The Ohio State University

Working Papers in Linguistics No. 55

Studies on the Interplay of Speech Perception and Phonology

Edited by
Elizabeth Hume
Keith Johnson

The Ohio State University
Department of Linguistics

222 Oxley Hall
1712 Neil Avenue
Columbus, Ohio 43210-1298 USA
lingadm@ling.ohio-state.edu

Winter 2001

The Ohio State University
WORKING PAPERS IN LINGUISTICS

Working Papers in Linguistics is an occasional publication of the Department of Linguistics of Ohio State University containing articles by members (students and faculty) of the department. The first issue was published in 1967 and since then one to three issues have been published each year. To subscribe to the OSU WPL series, or order a back issue, please see our web page at:

http://ling.ohio-state.edu/osu_wpl

or write to:

OSU WPL
Department of Linguistics
222 Oxley Hall, 1712 Neil Ave.
Columbus, OH 43210-1298

lingadm@ling.ohio-state.edu

Back issues of OSU-WPL available as of March 2001

54. (Autumn 2000): ($15.00) ***Varia***. Edited by Jennifer S. Muller, Tsan Huang, and Craige Roberts.
53. (Summer 2000): ($15.00) ***Varia/Historical Linguistics***. Edited by Amanda Miller-Ockhuizen, Robert Levine, and Anthony J. Gonsalves.
52. (Summer 1999): ($15.00) ***Historical Linguistics***. Edited by Brian Joseph.
51. (Summer 1998): ($15.00) ***Varia***. Edited by Mary Bradshaw, David Odden and Derek Wyckoff.
50. (Spring 1997): ($15.00) ***Papers from the Linguistics Laboratory***. Edited by Kim Ainsworth-Darnell and Mariapaola D'Imperio.
49. (Spring 1996): ($15.00) ***Papers in Semantics***. Edited by Jae-Hak Yoon and Andreas Kathol.
48. (Spring 1996): ($15.00) ***Papers in Phonology***. Edited by David Dowty, Rebecca Herman, Elizabeth Hume, and Panayiotis A. Pappas.
47. (Autumn 1995): ($12.00) ***Varia***. Edited by David Dowty, Rebecca Herman, Elizabeth Hume, and Panayiotis A. Pappas.
46. (October 1995): ($12.00) ***Studies in Synchronic and Diachronic Variation***. Edited by Elizabeth Hume, Robert Levine and Halyna Sydorenko.

45. (February 1995): ($15.00) ***Papers from the Linguistics Laboratory***. Edited by Stefanie Jannedy.
44. (April 1994): ($15.00) ***Papers from the Linguistics Laboratory***. Edited by Jennifer J. Venditti.
43. (January 1994): ($12.00) ***Papers from the Linguistics Laboratory***. Edited by Sook-hyang Lee and Sun-Ah Jun.
42. (September 1993): ($15.00) ***Papers in Syntax***. Edited by Andreas Kathol and Carl Pollard.
41. (December 1992): ($12.00) ***Papers in Phonology***. Edited by Elizabeth Hume.
40. (July, 1992): ($15.00) ***SALT II: Proceedings from the Second Conference on Semantics and Linguistic Theory, May 1 - 3, 1992***. Edited by Chris Barker and David Dowty.
39. (December 1990): ($15.00) ***When Verbs Collide: Papers from the 1990 Ohio State Mini-Conference on Serial Verbs. May 26-27, 1990.*** Edited by Brian D. Joseph and Arnold M. Zwicky:
38. (July 1990): ($10.00) ***Papers from the Linguistics Laboratory***. Edited by Gina Lee and Wayne Cowart.
37. (August 1989): ($10.00) ***Papers in Morphology and Syntax***. Edited by Joyce Powers, Uma Subramanian, and Arnold M. Zwicky.
36. (September 1987): ($10.00) ***Papers from the Linguistics Laboratory 1985-1987***. Edited by Mary Beckman and Gina Lee.
35. (May 1987): ($10.00) ***A Festschrift for Ilse Lehiste. Papers by colleagues of Ilse Lehiste at Ohio State University.*** Edited by Brian Joseph and Arnold Zwicky.
34. (December 1986): ($6.00) ***Studies on Language Change***. Edited by Brian Joseph.
33. (August 1986): ($6.00) ***Finnish Particle Clitics and General Clitic Theory***. Joel A. Nevis: OSU Ph.D. Dissertation.
32. (July 1986): ($6.00) ***Interfaces***. Arnold Zwicky.
31. (July 1985): ($6.00) ***Studies in Generalized Phrase Structure Grammar***. Edited by Michael Geis.
30. (July 1984): ($5.00) ***German Temporal Semantics: Three -Dimensional Tense Logic and a GPSG Fragment.*** John A. Nerbonne: OSU Ph.D. Dissertation.
29. (May 1984): ($5.00) ***Papers on Morphology.*** Edited by Arnold Zwicky & Rex Wallace:
28. (May 1983): ($5.00) ***Common Discourse Particles in English Conversation***. Lawrence Clifford Schourup: OSU Ph.D. Dissertation.
27. (May 1983): ($5.00) ***Papers in Historical Linguistics.*** Edited by Gregory T. Stump.
26. (May 1982): ($5.00) ***Grammatical Relations and Relational Grammar.*** Edited by Brian D. Joseph.
25. (January 1981): ($5.00) ***Papers in Phonology.*** Edited by Arnold Zwicky.
24. (March 1980): ($5.00) ***Clitics and Ellipsis.*** *Edited* by Arnold M. Zwicky.
22. (February 1977): (5.00) ***Papers in Psycholinguistics and Sociolinguistics.*** Edited by Olga Garnia.
21. (May 1976): ($5.00) ***Papers on Nonphonology.*** Edited by Arnold Zwicky.

Information Concerning OSDL
(Ohio State Dissertations in Linguistics)

Dissertations written by students in the OSU Linguistics Department since 1992 are distributed by the student-run organization OSDL. Below you will find a complete list of dissertations that we distribute. Any of these dissertations may be obtained for US $15 for addresses within the continental USA, and for $17 for addresses outside of the continental USA (includes shipping). If you are interested in obtaining any dissertation, please send a check (personal check, traveller's check, cash, or a money order -- we cannot process credit card payments -- in US-dollars drawn from a US-bank) made out to the "Ohio State University" to:

OSDL
Department of Linguistics
Ohio State University
222 Oxley Hall
1712 Neil Avenue
Columbus, OH 43210-1298
USA

If you have any questions about OSDL or any of the dissertations it distributes, please send e-mail to <osdl@ling.ohio-state.edu> or write to the address above.

Some dissertations are available for free download. Before you send money please visit our web site:

http://ling.ohio-state.edu/dissertations/

Dissertations available as of January 2001

Benjamin Xiaoping Ao (1993) ***Phonetics and Phonology of Nantong Chinese*** (Advisor: David Odden)

Kutz M.C. Arrieta (1998) ***Nominalizations in Basque: A Case in Language Attrition*** (Advisor: Brian D. Joseph)

Gayle Ayers (1996) ***Nuclear Accent Types and Prominence: Some Psycholinguistic Experiments*** (Advisor: Mary E. Beckman)

Mary Bradshaw (1999) ***A Crosslinguistic Study of Consonant-Tone Interaction*** (Advisor: David Odden)

Michael Cahill (1999) ***Aspects of the Morphology and Phonology of Konni*** (Advisor: David Odden)

Hee-Rahk Chae (1992) ***Lexically Triggered Unbounded Discontinuities in English: An Indexed Phrase Structure Grammar Approach*** (Advisor: Arnold Zwicky)

Chan Chung (1995) ***A Lexical Approach to Word Order Variation in Korean*** (Advisor: Carl Pollard)

John Xiang-ling Dai (1992) ***Chinese Morphology and its Interface with the Syntax*** (Advisor: Arnold Zwicky)

Kim Ainsworth-Darnell (1998) ***The Effects of Priming on Recognition Latencies to Familiar and Unfamiliar Orthographic Forms of Japanese Words*** (Advisor: Julie E. Boland)

Mariapaola D'Imperio (2000) ***The Role of Perception in Defining Tonal Targets and their Alignment*** (Advisors: Mary Beckman and Keith Johnson) <download available>

Qian Gao (2000) ***Argument Structure, HPSG and Chinese Grammar*** (Advisor: Carl J. Pollard)

Svetlana Godjevac (2000) ***Intonation, Word Order and Focus Projection in Serbo-Croatian*** (Advisor: Craige Roberts) <download available>

Gwang-Yoon Goh (2000) ***The Synchrony and Diachrony of the English Prepositional Passive: Form, Meaning, and Function*** (Advisor: Brian D. Joseph)

Karin Golde (1999) ***Binding Theory and Beyond: An Investigation into the English Pronominal System*** (Advisor: Carl Pollard)

Rebecca Herman (1998) ***Intonation and Discourse Structure in English: Phonological and Phonetic Markers of Local and Global Discourse Structure*** (Advisor: Mary Beckman)

Eun Jung (Yoo) Hong (1997) ***Quantifiers and Wh-Interrogatives in the Syntax-Semantics Interface*** (Advisor: Carl Pollard)

Sun-Ah Jun (1993) ***The Phonetics and Phonology of Korean Prosody*** (Advisor: Mary E. Beckman)

Hyeon-Seok Kang (1997) ***Phonological Variation in Glides and Diphthongs of Seoul Korean: Its Synchrony and Diachrony*** (Advisor: Donald C. Winford)

Hyeree Kim (1996) ***The Synchrony and Diachrony of English Impersonal Verbs: A Study in Syntactic and Lexical Change*** (Advisor: Brian D. Joseph)

No-Ju Kim (1996) ***Tone, Segments, and Their Interaction in North Kyungsang Korean: A Correspondence Theoretic Account*** (Advisor: David Odden) <download available>

Gina Maureen Lee (1993) ***Comparative, Diachronic and Experimental Perspectives on the Interaction between Tone and Vowel in Standard Cantonese*** (Advisor: Brian D. Joseph)

Sook-hyang Lee (1994) ***A Cross-Linguistic Study of the Role of the Jaw in Consonant Articulation*** (Advisor: Mary E. Beckman)

Julie McGory (1997) ***The Acquisition of Intonation Patterns in English by Native Speakers of Korean and Mandarin*** (Advisor: Robert Fox)

Bettina Migge (1998) ***Substrate Influence in the Formation of the Surinamese Plantation Creole: A Consideration of Sociohistorical Data and Linguistic Data from Ndyuka and Gbe*** (Advisor: Donald C. Winford)

Frederick B. Parkinson (1996) ***The Representation of Vowel Height in Phonology*** (Advisor: David Odden) <download available>

Shu-Hui Peng (1996) ***Phonetic Implementation and Perception of Place Coarticulation and Tone Sandhi*** (Advisor: Mary E. Beckman)

Elizabeth Strand (2000) ***Gender Stereotype Effects in Speech Processing*** (Advisor: Keith Johnson)

Jennifer Venditti (2000) ***Discourse Structure and Attentional Salience Effects on Japanese Intonation*** (Advisor: Mary Beckman) <download available>

Katherine Welker (1994) ***Plans in the Common Ground: Toward a Generative Account of Conversational Implicature*** (Advisor: Craige Roberts)

Eun Jung Yoo (1997) ***Quantifiers and WH-Interrogatives in the Syntax-Semantics Interface*** (Advisor: Carl J. Pollard)

Jae-Hak Yoon (1996) ***Temporal Adverbials and Aktionsarten in Korean*** (Advisor: Craige Roberts) <download available>

Ohio State University Working Papers in Linguistics No. 55

Studies on the Interplay of Speech Perception and Phonology

Table of Contents

OSU Working Papers in Linguistics 55, 1-22

A Model of the Interplay of Speech Perception and Phonology[1]

Elizabeth Hume
ehume@ling.ohio-state.edu

Keith Johnson
kjohnson@ling.ohio-state.edu

1. Introduction

It has proven practical over a long history of research on language sound systems to rationalize phonological units and processes in terms of speech articulation. The Sanskrit grammarians, for example, focused on vocal anatomy and articulatory processes to the exclusion of descriptions of acoustic or auditory impressions produced by speech sounds (Allen, 1953). Similarly, the 19th century linguists Bell (1867), Sweet (1877), Sievers (1881), Passy (1890), and Rousselot (1897-1901) all focused primarily on speech articulation to explain sound change, describe similarities and differences across languages and in language teaching. For example, the Sweet/Bell system of vowel classification (which is still widely used in phonological description) and their iconic phonetic alphabets were based on speech articulation. This tradition of articulatory

---

[1] We would like to thank the members of our phonetics/phonology seminars for their very valuable and thoughtful input on this research. Thanks to Tim Face, Tsan Huang, Scott Kiesling, Matt Makashay, Jeff Mielke, Amanda Miller-Ockhuizen, Jennifer Muller, Misun Seo, Georgios Tserdanelis, Steve Winters and Peggy Wong. We are also grateful to Jose Ignacio Hualde, Brian Joseph, Jaye Padgett, and the members of the audiences at the University of Chicago and at the 1999 ICPhS Satellite Meeting, 'The Role of Perception in Phonology' for their helpful comments. The authors' names are listed alphabetically.

phonetics also formed the basis for the structuralists' approach to phonetics and phonology (Pike, 1943).

It is arguably the case that this early and prolonged emphasis on the articulatory foundations of sound systems was due to the fact that the articulators are open to observation. The linguist can observe the movements of the lips, jaw, and (with a little more ingenuity) the tongue, and the availability of such observations provided an important point of reference for theories of phonology by making available a set of explanatory mechanisms that can be applied to phonological patterns.

Rationalization of language sound systems from the point of view of the listener has, however, had a more spotted history. Some of the more obvious auditory properties have been noted (e.g., sonority, Sievers, 1881), but it was only recently - after the development of the sound spectrograph - that a comprehensive approach to language sound structure in terms of acoustic/auditory properties was attempted (Jakobson, Fant & Halle, 1952). However, JFH's attempt was impeded by the newness of the available technology and the relative paucity of perceptual data (which at the time was limited to basic psychoacoustic measures of pitch, loudness, and duration together with the earliest works on speech intelligibility for voice transmission over telephone lines). In his book on acoustic phonetics, Joos (1948) suggested that linguists would not readily accept auditory/acoustic foundations in the rationalization of language sound systems. Concerning Jespersen's (1904) chapter 'Akustisch oder Genetisch', Joos said:

> [Jespersen] showed that, however desirable it might seem to base phonetic categories upon acoustic characteristics, it was then impossible to make any progress in that direction because of the incapacity of the known instruments to furnish adequate data. Making a virtue of necessity, phoneticians have developed phonetic theory entirely upon the articulatory ('genetisch') basis, and developed it to the point where inadequacy is seldom if ever noticed. Nothing happened to shake Jespersen's conclusion for nearly half a century. During this time the technicians produced no instrument which could deal with the central problem, and phonetic doctrine crystallized in the tradition that articulation can alone support linguistically useful phonetic categories. (Joos, 1948:7)

Joos' comments foreshadowed theoretical developments in the years following JFH in which linguists returned to the more established knowledge-base provided by the phonetic study of speech articulation (Chomsky & Halle, 1968). One change in attitude which has persisted, however, is that after JFH it is often assumed that phonological features have dual definitions both in terms of audition/acoustics and articulation (see, e.g. Hume 1994 regarding [coronal]). Yet, despite this acknowledged role for auditory aspects of speech, perceptual effects and auditory properties of sound have less commonly played a role in linguists' speculations on the role of phonetics in phonological patterns (though see, e.g., Bladon 1986; Donegan, 1978; Liljencrants & Lindblom, 1972; Lindblom, 1990; Martinet, 1955; Ohala, 1990, 1993).

It is significant, therefore, that the role of speech perception in language sound systems has recently seen a revival of interest among phonologists. This increasing interest appears to be driven by two factors. First, rapid technological advances over the last 10 to 15 years have made it feasible to collect a wide range of perceptual data both in the laboratory and in the field (e.g. Wright, 1996). This in turn has made it possible for researchers to work out some general properties of speech perception which appear to be relevant in stating phonological patterns. Second, the development of Optimality Theory (Prince & Smolensky 1993, McCarthy & Prince 1993) has allowed for the statement of perceptually grounded constraints which interact dynamically with constraints motivated by other general principles. As a result, there has been a new and growing interest in exploring the role of perceptual phenomena in accounting for cross-linguistic sound patterns (e.g. Boersma 1998, Côté 1997, Flemming 1995, Hume 1998, Jun 1995, Hayes 1999, Ovcharova 1999, Silverman, 1995, Steriade 1995, 1997). For instance, building on insights from, e.g., Kingston (1985) and Ohala (1981), in addition to the notion of phonetically grounded constraints (e.g., Archangeli & Pulleyblank 1994), Steriade's (1995, 1997) pioneering work in this area explores the extent to which phonological constraints grounded in perceptual cues account for cross-linguistic patterns of laryngeal neutralization and retroflexion. Regarding the former, Steriade argues that loss of laryngeal contrast occurs in contexts in which the perceptual cues to the specific contrast are relatively weak. Conversely, contrasts are maintained in positions that are high on the scale of perceptual salience.

These developments in speech perception and phonological research provide a solid foundation for continued and significant progress in understanding language sound systems. The time then seems ripe to consider the interplay of speech perception and phonology more closely. In this regard, there are at least three key research questions that we see as important starting points for this endeavor: first, to what extent does speech perception influence phonological systems?; second, to what extent does the phonological structure of language influence speech perception?; and third, where do speech perception phenomena belong in relation to a formal description of the sound structure of language? In the following sections we address each of these questions, first, by focusing on the interplay of phonology and speech perception, and then by laying out a general model for the study of the interaction of phonology with external forces such as speech perception.

2. The Interplay of Speech Perception and Phonology

In this section, we present a range of evidence, including new work from our lab, pointing to the influence of language sound structure on speech perception, as well as the influence of speech perception on phonological systems.

2.1 The Influence of Phonological Systems on Speech Perception

That phonological systems have an influence on speech perception is suggested by a variety of evidence. For example, studies in second language learning (e.g. Best et al. 1988; Polka & Werker, 1994) have found that listeners are more adept at perceiving sounds of their native language than those of a second language acquired later in life. Furthermore, first language acquisition research (e.g. Kuhl, et al., 1992) shows that perceptual learning occurs as babies' perceptual systems become tuned to language-specific phonetic patterns, such as typical vowel formant ranges. Additionally, model studies (Guenther & Gjaja, 1996; Makashay & Johnson, 1998) have explored auditory neural map formation mechanisms that may be involved in phonetic acquisition. Adaptive neural network models of perceptual learning show human-like patterns of phonetic tuning using idealized pseudo-phonetic data (Guenther & Gjaja, 1996) and using real phonetic data (Makashay & Johnson, 1998).

Phonological systems of contrast may also influence perception (e.g., Dupoux, Pallier, Sebastian & Mehler, 1997; Lee, Vakoch & Wurm, 1996). For example, experimental results from Hume, Johnson, Seo, Tserdanelis, & Winters (1999) indicate that for both Korean and American English listeners, transition stimuli have a greater amount of consonant place information than burst stimuli. However, it is interesting that for Korean listeners this difference between bursts and transitions was greater than it was for American English listeners. In other words, Korean listeners were better able to identify a consonant's place of articulation from the transition stimuli alone, than were American listeners. One explanation for this finding relates to differences in the system of phonological contrasts in each language. Unlike English, Korean includes the set of phonological contrasts among tense, lax, and aspirated stops, which is cued in part by the amplitude of aspiration. The presence of these phonological contrasts may lead Korean listeners to focus greater attention on the interval of time following the stop release burst; that is, on the transitions.

2.2 The Influence of Speech Perception on Phonological Systems

Speech perception plays at least three distinct roles in shaping language sound systems: a. failure to perceptually compensate for articulatory effects; b. avoidance of weakly perceptible contrasts; c. avoidance of noticeable alternations.

Ohala's (1981) account of the listener as a source of sound change is one of the most explicit accounts of a point of contact between speech perception and language sound structure. In this account, listeners may fail to perceptually compensate for coarticulation and come to use different articulatory targets in their own speech by misapprehending speech produced by others (see also, Beddor et al., 2001). This is illustrated in (1), where a speaker in uttering /xy/[2] produces [wy] because of coarticulation between [x] and [y]. The listener fails to compensate for the coarticulation and so presumes that the first speaker intended to say /wy/.

(1) /xy/ → [wy] → /wy/

The common process of palatalization (or rather, coronalization, see e.g. Hume 1994) may also have its roots in misperception. Chang et al. (2001) and Clements (1999) suggest that the common manner change of a velar stop to a palato-alveolar affricate before a front vowel is due to the listener's reinterpretation of the velar's aspiration as the frication noise of a strident consonant. Thus, synchronic variability or diachronic change in sound patterns may be due to listener's misperceptions, that is, a phonetics/phonology mismatch.

The second area in which speech perception exerts influence on phonological systems derives from the fact that contrasts of weak perceptibility tend to be avoided in language. For example, sound differences that are relatively imperceptible tend not to be used contrastively in language. In the extreme case this can be an absolute prohibition. Illustrations of such imperceptible contrasts include interdental [θ] versus dental [θ], concave versus convex tongue shape for lax vowels, etc. These are all pronounceable, but low salience, contrasts that are not used in language.

Contrast is relevant from both paradigmatic and syntagmatic perspectives, and weak contrast along either dimension may be avoided by enhancing, or optimizing, the contrast, on the one hand, or sacrificing it, on the other. This can be achieved by means of a variety of repair strategies, including epenthesis, metathesis, dissimilation, assimilation and deletion. Among these strategies, epenthesis, dissimilation and metathesis tend to optimize contrast, while with assimilation and deletion contrast is sacrificed.

To illustrate, **epenthesis** in Maltese can be seen as strengthening a length contrast among consonants. In this process, the vowel [i] is epenthesized before a word-intitial geminate consonant, created by the concatenation of the imperfective morpheme /t/ and a stem-initial coronal obstruent, e.g. /t+dierek/ [iddierek] 'to rise early, 3rd p. imperf.'

---

[2] In this discussion, x,y,w are used as variables over phonetic symbols.

(Aquilina, 1959; Hume, 1996). Since the perceptual cues to word-initial geminates, stops especially, are relatively weak (see, e.g. Abramson, 1987; Muller, in prep.), insertion of a vowel before the geminate enhances the perceptibility of consonant length, and hence, the identity of the imperfective morpheme. Contrast optimization also occurs in English plural noun formation where a vowel precedes the plural morpheme just in case the noun stem ends in a sibilant consonant, e.g. dishes, judges, cf. modems, cats. Since the plural morpheme is itself a sibilant, the appearance of a vowel between the two consonants renders the distinction between the segments more perceptible. This is all the more important given that the second sibilant alone carries the meaning of the plural morpheme. That contrast is strengthened in this manner follows from the view that large modulations in the speech signal serve to increase the salience of cues in the portion of the signal where the modulation takes place (Ohala 1993; Kawasaki 1982). It makes sense that modulation would enhance the perceptibility of fricative sequences because otherwise auditory masking would obscure place information in adjacent fricatives (Bladon, 1986).

Many cases of **dissimilation** receive the same account. In Greek, for instance, consonant clusters comprised of two stops or two fricatives optionally dissimilate resulting in variation among, for example, [pt] ~ [ft] (***epta ~ efta*** 'seven'); and [fθ] ~ [ft] (***fθinos ~ ftinos*** 'cheap' (masc. nom.) (Newton, 1972; Tserdanelis, 2001). Dissimilation effects a difference in manner of the two segments, enhancing syntagmatic contrast by increasing the modulation between adjacent segments.

Perceptibility can also be a trigger for **metathesis**. To cite but one example, in Faroese, the sequence /sk/ metathesizes just in case a stop consonant follows, e.g. /baisk + t/ [baikst] *[baiskt] 'bitter, neut.sg.' (Jacobsen & Matras 1961, Lockwood 1955, Rischel 1972; Seo & Hume 2001). Hume (1998, 2001) argues that consonant/consonant metathesis in Faroese, as in many other languages, serves to enhance both paradigmatic and syntagmatic contrast.[3] The problem with the unmetathesized sequence stems from the fact that a stop consonant would be sandwiched between two consonants, a context of poor perceptibility for the stop, in particular. To repair the sequence, the consonants switch order so that the weaker stop consonant is positioned in a more robust context. Thus, the perceptibility gain in the output is achieved by shifting the stop to postvocalic position, a context with more robust stop place cues. The fricative consonant, with stronger internal place cues, fares better in interconsonantal position. Winters (2001) also found evidence of a perceptibility gain for patterns of stop/stop metathesis observed cross-linguistically in VCCV sequences (see also Steriade, 2001).

---

[3] See Blevins & Garrett (1998) for discussion of the role of perception in consonant/vowel metathesis.

Contrary to the repair strategies above in which the avoidance of a weak contrast is achieved by perceptual optimization, in cases of total segment **assimilation** and **deletion** contrast is instead sacrificed. For example, in Korean the sequences /n+l/, /l+n/ are realized as [ll], e.g. /non-li/ [nolli] 'logic', **/səl-nal/ [səllal]** 'New Year's day' (see e.g., Davis & Shin, 1999; Seo, 2001). In Seo's (2001) discussion of the role of perception in Korean assimilation, she notes that the syntagmatic contrast of the nasal/lateral sequence is of low salience, given the acoustic/auditory similarity of the two segment types. The articulatory effort required to maintain a perceptually salient contrast between the two segments is outweighed by, what we speculate to be, the articulatory forces driving assimilation. The consequence is a loss of nasal-oral contrast in this context. For further discussion on the possible link between perception and assimilation, see Hura et al., 1992; Jun, 1995; Ohala, 1990; Steriade, 2001; Winters 2001.

The ultimate sacrifice in contrast occurs with segment deletion, such as Turkish /h/ deletion. Experimental evidence supports the perceptual basis of this type of deletion. As Mielke 2001 and Ovcharova 1999 show, /h/ optionally deletes in contexts in which it is relatively imperceptible, such as after an aspirated stop but not before ([ethem] ~ [etem] 'proper name'; [kahpe] *[ka:pe] 'harlot'), word-finally but not word-initially ([timsah][4] ~ [timsa:] 'crocodile'; [hava] *[ava] 'air'), and adjacent to a fricative ([safha] ~ [safa] 'sleep'; [tahsil] ~ [ta:sil] 'education').

The third area in which speech perception exerts influence on phonological systems concerns the avoidance of noticeable alternations. In this function, perception is seen as a type of filter on sound change. For example, Kohler 1990 states that changes are "only accepted (1) if they bear an auditory similarity to their points of departure, and (2) if the situational context does not force the speaker to rate the cost of a misunderstanding or a break down of communication very high" (p. 89). Note that the filter has two aspects, the first purely in terms of perceptual salience and the second in terms of the communicative context. Drawing on evidence from assimilation, Steriade (1999) interprets the communicative aspect of the filter in a more sociolinguistic manner: "innovation is channeled... in the direction that is least likely to yield blatant departures from the [established pronunciation] norm."

Huang's (2001) study of tone sandhi in Mandarin Chinese illustrates this effect. Mandarin has four lexical tones: level high (55); mid-rising (35); low-falling-rising (214); high-falling (51). (The numbers in parentheses indicate the pitch values of the tones on a five-level scale.) The phonological process under study concerns the well-known tone sandhi in which a low-falling-rising tone is simplified to mid-rising just in case it is followed by another low-falling-rising tone, i.e. /214 214/ → [35 214]. Huang argues

---

[4] Deletion of /h/ word-finally seems to be categorical for at least some speakers.

that this process is a case of perceptually tolerated articulatory simplification (Hura et al., 1992; Kohler, 1990; Steriade, 2001). In other words, the contour tone 214 is simplified to 35, rather than to 55 or 51, one of the other two "simpler" tones in the language, because 214 is more similar to 35 than it is to either of the other tones. The phonological change is, therefore, less noticeable. To test this hypothesis, native speakers of American English and Mandarin Chinese discriminated pairs of the four Mandarin Chinese tones. The results support Huang's hypothesis; listeners from both languages had the greatest difficulty distinguishing between 35 and 214, as shown in figure 1. It is interesting to note that this tendency was much more pronounced for Mandarin Chinese listeners, suggesting a further effect of phonology on perception (see section 4.2. for related discussion).

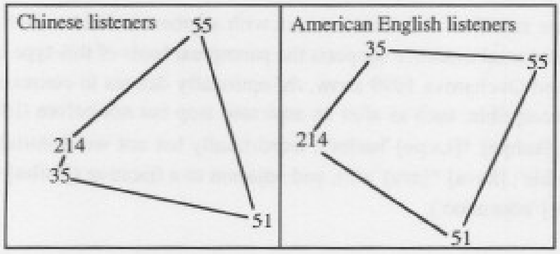


Figure 1. The four tones of Mandarin Chinese in perceptual space for Mandarin Chinese listeners and American English listeners. Multidimensional scaling data from Huang (2001).

While the preceding studies focus on the perceptual/communicative aspects of the filter, we interpret it more broadly, as including at least four external forces: perception, production, generalization, conformity. This can be illustrated in general terms in the context of the five phonological repair strategies noted above. As shown in figure 2, for every sound or sound sequence that is ripe for change (for perceptual, articulatory or other reasons), there are a variety of potential ways in which a sequence can be modified. For example, to repair a given sequence 'xy', any of the five repair strategies given below could be used. That is, a segment could be epenthesized between 'x' and 'y', the order of the two segments could be reversed, one of the segments could be deleted, and so on. There can also be more than one possible output for a given repair strategy. With respect to epenthesis, for example, the sequence 'xy' could be repaired by inserting a segment between the two sounds, before the entire sequence, or after it. All three patterns are observed cross-linguistically (see Broselow 1981, Kenstowicz 1994 for related discussion). The selection of the output is determined by filters, of which perception is one. How this filtering is implemented constitutes the focus of section 4.

