced. Based on these forms, the generalization "Labial is at the left edge of the word" is made, and this generalization becomes grammaticalized as a high-ranking constraint in the child's grammar. It is the constraint [Labial that is responsible for apparent cases of Labial consonant harmony. In a similar fashion, *[Dorsal emerges in the grammar. This constraint is consistent with the target language lexicon where a dorsal specification is relatively infrequent at C_1 and relatively frequent at C_2. PvK and TvK words are added to the lexicon, and are produced faithfully. Subsequently, *[Dorsal emerges high-ranked in the grammar.

The presence of these constraints in the grammar is reflected in unfaithful productions that promote labials and ban dorsals in initial position in the production data. U-shaped developmental patterns are a consequence of these emerging markedness constraints. As soon as, for example, *[Dorsal emerges, *koek* is realized as [tuk] rather than the earlier, faithful production [kuk]. It seems that at this point in development, learners build constraints into the grammar based on the structure of their individual lexicons or intake. A question that arises is whether we should regard all markedness constraints as emergent constraints (Boersma 1998). Is it possible to find a principled difference between an innate set of constant, stable markedness constraints, and these emergent markedness constraints that reflect the learner's focus on word-sized units? If there is such a difference, it might well be the one between prosodic markedness constraints, which are relatively uncontroversial in phonological theory, and the more elusive segmental markedness constraints.

With regard to the role of input frequency we found that at the stage where language learners are able to segmentalize the words in their lexical representation, input frequency appears to determine the learner's choice for certain lexical patterns. Input frequency does not affect the first "holistic" stage, as PP and KK patterns are not very frequent but are nevertheless produced early. However, when words are segmentalized, the most frequent pattern in the input, PT, appears first. Patterns of low frequency, like KP, appear late. The frequencies of the intermediate input patterns lie quite close together and children vary in how they expand their lexicons with respect to these patterns. It thus seems that early production patterns and input frequency conspire towards the emergence of markedness constraints in the grammar.

We have shown that children start out with faithful productions of targets, and thus, that Faithfulness – at least with respect to the underlying PoA structure – is active in the early stages of grammatical development. The generally assumed initial state of the grammar, Markedness >> Faithfulness, cannot account for emerging unfaithfulness in development. We might thus need to differentiate between universal markedness constraints and emergent markedness constraints that are based on phonological characteristics of the lexicon at a specific stage. This requires careful and detailed studies of developmental data, particularly of languages that have a different distribution of PoA patterns in words, as these are predicted to show different emerging constraints.

What still needs to be explained is the fact that children initially appear to select specific words for production. This is a problem for any theory of phonological acquisition. We will offer a hypothesis here. It is clear from studies like Stager & Werker 1997 that there can be a discrepancy between the learner's perceptual abilities and his or her representation of perceived features in the mental lexicon, i. e. more information is perceived than stored. When a word with an incomplete representation is produced, it is likely that this produced form will deviate from the form that was originally perceived by the learner. If the learner has a high perceptual standard, i. e., he or she knows what the word is supposed to sound like, the form that the production system comes up with will mismatch the perceptual standard. Initially, then, children prefer to play safe, phonologically speaking, producing only those forms that match their perceptual standard. However, since their production system develops slower than their communicative needs, at some point they trade phonological security for more expressive power and allow for mismatching productions.

To conclude, our analysis departs from classical OT accounts in two respects. First, lexical representations are not adult-like from the start; words appear to be unsegmentalized units at first. Segmentalization leads to the emergence of position-specific constraints in the grammar. Second, not all constraints are innate; constraints may emerge as children generalize over their lexicons. If it is indeed the case that both labial and dorsal harmony are the result of emergent constraints, a challenge for future work is to understand why in some child languages labial harmony is prevalent, while in others, notably English, dorsal harmony is more common. Our prediction is that different constraints emerge in different languages (see Fikkert, Levelt, and Van de Weijer, 2002) depending on the frequency of phonological patterns in the input and intake of the learners of these languages.

Finally, CH-like forms, i. e. forms that would previously have been analyzed as resulting from a harmony process between consonants, must be viewed as an epiphenomenon of the phonological contents of the lexicon, in combination with an immature planning and production system. The developmental lexicon is child-specific, therefore CH is child-language specific. There is no need to account for the fact that CH of primary place features does not occur in adult languages, since the constraints that produce this effect have become obsolete and are simply no longer part of adult grammars.

References

Barlow, Jessica
1997 A constraint-based account of syllable onsets: Evidence from developing systems. Ph. D. Dissertation. Indiana University.

Bernhard, Barbara and Joseph Stemberger
1998 *Handbook of Phonological Development: From the Perspective of Constraint-based Nonlinear Phonology.* San Diego: Academic Press.

Boersma, Paul
1998 Functional phonology: Formalizing the interactions between articulatory and perceptual drives. Ph. D. Dissertation. University of Amsterdam.

Boersma, Paul and Clara C. Levelt
2003 Optimality Theory and phonological acquisition. *Annual Review of Language Acquisition* 3: 1–50.

Boysson-Bardies, Bénédicte de and Marilyn Vihman
1991 Adaption to language: Evidence from babbling and first words in four languages. *Language* 67: 297–319.

Davis, Barbara L., Peter F. MacNeilage and Christine L. Matyear
2002 Acquisition of serial complexity in speech production: A comparison of phonetic and phonological approaches to first word production. *Phonetica* 59: 75–109.

Ferguson, Charles A. and Carol Farwell
1975 Words and sounds in early language acquisition. *Language* 51: 419–439.

Fikkert, Paula
1994 *On the Acquisition of Prosodic Structure*. Ph. D. Dissertation. University of Leiden. (Holland Institute of Generative Linguistics Dissertations in Linguistics 6). The Hague: Holland Academic Graphics.

Fikkert, Paula
2006 Developing representations and the emergence of phonology: Evidence from perception and production. *Proceedings of LabPhon 10*, Paris, France.

Fikkert, Paula, Marieke van Heugten, Philo Offermans and Tania S. Zamuner
2005 Rhymes as a window into grammar. *Proceedings of the 29th Annual Boston University Conference on Language Development*, 204–215. Somerville, Ma: Cascadilla Press.

Fikkert, Paula, Clara C. Levelt and Joost van de Weijer
2002 Input, intake and phonological development: The case of consonant harmony. Paper presented at GALA, Utrecht.

Fikkert, Paula, Clara C. Levelt and Tania Zamuner
2005 Learning underlying forms: Evidence from child perception and production. Paper presented at NELS 36: the Northeastern Linguistic Society. University of Massachusetts, Amherst, 28–30 September 2005.

Frisch Stefan A., Janet Pierrehumbert and Michael Broe
2004 Similarity Avoidance and the OCP. *Natural Language and Linguistic Theory* 22: 179–228.

Gnanadesikan, Amalia
2004 Markedness and faithfulness constraints in child phonology. In: René Kager, Joe Pater and Wim Zonneveld (eds.), *Constraints in Phonological Acquisition*, 73–108. Cambridge: Cambridge University Press.

Goad, Heather
1998 Consonant harmony in child language: An optimality-theoretic account. In: S. J. Hannahs and Martha Young-Scholten (eds.), *Focus on Phonological Acquisition*, 113–142. Amsterdam: John Benjamins.

Goad, Heather
2001 Assimilation phenomena and initial constraint ranking in early grammars. *Proceedings of the 25th Annual Boston University Conference on Language Development*, 307–318. Somerville, Ma: Cascadilla Press.

Goad, Heather
2003 Licensing and directional asymmetries in Consonant Harmony. Poster presented at the Child Phonology Conference, Vancouver, B. C. July 1–4, 2003.

Ingram, David
1974 Phonological rules in young children. *Journal of Child Language* 1: 29–64.

Jongstra, Wenckje
2003 Variation in reduction strategies in Dutch word-initial consonant clusters. Ph. D. Dissertation. University of Toronto.

Jusczyk, Peter W.
1997 *The Discovery of Spoken Language*. Cambridge, Ma: The MIT Press.

Jusczyk, Peter W., Paul Luce and Jan Charles-Luce
1994 Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language* 33: 630–645.

Kuhl, Patricia K.
2000 A new view of language acquisition. *Proceedings of the National Academy of Science* 97: 11 850–11 857.

Lahiri, Aditi and Vincent Evers
1991 Palatalization and coronality. In: Carol Paradis and Jean-François Prunet (eds.), *The Special Status of Coronals: Internal and External Evidence*. (Phonology and Phonetics 2), 79–100. San Diego: Academic Press.

Levelt, Clara C.
1994 *On the Acquisition of Place*. Ph. D. Dissertation. Leiden University. (Holland Institute of Generative Linguistics Dissertations in Linguistics 8). The Hague: Holland Academic Graphics.

Levelt, Clara C.
1995 The segmental structure of early words: Articulatory frames or phonological constraints. In: Eve Clark (ed.), *Proceedings of the 27th Child Language Research Forum*, 19–27. Stanford, CA: CSLI.

Macken, Marcy
1978 Permitted complexity in phonological development: One child's acquisition of Spanish consonants. *Lingua* 44: 219–253.

MacNeilage, Peter F. and Barbara L. Davis
2000 Origin of the internal structure of word forms. *Science* 288: 527–531.

MacWhinney, Brian
2000 *The CHILDES Project: Tools for Analyzing Talk. 3rd Ed. Vol. 2: The Databases*. Mahwah, NJ: Lawrence Erlbaum.

McCarthy, John and Alan Prince
1994 The emergence of the unmarked. In: Mercè Gonzàlez (ed.), *NELS 24: Proceedings of the Northeastern Linguistic Society, Vol. 2*, 333–379. Amherst, Ma: GSLA.

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

Moskowitz, Arlene J.
1973 The acquisition of phonology and syntax: A preliminary study. In: Jaakko Hintikka *et al*. (eds.), *Approaches to Natural Language: Proceedings of the 1970 Stanford Workshop on Grammar and Semantics*, 48–84. Dordrecht: Reidel Publishers.

Paradis, Carol and Jean-François Prunet (eds.)
1991 *Phonetics and Phonology: Vol 2. The Special Status of Coronals*. San Diego, CA: Academic Press.

Pater, Joe
2002 Form and substance in phonological development. In: Line Mikkelson and Chris Potts (eds.), *Proceedings of the West Coast Conference on Formal Linguistics 21*, 348–372. Somerville, Ma: Cascadilla Press.

Pater, Joe and Adam Werle
2001 Typology and variation in child consonant harmony. In: Caroline Féry, Antony Dubach Green and Ruben van de Vijver (eds.), *Proceedings of the Holland Institute of Generative Linguistics Phonology Conference (HILP) 5*, 119–139. University of Potsdam.

Pater, Joe and Adam Werle
2003 Direction of assimilation in child consonant harmony. *Canadian Journal of Linguistics* 48 (3/4): 385–408.

Pater, Joe, Christine Stager and Janet Werker
2004 The perceptual acquisition of phonological contrasts. *Language* 80: 384–402.

Prince, Alan and Paul Smolensky
2004 *Optimality Theory: Constraint Interaction in Generative Grammar.* Amsterdam: Blackwell.

Schaerlaekens, Annemarie, Dolf Kohnstamm and Maryline Lejaegere
1999 *Streefwoordenschat van zesjarigen: derde herziene versie, gebaseerd op nieuw onderzoek in Nederland en België*. [Expected vocabulary of

six-year olds: third revised edition, based on new research in The Netherlands and Belgium]. Lisse: Swets en Zeitlinger.

Stager, Christine L. and Janet F. Werker
1997 Infants listen for more phonetic detail in speech perception than in word learning tasks. *Nature* 388: 381–382.

Stemberger, Joe, Barbara Bernhardt and Carl Johnson
2001 "Regressions" ("u"-shaped learning) in the acquisition of prosodic structure. *Rutgers Optimality Archive* [ROA-471].

Torgerson, Warren
1963 *Theory and Methods of Scaling*. New York: Wiley.

Velleman, Shelley
1995 Metathesis highlights feature-by-position constraints. In: Barbara Bernardt, John Gilbert and David Ingram (eds.), *Proceedings of the University of British Columbia International Conference on Phonological Acquisition*, 173–186. Somerville, Ma: Cascadilla Press.

Vihman, Marilyn
1996 *Phonological Development. The Origins of Language in the Child*. Oxford: Blackwell.

Vihman, Marilyn and Shelley Velleman
2000 Phonetics and the origins of phonology. In Noel Burton-Roberts, Philip Carr and Gerard Docherty (eds.), *Phonological Knowledge*, 305–339. Oxford: Oxford University Press.

Vihman, Marilyn, Shelley Velleman and Lorraine McCune
1993 How abstract is child phonology? Towards an integration of linguistic and psychological approaches. In M. Yavas (ed.), *First and Second Language Phonology*, 9–44. San Diego: Singular Publishing Group.

Waterson, Natalie
1971 Child phonology: A prosodic view. *Journal of Linguistics* 7: 179–211.

Waterson, Natalie
1987 *Prosodic Phonology: The Theory and its Application to Language Acquisition and Speech Processing*. Newcastle upon Tyne: Grevatt & Grevatt.

Weijer, van de Joost
1998 *Language Input for Word Discovery*. Ph. D. Dissertation. Nijmegen: Max Planck Series in Psycholinguistics 9.

Werker, Janet F., Chris Fennell, Kathleen Corcoran and Christine L. Stager
2002 Age and vocabulary size influences on the phonological representation of newly learned words in infants aged 14 to 20 months. *Infancy* 3: 1–30.

Zamuner, Tania S., Lou-Ann Gerken and Michael Hammond
2004 Phonotactic probabilities in young children's production of coda consonants. *Journal of Child Language* 31: 515–536.

Zink, Inge
2001 *N-CDIs: Lijsten voor Communicanieve Ontwikkeling. Aanpassing en hernormering van de MacArthur CDI`s* [N-CDIs: Lists for communicative development. Adaptations and re-standardizing of the MacArthur CDIs]. Leuven: Acco.

Appendix A: PoA patterns in production

Tom

	PVP	TVT	PVT	PVK	TVK	KVK	TVP	KVT	KVP
I 1;2.14–1;3.24	PAP POP	TAT TIT							
II/III 1;3.24–1;5.0	PIP	TOT	PAT						
III/IV 1;5.0–1;5.28			PIT	PAK					
IV/V 1;5.28–1;6.11				PIK POK	TIK TAK TOK	KIK KAK KOK	TIP TAP TOP	KIT KAT KOT	KOP KAP KIP

Noortje

	PVP	TVT	KVK	PVT	PVK	TVK	TVP	KVT	KVP
I/II 1;7.14–2;2.21	PAP POP PIP	TAT TIT TOT	KAK KOK KIK						
III 2;2.21–2;4.4				PAT POT PIT					
IV 2;4.4–2;7.2					PAK PIK POK	TIK TAK TOK			
V 2;7.2–2;11			KIK KAK KOK				TIP TAP TOP		
V 2;11								KAT KOT	KOP KAP KIP

Eva

	PVP	TVT	KVK	PVT	PVK	TVK	TVP	KVT	KVP
I 1;4.12– 1;4.26	PAP POP	TAT TIT		PIT					

	PVP	TVT	KVK	PVT	PVK	TVK	TVP	KVT	KVP
II 1;4.26–1;8.12		TOT		POT PAT	POK PIK	TAK TIK TOK			
V 1;8.12–1;11.8	PIP				PAK		TOP TAP TIP		
V 1;11.8–			KIK KOK KAK						KOP

Appendix B: Targets

Eva

	PVP	TVT	PVT	PVK	TVK	TVP	KVK	KVT	KVP
I 1;4.12–1;4.26	PAP POP	TAT TIT	PAT POT PIT	PIK POK	TIK	TOP			
II 1;4.26–1;6.1				PAK			KOK KIK	KIT	
III 1;6.1–1;8.12		TOT			TOK	TIP		KAT	KOP
IV 1;8.12–1;11.8	PIP				TAK	TAP	KAK	KOT	KAP KIP

Noortje

	PVP	TVT	PVT	KVK	PVK	TVK	TVP	KVT	KVP
I 1;7.14–2;3.21	PAP POP	TAT TIT TOT	PIT PAT	KOK KIK					
II 2;3.21–2;5.23	PIP		POT						

	PVP	TVT	PVT	KVK	PVK	TVK	TVP	KVT	KVP
III 2;5.23–2;6.5					PAK PIK POK	TOK TAK TIK			
IV 2;6.5–							TIP TAP TOP	KOT KAT KIT	KAP KIP KOP

Tom

	PVP	TVT	PVT	KVK	PVK	TVK	TVP	KVT	KVP
I 1;0.24–1;3.14	PAP POP	TAT TIT	PAT	KOK					
II 1;3.14–1;5.28	PIP	TOT	POT PIT		PIK	TIK			
III 1;5.28–1;6.25				KAK KIK	PAK POK	TOK TAK	TAP	KIT KAT	KIP
IV 1;6.25–							TIP TOP	KOT	KAP KOP

Jarmo

	PVP	TVT	PVT	KVK	PVK	TVK	TVP	KVT	KVP
I 1;4.18–1;7.15	PAP	TAT TIT		KOK KIK					
II 1;7.15–1;9.23	PIP POP		PAT POT PIT		POK PIK	TIK	TIP TOP	KAT	KAP
III 1;9.23–1;11.20		TOT					TAP		
IV 1;11.20–					PAK	TOK		KOT	KIP KOP

Appendix C: Faithful productions

Table IV: Faithful vs. unfaithful use of PoA pattern

Eva

PoA pattern in production	First Faithful use	First Unfaithful use for Targets:	
TIT	1;4.12	1;4.12	PIT
			PIK
		1;4.26	KIT
		1;4.26	KIK
TAT	1;4.12	1;4.26	PAK
		1;6.1	PAT
		1;7.15	KAT
		1;8.12	KAK
PAP	1;4.12	1;6.1	PAT
		1;4.26	TAP
POP	1;4.12	1;4.12	POT
			TOP
			POK
		1;6.1	TOK
		1;6.1	KOP
PAT	1;4.12	1;7.15	PAK
*TIK	1;4.26	1;4.12	PIK
		1;6.1	KIT
		1;9.22	KIK
TOT	1;6.12	1;7.22	KOK
		1;9.8	KOT
		1;11.8	TOK
POT	1;6.12	1;6.12	TOT
		1;5.22	KOK
			POK
*TAK	1;8.12	1;6.1	PAK
			KAT
TOP	1;8.12	1;8.12	KOP

Noortje

PoA pattern in production	First Faithful use	First Unfaithful use for Targets:	
PAP	1;7.14	2;6.5	TAP
POP	2;1.17	2;6.5	KOP
TAT	2;2.21	2;7.2	KAT

PoA pattern in production	First Faithful use	First Unfaithful use for Targets:	
TIT	2;3.7	2;1.17 2;5.23	PIT KIT
TOT	2;3.7	2;7.2	KOT
PIP	2;3.21	2;6.5	TIP KIP
TOK	2;5.23	2;7.16	KOK
TIK	2;5.23	2;7.16	KIK

Tom

PoA pattern in production	First Faithful use	First Unfaithful use for Targets:	
PAP	1;1.21	1;3.14 1;5.28	PIK TAP
POP	1;5.0	1;5.28 1;6.11	POK POT
TAT	1;2.27	1;3.14 1;6.25	TIK KIT
TIK	1;5.28	1;5.28	KIK

Jarmo

PoA pattern in production	First Faithful use	First Unfaithful use for Targets:	
PAP	1;6.27	1;9.9 1;10.23	PAT TAP
TIT	1;6.27	1;8.26 1;9.23 1;10.9 2;1.22	TIK TIP KIK PIT
KOK	1;6.27	1;10.23 2;0.4	TOK KOT
TAT	1;7.28	1;7.28 2;0.28	PAT TIK
POP	1;8.26	1;9.9	POK
TOT	1;10.23	1;11.20	POT
*KOT	2;0.28	1;11.20 2;2.6	KOP POK
*KOP	2;2.27	2;0.28	POK KOT
*KAK	?	1;8.12 2;0.28	KAT PAK

Acquisition

Second language (L2) acquisition

Learning to perceive a smaller L2 vowel inventory: An Optimality Theory account

Paul Boersma and Paola Escudero

This paper gives an Optimality-Theoretic formalization of several aspects of the acquisition of phonological perception in a second language. The subject matter will be the acquisition of the Spanish vowel system by Dutch learners of Spanish, as evidenced in a listening experiment. Since an explanation of the learners' acquisition path requires knowledge of both the Dutch and the Spanish vowel system, the 12 Dutch and 5 Spanish vowels are presented in Figure 1. Along the vertical axis we find the auditory correlate of perceptual vowel height (first formant, F1), and along the horizontal axis the auditory correlate of perceptual vowel backness (second formant, F2), whose articulatory correlates are tongue backness and lip rounding. A third auditory dimension, duration, is implicit in the length sign ("ː") used for 4 of the 12 Dutch vowels.

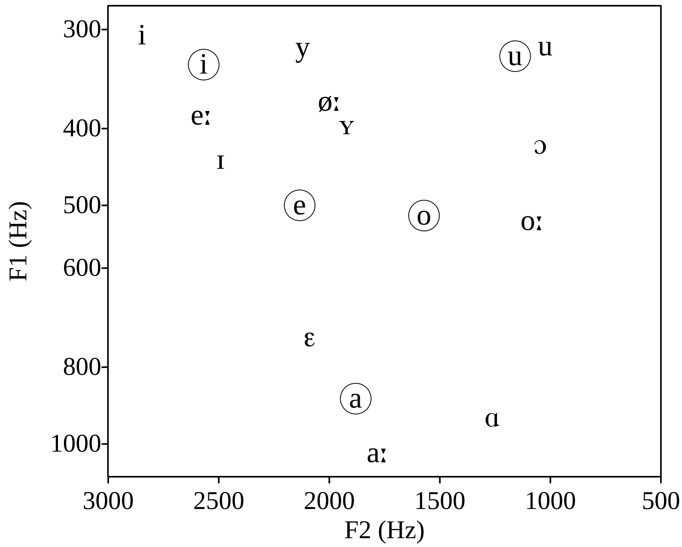


Figure 1. The 5 Spanish vowels (circled) amidst the 12 Dutch vowels.

To control for speaker-dependent vocal tract dimensions, we based the two sets of formant values in Figure 1 on the speech of a single speaker, a perfect

Spanish-Dutch bilingual (moved to the Netherlands when she was 12, currently a teacher of Spanish speaking proficiency at the University of Amsterdam, with no noticeable foreign accent in either Dutch or Spanish). We see the usual features of the Dutch vowel system: /i/, /y/ and /u/ at the same height, /eː/ and /øː/ at the same height, /ɪ/ and /ʏ/ at the same height, /ɛ/ more open than /ɔ/, /ɑ/ more open than /ɛ/ but somewhat closer than /aː/. As for most speakers of Dutch, /aː/ is front and /ɑ/ is back. As for many speakers, /ɪ/ and /ʏ/ are a bit lower than /eː/ and /øː/. The height of /ɔ/ shows that this speaker is from one of those large areas that merge the reflexes of both historical /ɔ/ and /ʊ/ into a single relatively high variant at the height of /ɪ/ and /ʏ/ (if this had been true of all speakers of Dutch, a better symbol for the phoneme /ɔ/ would have been /ʊ/). A more idiosyncratic feature of the speaker's regional accent is the low position in the chart of the vowel /oː/, which is due to its large degree of diphthongization (i. e., the three higher mid vowels are phonetically realized by this speaker as [ei], [øy], [ɔu]). As for this speaker's Spanish vowel system, we see that /a/ is rather front, that /e/ and /o/ are not close to any Dutch vowel, and that the extent of the Spanish vowel space is somewhat smaller than that of the Dutch vowel space, with a notable centralization of /o/. The patterns are compatible with what is known about Dutch (Pols, Tromp, and Plomp 1973; Koopmans-Van Beinum 1980), about Spanish (Bradlow 1995, 1996), and about the crosslinguistic correlation between the size of a language's auditory vowel space and the size of its vowel inventory (Liljencrants and Lindblom 1972; Lindblom 1986).