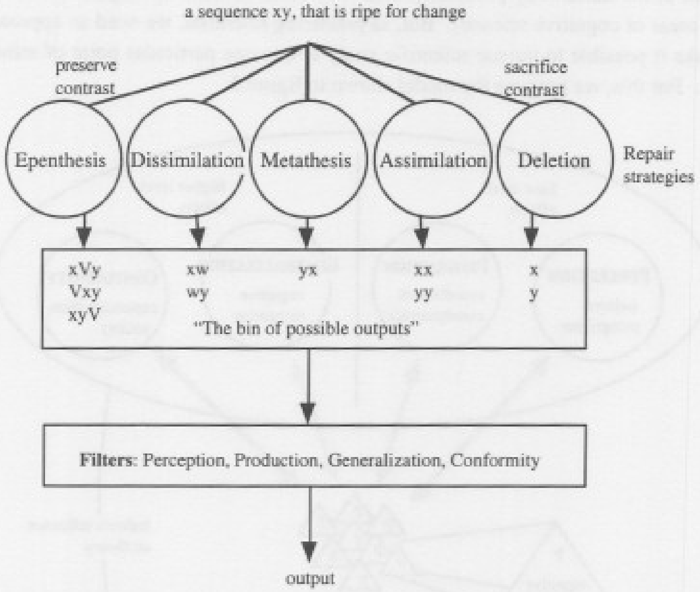


Figure 2. Characterization of phonological repair strategies, and the role of filters in selecting among possible outputs.

3. The Interaction of External Forces and Phonology

To study the role of speech perception in phonology it is necessary to conceive of ways that realities in the domain of speech perception interface with the cognitive symbolic representation of language sound structure. Realities in speech perception are tied up with physical acoustic descriptions of speech sounds and the auditory transduction of speech sounds in the auditory periphery. Phonological systems, on the other hand, are symbolic in nature, dissociated from any particular physical event in the world. Indeed, such is the independence of phonology from the physical world, that it can be said that two people share the same symbolic phonological system, speak the same language, even though their experience of physical events in the world does not overlap at all. Prior to mass communication this may have been the rule.

The problem is thus a classic one in the study of language sound systems, namely the relationship between phonetics and phonology. The phonetics/phonology interface problem is an instance also of the classic philosophical problem on the relationship between the mind and the body. Our strategy may or may not be relevant for other instances of the mind/body problem (whether in other domains in linguistics, or in more remote areas of cognitive science). But, as practicing scientists, we need an approach that will make it possible to pursue scientific study at this one particular point of mind/body contact. For this, we propose the model shown in figure 3.

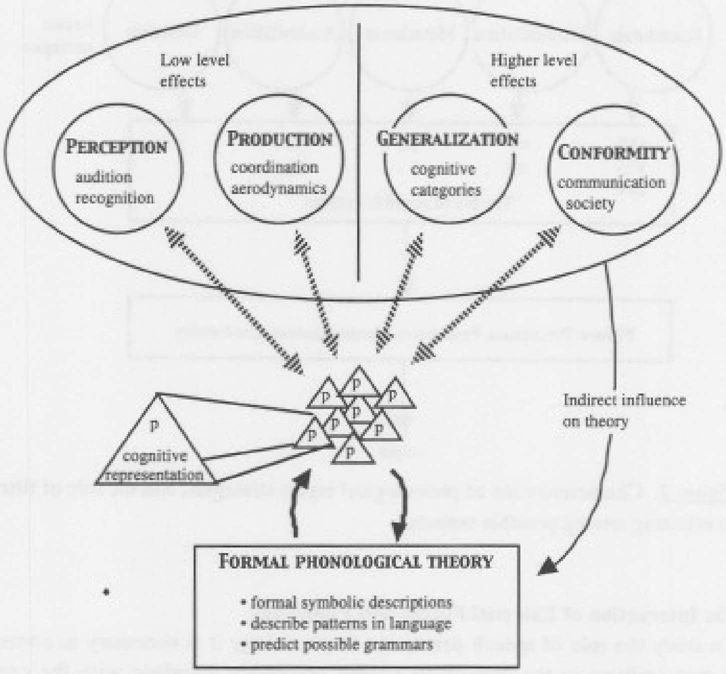


Figure 3. A general model of the interplay of external forces and phonology, broadly defined.

In the study of language sound systems, we work with two symbolic domains: the one cognitive, the other formal. The cognitive symbolic representation of a language's sound system, characterized as *p* in figure 3, is embodied in an individual's brain. We may assume that *p* is a component of *l*, the cognitive symbolic representation of a language. The linguistic sound system of a community of speakers/listeners can thus be defined as a collection of *p*'s. The formal symbolic domain defines the inventory of symbols and the procedures for symbol manipulation found in formal linguistic descriptions. The theory describes sound patterns observed in language, hence, the arrow pointing from *p* to

Formal Phonological Theory in figure 3. It is these sound patterns that constitute the data that the theory is based on. The arrow pointing from *Formal Phonological Theory* to *p* reflects the goal of phonological theory to predict possible grammars. A formal symbolic description is not the same as a cognitive symbolic representation. Nonetheless, formal descriptions that remain consistent with what is known about cognitive representation provide insight into the cognitive representation by providing a language for discussing the intricacies of the mind.

The relationship between external factors and the two symbolic domains is also illustrated in figure 3. Two familiar low-level effects in the model, *perception* and *production*, have been discussed for decades in functional accounts of sound patterns. The role of 'ease of perception' and 'ease of production' are widely cited, though specific proposals as to how they may influence language are rare. Notice that, in our view, perceptual and productive abilities can both influence the sound system of language as well as be influenced by one's language, hence the bi-directional arrows in the diagram between these effects and *p*. Examples of these influences are provided in section 4 (see also section 2 regarding perception). Also included in the model are two higher level effects, *generalization* and *conformity*. Generalization refers to the tendency to simplify cognitive representations relative to the sensory reality experienced. This tendency for generalization underlies category formation in cognitive systems generally, and we see it as related to linguistic processes such as paradigm leveling and analogy. Conformity relates to the social and communicative factors which play an important role in shaping language sound structure. From a social perspective, the need to conform to a linguistic norm, for example, can exert influence over an individual's cognitive language sound patterns. The need in a communicative system to use forms that others will identify and accept also influences sound systems. Further discussion of the bi-directional influence of the two higher level factors appears in section 4 below.[5]

In our view, cognitive language sound patterns (*p*) are directly influenced by these external forces. However, the connection between formal phonological theory and the external forces is indirect (for an alternative view see, e.g., Flemming 1995, Steriade, 2001). The formal theory describes patterns found in individual languages and from these, derives cross-linguistic generalizations about those patterns. To the extent that language sound patterns are caused by external factors such as speech perception, these factors are reflected in the formal phonological theory. Yet, to incorporate them directly into phonological theory erroneously implies that they are exclusive to language. On the contrary, the cognitive factor, generalization, for example, relates not only to linguistic

---

[5] We do not rule out the possibility of other external factors. For example, Karen Landahl has suggested to us that ecological factors may have an influence on language sound systems. We leave this topic open for future consideration. We also considered whether to add learnability to the inventory, but decided that this is subsumed under the other factors.

category formation, but to category formation in general. Similarly, speech perception uses perceptual abilities that are also relevant to general auditory and visual perception (Fowler 1986). We refer the reader to Hale & Reiss (2000) for related discussion.

We view the model outlined in figure 3 as a starting point for the study of the interplay of external forces and phonology, broadly defined. Each aspect of the model constitutes an important area of research which, together, will lead to a more comprehensive understanding of language sound structures.

4. Implementation

Section 2 provided evidence that speech perception influences phonology and vice versa, and section 3 outlined a rather abstract model of how external forces interact with phonology. This section explores in more detail how to implement this model.

The interplay of perception and phonology occurs in time because speech perception is a process that occurs in time - the process of word recognition has a measurable onset and offset. Similarly, speech production is also a process that occurs in time. The higher-level functions, generalization and conformity are also tied to events in time; generalization to the process of language acquisition and perhaps also aspects of continuing language use; and conformity to events of personal interaction involving language use. Therefore, because these external forces operate on events in time, our model of the interplay of perception and phonology is implemented over time. That is, perception exerts influence on an individual's cognitive domain at a particular point in time, resulting in a modified representation of the sound system in question. In more formal terms, we suggest that the interplay of speech perception and phonology is implemented as the mapping from p to p', where p is a cognitive symbolic sound system at some particular time t, and p' is a cognitive symbolic sound system at some later time $t+\delta$. The mapping $p > p'$ (figure 4) is made up of a set of parallel filters or transduction functions comprised of the external forces introduced in section 2.

To understand how perception filters p, suppose that p requires the perception of a distinction that is somewhat hard to hear. In some instances, the difficult distinction required by p will be missed, simply misheard, so p will undergo a change to p'. This is very much in the spirit of Ohala's 1981 account of the listener as a source of sound change. The filtering action imposed by production takes a similar form. The cognitive symbolic representation p requires that the speaker make a sound that is hard to say. In some instances the speaker will fail to produce the sound and say something else and in this way contribute to a change in p. The filtering action of generalization is a little different from these. Here p appears to have a regular pattern which the cognitive system captures by reorganizing p. The cognitive category formation mechanism which we envision forms generalizations at the lowest level of acoustic/phonetic categories up to abstract morphophonemic patterns. Finally, conformity tends to bring p into line with the linguistic norms of the community whenever p differs from those norms.

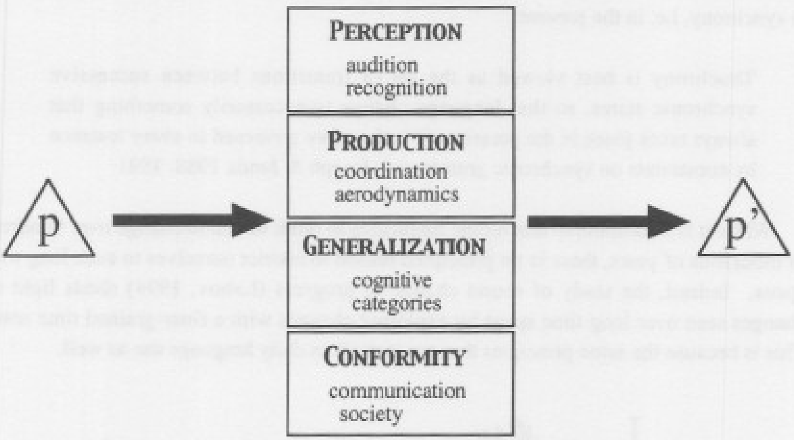


Figure 4. The mapping of p onto p' can be decomposed into a set of filters. Each component of the mapping process independently influences the relationship between p and p' and, hence, the structure of p'.

This model raises two important implementation issues. First, it is necessary to give an account of interactions among external forces in this model. How is the perceptual filter modulated by the production filter? How can conformity prevent changes that are motivated by perceptual or productive ease? Second, the language specificity of the external phonological forces (the upward-going arrows in figure 3) needs to be addressed. How are external forces dependent upon or shaped by the cognitive symbolic representation of language sound systems? We treat interactions among forces as a problem of understanding the time scale of phonetic mutation, and we treat language specificity by referring to p in the definition of the forces.

4.1 Interactions of external forces

The four filters in figure 4 (the external phonological forces) can be treated as completely independent of one another. Interactions of opposing tendencies in this model occur in cycling $p > p'(> p...)$ where the interval between cycles is very short. A change that reduces cost on one function may produce increased cost on another function and so be quickly reversed. For example, the sound pattern [nt] may be changed to [nd] in order to achieve lower articulatory cost (avoiding the modulation of voicing). In the next cycle,

[nd] may be changed back to [nt] because [nd] conflicts with conformity (e.g. [nd] diverges too much from the socially accepted pronunciation norm).

This view is consistent with Joseph & Janda's (1988) view that sound change occurs in synchrony, i.e. in the present.

> 'Diachrony is best viewed as the set of transitions between successive synchronic states, so that language change is necessarily something that always takes place in the present and is therefore governed in every instance by constraints on synchronic grammars.' (Joseph & Janda 1988: 194)

While it is traditional in diachronic linguistics to think of sound change over hundreds or thousands of years, there is no principled reason to restrict ourselves to such long time spans. Indeed, the study of sound change in progress (Labov, 1994) sheds light on changes seen over long time spans by exploring changes with a finer-grained time scale. This is because the same principles that are at work in daily language use as well.

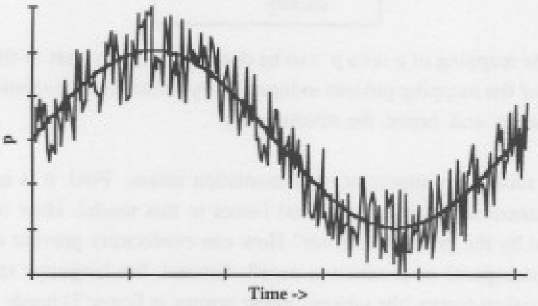


Figure 5. A coarse-grained time scale shows general tendencies, illustrated by the slowly changing line, while a fine-grained time scale shows rapidly fluctuating change. Time in this illustration is on the horizontal axis, and the vertical axis is meant to show, in an abstract one dimensional projection, the location *p* of a language in the space of possible languages.

Thus, unlike a view of sound change that uses a coarse-grained time scale, our model handles interactions among forces by adopting a fine-grained scale, as illustrated in figure 5. The function *p(t)*, which shows the development of sound *p* over time, has local noise overlaying global stability. Through the sequential interaction of forces, it is a self-organizing system that is nonetheless in constant flux.

4.2 Language Specificity

The model in figure 3 has bi-directional arrows between the cognitive symbolic representations *p* and each of the external forces. We saw in section 2 that there is some evidence that speech perception processes are language specific, influenced, for example, by the system of contrasts in a language. Further evidence of the language specificity of speech perception can be seen in Mielke's (2001) study of /h/ perception in English, French, Turkish, and Arabic. Figure 6 shows average sensitivity to /h/ in a variety of segmental contexts in Mielke's study. (Sensitivity was estimated using the signal detection measure d'.) Two aspects of these data are relevant in the current discussion. First, the cross-linguistic differences are striking. The two languages with limited /h/ distributions, English and French, show low /h/ salience, while the two languages with extensive /h/ distributions, Turkish and Arabic, show high /h/ salience. Second, despite these cross-linguistic differences, all four of the languages show similar patterns of salience as a function of different segmental environments.

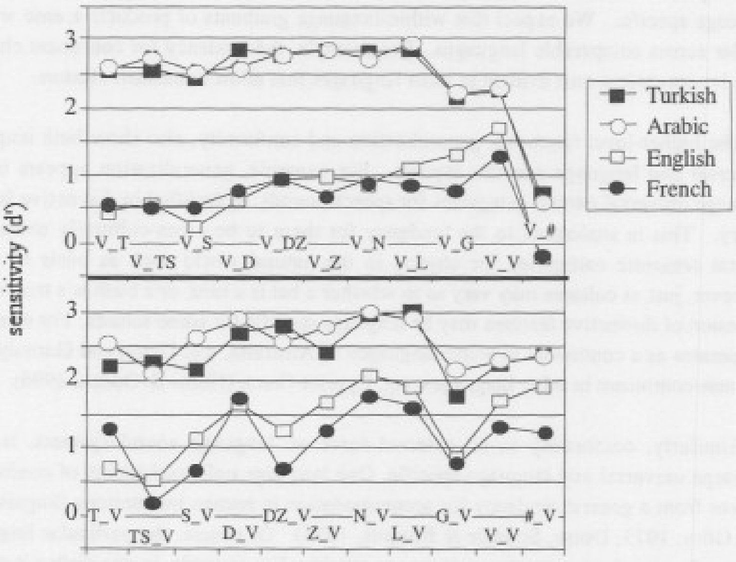


Figure 6. Perceptual sensitivity to /h/ in different segmental contexts by listeners of American English, Turkish, Arabic, and French. Data from Mielke (2001). Sensitivity (d') to [h] in postvocalic (top) and prevocalic positions (bottom). (T/D=voiced stop, TS/DZ=voiceless/voiced afficate, S/Z=voiceless/voiced fricative, N=nasal, L=liquid, G=glide, V= vowel)

With these data in mind, we could say that the perceptual influence on phonology is static because the pattern of perceptual salience in segmental context remains relatively constant across languages, but the perceptual influence on phonology is also dynamic because overall /h/ salience differs from language to language. The influence of perception on *p*, includes both the universal, static aspect of perception and the language specific, dynamic aspect. The upward pointing part of the bi-directional arrow from *p* to perception is meant to depict the fact that language sound systems shape perception.

There is also evidence that language sound systems can influence speech production, linguistic generalization, and social conformity. The language universal aspect of production has been a focus of research for over a century. However, it seems undeniable that any definition of easy or hard sounds or sound sequences must make reference to the native language(s) of the speaker. A post-alveolar click with velar accompaniment [!x] may be very hard for a person who doesn't speak !Xóõ, while it is perfectly natural to the native speaker. But as with perception, ease of production is both language universal and language specific. We expect that within-language gradients of productive ease will be similar across comparable languages. For example, the tendency for consonant clusters to be homorganic seems evident in most languages that allow consonant clusters.

The higher-level functions, generalization and conformity, also show both language universal and language specific aspects. For example, generalization appears to use language universal natural categories for speech sounds, as codified in distinctive feature theory. This is analogous to the tendency for there to be cross-culturally ubiquitous natural semantic categories for objects in the natural world such as birds or trees. However, just as cultures may vary as to whether a bat is a bird, or a bush is a tree, so the extension of distinctive features may be language specific for some sounds. For example, /l/ operates as a continuant in some languages of Australia, e.g. Djapu and Gurindji, and as a non-continuant in other languages, e.g. Cypriot Greek (Hume & Odden, 1996).

Similarly, conformity as an external force on language sound systems, is both language universal and language specific. One language universal aspect of conformity derives from a general tendency for accommodation in human interactions (linguistic or not; Giles, 1973; Doise, Sinclair & Bourhis, 1976). Of course, the particular linguistic norms of a speech community are language specific. For example, in one dialect 'cat' may be pronounced [**kæt**] while in another it is [**kæʔ**]. So, cognitive symbolic representations define norms, and conformity derives expectations based on those norms. But in addition to this, the drive for accommodation itself may be altered by *p*. It seems logical that if a community has a fairly diverse makeup such that people are exposed to a large range of linguistic variation, then the tendency for accommodation, and hence conformity, may be lessened.

To summarize, there is evidence that p influences each of the four external phonological forces. This justifies the bi-directional arrows in figure 3. However, in our sketch of the implementation (figure 4) there is no explicit account of bi-directionality.

We could implement language specificity as a type of cyclic filtering, where the external forces are altered (filtered) by p as schematized in (2). (2a) shows the idea that was presented earlier in figure 4. (2b) extends this notion to suggest that p also serves as a kind of function on the set of external forces.

(2) a) p -> filter -> p' a') $f(p) = p'$
b) filter -> p -> filter' b') $p(f) = f'$

However, notice that the language specificity of the external forces derives from the fact that we define each of them in terms of p. That is, /h/ is perceptually salient in languages that have extensive /h/ distributions. /!x/ is pronounceable in languages that have /!x/ in their system of phonological contrasts. Similarly generalization and conformity are both operations over the contents of p. So, by defining the external forces in terms of the cognitive symbolic representation of language sound structure (a system of contrasts and a lexicon of word forms that make use of those contrasts) we have built language specificity into them.

5. Conclusion

The model outlined above is presented as a starting point for the study of the interplay of speech perception and phonology, defined to include the cognitive and formal representations of phonological systems. The aim of this chapter has been to situate the study of the interplay of these two domains in a broader context - taking into account other factors such as speech production, linguistic cognition and social influence. While we recognize that this venture is necessarily programmatic, we see each aspect of the model as constituting an important area of research which, together, will lead to a more comprehensive understanding of language sound structures.

REFERENCES

Abramson, A. (1987) Word-initial consonant length in Pattani Malay. *ICPhS 11*, 68-70.
Allen, W.S. (1953) *Phonetics in Ancient India.* London: Oxford University Press.
Aquilina, Joseph. (1959) *The Structure of Maltese.* Valletta, Malta: Progress Press Co.
Archangeli, D. & D. Pulleyblank (1994) *Grounded Phonology.* Cambridge, MA: MIT Press.
Bell, A.M. (1867) *Visible Speech.* London: Simpkin, Marshall & Co.

Beddor, P., Krakow, R.A. & Lindemann, S. (2001) Patterns of perceptual compensation and their phonological consequences. In E. Hume & K. Johnson (eds.), *The Role of Speech Perception in Phonology* (pp. 55-78). New York: Academic Press.

Best, C.T., McRoberts, G. & Sithole, N. (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, 345-60.

Bladon, A. (1986) Phonetics for hearers. In G. McGregor (ed.), *Language for Hearers*. Oxford: Pergamon, pp. 1-24.

Blevins, J. & Garrett, A. (1998) The origins of consonant/vowel metathesis. *Language 74*, 508-553.

Boersma, P. (1998) Functional phonology. PhD dissertation. University of Amsterdam.

Broselow, Ellen. (1981) On predicting the interaction of stress and epenthesis. *Glossa 16*, 115-32.

Chang, S., Plauché, M. & Ohala, J. (2001) Markedness and consonant confusion asymmetries. In E. Hume & K. Johnson (eds.), *The Role of Speech Perception in Phonology* (pp. 79-101). New York: Academic Press.

Chomsky, N. & Halle, M. (1968) *The Sound Pattern of English*. New York: Harper & Row.

Clements, G. N. (1999) Affricates as non-contoured stops. In O. Fujimura, B. Joseph & B. Palek (eds.), *Proceedings of LP'98: Item order in language and speech.* (pp. 271-299). Charles University, Prague: The Karolinum Press.

Côté, M-H. (1997) Phonetic salience and consonant cluster simplification. In B. Bruening, Y. Kang and M. McGinnis (eds.), *MITWPL 30.* (pp. 229-262). Cambridge, MA: MIT.

Davis, S. & Shin, S-H. (1999) The syllable contact constraint in Korean: An optimality-theoretic analysis. *Journal of East Asian Linguistics 8*, 285-312.

Donegan, P. (1978) *The natural phonology of vowels*. PhD dissertation, Ohio State University. New York: Garland Press, 1985.

Doise, W., Sinclair, A., & Bourhis, R. (1976) Evaluation of accent convergence and divergence in cooperative and competitive intergroup situations. *British Journal of Social and Clinical Psychology 15*, 247-252.

Dupoux, E., Pallier, C., Sebastian, N. & Mehler, J. (1997) A destressing 'deafness' in French? *Journal of Memory and Language 36*, 406-421.

Flemming, E.S. (1995) *Auditory representations in phonology*. UCLA PhD dissertation.

Fowler, C.A. (1986) An event approach to the study of speech perception. *Journal of Phonetics 14*, 3-28.

Giles, H. (1973) Accent mobility: a model and some data. *Anthropological Linguistics 15*, 87-105.

Guenther, F. & Gjaja, M. (1996) The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America 100*, 1111-1121.

Hale, M. & Reiss, C. (2000) 'Substance abuse' and 'dysfunctionalism': Current trends in phonology. *Linguistic Inquiry 31*, 157-169.

Hayes, B. (1999) Phonetically driven phonology: The role of Optimality Theory and Inductive Grounding. In Darnell, M., E. Moravcsik, F. Newmeyer, M. Noonan and K. Wheatley (eds.), Frunctionalism and Formalism in Linguistics. Studies in Language Companion Series 41 (pp. I: 243-285). Amsterdam: Benjamins.

Huang. T. (2001) Tone perception by speakers of Mandarin Chinese and American English. *The Interplay of Speech Perception and Phonology*, OSUWPL, vol. 55.

Hume, E. (1994). *Front Vowels, Coronal Consonants and their Interaction in Nonlinear Phonology.* New York: Garland. (1992 Cornell Dissertation)

Hume, E. (1996) Coronal consonant, front vowel parallels in Maltese. *Natural Language and Linguistic Theory 14*, 163-203.

Hume, E. (1997) Towards an Explanation of Consonant/Consonant Metathesis. ms. OSU.

Hume, E. (1998) The Role of Perceptibility in Consonant/Consonant Metathesis. *Proceedings of West Coast Conference on Formal Linguistics 17*, 293-307.

Hume, E. (2001) Metathesis: formal and functional considerations. In E. Hume, N. Smith & J. van de Weijer (2001).

Hume, E., K. Johnson, M. Seo, G. Tserdanelis & S. Winters. (1999) A Cross-linguistic Study of Stop Place Perception. *Proceedings of the XIVth International Congress of Phonetic Sciences*. 2069-2072.

Hume, E. & Odden, D. (1996) Reconsidering [consonantal]. *Phonology 13*, 345-376.

Hume, E., Smith, N., & van de Weijer, J. (eds.) (2001) *Surface Syllable Structure and Segment Sequencing*. Leiden, NL: Holland Institute of Linguistics.

Hura, S.L., B. Lindblom & R.L. Diehl. (1992) On the role of perception in shaping phonological assimilation rules. *Language & Speech, 35.1, 2*, 59-72.

Jacobsen, M.A. & C. Matras. (1961) *Førosysk-Donsk Ordabók*. Tørshavn: Føroya Fródskaparfelag.

Jakobson, R., Fant, G. & Halle, M. (1952) *Preliminaries to speech analysis*. Cambridge, MA: MIT Press.

Jespersen, O. (1904) *Phonetische grundfragen*. Leipzig: B.G. Tuebner.

Joos, M. (1948) Acoustic phonetics. *Language*, 24(Suppl.), 1-137.

Joseph, B. & R. Janda (1988) The how and why of diachronic morphologization and demorphologization. In M. Hammond & M. Noonan (eds.), *Theoretical Morphology*. New York: Academic Press.

Jun, J. (1995) Place assimilation as the result of conflicting perceptual and articulatory constraints. *Proceedings of WCCFL 14*. 221-237.

Kawasaki, H. (1982) An acoustic basis for universal constraints on sound sequences. PhD dissertation, University of California, Berkeley.

Kenstowicz, M. 1994. Syllabification in Chukchee: a constraints-based analysis. In Alice Davison et al. (eds.), *Proceedings of the 4th annual meeting of the Formal Linguistics Society of the Midamerica*. University of Iowa, 160-81.

Kingston, John (1985) The phonetics and phonology of the timing of oral and glottal events. PhD dissertation, University of California, Berkeley.

Kohler, K. (1990) Segmental reduction in connected speech: Phonological facts and phonetic explanations. In W.J. Hardcastle & A. Marchal (eds.) *Speech Production and Speech Modeling* (pp. 69-92). Dordrecht: Kluwer Academic Publishers.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N. & Lindblom, B. (1992) Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255, 606-8.

Labov, W. (1994) *Principles of Linguistic Change: Internal Factors.* Oxford: Blackwell.

Lee, Y-S., Vakoch, D. & Wurm, L. (1996) Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research 25*, 527-542.

Liljencrants, Johan, and Björn Lindblom. (1972) Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language 48.4*, 839-862.

Lindblom, B. (1990) Explaining phonetic variation: A sketch of the H&H theory. In W.J. Hardcastle and A. Marchal (eds), *Speech Production and Speech Modeling*. (pp. 403-40) Dordrecht: Kluwer Academic Publishers.

Lockwood, W.B. (1955) *An Introduction to Modern Faroese*. Copenhagen: Ejnar Munksgaard.

Makashay, M. & Johnson, K. (1998) Surveying auditory space using vowel formant data. *The Sound of the Future: A Global View of Acoustics in the 21st Century, Proceedings 16th International Congress on Acoustics and 135th Meeting Acoustical Society of America* pp. 2037-8.

Martinet, André. (1955) *Economie des changements phonétiques*. Editions A. Francke AG Verlag Berne.

McCarthy, J. & Prince, A. (1993) Prosodic morphology. ms. UMass, Amherst and Rutgers University.

Mielke, J. (2001) A Perceptual Account of Turkish h-Deletion. *The Interplay of Speech Perception and Phonology*, OSUWPL, vol. 55.

Muller, J. (In preparation) The Phonetics and Phonology of Initial Geminates. PhD dissertation, Ohio State University.

Newton, B. (1972) *The Generative Interpretation of Dialect: A Study of Modern Greek Phonology*. Cambridge: Cambridge University Press.

Ohala, J. (1981) The listener as a source of sound change. In Masek, C.S., R.A. Hendrik, M. F. Miller (eds.), *Papers from the Parasession on Language and Behavior*: *Chicago Linguistics Society*. Chicago: CLS. 178-203.

Ohala, J. (1990) The phonetics and phonology of aspects of assimilation. In J. Kingston & M. Beckman (eds.), *Papers in Laboratory Phonology 1*. Cambridge: Cambridge University Press. 258-275.

Ohala, J. (1993) The perceptual basis of some sound patterns. In D.A. Connell & A. Arvaniti (eds), *Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence*. (pp. 87-94) Cambridge: Cambridge University Press.

Ovcharova, O. (1999) A perception-based study of consonant deletion in Turkish. Poster presented at the ICPhS Satellite Meeting, *The Role of Perception in Phonology*. San Francisco. July 1999.

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

Pike, K. (1943) *Phonetics*. Ann Arbor, MI: University of Michigan Press.

Polka, L. & Werker, J. (1994) Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance 20*, 421-435.

Prince, A. & Smolensky, P. (1993) Optimality theory. ms. Rutgers University and the University of Colorado at Boulder.

Rischel, Jørgen. (1972) Consonant Reduction in Faroese Noncompound Wordforms. In E. Scherabon Firchow, K. Grimstad, N. Hasselmo and W. O'Neil (eds*.), Studies for Einar Haugen*. The Hague: Mouton.

Rousselot, L'Abbe P.J. (1897-1901) *Principes de phonétique expérimentale*. 2 vols., Paris: H. Welter.

Seo, M. (2001) The realization of l/n sequences in Korean. *The Interplay of Speech Perception and Phonology*, OSUWPL, vol. 55.

Seo, M. & Hume, E. (2001) A comparative OT account of metathesis in Faroese and Lithuanian. In E. Hume, N. Smith & J. van de Weijer (2000)

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

Silverman, D. (1995) *Phasing and recoverability*. PhD dissertation, UCLA. New York: Garland Press.

Steriade, D. (1995) Licensing Retroflexion. ms. UCLA.

Steriade, D. (1997) Phonetics in phonology: The case of laryngeal neutralization. ms. UCLA.

Steriade, D. (1999) Perceptual factors in place assimilation. Paper presented at "The Role of Speech Perception Phenomena in Phonology," a satellite meeting of ICPhS99, San Francisco, CA.

Steriade, D. (2001) Directional asymmetries in assimilation: A directional account. In E. Hume & Johnson, K. (eds.), *The Role of Speech Perception in Phonology*. (pp. 219-250). New York: Academic Press.

Sweet, H. (1877) *A Handbook of Phonetics*. Oxford: Clarendon Press.

Tserdanelis, G. (2001) Manner dissimilation directionality in Greek. *The Interplay of Speech Perception and Phonology*, OSUWPL, vol. 55.

Winters, S. (2001) VCCV perception: Putting place in its place. In E. Hume, N. Smith & J. van de Weijer (2001).

Wright, R.A. (1996) Consonant clusters and cue preservation in Tsou. PhD dissertation, UCLA.

OSU Working Papers in Linguistics 55, 23-42

The Interplay of Perception and Phonology in Tone 3 Sandhi in Chinese Putonghua*

Huang, Tsan
huang@ling.ohio-state.edu

1. Introduction

The phenomenon of tone sandhi has long been noticed in the Chinese dialects (e.g. Chao 1948, 1968). Past studies allude that tone sandhis may be analyzed as processes leading to ease of articulation (Chao 1948, 1968; Cheng 1973; among others). But those analyses do not offer an explanation why one particular output is selected when other outputs are possible. Take for example the Tone 3 (or T3) sandhi process of standard mainland Mandarin Chinese, which is examined in this study. While simplification seems to be the correct analysis in that both native intuition and vocal physiology support the theory that it is hard to produce two dipping T3s in a row[1], it does not explain why T3 simplifies to T2 but not T1 or T4, the other two "simpler" tones in Mandarin Chinese. (See Section 2.1 for a detailed description of the tones in Standard Mandarin Chinese.)

In the present study, we hypothesized that the output of the third tone sandhi may be perceptually conditioned. In the words of Kohler (1990), Hura et al. (1992), and

---

* I would like to express my gratitude to my teachers Keith Johnson, Beth Hume, Michael Broe, and Dave Odden, and to members of the PiP seminar, Prof. San Duanmu at University of Michigan, Janice Fon, Martin Jansche, Georgios Tserdanelis, and all others in the OSU Linguistics Department who offered their help and support. Needless to say, all errors are mine.

[1] T3 surfaces as a low (falling) tone in most cases (see Section 2.1). The trigger of T3 sandhi seems to be the "lowness" of T3 (see also Shih 1997). C. C. Cheng (1968, cited in Shih 1997) reported that, even when Mandarin speakers code-switch between Chinese and English, a T3 would change into T2 when the first syllable of the following English word is unstressed – i.e. bearing a low tone:

/hao^{214} pro'fessor/ → [hao^{35} professor] "good professor"; but
/hao^{214} 'student/ → [hao^{21} student] "good student".

Steriade (2001), this may be a case of perceptually tolerated articulatory simplification. That is, T3 is selected as the sandhi form because it is perceptually more similar to T3 than T1 or T4 is, which makes the change relatively hard to detect in perception. Although there are quite a few studies on T3 sandhi and on the confusability of T3 and T2 in the literature, none of these studies compared the perceptual confusability of T3 and T2 with that of T3 and T1, nor with that of T3 and T4. Thus, none of them dealt with this issue directly. The present study tries to address this gap in the literature with a perceptual experiment of monosyllabic tonal pairs, which recorded both the "same"/"different" judgements made by the participants and their reaction time during response latency. We hope that the results of this study will help provide better understanding of this sandhi process and, more importantly, some insight into the interplay of perception and phonology (Hume & Johnson, 2001).

2. Background

2.1 The tone sandhi phenomenon in Standard Mandarin Chinese (or Putonghua)

In almost all Chinese dialects, underlying full tones may be modified under the influence of their tonal phonetic environment. This phenomenon is known as tone sandhi (see, for example, Chao 1948, Kratochvil, 1968). In this study, we looked at the tone sandhi phenomenon in standard mainland Mandarin Chinese, or Putonghua. This language has four lexical tones: level high [55, ˥], mid-rising [35, ˧˥], low falling-rising [214, ˨˩˦][2], and high falling [51, ˥˩], traditionally termed Tones 1, 2, 3, and 4, respectively. (The numbers in the square brackets indicate the pitch values of these tones on a five-level scale. And the drawings next to the numbers are graphic representations of those pitch values, termed Chao's tone letters; for a detailed discussion of the tone letter system, see Chao 1948 & 1968.) There is also a "fifth" tone, namely the inherent neutral tone, whose pitch value varies dependent on its preceding full tone.

As described in Chao (1948, 1968), the third-tone sandhi happens when T3 of Chinese Putonghua (– valued 214) becomes T2 (– valued 35) when immediately followed by another T3. Schematically, / ˨˩˦ ˨˩˦ /→ [˧˥ ˨˩˦].

It is claimed by many that morphological and syntactic boundaries are irrelevant here. Some examples are provided in (1) below:

(1) a. /hao^{214} mi^{214}/ → [hao^{35} mi^{214}] "good rice"
| |
modifier head noun

b. /mi^{214} hao^{214}/ → [mi^{35} hao^{214}] "The rice is good."
| |
subject predicate

---

[2] In Cheng (1973), T3 is described as having the value [315].

c. /mai^{214} mi^{214}/ → [mai^{35} mi^{214}] "to buy rice"
| |
verb object

Other phonetic variants of T3 include [21][3] and [214], the first of which appears before all full tones except T3 and the second of which appears in sentence-final position.

2.2 Perception and phonology

Phonologists have noticed the influence of perception on phonology from very early on (Trubetzkoy, 1969). In Jakobson, Fant, and Halle (1952), perceptual features are treated as primary (see also Jakobson and Halle, 1956). But the generative tradition of phonology since Chomsky & Halle (1968) centers around articulatory phonology. Now a revival of the view of the interplay of perception and phonology seems to be in process. People have been asking questions such as "To what extent do speech perception phenomena influence phonological system?" "To what extent does the phonological structure of language influence speech perception?" (Hume and Johnson, 2001).

Kohler (1990), Hura et al. (1992), and Steriade (2001) hold that phonological processes such as segmental reduction, deletion, and assimilation are perceptually tolerated articulatory simplification and that the direction of such processes is determined by perception. That is, these processes only take place when the output of such a change is found to be highly confusable with the input perceptually. For example, as place contrasts in sibilants are more salient than place contrasts in stops (Kohler 1990, Hura et al. 1992), the following patterns were found in the common retroflexization process in Sanskrit (Steriade 2001):

(2) Sanskrit apical assimilation in VC$_1$C$_2$(C$_3$)V sequence

a. same manner apical clusters: progressive assimilation
/VʈtV/ → [VʈʈV]
b. sibilant-stop clusters: progressive assimilation
/VṣtV/ → [VṣʈV]
c. stop-sibilant clusters: no assimilation
/VʈstV/ → [VʈstV]

In (2a) and (2b), the underlying dental stop /t/ surfaces as the retroflex [ʈ] as a result of assimilation to the place feature of the proceeding /ʈ/ or /ṣ/. In (2c), however, as place contrast between the dental /s/ and the retroflex /ṣ/ is more salient than the place contrast between /t/ and /ʈ/ -- thus, /s/ and /ṣ/ are less confusable than /t/ and /ʈ/ are -- assimilation does not happen and /s/ surfaces as [s].

---

[3] Some writers treat [21] as the underlying form of T3, as this is the most common surface shape. In fact, in the variety of Mandarin spoken in Taiwan, [21] surfaces in the final position. Sometimes, it may even surface in the final position in Putonghua.

If perception can influence segmental phonology in such a way, it seems reasonable to hypothesize that it may also have an impact on suprasegmental phonology and that it may play a role in the T3 sandhi process in Chinese Putonghua. Kiriloff (1969) found that syllables with Tone 2 and Tone 3 may be perceptually confusable. Fon (1997, MA thesis) and Fon et al. (1999 ms.) observed that both T2 and T3 have a dip in them (see also sources cited therein, e.g. Ho 1976, Yang 1995). Although the initial dip in T2 is usually ignored in phonological analysis, it has been shown to be perceptually important (see, for example, Gottfried & Suiter 1997). In a binary forced choice (T2 or T3) experiment where subjects were asked to label the tones whose pitch contours had been manipulated, Shen & Lin (1991) found that both the intrinsic duration of these tones and the turning points (i.e., where the rise starts in the contour) contribute to the confusability (see also Blicher et al. 1990, Chuang et al 1971, and Fon et al. 1999, ms.). Unfortunately, none of these studies can be used as convincing evidence to support our hypothesis that T3 sandhi is perceptually conditioned, as no comparison was made between the degree of confusability of T3 and T2, and that of any other tonal pairs in this language. In addition, we were also interested in finding out how phonology may influence listeners' perception of tones. Thus, the following experiment was designed to test our hypothesis directly.[4]

3. The experiment

3.1 Participants

Ten Chinese listeners (6 female, 4 male, average age 27.9) and thirteen American English listeners (7 female, 6 male, average age 21.8) were recruited from the Columbus campus of the Ohio State University (OSU). The Chinese listeners were graduate students (or their spouses) at OSU. Although a couple of them are not from the geographical regions where Mandarin is spoken, they were all fluent in the standard language due to their education background: they all received at least college education, and Putonghua is usually the language of instruction in most classrooms in mainland China. The English listeners were undergraduate students taking an introductory linguistics course at OSU. They were all native speakers of Ohio English. The Chinese were paid for their participation in the experiment, whereas the Americans earned extra credit points for their Linguistics 201 class.

---

[4] There is historical evidence that the T3 sandhi may have happened 700 years ago when the T3 contour could have been completely different from its current shape. This process may have been grammaticized in the Mandarin dialects and carried down to the present day. But it is not the case that all current Mandarin dialects preserve this sandhi rule. For example, it is no longer in my dialect, Rugaohua, a Jianghuai Mandarin dialect. We may speculate that a certain generation of Rugaohua speakers gave a second thought of the sandhi process and, due to a change in the tonal shape of T3, could not see why it was necessary to have the process. So, they decided to drop it. On the same basis, maybe speakers of Beijing Mandarin, the base language for Putonghua, did a similar analysis. But since it was still necessary to have the sandhi, it was reinvented. And the fact that Putonghua speakers apply the rule even when code-switching (see Footnote 1) is evidence that it is a productive synchronic phonological process. Thus, a synchronic analysis seems to be justified.

The American English listeners were included to see if T3 and T2 of Chinese Putonghua in the T3 sandhi environment share some property that makes them confusable to non-native listeners. We assume that, if there is no effect of the listener's native phonology on perception, phonetic universality should allow everybody to act the same. Previous studies have shown that it is feasible to include "non-native" listeners. Kiriloff (1969) found that, when asked to ignore the segmental part of the syllable and focus on the tones, non-native speakers' performance was quite good (an average of 17.5 correct identifications out of 20 stimuli, or 87.5%)[5]. On the other hand, if we do find a difference between native and non-native listeners' performance, it might help us gain insight into how phonology may influence perception.

3.2 Stimuli

The stimuli were constructed from recordings produced by a female Putonghua speaker in disyllabic nonsense sequences with 15 tonal combinations (– that is, all possible pairs except T3T3 which does not occur in natural speech. The segmental make-up of these recorded sequences was kept constant and was always /bao-fang/. The typical stress pattern[6] for a disyllabic full-toned sequence was used to get the appropriate pitch contours of the four tones in the environment where T3 sandhi occurs. Ten (10) randomized lists of these sequences were recorded. The original recordings were done in a sound-proof booth in the phonetics laboratory at the OSU Linguistics Department. The speaker read from the afore-mentioned 10 randomized lists and was recorded with a head-mounted microphone (Shure SM10A model) and a DAT recorder.

The recordings were digitized at 22,050Hz with 16 bit samples. The first syllable (i.e. /bao/) was cut from these sequences. The seven (7) best productions of these /bao/ syllables for each of the four tones (as determined subjectively by the author) were then chosen to splice the test stimulus pairs, while three (3) other productions were used in the training session.

Figures 1, 2, 3 and 4 show pitch tracks of these stimulus tones. Note that only the first "half" of the T3 tonal contour is realized, which is typical of T3 in this non-final position.

---

[5] Gottfried & Suiter (1997) did find some degree of performance difference in native versus non-native listeners. But it was a more difficult identification task and their four conditions (– namely, initial only, center only, silent center, and final only) all involved cutting off some part of the syllable. Lee et al. (1996) also found a small native speaker advantage.

[6] Yip (1980) and Zhang (1988) mentioned that T3 sandhi is conditioned by the metrical pattern of the utterance and that the T3 that undergoes the sandhi has to be in the weak branch of the stress matrix, i.e. the syllable bearing the sandhi tone must not be linked to a node at the highest/primary stress level. Thus, it is predicted that the first T3 in /xiao3jie3/ 'miss' (with a weak-strong pattern) would undergo the T3 sandhi and surface as [xiao2jie3], whereas that in /jie3jie3/ 'older sister' (with a strong-weak pattern) would not, yielding the surface form [jie3jie0]. But see Shih (1997), where she holds that stress does not play a role in the sandhi processes. We chose not to commit ourselves to any particular phonological framework here and tried to take into consideration all possible conditions for this sandhi process.

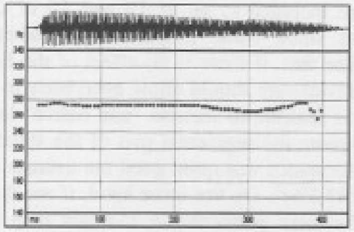


Figure 1. Pitch track of stimulus T1

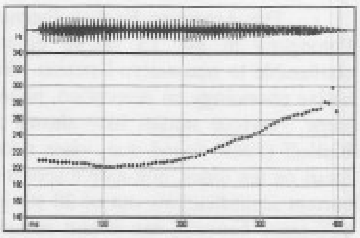


Figure 2. Pitch track of stimulus T2

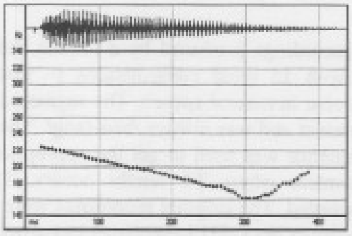


Figure 3. Pitch track of stimulus T3

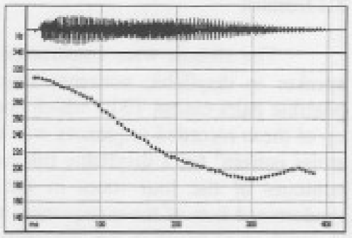


Figure 4. Pitch track of stimulus T4

The test session consisted of 7 sections, each of which contained 20 stimulus pairs. Thus, all participants listened to 20 × 7 = 140 pairs of the form /bao-bao/. The 20 pairs in each section included 12 different pairs (see the checked boxes, marked with x, in Table 1 below) and 8 identical pairs (i.e., each of the 4 identical pairs in the empty boxes in Table 1 was repeated twice in any of the test sections)[7]. Only the results of different pairs were analyzed. The identical pairs were included as fillers.

Table 1. Tonal combinations to be tested

	T1	T2	T3	T4
T1		x	x	x
T2	x		x	x
T3	x	x		x
T4	x	x	x	

---

[7] Each identical pair contains two repetitions of the same .wav file. Thus, the experiment emphasized psychoacoustic discriminability of tones.

3.3 Method

A discrimination task was used. Participants were tested in front of a computer one at a time in a sound-proof booth. The stimuli were presented to them through headphones, using the Micro Experimental Lab (MEL) program installed on a PC. While each stimulus pair was played with a 2000ms inter-pair interval, the words "same" and "different" were also displayed visually on the left and right sides of the computer screen, respectively. The participant input responses by pressing the "same" or "different" buttons on a button-box connected to the PC. Participants were asked to use their left and right index fingers to press the "same" and "different" buttons, respectively. Instructions, both given orally by the experimenter during the training session and displayed visually on the PC screen during the test session, asked the participant to respond as accurately and as quickly as possible. After each correct "same"/"different" judgment was made, the reaction time (RT) would appear on the screen as feedback; otherwise, the screen would display the words "wrong response". This made it clear to the subjects what a good performance was: one with shorter reaction time and fewer errors.

Both the "same-different" judgement accuracy and RT were recorded as experiment results. The measurement for RT started from the onset of the second syllable of the stimulus pair.[8]

3.4 Predictions

We predicted that, if T2 and T3 are more confusable, i.e. closer to each other in the perceptual space, then (i) people would make more mistakes when asked to tell whether they are the same or different, and (ii) people would take longer to make the judgment, that is, the shorter the perceptual distance, the longer the reaction time (RT) (see, for example, Shepard et al. 1975, Shepard 1978, Takane et al. 1983, Nosofsky 1992, although these authors disagree on what exactly the relationship between perceptual distance and RT is and how the transformation between them should be done. We shall postpone the discussion on these issues until Section 5.)

4. Results

The results basically support our hypothesis that T2 and T3 are perceptually more confusable. In terms of the mistakes that listeners made, there was no statistically significant difference between the tonal pairs, as error rates were very low in the responses of both the Chinese and English groups. But the pairs T2-T3 and T3-T2 did attract more errors than other pairs. Table 2 shows mean RT values of correct "different" responses and error rate in percentage for each non-identical tonal pair.[9]

---

[8] The mean duration measurements for all stimulus syllables are: T1=375.9ms, T2=414ms, T3=389.5ms., and T4=387.8ms. Such differences do not seem to be big enough to affect the RT measurements, as the adjusted RT data (with the duration of the second syllable subtracted) show a similar pattern.

[9] The median RT data– with or without the duration of the second syllable – reveal a pattern similar to the mean RT data (see Appendix I).

Table 2. Mean RT (ms) for correct "different" responses and percentage of errors

TONAL PAIRS	T1/T2		T1/T3		T1/T4	
	T1T2	T2T1	T1T3	T3T1	T1T4	T4T1
Chinese	568.9(4%)	556.7(7%)	572.8(3%)	584.2(6%)	602.4(4%)	572.6(4%)
English	558.1(1%)	671.5(1%)	516.6(2%)	556.1(2%)	606.8(5%)	594.0(2%)
TONAL PAIRS	T2/T3		T2/T4		T3/T4	
	T2T3	T3T2	T2T4	T4T2	T3T4	T4T3
Chinese	699.4(11%)	667.4(7%)	512.1(0%)	583.2(4%)	542.9(0%)	547.0(4%)
English	748.4(16%)	663.5(13%)	615.1(11%)	568.6(3%)	591.0(5%)	624.3(2%)

We can see that the Chinese listeners scored 62 correct responses out of all 70 T2T3 stimulus pairs (= 7 sections × 10 participants) with an error rate of 11% and 65 correct responses out of all 70 T3T2 stimuli with an error rate of 7%. On the other hand, the English listeners scored 76 correct responses out of all 91 T2T3 stimulus pairs (7 sections ×13 participants) with an error rate of 16% and 79 correct responses out of all 91 T3T2 stimulus pairs with an error rate of 13%.

Although error rates were too low to be significant, the RT data turn out to be very informative. The graphic representation in Figure 5 may help us see clearly what the RT values for the T2/T3 pairs are like compared to other tonal pairs. The points on the X-axis represent the non-identical pair types, and the numbers along the Y-axis show reaction time in milliseconds. The solid line represents the Chinese listeners' data, while the dashed line the English listeners'.

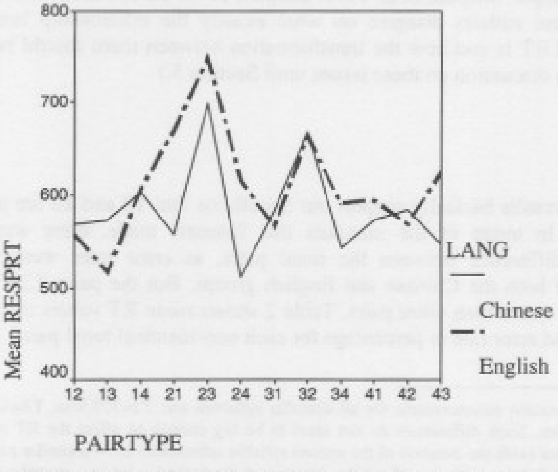


Figure 5. Mean RTs (in milliseconds) for the correct responses

As we expected, the slowest RT for the Chinese participants was found with the T2/T3 pairs (the two peaks in the solid line in Figure 5), with the RT for T2T3 being even longer than that for T3T2. One possible explanation for this pattern might be that, as this sequence is identical to the sandhi output where the T3 and T2 distinction is neutralized into T2 for the Chinese listeners, they are biased because of their native phonology. However, we see a similar picture with the American listeners: the RT for T2T3 is also the longest, and the RT for T3T2 is the third longest of all pairs (shorter than that for T2T1), which may be seen as evidence for saying that the way the Chinese reacted is not completely due to their native phonology: phonetically, there exists some universal perceptual distance between these tones for both the native and non-native listeners.

5. Analyses

5.1 Repeated measures analysis of variance and Independent-Samples T test

A repeated measures analysis of variance (ANOVA) was performed on the RT data of the listeners' correct "different" responses, with all 12 non-identical tonal pairs (i.e. T1T2, T2T1, T1T3, T3T1, T1T4, T4T1, T2T3, T3T2, T2T4, T4T2, T3T4, T4T3) as within subject variables, and language as between subject variable. No significant result was found between listener/langauge groups. But there was a significant effect with pair types, sig.[F(9.321, 1118.461) = 353343.626, $p < .001$, $\tilde{\eta}^2 = .106$]. There was also a significant effect with the interaction of language and pair, sig.[F(9.321, 1118.461) = 103801.579, $p < .001$, $\tilde{\eta}^2 = .034$].

A post-hoc test of pair-wise comparison, which compares the raw RT values for each pair against all other pairs within the same listener group, shows that pairs T2T3 and T3T2 are significantly different from the other pairs ($p<.05$) for both groups of listeners. For the Chinese listeners, pairs T2T3 and T3T2 are totally different things from the other pairs($p<.05$). Pair T2T3 was found to be significantly different from all pairs except pairs T1T4 and T3T2. Pair T3T2 is significantly different from all pairs except pairs T4T1, T4T2, T3T1, T1T4 and T2T3, showing a possible effect of phonology on perception, as no difference was found between any two of the other pairs. Interestingly, pair T2T3 was found to be significantly different from three more pairs than pair T3T2, which seems to make the effect of phonology even stronger, as T2T3 is the output of the T3 sandhi.

The English listeners, on the other hand, found pair T1T3 to be the least confusable and significantly different from all other pairs except T3T1 and T1T2 ($p<.05$). They also found pairs T2T3, T2T1 and T3T2 to be the most confusable ($p<.05$). This pattern seems to be more phonetic than phonological, as the English listeners seem to rely more on the phonetic shapes of these tones when making their decisions. If the pitch value of the ending point of the first syllable matches that of the starting point of the second syllable, the English listeners found them to be confusable. This behavior is different from that of the Chinese listeners' who found only the T2 and T3 pairs to be confusable. Again, this seems to provide more evidence that the Chinese listeners' perception is influenced by their native phonology.

An Independent-Samples T test was also performed on the RT data, with language as the grouping variable and RT as the test variable. The outliers in both listener groups were taken out. The results are as follows:

Table 3. Independent-Samples T test

Pair type			*meanRT Chinese*	*meanRT English*
T1T2	t(137) = 1.175,	p = .242	530.9	508.5
***T2T1**	**t(133) = -4.677,**	**p < .001**	**514.1**	**608.6**
***T1T3**	**t(138) = 2.41,**	**p = .017**	**543.5**	**496.1**
T3T1	t(138) =.665,	p = .507	550.8	536.4
T1T4	t(138) = -1.784,	p = .077	554.5	591.8
T4T1	t(138) = -1.359,	p = .176	543.8	572.6
T2T3	t(120) = -1.022,	p = .309	663.3	687.2
T3T2	t(135) = .345,	p = .731	662.7	654.4
***T2T4**	**t(128) = -4.667,**	**p < .001**	**486.8**	**578.3**
T4T2	t(138) = 1.098,	p = .274	575.6	551.4
***T3T4**	**t(141) = -3.02,**	**p = .003**	**513.5**	**579.7**
T4T3	t(135) = -1.688,	p = .094	532.9	568.9

The pairs with a significant between-language-group effect ($p<.05$) have been bold-faced and indicated with an asterisk in Table 3. The mean RT values (in milliseconds) make the pattern even more interesting. In general, the Chinese listeners did better than the English listeners. For some pairs that the English listeners found confusable, i.e., **T2T1**, **T2T4**, and **T3T4**, the Chinese listeners did not seem to have more difficulty distinguishing them at all. For pair **T1T3**, which the English listeners found to be the least confusable, the Chinese listeners did not seem to see it as an easier pair than the other pairs.

5.2 Multidimensional scaling (MDS)

In a sense, the RT data obtained reflect similarity between the tones: RT values increase as the tones get more similar. We need to find a way to transform RT (or similarity) values into perceptual distances (or dissimilarity). As no direct measurements can be made for either the physical or the perceived distances between these tones, RT measurements were converted into perceptual distances based on the assumption that the closer two "objects" are in the perceptual space, the longer it takes for people to tell them apart (see, for example, Shepard et al. 1975, Shepard 1978, Takane et al. 1983, Nosofsky 1992).

How exactly RT reflects perceptual or physical distance is still a question begging to be answered. In our case here, we would probably also need to take into account the influence of phonology on perception as well as the characteristics of the stimuli (i.e., the phonetic characteristics of the tones). Nevertheless, several approaches have been proposed to convert RT into distances. Curtis et al. (1973), Shepard et al. (1975), and

Shepard (1978) advocate for the reciprocal function. Their argument for that is, with correct "different" judgments, reaction time values have been found to be nearly reciprocal of distance values. Takane and Sergent (1983) and Nosofsky (1992) suggest the log normal function. Takane and Sergent's (1983) reason for choosing the log normal function over the reciprocal function is that it is not the case that correct "same" RT is reciprocal to distance. As only the RTs of correct "different" judgments were of interest in the present study, the choice of the reciprocal approach seems to be justified. In addition, this approach is well-supported by previous research.

In fact, we did make use of the log normal approach and found the MDS results to be very similar to the reciprocal approach. In addition to the reciprocal and the log normal functions, which turn linearly related RTs into a non-linear distribution of distances, we also tried a linear approach suggested by Michael Broe (personal communication). RTs were rescaled using the formula (Observed RT/Maximal RT) so that they now distribute along a scale of 0~1. Then, the 0~1 RT scale were turned into a 0~1 distance scale by subtracting the new "RT" values from 1 (i.e. distance = 1 - Observed RT/Maximal RT]). Again, the MDS results are surprisingly similar to the reciprocal approach. The calculated distances (=1/RT) are reported in Table 3 below.[10]

Table 4. Distances derived from RTs for correct responses for all different tonal pairs. Values are 10^3 times the original reciprocal values.

	T1T2	T1T3	T1T4	T2T1	T2T3	T2T4
Chinese	1.895	1.903	1.853	1.981	1.54	2.121
English	1.977	2.079	1.766	1.617	1.468	1.778
	T3T1	T3T2	T3T4	T4T1	T4T2	T4T3
Chinese	1.871	1.651	2.014	1.905	1.874	1.95
English	1.918	1.570	1.860	1.800	1.884	1.800

The mean distance for tonal pairs involving the same tones was taken to be the distance between these tones in the MDS analysis. For example, the mean value for the T1T2 and T2T1 distances was taken to be the distance between T1 and T2.

Table 5. Averaged distances for tonal pairs involving the same tones.

	T1/T2	T1/T3	T1/T4	T2/T3	T2/T4	T3/T4
Chinese	1.938	1.887	1.879	**1.596**	1.998	1.982
English	1.797	1.998	1.783	**1.519**	1.831	1.830

As shown in Table 4, the distance between T2 and T3 was found to be the shortest by both the Chinese-speaking and the English-speaking participants.

---

[10] Distances are averages of all reciprocal values of the original RT data, not the mean RT.

These data were then analyzed as dissimilarities using monotonic MDS, with the 2-dimensional MDS taken as the default.[11] And evaluation of a 2-dimensional scaling is very satisfactory: stress for the Chinese listeners' data is 0.0 and that for the English listeners' data 0.00174; values of the proportion of variance (RSQ) are 1.0 and 0.99 for these data sets, respectively[12]. A cluster tree analysis for both the Chinese and the English (Figure 8) listeners' data revealed that the groupings of the tones are exactly the same for both groups of listeners: T3 is grouped with T2, and T4 is grouped with T1.

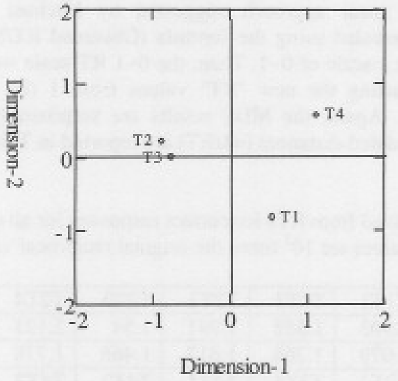


Figure 6. Two-dimensional scaling for Chinese listeners' RT data. [stress = 0.00, Proportion of variance (RSQ) = 1.0]

---

[11] The error data in percentage (see Table 2) was also analyzed as similarity data; that is, it was assumed that the more similar the two objects are, the higher the error rate is. The MDS results turned out to be very similar to the distance analysis (see Appendix II).

[12] As we only have four (4) objects in the analysis, the stress is low in a 1-dimensional analysis, too. But it does show some improvement for the English listeners' data when we change the number of dimensions from 1 to 2.

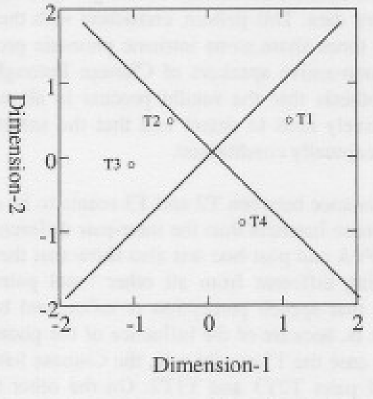


Figure 7. Two-dimensional scaling for the English listeners' RT data. [stress = 0.001729, Proportion of variance (RSQ) = 0.99902]

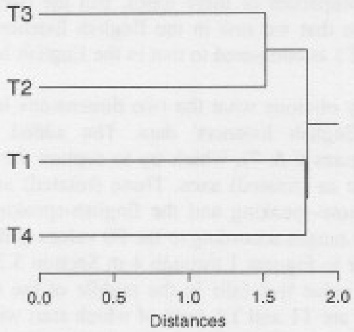


Figure 8. Groupings of the four Putonghua tones in Chinese/English data.