1. Ease and difficulty for Dutch learners of Spanish vowels

For Dutch learners of Spanish who want to master the Spanish vowel system, there is something easy as well as something difficult about it. The ease lies in creating lexical representations for Spanish vowels, while the difficulty lies in perception, i. e. in the mapping from raw auditory data to discrete representations that can be used for lexical access.

1.1. Easy: Lexical symbols for L2 vowels

When native speakers of Dutch learn to use the vowel system of the Spanish language, they seem to have the advantage that the target language has fewer vowels than their native language, so that they have the option of reusing a subset of their native vowel categories for the storage of Spanish lexemes. The phonological representations of entries in the Spanish lexicon can get by with only five vowel categories, which we will denote as |a|$_{\text{S}}$, |e|$_{\text{S}}$, |i|$_{\text{S}}$, |o|$_{\text{S}}$, and |u|$_{\text{S}}$ (in our notations, subscript S is used for structures in the minds of native speakers of Spanish, and underlying forms are given within pipes).[1] Thus, the lexical representation of the word *centrifugado* 'centrifugated' is |θentɾifuɣaðo|$_{\text{S}}$ for native speakers of (European) Spanish. Native speakers of Dutch have to maintain at least 12 vowel categories in their native lexical representations: |ɑ|$_{\text{D}}$, |aː|$_{\text{D}}$, |ɛ|$_{\text{D}}$, |ɪ|$_{\text{D}}$, |eː|$_{\text{D}}$, |i|$_{\text{D}}$, |ʏ|$_{\text{D}}$, |øː|$_{\text{D}}$, |y|$_{\text{D}}$, |ɔ|$_{\text{D}}$, |oː|$_{\text{D}}$, |u|$_{\text{D}}$ (subscript D for structures in the minds of native speakers of Dutch). When learning Spanish, then, they could simply[2] reuse five of these for representing their L2 Spanish lexemes; no category split, no category creation would be necessary. As we will see when discussing the results of our listening experiment (§ 1.5), this is what the learners indeed seem to do. The following simplified list shows which Dutch vowels are reused for which Spanish vowels in the interlanguage:

(1) Identification of lexical symbols for Dutch learners of Spanish

|ɑ|$_{\text{D}}$ – |a|$_{\text{S}}$
|ɛ|$_{\text{D}}$ – |e|$_{\text{S}}$
|i|$_{\text{D}}$ – |i|$_{\text{S}}$
|ɔ|$_{\text{D}}$ – |o|$_{\text{S}}$
|u|$_{\text{D}}$ – |u|$_{\text{S}}$

Note that this identification does not describe the knowledge of the learners; rather, it is an observation that we as linguists can infer from experimental tasks (as we do in § 1.5). The identification in (1) means, for instance, that the Dutch learner's underlying representation of Spanish *centrifugado* is |θɛntɾifugɑdɔ|$_{\text{D}}$. Also note that our use of vowel symbols is not meant to suggest crosslinguistic identity: |u|$_{\text{D}}$ is not a priori more similar to |u|$_{\text{S}}$ than |ɑ|$_{\text{D}}$ is to |a|$_{\text{S}}$.[3]

1 We use pipes in order to distinguish underlying forms from phonological surface structures, which are given between /slashes/, and auditory phonetic forms, which are given in an approximate IPA transcription between [square brackets].

2 Escudero (2005: 214–236) investigates and models (in Optimality Theory) the possibility that category reuse is not an "easy" instantaneous act that occurs magically at the start of L2 acquisition after all. Escudero proposes instead that category reuse gradually emerges as an automatic result of an initial creation of lexical items with multiple underlying phonological representations and a subsequent reduction of this lexical variability by the process of message-driven learning of recognition.

3 Nor *less* similar. A theory of phonology that regards all vowels as a combination of innate (hence crosslinguistically identical) phonological feature values may even consider every vowel at the left in (1) as featurally identical to its counterpart at the right.

1.2. Difficult: Perceptual boundaries of L2 vowels

While the reuse of existing categories is advantageous in itself, there is an additional gain in the identifications in (1), which are far from arbitrary. This section first shows that these identifications are largely based on language-specific perceived (auditory and structural) similarity, and then shows why such an identification strategy is advantageous.

Typical tokens of an intended native Spanish |a|$_{\mathrm{S}}$ tend to sound like a short somewhat front open vowel, which in a narrow auditory-phonetic transcription is [a] or [a̠]. The spectral quality (F1 and F2) of these tokens is close to that of typical tokens of Dutch |aː|$_{\mathrm{D}}$, which are phonetically realized like the long cardinal IPA open front vowel [aː]; the duration of the Spanish tokens, however, is close to that of typical tokens of Dutch |ɑ|$_{\mathrm{D}}$, which typically sound like the slightly rounded low back vowel [ɑ̹]. Since Dutch listeners, when having to categorize sounds in the [a]-[ɑ̹]-[ɑ̹ː]-[aː] region, weigh the duration cue much higher than the spectral cues (Gerrits 2001: 89), they will classify the Spanish [a]-like tokens as /ɑ/$_{\mathrm{D}}$ rather than as /aː/$_{\mathrm{D}}$.[4] Another option is to perceive these tokens as /ɛ/$_{\mathrm{D}}$, whose typical realizations in Dutch sound like the cardinal IPA open mid front vowel [ɛ]. In the listening experiment partly discussed below we found that non-Spanish-learning speakers of Dutch perceived Spanish |a|$_{\mathrm{S}}$ as /ɑ/$_{\mathrm{D}}$ 60 percent of the time, as /ɛ/$_{\mathrm{D}}$ 27 percent of the time, and as /aː/$_{\mathrm{D}}$ 4 percent of the time. So it seems that language-specifically perceived similarity, with duration as the main determining cue, largely explains the identifications in (1).[5]

So why would learners choose to base their identifications on perceived similarity, i.e. what advantage does it give them to reuse Dutch categories whose auditory distributions include the most typical tokens of the Spanish correspondents, as in (1)? To answer this, we have to consider what is involved in the listener's *comprehension* task, i.e. her mapping from auditory information to

---

4 We use slashes ("/") for perceived phonological surface representations. We assume that these representations consist of the same kinds of discrete arbitrary symbols as lexical representations, because the task of the perception process is to turn raw auditory data into discrete representations that are maximally suited for lexical access. See (2) for an explicit model.

5 Deeper mechanisms than perceived similarity may play a role as well, such as choosing categories that are peripheral in the L1, in order to improve *production* in such a way that other listeners' comprehension improves. This may contribute to linking |a|$_{\mathrm{S}}$ to |ɑ|$_{\mathrm{D}}$ rather than to |ɛ|$_{\mathrm{D}}$. Such a bias towards peripherality also follows automatically (i.e. without goal orientation) from Escudero's (2005: 214–236) model of selecting underlying representations (cf. fn. 2).

lexical representations that make contact with meaning. In several theories of phonological comprehension (for an overview, see McQueen and Cutler 1997 and McQueen 2005), the process consists of two sequential levels, which can be called *perception* and *recognition*. The ("prelexical") perception process maps auditory to phonological surface representations without accessing the lexicon, and the recognition process maps the phonological surface representations to underlying forms in the lexicon and is heavily influenced by the semantic and pragmatic context.

(2) Two-stage comprehension model

	perception		recognition	
auditory representation	→	phonological representation	→	lexical representation
e. g. [kæso]		/kɛsɔ/$_D$		|kɑsɔ|$_D$ 'case'

The advantage of reusing lexical categories now becomes clear: the learner will exhibit some initial proficiency in her comprehension, at least if she transfers the perception system to her interlanguage system as well. Suppose, for instance, that the learner is in a stage at which she has already correctly stored the Spanish words |kaso|$_S$ 'case' and |keso|$_S$ 'cheese' into her interlanguage lexicon as |kɑsɔ|$_D$ 'case' and |kɛsɔ|$_D$ 'cheese'. A hundred native tokens of an intended |kaso|$_S$ will have a distribution of vowel formants (for the |a|$_S$ part) that is centred around values that are typical of a low front vowel. As suggested above, Dutch monolinguals may hear 60 of these vowel tokens as /ɑ/$_D$, 27 as /ɛ/$_D$. If learners transfer this perception to their interlanguage, they will perceive 60 instances of |kaso|$_S$ as /kɑsɔ/$_D$, 27 as /kɛsɔ/$_D$. In the majority of the cases, then, a beginning learner will perceive /kɑsɔ/$_D$, from which the lexical item |kɑsɔ|$_D$ 'case' can be retrieved quite easily. Thus, comprehension is well served by an initial transfer of native perception (which presupposes an initial transfer of native lexical symbols) to the interlanguage.

But an interlanguage perception system that is identical to the native perception system is not perfect yet. In the example above, 27 percent of intended |kaso|$_S$ tokens, perhaps the most fronted and raised ones, will be perceived as /kɛsɔ/$_D$, from which it is not so easy to retrieve the lexical item |kɑsɔ|$_D$ 'case'.[6] To improve, the learner will have to learn to perceive tokens in the auditory [æ] region as /ɑ/$_D$ rather than as /ɛ/$_D$ when listening to Spanish. Preferably, though,

6 In Optimality-Theoretic terms, having to map a perceived /kɛsɔ/ to an underlying |kɑsɔ| can be said to involve a faithfulness violation in the recognition grammar (Boersma 2001).

tokens in that same region of auditory space should continue to be perceived as /ɛ/$_D$ if the learner is listening to Dutch. The following table sums up the ways in which [æ] would then be perceived in the five cases we discussed:

(3) Five perceptions of the auditory form [æ]
Monolingual Spanish: [æ] → /a/$_S$
Monolingual Dutch: [æ] → /ɛ/$_D$
Beginning learners when listening to Spanish: [æ] → /ɛ/$_D$ (transfer)
Proficient learners when listening to Spanish : [æ] → /ɑ/$_D$ (native-like)
All learners when listening to Dutch: [æ] → /ɛ/$_D$ (double perception systems)

The situation in (3) would require a duplication of the learner's perception system, where the interlanguage perception system starts out as a clone of the native perception system but subsequently develops towards something more appropriate for the comprehension of the target language, without affecting the L1 perception system (Escudero & Boersma 2002). The experiment described below, in which we show that Dutch learners of Spanish exhibit different perceptual behaviour when they think they are listening to Dutch than when they think they are listening to Spanish, provides evidence for two separate perception systems in L2 learners.

1.3. The listening experiment: Method

The method (stimulus material, subjects, tasks) was described before in Escudero & Boersma (2002). We repeat here only what is relevant for the present paper.

Stimulus material. The same bilingual speaker as in Figure 1 read aloud a Spanish text, from which we cut 125 CVC (consonant-vowel-consonant) tokens. The consonants were selected in such a way that each of the 125 CVC tokens could pass for a licit Dutch syllable (apart from the vowel).

Subjects. Thirty-eight Dutch learners of Spanish performed the three tasks described below. The learners were from various parts of the Netherlands, so that their vowel systems may differ from the one in Figure 1 mainly in the location of /ɔ/ (which for many speakers has [ɔ]- and [ʊ]-like positional variants) and in the location of /oː/ (which for many speakers has the same degree of diphthongization, and the same height, as /eː/ and /øː/). There were two control groups: 11 Dutch non-learners of Spanish performed the first and second tasks only, and 44 native speakers of Spanish performed the third task only.

First task. In the first task the subjects were told that they were going to listen to a number of Dutch CVC syllables and had to classify the vowel into the Dutch classes /ɑ/, /aː/, /ɛ/, /ɪ/, /eː/, /i/, /ʏ/, /øː/, /y/, /ɔ/, /oː/, /u/. But what the subjects actually heard was a randomized set of the 125 Spanish tokens. To enhance the Dutch *perception mode*, the tokens were interspersed with 55 CVC tokens that were cut from a Dutch text spoken by the same bilingual speaker; many of these 55 tokens contained very Dutch-sounding vowels and consonants, often corresponding to a recognizable Dutch word, e. g. /ɦøːs/ 'really'. Also, the 180 CVC tokens were embedded within a Dutch carrier phrase (*luister naar…*).

Second task. The second task differed from the first only in the perception mode that we wanted to bring the subjects in. So we told the subjects (correctly, this time) that they were going to listen to Spanish CVC sequences, and we interspersed the 125 CVC tokens (which were the same as in the first task) with 55 very Spanish-sounding tokens (e. g. /roɾ/) and embedded the 180 stimuli within a Spanish carrier phrase (*la palabra*…). Importantly, though, we told the listeners to try to "listen with Dutch ears" to these stimuli and to classify the 180 tokens into the 12 Dutch vowel classes.

Third task. The third task differed from the second only in that we told the listeners to listen with Spanish ears and to classify the 180 tokens into the 5 Spanish vowel classes. This task, then, simply tested the learners' proficiency in the perception of the target language.

1.4. The listening experiment: Results

When the subjects thought that the language they were hearing was Dutch (Task 1), they responded differently from when they thought the language was Spanish (Task 2): they turned out not to be able to completely "listen with Dutch ears" in Task 2. For details, see Escudero & Boersma (2002, to appear). We now describe the three main differences between the results of the two tasks. In Task 2, the group of 38 listeners avoided responding with "ɪ". Although most tokens that were scored as "ɪ" in the first task were still scored as "ɪ" in the second (namely 599), many tokens that were scored as "ɪ" in the first task were scored as "i" or "ɛ" in the second (namely, 120 and 101, respectively). The reverse drift was much smaller: the number of tokens that were scored as "i" or "ɛ" in the first task but as "ɪ" in the second were only 27 and 57, respectively. Since the differences between 120 and 27 and between 101 and 57 are significantly greater than zero (see Escudero & Boersma to appear for the statistical tests), we can reliably say that the listener group shied away from the "ɪ" response in the second task. The learners showed an analogous behaviour for "ʏ" responses, which were avoided in the second task, where many of them were replaced with "u" and "ɔ" responses. A third reliable effect was the shift of the "ɑ" response: many tokens that were scored as "ɛ" when the listeners were fooled into thinking the language was Dutch were scored as "ɑ" when the listeners knew it was Spanish, and many tokens that were scored as "ɑ" in the first task were scored as "ɔ" in the second. Finally, the long vowels "aː", "eː", "oː" and "øː" were generally avoided in the responses in Task 2.

The learners showed developmental effects. The degree of "ɪ" avoidance in Task 2 relative to Task 1 correlated with the experience level of the learners (who were divided into 11 beginners, 18 intermediate, and 9 advanced on the basis of an independent language background questionnaire) as well as with the perceptual proficiency level as measured in Task 3 (Escudero & Boersma 2002).

1.5. The listening experiment: Interpretation

The shift from "ɛ" responses in the first task toward "ɑ" responses in the second shows that the learners reused their Dutch $/ɑ/_D$ category for perceiving Spanish $/a/_S$. We can explain this shift by assuming that for [æ]-like auditory forms some of the learners follow the mode-dependent strategies predicted in (3) for proficient learners:

(4) Two separate language modes for a proficient Dutch learner of Spanish

Language mode	Token	Perception	Response
Dutch	[æ]	$/ɛ/_D$	"ɛ"
Spanish	[æ]	$/ɑ/_D$	"ɑ"

For the Spanish vowel $/i/_S$, which could in principle have been identified with Dutch $/ɪ/_D$ or with Dutch $/i/_D$, the avoidance of "ɪ" in the second task shows that in fact Spanish $/i/_S$ was identified with Dutch $/i/_D$. This shows that (1) is correct. The avoidance of the four long vowels in both the first and second tasks confirms the expectation mentioned in § 1.2 that duration is a strong auditory cue that can override any spectral similarity.

The developmental effects can be explained by an initial *transfer* of the native perception system to the interlanguage, followed by *lexicon-guided learning*. Thus, the Dutch-appropriate perception of [æ] as $/ɛ/_D$ is transferred to the initial state of the learner's interlanguage, so that a beginning Dutch learner of Spanish will perceive [æ] as $/ɛ/_D$, regardless of whether she listens to Dutch or to Spanish. When she is listening to Spanish, however, the lexicon will often issue an error message. If the learner perceives an incoming [kæso] as /kɛsɔ/$_D$, for instance, higher conceptual processing may force the lexicon to recognize /kɛsɔ/$_D$ as |kɑsɔ|$_D$ 'case'. If that happens, the lexicon can 'tell' the perception system to modify itself in such a way that a /kɑsɔ/$_D$ perception becomes more likely in the future (note that the existence of minimal pairs is not required). Both the perception system and lexicon-guided learning are formally modelled in the following sections.

2. An explicit phonological model of perception

Perception researchers agree that prelexical perception, i. e. the mapping from auditory to phonological representations, is a language-dependent process for all speakers from about 9 months of age (Werker and Tees 1984; Jusczyk, Cutler, and Redantz 1993; Polka and Werker 1994). This language dependence is enough reason for us as linguists to want to model prelexical perception by linguistic means, e. g. to model it by Optimality-Theoretic constraint ranking, as has been done before by Boersma (1997, 1998, 1999, 2000), Hayes (2001), Escudero and Boersma (2003, 2004), and Pater (2004).[7] Tesar's (1997, 1998) and Tesar & Smolensky's (2000) Optimality-Theoretic modelling of the process of *robust interpretive parsing*, i. e. a mapping from unanalysed ("overt") sequences of syllables with stress marks to full abstract hierarchical foot structures, can also be seen as a case of Optimality-Theoretic modelling of perception, an idea that was pursued by Apoussidou & Boersma (2003, 2004).[8]

7 Not included in this list are those who model comprehension as a single mapping in Optimality Theory, namely Smolensky (1996), Kenstowicz (2001), Broselow (2003), and Yip (2006), nor developments more recent than the present paper, such as Boersma (2007) and Escudero (2005).

8 We have to point out that Smolensky (p. c.) does not consider perception and robust interpretive parsing to be the same, because our auditory form is more peripheral and continuous than Tesar & Smolensky's overt form, which has already been analysed into discrete syllables. However, we see no reason why the language-specific construction of feet should not be handled in parallel with more peripheral-looking processes like the language-specific mapping from vowel duration to e. g. stress in Italian or to vowel length in Czech. Until there is evidence for prelexical sequential modularity, we will subsume all these processes under the single umbrella of "perception". The literature on the perception of foot structure by infants (e. g. Jusczyk, Houston, and Newsome 1999; Polka, Sundara, and Blue 2002; Curtin, Mintz, and Christiansen 2005) usually talks about "word segmentation", but uses perceptual terminology like "cue weighting".

In our special case of L2 acquisition, perception can depend on the language that learners think they are listening to: the likelihood of mapping [æ] to the Dutch lexical vowel symbol $/\text{ɛ}/_{\text{D}}$ depends on whether the learner thinks she is hearing Dutch (more likely) or Spanish (less likely), as we mentioned in § 1.4. We therefore model the behaviour of the learner with two separate perception grammars, one for her Dutch perception, which does not change during her learning of Spanish, and one for her Spanish perception, which starts out as a clone of her Dutch perception grammar and subsequently develops towards a more Spanish-appropriate grammar by the lexicon-driven optimization we introduced in § 1.5.

2.1. Tableaus and constraints that model perception

Optimality-Theoretic perception grammars use the same decision scheme as the more usual Optimality-Theoretic production grammars. Whereas a production grammar takes an underlying lexical representation as its input and yields a pronunciation or surface structure as its output (Prince and Smolensky 1993, McCarthy and Prince 1995), a perception grammar takes an auditory representation as its input and yields a phonological surface structure as its output.

The perceptual process that we restrict ourselves to in this paper is static categorization, where the inputs are static (temporally constant) values of auditory features and the output candidates are language-specific phonological features or phonemes. Escudero & Boersma (2003) proposed that this mapping is evaluated by the negatively formulated constraint template in (5), which directly relates auditory feature values to phonological categories. The reason for its negative formulation will be discussed in § 4.5.

(5) Arbitrary cue constraints
“A value *x* on the auditory continuum *f* should not be mapped to the phonological category *y*.”

For our case, the perception of Dutch and Spanish vowels, the relevant auditory continua are the first formant (F1), the second formant (F2), and duration, and the relevant phonological categories are the 12 Dutch vowel symbols. Examples of the relevant *cue constraints* (the term is by Boersma 2007 and Escudero 2005) are therefore “an F1 of 531 Hz is not $/\text{ɔ}/_{\text{D}}$”, or “an F2 of 1585 Hz is not $/\text{eː}/_{\text{D}}$”, or “a duration of 150 ms is not $/\text{y}/_{\text{D}}$”. We propose that these cue constraints are *arbitrary*, i. e. they exist for any auditory value and any vowel category, regardless of whether that auditory value is a plausible cue for that vowel category. Thus while a typical F1 value for /i/$_D$ is 280 Hz, we indiscriminately allow the presence of constraints like "an F1 of 280 Hz is not /i/$_D$" and "an F1 of 900 Hz is not /i/$_D$". It is the *ranking* of these constraints, not their presence, that determines what auditory values map to what vowel categories. Thus, in order to make it unlikely that an auditory input with an F1 of 900 Hz will ever be perceived as /i/$_D$, the constraint "an F1 of 900 Hz is not /i/$_D$" should be ranked very high, and in order to allow that [i]-like auditory events can be perceived as /i/$_D$ at all, the constraint "an F1 of 280 Hz is not /i/$_D$" should be ranked rather low.

As an example, consider the perception of the typical token of the Spanish vowel |a|$_S$, namely an [a]-like auditory event with an F1 of 877 Hz, an F2 of 1881 Hz, and a duration of 70 ms. In tableau (6) we see that the two spectral cues favour the perception of |aː|$_D$, but that in line with the finding in § 1.5 these cues are overridden by the duration constraints, which assert that an overtly short vowel token (e. g. 70 ms long) should not be perceived as the vowel /aː/$_D$.

(6) Dutch cross-language perception of a typical token of Spanish |a|$_S$

[a], i. e. [F1=877, F2=1881, dur=70]	[dur=70] is not /aː/$_D$	[F1=877] is not /ɛ/$_D$	[F2=1881] is not /ɑ/$_D$	[F1=877] is not /aː/$_D$	[F1=877] is not /ɑ/$_D$	[F2=1881] is not /ɛ/$_D$, not /aː/$_D$	[dur=70] is not /ɑ/$_D$, not /ɛ/$_D$
/aː/$_D$	*!			*		*	
☞ /ɑ/$_D$			*		*		*
/ɛ/$_D$		*!				*	*

With (6) we can describe the behaviour of the non-Spanish-learning Dutch listeners in the experiment. There are two reasons why the listeners' responses are variable. First, the 25 |a|$_S$ tokens in the experiment were all different, so that some will have been closer to [ɛ], some to [ɑ]. Secondly, listeners are expected to show variable behaviour even for repeated responses to the same token. We model this by using *Stochastic Optimality Theory* (Boersma 1997, 1998; Boersma and Hayes 2001), in which constraints have *ranking values* along a continuous scale and in which some *evaluation noise* is temporarily added to the ranking of a constraint at each evaluation. In tableau (6) this will mean that candidate /ɑ/$_D$ will win most of the time, followed by candidate /ɛ/$_D$.

In general, the candidates in a tableau should be all 12 vowels. Since that would require including all 36 relevant cue constraints, we simplified tableau

(6) to include only three candidates, so that we need only consider 9 constraints. The remaining nine candidate vowels can be ruled out by constraints such as "an F1 of 877 Hz is not /i/$_D$" and "an F2 of 1881 Hz is not /ɪ/$_D$", which are probably ranked far above "a duration of 70 ms is not /aː/$_D$", since there were no "i" or "ɪ" responses at all for intended |a|$_S$. Tableau (6) also abstracts away from constraints such as "an F1 of 280 Hz is not /ɔ/$_D$" that refer to auditory feature values that do not occur in the input of this tableau. Such constraints do exist and are ranked along the same continuum as the nine constraints in (6); the constraint "an F1 of 280 Hz is not /ɔ/$_D$" can interact with six of the nine constraints in (6), namely when the input contains a combination of an F1 of 280 Hz with either an F2 of 1881 Hz or a duration of 70 ms.

Since the four long Dutch vowels play no role in the identifications in (1) or in the perception experiment reported in § 1.4, we will from now on ignore these long vowels and consider only the eight short vowels as possible candidates. This allows us to ignore the duration constraints and to focus on the spectral cues alone.

2.2. Lexicon-driven perceptual learning in Optimality Theory

A tableau is just a *description* of how perception can be modelled in Optimality Theory. A more *explanatory* account involves showing how the ranking of so many constraints can be learned. This section describes Boersma's (1997, 1998) proposal for lexicon-driven optimization of an Optimality-Theoretic perception grammar, as it was first applied to the ranking of arbitrary cue constraints in L1 and L2 acquisition by Escudero & Boersma (2003, 2004).

Throughout our modelling of perception we assume that the learner has already established correct representations in her lexicon. This means that the listener's recognition system (see (2)) can often reconstruct the speaker's intended vowel category, even if the original perception was incorrect. After all, the listener's recognition system will only come up with candidate underlying forms that are actually in the lexicon, and in cases of ambiguity will also be helped by the semantic context (see Boersma 2001 and Escudero 2005: 214–236 for Optimality-Theoretic solutions). If the resulting underlying form differs from the perceived surface form, the recognition system can signal to the perception system that the perception has been "incorrect". We will denote such situations by marking the speaker's intention (as recognized by the listener) in the listener's perception tableau with a check mark, as in (7).

(7) A beginning learner's misperception of a high front token of Spanish |a|$_S$

[F1=800, F2=1900]	[F1=800] is not /ɪ/$_D$	[F2=1900] is not /ɔ/$_D$	[F1=800] is not /ɔ/$_D$	[F2=1900] is not /ɑ/$_D$	[F1=800] is not /ɛ/$_D$	[F1=800] is not /ɑ/$_D$	[F2=1900] is not /ɛ/$_D$
√ /ɑ/$_D$				*!→		*→	
☞ /ɛ/$_D$					←*		←*
/ɔ/$_D$		*!	*				
/ɪ/$_D$	*!						

We can assume that the constraint "an F1 of 800 Hz is not /ɛ/$_D$" in (7) is ranked lower than the constraint "an F1 of 877 Hz is not /ɛ/$_D$" in (6), because 800 Hz is closer to typical F1 values of |ɛ|$_D$ than 877 Hz is. By this lower ranking, the constraint "an F1 of 800 Hz is not /ɛ/$_D$" can be ranked below "an F2 of 1900 Hz is not /ɑ/$_D$", which is of course ranked at nearly the same height as "an F2 of 1881 Hz is not /ɑ/$_D$" in (6). This difference between (6) and (7) now makes /ɛ/$_D$ the winner. However, if the learner's postperceptual recognition tells her she should have perceived /ɑ/$_D$ because the recognized lexeme contains the vowel |ɑ|$_D$, she can mark this candidate in the tableau ("√"), and when she notices that this form is different from her winning candidate /ɛ/$_D$, she can take action by changing her perception system. The changes are depicted in the tableau by arrows: the learner will raise the ranking of the two constraints that prefer the form she considers correct ("←") and lower the ranking of the two constraints that prefer her incorrectly winning candidate ("→"), thus making it more probable that auditory events with an F1 of 800 Hz or an F2 of 1900 Hz will be perceived as /ɑ/$_D$ at future occasions, at least when she is listening to Spanish.

In order to prove that the learning algorithm just described works for Dutch learners of Spanish throughout their L1 and L2 acquisition, we will show two computer simulations. Section 3 will simulate a simplified problem, namely the L1 and L2 acquisition of the mapping from a single auditory continuum (F1) to four vowel heights (exemplified by /ɑ/$_D$, /ɛ/$_D$, /ɪ/$_D$, and /i/$_D$). Section 4 will fully simulate the L1 and L2 acquisition of the mapping from two auditory continua (F1 and F2) to the 12 Dutch vowels and, later, the 5 Spanish vowels of Figure 1.

3. One-dimensional vowel loss

We will first simulate the acquisition of a simplified vowel system, one in which a single auditory continuum, namely F1, is mapped to only four vowels. This initial simplification is necessary in order for us to be able to illustrate with explicit graphics how constraint rankings in the perception grammar can lead to an optimal perception in L1 and L2. The two-dimensional case of § 4 will then be a straightforward extension.

3.1. The L1 language environment

The L1 at hand is a language with only four vowels, simplified Dutch. The vowels carry the familiar labels $/\text{ɑ}/_D$, $/\text{ɛ}/_D$, $/\text{ɪ}/_D$, and $/\text{i}/_D$, but they are distinguished only by their F1 values. We assume that the token distributions of the four intended vowels $|\text{ɑ}|_D$, $|\text{ɛ}|_D$, $|\text{ɪ}|_D$, and $|\text{i}|_D$ have Gaussian shapes around their mean values along a logarithmic F1 axis, as in Figure 2. The mean values (i. e. the locations of the peaks in Figure 2) are the same as the median F1 values of Figure 1, namely 926, 733, 438, and 305 Hz, and the standard deviation is 0.05 along a base-10 logarithmic scale (i. e. 0.166 octaves). This leads to the curves in Figure 2, where we assume for simplicity that all four vowels occur equally frequently, so that the four peaks are equally high.

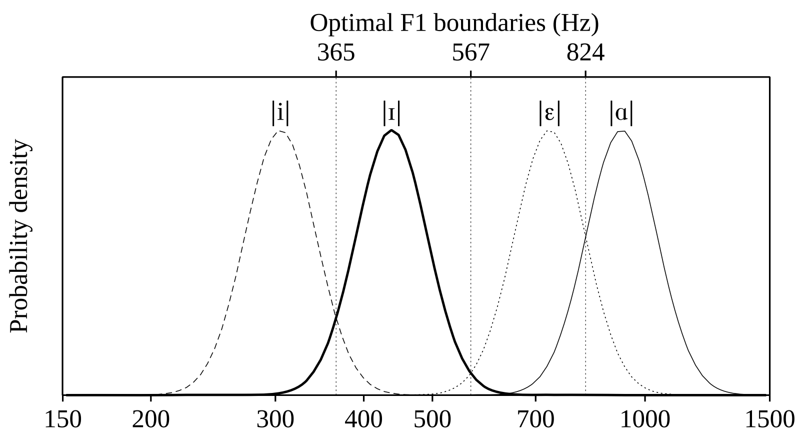


Figure 2. Idealized token distributions for four short Dutch vowels.

3.2. Optimal L1 perception

Figure 2, then, describes the distributions of *speakers'* productions of the four intended vowels in a large corpus of one-dimensional Dutch. The task of the *listeners* is to map each incoming F1 value onto one of the vowel categories $/\text{ɑ}/_D$, $/\text{ɛ}/_D$, $/\text{ɪ}/_D$, and $/\text{i}/_D$, in preparation for subsequent access of a word containing one of the underlying vowels $|\text{ɑ}|_\text{D}$, $|\text{ɛ}|_\text{D}$, $|\text{ɪ}|_\text{D}$, and $|\text{i}|_\text{D}$. The question now is: what would be an optimal strategy for a listener? We propose that the optimal strategy is to minimize the discrepancy between the perceived vowel and the recognized vowel, i. e. to minimize the number of cases where the listener perceives a certain vowel (e. g. $/\text{ɛ}/_\text{D}$) but subsequently finds a different vowel (e. g. $|\text{ɪ}|_\text{D}$) in her lexicon (we call such a situation a *perception error*).

A general strategy that achieves this minimization of the number of perception errors is the *maximum likelihood* strategy (Helmholtz 1910), where the listener perceives any given F1 value as the vowel that was most likely to have been intended by the speaker. In Figure 2 we see that if a listener hears an F1 value of 400 Hz, it is most likely that this was a token of an intended vowel $|\text{ɪ}|_\text{D}$. We know this because for an F1 of 400 Hz the distribution curve for $|\text{ɪ}|_\text{D}$ lies above the distribution curves for the other three vowels. In general, any F1 value should be perceived as the vowel whose curve is highest. Which curve is highest in Figure 2 is determined by the three main cutting points of the curves, which lie at 365, 567, and 824 Hz. Given the distributions in Figure 2, then, a maximum-likelihood strategy entails that the listener should perceive all incoming F1 values below 365 Hz as $/\text{i}/_\text{D}$, all F1 values between 365 and 567 Hz as $/\text{ɪ}/_\text{D}$, all F1 values between 567 and 824 Hz as $/\text{ɛ}/_\text{D}$, and all F1 values above 824 Hz as $/\text{ɑ}/_\text{D}$. If the listener indeed uses these three *optimal boundaries* as her criteria for perception, she will achieve a correctness percentage of 90.5. That is, of all F1 values that will be drawn according to the distributions of Figure 2 (with equal probabilities for each of the four intended vowels) she will perceive 90.5 percent as the same vowel as she will subsequently find in her lexicon. The remaining 9.5 percent are cases of perception errors, caused by the overlap in the curves of Figure 2 (i. e. in 9.5 percent of the productions an F1 value crosses the boundary with a neighbouring vowel).

The reader will have noticed that our definition of optimal perception (minimizing the number of perception errors) is related to our operationalization of lexicon-driven learning (§ 2.2), which changes the perception grammar every time the listener makes a perception error. The simulation of the following section will show that lexicon-driven perceptual learning with the GLA indeed leads to optimal boundaries in the listener.

3.3. L1 acquisition of the perception of one-dimensional Dutch

In order to be able to do a computer simulation of the F1-only simplified Dutch vowel system, we divide up the F1 continuum between 150 and 1500 Hz into 100 values equally spaced along a logarithmic scale: 152, 155, 159, …, 1416, 1449, and 1483 Hz. We will assume that only these 100 frequencies are possible incoming F1 values. According to § 2.1, we therefore need 400 cue constraints (100 F1 values × 4 vowel categories) that can be formulated like "[F1 = 1416 Hz] is not /ɪ/$_{\mathrm{D}}$".[9]

We assume that in the initial state of our learner all lexical representations are already correct, so that lexicon-driven learning according to tableaus like (7) works flawlessly. We further assume that all 400 cue constraints are initially ranked at the same height, namely at 100.0, so that any F1 value has a probability of 25 percent of being perceived as any of the four vowels. This combination of assumptions is obviously a severe simplification, since a correct lexicalization must depend on a reasonably good perception system, i. e. one whose percentage correct is much higher than 25. Such a reasonably good perception system could be obtained by an Optimality-Theoretic distributional learning method for infants such as the one described by Boersma, Escudero & Hayes (2003), but we will not pursue this here since we are mainly interested in what happens later in life.

We feed our simulated learner with 10,000 F1 values per virtual year, drawn from the distributions in Figure 2 (i. e. more F1 values near the peaks than near the valleys), always telling the learner, as in (7), what would have been the correct perception. Every time there is a mismatch between the perceived vowel and the correct vowel (i. e. the vowel intended by the speaker, as recognized by the listener's lexicon), some rankings change by a small amount, which Stochastic Optimality Theory refers to as the *plasticity* (or *learning step*). The plasticity is 1.0 during the first year, then decreases by a factor of 0.7 every year, ending up as a plasticity of 0.0023 during the 18th virtual year. With a constant evaluation noise of 2.0, this plasticity scheme causes learning to be initially fast but imprecise, and later on slow but accurate.

The left side of Figure 3 shows the development of the grammars and is to be interpreted as follows. For every F1 value it is the lowest-ranked constraint that determines into which vowel category the F1 value will most often be classified. For instance, for an F1 of 400 Hz the lowest ranked constraint (the thick curve) is "[F1 = 400 Hz] is not /ɪ/$_{\mathrm{D}}$". Tableau (8) shows that the low ranking of this constraint determines the winning candidate, irrespective of the relative ranking of the other three relevant constraints.

---

9 A more sophisticated discretization of the F1 continuum, as used by Boersma (1997), would involve taking many more F1 values and allowing the learning algorithm to change the ranking of some neighbouring constraints by a value that decreases with the distance to the incoming F1. This would lead to results similar to those obtained by the simplified discretization of the present paper.

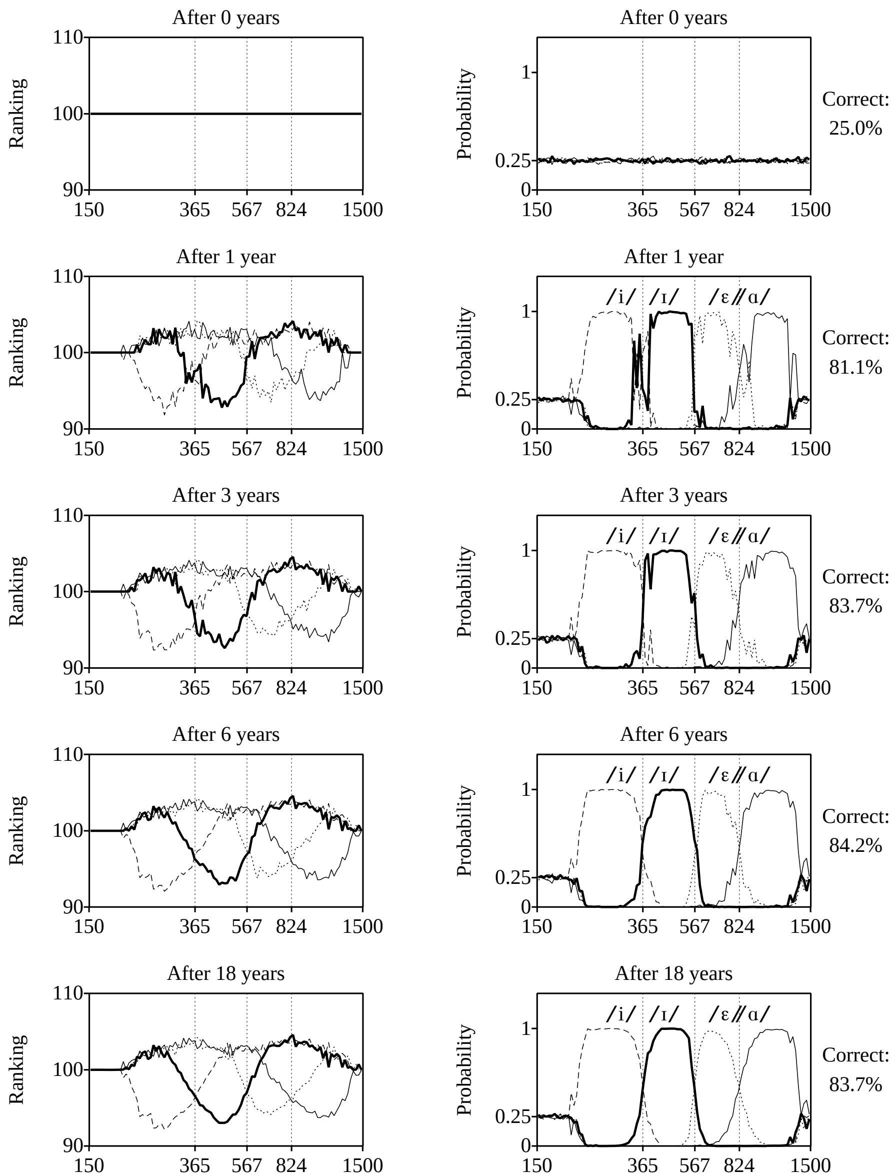


Figure 3. Simulated L1 acquisition of Dutch.
Left: the rankings of the four constraint families "[F1=*x*] is not /*vowel*/$_D$".
Right: the identification curves.
Dashed: /i/$_D$; plain thick: /ɪ/$_D$; dotted: /ɛ/$_D$; plain thin: /ɑ/$_D$.

(8) Perception determined by the lowest curve

[F1=400]	[F1=400] is not /ɑ/$_{\mathrm{D}}$	[F1=400] is not /ɛ/$_{\mathrm{D}}$	[F1=400] is not /i/$_{\mathrm{D}}$	[F1=400] is not /ɪ/$_{\mathrm{D}}$
/ɑ/$_{\mathrm{D}}$	*!			
/ɛ/$_{\mathrm{D}}$		*!		
☞ /ɪ/$_{\mathrm{D}}$				*
/i/$_{\mathrm{D}}$			*!	

Every grammar leads to its own perception pattern. In the course of the 18 virtual years we see that the crossing points of the constraint curves come to lie close to the optimal boundaries of 365, 567, and 824 Hz. If a listener with the 18th-year grammar in Figure 3 were to have an evaluation noise of zero, her percentage correct would be about 90.5, just as for the maximum-likelihood listener in § 3.2 (the percentage correct can be estimated by running 100,000 F1 values, distributed as in Figure 2, through the grammar and counting the number of correct output vowels). If we assume, however, that the listener has an evaluation noise of 2.0, just as during learning, the percentage correct is a bit lower. It can be shown (Boersma 1997) that in the one-dimensional case the resulting perception grammar is *probability matching*, i. e. the probability of perceiving a certain F1 value as a certain vowel comes to approximate the probability that this F1 value had been intended as that vowel. For instance, we can read off Figure 2 that an F1 value of 400 Hz has 90 percent chance of having been intended as |ɪ|$_{\mathrm{D}}$ and 10 percent chance of having been intended as |i|$_{\mathrm{D}}$. When confronted with an auditory input of 400 Hz, a probability-matching listener will perceive it 90 percent of the time as /ɪ/$_{\mathrm{D}}$ and 10 percent of the time as /i/$_{\mathrm{D}}$. This is exactly what our learner comes to do, improving her perception of the whole distribution from 25 percent correct to 83.7 percent correct, which is the same value that can be computed from Figure 2.[10] In the rest of this paper we will call probability-matching behaviour "optimal", and forget about maximum-likelihood behaviour, which never occurs in practice anyway.

The right side of Figure 3 shows our virtual listener's *identification curves* (as known from many perception experiments with real listeners), i. e. for each

10 Given a distribution where $p(f, v)$ denotes the probability that a token drawn randomly from the language environment has an F1 of f Hz and was intended as the vowel v (i. e. $\Sigma_{f,v}\, p(f, v) = 1$), the fraction correct for a maximum-likelihood listener can be computed as $\Sigma_f \max_v p(f, v)$, and the fraction correct for a probability-matching listener can be computed as $\Sigma_f\, (\Sigma_v\, p(f, v)^2 / \Sigma_v\, p(f, v))$.

of the four vowels a curve that shows for every F1 value how often that F1 value is perceived as that vowel. These curves are computed by running each of the 100 F1 values through the grammar 1,000 times and counting how often each of the four possible vowels is the winner. The virtual learner grows increasingly confident of her category boundaries, which become optimal for her language environment.

3.4. L2 acquisition of the perception of one-dimensional Spanish

After having learned Dutch for 18 years, our virtual learner starts learning Spanish. Our one-dimensional Spanish has the three vowels $|a|_S$, $|e|_S$, and $|i|_S$, whose F1 distributions are centred around the median F1 of Figure 1, again with a logarithmic standard deviation of 0.05. The learner equates the three Spanish vowels with her Dutch categories $|ɑ|_D$, $|ɛ|_D$, and $|i|_D$, respectively, as do the real learners of § 1.4. Her L2 language environment can thus be described by the curves in Figure 4.

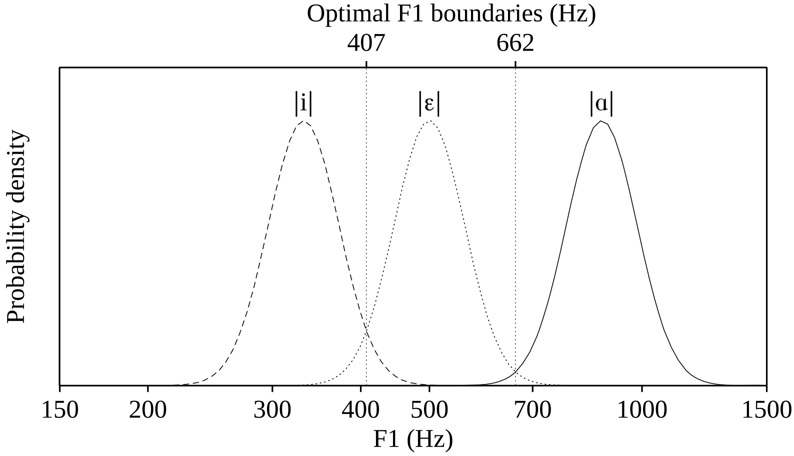


Figure 4. The Spanish vowel environment, with Dutch labels.

The learner's initial interlanguage grammar has to be a copy of her current grammar of Dutch (§ 1.2), so the picture in the upper left of Figure 5 is identical to the picture in the lower left of Figure 3. Such a grammar handles Spanish better than an infant-like grammar where all constraints are ranked at the same height. Whereas an infant-like grammar (with the four Dutch categories) would score 25 percent correct, the copied Dutch grammar already scores 53.1

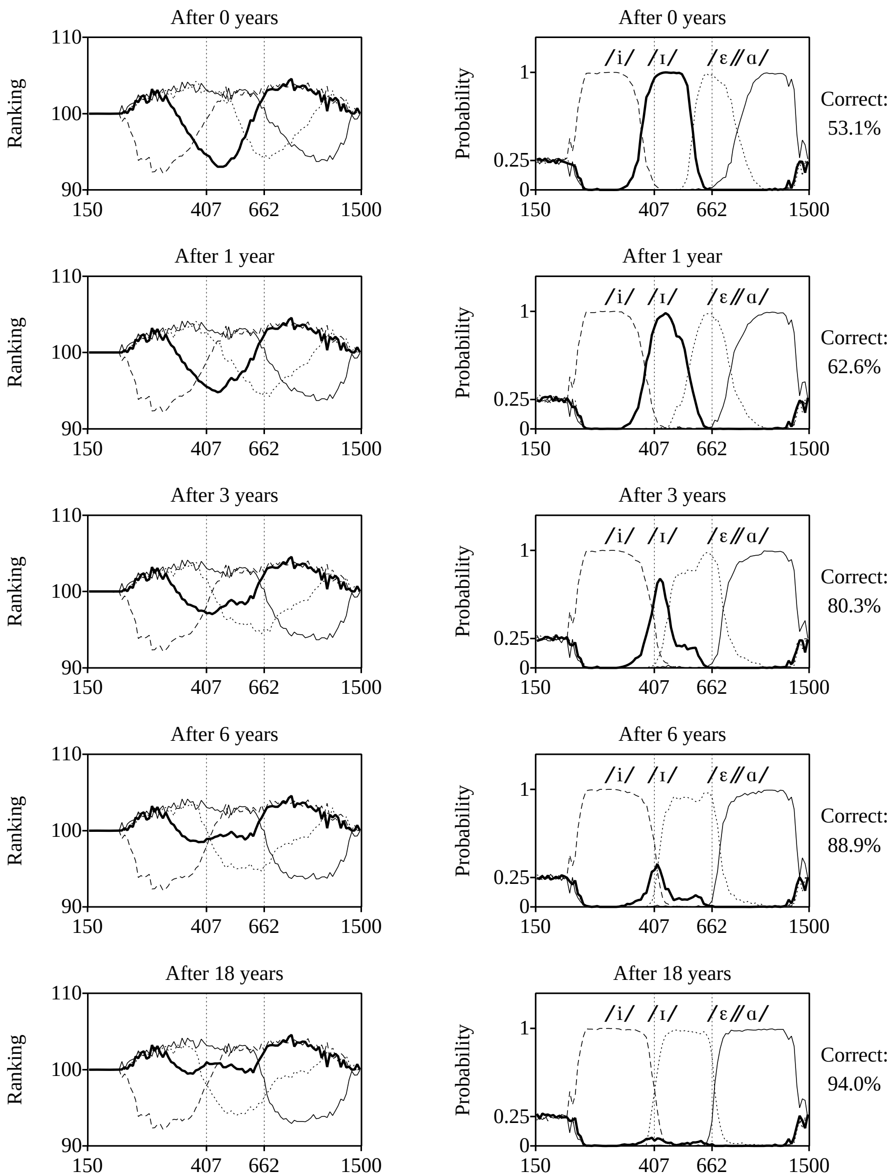


Figure 5. Simulated L2 acquisition of Spanish.
Left: the rankings of the four constraint families “[F1=*x*] is not /*vowel*/$_{\mathrm{D}}$”.
Right: the identification curves.
Dashed: /i/$_{\mathrm{D}}$; plain thick: /ɪ/$_{\mathrm{D}}$; dotted: /ɛ/$_{\mathrm{D}}$; plain thin: /ɑ/$_{\mathrm{D}}$.

percent correct. Nevertheless, this score is far from nativelike, since an adult probability-matching listener of Spanish will achieve 95.5 percent correct (as computed from Figure 4). If she is to gain more accuracy in her L2 environment, our virtual listener will have to learn.