6. Discussion

Combining the information of the distances between the tones along the two dimensions in Figures 6 and 7 and the information of their groupings in Figure 8, we have a very telling picture. First, notice that T2 and T3 are grouped together in both the Chinese and the English listeners' data. This pattern, consistent with the pattern shown in Figure 5, shows that these two tones share some intrinsic phonetic property that can be perceived by both native and non-native speakers of Chinese Putonghua. This may be seen as evidence for our hypothesis that the sandhi process is allowed to take place because such a change is relatively hard to detect and that the selection of T2 as the output of the sandhi may be perceptually conditioned.

Second, the perceived distance between T2 and T3 seems to be smaller relative to the other tonal pairs for the Chinese listeners than the inter-pair difference for the English listeners.[13] Recall that the ANOVA and post-hoc test also show that the Chinese listeners treated the T2/T3 pairs as being different from all other tonal pairs. These findings provide evidence for the view that speech perception is influenced by phonology (see Hume and Johnson 2001). That is, because of the influence of the phonological structure of their native language, in this case the T3 sandhi rule, the Chinese listeners were highly biased toward the similarity of pairs T2T3 and T3T2. On the other hand, the English listeners were dealing with mostly the phonetic characteristics of the tones. (Basically, as was mentioned before, if the ending pitch of the first syllable matches the starting pitch value of the second syllable, for example T2T1, or if two tones in a pair share a similar tonal contour, for example T3T4, the pair was found to be more confusable.) If there is no phonological effect in addition to familiarity, the MDS for the Chinese listeners would look different: one would expect the Chinese listeners' perceived distance between any tonal pair to be longer than the English listeners' due to familiarity. The distance between T2 and T3 might still be short for the Chinese listeners relative to the other tonal pairs because of the intrinsic properties of these tones. But the overall MDS pattern should look similar to the pattern that we saw in the English listeners' data but with a longer distance between T2 and T3 as compared to that in the English listeners' MDS.

It may not be very obvious what the two dimensions in the MDS configuration are, especially in the English listeners' data. The added (diagonal) lines in the configuration figures (Figures 6 & 7), which try to capture the information given in the cluster trees, may be seen as (rotated) axes. These (rotated) axes show that, along one dimension, both the Chinese-speaking and the English-speaking listeners have divided the tones into two register ranges according to the F0 values at the beginning of the tones. (For the pitch tracks, refer to Figures 1 through 4 in Section 3.2.) Thus, T2 and T3, both of which start with a F0 value that falls in the middle of the speaker's pitch range, are grouped together. And so are T1 and T4, both of which start with a F0 value that falls in the upper level of the speaker's pitch range. Along the other dimension, the tones seem to have been grouped together according to the characteristics of their pitch contours. Thus,

---

[13] The absolute T2/T3 distance value for the Chinese listeners is longer than that of the English listeners, which may be another effect of phonology on perception, as they perceive tones better than the English listeners who speak a non-tone language.

T1 is set apart from T2, T3 and T4 because T1 is a level tone with a static pitch level while the other three tones are contour tones with dynamic pitch movements. In other words, MDS reveals that both the tonal contour and the starting pitch point are important cues for tonal perception.

The patterns in the result of the Independent-Samples T test are also very revealing. It provides evidence for the view that the Chinese listeners treat each tonal contour as an indivisible unit (see also Jansche 1999 ms.), as neither the phonetic pitch level of the starting or ending point of the contour nor the similarity in tonal contours seems to contribute much to the confusability or distinctiveness of the tones. Unlike the English listeners who were using these characteristics of the tones as important phonetic cues to distinguish the tones, the perception of the Chinese listeners seemed to be independent of these cues to a certain extent. In other words, the Chinese listeners' phonological knowledge seems to have "transcended" their phonetic knowledge. The fact that, in one case, with pair **T1T3**, the Chinese even "suffered" from their phonological knowledge – i.e., failed to use the phonetic cues as effectively as the English listeners did – also shows that tonal perception is not influenced merely by familiarity; otherwise, one should expect the Chinese listeners to do better in all cases.[14] They did not. As can be seen in Table 3, they treated pair **T1T3** as an "average" pair. They performed almost as poorly as the English listeners did on pairs T2T3 and T3T2. We are not denying that familiarity played a role here, as the Chinese did better in general. Familiarity might have interacted with phonology, as the Chinese listeners did perceive the T2T3 and T3T2 pairs slightly better than the English listeners did, although, given their native phonology, one might not have been totally surprised should the Chinese have appeared to be "blind" to the distinction between T2 and T3. If one takes a second look at the error data shown in Table 2, he may find a similar pattern there. That is, the mistakes that the English listeners made were more phonetically-driven, while the mistakes in the Chinese listeners' data point to the influence from the Chinese tonal phonology.

7. **Conclusion**

To sum up, we examined Chinese Putonghua tones with a "same"/"different" discrimination task in this study. Distances between the tones were derived from the reaction time data by the reciprocal function. The MDS analysis on both the RT data and the error data shows that T3 is perceived as closer to T2 than it is to T1 or T4, which supports our hypothesis that T2 is chosen as the sandhi form for T3 because such a change is perceptually tolerated. The MDS analysis also shows that phonology influences speech perception, as the Chinese listeners perceived an even shorter relative distance between T3 and T2 than the American English listeners did. This is further supported by the post-hoc test result which shows that the Chinese listeners found only the T2/T3 pairs to be significantly different from all other tonal pairs. The Independent-Samples T test also reveals a phonological effect on perception as the Chinese appeared to have treated

---

[14] In two experiments involving more complicated tasks of tonal discrimination, Lee et al. (1996) found that native tone language speakers did better than speakers of a different tonal language, who, in turn, did better than nontone language speakers.

each tonal contour as an indivisible unit and ignored some important phonetic cues. It is evident from these results that there clearly is an interplay between perception and phonology and that the two may interact to constrain changes in the phonological structure of the language.

REFERENCES

Blicher, D. L., Diehl, R. L., & L. B. Cohen, 1990. Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: evidence of auditory enhancement. *Journal of Phonetics*, 18 (1), 37-49.

Chao, Yuen Ren. 1948. *Mandarin Primer.* Cambridge, MA: Harvard University Press.

----------. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

Chen, Matthew Y. 2000. *Tone Sandi patterns across Chinese dialects.* Cambridge, UK: Cambridge University Press.

Cheng, Chin-chuan. 1973. *A Synchronic Phonology of Mandarin Chinese.* The Hague: Mouton.

Chomsky, N. & M. Halle, 1968. *The sound pattern of English.* New York, Harper & Row.

Chuang, C.K., S. Hiki, T. Sone, & T. Nimura, 1971. The acoustic features and perceptual cues of the four tones of standard colloquial Chinese. 7th International Congress on Acoustics, Budapest 1971.

Curtis, D. W., Paulos, M. A., & S. J. Rule, 1973. Relation between disjunctive reaction time and stimulus difference. *Journal of Experimental Psychology*.

Fon, Y-J. 1997. What are tones really like? – An acoustic-based study of Taiwan Mandarin tones. Master's thesis. National Taiwan University.

Fon, J., Chiang W.-Y., & H. Cheung, 1999. Production and Perception of T2 and T3 in Taiwan Mandarin. Ms.

Gottfried, T. L. & T. L. Suiter, 1997. Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics* 25: 207-231.

Hume, E. & K. Johnson, in press. A Model of the Interplay of Speech Perception and Phonology. In E. Hume & K. Johnson (eds.) *The Role of Speech Perception in Phonology* (pp 3-26). New York: Academic Press.

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

Jakobson, R., Fant, G., and M. Halle, 1952. *Preliminaries to speech analysis*: the distinctive features and their correlates. Cambridge: Acoustics Laboratory, Massachusetts Institute of Technology.

Jakobson, R. and M. Halle, 1956. *Fundamentals of language*. Gravenhage: Mouton & Co..

Jansche, Martin. 1999. The Mandarin third tone sandhi across dialects. The Ohio State University, ms.

Kiriloff, C. 1969. On the Auditory perception of Tones in Mandarin. *Phonetica* 20: 63-67.

Kohler K. 1990. Segmental reduction in connected speech: Phonological facts and phonetic explanations. In W.J. Hardcastle & A. Marchal (eds.) *Speech Production and Speech Modeling* (pp. 69-72). Dordrecht: Kluwer Academic Publishers.

Kratochvil, Paul, 1968. *The Chinese Language Today: features of an emerging standard.* London: Hutchinson.

Lee, Y-S, Vakoch, D.A., & L. H. Wurm, 1996. Tone Perception in Cantonese and Mandarin: A Cross-Linguistic Comparison. *Journal of Psycholinguistic Research.*

Nosofsky, R. M. 1992. Similarity scaling and cognitive process models. *Annual Review of Psychology*, 1992, 43, 25-53.

Shen, X. S. & Lin, M. 1991. A perceptual study of Mandarin Tone 2 and 3. *Language and Speech*, 34 (2), 145-156.

Shepard, R. N. The circumplex and related topological manifolds in the study of perception. In Shye, S. (Ed.), *Theory construction and data analysis in the social sciences.* San Francisco: Jossey-Bass, 1978.

Shepard, R. N., Kilpatric, D. W., & J. P. Cunningham, 1975. The internal representation of numbers. *Cognitive Psychology*, 1975, 7, 82-138.

Shih, Chi-lin, 1997. Mandarin third tone sandhi and prosodic structure. In Wang Jialing and Norval Smith (eds.) *Studies in Chinese Phonology.* Berlin & New York: Mouton de Gruyter, 1997.

Steriade, D. 2001. A perceptual account of directional asymmetries in assimilation and cluster reduction. In E. Hume & K. Johnson (eds.) *The Role of Speech Perception in Phonology* (pp 219-250). New York: Academic Press.

Takane Y. & J. Sergent, 1983. Multidimensional Scaling Models for Reaction Times and Same-different Judgments. *Psychometrika*, 1983, Vol. 48, No. 3, 393-423.

Trubetzkoy, N. S. 1969. *Principles of phonology.* C. Baltaxe (translator). Berkeley and Los Angeles: University of California Press, 1969.

Yip, M. 1980. The Tonal Phonology of Chinese. Ph.D. dissertation. MIT. Distributed by the Indiana University Linguistics Club.

Zhang, Zheng-sheng, 1988. Tone and Tone Sandhi in Chinese. OSU Ph.D. dissertation.

Appendix I – Multidimensional Scaling of the median data

Table 6. Median RT values for correct responses (in milliseconds)

tonal pairs	T1/T2		T1/T3		T1/T4	
	T1T2	T2T1	T1T3	T3T1	T1T4	T4T1
Chinese	546	509	536.5	540	551	554
English	504	621	493	520	575.5	566
tonal pairs	T2/T3		T2/T4		T3/T4	
	T2T3	T3T2	T2T4	T4T2	T3T4	T4T3
Chinese	**656**	**675**	484	559	502	519
English	**698**	**627**	587	547	583	590

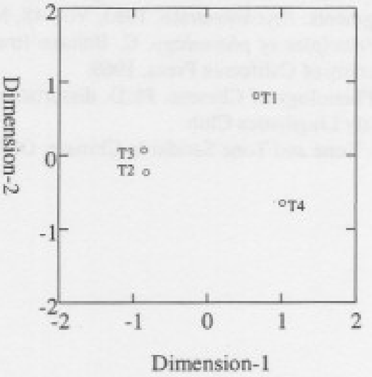


Figure 9. MDS analysis on median reaction time data of the Chinese listeners. [stress = 0.00, Proportion of variance (RSQ) = 1.0]

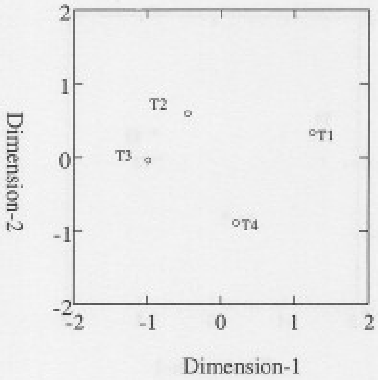


Figure 10. MDS analysis on median reaction time data of the English listeners. [stress = 0.00, Proportion of variance (RSQ) = 1.0]

Appendix II – Multidimensional Scaling of the error data

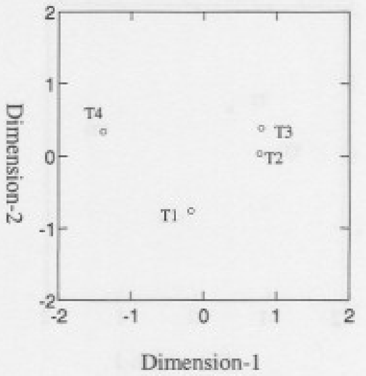


Figure 11. MDS analysis on the error data of the Chinese listeners. [stress = 0.00, Proportion of variance (RSQ) = 1.0]

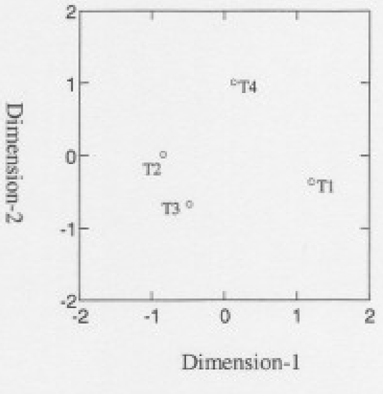


Figure 12. MDS analysis on the error data of the English listeners. [stress = 0.00, Proportion of variance (RSQ) = 1.0]

OSU Working Papers in Linguistics 55, 43-69.

A Perception-based Study of Sonorant Assimilation in Korean*

Misun Seo
mseo@ling.ohio-state.edu

1. Introduction

Speech perception phenomena have been drawn on by researchers in the area of phonological theory to elucidate synchronic phonological processes such as neutralization (Steriade 1995, 1997), consonant/consonant metathesis (Hume 1998, 2001), place assimilation (Jun 1995), etc. It has also been found that listeners' perceptual abilities are influenced by their native language experience (e.g. Hume et al. 1999).

In this paper, this bidirectional interplay between speech perception and phonology is investigated further. The influence of speech perception is examined as a possible means of understanding sonorant assimilation in Korean. Two types of sonorant assimilation are attested in Korean. First, when a nasal /n/ is adjacent to a lateral, the nasal is lateralized, as in /non + li/ → [nolli] 'logic', /səl + nal/ → [səllal] 'New Year's day'. Korean is not the only language with n-lateralization in n/l sequences. This process is attested in a wide range of languages such as Klamath, Ponapean, Toba Batak, Leti, Teralfene dialect of Flemish, Rendille, Somali, and Udi. In different languages such as Tatar and Yakut, l-

---

* This work was supported by Graduate Students Travel Grants from International Studies at the Ohio State University. My thanks to Elizabeth Hume, Keith Johnson and the PIP discussion group at the Ohio State University for their valuable comments and suggestions on this research. I would also like to thank Professor Gösta Bruce for his assistance with the perception experiment run in Lund, Sweden. Of course, any errors are mine.

nasalization is observed in n/l sequences. Second, when a lateral is preceded by a noncoronal nasal /m/ or /ŋ/, the lateral is nasalized, as can be seen in /sam + lyu/ →[samnyu] 'third rate', /yəŋ + li/ →[yəŋni] 'profit'. L-nasalization in /ml/ is also found in Tatar and Yakut[1]. However, unlike n-lateralization in n/l sequences, l-nasalization in /ml/ and /ŋl/ sequences is a rather uncommon process cross-linguistically and to our knowledge, there is no language with nasal-lateralization in those sequences.

Four important questions can be raised concerning n-lateralization and l-nasalization. First, why is n-lateralization or l-nasalization in n/l sequences a cross-linguistically common process while l-nasalization in /ml/ and /ŋl/ is not? Second, with respect to the n/l sequences, why are both n-lateralization and l-nasalization attested in different languages? Third, why is l-nasalization rather than nasal-lateralization observed in the /ml/ and /ŋl/ sequences in Korean? In this paper, we propose that these questions can be answered by recourse to speech perception. Following Koher (1990), Hura et al. (1992) and Steriade (2001), we hypothesize that both n-lateralization and l-nasalization occur in n/l sequences since both are perceptually licensed changes. We hypothesize that the change of /l/ to [n] rather than the change of /m/ or /ŋ/ to [l] occurs in /ml/ and /ŋl/ sequences since the former is a perceptually less noticeable change. Another important question to be raised is why the change of /n/ to [l] is preferred to the change of /l/ to [n] in Korean, given that both n-lateralization and l-nasalization are perceptually allowed changes and thus both are attested in n/l sequences cross-linguistically? We propose that listeners' perceptual abilities are not the only factors shaping the phonological patterns of languages and that this question can be answered by taking into account articulation as well as perception.

The influence of phonology on speech perception is also explored in this paper by comparing Korean listeners' perceptual abilities with those of Moroccan Arabic and Swedish listeners, whose native languages show different phonological patterns in the nasal/liquid sequences from Korean. According to the P-map hypothesis (Steriade, 2001), listeners' perceptual abilities are the same regardless of the native language of a listener and sound patterns found in that language. This paper provides a test of this hypothesis through perception experiments involving Korean, Moroccan Arabic and Swedish listeners.

Our results suggest that l-nasalization, i.e. the change of /ml/ and /ŋl/ to [mn] and [ŋn], respectively, is driven by perceptual considerations. Our results from Korean listeners show that the change of /l/ to [n] is perceptually less noticeable than nasal-lateralization in the /ml/ and /ŋl/ sequences, as we expected. Our results show no significant difference in n/l sequences between the change of /l/ to [n] and that of /n/ to [l], suggesting that both n-lateralization and l-nasalization are perceptually allowed changes.

---

[1] Tatar: /khanım + lÃr/ → [khanımnar] 'ladies' (Poppe 1963) (/Ã/ is the archiphoneme of [ä] and [a].)
Yakut: /olom + lAr/ → [olomnor] 'fords' (Krueger 1962) (/A/ is the archiphoneme of [a], [e], [o], and [ö].)

This paper is organized as follows. In section 2, assimilation patterns in /nl/, /ln/, /ml/ and /ŋl/ sequences found in Korean are given. In section 3, the realization of n/l sequences in other languages is presented. We discuss the possible motivation behind the patterns of assimilation in nasal/liquid sequences and posit a hypothesis based on that in section 4. In section 5, the realization of nasal/lateral sequences in Moroccan Arabic and Swedish is given. In section 6, an outline of the perception experiment, which is designed to test the hypothesis posited in section 4, is given, and the results of the perception experiment and discussion are given in section 7. The results and discussion of the linear regression analysis, which was done to investigate which language-universal and language-specific cues are involved in speech perception and how much their influence is, are given in section 8. In section 9, we provide an account of sonorant assimilation patterns in Korean based on our perception experiment results.

2. Sonorant Assimilation in Korean

In Korean, two types of assimilation processes are attested when a lateral is adjacent to a nasal. First of all, when a nasal /n/ is followed by a lateral, the nasal is lateralized, as the examples in (1) illustrate[2].

(1) n-lateralization in /nl/

Input	**Output**	**Gloss**	**Related forms**	
/non + li/	[nolli]	'logic'	[non]	'discussion'
			/li/ ([i])	'reason'
			[nonceŋ]	'dispute'
			[illi]	'some reason'
			[c^huri]	'reasoning'
/han + lyaŋ/	[hallyaŋ]	'limit'	[han]	'limit'
			/lyaŋ/ ([yaŋ])	'quantity'
			[hangye]	'boundary'
			[toyraŋhyəŋ]	'weights and measures'
/c^hən + li/	[c^həlli]	'natural law'	[c^hən]	'sky'

This n-lateralization process is also attested when a lateral precedes a nasal /n/, as the examples in (2) illustrate.

---

[2] Components of each compound are included as related forms and their meanings are indicated. In Korean, a lateral is deleted in word-initial onset position as shown in the example 'reason', and it is realised as a flap intervocalically as the example 'reasoning' illustrates. In the case of a geminate lateral, it can occur in word-medial onset position as shown in the example 'some reason'. Other examples, which will be helpful in figuring out the underlying form, are also included as related forms.

(2) n-lateralization in /ln/

Input	Output	Gloss	Related forms	
/səl + nal/	[səllal]	'New Year's Day'	[səl]	'New Year'
			[nal]	'day'
/tʰɨl + ni/	[tʰɨlli]	'denture'	[tʰɨl]	'frame'
			/ni/ ([i])	'teeth'
			[saraŋni]	'wisdom teeth'
/pul + nɨŋ/	[pullɨŋ]	'incapability'	[pul]	'not'
			[nɨŋ]	'capability'
			[pulmyən]	'insomnia'
			[nɨŋnyək]	'capability'

However, when a lateral is preceded by a noncoronal nasal /m/ or /ŋ/, the lateral is nasalized as the examples in (3) illustrate[3].

(3) l-nasalization after a non-coronal nasal

Input	Output	Gloss	Related forms	
/sam + lyu/	[samnyu]	'third rate'	[sam]	'three'
			/lyu/ ([yu])	'rate, class'
			[iɾyu]	'second rate'
/tam + lyək/	[tamnyək]	'courage'	[tam]	'gall'
			/lyək/ ([yək])	'power'
			[siɾyək]	'eyesight'
/yəŋ + li/	[yəŋni]	'profit'	[yəŋ]	'administration'
			/li/ ([i])	'profit'
			[soɾi]	'a small profit'
/saŋ + lyu/	[saŋnyu]	'the upper stream'	[saŋ]	'upper'
			/lyu/ ([yu])	'stream'
			[haɾyu]	'the downstream'

As for these two different sonorant assimilation processes, we can raise the question why it is that the nasal assimilates to the lateral when the nasal is coronal, but when the nasal is velar or labial, it is the lateral that assimilates in manner to the nasal.

3. The Realization of n/l sequences in other languages

While l-nasalization in /ml/ and /ŋl/ sequences is a rather uncommon process, n-lateralization in n/l sequences is attested not only in Korean but in a wide range of different languages. N-lateralization before a lateral is observed in languages such as

---

[3] No assimilation process is found in the /lm/ sequence as in /pal + mok/ → [palmok] 'ankle'. I leave this for future study and will not discuss it in this paper.

Klamath, Ponapean, Toba Batak, Moroccan Arabic, and Leti. Tatar and Yakut show a different alternation pattern. Unlike other languages, l-nasalization occurs in /nl/.

(4) a. n-lateralization in /nl/

Klamath: /honlina/ → [hollina] 'flies along the bank' (Barker 1964, Rice & Avery 1991)
Ponapean: /nan-lɛŋ/ → [nallɛŋ] 'heaven' (Rehg & Sohl 1981, Rive & Avery 1991)
Toba Batak: /lɛan lali/ → [lɛal lali] 'give a hen-harrier' (Hayes 1986)
Moroccan Arabic: /ban + li/ → [balli] 'it seemed to me' (Amakhmakh 1997)
Leti: /na + losir/ → [llosir] '3sg, to follow' (Hume et al., 1997)

b. l-nasalization in /nl/

Tatar: /khayvan + lÄr/ → [khayvannar] 'animals' (Poppe 1963)
Yakut: /oron + lAr/ → [oronnor] 'beds' (Krueger 1962)

Languages such as Leti, Teralfene dialect of Flemish, and Rendille illustrate n-lateralization after a lateral. Moroccan Arabic shows a different alternation pattern. Unlike other languages, the underlying /ln/ sequence surfaces as [nn].

(5) a. n-lateralization in /ln/

Leti: /vulan/ → [vulla] 'moon' (Hume et al., 1997)
Teralfene dialect of Flemish: /spe:l-n/ → [spe:ll,] 'to play' (Levin 1988)
Rendille: /yeel-n-e/ → [yeelle] 'we carved' (Sim 1981)
Somali: /dil + -nay/ → [dillay] '(we) killed' (Zorc and Osman 1993)
Udi: /k'alnexa/ → [k'allexa] '(s)he calls' (Schulze 2001)

b. l-nasalization in /ln/

Moroccan Arabic: /dyal + na/ --> [dyanna] 'ours' (Amakhmakh 1997)

Related to n-lateralization and l-nasalization in n/l sequences and l-nasalization in the /ml/ and /ŋl/ sequences, we can raise two important questions. First, why do n-lateralization and l-nasalization occur in n/l sequences cross-linguistically while l-nasalization in the /ml/ and /ŋl/ sequences is not attested frequently in other languages? Second, why can both n-lateralization and l-nasalization occur in n/l sequences cross-linguistically?

4. Proposal: Perceptual Account[4]

Questions related to n-lateralization and l-nasalization in n/l sequences and l-nasalization in /ml/ and /ŋl/ may be answered by recourse to speech perception. Kohler (1990) and Hura et al. (1992) view assimilation as perceptually tolerated articulatory simplification. According to them, assimilation tends not to occur when the members of a consonant class are relatively distinctive perceptually, such that their articulatory reduction would be particularly salient. Steriade (2001) also provides a perceptually motivated account for phonological change such as assimilation by claiming that a sequence of acoustically similar segments is more likely to be selected for assimilation. According to Steriade, speakers' behavior is guided by a model of the generic listeners' perceptual abilities and biases. This model of the generic listener, which is called the P-map, has the function of identifying regions of relative safety within which a speaker can deviate from established pronunciation norms while minimizing the risk of being noticed. Thus, as the result of modification, among possible output forms for a given input, the one that differs least from the unmodified input form is preferred[5].

Following this, we hypothesize that [l] and [n] are difficult to distinguish in sequence since they are acoustically and auditorily similar, and this permits assimilation to occur. This hypothesis is supported by Borden & Harris (1984), and Johnson (1997). According to them, [l] and [n] are acoustically and auditorily similar in that both have formant structure and the same place of articulation. They are distinguished only by small differences in the frequencies of the formants and antiformants during the consonant closure.

Based on Kohler (1990), Hura et al. (1992), and Steriade (2001), we can further hypothesize that, as the result of assimilation, either a geminate lateral or a geminate nasal is obtained since both [ll] and [nn] are like /nl/ and /ln/. In other words, both changing [l] to [n] and changing [n] to [l] are perceptually allowed changes in n/l sequences.

As for assimilation in the /ml/ and /ŋl/ sequences, we hypothesize that it is attested less commonly cross-linguistically compared with assimilation in n/l sequences since the noncoronal nasals and lateral are less similar acoustically and auditorily. Although the noncoronal nasals and lateral all have formant structure, they differ in terms of place of articulation. Thus, it is expected that assimilation is frequently observed when more similar /n/ and /l/ segments are adjacent, while assimilation is less frequently observed when less similar /m/ or /ŋ/ and /l/ are adjacent. When assimilation is attested in the /ml/ and /ŋl/ sequences, we hypothesize that a lateral is nasalized since [mn] and [ŋn] are more similar to /ml/ and /ŋl/, respectively, than they are to [ll]. In other words, changing the

---

[4] Davis (1999) provides an account of sonorant assimilation in Korean based on a syllable contact constraint prohibiting rising sonority over a syllable boundary. His analysis is problematic, among other things, in that it cannot generalize to cases in which syllable contact is not relevant. For example, in Leti, tautosyllabic /nl/ surfaces as [ll], as in /na + losir/ --> [llosir] '3sg, to follow' (Hume et al., 1997).

[5] A similar account is also proposed by Kohler (1990).

lateral to a nasal results in a less noticeable change than would be the case if the nasal changed to a lateral. A perception experiment was run to test this hypothesis.

Given that the perceptual abilities (i.e. the P-map) proposed by Steriade are claimed to be universal, it is predicted that listeners' perception patterns will be the same regardless of the native language of a listener and sound patterns found in that language. That is, it is predicted that listeners will perceive both a geminate lateral and a geminate nasal as like /nl/ and /ln/, and both [mn] and [ŋn] as more like /ml/ and /ŋl/, respectively, than a geminate lateral. To test this hypothesis, a perception experiment was run. The universality of the P-map can be tested by comparing the results of the perception experiment obtained from Korean listeners with the results obtained from listeners of other languages which illustrate different phonological patterns in /nl/, /ln/, /ml/, and /ŋl/.

5. The Realization of nasal/lateral sequences in Moroccan Arabic and Swedish

Moroccan Arabic and Swedish differ from Korean in terms of the phonological patterns involving /nl/, /ln/, /ml/, and /ŋl/ sequences. Thus, a comparison of the results of the perception experiments obtained from Korean, Moroccan Arabic, and Swedish listeners will provide a good test of the P-map hypothesis.

5.1 Moroccan Arabic[6]

Moroccan Arabic is similar to Korean in that a nasal /n/ is lateralized before a lateral.

(6) n-lateralization in /nl/ (Amakhmakh 1997)

Input	Output	Gloss
/ban + li/	[balli]	'it seemed to me'
/mən + lhih/	[məllhih]	'from there'

However, Moroccan Arabic also shows a number of differences from Korean. First, while the /ln/ sequence is realized as [ll] in Korean, the sequence is realized as [nn] in Moroccan Arabic, as the following examples illustrate.

(7) l-nasalization in /ln/ (Amakhmakh 1997)

Input	Output	Gloss
/l + na/	[nna]	'to us'
/mal + na/	[manna]	'what's the matter with us'

Second, unlike Korean, the velar nasal /ŋ/ does not belong to the Moroccan Arabic sound inventory. Thus, while the sequence /ŋl/ surfaces as [ŋn] in Korean, no consonant sequence having the velar nasal as one of its components is attested in Moroccan Arabic.

---

[6] In Moroccan Arabic, n-lateralization and l-nasalization are morphologically conditioned. They occur with specific types of morphemes such as definite prefix /l-/ (Heath 1987; Keegan 1986).

Third, with respect to the sequence /ml/, in Moroccan Arabic it surfaces as [ml] without any change; in Korean, on the other hand, it is pronounced as [mn]. Thus, both [ml] and [mn] are possible surface forms in Moroccan Arabic, as the examples in (8) illustrate (Harrell 1962; Heath 1987).

(8) **Input**	**Output**	**Gloss**
/šrahom + lek/	[šrahomlek]	'he bought tem for you (sg.)'
/ẓṭm + l + ha /	[ẓṭmlha]	'he stepped on her (foot, etc.)'
/qeddem + na/	[qeddemna]	'we presented'
/sellem + na/	[sellemna]	'we greeted'

5.2 Swedish

Swedish is different from Korean and Moroccan Arabic in that no alternation patterns are found when a lateral is adjacent to another sonorant, although a geminate lateral and nasal exist, as shown in the examples [alla] 'all' (Pyun 1987) and [hennes] 'her' (NTC Publishing Group, 1997). Thus, the sequences /nl/ and /ln/ surface as [nl] and [ln], respectively, without undergoing assimilation.

(9) **Input**	**Output**	**Gloss**
/vanlig/	[vanlig]	'usual'
/manlig/	[manlig]	'male'
/molnig/	[molnig]	'cloudy'
/falna/	[falna]	'die down'

In addition, no alternations are found when a lateral is preceded by a noncoronal nasal. Thus, the underlying sequences /ml/ and /ŋl/ surface without undergoing assimilation. The sequences [mn] and [ŋn] are also possible surface forms.

(10) **Input**	**Output**	**Gloss**
/hemlig/	[hemlig]	'secret'
/næ mna/	[næmna]	'mention'
/viŋla/	[viŋla]	'stagger'
/reŋna/	[reŋna]	'rain'

5.3 Summary

We can summarize the phonological patterns involving /nl/, /ln/, /ml/, /ŋl/ sequences found in Korean, Moroccan Arabic, Swedish as follows:

Table 1. A summary of the patterns

Underlying Sequence	Surface Realization		
	Korean	Moroccan Arabic	Swedish
/nl/	[ll]	[ll]	[nl]
/ln/	[ll]	[nn]	[ln]
/ml/	[mn]	[ml]	[ml]
/ɳl/	[ɳn]		[ɳl]

5.4 Testing the P-Map Hypothesis

According to the P-map hypothesis (Steriade 2001), listeners' perceptual abilities are the same regardless of the native language of a listener and sound patterns found in that language. As shown above, Moroccan Arabic and Swedish show different phonological patterns in nasal/lateral sequences from Korean. If Korean, Moroccan Arabic, and Swedish listeners show the same perception patterns in the perception experiment, the P-map hypothesis will be supported. However, if it is hypothesized that the input and output of assimilation are perceptually confusable, it is expected that language-particular sound patterns influence listeners' perceptual abilities. Thus, according to this hypothesis, it is expected that Swedish listeners will perform the perception experiment better overall, compared with Korean and Moroccan Arabic listeners since all nasal/lateral sequences concerned exist as surface forms in Swedish. In addition, Korean listeners are expected to perform the perception experiment worse since /nl/, /ln/, /ml/ and /ɳl/ do not surface in Korean. Since /ln/ is realized as [ll] in Korean, it is expected that Korean listeners perceive /ln/ as more similar to [ll] than to [nn]. In the cases of Moroccan Arabic listeners, it is expected that they will show different perception patterns from Korean listeners by perceiving /ln/ as more similar to [nn] than to [ll], considering that /ln/ is realized as [nn] in Moroccan Arabic. Since /nl/ is realized as [ll] in Moroccan Arabic as it is in Korean, it is expected that Moroccan Arabic and Korean listeners will perceive /nl/ as more similar to [ll] than to [nn].

6. Methods

6.1 Stimuli

For the experiment, recordings of ten repetitions of the sequences [anla], [alna], [anna], [alla], [amla], [aɳla], [amna], and [aɳna] were made by a Hindi speaker in a sound-attenuated booth at the Ohio State University and digitized at 22050 Hz. For the recordings, a head-mounted microphone (SM10A SHURE) and a TEAC V-427C tape recorder were used. Stimuli recordings were made by a Hindi speaker since every consonant sequence in the stimuli occurs in Hindi[7].

The types of stimuli presented to listeners are as follows:

---

[7] The perception experiment was not run with Hindi listeners to avoid potential native language effects. For the same reason, stimuli recordings were made by a Hindi speaker rather than by a Swedish speaker, although every sequence used in this perception experiment occurs in Swedish.

(11) [anla]/[anla]
[anla]/[anna]
[anla]/[alla]

[alna]/[alna]
[alna]/[anna]
[alna]/[alla]

[anna]/[anna]
[alla]/[alla]

[amla]/[amla]
[amla]/[amna]
[amla]/[alla]

[aŋla]/[aŋla]
[aŋla]/[aŋna]
[aŋla]/[alla]

To make these paired stimuli, four repetitions were taken for permutation from ten repetitions digitized for each stimulus. For each pair with the same stimuli (i.e. [anla] /[anla], etc.), twelve pairs of stimuli were obtained after permutation, while sixteen pairs of stimuli were obtained in the case of a pair with different stimuli (i.e. [anla]/[anna], etc.). As illustrated in (11), there are 6 pairs consisting of the same stimuli and 8 pairs consisting of different stimuli. To have the same number of 'same' and 'different' stimuli, 8 pairs were taken from the twelve pairs obtained after permutation for each identical stimulus and 6 pairs were taken from the sixteen pairs for each different stimulus. By this procedure, 48 pairs were obtained for identical stimuli, and another 48 pairs for different stimuli, totalling 96 pairs in all. All types of stimuli were presented to listeners in the order illustrated in (11). That is, in the presentation of the stimulus pair [anla]/[anna], for example, [anla] was presented before [anna].

6.2 Listeners

The listeners were 20 native speakers of Seoul Korean (10 males, 10 females), 22 native speakers of Swedish (8 males, 14 females), and 7 native speakers of Moroccan Arabic (6 males, 1 female). The age range for the Korean listeners was 20 to 36 years, with all having lived in the U.S. between 1 and 15 years (average 4 years). Two Swedish female listeners were students at the Ohio State University, 25 and 31 years old, both of whom have lived in the U.S. for 5 years. Twenty Swedish speakers were students at Lund University in Sweden, ranging in age from 20 to 57 years. The age range for the Moroccan listeners was 24 to 37 years, with all having lived in the U.S. between 1 and 2 years.

6.3 Task

The task of the listeners was to determine whether members of a pair of stimuli they heard are the same or different. Error rate and reaction time were recorded to use as an indicator of ease of perception in the analysis. For the discrimination test, the MEL program was used. The inter-stimulus interval was 500 ms and the reaction time clock started at the onset of the second stimulus in each pair. For the reaction time, the program was set to measure the time between the onset of the second stimulus and the point that a listener pressed the response button. Listeners were instructed to press the SAME button if they thought the pair of stimuli they heard were the same or the DIFFERENT button if they thought the pair of stimuli they heard were different. They were instructed to respond as accurately and quickly as possible and to look at the screen during the experiment. To enable a quick response, listeners were asked to use one hand to press the

SAME button and the other hand to press the DIFFERENT button. The stimuli were played through the headphones in a sound-attenuated booth.

Listeners heard 96 pairs of stimuli twice: once at approximately the speech reception threshold (40dB) and once again at a comfortable listening level (70dB)[8]. Listeners worked through the experiment at the speech reception threshold before they heard the stimuli again at a comfortable listening level. The experiment at the speech reception threshold (that is, at 40dB) was done for the purpose of eliciting mistakes.

6.4 Predictions

Based on the claims made by Kohler (1990), Hura et al. (1992), and Steriade (2001), several predictions can be made. First of all, since both n-lateralization and l-nasalization are attested in n/l sequences cross-linguistically, and l-nasalization rather than nasal-lateralization occurs in /ml/ and /ŋl/ in Korean, Korean listeners are predicted to perceive the members of the [anla]/[alla], [anla]/[anna], [alna]/[alla], [alna]/[anna], [amla]/[amna], and [aŋla]/[aŋna] as more similar to each other than the pairs [amla]/[alla] and [aŋla]/[alla]. Furthermore, it is predicted that it will take more time to discriminate the former pairs than the latter pairs under the assumption that a similar pair will be harder to discriminate. Finally, based on the proposed universality of the P-map, it is predicted that the results of the discrimination test will not show language effects. Thus, Swedish and Moroccan Arabic listeners will give the same responses as Korean listeners.

7. Results and Discussion

7.1 Reaction Time

For the reaction time analysis, only the reaction time of the correct responses for different stimuli was considered. To avoid the influence of outliers on the results, each subject's median reaction time was calculated for each pair type. There were cases in which subjects gave no correct response, and thus no median reaction time can be given. In such cases, reaction time was filled in by an expected value obtained by the addition of a mean deviation for pair type, which is calculated by subtracting a pair type mean from the total mean, and a mean of the subject's median reaction time.

We analyzed the obtained data in a repeated-measures analysis of variance (ANOVA) having one between-listeners factor (**language**: Korean, Moroccan Arabic or Swedish), and two within listeners factors (**loudness**: 40dB or 70dB; **pair type**). The results of the analysis of variance are shown in Table 2.

---

[8] We used 40 dB as the speech reception threshold level based on Winters (2000).

Table 2. Repeated Measures of ANOVA of Reaction Time. (The effects marked with bold face and * were significant at $p < 0.01$.)

Source of Variance	DF	F
Between listeners		
Language	2	.6
Within listeners		
Loudness	1	29.7*
Pair type	7	23.6*
Loudness * Language	2	3.1
Pair type * Language	14	1.6
Loudness * Pair type	7	2.7*
Loudness * Pair type * Language	14	1.3

The effect of loudness was reliable in this analysis. On average, the mean reaction times were 851.6 ms for 40 dB and 747.5 ms for 70 dB. Thus, listeners' reaction time was slower when the stimuli were played at the reception threshold level, that is, at 40 dB.

There was also a main effect of **pair type**.

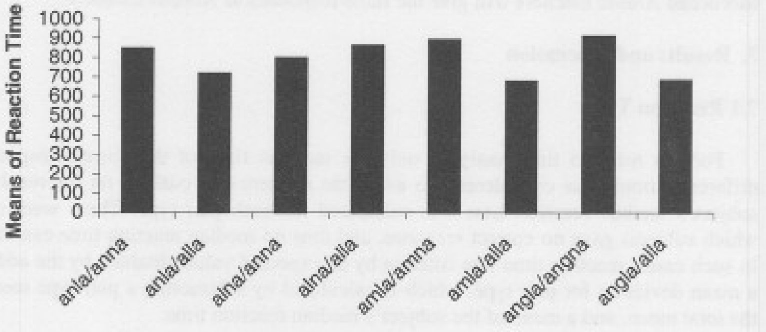


Figure 1. Pairtype

As can be seen in figure 1, the pair types [amla]/[amna] and [aŋla]/[aŋna] ('ŋ' is written as 'ng' in the figure) have longer reaction times than the pair types [amla]/[alla] and [aŋla]/[alla], respectively. This is the expected result from our hypothesis: the input and output of an assimilation process is confusable perceptually and, thus, it will take more time to distinguish them. However, when it comes to the pair type [anla]/[alla], contrary to our hypothesis that there will be no perceptual difference between [anla]/[alla] and [anla]/[anna] since both n-lateralization and l-nasalization are attested in /nl/ cross-linguistically, it has shorter reaction time than the pair type [anla]/[anna]. For the pair types [alna]/[alla] and [alna]/[anna], contrary to our hypothesis, listeners showed longer reaction time in discriminating [alna]/[alla].

Another main effect in the analysis was the **loudness * pair type** interaction.

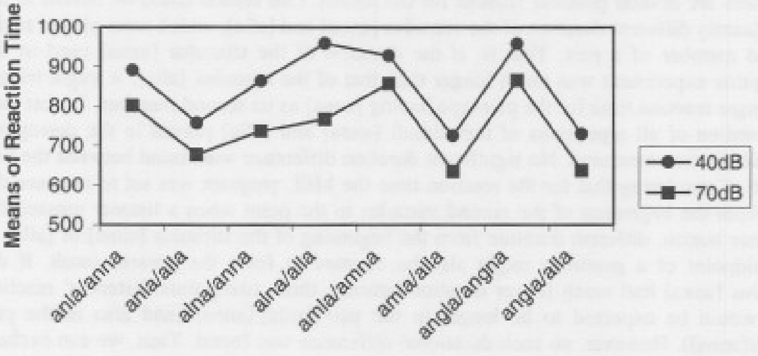


Figure 2. Loudness * Pair type interaction

As we can see in figure 2, for every pair type the mean reaction time decreases as the loudness changes from 40 dB to 70dB. The overall pattern of each pair type is consistent at both 40 dB and 70 dB levels. Thus, the pair type [anla]/[alla], [amla]/[alla], and [aŋla]/[alla] have shorter reaction times at both loudness levels. In the cases of [alna]/[anna] and [alna]/[alla], they showed no significantly different reaction time at 70 dB. However, [alna]/[alla] showed a longer reaction time than [alna]/[anna] at 40 dB.

7.2 Discussion

The reaction time results are consistent with the hypothesis that l-nasalization in the /ml/ and /ŋl/ sequences occurs since it is a perceptually less noticeable change than nasal-lateralization. As we expected, it took longer for listeners to discriminate between members of the pairs [amla]/[amna] and [aŋla]/[aŋna] than between members of the pairs [amla]/[alla] and [aŋla]/[alla]. The analysis also showed no language effect, thus, supporting Steriade's hypothesis that listeners' perceptual abilities are universal regardless of the native language of a listener and sound patterns found in that language.

However, the results contradict our hypothesis concerning n-lateralization and l-nasalization in n/l sequences. According to our hypothesis, it was predicted that either /n/ is lateralized or /l/ is nasalized in n/l sequences since both a geminate lateral and a geminate nasal are similar to /nl/ and /ln/. In the cases of the pair types [alna]/[alla] and [alna]/[anna], contrary to our hypothesis, listeners showed longer reaction time in discriminating [alna]/[alla] than [alna]/[anna]. For the pair types [anla]/[alla] and [anla]/[anna], although it was expected that it would take the same time to discriminate between members of the pair [anla]/[alla] and between members of the pair [anla]/[anna], listeners were better at discriminating the pair [anla]/[alla] than the pair [anla]/[anna] as shown in figure 1.

There are several possible reasons for the results. One reason could be related to the significantly different duration of the stimulus [anna] and [alla], which were played as the second member of a pair. That is, if the duration of the stimulus [anna] used in the perception experiment was much longer than that of the stimulus [alla], it might trigger the longer reaction time for the pair type having [anna] as its second member. To test this, the duration of all repetitions of the stimuli [anna] and [alla] played in the perception experiment was measured. No significant duration difference was found between the two stimuli. Considering that for the reaction time the MEL program was set to measure the time from the beginning of the second stimulus to the point when a listener pressed the response button, different duration from the beginning of the stimulus [anna] or [alla] to the midpoint of a geminate might also be claimed to force the present result. If the stimulus [anna] had much longer duration between those two points, listeners' reaction time would be expected to be longer in the pair [anla]/[anna] (and also in the pair [alna]/[anna]). However, no such durational difference was found. Thus, we can exclude the possibility that the result was obtained due to the different durations of the stimuli [anna] and [alla].

Another possible reason for the result could be related to which consonant is different in each pair type; that is, whether the first or second consonant is different. If the first consonant of two stimuli is different as in the pair type [anla]/[alla], the reaction time is expected to be faster since listeners can decide whether two stimuli are different as soon as they listen to the first consonant of the second stimulus. However, when a pair type is composed of two stimuli differing in the second consonant as in [anla]/[anna], listeners would have to wait until they listened to the second consonant of the second stimulus, and thus it is expected that their reaction time will be longer. The longer reaction times of the pairs [amla]/[amna], [aŋla]/[aŋna] and [alna]/[anna] as compared to the pairs [amla]/[alla], [aŋla]/[alla] and [alna]/[alla] could have been influenced by this factor.

Thus, the reaction time analysis seems to be influenced by the position of the different consonant in a pair type. To see what the results are when the influence of this factor is excluded, an analysis considering listeners' sensitivity measure was done.

7.3 Sensitivity (d') measure

In performing the analysis of the results based on listeners' sensitivity measure, listeners' perceptual abilities were measured for each pair type using a sensitivity measure d'. This sensitivity measure takes into account a listener's bias to choose a particular response alternative by calibrating the 'hit rate' (the proportion of correctly recognized different stimuli) with the 'false alarm rate' (the mean proportion of incorrectly recognized same stimuli)[9]. The d' analysis might tell us more things about

---

[9] Hit = correct use of response "different"
False alarm = incorrect use of "different"

our discrimination test since it takes into account false alarm rate as well as hit rate while the reaction time analysis given above takes into account the reaction time of correct responses only.

The formula for d' is as follows:

(12) $d' = z(H) - z(F)$,
where H is a hit rate, F is a false alarm rate, z means z-score.

Instead of getting d' values using the above formula, we used the tables in Kaplan, MacMillan & Creelman (1978) which show a d' value for given hit and false alarm rates (in the AX discrimination task).

The sensitivity data were analyzed using a univariate analysis of variance (ANOVA) having three between-listeners factors (**language**: Korean, Moroccan Arabic or Swedish; **loudness**: 40 dB or 70 dB; **pair type**: [anla]/[anna], [anla]/[alla], [alna]/[anna], [alna]/[alla], [amla]/[amna], [amla]/[alla], [aŋla]/[aŋna] and [aŋla]/[alla]). The results of the analysis of variance are shown in Table 3.

Table 3. Univariate Analysis of Variance for Sensitivity (d'). (The effects marked with boldface and * were significant at $p < 0.01$)

Source of Variance	DF	F
Language	2	31.9*
Loudness	1	110.4*
Pair type	7	29.2*
Language * Loudness	2	10.1*
Language * Pair type	14	3.1*
Loudness * Pair type	7	2.8*
Language * Loudness * Pair type	14	3.9

One of the main effects in this analysis is the effect of **language**. As shown in figure 3, Moroccan Arabic and Swedish listeners showed higher sensitivity than Korean listeners.

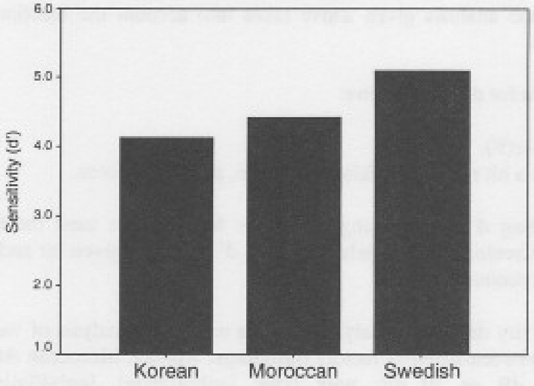


Figure 3. Language

Sensitivity (d') values were also significantly different for different **loudness** levels. The sensitivity (d') value was 3.9 at 40 dB and 5.1 at 70 dB. The effect of **pair type** was also reliable, as shown in Figure 4.

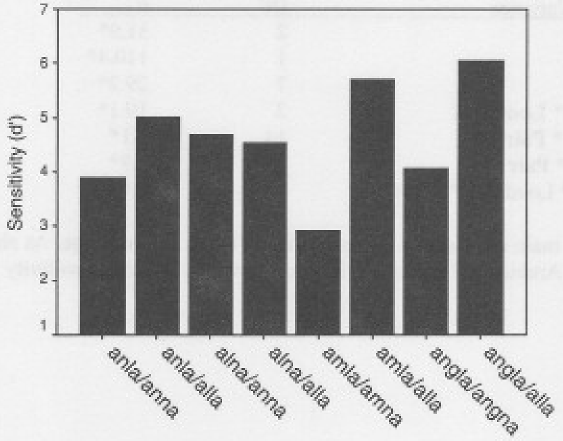


Figure 4. Pair type

The results from the sensitivity data are similar to those from the reaction time data. Listeners showed significantly lower sensitivity to the pairs [amla]/[amna] and [aŋla]/[aŋna] than to the pairs [amla]/[alla] and [aŋla]/[alla], respectively. Contrary to our hypothesis, the pair type [anla]/[anna] is higher in sensitivity than the pair type

[anla]/[alla], and, unlike the results from the reaction time data, but as we expected, the pair types [alna]/[anna] and [alna]/[alla] showed no significant difference in sensitivity.

The **language * loudness** interaction showed that Swedish listeners were better at 40 dB and their sensitivity increased much more at 70 dB than it did for Korean and Moroccan Arabic listeners, as illustrated in figure 5.

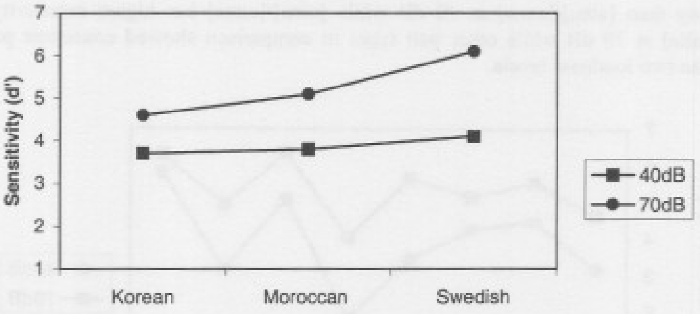


Figure 5. Language * Loudness Interaction

The **language * pair type** interaction is shown in figure 6.

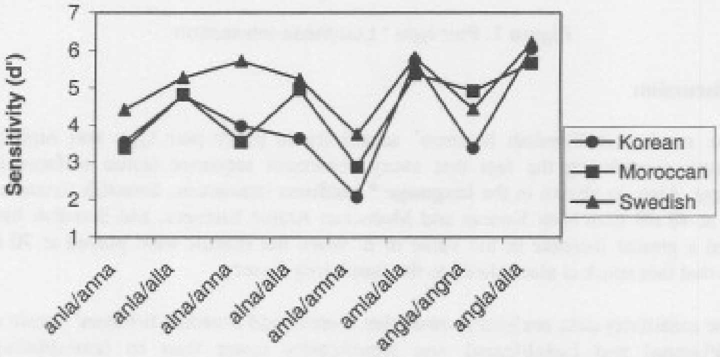


Figure 6. Language * Pair type Interaction

Swedish listeners showed relatively higher sensitivity overall across the pair types. Korean and Swedish listeners had significantly lower sensitivity to the pairs [amla]/[amna] and [aŋla]/[aŋna] than to the pairs [amla]/[alla] and [aŋla]/[alla]. Listeners' sensitivity difference between [anla]/[anna] and [anla]/[alla] was not significantly different except in the case of Swedish listeners who showed higher sensitivity to [anla]/[alla] than to [anla]/[anna]. In the case of [alna]/[anna] and [alna]/[alla], as we expected, there was no significant sensitivity difference for any group of listeners.

The **pair type * loudness** interaction showed that listeners' sensitivity to the pair types [alna]/[alla], [amla]/[amna], and [aŋla]/[aŋna] increased more, compared with other pair types when they were heard at 70 dB. It is interesting that [alna]/[alla] has higher sensitivity than [alna]/[anna] at 40 dB while [alna]/[anna] has higher sensitivity than [alna]/[alla] at 70 dB while other pair types in comparison showed consistent patterns across the two loudness levels.

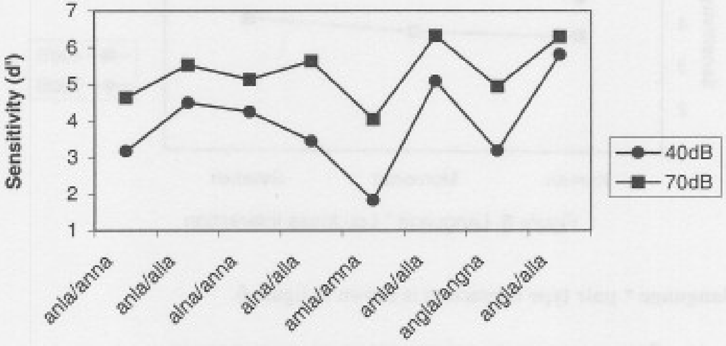


Figure 7. Pair type * Loudness Interaction

7.4 Discussion

The result that Swedish listeners' sensitivity to every pair type was high is not surprising considering the fact that every consonant sequence tested surfaces in the language. Also, as shown in the **language * loudness** interaction, Swedish listeners were better at 40 dB than both Korean and Moroccan Arabic listeners, and Swedish listeners showed a greater increase in the value of d' when the stimuli were played at 70 dB. It seems that this result is also related to the familiarity factor.

The sensitivity data analysis showed that Korean and Swedish listeners' sensitivity to [amla]/[amna] and [aŋla]/[aŋna] was significantly lower than to [amla]/[alla] and [aŋla]/[alla], respectively. This then supports the hypothesis that l-nasalization in the /ml/ and /ŋl/ sequences is influenced by listeners' perceptual abilities since they perceive [amla] and [aŋla] as more similar to [amna] and [aŋna], respectively. It is especially interesting that Swedish listeners showed the same perception pattern as Korean listeners even though the consonant sequences [ml], [mn], [ŋl], and [ŋn] all occur in Swedish. On the other hand, Moroccan Arabic listeners showed no significant difference in sensitivity between [amla]/[amna] and [amla]/[alla], and between [aŋla]/[aŋna] and [aŋla]/[alla].

Thus, the hypothesis that listeners' perceptual abilities are the same regardless of the phonological patterns of their native languages is supported by the results from Swedish and Korean listeners while it is not supported by those from Moroccan Arabic listeners.

It seems that the phonological patterns of a listener's native language also influence speech perception when we consider the fact that Korean listeners' sensitivity was lower, compared with that of Moroccan Arabic and Swedish listeners. Among the three languages, Korean is the only language neutralizing every nasal/liquid sequence discussed here. Thus, under our hypothesis that the input and output of neutralization are confusable, Korean listeners' lower sensitivity might be influenced by such Korean phonological patterns. Also, the result that there is no significant difference between [aŋla]/[aŋna] and [aŋla]/[alla] in Moroccan Arabic listeners might be the influence of the phonology of Moroccan Arabic in which /ŋ/ is not a possible speech sound.

Korean and Moroccan Arabic listeners showed no significant difference in sensitivity between [anla]/[anna] and [anla]/[alla], as we expected, while Swedish listeners showed higher sensitivity to [anla]/[alla] than to [anla]/[anna]. These results support our hypothesis in part that [anla]/[alla] is as confusable as [anla]/[anna] and thus both n-lateralization and l-nasalization are attested in different languages. One reason for the results from Korean and Moroccan Arabic listeners may relate to the vowel nasalization cue. In [anla]/[alla], the first vowels of the two stimuli are different phonetically due to the presence or absence of vowel nasalization, respectively. That is, in [anla], the first vowel is nasalized before a nasal consonant while there is no such nasalization in the first vowel of the stimulus [alla]. Thus, if only language universal cues are involved in the discrimination of [anla]/[anna] and [anla]/[alla], it is expected that the pair type [anla]/[alla] will be easier to be distinguished than the pair type [anla]/[anna], in which the first vowels of both stimuli are nasalized.

Vowel nasalization may also provide insight as to why, in figure 7, listeners' sensitivity to [anla]/[anna], [alna]/[alla], [amla]/[amna], and [aŋla]/[aŋna] is lower than to other pair types at 40 dB. In each of these pair types, the first vowels have the same status regarding the nasalization cue; they are either both nasalized or both nonnasalized. Thus, considering the vowel nasalization, we can say that these pair types are more confusable. Under the assumption that more confusable stimuli will be even harder to discriminate at the reception threshold level (40 dB), this result is expected and it suggests that vowel nasalization played an important role when discriminating between two stimuli in this test.

If the vowel nasalization factor is considered in the case of the pair type [alna]/[anna] and [alna]/[alla], it is expected that [alna]/[anna] will show a higher d' value than [alna]/[alla]. However, as mentioned before, the listeners showed no difference in the sensitivity to either pair types. Also, Swedish listeners showed higher sensitivity to [anla]/[alla] than to [anla]/[anna], while Korean and Moroccan Arabic listeners showed no significant sensitivity difference between those two stimuli. These results suggest that some other language-universal or language-particular factors might also influence speech perception. Since, in this analysis, we are unable to determine which phonetic or language-specific factors influenced the value of d' and how much their influence is, we performed a linear regression analysis separately for each language.

8. Linear Regression

8.1 Hypothesis

We hypothesize that the presence or absence of a vowel-nasalization cue, a place contrast, and a geminate consonant in the stimuli, as well as different levels of loudness (that is, 40 dB or 70 dB) will be language-universal phonetic factors influencing listeners' perceptual abilities. If there is a vowel-nasalizaion contrast in a pair type as in [alna]/[anna], it is expected that listeners' sensitivity will increase. Also, if two stimuli have a place contrast, as in [amla]/[alla] or [aŋla]/[alla], it is predicted that it will be easier for listeners to discriminate between the members of the pair. It is also predicted that the presence of a geminate in a pair type such as [anla]/[anna] will increase listeners' sensitivity to that pair type due to the longer duration of a geminate consonant. Finally, the prediction is that stimuli played at 70 dB will increase listeners' sensitivity.

Table 4 shows how each pair type was coded for the linear regression analysis based on the proposed language-universal factors.

Table 4.

Pair Type	Place Contrast	Vowel Nasalization	Geminate	Loudness
[anla]/[anna]	No (0)	No (0)	Yes (1)	40 dB (0)
[alna]/[anna]	No (0)	Yes (1)	Yes (1)	70 dB (1)
[amla]/[amna]	No (0)	No (0)	No (0)	
[aŋla]/[aŋna]	No (0)	No (0)	No (0)	
[anla]/[alla]	No (0)	Yes (1)	Yes (1)	
[alna]/[alla]	No (0)	No (0)	Yes (1)	
[amla]/[alla]	Yes (1)	Yes (1)	Yes (1)	
[aŋla]/[alla]	Yes (1)	Yes (1)	Yes (1)	

As for place contrast and vowel nasalization, if a pair type has a cue, the value '1' was assigned, if the cue is absent, the value '0' was assigned. If a pair type contains a geminate, the value '1' was assigned and if not, the value '0' was assigned. Finally, for loudness, the value '0' was assigned to a 40 dB level and the value '1' to a 70 dB level.