We immerse our virtual learner in a rich Spanish environment where she hears 10,000 vowel tokens a year, as many as during her L1 acquisition. Acknowledging her high motivation, we endow her with a plasticity of 0.01, which is over four times as high as her final L1 plasticity of 0.0023 but of course still only a tiny fraction of her initial L1 plasticity of 1. The development of the virtual L2 learner is shown in Figure 5.

The main feature of the development is the fall of the /ɪ/$_{\mathrm{D}}$ category. Whenever the learner perceives an incoming F1 value as /ɪ/$_{\mathrm{D}}$, the interlanguage lexicon, which does not contain any instances of |ɪ|$_{\mathrm{D}}$, will tell her that she should have perceived a different vowel, most often /i/$_{\mathrm{D}}$ or /e/$_{\mathrm{D}}$. In all these cases, one of the constraints "[F1=*x*] is not /ɪ/$_{\mathrm{D}}$" will rise along the ranking scale, thus making it less likely that the next occcurrence of the same F1 value will again be perceived as /ɪ/$_{\mathrm{D}}$.

The learner's proficiency clearly improves, although despite her complete immersion in her L2 environment, despite her raised motivation, and despite her full access to an L1-like learning mechanism (the GLA), she has trouble achieving complete nativelike competence (i. e. 95.5 %), even in 18 years. This small failure is mainly due to the plasticity of 0.01, which stresses adultlike precision rather than infantlike learning speed.

4. Two-dimensional vowel loss and shift of |ɑ|$_{\mathrm{D}}$

After the oversimplification of § 3, our second simulation reflects a more realistic situation, in which two auditory cues, namely both F1 ('height') and F2 ('place'), contribute to the perception of the whole Dutch system of short vowels. We divide both continua into 21 values, as shown in Figure 6. Some height-place combinations cannot occur articulatorily (frog-like sounds in the bottom left) or by definition (the bottom right, where F1 is greater than F2); these are left blank in the figure.

4.1. The 2-dimensional L1 language environment

Figure 6 summarizes the height and place distributions for native speakers of Dutch. The circles represent the centres of the token distributions of the eight vowels. Their locations are similar to those in Figure 1, but for the purposes of the present section we have made each of them coincide exactly with one of the 21×21 possible height-place values. We assume that the standard deviation of the Gaussian place distribution is 2.0 columns along the horizontal axis, and that the standard deviation of the Gaussian height distribution is 2.0 rows along the vertical axis. We also simplifyingly assume that all short vowels are equally common, except $|\text{y}|_\text{D}$, which we take to be five times less common in this simplified Dutch inventory than every other short vowel. Figure 6 then shows for each F1-F2 combination what the most likely intended vowel is. The regions thus attributed to each vowel are delimited by dotted lines in the figure. These "production boundaries" turn out to run at equal distances to the nearest vowels, except for the boundaries around the $|\text{y}|_\text{D}$ area, which reflect the low token frequency of this vowel.

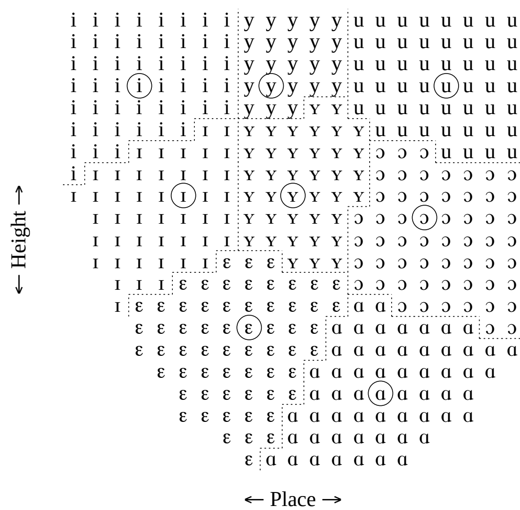


Figure 6. Circles: the centres of the token distributions of the eight short Dutch vowels. Phonetic symbols: the most likely intended vowel for every place-height combination.

4.2. Optimal 2-dimensional perception

Since Figure 6 shows the most likely intended productions, the production boundaries in this figure must indicate the optimal boundaries for perception as well. We can compute that a probability-matching listener would score 78.2 % correct. The following section shows that GLA learners can achieve this optimal perception.

4.3. L1 acquisition of the perception of 2-dimensional Dutch

Analogously to § 3.3, we feed a virtual Dutch listener 10,000 F1-F2 tokens a year, drawn randomly from the distribution in Figure 6 (i. e. fewer tokens far away from the vowel centres than close to them, and fewer tokens of $|\text{y}|_{\text{D}}$ than of every other vowel).

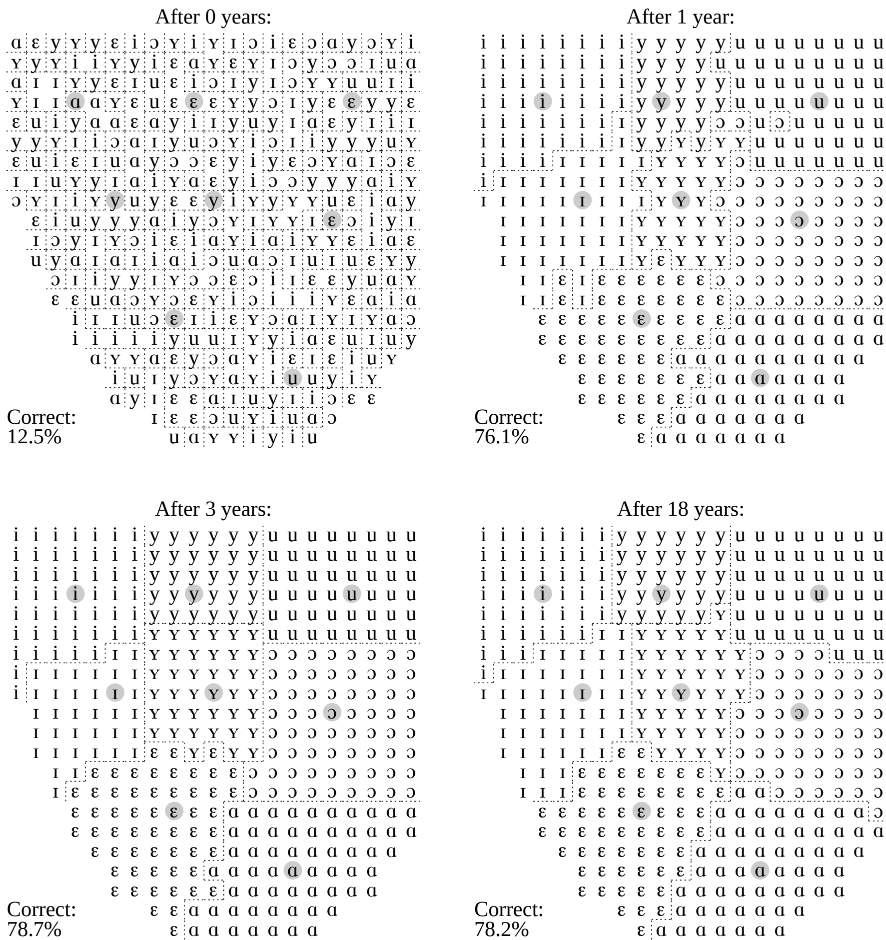


Figure 7. Simulated L1 Dutch vowel classification after 0, 1, 3, and 18 years. Grey disks: the eight Dutch short vowel centres in production.

The virtual learner's grammar contains 336 cue constraints (= (21 height values + 21 place values) × 8 vowels), which start out being ranked at the same height. Subsequent learning is performed, as before, via 180,000 tableaus, which in case of a misperception cause a learning step analogous to that in tableau (7). The evaluation noise and plasticity regime are as in § 3.3. There is no simple way to show the grammars or identification curves, as there was in the 1-dimensional case of § 3.3, but we can compute for every F1-F2 combination what the most likely perceived vowel is, by running each F1-F2 combination through the grammar 1000 times. The results are in Figure 7, which shows the development of the learner's performance. While after one year the "perception boundaries" (the dotted lines that delimit the most-likely-vowel areas) are still rather ragged, after 18 years they are smooth and very close to the production boundaries of Figure 6, leading to fractions correct that compare very well with the optimum reported in § 4.2. It turns out that the GLA is indeed capable of creating a stochastic Optimality-Theoretic grammar that exhibits optimal perceptual behaviour.

4.4. L2 acquisition of the perception of 2-dimensional Spanish

When the learner moves to Spain, her language environment becomes that of Figure 8, which shows the most likely intended Spanish vowels, under the assumption that the five vowels have equal token frequencies. When the learner copies her Dutch constraint ranking (i. e. the grammar in Figure 7, bottom right) to her Spanish interlanguage grammar, her fraction correct, given the distributions in Figure 8, is 47.6 % (cf. 56.6 % for the 1-dimensional case of § 3.4).

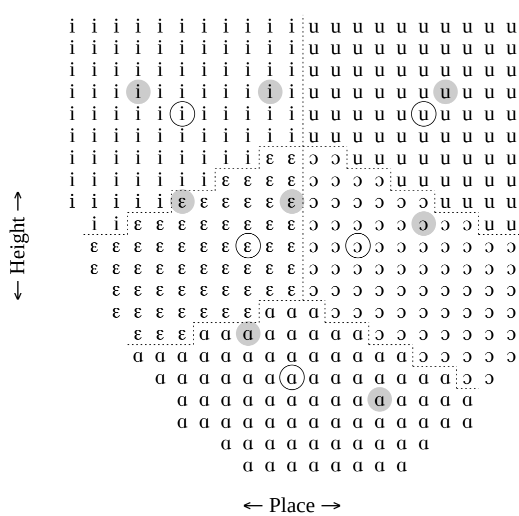


Figure 8. The Spanish vowel environment, with Dutch labels. Circles: the Spanish vowel centers. Grey disks: Dutch short vowel centres.

As with the 1-dimensional case of § 3.4, we immerse the learner in Spanish (10,000 tokens a year, drawn from the distributions in Figure 8, with lexicon-guided correction) with a plasticity of 0.01. The development of classification behaviour is shown in Figure 9. We see that the learner gradually loses her /ɪ/$_{\mathrm{D}}$, /ʏ/$_{\mathrm{D}}$, and /y/$_{\mathrm{D}}$ categories and shifts her /ɑ/$_{\mathrm{D}}$ category towards the front, just as the real human subjects did in our listening experiment (/ɪ/$_{\mathrm{D}}$ and /y/$_{\mathrm{D}}$ never fade entirely, continuing to occupy regions where the Spanish learning environment has offered very few tokens). Nativelike behaviour, which should follow the optimal boundaries in Figure 8 (and reach a fraction correct of 83.7 %), is closely approached but never completely attained, mainly as a result of the low plasticity relative to that of infants.

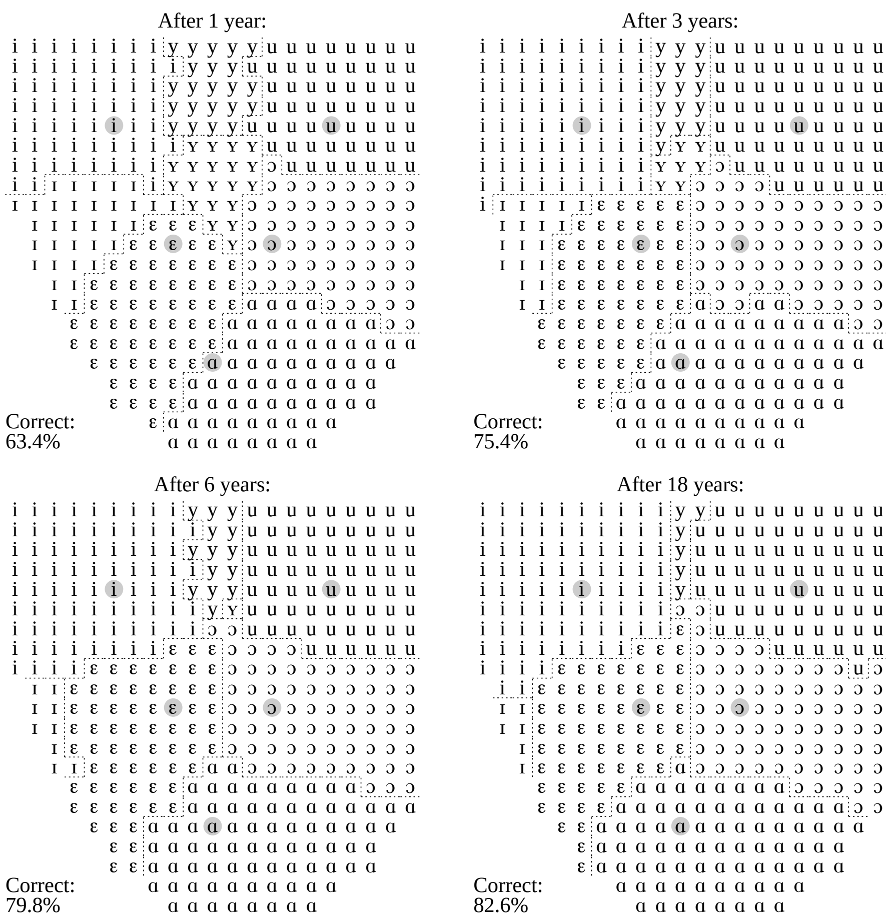


Figure 9. The perception of Spanish by a Dutch learner after 1, 3, 6, and 18 years. Grey disks: the Spanish vowel centres.

4.5. The need for negatively formulated cue constraints

In the present paper we have been using cue constraints with negative formulations, such as "an F1 of 400 Hz is not /ɑ/$_D$". Couldn't we just have used positively formulated cue constraints instead, like "an F1 of 400 Hz is /ɑ/$_D$"? There are two cases in which this makes no difference. The first case is that of a single auditory continuum, as in § 3: in tableau (8), in which every candidate violates a single constraint, we can simply rank positively formulated constraints in the reverse order of their negatively formulated counterparts, and the outcome will be the same. The second case is that of multiple auditory continua but only two different vowel categories (Escudero & Boersma 2003, 2004): if we have only two categories /A/ and /B/, the constraint "an F1 of 400 Hz is not /A/" is simply equivalent to the constraint "an F1 of 400 Hz is /B/".

But the equivalence does not generalize to cases with two (or more) auditory continua and more than two categories. For instance, an 18-year simulation of the acquisition of L1 Dutch with positively formulated cue constraints leads to a grammar that exhibits the behaviour in Figure 10, with a fraction correct of 44.9 % for the perception of Dutch, an achievement dramatically worse than that of the negatively formulated constraints of Figure 7, which scored 78.2 %. In Figure 10, the highest-ranked positively formulated constraint is "[height=6] is /ɛ/$_D$"; an entire row of epsilons (the sixth row from below) shows that this constraint has a non-local influence throughout the place continuum. The second-highest constraint is "[place=3] is /i/$_D$"; a complete column of i's (the third column from the left) shows that it has a non-local influence throughout the height continuum.[11] It can easily be seen that there exists no ranking of these positively formulated constraints that yields a separation into locally confined areas like those that appear in Figure 7: the top-ranked ones always determine the perception of entire rows or columns in the vowel grid.[12]

11 Computationally inclined readers may wonder why one cannot successively erase lines and columns with identical symbols from Figure 10 until the figure is empty. This is because Figure 10 is based on repeated stochastic evaluations (§ 4.3), not on a fixed ranking.

12 We repeated the same simulations with OT's predecessor Harmonic Grammar (HG; Legendre, Miyata, and Smolensky 1990), where the ranking values are additive weights. With the same type of evaluation noise that turns OT into Stochastic OT, our "Stochastic HG" learners end up with a good separation of the categories, scoring about 78 % correct, both for negatively and positively formulated cue constraints. Whether real humans use OT with negative constraints or HG with negative or positive constraints cannot be assessed on the basis of our data or simulations. Biological reality may well be more complex than both OT and HG.

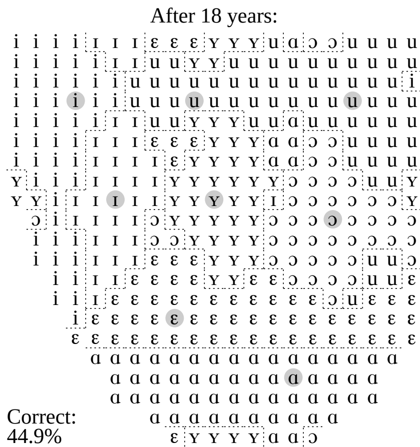


Figure 10. The failure of learning L1 Dutch with positively formulated cue constraints.

5. Discussion

Negatively formulated Optimality-Theoretic constraints can handle the categorization of both 1-dimensional and 2-dimensional auditory continua as attested in listening experiments, at least if every category spans a compact local region in the auditory space. Our Optimality-Theoretic perception model shares this property with several connectionist models, starting with the perceptron (Rosenblatt 1962), and with Massaro's (1987) fuzzy logical model of perception. But unlike these other models of perception, it makes a connection with phenomena that phonologists have traditionally been interested in, as witnessed by the perceptual processes that have been modelled in Optimality Theory: the interpretation of metrical feet, which requires structural constraints like IAMBIC and WEIGHT-TO-STRESS (Tesar 1997, 1998; Tesar and Smolensky 2000; Apoussidou and Boersma 2003, 2004); sequential abstraction, which can be handled by the interaction of structural constraints and cue constraints like the Obligatory Contour Principle and the Line Crossing Constraint (Boersma 1998, 2000); the interaction of structural constraints and auditory faithfulness in the categorization of vowel height (Boersma 1998) or consonant length (Hayes 2001); truncation by infants, which requires structural constraints like WORDSIZE (Pater 2004); and ghost segments, which can be handled by the interaction of structural and cue constraints (Boersma 2007).

The general usefulness of modelling perception in Optimality Theory extends to the specific kinds of cue constraints described here, which are not specific to the task of learning a smaller L2 vowel system. The same kind of constraints have been applied to learning to perceive a *larger* L2 vowel system, i. e. an inventory with new sounds (from Spanish to English: Escudero & Boersma 2004), and to learning an equally large L2 vowel system, i. e. an inventory with similar but non-identical sounds (from Canadian English to Canadian French: Escudero 2005), and they have been combined with auditory-to-auditory constraints in the modelling of L1 category formation (Boersma, Escudero & Hayes 2003).

Optimality-Theoretic accounts of perception and its acquisition thus bridge the gap between phonological theory and the computational modelling of human speech processing.

References

Apoussidou, Diana and Paul Boersma
2003 The learnability of Latin stress. *Proceedings of the Institute of Phonetic Sciences Amsterdam* 25: 101–148.

Apoussidou, Diana and Paul Boersma
2004 Comparing two Optimality-Theoretic learning algorithms for Latin stress. In Vineeta Chand, Ann Kelleher, Angelo J. Rodríguez and Benjamin Schmeiser (eds.), *Proceedings of the 23rd West Coast Conference of Formal Linguistics*, 29–42. Somerville, MA: Cascadilla.

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

Boersma, Paul
1998 *Functional Phonology*. PhD dissertation, University of Amsterdam. The Hague: Holland Academic Graphics.

Boersma, Paul
1999 On the need for a separate perception grammar. Manuscript, University of Amsterdam. [*Rutgers Optimality Archive* 358]

Boersma, Paul
2000 The OCP in the perception grammar. Manuscript, University of Amsterdam. [*Rutgers Optimality Archive* 435]

Boersma, Paul
2001 Phonology-semantics interaction in Optimality Theory, and its acquisition. In Robert Kirchner, Wolf Wikeley and Joe Pater (eds.), *Papers in Experimental and Theoretical Linguistics*, Volume 6, 24–35. Edmonton: University of Alberta.

Boersma, Paul
2007 Some listener-oriented accounts of *h-aspiré* in French. *Lingua* 117: 1989–2054.

Boersma, Paul, Paola Escudero and Rachel Hayes
2003 Learning abstract phonological from auditory phonetic categories: An integrated model for the acquisition of language-specific sound categories. *Proceedings of the 15th International Congress of Phonetic Sciences*, 1013–1016.

Boersma, Paul and Bruce Hayes
2001 Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32: 45–86.

Bradlow, Ann
1995 A comparative study of English and Spanish vowels. *Journal of the Acoustical Society of America* 97: 1916–1924.

Bradlow, Ann
1996 A perceptual comparison of the |i|-|e| and |u|-|o| contrasts in English and in Spanish: Universal and language-specific aspects. *Phonetica* 53: 55–85.

Broselow, Ellen
2003 Language contact phonology: Richness of the stimulus, poverty of the base. In Keir Moulton and Matthew Wolf (eds.), *NELS 34: Proceedings of the 34th Annual Meeting of the North-Eastern Linguistic Society.* Amherst: Graduate Linguistic Student Association of the University of Massachusetts.

Curtin, Suzanne, Toben H. Mintz and M. H. Christiansen
2005 Stress changes the representational landscape: Evidence from word segmentation. *Cognition* 96: 233–262.

Escudero, Paola
2005 *The Attainment of Optimal Perception in Second-Language Acquisition*. Ph. D. dissertation, University of Utrecht. Utrecht: Landelijke Onderzoeksschool Taalwetenschap.

Escudero, Paola and Paul Boersma
2002 The subset problem in L2 perceptual development: Multiple-category assimilation by Dutch learners of Spanish. In Barbora Skarabela, Sarah Fish and Anna H.-J. Do (eds.), *Proceedings of the 26th annual Boston University Conference on Language Development*, 208–219. Somerville, MA: Cascadilla.

Escudero, Paola and Paul Boersma
2003 Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. In Sudha Arunachalam, Elsi Kaiser and Alexander Williams (eds.), *Proceedings of the 25th Annual Penn Linguistics Colloquium. Penn Working Papers in Linguistics* 8.1, 71–85.

Escudero, Paola and Paul Boersma
2004 Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26: 551–585.

Escudero, Paola and Paul Boersma
to appear Language modes and perceptual development in Dutch learners of Spanish.

Gerrits, Ellen
2001 The categorisation of speech sounds by adults and children. PhD dissertation, University of Utrecht.

Hayes, Rachel
2001 An Optimality-Theoretic account of novel phonetic category formation in second language learning. Manuscript, University of Arizona.

Helmholtz, H. von
1910 *Handbuch der physiologischen Optik. Vol. 3.* Hamburg: Leopold Voss.

Jusczyk, Peter W., Anne Cutler and N. J. Redanz
1993 Infants' preference for the predominant stress patterns of English words. *Child Development* 64: 675–687.

Jusczyk, Peter W., Derek M. Houston and M. Newsome
1999 The beginnings of word segmentation in English-learning infants. *Cognitive Psychology* 39: 159–207.

Kenstowicz, Michael
2001 The role of perception in loanword phonology. *Linguistique africaine* 20.

Koopmans-van Beinum, Florien J.
1980 Vowel contrast reduction. An acoustic and perceptual study of Dutch vowels in various speech conditions. PhD dissertation, University of Amsterdam.