Since language-specific factors might also influence listeners' perception, we considered the issue of neutralization: whether or not consonant sequences in two stimuli contrast with each other. Our hypothesis is that listeners will have a hard time discriminating between two stimuli if one stimulus is neutralized to the other in their native language. That is, Korean listeners' sensitivity to [alna]/[alla] is expected to be lower than to [alna]/[anna] since /alna/ is neutralized to [alla] in Korean. However, it is expected that Moroccan Arabic listeners will show less sensitivity to [alna]/[anna] since /alna/ is neutralized to [anna] in Moroccan Arabic.

Different patterns of contrast in consonant clusters in each language may also influence speech perception. When consonant sequences in two stimuli contrast in one language, listeners of that language are expected to show higher sensitivity to the pair type consisting of those two stimuli. For example, the consonant sequence /ŋl/ and /ll/ do not contrast in Korean and Moroccan Arabic since /ŋl/ surfaces as [ŋn] in Korean and /ŋ/ does not belong to the Moroccan Arabic sound inventory. However, they do contrast in Swedish since both [ŋl] and [ll] surface. Thus, it is expected that Swedish listeners will show higher sensitivity to the pair type [aŋla]/[alla].

Table 5 shows how each pair type in Korean, Moroccan Arabic and Swedish was coded according to the language-specific factors: contrast and neutralization.

Table 5.

Pair Type	Contrast			Neutralization		
	Korean	Moroccan	Swedish	Korean	Moroccan	Swedish
[anla]/[anna]	No (0)	No (0)	Yes (1)	No (1)	No (1)	No (1)
[alna]/[anna]	No (0)	No (0)	Yes (1)	No (1)	Yes (0)	No (1)
[amla]/[amna]	No (0)	Yes (1)	Yes (1)	Yes (0)	No (1)	No (1)
[aŋla]/[aŋna]	No (0)	No (0)	Yes (1)	Yes (0)	No (1)	No (1)
[anla]/[alla]	No (0)	No (0)	Yes (1)	Yes (0)	Yes (0)	No (1)
[alna]/[alla]	No (0)	No (0)	Yes (1)	Yes (0)	No (1)	No (1)
[amla]/[alla]	No (0)	Yes (1)	Yes (1)	No (1)	No (1)	No (1)
[aŋla]/[alla]	No (0)	No (0)	Yes (1)	No (1)	No (1)	No (1)

If a pair type has two consonant sequences which contrast in a language, the value '1' was assigned, and if not, the value '0' was assigned. If a pair type consists of the input and output of manner neutralization in a language, the value '0' was assigned, otherwise the value '1' was assigned.

8.2 Results

The results of the linear regression analysis are shown in Table 6.

Table 6. Linear Regression of Sensitivity (d')

	Korean	Moroccan	Swedish
R Square	**.35**	**.20**	**.41**
Constant	**2.28**	**3.6**	**3.1**
place contrast	1.48	1.77	.51
geminate	.86		.73
loudness	.87	1.24	1.96
vowel nasalization	.81		.67
contrast		-1.02	
neutralization			

The value of R square measures the proportion of d' variability that can be predicted by the given model. That is, in the case of Korean, the model including place contrast, geminate, loudness, and vowel nasalization factors can predict 35% of the observed variation of d'. The remaining variation is due to other factors such as individual differences between listeners.

Since no paired stimulus used in the perception experiment shows any consonant sequence contrast in Korean, the contrast cue was deleted from the analysis. Thus, in Korean, place contrast, geminate, loudness and vowel nasalization are the cues influencing listeners' sensitivity. As we hypothesized, when such cues are present in a stimulus, listeners' sensitivity to that stimulus increased. Korean listeners' sensitivity was influenced most by the place contrast cue while the influence of loudness was relatively low in Korean listeners, as compared with Moroccan Arabic and Swedish listeners.

In Moroccan, the influence of the place contrast was the greatest. It is interesting that listeners' sensitivity to the stimuli decreased when the contrast cue was present.

The neutralization and contrast cues were deleted from the analysis of Swedish listeners' sensitivity data since the two members in every pair type contrast in Swedish. As in the results from Korean listeners, the presence of the cues place contrast, geminate, loudness, and vowel nasalization increased listeners' sensitivity. However, Swedish listeners were influenced most by the loudness cues.

8.3 Discussion

Korean and Swedish listeners' sensitivity to a stimulus was influenced by the language-universal phonetic cues: place contrast, geminate, and vowel nasalization. The presence or absence of phonetic cues influencing Korean and Swedish listeners' discrimination of each stimulus can be summarized as follows.

Table 7.

[amla]/[amna]	[amla]/[alla]	[aŋla]/[aŋna]	[aŋla]/[alla]
	place contrast geminate vowel nasalization		place contrast geminate vowel nasalization
[anla]/[anna]	[anla]/[alla]	[alna]/[anna]	[alna]/[alla]
geminate	geminate vowel nasalization	vowel nasalization geminate	geminate

The above table shows in detail which language-universal or language-particular cues influenced Korean and Swedish listeners' sensitivity (d') value. Listeners' significantly higher sensitivity to [amla]/[alla] and [aŋla]/[alla] than to [amla]/[amna] and [aŋla]/[aŋna], respectively, is caused by the phonetic cues such as place contrast, geminate, and vowel nasalization, which are all present in [amla]/[alla] and [aŋla]/[alla], and all absent in [amla]/[amna] and [aŋla]/[aŋna].

The results of the linear regression analyses also give insight into why Swedish listeners perceived the stimulus [anla]/[alla] more easily than the stimulus [anla]/[anna]. The former has an extra phonetic cue, vowel nasalizaion, which helps discriminate between [anla] and [alla]. However, it is still not clear why Korean listeners showed no significant sensitivity difference between [anla]/[anna] and [anla]/[alla]. In the cases of the stimuli [alna]/[anna] and [alna]/[alla], [alna]/[anna] has an extra phonetic cue, vowel nasalization. But Korean and Swedish listeners' sensitivity difference between [alna]/[anna] and [alna]/[alla] was not significant and it is not clear which factors triggered this result.

Moroccan Arabic listeners' sensitivity to a stimulus was influenced by the language-universal cues: place contrast and loudness and the language-specific cue, contrast. The presence or absence of phonetic cues influencing Moroccan Arabic listeners' discrimination of each stimulus is summarized in Table 8.

Table 8.

[amla]/[amna]	[amla]/[alla]	[aŋla]/[aŋna]	[aŋla]/[alla]
contrast	place contrast contrast		place contrast
[anla]/[anna]	[anla]/[alla]	[alna]/[anna]	[alna]/[alla]

By considering both the results of the linear regression analysis for Moroccan Arabic listeners given in Table 6 and phonetic cues influencing the discrimination of each stimulus given in Table 8, we can account for Moroccan Arabic listeners' perception patterns to some extent. In the case of the pairs [anla]/[anna], [anla]/[alla], [alna]/[anna] and [alna]/[alla], they do not include any phonetic or language specific cue influencing listeners' sensitivity. Thus, it is expected that listeners will show no difference in sensitivity between the pairs being compared. In the case of the pairs [amla]/[alla] and [aŋla]/[alla], they contain the place constrast cue which enhances listeners' sensitivity while the pairs [amla]/[amna] and [aŋla]/[aŋna] do not contain it. Thus, it is expected that listeners' sensitivity to [amla]/[alla] and [aŋla]/[alla] will be higher. However, Moroccan Arabic listeners displayed no significant difference between the two stimuli being compared and it is not clear which other factors triggered this result.

9. General Discussion

As an indicator of ease or difficulty of perception, we used reaction time and sensitivity (d') data. As mentioned above, it seems that listeners' reaction time was influenced by some non-linguistic factors such as the position in which different consonants are located in a paired stimulus. Thus, it seems that it will be better to discuss sonorant assimilation patterns based on the results from the sensitivity data, although the results from the two analyses are almost identical.

The sensitivity (d') data results suggested that l-nasalization in /ml/ and /ŋl/ is influenced by speech perception. As expected, [amla]/[amna] was more confusable than [amla]/[alla], and [aŋla]/[aŋna] was more confusable than [aŋla]/[alla] in Korean. Since changing /l/ to [n] is a less noticeable change in /ml/ and /ŋl/ perceptually, it is expected that the lateral /l/ is nasalized when sonorant assimilation occurs in those sequences, as in Korean.

Korean listeners showed no significant sensitivity difference between [anla]/[anna] and [anla]/[alla], and between [alna]/[anna] and [alna]/[alla]. This result supports our hypothesis that the degree of noticeability or salience between the two stimuli being compared is the same. That is, both changing /l/ to [n] and /n/ to [l] in n/l sequences are perceptually allowed changes, and thus both changes are attested in languages. If this is the case, we can then raise the question concerning which non-perceptual factors force l-nasalization, n-lateralization, or both in one language. As one possible factor, we can consider speakers' tendency to reduce speaking effort. In the production of [anna], extra articulatory effort to lower the velum is required, while no such effort is required in producing [alla]. Thus, we can say that speakers may prefer changing perceptually bad /nl/ and /ln/ sequences to a geminate lateral, as in Korean. Given this, why do /ml/ and /ŋl/ sequences surface as [mn] and [ŋn], respectively, if the production of a geminate lateral requires less speaking effort? In this case, it is because a speaker can deviate from established pronunciation norms while minimizing the risk of being noticed. As our perception experiment results showed, changing /ml/ and /ŋl/ to a geminate lateral is more noticeable than changing /l/ to [n]. Thus, /ml/ and /ŋl/ surface as [mn] and [ŋn], respectively, in this case.

In Moroccan Arabic, a nasal /n/ is lateralized in the /nl/ sequence while a lateral /l/ is nasalized in the /ln/ sequence. According to our experiment results from Moroccan Arabic listeners, both are perceptually allowed changes since no significant sensitivity difference was found between changing /l/ to [n] and changing /n/ to [l] in n/l sequences. Given speakers' tendency to reduce speaking effort discussed above, however, it is expected that both /nl/ and /ln/ be realized as [ll]. We can then raise the question why n-lateralization is attested in /nl/ and l-nasalization in /ln/. As one possible answer to this, the force preserving a prevocalic consonant which has been claimed to be perceptually stronger than a coda consonant by many researchers, can be considered. Thus, /VnlV/ changes to [VllV] and /VlnV/ to [VnnV] in Moroccan Arabic[10].

10. Conclusion

This paper investigated the influence of speech perception as a source of explaining sonorant assimilation in Korean. The results suggest that l-nasalization in the changes from /ml/ and /ŋl/ to [mn] and [ŋn], respectively, is driven by perceptual considerations. However, the perception experiment results and assimilation patterns in n/l sequences suggest that listeners' perceptual abilities are not the only factors shaping the phonological patterns of languages. Rather, articulation as well as perception is relevant. That is, not only the change made by assimilation should be perceptually less noticeable, but it should also be articulatorily easy.

Our perception experiment results also showed that the universality of Steriade's P-map is supported in part. In general, universal speech perception patterns are attested between two groups of listeners whose native languages are different. However, at the same time, languages may deviate from a universal perception pattern. In this study, Swedish listeners showed different perception patterns from Korean and Moroccan Arabic listeners by having higher sensitivity to [anla]/[alla] than to [anla]/[anna] while Korean and Moroccan Arabic listeners had no sensitivity difference between those two stimuli. Moroccan Arabic listeners displayed different perception patterns from Korean and Swedish listeners by showing no sensitivity difference between [amla]/[amna] and [amla]/[alla], and [aŋla]/[aŋna] and [aŋla]/[alla] while Korean and Swedish listeners showed higher sensitivity to [amla]/[alla] and [aŋla]/[alla].

Finally, our perception experiment results showed that speech perception is influenced by phonological patterns of a listener's native language. We speculate that Korean listeners had lower sensitivity compared with other listeners due to the phonological patterns in Korean where /nl/, /ln/, /ml/ and /ŋl/ never surface as the result of neutralization.

---

[10] In Tatar and Yakut, the underlying /nl/ is realized as [nn], as mentioned in section 3. According to our perception experiment results, this is a perceptually allowed change. However, it is not clear which factor forces the prevocalic /l/ to be changed to [n] resulting in a geminate nasal which requires more articulatory effort than a geminate lateral. I leave this for future study.

References

Abdel-Massih, Ernest T. 1982. An Introduction to Moroccan Arabic. The University of Michigan: Ann Arbor.

Amakhmakh, Naceur. 1997. Non-Linear Phonology of a Moroccan Arabic Dialect. PhD. dissertation. University of Wisconsin-Madison.

Barker, M. 1964. Klamath Grammar. University of California Press, Berkeley.

Borden, Gloria & Katherine Harris. 1984. Speech Science Primer: Physiology, Acoustics and Perception of Speech. 2nd edition. Baltimore: Williams and Wilkins.

Davis, Stuart. 1999. A Comprehensive Optimality-Theoretic Analysis of Syllable Contact in Korean. Paper presented at MCWOP 99.

Hayes, Bruce. 1986. Assimilation as Spreading in Toba Batak. Linguistic Inquiry 17, 467-500.

Harrell, Richard S. 1962. A Short Reference Grammar of Moroccan Arabic. Georgetown University Press: Washington, D. C.

Heath, Jeffrey. 1987. Ablaut and Ambiguity: Phonology of a Moroccan Arabic Dialect. State University of New York Press: Albany.

Hume, Elizabeth, Jennifer Muller & Aone van Engelenhoven. 1997. Nonmoraic Geminates in Leti. Phonology 14.3.

Hume, Elizabeth. 1998. The Role of Perceptibility in Consonant/Consonant Metathesis. Proceedings of West Coast Conference on Formal Linguistics 17. 293-307.

Hume, E., K. Johnson, M. Seo, G. Tserdanelis & S. Winters.1999. A Cross-Linguistic Study of Stop Place Perception. Proceedings of the XIVth International Conference of the Phonetic Sciences. 2069-2072.

Hume, Elizabeth. 2001. Metathesis: formal and functional considerations. In E. Hume, N. Smith & J. van de Weijer (eds.), Surface Syllable Structure and Segment Sequencing. Leiden, NL: Holland Institute of Linguistics.

Hura, S. L., Lindblom, B. & Diehl, R. 1992. On the role of perception in shaping phonological assimilation rules. Language & Speech 35, 59-72.

Johnson, Keith. 1997. Acoustic and Auditory Phonetics. Blackwell Publishers Ltd.

Jun, Jongho. 1995. Place assimilation as the result of conflicting perceptual and articulatory constraints. Proceedings of WCCFL 14, 221-237.

Kaplan, Howard L., Macmillan, Neil A. & C. Douglas Creelman. 1978. Methods & Designs: Tables of d' for variable-standard discrimination paradigms. Behavior Research Methods & Instrumentation. Vol. 10[6], 796-813.

Keegan, John M. 1986. The Phonology and Morphology of Moroccan Arabic. PhD. Dissertation. City University of New York.

Kohler, K. 1990. Segmental reduction in connected speech: Phonological facts and phonetic explanations. In W.J. Hardcastle & A. Marchal (eds.) Speech Production and Speech Modeling (pp. 69-92). Dordrecht: Kluwer Academic Publishers.

Krueger, John R. 1962. Yakut Manual. Bloomington, IN: IU Publications.

Levin, Juliette. 1988. A Place for Lateral in the Feature Geometry. NTC Publishing Group. 1997. NTC's Compact Swedish and English Dictionary. Lincolnwood, Illinois USA.

Poppe, Nicholas. 1963. Tatar Manual. Bloomington, IN: IU Publications.

n, Kwang-soo. 1987. Korean-Swedish Interlanguage Phonology. PhD. dissertation. University of Stockholm.

hg, K. L.& D. G. Sohl. 1981. Ponapean Reference Grammar. University Press of Hawaii, Honolulu.

Rice, K. & P. Avery. 1991. On the relationship between laterality and coronality. Phonetics and Phonology 2. The special status of coronals: internal and external evidence. Carole Paradis and Jean-Francois Prunet (eds.). 101-124.

Schulze, Wolfgang. 2001. An Online Grammar of Udi. Ludwig-Maximilians-Universität München. (http://www.lrz-muenchen.de/~wschulze/udinhalt.htm)

Sim, Ronald J., 1981. Morphophonemics of the Verb in Rendille. Afroasiatic Linguistics 8 (1), 1-33.

Steriade, Donca. 1995. Licensing retroflexion. Ms. UCLA.

Steriade, Donca. 1997. Phonetics in Phonology: The case of laryngeal neutralization. ms. UCLA.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: a perceptual account. In E. Hume & K. Johnson (eds.) The Role of Perception in Phonology. New York: Academic Press. 219-250.

Winters, Steve. 2000. Turning phonology inside out, or Testing the relative salience of audio-visual cues for place of articulation. OSU Working Papers in Linguistics 53, 168-198.

Zorc, R. David and Madina M. Osman. 1993. Somali-English Dictionary with English Index. Dunwoody Press: Kensington, Maryland.

OSU Working Papers in Linguistics 55, 70-87

VCCV Perception: Putting Place in its Place

Steve Winters
swinters@ling.ohio-state.edu

Abstract

Jun (1995) and Hume (1998) incorporate perception into analysis of cross-linguistic trends in place assimilation and metathesis by claiming that the perceptual salience of specific segments motivates the ranking of relevant OT constraints. This study investigates the specific claims Jun and Hume make concerning the perceptual salience of cues for stop place of articulation to determine whether their salience actually could motivate the proposed OT rankings. Since both Jun and Hume based their proposals on a consideration of cues for stop place of articulation in the appropriate (VCCV) context for place assimilation and metathesis, this study only tested the salience of stops in this context. Listeners heard unreleased stops of three places of articulation (labial, coronal, dorsal) and two manners (oral, nasal) in two stress patterns *preceding* pre-vocalic oral stops of three other places of articulation. The perceptual salience (as measured in d') of stops in this context did not always bear out the predictions made by Jun and Hume. Interestingly, labials were generally the most salient place of articulation while dorsals were the worst. Nasal stops also turned out to be more salient than oral stops. Less surprisingly, pre-vocalic stops were more salient than post-vocalic stops, and place salience was highest for stops preceding coronals in pre-vocalic position. The variable success of Jun's and Hume's proposed hierarchies of place salience underscores the need to test the empirical validity of hypotheses concerning the interaction of phonology and perception.

ntroduction

e role of perception in phonology has a long but largely unsung history, dating back to t least the early 1970's work of Björn Lindblom and his theories of adaptive dispersion in vowel spaces. The more recent influence of Optimality Theory in linguistic circles provides some new perspectives on how perception might influence phonology. Studies such as Jun (1995) and Hume (1998), for example, attempt to account for phonological processes such as place assimilation and metathesis by appealing to the perceptual salience of specific cues for place of articulation in stop consonants. Though the specifics of their accounts differ, both Jun and Hume propose that differences in perceptual salience can lead to different rankings of phonological constraints (which are encoded in terms of the articulatory intentions which define a speaker's grammar). Note that this subtle gap between perceptual salience and articulatory intentions implies that perception is not literally a part of phonology, but rather has an indirect influence on grammatical possibilities. However, the fact that different constraint rankings constitute different grammars in Optimality Theory has an interesting implication: if the relative perceptual salience of various kinds of sounds is universal, their corresponding constraint rankings would provide some limitation on the kinds of grammars that could possibly exist.

That Jun and Hume both draw phonological conclusions based on cues for place of articulation in stop consonants has interesting implications for the study of stop place perception. Universal or context-invariant acoustic information in the cues for specific places of articulation has been notoriously difficult to find (though note the work of Stevens & Blumstein 1978). This contrasts sharply with the acoustic characteristics of vowels, each of which has a comparatively uniform and easily identifiable pattern of formant frequencies to which it might be assumed the human perceptual mechanism directly responds. The apparent lack of invariant acoustic information for the place of stop consonants has led some to conclude that the perception of these sounds takes place not so much on the basis of a reaction to something that is "out there" but rather as the result of complex and highly specialized perceptual processing in the human brain (Liberman & Mattingly, 1985).

Nevertheless, researchers such as Miller and Nicely (1955), Wang and Bilger (1973) and Winters (2000) have attempted to identify universal patterns in the perception of various stop places on the basis of merely what a listener can hear coming in from "outside." Perhaps unsurprisingly, studies of this nature have yielded conflicting results concerning the relative perceptual "salience" for different stop places of articulation. With respect to the places labial, coronal and dorsal in pre-vocalic position, for instance, Miller and Nicely (1955) found that salience was highest for coronals, but not substantially different between dorsals and labials. Wang and Bilger (1973), in turn, found that salience was equally high for labials and coronals but lower for dorsals. Not to be outdone, Winters (2000) concluded that salience was highest for labials and dorsals but lower for coronal stops.

What to make of such empirical confusion? One difficulty in comparing res across these experiments is that each study used different methods, which may conc underlying commonalities in the results. Another problem is that taking such broa swipes at determining the universal "salience" of various places of articulation may ignore the troublesome context-based variance in the acoustic cues which signal stop place. Some variance may disappear when looking at individual contexts, or--even more specifically--at particular "packages" of acoustic cues for stop place. A listener may not necessarily generalize perceptual information across such contexts and packages in developing perceptually-based constraints for their Optimality Theoretic phonologies.

2. Theoretical Proposals

Interestingly, the perceptually-based OT constraints proposed by Jun and Hume only address the salience of various cues for stop place in specific contexts. Jun, for instance, accounts for cross-linguistic patterns in place assimilation by only considering what cues for place are present in the appropriate VCCV context for this process. If the first consonant in the CC sequence is a stop, the release burst of the first consonant is commonly dropped, thereby making the vowel-to-consonant transition the only cue for the first stop's place of articulation. From this observation, Jun concludes that the salience of stop place in post-vocalic position depends entirely on the relatively uniform acoustic characteristics of transition cues for the different places of articulation. In coronals, for instance, "Tongue tip gestures are rapid; thus, they have rapid transition cues. In contrast, tongue dorsum and lip gestures are more sluggish; thus, they have long transitions. Consequently, noncoronals have more robust perceptual cues than coronals." Jun's reasoning here seems to be based on the not unintuitive idea that extending the duration of acoustic information will increase the salience of that acoustic cue.

Jun's reasoning in comparing the relative salience of dorsal and labial cues, however, is slightly more complex.

> "Unlike labials and coronals, velars have an acoustic attribute, i.e., compactness (Jakobson, Fant and Halle, 1963). Velars can be characterized by a noticeable convergence of F2 and F3 of a neighboring vowel. These two formants can form a prominence in the midfrequency range. As argued and discussed by Stevens (1989), such a midfrequency prominence of velars can form a robust cue for place of articulation...Based on Stevens' claim, we assume that velars have an additional acoustic cue, i.e., compactness, for place of articulation, compared to coronals and labials."

By virtue of this reasoning, then, Jun claims that post-vocalic unreleased dorsal stops are more salient than labials, which are, in turn, more salient than coronals.

If these claims are true, they do a neat job of accounting for certain cross-linguistic tendencies to assimilate such unreleased, post-vocalic stops. In the spirit of Mohanan (1993), Jun performed a cross-linguistic survey of assimilation processes and noted a number of intriguing implicational relationships. For one, Jun notes that neither dorsals nor labials are targets of place assimilation unless coronals are as well. Secondly, he notes that dorsals do not assimilate unless labials do so, too. The pattern seems clear: a more salient place of articulation will not assimilate unless a less salient place already does so.[1]

Jun assumes that such patterns are assimilated into a speaker's grammar under the rubric of "preservation" constraints, which he defines as:

> (1) "Pres(X(Y)): Preserve perceptual cues for X (place or manner of articulation) of Y (a segmental class).
>
> Universal ranking: Pres(M(N)) >> Pres(M(R)),
> where N's acoustic cues for M are stronger than R's cues for M."

The appropriate ranking of preservation constraints for place in unreleased stops, then, would be:

> (2) Pres(pl(dor^{-}]) >> Pres(pl(lab^{-})) >> Pres(pl(cor^{-}))

Jun does not stop there; he also looks at patterns in place assimilation with regard to manner, syllabic position and trigger place Jun proposes the following constraint rankings for the relevant groups of sounds:

> (3) Manner: Pres(pl([stop]C)) >> Pres(pl[nasal]C))
> (4) Position: Pres(pl(onset)) >> Pres(pl(coda))
> (5) Trigger: Pres(pl(__cor)) >> Pres(pl(__noncor))

These rankings are also based on Jun's analysis of the relative salience for each sound group's context-dependent acoustic cues. Since his analysis of perceptual salience is based on speculation and not experimentation, however, it seems fair enough to ask if these conclusions are really valid. Is this really an example of perception influencing phonology or are these patterns the result of some other cross-linguistic influence?

Such questions seem even more relevant when looking at Hume's (1998) analysis of consonant/consonant metathesis. Hume proposes that this process may often be driven by perceptual factors; specifically, she claims that,

---

[1] Though see Tserdanelis and Hume (2000) for potential counterevidence to these assimilation patterns.

> "...by metathesis, a perceptibly vulnerable consonant shifts to a context in which the phonetic cues to the sound's identification are more robust, thereby enhancing the consonant's auditory prominence and, in turn, strengthening syntagmatic and paradigmatic contrast among sounds in a given language. By perceptibly vulnerable, I refer to a consonant with comparatively weak segment internal and/or contextual cues to, e.g., place and/or manner of articulation." (295-296)

The proposed role that the salience of acoustic cues plays in shaping phonological structures is slightly different here; instead of weakly-cued segments being eliminated (as in Jun), they are shifted into a context in which they would be more salient. The formal mechanism whereby such perceptual optimization is implemented is a family of "AVOID" constraints, which Hume defines as:

> (6) AVOID C/X: Avoid positioning a consonant (C) in a context (X) in which it is perceptually weak.

Whether or not perception influences phonology through a strategy of "avoidance" or "preservation"--or even some other strategy--is an interesting (and difficult) research question in its own right. But in this case a more tractable question is whether or not it really is the relative perceptual salience of different cues for stop place that is influencing cross-linguistic patterns in metathesis and place assimilation. Hume proposes that labials have relatively low salience in certain contexts, which can motivate their metathesis into a more salient context. To wit, Hume notes: "...labials can be considered particularly vulnerable given inherently short vowel transitions and relatively weak bursts, as compared to coronals and velars." This analysis can account for stop/stop metathesis in a language like Kui, where labials only metathesize when preceded by a dorsal in a stressed coda position. "The shift of the labial stop from an unstressed to stressed position at the expense of a velar in Kui is therefore not surprising, given that prosodic prominence in the language results in greater duration of transitions into the labial" (296).

However, Hume's claims about the "vulnerability" of labials seems to be at odds with Jun's proposal that labials are *not* the least salient place of articulation (in precisely the same context!). Part of the confusion here may stem from the fact that both researchers are *speculating* about what place cues are more or less salient; the rest of the confusion only follows from the lack of empirical data on which places (and cues) for stops are more or less salient.

Assertions about the relative "strength" or "weakness" of various acoustic cues for place beg the question of how, exactly, we might know whether an acoustic cue is "weak" or "strong". Hume and Jun base their claims on spectrographic analyses of typical labial, coronal or dorsal productions, but listeners who are actually in the business of acquiring and using phonologies have no such electro-mechanical luxury. These listeners have to base their categorical decisions of place salience upon their own auditory experiences--whatever their evaluative mechanism might be. As a matter of operational fact, then, the "strength" or "weakness" of acoustic cues could only affect phonological structure insomuch as they are reflected in listeners' perceived experiences of acoustic reality.

We need not speculate blindly about such experiences; listeners themselves can let us know what they are (within certain limits). So, given the proper interpretation of such experiences within some experimental paradigm, it should be possible to establish empirically what the relative strengths and weaknesses of various acoustic cues are. It should be possible, for instance, to investigate hypotheses such as those of Hume and Jun and determine whether their rankings of salience might genuinely serve as the motivation behind the phonological patterns they have found.

3. Methods

This study tested Jun's and Hume's claims about the perceptual salience of cues for stop place by investigating listeners' perception of vowel-stop-stop-vowel sequences. The utterances used to create the stimuli in this study were borrowed from Winters' (2000) study of audio and visual cues for place of articulation. All the original stimuli were of the form CVhVC, where the initial and final consonants were always identical, and both vowels were always [a]. These consonants could be either voiced oral or nasal voiced stops and could have either labial, coronal or dorsal places of articulation. There was also stress on either the first or second syllable of the nonsense CahaC word. Varying all of these factors made it possible to test Jun's and Hume's combined claims about salience in different syllabic positions, for different places and manners of articulation, and in stressed and unstressed syllables.

Two speakers, one male and one female, produced all of the relevant CahaC tokens while being videotaped in a sound-proof booth (for recording details, see Winters (2000)). For the original study, the videorecording of these production tokens was then digitized and edited into audio-visual and audio-only VC or CV tokens; the current study simply appropriated the audio-only tokens and digitally spliced them together to form the desired VCCV stimuli. Crucially, this study also eliminated stop bursts from coda position, since Jun's original proposals only considered the salience of unreleased coda stops. Practically speaking, this meant that the VC portion of the to-be-spliced-together VCCV stimuli contained the entire VC articulation up until a release burst (if any) or the offset of any noticeable closure voicing in the waveform. The CV tokens were then spliced directly after these edited VC tokens. The interval between the first vowel's offset and the second vowel's onset was a uniform 150 ms; in certain cases silence had to be inserted after the stop closure to augment the intervocalic duration (see Figure 1a). This particular time interval of 150 ms was chosen after it was found that shorter intervals generally induced a percept of only one consonant between the two vowels. Tokens with nasals in the coda position included nasal murmur during some of the 150 ms of intervocalic closure (see Figure 1b).

The resultant VCCV tokens could vary in place for both consonants, and could have stress on either the first or second syllable. Manner only varied in the first consonant, though, since not all nasals can appear in onset position in English. This meant that there were 3 (C1 place) x 3 (C2 place) x 2 (nasal/oral) x 2 (stressed/unstressed) = 36 token types; since these were produced by two different speakers, this amounted to 72 basic stimuli. These stimuli were randomized by computer and each presented twice to listeners in a sound-proof booth over headphones. After hearing each stimulus, a computer presented them with the question "What did the speaker say?" and gave them nine different responses to choose from (see Figure 2). These alternatives were written as if they were two words (e.g., 'ab da') and differed only in place of articulation for the coda and onset consonants. In order to reduce the listener's task to this point, the stimuli were presented to the listeners in separate blocks with nasal stops and oral stops.

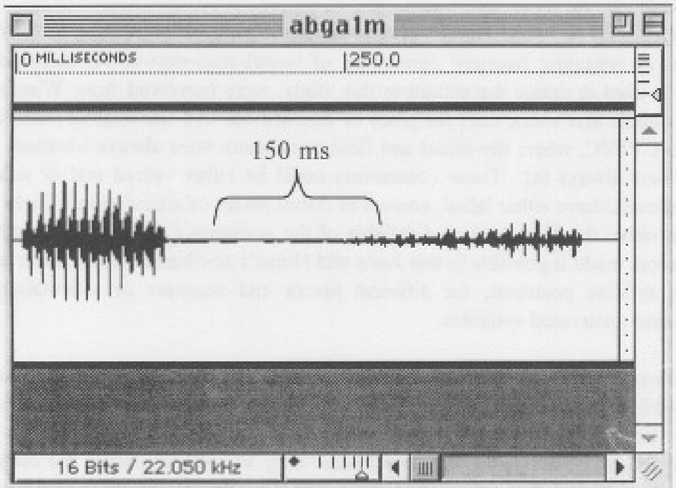


Figure 1a: Waveform for male production of “abga,” with stress on first syllable.

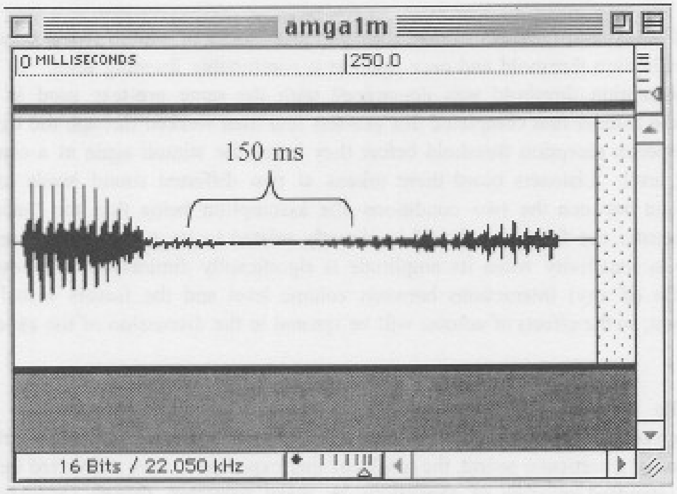


Figure 1b: Waveform for male production of “amga,” with stress on first syllable.

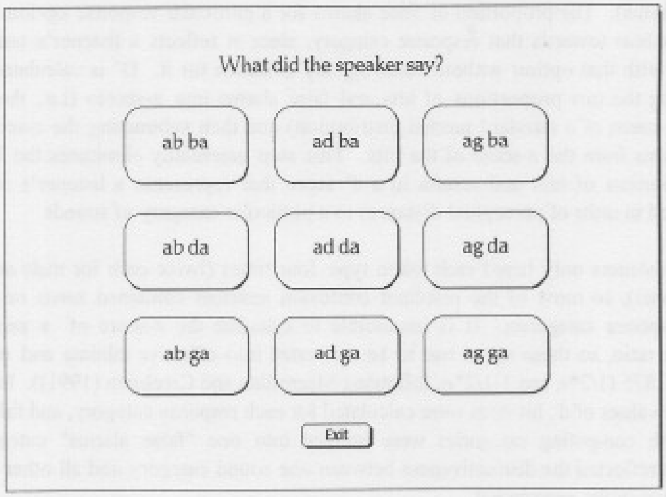


Figure 2: Presentation of experimental response alternatives

Twenty-four listeners worked through these blocks of stimuli twice: once at their speech reception threshold and once again at a comfortable listening level. A listener's speech reception threshold was determined with the same pre-test used in Winters (2000); the listener first completed this pre-test and then worked through the experiment at their speech reception threshold before they heard the stimuli again at a comfortable listening level. Listeners heard these tokens at two different sound levels to elicit a comparison between the two conditions--the assumption being that the "salience" of some acoustic cue for place should be directly related to its robustness in resisting a decrease in sensitivity when its amplitude is significantly diminished. However, there were little (if any) interactions between volume level and the factors tested in this experiment, so the effects of volume will be ignored in the discussion of the experimental results.

4. Results

Working on the assumption that measures of sensitivity most accurately reflect the "salience" of a particular sound, the results of this experiment were converted into scores of d', a standard measure of sensitivity in signal detection theory (MacMillan and Creelman, 1991), for each token type. Calculating d' involves eliminating listener *bias* in the experimental response options. Every time a listener gives a particular response (e.g., 'ab'), that response was either a *hit* (i.e., an 'ab' stimulus) or a *false alarm* (i.e., not an 'ab' stimulus). The proportion of false alarms for a particular response option reflects a listener's *bias* towards that response category, since it reflects a listener's tendency to respond with that option without receiving any evidence for it. D' is calculated by first converting the raw proportions of hits and false alarms into z-scores (i.e., the distance from the mean of a standard normal distribution) and then subtracting the z-score of the false alarms from the z-score of the hits. This step essentially eliminates the bias from the proportion of hits and results in a d' score that represents a listener's sensitivity (measured in units of perceptual distance) to a particular category of sounds.

Listeners only heard each token type four times (twice each for male and female productions), so most of the resultant confusion matrices contained zeros or fours for some response categories. It is impossible to calculate the z-score of a zero or one response ratio, so these ratios had to be converted into effective minima and maxima of .125 and .875 (1/2*n and 1-1/2*n, following Macmillan and Creelman (1991)). In order to calculate values of d', hit rates were calculated for each response category, and false alarms from both competing categories were lumped into one "false alarms" category. D' therefore reflected the distinctiveness between one sound category and all other response alternatives in the experiment.

As in the Winters (2000) study, the data yielded conflicting results concerning Jun's and Hume's proposals about the relative salience of different places of articulation. Appendix I gives raw confusion matrices for listener responses in all conditions, while the following figures show average d' values across listeners for the theoretically relevant conditions. Figure 3, for instance, shows listener sensitivity in D' to labial, coronal and dorsal places of articulation in both post-vocalic (coda) and pre-vocalic (onset) positions. Unsurprisingly, sensitivity to stop place was significantly higher in onset position, thereby verifying Jun's least controversial hypothesis (4). (See Appendix II for a description of statistical methods and specific results). However, the relative sensitivity of unreleased place in coda position contradicted Jun's assumptions in (2)--labial was the most salient place in this condition, followed by coronal, and then dorsal.

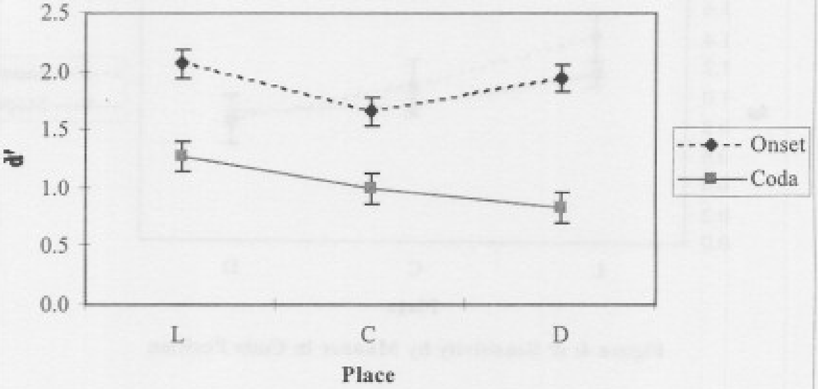


Figure 3: d' Sensitivity by Syllabic Position

The results also failed to bear out Jun's claims about the comparatively higher salience of oral stops over nasal stops in coda position (3). In fact, it re-confirmed the surprising result of Winters' (2000) study that, if anything, nasals are more salient than stops in this position. Measured in d' (Figure 4), there is no significant difference between sensitivity for nasal stops and oral stops; nasal stops just enjoy a slight sensitivity advantage. (Superimposed on these results is the same labial > coronal > dorsal pattern in sensitivity that was seen in Figure 3. However, coronals are significantly more salient than dorsals only in nasal stops.) For some reason not apparent in the acoustic signal, listeners actually seem to be more sensitive to place information in nasals than in oral stops in coda position.

One of Jun's more interesting claims was that place sensitivity in coda position was itself sensitive to the place of a following stop consonant (5); Jun assumed salience would be higher before coronal stops than non-coronals, due (again) to the rapidity of coronal gestures and the corresponding lack of articulatory overlap in comparison to non-coronal gestures. For a d' analysis (Figure 5), the results generally supported this hypothesis; Onset Place was a significant ANOVA factor (F = 39.802; df = 2,22; p < .001).

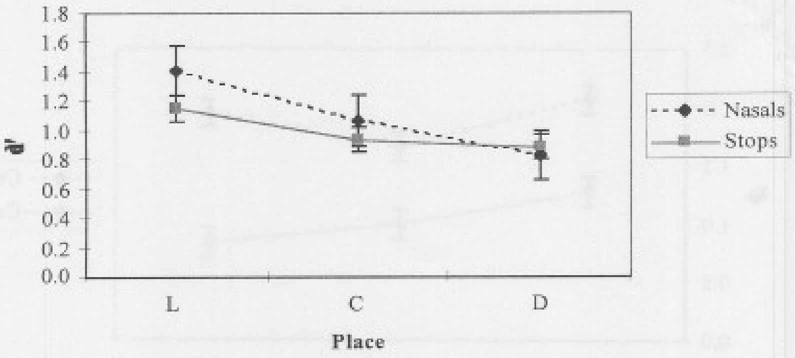


Figure 4: d' Sensitivity by Manner in Coda Position

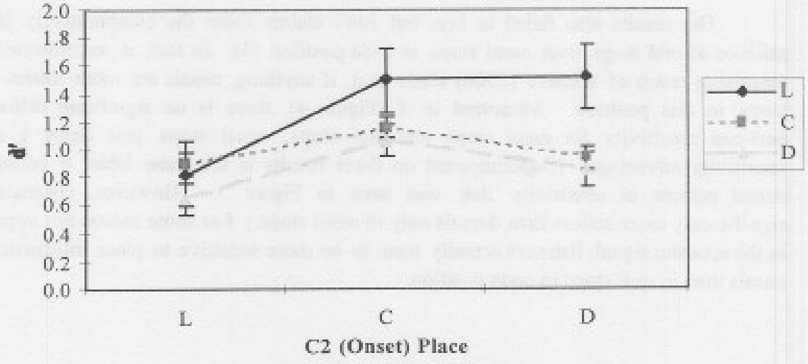


Figure 5: C1 (Coda) d' Sensitivity by C2 (Onset) Place

Post-hoc tests also showed that sensitivity was almost always significantly higher before coronal stops than before dorsals or labials; the only exception here were labial stops, which were not significantly more salient before coronals than before dorsals.

Though some of Jun's hypotheses seem to be supported by these results, the numbers do not bode well for Hume's hypothesis of labial stop "vulnerability". Labial salience seems to be particularly strong in any context, thereby seemingly invalidating any motivation to metathesize these segments into some more salient position. Looking at the specific context for dorsal-labial metathesis in Kui, however--a stop in a stressed coda followed by a stop in an unstressed onset--seems to show that the perceptual optimization of the *dorsal* stop may motivate this process. Figure 6 shows that (oral) dorsal stops in stressed codas have remarkably low salience in comparison to (oral) labial stops in the same position--a fact which is, of course, consistent with the results from Figure 3. In unstressed onset position, however, dorsal salience increases significantly while labial salience does not change drastically. The overall salience of a labial-dorsal stop sequence in this prosodic context would therefore be significantly higher than the overall salience of a dorsal-labial sequence--and it is precisely the more salient sequence that the speakers of Kui choose to produce. Although the labial stop vulnerability hypothesis may be incorrect, Hume's analysis of *why* this process occurs may be appropriate--Kui may be avoiding the production of dorsals in the weak (coda) context.

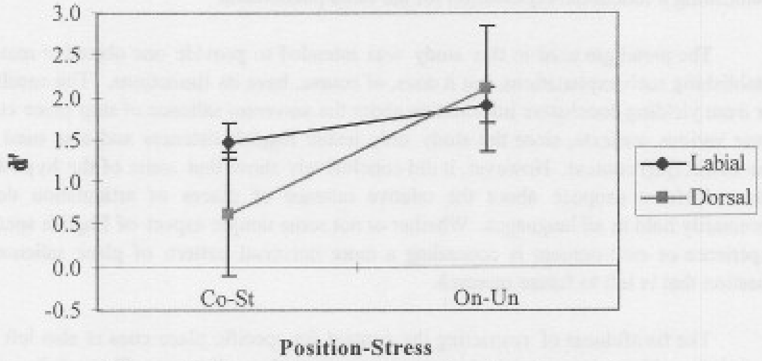


Figure 6: Place d' Sensitivity for stops by place and stress context

5. Discussion

The fact that communication is language's primary function no doubt plays a role in the kinds of phonological patterns we find in languages throughout the world. It is not unreasonable to suggest that the drive for communicative ease may spawn phonological processes that seem to be articulatory simplifications or acoustic enhancements. Nor is it unreasonable to expect that sound inventories will more commonly include articulatorily simple segments or vowels that are maximally dispersed throughout acoustic space (as in, e.g., Liljencrants and Lindblom 1972). These tendencies do not, of course, preclude the formal possibility for more complex articulations or vowels with unlikely formant patterns--but this is no reason to deny such tendencies any place in the theoretical analysis of language. Explaining grammatical patterns in language on the basis of their communicative function is no less valid (or interesting) than explaining them as purely formal entities. All that is really crucial--in *both* approaches--is establishing the empirical validity of the proposed explanation.

This is where functional analysis can run into trouble. The functional accounts proffered by Jun and Hume for cross-linguistic patterns in metathesis and place assimilation are easy enough to accept on an intuitive basis--who, for example, would not believe that cues for nasal stops are less salient than cues for oral stops? Without the empirical justification provided by studies such as this one, however, such assumptions may just as likely be untrue. Understanding that most language use takes the form of communication provides the linguistic imagination with a wealth of hypotheses about why we find the patterns in phonology that we do--but this is only the first step towards establishing a functional *explanation* for the same phenomena.

The paradigm used in this study was intended to provide one objective means of establishing such explanations, but it does, of course, have its limitations. The results are far from yielding conclusive information about the *universal* salience of stop place cues in these various contexts, since the study only tested English listeners and also used only one vowel ([a]) context. However, it did conclusively show that some of the hypotheses Jun and Hume propose about the relative salience of places of articulation do not necessarily hold in *all* languages. Whether or not some unique aspect of English speakers' experience or environment is concealing a more universal pattern of place salience is a question that is left to future research.

The fruitfulness of restricting the context for specific place cues is also left open to question. There is no *a priori* reason to presume that a listener will generalize across the acoustic manifestations of a particular sound in various contexts; the amount of perceptual detail that is available to a listener in constructing a constraint-based phonology is potentially limited only by the listener's psychophysical capabilities. Discovering what connections there may be between the psychophysical input in speech communication and the formal structures a listener develops in constructing a grammar is the exciting possibility offered by this line of speech perception research. Finding out what limits there might be to these connections--and thereby addressing the issue of *granularity* (Pierrehumbert, 1999)--is the further knowledge that this research may reveal for the study of cognition.

Acknowledgements

I would like to thank Keith Johnson, Beth Hume, Mary Beckman, Ilse Lehiste, Pauline Welby, Kiyoko Yoneyama, Jenny Vannest, Giorgos Tserdanelis, Jeff Mielke, Misun Seo, Matt Makashay, and all of Ohio State's PiPsters for their input on this work. I would also like to thank the participants of the S5 conference for their helpful comments on the first presentation of this paper. Part of this material is based on work supported under a National Science Foundation Graduate Fellowship.

References

Hume, Elizabeth (1998). The role of perceptibility in consonant/consonant metathesis. In K. Shahin et al. (eds.), Proceedings of the West Coast Conference of Formal Linguistics 17: 293-307.

Jun, Jongho (1995). Place assimilation as the result of conflicting perceptual and articulatory constraints. *Proceedings of the West Coast Conference of Formal Linguistics* 14: 221-237.

Liberman, Alvin and Mattingly, Ignatius (1985). The motor theory of speech perception revised. *Cognition*, **21**, 1-36.

Liljencrants, Johan and Lindblom, Björn (1972). Numerical simulation of vowel quality systems. *Language*, **48 (4)**, 839-862.

MacMillan, Neil and Creelman, Douglas (1991). Detection theory: a user's guide. Cambridge: Cambridge University Press.

Miller, G.A. and Nicely, P.E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, **27**, 338-352.

Pierrehumbert, Janet (1999). Formalizing Functionalism. In M. Darnell, E. Moravscsik, F. Newmeyer, M. Noonan and K. Wheatley (eds). Formalism and Functionalism in Linguistics, John Benjamins, Amsterdam. Vol. I, 287-305.

Stevens, Kenneth and Blumstein, Sheila (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, **64**, 1358-1368.

Tserdanelis, Georgios and Hume, Elizabeth (2000) Nasal place assimilation in Sri Lankan Portugese Creole: implications for markedness theory. Paper presented at MOT Workshop in Phonology, York University, Toronto.

Wang, M.D. and Bilger, R.C. (1973). Consonant confusion in noise: A study of perceptual features. *Journal of the Acoustical Society of America*, **54**, 1248-1266.

Winters, Stephen (2000). Turning phonology inside out: testing the relative salience of audio and visual cues for place of articulation. In Levine, R., A. Miller-Ockhuizen, T. Gonsalvez (eds.), *Ohio State Working Papers in Linguistics* **53**: 168-199.

			C2=Labial				C2=Coronal				C2=Dorsal	
			Respond				Respond				Respond	
	Unstressed	Labial	Coronal	Dorsal	Unstressed	Labial	Coronal	Dorsal	Unstressed	Labial	Coronal	Dorsal
	ab	101	41	50	ab	96	40	56	ab	109	31	53
Given	ad	41	89	62	ad	27	99	66	ad	35	89	68
	ag	39	46	107	ag	25	50	117	ag	31	57	104
	Stressed	Labial	Coronal	Dorsal	Stressed	Labial	Coronal	Dorsal	Stressed	Labial	Coronal	Dorsal
	ab	138	27	27	ab	133	21	38	ab	128	22	42
Given	ad	45	95	52	ad	30	113	49	ad	33	98	61
	ag	37	32	123	ag	23	35	134	ag	20	34	138
	Unstressed	Labial	Coronal	Dorsal	Unstressed	Labial	Coronal	Dorsal	Unstressed	Labial	Coronal	Dorsal
	am	135	30	27	am	148	26	18	am	151	15	26
Given	an	38	96	58	an	20	122	50	an	23	102	67
	ang	48	55	89	ang	38	43	111	ang	32	61	99
	Stressed	Labial	Coronal	Dorsal	Stressed	Labial	Coronal	Dorsal	Stressed	Labial	Coronal	Dorsal
	am	135	35	22	am	157	17	18	am	151	15	26
Given	an	29	108	55	an	24	110	58	an	19	105	68
	ang	68	39	85	ang	48	36	108	ang	39	54	99

			C1=Labial				C1=Coronal				C1=Dorsal	
			Respond				Respond				Respond	
	Unstressed	Labial	Coronal	Dorsal	Unstressed	Labial	Coronal	Dorsal	Unstressed	Labial	Coronal	Dorsal
	ba	292	72	20	ba	298	70	16	ba	276	92	16
Given	da	12	299	73	da	10	303	71	da	7	298	79
	ga	9	22	353	ga	9	36	339	ga	7	38	339
	Stressed	Labial	Coronal	Dorsal	Stressed	Labial	Coronal	Dorsal	Stressed	Labial	Coronal	Dorsal
	ba	306	48	30	ba	303	51	30	ba	279	52	53
Given	da	7	256	121	da	12	266	106	da	7	234	143
	ga	9	27	349	ga	11	25	348	ga	8	29	347

Appendix II

Once d' values had been calculated for each stimulus type (by subject), ANOVAs were run in order to determine which factors had significant effects on the variation in the d' values. The experiment included six basic factors:

1. Volume (Speech reception threshold/Comfortable listening level)
2. Manner (Oral/Nasal)
3. Place (Labial/Coronal/Dorsal)
4. Stress (Stressed/Unstressed)
5. Position (Onset/Coda)
6. Place in adjacent position: a. OnsetPlace (Labial/Coronal/Dorsal)
   b. CodaPlace (Labial/Coronal/Dorsal)

The last factor was included as a means of investigating Jun's claims about the effects of an adjacent place of articulation on the perceptibility of place in, for example, coda position. Unfortunately, including this factor in a six-factor ANOVA of the results would potentially yield significant but uninterpretable OnsetPlace*Place*Position or CodaPlace*Place*Position interactions. In order to avoid this problem, the six-factor ANOVA was boiled down to two five-factor ANOVAs and one four-factor ANOVA.

The first two, five-factor ANOVAs examined variance in sensitivity only for one syllabic position at a time--Place (in coda position) was one factor while OnsetPlace served as the adjacent place factor, for example. Table IA shows the results for this analysis of sensitivity for coda stops. In table IB, place in onset position served as a main factor while CodaPlace functioned as the adjacent place factor. In the third ANOVA, place was not considered as a factor at all, but syllabic position was. The results for these ANOVAs can be found in table IC.

In order to follow up on the significant results from the ANOVA testing, two-tailed t-tests were performed on the averages that were graphed in Figures 3-6. The t-tests made pairwise comparisons of the d' scores from which the graphed means had been drawn (for instance, comparing labial and coronals in coda position, Figure 3) and determined the likelihood of the two sets of scores having come from the same population. A probability of less than .01 was taken as signifying that the sets did not arise from the same population. These comparisons are shown in table II.

Table I:D' ANOVAs

A. Significant factors in Coda position

Factor	**F**	**df**	**p**
Volume	119.269	1,23	<.001
OnsetPlace	39.802	2,22	<.001
Stress	43.151	1,23	<.001
Place	39.584	2,22	<.001
Manner*Stress	14.365	1,23	0.001
Volume*Place	21.207	2,22	<.001
Manner*Place	13.610	2,22	<.001
OnsetPlace*Place	24.671	4,20	<.001
Volume*OnsetPlace*Place	6.909	4,20	0.001
Stress*Place	18.282	2,22	<.001
OnsetPlace*Stress*Place	23.705	4,20	<.001
B. Significant factors in Onset Position			
Volume	206.083	1,23	<.001
CodaPlace	11.110	2,22	<.001
Stress	8.563	1,23	0.008
Place	60.579	2,22	<.001
Manner*CodaPlace	6.973	2,22	0.005
Volume*Stress	11.111	1,23	0.003
Volume*Place	38.535	2,22	<.001
Stress*Place	17.242	2,22	<.001
Volume*Stress*Place	7.326	2,22	0.004
C. Significant factors across Positions			
Position	246.709	1,23	<.001
Volume	352.525	1,23	<.001
Manner	8.732	1,23	0.007
Stress	65.409	1,23	<.001
Position*Stress	10.297	1,23	0.004
Volume*Stress	16.208	1,23	0.001
Position*Manner*Stress	7.895	1,23	0.010

Table II: Probability (two-tailed t-test) that mean d' values graphed in each of the figures are the same

Figure 3	Coda	Onset
L-C	**0.00**	**0.00**
C-D	**0.00**	**0.00**
L-D	**0.00**	0.08

Figure 4	Orals	Nasals
L-C	**0.00**	**0.00**
C-D	0.39	**0.00**
L-D	**0.00**	**0.00**

Figure 5	L	C	D
L-C	**0.00**	**0.00**	**0.00**
L-D	**0.00**	0.49	**0.01**
C-D	0.63	**0.01**	**0.01**

Figure 6	Coda-St	Onset-Un
L-D	**0.00**	0.18

OSU Working Papers in Linguistics 55, 88-116

Lexical Effects in the Perception of Obstruent Ordering[1]

Matthew J. Makashay
makashay@ling.ohio-state.edu

1. INTRODUCTION

Cross-linguistic study of obstruent metathesis (Bailey 1970; Ultan 1971; Silva 1973; Hock 1985; Hume 1998, 2001; Steriade 2001) has attempted to understand this process. Many accounts delve into human auditory perception to aid in the explanation of these seemingly complex patterns of re-ordering.[2] Hock proposed a perceptual motivation for the preference of the ordering fricative-stop word initially (prevocalically), as there are clearer transitions between stops and vowels than between stops and fricatives.

Hume provided a systematic perceptual account for both place and continuant metathesis. The strengths and weaknesses of cues in different consonantal positions could give impetus for a consonant to switch to a position with better cues if there is little detriment to the other consonant, resulting in overall better perception of the sequence. An example of this process is the metathesis of /VkpV/ to [VpkV] in Kui (Hume 2001), shown in Table 1. The explanation for this place re-ordering involves the strong burst for

---

[1] Thanks to Elizabeth Hume, Keith Johnson, Mary Beckman, Michael Broe, Lisa Shoaf, Amanda Miller-Ockhuizen, Thomas Stewart, Kiyoko Yoneyama, Jeff Mielke, Steve Winters, Misun Seo, Huang Tsan, Georgios Tserdanelis, and Peggy Wong for their input. This material is based upon work supported under a National Science Foundation Graduate Fellowship.

[2] Articulatory explanations, which will not be focussed on in this paper, have also been proposed. For example, Bailey's observation that metathesis can result in apicals following nonapicals (and nondorsals following dorsals) led to his proposal that metathesis may be driven by a preference for a "natural" physiological ordering.

prevocalic velars (vs. the weak burst for labials) and the good vowel formant transitions to labials in postvocalic position (vs. good formant transitions to velars after certain vowels). The /k/ in postvocalic position ([VkpV]) is not as good perceptually as it would be in prevocalic position (Winters 2001). The gain of cues for /k/ by switching to prevocalic position with its burst more than compensates for the loss of linearity, since the consonant might have been perceptually lost otherwise. /p/ is relatively robust in its new postvocalic position, barely losing any cues besides its weak burst. Thus, a cluster containing /p/ and /k/ has the best overall perceptual cues identifying each segment if they are placed in this "optimal" order as [VpkV]. Conversely, the cluster [VkpV], with poorer cues overall in identifying each segment, is "non-optimal."

Table 1. Example of [kp] metathesis in Kui (Hume 2001)

Verb Stem	*Past* -te	*Pres. Part.* -pi	*Gloss*
ah-	ahte	ahpi	'to hold'
lek-	lekte	lepki	'to break'

The metathesis of stops and fricatives is accounted for by the fact that fricatives have strong internal cues (as well as external), while stops only have external cues such as vowel transitions or bursts. Therefore, a stop will gain better perceptibility if it can switch to a position with better external cues. Steriade (2001) states that sibilant-stop (ST) and stop-sibilant (TS) sequences are confusable, citing evidence from Pickett (1958) and Fay (1966) which she claims indicates that the linear order of adjacent consonants that share manner features (such as obstruency and continuency) is highly non-salient. Incorporating this with Hume's perceptual account forms her hypothesis that metathesis could arrive from listener error: Mishearing an obstruent cluster of a stop and a sibilant would be constrained by perceptual optimization, resulting in a sibilant-stop cluster if prevocalic, otherwise in a stop-sibilant cluster (if postvocalic). A stop will change positions to gain a good burst prevocalically. If there is no prevocalic position, it will switch to a postvocalic position to gain vowel formant transition cues.

However, while they are certainly less systematic than the optimal result, both prevocalic TS[3] and postvocalic ST[4] metathesis results have occurred historically across languages, but very rarely. Silva (1973) proposes that the cases involving metathesis to stop-sibilant only occur in languages with affricates[5], so speakers are previously accustomed to stop-sibilant sequences. Even though some data go against perceptual optimization, since there are so few examples, they could be the result of chance misperceptions that were learned. Although one cluster ordering is more "optimal" than

---

[3] Nakao (1986) has these examples from OE: ā[sk]ian > ā[ks]ian 'ask' (but also ā[ks]ian > ā[sk]ian), and a[sk]e > a[ks]e 'ashes'. Silva (citing Collinder 1960) notes a case in the development of Lappish into Mordvin: boaške 'the small of the leg' > pukšo 'the thick flesh; thigh, buttock'.

[4] Nakao has more examples from OE: tu[ks] > tū[sk] 'grinder' and wæps > wæsp 'wasp'. Silva has a case in the change from OE to ME: do[ks] > do[sk] 'dusk'.

[5] Which is true for all the languages in footnote 3.

another by having better perceptual cues, it just has a better chance of becoming an output, not the only chance.

Perceptual cues may not be the only factor involved in metathesis, as lexical effects may influence the ordering of obstruents as well. Much of the evidence used to support cue saliency of different consonants in various positions comes from perceptual studies. However, many of these studies do not take into account lexical effects, such as word frequency and phonotactics. These effects have been shown to affect perception (Luce 1986, Luce and Pisoni 1998, Pitt and McQueen 1998, Vitevitch and Luce 1999, Frisch et al 2000). If metathesis is motivated by the re-ordering of consonants to provide overall better cues to their identity, this should influence the cluster inventory of the lexicon. However, the lexicon may influence the perception of consonant clusters toward orders that occur more frequently.

This study will attempt to determine whether perceptual cues, lexical effects, or both influence the ordering of obstruents in American English. After a brief overview of possible acoustic cues that perceptually distinguish one consonant from another in English medial obstruent clusters, previous perceptual experiments on these clusters will be summarized, then the frequency of medial obstruent clusters that occur in English words will be examined, followed by a report of a perceptual experiment factoring in both acoustic and lexical information. English was chosen as the language of study because of the extensive previous research on it both phonetically and lexically. While phonetic research has been conducted cross-linguistically, only a handful of languages have enough analyzed corpora to yield spoken word frequencies, word familiarities, and other aspects of the lexicon that may influence the recognition of a word. In order to determine if different acoustics have a perceptual effect, any lexical factors need to be taken into account.

2. ACOUSTIC CUES

Before hypothesizing about how perception plays a role in metathesis, the assumed acoustic basis of the perceptual cues needs to be discussed. For the obstruents in question, stops and fricatives, the acoustic cues vary by phone and also by position within the cluster. Due to restrictions of the English lexicon—some clusters do not occur in enough words to have a large enough sample for testing—only stop-stop clusters composed of /p/, /t/, and /k/, and fricative-stop and stop-fricative clusters (fricative%stop clusters) composed of /p/, /t/, /k/, and /s/ will be investigated in the study.