Legendre, Géraldine, Yoshiro Miyata and Paul Smolensky
1990 Harmonic Grammar – a formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, 884–891. Cambridge, MA: Erlbaum.

Liljencrants, Johan and Björn Lindblom
1972 Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48: 839–862.

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

McQueen, James M.
2005 Speech perception. In Koen Lamberts and Robert L. Goldstone (eds.), *The Handbook of Cognition*, 255–275. London: Sage Publications.

McQueen, James M. and Anne Cutler
1997 Cognitive processes in speech perception. In William J. Hardcastle and John Laver (eds.), *The Handbook of Phonetic Sciences*, 566–585. Oxford: Blackwell.

Massaro, Dominic William
1987 *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry.* Hillsdale: Lawrence Erlbaum.

Pater, Joe
2004 Bridging the gap between perception and production with minimally violable constraints. In René Kager, Joe Pater and Wim Zonneveld (eds.), *Constraints in Phonological Acquisition*, 219–244. Cambridge: Cambridge University Press.

Polka, Linda and Janet F. Werker
1994 Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance* 20: 421–435.

Polka, Linda, Megha Sundara and Stephanie Blue
2002 The role of language experience in word segmentation: A comparison of English, French, and bilingual infants. Paper presented at the 143rd Meeting of the Acoustical Society of America: Special Session in Memory of Peter Jusczyk, Pittsburgh, Pennsylvania.

Pols, Louis C. W., H. R. C. Tromp and Reinier Plomp
1973 Frequency analysis of Dutch vowels from 50 male speakers. *Journal of the Acoustical Society of America* 53: 1093–1101.

Rosenblatt, Frank
1962 *Principles of Neurodynamics; Perceptrons and the Theory of Brain Mechanisms*. Washington: Spartan Books.

Smolensky, Paul
1996 On the comprehension/production dilemma in child language. *Linguistic Inquiry* 27: 720–731.

Tesar, Bruce
1997 An iterative strategy for learning metrical stress in Optimality Theory. In Elizabeth Hughes, Mary Hughes and Annabel Greenhill (eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development*, 615–626. Somerville, MA: Cascadilla.

Tesar, Bruce
1998 An iterative strategy for language learning. *Lingua* 104: 131–145.

Tesar, Bruce and Paul Smolensky
2000 *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.

Werker, Janet F. and R. C. Tees
1984 Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7: 49–63.

Yip, Moira
2006 The symbiosis between perception and grammar in loanword phonology. *Lingua* 116: 950–975.

The effect of perceptual factors in the acquisition of an L2 vowel contrast

Juli Cebrian

1. Introduction

Second language (L2) speech is commonly characterized by the failure to sound like native speech, particularly when L2 learning starts after childhood (Scovel 1988; Long 1990; Flege, Munro and MacKay 1995, among others). One factor responsible for L2 learners' difficulty to establish accurate, target-like categories for L2 sounds is the influence of the learners' first language (L1). For instance, Trubetzkoy ([1939] 1969) argued that the phonology of the L1 may cause learners to filter out acoustic differences that are not phonemically relevant in the L1. More recent work relates this difficulty to the loss of perceptual sensitivity to non-native sounds in the course of L1 acquisition (Rochet 1995; Strange 1995; Iverson et al. 2003). Adult L2 learners tend to perceive non-native sounds in terms of their native categories and consequently hear L2 sounds as instances of L1 sounds, that is to say, they "assimilate" non-native sounds to L1 categories (Best and Strange 1992; Best 1995). Thus research on L2 phonology has focused on the relationship between cross-language phonetic distance and sensitivity to non-native sounds. According to the Contrastive Analysis Hypothesis (Lado 1957) L2 sounds that are closer or similar to L1 sounds will be easier to learn than more distant or newer sounds. This view is challenged by more recent models. Best and colleagues' Perceptual Assimilation Model proposes that category formation for L2 sounds is more likely to occur in the case of L2 sounds that are moderately similar to L1 sounds than in the case of sounds that are very similar to L1 sounds or else are too dissimilar to be assimilated to any L1 category (Best and Strange 1992). According to Flege's Speech Learning Model (1995, 2003), the greater the perceived dissimilarity between L1 and L2 sounds, the greater the likelihood that learners will establish target-like categories, given sufficient exposure to and experience with the target language.

Non-native speakers may also fail to perceive and produce L2 sounds accurately if they differ from native speakers in their use of acoustic information, or cues, in the formation of target L2 sound categories. For instance, studies have found that whereas American English speakers attend to F_2 and F_3 differences as the main cue to the /ɹ/-/l/ distinction, Japanese learners of English rely on duration differences (Underbakke et al. 1988) or variations in F_2 only (Iverson et al. 2003). One question that arises is whether non-native cues are available to L2 learners. For example, some studies have examined the acquisition of temporal contrasts by L2 speakers whose L1 has no such contrast. McAllister, Flege and Piske (2002) examined the acquisition of the Swedish phonemic length contrast by speakers of Estonian, English and Spanish, and found that success in learning the contrast was related to the role of duration in the L1. These results supported their Feature Hypothesis, which claims that an L2 contrastive category will be difficult to acquire if it is based on a phonetic feature not exploited in the L1. A contrasting approach is Bohn's (1995) Desensitization Hypothesis, which claims that late learners can detect temporal differences between a pair of unfamiliar L2 vowels more readily than spectral differences. Supporting evidence comes from studies which show that L2 English speakers exploit temporal cues to a greater extent than spectral cues in differentiating between /i/ and /ɪ/. This is found with learners whose L1 has temporal contrasts, e. g., Hungarian, Arabic and Japanese speakers (Altenberg and Vago 1987; Munro 1993; Minnick-Fox and Maeda 1999), but crucially also with learners whose L1 does not make use of duration, such as Spanish, Korean and Mandarin Chinese speakers (Flege, Bohn and Jang 1997; Wang and Munro 1999).

The general goal of this paper is to examine the role of native and non-native cues in the categorization of a second language contrast. The specific goal is to evaluate the use that Catalan adult learners of English make of spectral and temporal cues in the perception and production of the English tense vs. lax vowel contrast. First, the perceptual similarity between English high and mid front vowels (/i/, /ɪ/, /eɪ/ and /ɛ/) and the acoustically closest Catalan vowels is examined in order to assess whether the L2 contrasts have a match in the L1. The learners' categorization of the English vowel contrast is then examined in a series of experiments. A perception experiment evaluates the relative weighting of acoustic cues by means of synthetic stimuli varying in spectral and temporal characteristics. The production of the target L2 vowels is assessed acoustically and by means of intelligibility tests, and the results are compared to the previous experiments.

2. The L1 and the L2

The target feature in this study is the so-called lax-tense contrast in English, particularly with respect to the English high and mid front vowels (i. e., /i/-/ɪ/,

/eɪ/-/ɛ/). This opposition is associated with variations in height and backness and a difference in vowel duration (Ladefoged and Maddieson 1996). The L1 is the Eastern variety of Catalan. The Catalan vowel inventory consists of seven vowels (/i, e, ɛ, a, ɔ, o, u/) plus the reduced unstressed vowel [ə], with four degrees of height, and no lax-tense or temporal contrast (Recasens 1993). In addition to the monophthongs, Catalan has a number of diphthongs involving the high glides such as /ej/ (e. g., *rei* 'king'). A comparison of the first and second formants of the high and mid vowels reveals that Catalan /i/, /e/ and /ɛ/ fall within the acoustic vowel space of the English vowels /i/, /ɪ/ and /ɛ/, respectively (see Table 1).

Table 1. F1 and F2 values of high and mid front vowels for male speakers of Catalan (Recasens 1984), American English (a = Peterson and Barney 1952; b = Hillenbrand et al. 1995), and British English (c = Deterding 1997).

Catalan				English						
Vowel	F1	F2	Examples	Vowel	F1[a]	F2[a]	F1[b]	F2[b]	F1[c]	F2[c]
/i/	276	2156	*nit* 'night'	/i/	270	2290	342	2322	275	2221
/e/	397	1982	*nét* 'grandson'	/ɪ/	390	1990	427	2034	382	1958
				/eɪ/			476	2089		
/ɛ/	544	1811	*net* 'clean'	/ɛ/	530	1840	580	1799	560	1797

English native speakers have been found to rely mostly on spectral cues in differentiating tense and lax vowels. In a study on the role of duration in vowel recognition, Hillenbrand, Clark and Houde (2000) found that the lax-tense vowel pairs in high vowels (/i/-/ɪ/ and /u/-/ʊ/) are in fact minimally affected by duration. Native English speakers' reliance on spectral cues is also reported in a number of works evaluating the relative weighting of cues in native and L2 English, as discussed above (these studies mostly involve North American English speakers; see Escudero (2001) for a study on Scottish English and Southern England English speakers).

3. Perceptual assimilation task

Researchers have increasingly been employing perceptual assimilation tasks to determine the degree of cross-language phonetic similarity (Schmidt 1996; Flege, Bohn and Jang 1997; Ingram and Park 1997;

Strange et al. 1998, 2001, 2005; Guion et al. 2000, among others). In these tasks, listeners with no L2 experience are presented with L2 speech stimuli, and asked to indicate to which L1 phonetic category each L2 token is most similar, and rate its "goodness" as an exemplar of that category, as discussed below.

3.1. Subjects. 20 native speakers of Catalan with little or no knowledge of English (mean age: 28, range: 19–47) participated in the experiment.

3.2. Stimuli. English vowel stimuli were elicited from two male speakers of Canadian English. Each talker read a list containing English target vowels in monosyllabic words of the form /h/ + vowel (e. g., *hee*) in the case of the tense vowels and /h/ + vowel + /b/ in the case of the lax vowels (e. g., *hib*, *heb*). This particular consonant environment was chosen in order to minimize C to V and V to C tongue coarticulation (Strange et al. 1998), facilitating the extraction of the vowel in order to prepare the vowel stimuli for the experiment. Data were digitized at a 10 kHz sampling rate and normalized for peak intensity. The vowel portions were then edited out from each test word so as to minimize the effect of consonant properties that might create an impression of a foreign accent.

3.3. Procedure. The subjects were presented with randomized tokens of the four English vowels and had to choose from four options representing the Catalan high and mid front vowels and the Catalan diphthong /ej/ in conventional Catalan orthography, namely, *i*, *é*, *è* and *ei* (representing /i/, /e/, /ɛ/ and /ej/, respectively). Subjects chose the alternative that best corresponded to the vowel they heard. After selecting an option, the same stimulus was heard again and subjects selected a goodness rating from a 7-point scale according to how closely that sound approximated the Catalan vowel they had just selected. A score of '1' corresponded to a poor exemplar of the chosen response vowel, and '7' corresponded to a good exemplar. Subjects also heard and rated actual Catalan vowels, which were mixed in the task with the English vowels, for control purposes (see Cebrian (2006) for further details). The task was preceded by a training period to familiarize the subjects with the procedure and adjust the listening level.

3.4. Results. Table 2 shows the percentage of time that each English vowel was heard as, or assimilated to, a Catalan vowel (i. e., assimilation scores) and the mean goodness rating obtained by each English vowel. Although the acoustic comparison showed a very close correspondence between Catalan and English vowels (see Table 1 above), the perceptual comparison suggests that the English vowels are assimilated to the Catalan vowels to different degrees. This discrepancy between perceived similarity and acoustic distance is consistent with the findings in previous studies (Flege 1991; Stevens et al. 1996; Bohn, Strange and Trent 1999), and illustrates that direct assessment of assimilation patterns is necessary for measuring cross-linguistic similarity. Other acoustic properties in addition to steady-state F1 and F2 values may need to be evaluated when predicting perceptual distance.

Table 2. Perceptual assimilation of English vowels to Catalan vowels and goodness ratings.

	English Target Vowels							
	/i/		/eɪ/		/ɪ/		/ɛ/	
Responses	%	Rating	%	Rating	%	Rating	%	Rating
i (/i/)	99	6.2			14	2.7		
ei (/ej/)	1	1.0	84	4.6				
é (/e/)			13	4.0	66	3.5	7	3.8
è (/ɛ/)			3	2.8	20	3.0	93	4.2

The results indicate that the English vowel /i/ obtained the highest assimilation scores to the acoustically closest Catalan vowel (/i/) and the highest goodness ratings (99 % and 6.2, respectively). Therefore, English /i/ is perceived as nearly identical to Catalan /i/. The high vowel was followed in perceived similarity by English /ɛ/, strongly assimilated to Catalan /ɛ/. English /eɪ/ was consistently identified with the Catalan diphthong /ej/ rather than with the monophthong /e/, indicating that the high offglide is a crucial cue to its identification. Finally, vowel /ɪ/ was the least readily assimilated to an L1 vowel and obtained the lowest goodness ratings, patterning as the most dissimilar vowel. If, as predicted by recent theories (Best 1995; Flege 1995), perceived similarity has an effect on the ability to create accurate L2 categories, we will expect the most similar vowels (/i, ɛ, eɪ/) to pattern differently from the dissimilar vowel /ɪ/ in the perception and production experiments.

4. Perception of L2 vowels

Perception of the tense-lax vowel contrast was assessed using a vowel identification task involving synthetic stimuli, as described below.

4.1. Subjects. 30 Catalan learners of English and 20 native Southern Ontario English speakers participated in the experiment. The Catalan speakers were undergraduate students in English Philology at the Universitat Autònoma de Barcelona, Spain. They were in their third or fourth university year (mean age: 22 years). English was the language of instruction and study in most of their courses. In addition, many had spent between a few weeks and a year in an English speaking country. They were bilingual in Catalan and Spanish, but they were Catalan-dominant bilinguals as assessed by a questionnaire and a brief interview with the experimenter. Finally, the English native speaker group was mostly made up of undergraduate and graduate students at the University of Toronto (mean age: 36).

4.2. Stimuli. The two-dimensional /i/-/ɪ/-/ɛ/ English continuum consisted of 11 vowel quality steps and four temporal steps. With respect to the 11 quality steps, vowels 1, 6 and 11 corresponded to the prototypical spectral values for English /i/, /ɪ/ and /ɛ/, respectively, based on Peterson and Barney's (1952) values for male voices (see Table 1 above). Intermediate vowels were calculated in linear steps. The four different durations were 100, 150, 200 and 250 ms. The synthetic continuum was created following Klatt's (1980) parameters for vowels in isolation, and using Computerized Speech Lab software.[1]

4.3. Procedure. The task was a three-alternative forced-choice task in which each response alternative consisted of English words written in English orthography and representing one of the three target English vowels (/i/, /ɪ/, /ɛ/), namely *beat*, *bit*, and *bet*. A trial consisted of two presentations of each stimulus with an inter-stimulus interval of two seconds. After hearing a given stimulus, subjects selected a response from the alternatives presented on the screen. Each stimulus appeared five times (five trials). The task was preceded by a practice period to familiarize the subjects with the procedure.

4.4. Results. The results for the native English speakers and the Catalan learners of English are illustrated in Figures 1–6, which provide the percentages of *beat* (Figures 1 and 2), *bit* (Figures 3 and 4) and *bet* (Figures 5 and 6) responses for each vowel stimulus. The eleven vowel quality steps from /i/ to /ɪ/ to /ɛ/ are represented on the x-axis while the four durations are represented by the lines in the graph.

---

1 Vowel /eᶦ/ was not included in this test for two reasons. First, this vowel is strongly assimilated to an L1 diphthong rather than a monophthong so that the main cue to its identification, a change in quality, is probably different from the main cue for the other L2 vowels. Secondly, the acoustic characteristics of /eᶦ/, i.e., the overlap between the steady state format values for /eᶦ/ and /ɪ/, and /eᶦ/'s greater formant transitions, complicate its inclusion in the continuum.

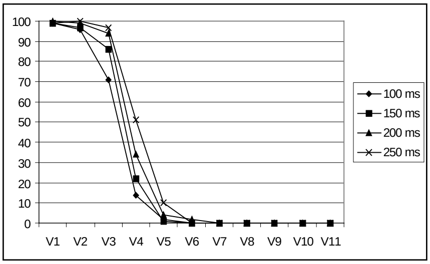


Figure 1. Percentages of *beat* responses for native English speakers.

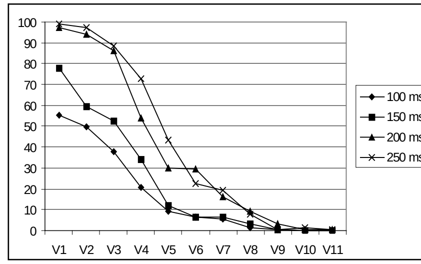


Figure 2. Percentages of *beat* responses for Catalan learners of English.

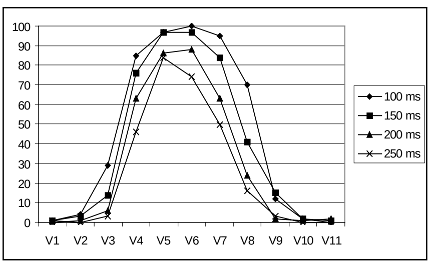


Figure 3. Percentages of *bit* responses for native English speakers.

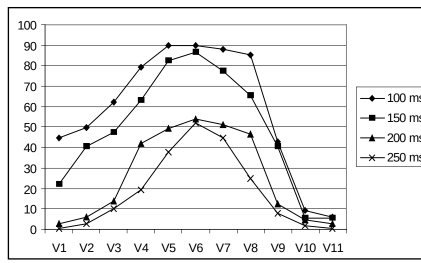


Figure 4. Percentages of *bit* responses for Catalan learners of English.

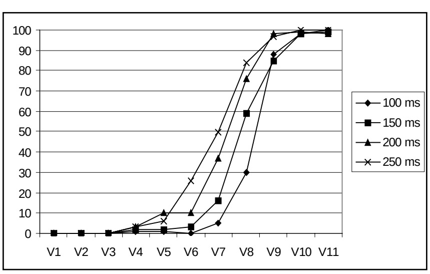


Figure 5. Percentages of *bet* responses for English native speakers.

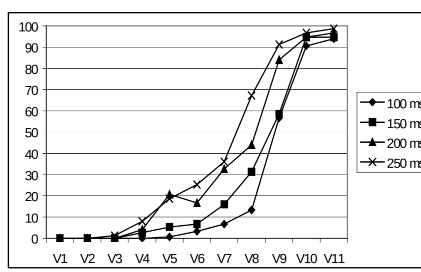


Figure 6. Percentages of *bet* responses for Catalan learners of English.

As shown in Figures 1, 3 and 5, the English speaking group displayed clear crossovers from /i/ to /ɪ/ and from /ɪ/ to /ɛ/, indicating a consistent pattern of vowel categorization based on vowel quality rather than duration. The L2 learners' results for /ɛ/ closely resemble the native speakers' (Figure 6). However, their identification scores for /i/ and /ɪ/ appear to be affected by temporal differences. For example, the number of *beat* responses for the prototypical vowel /i/ (left end of the continuum in Figure 2) increases as a function of duration. In contrast, in the case of vowel /ɪ/ (Vowel 6 in the continuum, Figure 4), shorter tokens obtain higher *bit* responses than longer tokens. Thus, with equal spectral characteristics, the shorter the vowel, the more likely it is to be identified as the vowel in *bit*, whereas the longer the vowel, the greater the number of *beat* responses (see also Table 3).

Table 3. Correct identification of each prototypical vowel in each duration condition.

Correct Identification	Group	100ms	150ms	200ms	250ms	Means
% *beat* resp. for Vowel 1 (/i/)	Catalans	55	78	97	99	83
	English	99	99	100	99	99
% *bit* resp. for Vowel 6 (/ɪ/)	Catalans	90	87	54	52	71
	English	100	97	88	74	90
% *bet* resp. for Vowel 11 (/ɛ/)	Catalans	94	95	97	99	96
	English	100	99	98	100	99

A statistical analysis was conducted on the percentage of *beat* responses obtained for the prototypical vowel /i/ (Vowel 1), the percentage of *bit* responses for /ɪ/ (Vowel 6) and of *bet* responses for /ɛ/ (Vowel 11); in other words, on the percentage correct responses for each prototypical vowel. A three way ANOVA was performed with Language as a between groups factor (English speakers and Catalan speakers), and Duration (four durations) and Vowel (/i/, /ɪ/ and /ɛ/) as within groups factors. All main effects and interactions proved significant. The significant effect of Language was reflected in the higher identification scores obtained by the English native speakers ($F(1,48) = 18.28$, $p < .001$). The overall higher identification scores for /ɛ/ explain the significance of Vowel ($F(2,96) = 22.91$, $p < .001$), and the overall higher scores for shorter vowel tokens accounts for the statistical significance of Duration ($F(3,144) = 5.18$, $p < .01$). Vowel and Duration appeared to have an effect for the Catalan group but not for the English-speaking group, thus the Language x Vowel interaction ($F(2,96) = 5.42$, $p < .01$) and Language x Duration interaction ($F(3,144) = 7.1$, $p < .001$)). The two way Duration x Vowel interaction ($F(6,288) = 23.96$, $p < .001$) reflects the absence of the Duration effect with vowel /ɛ/. Finally, the three way interaction ($F(6,288) = 8.26$, $p < .001$) is due to the fact that Duration had an effect for only two of the three vowels and for only one of the two groups. Tukey HSD post hoc tests showed that the Catalans differed from the English group in the correct identification scores for vowel /i/ and /ɪ/ ($p < .001$) but not for /ɛ/ and that Duration was significant for the non-native group with respect to vowels /i/ and /ɪ/. Duration in fact influenced the English speakers' responses as well, since the identification scores dropped from 100 % for 100 ms-long /ɪ/ to 74 % in the case of 250 ms-long /ɪ/. This means that very long /ɪ/ tokens sound less natural to native speakers than average or shorter /ɪ/ tokens. Nevertheless, Catalan speakers still pattern quite differently from native English speakers since Catalans' identification rates are much lower, showing a greater effect of duration.

These results show that native Southern Ontario English speakers rely mostly on spectral cues in their identification of the English vowels /i/, /ɪ/ and /ɛ/. This outcome is in accordance with previous studies on native English speakers (Hillenbrand et al. 2000). The fact that durational differences can be salient for L2 learners even if not part of their L1 replicates previous findings (Flege, Bohn and Jang 1997; Wang and Munro 1999), and lends support to Bohn's (1995) Desensitization Hypothesis. In contrast, it argues against McAllister et al.'s (2002) Feature Hypothesis, which does not predict the availability of a non-L1 feature to adult learners. With respect to the relation between phonetic similarity and perception ability, English /ɛ/, which was strongly assimilated to Catalan /ɛ/ in the first experiment, obtains very high identification scores. However, two English vowels which obtained different assimilation scores to Catalan vowels, the near-identical vowel /i/ and the dissimilar vowel /ɪ/, appear to obtain comparable identification scores, casting some doubt on the effect of phonetic similarity and pointing to other factors at play in the categorization of L2 sounds. Before discussing this issue further, we will evaluate the production of the L2 vowels.

5. Production of L2 vowels

Production of the four target English vowels was examined by means of acoustic measurements of L2 production (first and second formant values, and duration), and listening tests with native English speakers as evaluators (an identification test and a goodness rating task). The same Catalan subjects that participated in the perception experiment took part in the production experiment. Four native Canadian English speakers also participated for control purposes.

Production was tested by eliciting the vowels in the nonsense word frame /h_b/ (e. g., *heb* (/hɛb/)). The method of elicitation was a mixture of repetition and vowel insertion. Subjects first repeated an h_d word and then produced new sets of words by inserting the vowel in the h_d word into the nonsense word frames. This method was used to avoid orthographic interference and contamination from effects of word frequency. Finally, the h_b consonant environment was chosen in order to minimize C to V and V to C tongue coarticulation (Strange et al 1998), permitting a clearer analysis and extraction of the vowel portion for the listening tests. Subjects' responses were digitized at a 10 kHz sampling rate, and saved as audio files for both acoustic analysis and to provide the stimuli for the identification and goodness rating tasks.

5.1. Acoustic analysis

The steady state F1 and F2 frequencies of the English vowels produced by the Catalan subjects are plotted in Figure 7. Results for /eɪ/ are not included given the acoustic overlap between this vowel and /ɪ/ and the fact that formant movement is a crucial cue to its identification.

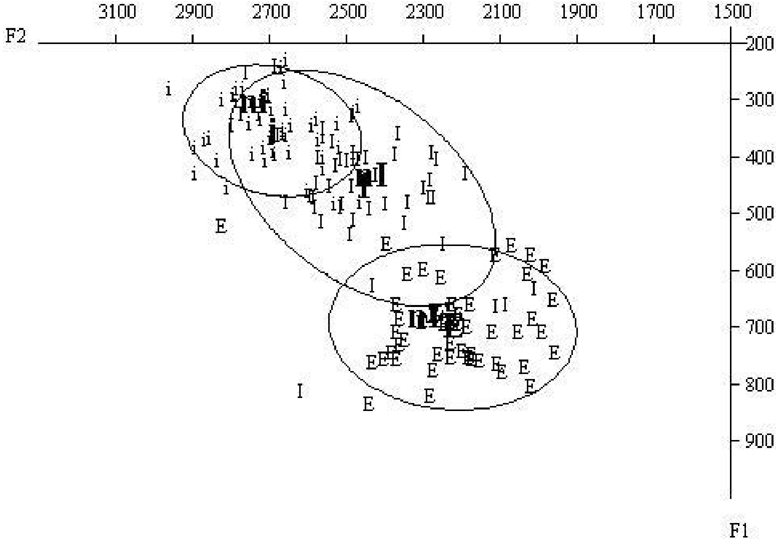


Figure 7. Catalan speakers' F1 and F2 values for English /i/, /ɪ/ and /ɛ/ (boldface symbols indicate the group's mean values; the corresponding means for the control native English group are also included, indicated by boldface symbols preceded by an "n").

The results show that the mean F1 and F2 values for the Catalan group (boldface symbols) are very close to the native English values (boldface symbols preceded by "n"). However, there is great variability in vowel quality as illustrated by the size and overlapping areas of the vowels' acoustic space, especially in the case of /ɪ/, with a large /ɪ/-/i/ overlap. This indicates that most Catalans failed to produce a spectral contrast between these two vowels. Vowel /ɛ/ appears to be more clearly differentiated spectrally.

With respect to vowel duration, learners evidenced a native-like implementation of duration distinctions. These are given in Table 4, which includes standard deviations in parentheses. These values are comparable to those obtained by the English control group (/i/: 257 ms, /ɪ/: 144 ms, /eɪ/: 292 ms, /ɛ/: 169 ms). The duration differences between vowels reached significance in a one-way analysis of variance ($F(3,84) = 60.39$, $p < .001$). This pattern of vowel differences is consistent with results for native English speakers (e. g., Hillenbrand et al. 1995; Giegerich 1992; Lindsey 1990, among others). Thus, perhaps the most consistent finding of the production and perception experiments has to do with the use of duration by Catalan learners of English, in contrast with the absence of systematic temporal differences in the L1, where /i/ tends to be shorter (78 ms) than /e/ (86 ms) and /ɛ/ (102 ms) (Recasens 1984). As with the perception data, this finding is consistent with Bohn's (1995) Desensitization Hypothesis.

Table 4. Mean durations in ms for each English vowel and mean tense/lax vowel duration ratios in the L2 data (standard deviations are given in parentheses).

/i/	/ɪ/	/i/-/ɪ/ ratio	/eɪ/	/ɛ/	/eɪ/- /ɛ/ ratio
243 (64)	153 (33)	1.62	257 (45)	183 (36)	1.44

5.2. Intelligibility tests

Previous research has shown it is important to complement acoustic analyses with perceptual measures (e. g., Munro 1993; Munro, Flege and MacKay 1996; Flege 1997; Flege, Bohn and Jang 1997). However, it is clear that not all utterances that are identified with the same target vowel in a listening test are in fact phonetically equivalent (Hillenbrand et al. 1995). L2 speech may be accented yet highly identifiable (Munro and Derwing 1995; Munro, Flege and MacKay 1996). In order to distinguish between intelligibility and native-like production, the L2 productions were assessed for accuracy by means of both a vowel identification task and a goodness rating task.

5.2.1. Stimulus preparation. Stimuli consisted of the vowel portions edited out from each h_b test word so as to minimize the effect of consonant properties that might create an impression of a foreign accent. The vowel portion was extracted leaving intact cues to vowel identity insofar as possible. Signal editing was carried out visually and aurally with Praat software by examining both the amplitude of a waveform and the formant structure on a spectrogram. The vowel portion comprised from the first to the last positive peak in the periodic portion of the signal as indicated by an increase/decrease in overall amplitude and waveform complexity. The selected vowel portion was windowed out smoothing the onset and offset with a ramping function and making the initial and final splices at zero crossings.

5.2.2. Procedure. The first listening test was a forced choice vowel identification task. Listeners were asked to identify the vowels they heard as the vowel in one of the six alternatives that appeared on the computer screen, namely, 'had,' 'heed,' 'hid,' 'hayed,' 'head' and 'hub'. Vowels were presented in a randomized order. In the second listening task, the goodness rating task, stimuli were grouped in blocks of the same intended vowel and indicated with the corresponding IPA phonetic symbol. Listeners were asked to provide a goodness rating using a 7-point scale: '1' represented a poor exemplar of the target vowel, '7' represented a good English-sounding vowel.

5.2.3. Listeners. Eight native Canadian English speakers participated in the vowel identification task, mostly undergraduate students at the University of Toronto. A different set of eight listeners participated in the goodness rating task. In this case, the listeners were graduate students and other members of the Linguistics Department at the University of Toronto.

5.2.4. Results. The results of the two listening tests are summarized in Table 5 below, which provides the mean identification scores and goodness ratings (standard deviations are given in parentheses). The results for the English control group ranged from 87 to 96 % in identification scores and 5.5 to 5.7 in goodness ratings.

The vowel /eᴵ/ obtained the highest identification scores and goodness ratings, followed by /ɛ/, while /i/ and /ɪ/ obtained the lowest identification and goodness ratings scores. The goodness ratings for /i/ were somewhat higher than those for /ɪ/. The more frequently identified vowels obtained the highest goodness ratings. A statistical analysis yielded significant correlations between identification scores and goodness ratings for vowels /i/ and /ɪ/ ($r(29) = .91$, $p < .001$ and $r(29) = .76$, $p < .001$, respectively). The correlations were not significant in the case of the other two vowels probably due to small variance given their high scores.

An analysis of the acoustic characteristics of the production data indicates that /i/ tokens with longer duration, lower F1 and greater F2-F1 difference were better identified and rated, whereas in the case of vowel /ɪ/, better results in the listening tests corresponded to the opposite characteristics, that is, shorter vowel duration, higher F1, lower F2 and smaller F2-F1 difference. With respect to the relationship between perception and production, no significant correlations were found in the current study. However, in a study including the same group of subjects, significant correlations were obtained between perception and production data with a different set of perception experiments involving natural stimuli (Cebrian 2002). In that case perception results were better than production results, supporting the view that accurate perception precedes accurate production in L2 (Flege 1995, 1999; Rochet 1995).

Table 5. Percentages of correct identification and goodness ratings of the L2 vowels produced by the Catalan subjects.

Correct Identification				Goodness Rating			
/i/	/ɪ/	/eɪ/	/ɛ/	/i/	/ɪ/	/eɪ/	/ɛ/
73	71	99	86	4.5	3.9	5.5	5.2
(33)	(33)	(4)	(9)	(1.3)	(1.3)	(0.5)	(0.7)

6. Discussion

The results of the perception experiment showed that Catalans categorized the English vowel /ɛ/ in a target-like fashion and perceived it on the basis of vowel quality. This is also observed in the production data, where both /eɪ/ and /ɛ/ are the most accurately produced L2 vowels. Recall that English /eɪ/ and /ɛ/ were consistently assimilated to (that is, heard as good instances of) Catalan /ej/ and /ɛ/, respectively, in the perceptual assimilation task. Thus, Catalans may be using the L1 vowel categories when perceiving and producing the strongly assimilated English vowels /eɪ/ and /ɛ/. It is possible that the Catalan vowels are perceived as acceptable instances of English /eɪ/ and /ɛ/ by English native speakers so that Catalans' use of Catalan /ej/ and /ɛ/ in English goes mostly unnoticed by English-speaking listeners. The results for these two vowels are in agreement with the theories reviewed above. The Contrastive Analysis Hypothesis (Lado 1957) predicts that similar vowels will be easier to learn. The Speech Learning Model (Flege 1995, 2003) and the Perceptual Assimilation Model (Best and Strange 1992; Best 1995) allow for L2 vowels that are very close to L1 vowels, or 'near-identical' vowels, to pass as good instances of target L2 vowels.

With respect to the other two vowels, in general Catalan subjects were not very successful at forming a target-like category for the weakly assimilated vowel /ɪ/. This would be predicted by a contrastive analysis approach. According to Best and Flege's models this vowel may be authentically categorized given enough experience with the L2. This prediction cannot be evaluated here since this study does not examine the role of experience (see Cebrian (2006) for an examination of experience). The learners in this study may be at a stage where they have not had enough exposure to form an accurate category for this new vowel. Importantly, the results for English /i/ do challenge the predictions of the different models: despite the high degree of assimilation of English /i/ to the Catalan /i/ (99 % assimilation scores and a goodness rating of 6.2 out of 7), this vowel obtained perception and production scores that are comparable to those for the weakly assimilated vowel /ɪ/ rather than the strongly assimilated /eɪ/ and /ɛ/. Similar results are reported in other studies. For example, Flege (1992) reports that identification rates of Spanish speakers' production of English /i/ and /ɪ/, considered on the basis of spectral measurements to be similar and new, respectively, was 57 % for /i/ and 61 % for /ɪ/ for experienced late learners and 69 % /i/ and 51 % /ɪ/ for inexperienced late learners, whereas the similar vowel /ɛ/ obtained 91–99 %. The failure to obtain different results for L2 vowels with different degrees of similarity to L1 vowels runs counter to the predictions of Best and Flege's models (Best and Strange 1992; Best 1995; Flege 1995). Neither do the current results for /i/ and /ɪ/ support a contrastive analysis approach (Lado 1957), which would predict better performance with /i/ than /ɪ/.

Perceived similarity alone thus does not appear to be a good predictor for production and perception accuracy. Other factors must play a role in the categorization of L2 vowels, namely, the fact that sound categories are established upon contrastive properties that distinguish them from neighbouring sound categories. The results for both perception and production show that generally learners fail to establish a spectral distinction between /i/ and /ɪ/ the way native English speakers do. Instead, as predicted by the Desensitization Hypothesis (Bohn 1995), learners tend to distinguish these vowels temporally because there is no clear spectral match for /ɪ/ in the L1. The categorization of the /i/-/ɪ/ contrast as a temporal opposition, which is crucially not an L1 feature, may result in the fact that, despite its strong assimilation to a Catalan vowel in the perceptual similarity test, English /i/ is not categorized in terms of the closest L1 vowel. Clearly, the fact that Catalans have greater difficulty categorizing the English /i/-/ɪ/ contrast than the English /eɪ/-/ɛ/ contrast results from the fact that the former has no parallel phonemic contrast in Catalan whereas the latter does (i. e., Catalan /ej/-/ɛ/). Consequently, performance on /i/ both in perception and production is affected by the need to establish a new contrast in the inventory (/i/-/ɪ/). This illustrates that L2 vowels are acquired as part of a system in which the need to establish oppositions with neighbouring vowels, as in the case of the /i/-/ɪ/ contrast, may take precedence over L1-L2 vowel assimilation patterns, at least at some stages in the acquisition of the L2.

7. Summary and conclusions

This study has evaluated the acquisition of a contrast based on temporal and spectral differences, the English tense-lax vowel contrast, by speakers of Catalan, a language which distinguishes vowels only spectrally. A series of experiments have assessed the perceived distance between English and Catalan high and mid front vowels, and the use of spectral and temporal properties in the perception and production of the English vowel contrast by Catalan learners of English. Importantly, the L2 learners appear to make use of duration cues in their perception and production of English /i/ and /ɪ/. These results argue against McAllister, Flege and Piske's (2002) Feature Hypothesis, who posit that adult L2 learners are unlikely to perceive and exploit L2 features not used to signal phonological contrast in the L1. In contrast, the results are consistent with the Desensitization Hypothesis, which claims that adult learners may implement a non-L1 duration distinction to establish a contrast that has no spectral counterpart in the L1 (Bohn 1995).

The finding that a strongly assimilated vowel and a weakly assimilated vowel (i. e., /i/ and /ɪ/) obtain the same results in L2 perception and production or that strongly assimilated vowels yield different results (i. e., /i/ vs. /eɪ/ and /ɛ/) poses a problem to models that relate the likelihood of L2 vowel category formation to the degree of similarity between L1 and L2 sounds (Flege 1995; Best 1995; Lado 1957). These results underscore the importance of factors that interact with perceived similarity such as adequate cue weighting and categorization of neighbouring vowels. The categorization of the weakly assimilated vowel /ɪ/ and the strongly assimilated /i/ in terms of a temporal contrast affects the category formation for /i/, which no longer patterns as a strongly assimilated vowel. This emphasizes that vowels are not acquired individually but as part of a system of contrasting categories with the consequence that the formation of one vowel category can directly affect the categorization of another vowel.

Acknowledgements

This research was supported by research grants from the Department of Linguistics at the University of Toronto, by grant HUM2005–02746/FILO from the Spanish Ministry of Education and Science and by the research group grant 2005SGR00864 from the Catalan Government.

References

Altenberg, Evelyn and Robert M. Vago
1987 Theoretical implications of an error analysis of second language phonology production. In: Georgette Ioup and Steven H. Weinberger (eds.), *Interlanguage Phonology. The Acquisition of a Second Language Sound System. Series on Issues in Second Language Research*, 148–164. Cambridge, MA: Newbury House Publishers.

Best, Catherine
1995 A direct realist view of cross-language speech perception. In: Winifred Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, 233–277. Timonium, MD: York Press.

Best, Catherine and Winifred Strange
1992 Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20, 305–331.

Bohn, Ocke-Schwen
1995 Cross-language perception in adults: First language transfer doesn't tell it all. In: Winifred Strange (ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, 379–410. Timonium, MD: York Press.

Bohn, Ocke-Schwen, Winifred Strange and Sonja A. Trent
1999 On what it takes to predict perceptual difficulty in cross-language vowel perception. *Journal of the Acoustical Society of America* 105, No. 2, Pt. 2.

Cebrian, Juli
2002 Phonetic similarity, syllabification and phonotactic constraints in the acquisition of a second language contrast. PhD dissertation. Toronto Working Papers in Linguistics Dissertation Series. Toronto, Canada: Department of Linguistics, University of Toronto.

Cebrian, Juli
2006 Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics* 34, 372–387.

Deterding, David
1997 The formants of monophthong vowels in Standard Southern British English pronunciation. *Journal of the International Phonetic Association* 27, 47–55.

Escudero, Paola
2001 The role of the input in the development of L1 and L2 sound contrasts: Language-specific cue weighting for vowels. In: A. H-J. Do, L. Domínguez and A. Johansen (eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development*, Vol 1, 250–261. Somerville, MA: Cascadilla Press.

Flege, James Emil
1991 The interlingual identification of Spanish and English vowels: Orthographic evidence. *Quarterly Journal of Experimental Psychology* 43A, 701–731.

Flege, James Emil
1992 Speech learning in a second language. In: Charles A. Ferguson, Lise Menn and Carol Stoel-Gammon (eds.), *Phonological Development: Models, Research, Implications*, 565–604. Timonium, MD: York Press.

Flege, James Emil
1995 Second language speech learning: Theory, findings and problems. In: Winifred Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, 233–277. Timonium, MD: York Press.

Flege, James Emil
1997 English vowel production by Dutch talkers: More evidence for the "similar" vs. "new" distinction. In: Allan James and Jonathan Leather (eds.), *Second Language Speech: Structure and Process,* 11–52. Berlin/New York: Mouton de Gruyter.

Flege, James Emil
1999 The relation between L2 production and perception. *Proceedings of the International Congress of Phonetic Science*s, 1273–1276. San Francisco, CA.

Flege, James Emil
2003 Assessing constraints on second language segmental production and perception. In: Niels O. Schiller and Antje S. Meyer (eds.), *Phonetics and Phonology in Language Comprehension and Production, Differences and Similarities*, 319–355. Berlin: Mouton de Gruyter.

Flege, James Emil, Murray J. Munro and Ian R. A. MacKay
1995 Factors affecting degree of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 97, 3125–3134.

Flege, James Emil, Ocke-Schwen Bohn and Sunyong Jang
1997 Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25, 437–470.

Giegerich, Heinz J.
1992 *English Phonology – An introduction*. Cambridge: Cambridge University Press.

Guion, Susan G., James E. Flege, Reiko Akahane-Yamada and John S. Pruitt
2000 An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America* 107 (5) Pt. 1., 2711–2724.

Hillenbrand, James M., Laura A. Getty, Michael J. Clark and Kimberlee Wheeler
1995 Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America* 97, 3099–3111

Hillenbrand, James M., Michael J. Clark and Robert A. Houde
2000 Some effect of duration on vowel recognition. *Journal of the Acoustical Society of America* 108, 3013–3022.

Ingram, John C. L. and See-Gyoon Park
1997 Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics* 25, 343–370.

Iverson, Paul, Patricia K. Kuhl, Reiko Akahane-Yamada, Eugen Diesch, Yohich Tohkura, Andreas Kettermann and Claudia Siebert
2003 A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, B47-B57.
Klatt, Dennis H.
1980 Software for a cascade/parallel format synthesizer. *Journal of the Acoustical Society of America* 6, 971–993.
Ladefoged, Peter and Ian Maddieson
1996 *The Sounds of the World's Languages*. Oxford: Blackwell.
Lado, R.
1957 *Linguistics Across Cultures*. Ann Arbor: University of Michigan Press.
Lindsey, Geoff
1990 Quantity and quality in British and American vowel systems. In: Susan Ramsaran (ed.), *Studies in the Pronunciation of English: A Commemorative Volume in Honour of A. C. Gimson*, 106–118. New York/London: Routledge.
Long, Michael
1990 Maturational constraints on language development, *Studies in Second Language Acquisition* 12, 251–286.
McAllister, Robert, James Emil Flege and Thorsten Piske
2002 The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics* 30, 229–258. doi:10.1006/jpho.2002.0174.
Minnick-Fox, Michelle and Kazuaki Maeda
1999 Categorization of American English vowels by Japanese speakers. In: John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville and Ashlee Baily (eds.), *Proceedings of the 14th International Congress of Phonetic Sciences*, 1437–40. Berkeley, CA: University of California.
Munro, Murray J.
1993 Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech* 36, 39–66.
Munro, Murray J., James Emil Flege and Ian R. A. MacKay
1996 The effects of age of second-language learning on the production of English vowels. *Applied Psycholinguistics* 17, 313–334.
Munro, Murray J. and Tracey Derwing
1995 Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45:1, 73–97.
Peterson, Gordon E. and Harold L. Barney
1952 Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24, 175–184.
Recasens, Daniel
1984 Coarticulació de vocals i consonants del català en el decurs. Ph. D. dissertation, Universitat de Barcelona.

Recasens, Daniel
1993 *Fonètica i fonologia*. Barcelona: Enciclopèdia Catalana.

Rochet, Bernard L.
1995 Perception and production of L2 speech sounds by adults. In: Winifred Strange (ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, 379–410. Timonium, MD: York Press.

Scovel, Thomas
1988 *A Time to Speak: A Psycholinguistic Inquiry into the Critical Period for Human Speech*. Rowley, Mass.: Newbury House.

Schmidt, Anne M.
1996 Cross-language identification of consonants. Part 1. Korean perception of English. *Journal of the Acoustical Society of America* 99 (5), 3201–3211.

Strange, Winifred
1995 Phonetics of second-language acquisition: Past, present, future. In: Kjell Elenius and Peter Branderud (eds.), *Proceedings of the 1995 International Congress of Phonetic Sciences*, Vol. 4, 84–91. Stockholm: Arne Stomberg.

Strange, Winifred, Reiko Akahane-Yamada, Rieko Kubo, Sonja A. Trent, Kanae Nishi and James J. Jenkins
1998 Perceptual assimilation of American English vowels by Japanese speakers. *Journal of Phonetics* 26, 311–344.

Strange, Winifred, Reiko Akahane-Yamada, Rieko Kubo, Sonja A. Trent and Kanae Nishi
2001 Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *Journal of the Acoustical Society of America* 109, 1691.

Strange, Winifred, Ocke-Schwen Bohn, Kanae Nishi and Sonja A. Trent
2005 Contextual variation in the acoustic and perceptual similarity of North German and American English vowels. *Journal of the Acoustical Society of America* 118, 1751–1762.

Stevens, Kenneth N., Alvin Liberman, Michael Studdert-Kennedy and Sven Öhman
1996 Cross language study of vowel perception. *Language and Speech* 12, 1–23.

Trubetzkoy, N. S.
[1939] 1969 *Principles of Phonology.* Translated by C. A. Baltaxe. Berkeley, CA: University of California Press.

Underbakke, Melva, Linda Polka, Terry L. Gottfried and Winifred Strange
1988 Trading relations in the perception of /r/ and /l/ by Japanese learners of English. *Journal of the Acoustical Society of America* 84, 90–100.

Wang, Xinchun and Murray J. Munro
1999 The perception of English tense-lax vowel pairs by native Mandarin speakers: The effect of training on attention to temporal and spectral cues. In: John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville and Ashlee Baily (eds.), *Proceedings of the 14th International Congress of Phonetic Sciences*, 125–129. Berkeley, CA: University of California.

Some reflections on abstractness and the shape of inputs: The case of aspiration in English[1]

Heather Goad

1. Preliminaries

Modern phonological theory has typically aimed to provide a unique underlying representation for a given morpheme in spite of the presence of morphophonemic alternation (cf. the historical overview in Anderson 1985). The result is a one-to-many mapping between levels of representation and, accordingly, the question of what information is present in inputs has been of central importance in theory development. While early generative phonology held the view that inputs are abstract (Chomsky and Halle 1968), the advent of Optimality Theory (Prince and Smolensky 1993/2004) has marked a shift away from this position. Although Optimality Theory includes the assumption that there are no constraints on the shapes of inputs, Lexicon Optimization guides learners in the usual case to select inputs which correspond to one of the surface forms attested in the language, that is, inputs which are not underspecified. This line of thinking has been taken a step further in the work of researchers who adopt the position that the phonetics and phonology form a single module of the grammar; inputs are phonetically enriched, inconsistent with their being underspecified (see, e. g., Boersma 1998, Steriade 2000, Flemming 2001, Curtin 2002 for proposals along these lines).

In this paper, I address the question of the shapes of inputs from the vantage point of second language acquisition. The principal goal is to determine the kind of information that is stored in native-language input representations

1 I would like to thank members of the audience at the Second International Conference on Contrast in Phonology for questions and comments on an earlier version of this paper. The paper has also benefitted from discussions with Jonathan Bobaljik, Kathleen Brannen, Suzanne Curtin, Elan Dresher, Jill Heather Flegg, Joe Pater, and Linda Polka and from comments from two anonymous reviewers. I take full responsibility for the content; in particular, Curtin and Pater are in no way responsible for the (sometimes outlandish) interpretations of the data from Curtin, Goad, and Pater (1998) and Pater (2003). This research was supported by grants from SSHRC and FQRSC.

through observing the effects of transfer from the first language into the second language. Using experimentally-obtained results on the second language acquisition of laryngeal contrasts by English learners of Thai, I will attempt to demonstrate that inputs must be abstract. Specifically, despite the presence of aspiration in the onset of stressed syllables in English, I will argue from the patterns of behaviour that emerge in the second language data that English cannot have the feature which formally marks aspiration present in inputs. A more general goal of the paper is to draw attention to the issues that the data under investigation raise concerning abstractness, in the context of current thinking in phonology.

2. Outline of the issues

Most of the empirical generalizations discussed here come from earlier collaborative work with Suzanne Curtin and Joe Pater (Curtin, Goad, and Pater 1998). Curtin, Goad, and Pater report on an experiment where English- and French-speaking subjects were taught Thai words which exploit the three-way laryngeal contrast found in this language. To provide a context for the issues to be discussed, I begin by briefly presenting the principal finding of Curtin, Goad, and Pater. When anglophones were tested using a methodology that taps lexical representations (Minimal Pair Identification task[2]), they performed significantly better on the Voiced-Plain contrast than on Plain-Aspirated. In fact, their performance on Plain-Aspirated was poor enough to suggest that this contrast is funnelled into a single input representation, as schematized in (1) for labials.

(1) Minimal Pair task:

Stimuli:	[b]	[p]	[p^h]
Identified as:	/b/	/p/	

In research on second language acquisition, the generally-held view is that learners initially transfer properties from their native language grammar into the second language. Accordingly, Curtin, Goad, and Pater argue that the re-

2 In this task, subjects hear a word which is the correct name for one of three pictures. Names for two of the pictures form a minimal pair while the third is a foil. Subjects must select the picture which corresponds to the auditory stimulus (see § 3.2 for further details).

sults in (1) support the view that English speakers' inputs for Thai are underspecified for [spread glottis], the feature marking aspiration, defined as presence/(absence) of significant glottal width at the point of release of a stop. Inputs are only specified for what is contrastive in English, namely [voice], which indicates presence/(absence) of vocal cord vibration. If this is the correct interpretation of (1), it speaks against Lexicon Optimization: as voiceless stops in English are aspirated foot-initially, Lexicon Optimization will favour the input specification of [spread glottis] in this position (§ 4). It is also inconsistent with the view that inputs are phonetically-enriched; the latter would favour the inclusion in inputs of the set of phonetic properties which together mark aspiration. Finally, it is inconsistent with proposals which consider English to be a language in which [spread glottis] (or its equivalent) is underlyingly present and [voice] (or its equivalent) is not specified (e. g., Harris 1994, Iverson and Salmons 1995, Avery 1996).

While a logical conclusion to draw from (1) is that inputs are underspecified for [spread glottis], the validity of this interpretation is questioned when the additional results in (2) are considered; all appear to demonstrate a role for [spread glottis]:

(2) a. In the Minimal Pair task, English speakers performed significantly better on Aspirated-Voiced than on Plain-Voiced;

b. A subset of English speakers performed well on Aspirated-Plain late in the experiment;

c. Good results on Aspirated-Plain were obtained in the ABX task, in contrast to the Minimal Pair task;

d. Good results on Aspirated-Plain were obtained in Pater's (2003) replication of Curtin, Goad, and Pater using a methodology that taps lexical representations.

My goal will be to demonstrate that the position that inputs are unspecified for [spread glottis] can be upheld, in spite of the observations in (2).

3. Curtin, Goad, and Pater's experiment

3.1. Predictions

As mentioned in § 2, Thai has a three-way laryngeal contrast; both [voice] and [spread glottis] are distinctive. English and French only exhibit a two-way contrast, usually described as involving the feature [voice]. These languages differ, though, in that aspiration is absent from French but contextually present in English: voiceless stops are aspirated foot-initially ([rǽpəd]–[rəp^{h}ídəti] ‘rapid’–’rapidity’). In theories of generative phonology which assume that inputs only contain contrastive material and that [voice] is the relevant distinctive feature in English, voiceless stops are underlyingly represented as unaspirated, and [spread glottis] is supplied by rule. When considering adult English speakers who are attempting to learn the three-way contrast in Thai, this approach predicts that voicing should emerge first in the interlanguage grammar; as [voice] is stored in English inputs, it should be the laryngeal feature available for transfer. Accordingly, aspirated and plain stimuli, both of which are [-voice], should initially be funnelled into a single category in contrast to voiced stimuli.

This prediction appears to be challenged by findings from the speech perception literature. As schematized in (3) for labials, when anglophones are presented with synthesized Voice Onset Time correlates of the Thai Voiced-Plain and Plain-Aspirated contrasts, they identify stimuli whose Voice Onset Time values correspond to Thai plain [p] as ‘b’, not as ‘p’ (Abramson and Lisker 1970; replicated by Strange 1972, Pisoni et al. 1982, among others).

(3) English speakers’ identification of Voice Onset Time correlates of Thai [voice] and [spread glottis]:

Stimuli:	[b]	[p]	[p^{h}]
Identified as:	‘b’		‘p’

This finding is not surprising when the Voice Onset Time values obtained by Lisker and Abramson (1964) for Thai and English are compared. Table 1 reveals that English /b, d/ align most closely with Thai /p, t/, while /p, t/ align most closely with /p^{h}, t^{h}/.[3,4]

3 Note that Thai has no /g/; thus, the focus of the discussion throughout this paper is on labial and coronal stops only.

4 The English values for /b, d/ in Table 1 come from 114 tokens produced almost exclusively by three of the four speakers in Lisker and Abramson. The fourth speaker produced virtually all of his voiced stops with voicing lead and was responsible for 95 % of all cases of voicing lead in the sample.

Table 1. Voice Onset Time in msec.

Thai (3 speakers)	/b/	/p/	/pʰ/	/d/	/t/	/tʰ/
Average	-97	6	64	-78	9	65
Range	-165:-40	0:20	25:100	-165:-40	0:25	25:125
English (4 speakers)		/b/	/p/		/d/	/t/
Average		1	58		5	70
Range		0:5	20:120		0:25	30:105

The results in (3) demonstrate that English speakers can perceive aspiration more easily than voicing, at least in terms of Voice Onset Time. This may suggest that [spread glottis] (or the corresponding Voice Onset Time range) rather than [voice] is stored in inputs, as has recently been proposed by Harris (1994), Iverson and Salmons (1995), and Avery (1996), as mentioned above. Before underlying [voice] can be rejected, however, it is important to consider the type of methodology employed in the speech perception literature. These studies use phoneme identification and discrimination tasks which require that subjects distinguish minimally different sounds, either by labelling the sounds with orthographic symbols, or by indicating whether two sounds are the same or different. They do not require access to stored representations, as does the methodology employed by Curtin, Goad, and Pater (§ 3.2). Nevertheless, if the order of acquisition of stored contrasts in a second language correlates with relative perceptibility, then [spread glottis] should emerge first, contra the prediction of phonological approaches where English inputs only contain contrastive [voice].

3.2. Methodological concerns

In order to investigate the divergent predictions outlined above, Curtin, Goad, and Pater required that subjects learn 18 Thai words (6 Aspirated-Plain-Voiced minimal sets). The main indicator of subjects' discrimination abilities was considered to be a task that taps underlying representations, the Minimal Pair task described in (4a).

(4) a. Minimal Pair task:
Subjects hear a Thai word which is the correct label for one of three pictures displayed on a computer screen. Names for two of the pictures form a minimal pair; the third is a foil. Subjects press the key which corresponds to the correct picture.

b. ABX task:
A minimal pair AB is presented aurally, followed by a third word X that matches either A or B. Subjects press the key which indicates that X is most like A or most like B.

The ABX task described in (4b), which used exactly the same stimuli as the Minimal Pair task, was also designed to tap stored representations: the tokens for A, B and X were produced by different speakers, and the interstimulus interval between B and X was relatively long. However, the methodology does not necessitate access to stored representations, a point which will be returned to in § 5.3.

3.3. Minimal Pair results and interpretation

The results of Curtin, Goad, and Pater's Minimal Pair task are in Table 2. English and French speakers performed strikingly similarly on this task; indeed, in an Analysis of Variance examining contrast, language, and testing day, no effect was found for language, only for contrast.

Table 2. Proportion correct in Minimal Pair task

Testing Day	Aspirated-Plain		Plain-Voiced		Aspirated-Voiced	
	English	French	English	French	English	French
2	.59	.60	.82	.75	.93	.91
4	.63	.60	.77	.81	.95	.94
11	.68	.59	.82	.81	.95	.96

Concerning the latter, performance on Plain-Voiced was significantly better than Aspirated-Plain for both groups of learners. In fact, both groups discriminated Aspirated-Plain at only slightly better than chance.

The Minimal Pair results are not consistent with the speech perception literature which, recall, found better results on Voice Onset Time correlates of aspiration, not voice. Since correct responses on the Minimal Pair task must be made on the basis of stored representations, Curtin, Goad, and Pater maintain that the results reveal that [voice], not [spread glottis], is what English (and French) speakers transfer and thus initially represent when acquiring Thai.[5] These results sup-

5 Importantly, the Voice Onset Time values of the Thai stops in Curtin, Goad, and Pater are comparable to those of Lisker and Abramson (1964). The question that arises is what phonetic cue(s), other than Voice Onset Time, is particularly promi-

port the view that learners do not first acquire the contrast that is most perceptible but, instead, that which corresponds to what many generative phonologists treat as underlying, namely [voice]. Accordingly, English inputs are underspecified for [spread glottis], despite the presence of surface aspiration in this language. The consequences of this for Lexicon Optimization are discussed next.

4. Lexicon Optimization

As mentioned in § 1, Optimality Theory does not place any constraints on the shapes of inputs (what is referred to as Richness of the Base (Prince and Smolensky 1993/2004)). The burden of selecting correct outputs is placed entirely on ranking. The result is a potentially infinite set of inputs for a given output. Below, we will investigate how the learner selects appropriate input-output pairings, focussing on [spread glottis] in English.

4.1. (Under)specification of [spread glottis]

Two grammars capturing the distribution of aspiration in English are in (6).[6] The necessary constraints are first defined (informally) in (5).

(5) $_{\mathrm{Ft}}$[SG: Voiceless stops are enhanced by aspiration foot-initially
*SG: Stops are not aspirated
IDENT-IO(SG): Correspondent segments have identical values for [spread glottis]

---

nent in the Thai stimuli which leads speakers to group together plain and aspirated stops in contrast to voiced stops. When the Thai stimuli were examined for burst intensity (Goad 2000), it was found that the voiced stops have much bigger bursts than the plain and aspirated stops and so this is a likely candidate (average for labial and coronal voiced stops: .103 (RMS, expressed in Pascals); plain stops: .043; aspirated stops: .043). While speakers' sensitivity to burst intensity can account for Curtin, Goad, and Pater's Minimal Pair task results, it cannot account for their ABX results where good performance was observed on *both* Aspirated-Plain and Plain-Voiced (§ 5.3). This suggests that methodological considerations, rather than particulars of the stimuli employed, are responsible for Curtin, Goad, and Pater's Minimal Pair results. This question, however, clearly requires further examination.

6 The ranking *SG >> IDENT-IO(SG) is not evident from (6). It emerges when voiceless stops surface as plain. For example, to ensure that an input like /hæp^{h}i/ (permitted by Richness of the Base) surfaces as [hǽpi], *SG must be dominant.

(6)

Grammar 1: /pʰæt/	$_{\mathrm{Ft}}$[SG	*SG	IDENT (SG)
a. [pæt]	*!		*
☞ b. [pʰæt]		*	

Grammar 2: /pæt/	$_{\mathrm{Ft}}$[SG	*SG	IDENT (SG)
a. [pæt]	*!		
☞ b. [pʰæt]		*	*

How do learners select among alternative grammars like those in (6)? Following Smith (1973), the most commonly-held view in the literature on first language acquisition is that the child's input is equivalent to the adult's output (but cf. Macken 1980, Rice and Avery 1995, Brown and Matthews 1997), until evidence to the contrary is encountered. This will lead to the child selecting Grammar 1 at Stage 1. Concerning later developmental stages, exactly what constitutes evidence to the contrary depends on the theory adopted: absence of contrast or absence of alternations. In underspecification theory, the former is (explicitly or implicitly) relevant: inputs only contain contrastive features. As aspiration does not have this status in English, it will be underlyingly unspecified, leading to selection of Grammar 2.

In Optimality Theory, by contrast, Lexicon Optimization typically steers learners toward inputs that are not underspecified: in the absence of alternations, it reconciles learners to the input-output pairing where faithfulness is maximally respected (Prince and Smolensky 1993/2004, Inkelas 1994, Itô, Mester, and Padgett 1995). In the presence of alternations, inputs may be underspecified, but only in those contexts where the alternations are observed. Since voiceless stops are aspirated foot-initially in English, Lexicon Optimization favours the specification of [spread glottis] in this position in non-alternating forms like 'pat', leading to the selection of Grammar 1. (For alternating forms like 'rapid'–'rapidity', Grammar 2 will be selected.)

The laryngeal contrasts in Curtin, Goad, and Pater's Thai stimuli were in word- and foot-initial position (and displayed no alternations). Accordingly, the presence/absence of input [spread glottis] in this position should transfer to the English learners' grammar of Thai. If [spread glottis] is specified as per Lexicon Optimization (Grammar 1), the Thai plain-aspirated contrast should be perceptible to English speakers. If [spread glottis] is underspecified (Grammar 2), English speakers should collapse plain and aspirated stimuli into a single category. Only the latter correctly predicts the asymmetry observed by Curtin, Goad, and Pater in (1): plain and aspirated stimuli are perceived as the same by anglophones, in contrast to voiced stimuli, contra the predictions of Lexicon Optimization. We attempt to resolve this problem below.

4.2. Selecting underspecified inputs

Thus far, we have discussed how the finding in (1) reveals that anglophones cannot have [spread glottis] present in inputs. Since it has already been observed that both grammars in (6), where inputs do and do not contain [spread glottis] respectively, will select the correct output in production, the challenge is for the ranking in (6) to lead to the removal of [spread glottis] in perception, appropriately resulting in underspecified inputs.

Figure 1. shows the connection between perception and production within a single grammar, as envisaged here.

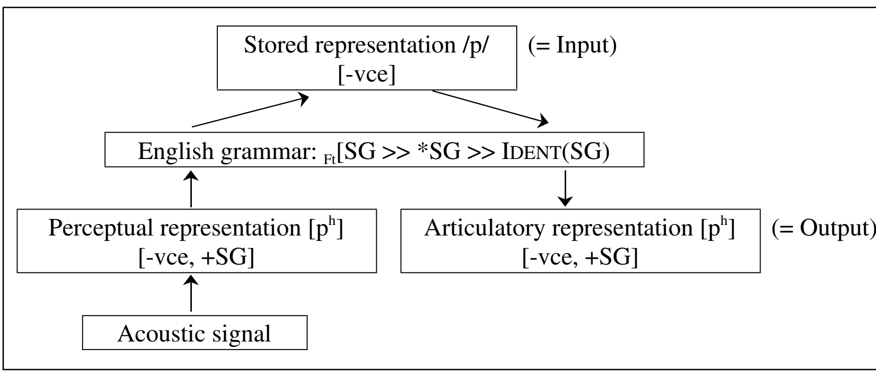


Figure 1. Perception and production in a single grammar

Focusing on perception, the processor must extract from the acoustic signal the correlates of [-voice] and [+spread glottis] which are part of the perceptual representation (Output) for [p^h]. When this form is passed up through the grammar, aspiration must be removed from [p^h], on its way to being mapped to the abstract form (Input) /p/.

I suggest that removal of aspiration occurs because of the type of constraint responsible for the presence of [spread glottis] in English. Since aspiration is contextually-determined in this language, a position-sensitive constraint, $_{Ft}$[SG, outranks *SG. Importantly, the context where [spread glottis] surfaces in English is prosodically- rather than morphologically-determined. If inputs are not prosodified, as is standardly assumed,[7] then $_{Ft}$[SG will have no impact on the shapes of inputs. Only *SG, the next constraint in the ranking, will play

7 While this is the standard position, it is counter to what is argued for in Goad and Rose (2004); at this point, I do not know how to resolve this.

a role, thereby resulting in the removal of [spread glottis] from inputs, the desired result.

5. Evidence from Curtin, Goad, and Pater that aspiration is specified in inputs?

We have just seen that, by considering the type of markedness constraint involved, it is possible to select as optimal inputs which are unspecified for [spread glottis] even when outputs are uniformly aspirated. The approach was motivated by the principal finding from Curtin, Goad, and Pater from which it was concluded that English speakers (learners of Thai) cannot have [spread glottis] present underlyingly. Recall from § 2, however, that there are additional results, in (2), which may lead us to question this conclusion: all of them appear to demonstrate a role for [spread glottis] in the English grammar. In the following sections, I return to these results, addressing for each whether [spread glottis] must be posited in inputs. I begin with (2a), performance on Aspirated-Voiced in the Minimal Pair task.

5.1. Aspirated-Voiced condition

Recall from § 3.3 that in the Minimal Pair task, performance on Plain-Voiced was significantly better than Aspirated-Plain for both groups of learners. At that point, there was no discussion of Aspirated-Voiced; however, Table 2 reveals that performance on this contrast is near ceiling. Indeed, Aspirated-Voiced vs. Aspirated-Plain reaches a higher level of significance than Plain-Voiced vs. Aspirated-Plain. Curtin, Goad, and Pater attribute this to the observation that aspirated stops cue the voiced-voiceless contrast better than plain voiceless stops. While they specifically say that this does not indicate that both [voice] and [spread glottis] are present underlyingly, they do not address the following problem: if aspirated stops signal the voicing contrast better than plain stops, how can this information be accessible to learners if inputs, the level targeted in the Minimal Pair task, have no access to [spread glottis] (as, for example, in the model in Figure 1)?

Expressed differently, does ceiling performance on Aspirated-Voiced force [spread glottis] to be present in English inputs? The source of the answer to this lies in the performance of the francophones on the Minimal Pair task. Table 2 shows that the francophones do as well as the anglophones on Aspirated-Voiced. As [spread glottis] plays no role in the French grammar, the question cannot be reduced to the status of [spread glottis] – as allophonic – in the English grammar. Accordingly, the issue does not concern Lexicon Optimization, determining whether [spread glottis] is present in English inputs and, thus, in the transferred grammar that English speakers build for Thai. Instead, if performance on Aspirated-Voiced leads to the input specification of [spread glottis] in English, it must be present in French as well. The question thus concerns whether or not inputs are phonetically enriched. If they are, aspiration would better cue the voicing contrast because the acoustic correlates of [spread glottis] present in the signal become part of the input, independent of the language.

The numbers in Table 2 clearly reflect the fact that there is gradience in the acoustic signal, simplified somewhat, on the Voice Onset Time dimension. The gradience must map onto a set of formal objects (features), but what do these features look like? For present purposes, I will consider the two options in Figure 2. In (a), the signal is gradient, but phonological features are binary, because perception is deemed to be categorical.[8] In (b), features ([Voice Onset Time] and others) are gradient, because perception is deemed to be continuous.

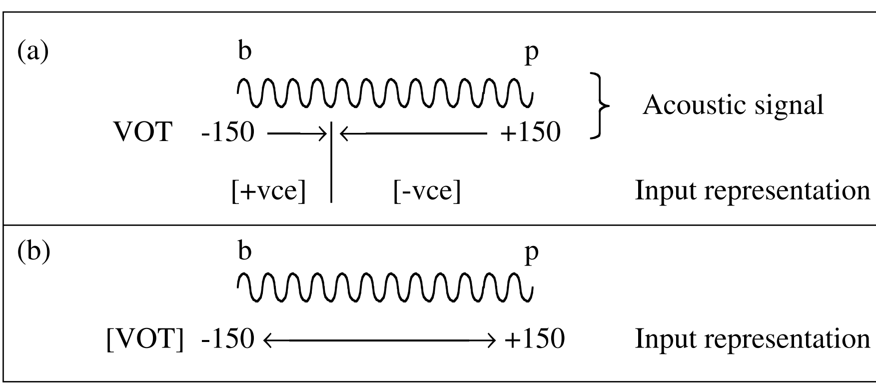


Figure 2. Input representations, using (a) binary and (b) gradient features

Given the findings from the Minimal Pair task – that Aspirated-Voiced vs. Aspirated-Plain reaches a higher level of significance than Plain-Voiced vs.

8 In (a), the cross-over point between "b" and "p" is given as -30, following Abramson and Lisker's (1970) results for Thai speakers. This is somewhat misleading, as their results were arrived at through phoneme discrimination and identification tasks (§ 3.1). As we are discussing underlying representations for English speakers, the appropriate boundary should be determined using tasks that tap inputs.

Aspirated-Plain – we might be tempted to conclude that perception is continuous and must be reflected in the grammar as in (b) in Figure 2. To assess this, we turn briefly to consider the research on Categorical Perception. Repp (1984: 251–252) defines Categorical Perception as "the experience of discontinuity as a continuously changing series of stimuli crosses a category boundary, together with the absence of clearly perceived changes within a category". In the perception of speech, this research has looked at the perceptual reality of discrete segments which (more or less) correspond to phonemes.

Concerning voicing in stops, Categorical Perception effects are particularly robust. While one might thus be tempted to conclude that (a) in Figure 2 is correct, there is also a large literature which has found that perception can be continuous (see Repp 1984 for a review). This work has focussed on determining the experimental conditions that can be manipulated to lead to either categorical or continuous perception. Does this research argue against perception as categorical and thus in favour of (b) in Figure 2? The answer, I believe, is no. What it does show is that while Categorical Perception effects are widely observed, the strongest version of the Categorical Perception hypothesis cannot be maintained, as there are experimental conditions under which listeners can discriminate within-category differences.

At this point, one might conclude that a decision between (a) and (b) in Figure 2 cannot be made. It is not obvious, however, how (b) would predict Categorical Perception effects at all, whereas (a) does allow for diversions from Categorical Perception. To explore how (a) permits such diversions, we turn to consider the different processing levels proposed by Werker and Logan (1985). Werker and Logan demonstrate that listeners can exploit different processing strategies, depending on experimental conditions, especially interstimulus interval, and also practice gained during the experiment itself. See (7):

(7) a. Phonemic: Stimuli perceived according to native language phonemic categories;
b. Phonetic: Sub-phonemic information perceived;
c. Acoustic: Finer acoustic detail between stimuli perceived.

Let us consider (7a-b) in the context of Figure 1 above. Phonemic processing will only access what is available in the stored representation; phonetic processing will access non-contrastive information as well, available in the perceptual representation. The essential point, then, is that while experiments can be designed to tap different levels of representation, stimuli are funnelled into native phonemic categories, once the information available in the phonetic code has decayed. Accordingly, there must be a level of representation that reflects the type of information that is perceived under such conditions – the Input in Figure 1.

Since Curtin, Goad, and Pater's Minimal Pair task requires access to inputs, it must involve phonemic processing. The results should therefore support the Categorical Perception hypothesis, (a) in Figure 2. I believe that they do. Recall from Table 2 that Aspirated-Plain was discriminated only slightly better than chance. This indicates that these stimuli form one category, [-voice]; however, some members of this category, the aspirates, are better instances of [-voice] than other members, resulting in ceiling performance on Aspirated-Voiced. In short, while some types of information in the acoustic signal (the phonetic correlates of aspiration) cue the voiced-voiceless contrast particularly well, poor performance on Aspirated-Plain strongly suggests that this information is not encoded in inputs.

5.2. Performance on Day 11

In this section, we turn to examine the performance on Aspirated-Plain at Day 11 where some improvement is observed among the anglophones (see (2b)).[9] The overarching question, as before, is whether these results demonstrate a role for [spread glottis] in inputs.

One question posed by Curtin, Goad, and Pater is whether surface aspiration in English has any positive effect on speakers' ability to underlyingly represent this feature. Recall from § 3.3 that an Analysis of Variance did not find the improvement on aspiration observed at Day 11 to be significant. To further explore the issue of whether the improvement reflected genuine development, Curtin, Goad, and Pater subtracted the participants' Day 2 scores from their Day 11 scores and subjected the scores to a Kolmogorov-Smirnov test. The difference between the anglophones and francophones was significant. However, development was only observed for three anglophones: they showed an average improvement of 24 %; the remaining five showed no improvement overall. A Kolmogorov-Smirnov test considered these two groups of anglophones to be significantly different.

Do these results suggest that [spread glottis] is present in native English inputs? The answer, I argue, is no. First, the presence of [spread glottis] – as mandated by Lexicon Optimization – cannot account for the observation that on Days 2 and 4, the anglophones only performed slightly above chance on

9 Day 11 is one week after no exposure to Thai.

Aspirated-Plain. Second, their performance on Days 2 and 4 is the same as the francophones who do not have [spread glottis] in their grammar. Finally, as just mentioned, the improvement at Day 11 is only observed for a subset of anglophones.

The presence of surface aspiration in English can have an effect on speakers' ability to *eventually* store this feature in their second language inputs. Indeed, the findings for Day 11 suggest that [spread glottis] has truly been phonologized in the grammars of the anglophone individuals involved. However, the presence of surface aspiration cannot, I suggest, have an effect at the outset of acquisition. The developmental scenario for second language acquisition is outlined in Figure 3. Stage 1 (Days 2 and 4) represents the transferred English grammar. The feature [spread glottis] has the same status as in the native English grammar: it is absent from inputs. Because of the two-way contrast in voicing in the transferred grammar, Thai [p] and [p^h] are mapped to a single category /p/. It is hypothesized that outputs will show aspiration for target [p] and [p^h], due to high-ranking $_{\mathrm{Ft}}$[SG (production was not tested). Stage 2 reflects the development exhibited by the three anglophones (Day 11). The three-way contrast is now perceptible as reflected by the demotion of *SG below the faithfulness constraint IDENT(SG). Without demotion of $_{\mathrm{Ft}}$[SG, production outputs are, for all intents and purposes, unaffected. This developmental path, that production lags behind perception, is commonly observed in first language development.

Three principal claims are being made here. One, development over time in Optimality Theory involves the elaboration of inputs (Goad and Rose 2004), not just constraint reranking. Two, the lexicalization of new features can only occur over time. Indeed, there were no English speakers in the Curtin, Goad, and Pater study who were able to perceive the Thai three-way voicing contrast from the outset. Three, there is a relationship between the presence of allophonic aspiration in the native language and the ability to lexicalize this feature in the second language. This is in the spirit of Brown (1998) but represents a weakening of her proposal. Brown hypothesizes that beyond the transfer stage, only features which are contrastive in the native language grammar can be combined to build new segments in a second language. This proposal is being extended here to include non-contrastive features.

As [spread glottis] has no status in French, the predictions made for this population of speakers are the same as Brown: the Thai three-way contrast should never be lexicalized. That is, in Figure 1, [spread glottis] will never be mapped from the acoustic signal into the perceptual representation. Whether or not this prediction can be upheld remains to be investigated.

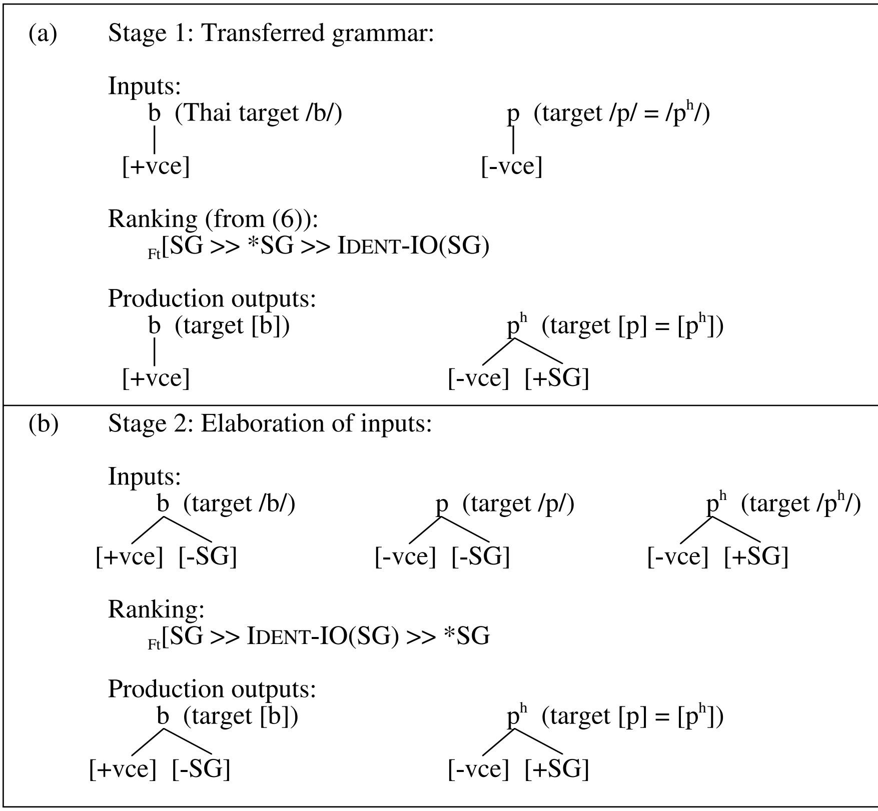


Figure 3. Stages in development

5.3. Curtin, Goad, and Pater's ABX results

In this section, I address the third issue concerning the role of [spread glottis] in the grammar transferred from English to Thai, that better results on [spread glottis] were obtained on Curtin, Goad, and Pater's ABX task than on their Minimal Pair task (see (2c)). Compare Table 3 with Table 2 from § 3.3.

What is striking about these results, when compared with the Minimal Pair results, is that there are differences across languages in the Aspirated-Plain and Plain-Voiced conditions: francophones performed better on Plain-Voiced than on Aspirated-Plain, as they did in the Minimal Pair task, but anglophones performed similarly on these two contrasts, unlike in the Minimal Pair task. Aspirated-Plain vs. Plain-Voiced was significant for the francophones only; thus, while the numbers in Table 3 may suggest that the anglophones are performing better on Aspirated-Plain than on Plain-Voiced, this is not significant.

Table 3. Proportion correct in ABX task

Testing Day	Aspirated-Plain		Plain-Voiced		Aspirated-Voiced	
	English	French	English	French	English	French
2	.84	.64	.83	.78	.99	.96
4	.77	.67	.73	.77	.99	.98
11	.79	.59	.70	.88	.88	.98

Why do the anglophones perform better on Plain-Voiced than on Aspirated-Plain in the Minimal Pair task, but not in the ABX? And while the ABX was designed to tap inputs, performance on Aspirated-Plain is much better than expected if [spread glottis] is not underlyingly specified; does this finding suggest that [spread glottis] is present in inputs?

Although the ABX task was designed to tap inputs, the methodology does not *require* lexical access, as subjects are presented with auditory stimuli only; thus, judgements can be based on phonetic similarity alone. In Curtin, Goad, and Pater, we suggested that the results on this task were due to subjects sometimes relying on their lexical representations ([±voice]) and sometimes on surface representations ([±spread glottis]). Given the position-sensitive nature of voicing and aspiration in English, we did not consider the possibility that tapping surface representations could result in a *three*-way distinction. That is, we did not consider the possibility that speakers might process stimuli in the ABX at Werker and Logan's (1985) phonetic level (7b), where within-category decisions can be made. The means in Table 4 suggest perception of a three-way contrast: performance on *both* Aspirated-Plain and Plain-Voiced in the ABX is as good as performance on Plain-Voiced in the Minimal Pair task.

Table 4. Anglophone means in ABX and Minimal Pair tasks

	ABX		Minimal Pair	
Aspirated-Plain	.80	not significant	.60	significant
Plain-Voiced	.75		.80	

Two questions arise at this point: (i) Are Curtin, Goad, and Pater correct in concluding that the ABX is sometimes tapping lexical representations and sometimes surface representations? (ii) Do the ABX results suggest that [spread glottis] is present in English inputs? I believe that the answer to both questions is no. Concerning (i), the ABX methodology is not well-suited to eliciting phonemic judgements; it favours within-category processing, even when the experiment is designed to elicit cross-category judgements (Werker and Logan 1985, Brannen 2002). In short, the ABX methodology enables listeners to perceive the three-way Aspirated-Plain-Voiced distinction. Following from this, concerning question (ii), the results do not indicate that [spread glottis] is present in English inputs: as we have just suggested, this task is not tapping inputs.

6. Evidence from Pater's replication that aspiration is specified in inputs?

Thus far, three potential sources of evidence for the input specification of [spread glottis] in English have been examined from the results obtained by Curtin, Goad, and Pater. It has been argued for each that, counter to appearance, [spread glottis] is not present in inputs. In this section, we turn finally to Pater's (2003) replication of Curtin, Goad, and Pater which found better results for Aspirated-Plain than Plain-Voiced on a task that taps lexical representations (see (2d)). These results appear to require that [spread glottis] be specified in English inputs, contra the conclusion reached so far.

In Curtin, Goad, and Pater's study, the Minimal Pair and ABX tasks were methodologically quite different from each other. Pater attempted to rectify this by modifying the methodology as in (8). (All subjects were anglophones; stimuli were the same as in Curtin, Goad, and Pater.)

(8) XAB discrimination tasks (Pater 2003):
a. Sound-Sound-Sound
b. Picture-Sound-Sound
c. Sound-Picture-Picture

Sound-Sound-Sound is most like Curtin, Goad, and Pater's ABX task, while Sound-Picture-Picture is most like their Minimal Pair task. Picture-Sound-Sound and Sound-Picture-Picture both require lexical access.

The results, averaged across subjects, are in Table 5.

Table 5. Means in Pater's XAB tasks

	Sound-Sound-Sound	Picture-Sound-Sound	Sound-Picture-Picture
Aspirated-Plain	.84	.83	.52
Plain-Voiced	.71	.72	.53

The most conspicuous result is that subjects performed only at chance on Sound-Picture-Picture. Pater is puzzled by this and thus excludes the task from further discussion; I return to this below. Second, performance is the same on both Sound-Sound-Sound and Picture-Sound-Sound, even though only the latter requires lexical access. Finally, Aspirated-Plain is significantly better than Plain-Voiced on both Sound-Sound-Sound and Picture-Sound-Sound.

A comparison of Tables 4 and 5 reveals two striking differences between Pater's and Curtin, Goad, and Pater's results. First, Pater's Picture-Sound-Sound most closely parallels Curtin, Goad, and Pater's ABX results; better performance is observed on Aspirated-Plain. As Picture-Sound-Sound requires access to inputs, we must consider whether Pater's results indicate that [spread glottis] is stored. Second, neither of Pater's tasks which require lexical access, Picture-Sound-Sound and Sound-Picture-Picture, mirror the results of Curtin, Goad, and Pater's Minimal Pair task – better performance on Plain-Voiced than on Aspirated-Plain which Curtin, Goad, and Pater use to argue against input [spread glottis].

In the following lines, I suggest that these differences arise from methodological considerations, that Pater's study is not a true replication of Curtin, Goad, and Pater. I hypothesize further that the Sound-Picture-Picture results indicate that [spread glottis] is not stored in inputs, at least not in the compositional way that native speakers store features (see below).

I begin with the duration of the experiment. Pater mentions that subjects were trained one day and tested the next. In Curtin, Goad, and Pater, subjects were similarly tested for the first time on Day 2. However, Curtin, Goad, and Pater also included a pre-test (Day 0) where subjects were tested on 18 different Thai stimuli. Although subjects were not taught the meanings of these words, they were given positive feedback on discrimination tasks. This additional exposure to Thai may have helped learners establish native-like representations for these segments.

In this context, one must question whether the subjects in Pater's experiment had enough opportunity to truly learn the words – to store them using the same set of primitives available to end-state grammars. In Sound-Picture-Picture, where performance was at chance, Pater mentions that on the foils, subjects performed near ceiling; accordingly, he concludes that they did learn the words. However, there are several cues to distinguish foils from test items; the former differed from the latter in the initial consonant's place of articulation and for at least one segment in the rhyme (all stimuli were Consonant-Vowel-Consonant in shape).

While excellent performance on the foils reveals that they are stored differently from the test stimuli, it does not tell us *how* the various stimuli are stored. We turn to this issue now. In the acquisition literature, a distinction is commonly drawn between holistic and analytic learning (e. g., Cruttenden 1981, Peters 1983 on first language acquisition; Wray 2002 on second language acquisition). An important theme that emerges from this literature is that holistic learning is common at the earliest stages in acquisition. This observation is extended to the present context as follows: at the immediate onset of perception in a second language, transfer is not yet a consideration, as stimuli are stored in holistic rather than analytic form. That is, as much information as can be extracted from the acoustic signal is stored, but this information is not yet mapped to a set of formal objects of analysis (features).

Two results suggest that a holistic, rather than compositional, analysis has been undertaken by the subjects in Pater's experiment. First, recall that performance on Sound-Picture-Picture is at chance. If subjects have not yet undertaken a featural analysis of the stimuli, good performance will require a comparison of at least two auditory stimuli, rather than an assessment based on a single stimulus as in Sound-Picture-Picture. Contrastingly, if subjects have had time to analyse and store the stimuli featurally, such a comparison will not be necessary: in Curtin, Goad, and Pater's Minimal Pair task, subjects were presented with a single auditory stimulus, and performance on one contrast, Plain-Voiced, was significantly better than chance. It must still be explained why, on the holistic view, Aspirated-Plain is perceived more accurately than Plain-Voiced on Pater's Sound-Sound-Sound and Picture-Sound-Sound tasks. This, I believe, follows from the observation that, ceteris paribus, [spread glottis] is more perceptible than [voice] (§ 3.1).[10]

The second result which suggests that the subjects in Pater's experiment have undertaken a holistic analysis is that there is a strong effect for place. Table 6 shows that, for both Sound-Sound-Sound and Picture-Sound-Sound, discrimination of aspiration is better for labials than for alveolars, while discrimination of voice is better for alveolars than for labials (both are significant).

Table 6. Means in Pater's tasks by place

	Sound-Sound-Sound		Picture-Sound-Sound	
	Labial	Alveolar	Labial	Alveolar
Plain-Voiced	.63	.78	.62	.81
Aspirated-Plain	.90	.78	.90	.77

10 As an anonymous reviewer points out, it could also be that the plain stops are being perceived as voiced in these tasks. Indeed, this could perhaps lead to an explanation of why performance on Aspirated-Plain versus Plain-Voiced in Pater's Sound-Sound-Sound and Picture-Sound-Sound tasks was significant, while performance on these same contrasts in Curtin, Goad, and Pater's ABX task was not.

If speakers have done an abstract featural analysis and display phonemic processing, place effects should not be found. Since phonemic processing accesses representations at the level of contrast, these representations will contain features for place, features for voicing, and their combinatorial possibilities, but differences in degree of voicing which are sensitive to place of articulation will not be accessible.[11] Place effects should only be present under phonetic processing which accesses non-contrastive information available in the perceptual representation, or under a holistic analysis where small differences in degree of voicing observed for different places of articulation will be stored.

If this approach is correct, place effects should be present in Curtin, Goad, and Pater's ABX task but not in their Minimal Pair task. To investigate this, we turn to Curtin (1997). Curtin observed place effects in the data collected by Curtin, Goad, and Pater, but they were largely dependent on task.

Table 7. Means in Curtin, Goad, and Pater's tasks by place

	ABX				Minimal Pair			
	English		French		English		French	
	Labial	Alveolar	Labial	Alveolar	Labial	Alveolar	Labial	Alveolar
Plain-Voiced	.64	.86	.69	.92	.75	.87	.75	.84
Aspirated-Plain	.93	.67	.73	.54	.65	.63	.60	.58

Table 7 reveals that, as expected, in the ABX, place effects were robust: performance on Aspirated-Plain was significantly better for labials than for alveolars for both language groups; Plain-Voiced exhibited the opposite pattern, with significantly better performance for alveolars. As expected, in the Minimal Pair task, labial does not enhance the perception of Aspirated-Plain, in contrast to the ABX. Unexpectedly, though, there were place effects for alveolars, with both groups performing significantly better on alveolar in the Plain-Voiced condition. Importantly, however, in contrast to the ABX, no particular place enhances the perception of Aspirated-Plain; this is consistent with the proposal that [spread glottis] is not present underlyingly. In short, the results for place are in the right direction: place effects are stronger in the ABX than

11 As an anonymous reviewer points out, this position may be too strong; for example, aspiration is much more salient in velars than in labials. If this perceptual effect has *phonological* consequences, then my position will have to be weakened. One possible phonological consequence would involve a language where /k/ has been singled out for spirantization, if this type of process arises from one noise source (burst) being misperceived as another (turbulence).

in the Minimal Pair task; and perception of Aspirated-Plain is not enhanced by place in the latter.

7. Conclusion

In this paper, I have argued that inputs are abstract and, thus, that the phonology (i. e., stored representations) does not necessarily align with the phonetics. Following from this, once there has been sufficient exposure to a second language, learners' inputs will show effects of transfer where their inputs are shaped by what is stored in the first language grammar. In the present case, inputs for English learners of Thai are specified for [voice] only, not [spread glottis], as revealed by Curtin, Goad, and Pater's Minimal Pair task.

Three sources of evidence which challenge the view that [spread glottis] is absent from English inputs were examined from Curtin, Goad, and Pater's results; for each, it was argued that, counter to appearance, [spread glottis] is not underlingly specified. First, although in the Minimal Pair task, Aspirated-Voiced was perceived better than Plain-Voiced, it was argued that this reflects gradience in the acoustic signal, where this gradience maps onto abstractly-represented features, leading to categorical perception effects.

Second, concerning the acquisition of non-contrastive features like [spread glottis], it was argued that the presence of such features in native language outputs can aid in their eventual lexicalization in a second language. However, the lexicalization of such features can only be observed at non-initial stages in acquisition, consistent with [spread glottis] being absent from transferred English inputs.

Third, good performance on pairs of stimuli involving features which are not contrastive can be observed under certain experimental conditions, but this does not lead to the conclusion that such features must be stored. Specifically, better performance for anglophones on Aspirated-Plain on Curtin, Goad, and Pater's ABX task than on the Minimal Pair task does not indicate that [spread glottis] is stored. The ABX task involves phonetic processing where within-category effects are expected, leading to across-the-board good performance.

Similar across-the-board good performance was argued to have been observed for learners who have stored stimuli as featurally-unanalysed, in tasks that involve a comparison between at least two auditory stimuli, as in Pater's Picture-Sound-Sound and Sound-Sound-Sound. For such learners, it was argued to follow that poor performance will be observed on tasks where subjects are exposed to one auditory stimulus only, as in Pater's Sound-Picture-Picture.

Additional stimulus effects, such as an interaction between voicing and place, were argued to be expected when stimuli are stored in holistic fashion.

A remaining question that has been left largely unaddressed concerns the weighting of the phonetic cues to onset voicing present in the Thai stimuli. If the hypothesis advanced in this paper proves to be correct, that representations are abstract, it is still of course the case that some cue or cues must have led to the profile of results obtained, notably that the anglophones in Curtin, Goad, and Pater's Minimal Pair task group together plain and aspirated stops in contrast to voiced stops. Although burst intensity leads to the right results for this task (see note 4), this is not the case for Curtin, Goad, and Pater's ABX task nor for any of Pater's tasks, where markedly different results were found. How do methodological considerations interact with the particulars of the stimuli employed to lead to the various different patterns of behaviour obtained? I leave this question to future research.

References

Abramson, Arthur and Leigh Lisker
1970 Discriminability along the voicing continuum. In: Bohuslav Hala, Milan Romportl and Premysl Janota (eds.), *Proceedings of the Sixth International Congress of Phonetic Sciences,* 569–573. Prague: Academia.

Anderson, Stephen
1985 *Phonology in the Twentieth Century.* Chicago: University of Chicago Press.

Avery, J. Peter
1996 The representation of voicing contrasts. Ph. D. dissertation, Department of Linguistics, University of Toronto.

Boersma, Paul
1998 *Functional Phonology.* The Hague: Holland Academic Graphics.

Brannen, Kathleen
2002 The role of perception in differential substitution. *Canadian Journal of Linguistics* 47: 1–46.

Brown, Cynthia
1998 The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research* 14: 136–193.

Brown, Cynthia and John Matthews
1997 The role of feature geometry in the development of phonemic contrasts. In: S. J. Hannahs and Martha Young-Scholten (eds.), *Focus on Phonological Acquisition*, 67–112. Amsterdam: John Benjamins.

Chomsky, Noam and Morris Halle
1968 *The Sound Pattern of English.* New York: Harper & Row.

Cruttenden, Alan
1981 Item-learning and system-learning. *Journal of Psycholinguistic Research* 10: 79–88.

Curtin, Suzanne
1997 The effects of place on the perception of voicing and aspiration contrasts. Ms., Department of Linguistics, University of Southern California.

Curtin, Suzanne
2002 Representational richness in phonological development. Ph. D. dissertation, Department of Linguistics, University of Southern California.

Curtin, Suzanne, Heather Goad and Joe Pater
1998 Phonological transfer and levels of representation: The perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research* 14: 389–405.

Flemming, Edward
2001 Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18: 7–44.

Goad, Heather
2000 The acquisition of voicing and aspiration contrasts by English and French learners of Thai: What's wrong with VOT? Paper presented at Department of Phonetics and Linguistics, University College London, March.

Goad, Heather and Yvan Rose
2004 Input elaboration, head faithfulness and evidence for representation in the acquisition of left-edge clusters in West Germanic. In: René Kager, Joe Pater and Wim Zonneveld (eds.), *Constraints in Phonological Acquisition*, 109–157. Cambridge: Cambridge University Press.

Harris, John
1994 *English Sound Structure*. Oxford: Blackwell.

Inkelas, Sharon
1994 The consequences of optimization for underspecification. Ms., Department of Linguistics, University of California, Berkeley.

Itô, Junko, R. Armin Mester and Jaye Padgett
1995 Licensing and underspecification in Optimality Theory. *Linguistic Inquiry* 26: 571–613.

Iverson, Gregory and Joseph Salmons
1995 Aspiration and laryngeal representation in Germanic. *Phonology* 12: 369–396.

Lisker, Leigh and Arthur Abramson
1964 A cross-language study of voicing in initial stops. *Word* 20: 384–422.

Macken, Marlys
1980 The child's lexical representation: The 'puzzle-puddle-pickle' evidence. *Journal of Linguistics* 16: 1–17.

Pater, Joe
2003 The perceptual acquisition of Thai phonology by English speakers: Task and stimulus effects. *Second Language Research* 19: 209–223.

Peters, Ann
1983 *The Units of Language Acquisition.* Cambridge: Cambridge University Press.

Pisoni, David, Richard Aslin, Alan Perey and Beth Hennessy
1982 Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance* 8: 297–314.

Prince, Alan and Paul Smolensky
1993 *Optimality Theory: Constraint Interaction in Generative Grammar.* Ms., Rutgers University and University of Colorado. Published Oxford: Blackwell [2004].

Repp, Bruno H.
1984 Categorical perception: Issues, methods, findings. In: Norman J. Lass (ed.), *Speech and Language: Advances in Basic Research and Practice,* Volume 10, 243–335. New York: Academic.

Rice, Keren and J. Peter Avery
1995 Variability in a deterministic model of language acquisition. In: John Archibald (ed.), *Phonological Acquisition and Phonological Theory,* 23–42. Hillsdale, NJ: Erlbaum.

Smith, Neil
1973 *The Acquisition of Phonology.* Cambridge: Cambridge University Press.

Steriade, Donca
2000 Paradigm uniformity and the phonetics-phonology boundary. In: Michael Broe and Janet Pierrehumbert (eds.), *Papers in Laboratory Phonology V,* 313–334. Cambridge: Cambridge University Press.

Strange, Winifred
1972 The effects of training on the perception of synthetic speech sounds: Voice Onset Time. Ph. D. dissertation, Department of Speech-Language-Hearing Sciences, University of Minnesota.

Werker, Janet and John Logan
1985 Cross-language evidence for three factors in speech perception. *Perception and Psychophysics* 37: 35–44.

Wray, Alison
2002 *Formulaic Language and the Lexicon.* Cambridge: Cambridge University Press.

Language Index

Subject Index