2.1. Stop-stop clusters

Much research has been performed on determining the acoustic cues for stops (Delattre et al 1955, Öhman 1965, Blumstein and Stevens 1979, Kewley-Port 1983, Lahiri et al 1984, Stevens 1989, Wright 2001, for example). One conclusion that may be drawn for American English is that labials and velars have better vowel formant transition cues in postvocalic position than coronals. This is due to acoustic indications for place, such as the lowering of formants for labials, and the "velar pinch" for velars after front vowels, which are illustrated in Figure 1. While American English /t/ may have a good cue to its identity by being glottalized, (sometimes also surfacing as a glottal stop), some talkers glottalize their vowels, rendering this cue useless in many cases.

Another conclusion that may be drawn from this research is that velars and coronals have better burst cues in prevocalic position than labials. Coronals have a higher frequency of burst energy than velars and labials, and velars may have two or more bursts, shown in Figure 1. But labial bursts have weaker, more diffuse energy than velars or coronals. Vowel formant transition cues are also good in prevocalic position, being almost mirror images of postvocalic transitions, but can be obscured by aspiration after voiceless stops in English. Another factor is that labials have shorter transitions to the following vowel, which makes them more likely to be masked than velars or coronals.

Overall, prevocalic position is better than postvocalic position for the identification of stop place (Blumstein and Stevens 1979, Wright 2001). One indication of this is that CV syllable structures are preferred over VC cross-linguistically.

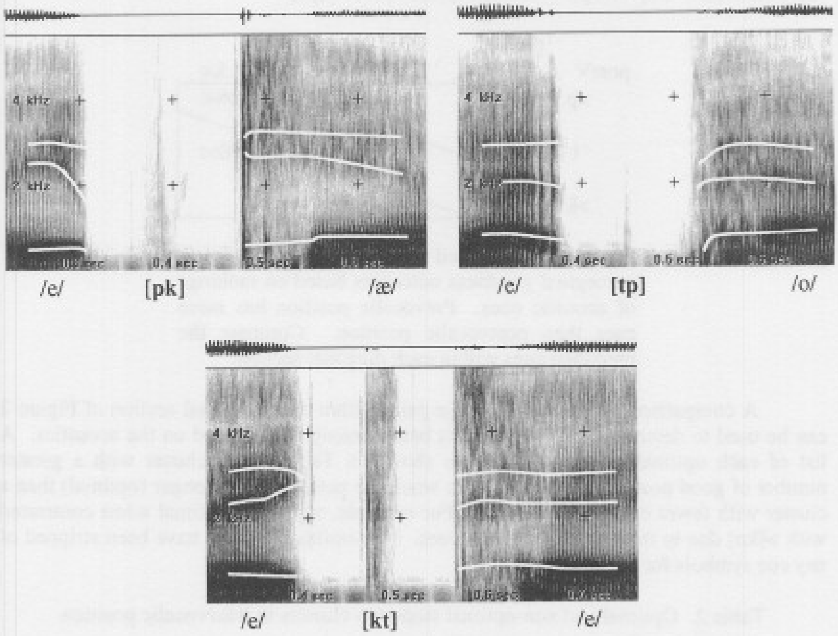


Figure 1. Note the falling of (traced) formant frequencies for postvocalic [p], the pinch of formants F2 and F3 for postvocalic [k] (for front vowels), and the glottalization that marks a following [t]. For the prevocalics, notice the strong double burst for [k] and the burst for [t], while [p]'s burst is very weak. Although the vowel formant transitions are obscured by aspiration, they are roughly symmetrical to those preceding the corresponding stop in postvocalic position.

Prevocalic position has burst information as well as transition cues, while postvocalic stops may be unreleased. Also, American English /t/ is not glottalized prevocalically, so it has good identification cues from its burst and following formant transitions. Figure 1 displays token utterances of an American English female illustrating the acoustic cues mentioned above for postvocalic and prevocalic [p], [t] and [k].

Based on the acoustic evidence discussed, predictions can be made on optimally positioned stops, and the clusters they compose. An optimal postvocalic stop will be indicated by a preceding '>' that symbolizes perceptually good, right-pointing cues (e.g. >[k]), and an optimal prevocalic stop will be indicated by a following '<' that symbolizes perceptually good, left-pointing cues (e.g. [k]<). As illustrated in Figure 2, postvocalically, labials and velar stops have better place cues than coronals, so >[p] and >[k] outrank [t]. Prevocalically, coronals and velars have better cues than labials, so [t]< and [k]< outrank [p]. Furthering the symbolism, optimal clusters will be encased in '> <'. For example, since >[p] has good postvocalic cues and [t]< has good prevocalic cues, the cluster they compose will be represented as >[pt]<.

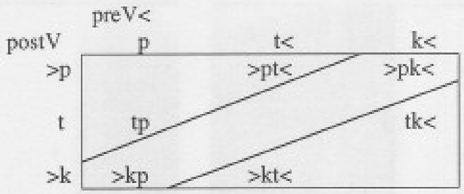


Figure 2. Predicted American English cluster perceptual goodness outcomes based on rankings of acoustic cues. Prevocalic position has more cues than postvocalic position. Compare the predicted cues within each diagonal section.

A comparison of the mirror-image pairs within each diagonal section of Figure 2 can be used to determine which cluster is better perceptually, based on the acoustics. A list of each optimal/non-optimal pair is shown in Table 2. A cluster with a greater number of good post and prevocalic cues would be perceptually stronger (optimal) than a cluster with fewer cues (non-optimal). For example, >[pk]< is optimal when contrasted with >[kp] due to the number of better cues. Non-optimal clusters have been stripped of any cue symbols for easier readability.

Table 2. Optimal and non-optimal stop-stop clusters in intervocalic position

optimal	non-optimal
>pt<	tp
>pk<	kp
>kt<	tk

2.2. Fricative-stop vs. stop-fricative intervocalic clusters

Fricatives always have internal fricative noise frequencies as a place cue, whether they are pre or postvocalic. Stops, on the other hand are better in prevocalic position, as shown above. Determining the ordering of intervocalic fricatives and stops is better when the stop follows the fricative because there is separation between frication and burst

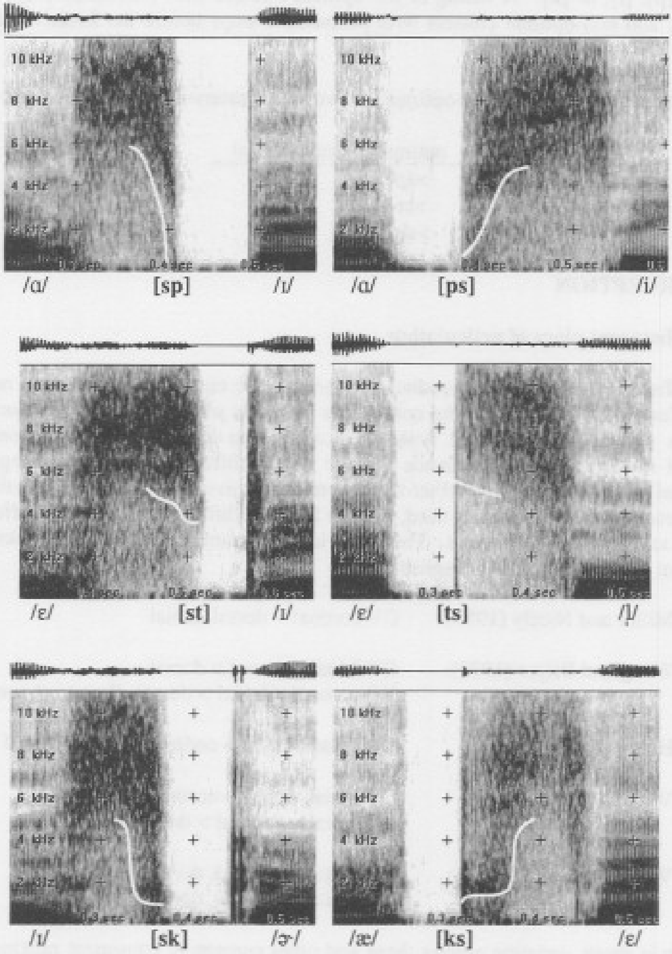


Figure 3. [s] always has internal noise as evidence to its place, but also has transitions to and from stops of a different place. White lines underscore the lowest peak of spectral energy in [s].

(frication followed by silence, then the stop burst) as opposed to when the stop precedes the fricative (silence followed by the burst, then frication) (Wright 2001). Although forward masking can occur for up to 75 to 100ms after the fricative ends, which could still overtake many following stop bursts, the burst of a preceding stop is well within the 50ms limit for backward masking (Yost, 1994). There are also stop place cues in the transitions to or from a neighboring fricative. Figure 3 displays examples of clusters with [s] and [p], [t], or [k]. A listing of the optimal clusters with prevocalic stops after the fricative, and non-optimal clusters with postvocalic stops before the fricative are shown in Table 3.

Table 3. Optimal and non-optimal intervocalic clusters with fricatives and stops

optimal	non-optimal
>sp<	ps
>st<	ts
>sk<	ks

3. PERCEPTION

3.1. Obstruent place of articulation

Testing the perceptual goodness of the acoustic cues of obstruents has not been a straightforward affair. Presenting consonants in words and non-words to subjects at low volumes and with background noise to examine the listening errors has resulted in different rankings of place salience. Some of the differences in the findings can be explained by the inventory (whether there were fricatives or voiced stops), whether bursts were present or not, the vowels used, and the fact that different talkers have different oral cavities and articulation patterns. The following are a sampling of salience rankings from studies of place perception in English:

(1) Miller and Nicely (1955): CV coronal > dorsal, labial

(2) Wang and Bilger (1973): CV labial, coronal > dorsal
VC coronal > labial > dorsal (burst not specified)

(3) Hume et al (1999): CV dorsal, labial > coronal (English and Korean)

(4) Winters (2001): CV labial, dorsal > coronal
VC labial > coronal > dorsal (burstless)

(5) Wright (2001): CV labial > coronal, dorsal (burstless)
VC labial > coronal, dorsal (burstless)

As there is much variation among these and other perceptual consonant ranking studies, the apparent overall trend of the ranking positions, modified by American English acoustics and the particular talker's speech characteristics, led to a set of optimal perceptual clusters identical to the optimal acoustic clusters in Table 2.

3.2. Segment ordering

One perceptual perspective that accounts for temporal disordering of stops and fricatives, as Steriade (2001) claims, is auditory streaming (Bregman 1990). In this account, the high frequency of fricative noise is perceptually far enough away from vowel formants to separate speech into separate streams—one containing fricatives, and one containing vowels and other sonorants. Temporal ordering across streams is difficult, as there are few acoustic cues that line up in both streams: Vowel formant transitions that give stop place cues are lower than frication frequencies, and a stop before a fricative could have its burst masked by the fricative. Switching the stop to a position in which it has a strong burst would bring it into the fricative stream, which is the expected result of metathesis prevocalically ([STV]). A stop that is preobstruent or phrase final may be unreleased, resulting in lack of evidence for it following the fricative. This may increase its chance of being ordered before the fricative ([VTS{C,#}]), instead of after ([VST˺{C,#}]), similar to the patterns observed in Faroese and Lithuanian (Seo and Hume 2001).

There also have been perceptual studies that have observed metathesis errors by subjects listening to clusters, and a few have tested aspects of the linear ordering of segments in clusters. One example is Pickett (1958), as noted above, which tested the perception of consonant clusters in noise (using flat noise with a signal-to-noise ratio at -4 dB and at +6 dB, and low-frequency noise with a spectrum slope of -12 dB per octave with a signal-to-noise ratio at -30 dB). Only the final consonant cluster syllables—bVCC—had alternate sibilant-stop pairs ([ts] and [st], as well as [ks] but not its pair). The largest reported listener error for the coronal pairs in the -4 dB flat noise was [ks], and the second was perceptual metathesis[6] (with a higher rate for [st]). In the low-frequency noise, the largest error for [ts] was [ks], followed by [ls] and then [st]. [st] did not have as high error rates. However, some of the stimuli used were actual English words, while others were not. English words containing [i], [ɑ], and [o] formed by bVks (*beaks* and *box*) have spoken and written frequencies over four times higher than bVst words (*beast*, *bossed*, and *boast*), which had over 16 times higher written frequencies (but similar spoken) than bVts words (*beets/beats*, and *boats*). Also, the responses were forced choice, preventing alternatives such as [sk] or [p].

Fay (1966) investigated subjects' temporal resolution of voiced non-plosive pairs (including nasals, fricatives, and liquids) and pure-tone pairs in noise, with no surrounding context. Staggered onsets with different lag and lead times of voiced non-plosives were played to subjects' right ears with equal offset times. The task was to determine which consonant came first, with onset lead and lag times of 70, 50, 30, 10, and 0 ms. Although stops were not used, nasal-fricative sequences seemed to break the expected pattern of fricative-stop clusters being easier to perceive word initially, with the timing of nasal-fricatives ([nz] and [nð]) perceived correctly more often than corresponding fricative-nasals. (Though there is a bias in hearing [ð] first in the [nð] pair.) The median scores were 100% for seven phoneme pairs out of twelve, and three pure-tone pairs out of four, but half the phoneme pairs had better temporal resolution than

---

[6] The term "perceptual metathesis" is used to indicate that the process is not incorporated in the grammar, as it was heard in manipulated laboratory speech and is not used systematically.

pure-tone pairs. Fay explains this by suggesting that linguistic experience gave subjects higher accuracy on temporal resolution of individual phoneme pairs than less natural pure-tone pairs.

Bond (1971) performed a perceptual experiment on the perception of stop-sibilant and sibilant-stop clusters ({p,t,k}s, s{p,t,k}) inter-vocalically and postvocalically in English words. White noise was added to spoken words to attain different signal-to-noise ratios of 0 dB, +12 dB, and -6 dB. Subjects were told before the test that some of the words were unusual, and were shown them. The test was presented twice, with responses written the first time, and spoken the second. Bond found that the most common error is perceptual metathesis, with the sibilant-stop clusters perceived correctly less often than the corresponding stop-sibilant. However, in inter-vocalic position the sibilant-stops were heard as stop-sibilants less often than the stop-sibilants were heard as sibilant-stops. But the sibilant-stop tokens used (*Caspian*, *blister*, and *asking*) have higher Kucera-Francis written frequencies and Brown verbal frequencies than the stop-sibilant tokens (*Capsian*, *blitzer*, and *axing*), so the result could be a lexical frequency effect, as the subjects would expect to hear the more common words.

Although many studies have tested the perceptual cues of obstruents, it appears that some of the results are contradictory. Place salience rankings are not in agreement with each other, and neither are the preferences for stop-fricative orderings. One possible explanation for this variation could be lexical effects, as the number of words that contain an obstruent sequence may be as important for recognition as its perceptual cues.

4. LEXICAL COUNTS AND FREQUENCIES

4.1. Counts

Given the acoustic and perceptual ranking of obstruent clusters in Table 2 and Table 3, the prediction is that the number of optimal clusters is greater than the number of non-optimal clusters in English words. All else being equal, if there are diachronic changes due to misperceptions, the optimal clusters should be more stable and therefore be in more lexical entries than the non-optimal clusters. These predictions are represented in Table 4.

Table 4. Predictions of lexical counts of words with optimal and non-optimal intervocalic clusters.

Number of words with optimal cluster	>	Number of words with non-optimal cluster
>pt<	>	tp
>pk<	>	kp
>kt<	>	tk
>sp<	>	ps
>st<	>	ts
>sk<	>	ks

Tallying the number of English words with medial obstruent clusters listed in the CELEX lexical database (Baayen, Piepenbrock, and Gulikers, 1995) results in totals that mostly support these predictions. For both all English words (Figure 4) and only monomorphemic English words (Figure 5) the optimal clusters >/pt/<, >/kt/<, >/sp/<, and >/st/< occur more often than their non-optimal counterparts, while >/pk/< does not appear to be much more common in words than /kp/. /ks/, the only cluster in English with one alphabetic letter, *x*, appears in more words than optimal >/sk/<, and also occurs in more words than all other non-optimal clusters combined. Because of orthography, and large lexical representation, /ks/ may be a better perceptual unit for English speakers than >/sk/<.

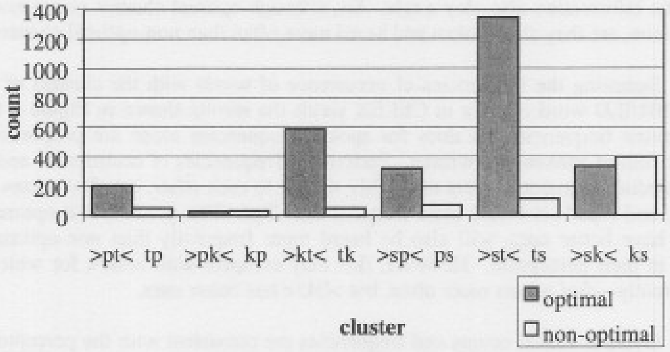


Figure 4. Word count of *VCCV* English words in the 52.5 thousand word pronunciation dictionary in CELEX.

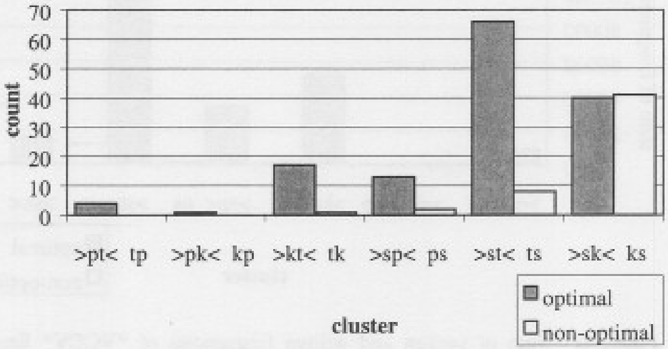


Figure 5. Word count of monomorphemic *VCCV* English words in the 52.5 thousand word pronunciation dictionary in CELEX.

4.2. Frequencies

While the ratio of words with optimal and non-optimal clusters in the lexicon supports the prediction, the usage of these words could affect their perceptibility. Clusters that are spoken more often will be heard more often, perhaps tuning the perceptual system towards detecting their cues more accurately than for clusters heard less often (Frisch et al 2000). Perception of words is affected by their frequency of occurrence, the number of neighboring words that are phonetically similar to them, the predictability of the segment sequences, and how familiar they are to the listener, among other factors (Pollack et al 1959, Savin 1963, Luce 1986, Luce and Pisoni 1998, Pitt and McQueen 1998, Vitevitch and Luce 1999, Frisch et al 2000). Pragmatic, semantic, and syntactic information also play a role. So, although optimal clusters occur more often in the lexicon, are they also spoken and heard more often than non-optimal clusters?

Summing the frequencies of occurrence of words with the clusters of interest in the COBUILD word corpora in CELEX yields the results shown in Figure 6 for spoken and written frequencies. Values for spoken frequencies alone are proportional to the overall sum of spoken and written. Patterns of frequencies of occurrence, and counts of word medial obstruent clusters are highly similar to each other, e.g. /ks/ occurs more than >/sk/<, and >/pk/< is barely more frequent than /kp/. This predicts that optimal clusters, which have better cues, will also be heard more frequently than non-optimal clusters, aiding in their perception. However, /ks/ may compete with >/sk/< for which is better perceptually—/ks/ occurs more often, but >/sk/< has better cues.

Overall, lexical counts and frequencies are consistent with the perceptual account of optimality. But, since there are cases in which they are at odds with each other, both perceptual and lexical effects were controlled for in the experiment.

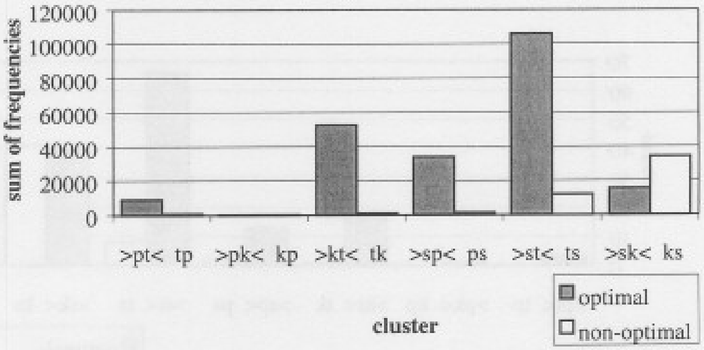


Figure 6. Sum of spoken and written frequencies of *VCCV* English words in the COBUILD 16.6 million written word corpus and in the COBUILD 1.3 million spoken word corpus in CELEX.

5. EXPERIMENT

To test whether the perceptibility of clusters depends on acoustic and lexical information, a lexical decision and repetition task based on natural, spoken American English words containing word-medial obstruent clusters ((C)*VCCV(C)*, e.g. *napkin*) was performed. Actual spoken words were used instead of nonsense syllables, since in ordinary speech communication listeners are trying to perceive meaning in words transmitted acoustically, not the order of consonants. The number of lexical items that contain a cluster may influence the perception of that cluster, which would be the result of word frequency and neighborhood effects (Luce 1986, Luce and Pisoni 1998, Pitt and McQueen 1998, Vitevitch and Luce 1999, Frisch et al 2000). Thus, words with high and low spoken and written frequency were used. Non-words that are the metathesized pairs of these words (e.g. [nækpɪn] for *napkin*) tested if subjects perceived the obstruent order correctly or not. If they heard the order correctly, the subjects would have decided that the token is a non-word ("*nakpin*"). If the subjects did not perceive the obstruents correctly and perceptually metathesized them, they would have decided that the token is a word (*napkin*). This is similar to a mispronunciation detection task.

The types of obstruent clusters used in the experiment had place differences for the stop-stop clusters ({p,t,k}-{p,t,k}), and continuant (and sometimes place) differences for the fricative-stop clusters (s-{p,t,k}) and stop-fricative clusters ({p,t,k}-s). The stop clusters tested the effects of place on confusability, while the clusters with fricatives and stops tested if the perceptual optimization hypothesis is supported by confusions resulting in metathesis only surfacing as fricative-stop before a vowel. Steriade's claim that the sharing of manner features corresponds with the non-saliency of the linear order of two consonants was also tested, and would be supported if the clusters with fricatives and stops metathesized less than clusters with only stops when the subjects identified the stimuli as words or not.

5.1. Predictions

A non-word token will more likely be perceptually metathesized to form a real English word if the resulting cluster is optimal (controlling for word frequency and neighborhood density), as demonstrated in Figure 7. If a subject hears a non-word token with a non-optimal cluster, there may be confusion as to the ordering of the cluster. If there is a real word that can be formed by metathesizing to an optimal cluster, the subject may decide the real word was what was actually spoken. Or there may be confusion, causing a longer reaction time, but the subject finally decides that the token is not a word. If a subject hears a non-word token with an optimal cluster, there should be little confusion as to the ordering of the cluster. The subject should quickly decide that the token is not a word. Subjects' responses should parallel the outcome from historical and grammatical metathesis: non-optimal clusters switching to optimal, and optimal clusters being maintained should be the overwhelming pattern.

Perceptual improvement effect

non-optimal => optimal	optimal ≠> non-optimal
False alarm or slow	Correct rejection & fast
Stimulus => Percept	Stimulus ≠> Percept
ks => >sk<	>sk< ≠> ks
[wɪksi] => whiskey	[tæski] ≠> taxi

Figure 7. Non-words containing clusters with "poor" acoustic cues like [wɪksi] should metathesize to English words with "good" acoustic cues. Non-words containing clusters with "good" acoustic cues like [tæski] should not metathesize to English words with "poor" acoustic cues.

The expected word frequency effect will be controlled for by balancing the overall group frequencies between optimal and non-optimal word pairs, but this cannot be done for the cluster frequencies themselves. As discussed in section 4, the number of words that contain a particular cluster and how frequently these clusters are used in speech can vary widely between optimal and non-optimal pairs, in some instances by a factor of ten. The result may be that there is a cluster frequency effect, which would be demonstrated by the listeners' faster reactions to or higher accuracy for words that contain high frequency clusters than for words with low frequency clusters. If this were the case, the results should be the same as for the optimality condition, since for most cluster pairs the optimal one also is the most frequent. The two cluster pairs that would go against this pattern are >[pk]< and [kp], which have roughly the same count and frequency and therefore should show no effect, and >[sk]< and [ks], in which the non-optimal cluster has a higher count and frequency and therefore would aid better performance.

5.2. Methods

Stimuli The stimuli were composed of targets and two types of foils. In order to minimize possible word-level stress effects, the attempt was made to only use words with the same stress pattern. The CELEX database was used to find trochaic[7] English words that also met the required cluster criteria. The targets were non-words produced by metathesizing the medial obstruents in these words. For example, [tæski] was created by metathesizing the [ks] in *taxi*, and [rɛtpa^{ɪ}l] was created by metathesizing the [pt] in *reptile*. Other English words with medial obstruents were used for real word foils (e.g. *ritzy* and *dropkick*). The non-word foils had zero phonological neighbors (by addition, subtraction, or substitution of a phone) and came from the substitution of medial obstruent clusters into English words with zero frequency and zero neighbors. For example, [flæspʌn] was created by substituting [sp] into *flashgun* and [ha^{ʊ}tkɔg] was created by substituting [tk] into *housedog*.

---

[7] As there were not enough trochaic words to provide an adequate number of tokens in each cluster group, some compounds with primary stress on the first syllable were used as well.

The list of 120 targets and their lexical sources appears in Appendix A, grouped by the resulting metathesized clusters. The clusters used are >[pt]< and [tp], >[pk]< and [kp], >[kt]< and [tk], >[sp]< and [ps], >[st]< and [ts], and >[sk]< and [ks]. Ten words per cluster for twelve clusters yield 120 targets. Cluster pairs (e.g. >[sk]< and [ks]) are balanced so they have similar word onsets and offsets and have similar total frequencies of occurrence.

The 120 English word foils are shown in Appendix B. These words have the same clusters used in the metathesized tokens, and have similar word onsets and offsets. However, due to the limited number of English words with the VCCV pattern, some of these tokens have consonants adjacent to the medial obstruents. Nonetheless, the word foils were constructed to be phonetically similar to the targets.

The list of 120 non-word foils and their lexical sources are shown in Appendix C. None of the English source words have a medial obstruent cluster used in the experiment. These words have a zero frequency of occurrence, and have a neighborhood density of zero (since no word in the CELEX database was phonemically similar based on the additions, subtractions, or substitutions of a single segment). Similarly, the non-words created by substituting the medial cluster with one of the 12 clusters used in the experiment have zero neighbors as well. Each obstruent in the substituting cluster was the result of a change in place or manner (and optionally voicing)[8] of the original obstruent. For example, the medial cluster in *squad car* was changed to [st] to make the non-word foil [skwɑs.tɑr] by changing the manner (and voicing) of /d/ to yield [s], and the place of /k/ to yield [t]. These non-words vary in degrees of "word-likeness" as an attempt to increase the difficulty of separating words from non-words. The first two tokens for each cluster do not contain a word for either syllable (e.g. [slut]-[pælv] from *sluice-valve* in the [tp] group). The following three tokens contain a word only in the second syllable (e.g. [stɑt]-*par* from *stockcar*). The next three tokens contain a word only in the first syllable (e.g. *pit*-[pot] from *pigboat*). The final two tokens contain words in both syllables (e.g. *greet-pun* from *grease-gun*).

In total, there are 360 stimuli in the lexical decision task: one-third are real word foils the subjects should reply YES to, one-third are non-word foils the subjects should reply NO to, and the remaining third are metathesized targets the subjects may reply YES or NO to depending on whether or not the tokens are perceptually metathesized to form real words. The foils also helped determine if the subject performed the task correctly.

Talker The talker was a female native Ohio English speaker with phonetic knowledge and no known speech or hearing disorders.

Procedure for talker Randomized lists of the tokens were read at a steady rate until three accurate repetitions were achieved. For the non-word targets and foils, the intended pronunciation was elicited by displaying the English word, followed by the cluster to substitute word medially:

---

[8] Some of these non-words resulted from a change only in voicing, as there were not enough English words that satisfied the 0 frequency/0 neighbors constraint.

(6) pizza
st

root beer
pk

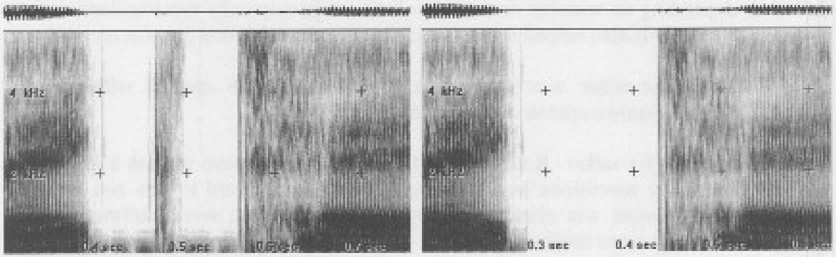
The talker would say the English word, followed by the non-word with the substitution. Stimuli were recorded onto DAT-tape using a Shure SM10A head-mounted microphone, and re-digitized to create computer soundfiles at 22.05 kHz.

The digitized words were edited to ensure that the soundfiles began with word onset and ended with word offset. Since the talker did not release all postvocalic stops in stop-stop clusters (and since this occurs in natural speech), the amplitudes of all stop-stop closures were reduced to zero, even if there was no detectable burst. This is demonstrated in Figure 8. Clusters with stops and fricatives were unaltered.

Two phonetically trained researchers naïve to the purpose of the experiment judged the accuracy of the pronunciations based on a provided list of transcriptions. Items that did not score 4 or above on a 5-point goodness scale by both judges were discarded, which only occurred for less than 1% of the cases.

Listeners The listeners were 30 native Ohio English speakers who were undergraduates at The Ohio State University with no known speech or hearing disorders. 20 of them heard the stimuli at a comfortable listening level (CLL group), and 10 of them heard the stimuli nearly at their speech reception threshold (SRT group). The listeners received partial course credit for their participation.

Procedure for listeners The experiment involved two tasks—an auditory lexical decision task (Goldinger 1996) and a repetition task. The purpose of the repetition task was to confirm that if a subject decided a metathesized target was a word, then the subject had indeed metathesized it to the intended word, and did not make a different error to create some unrelated word. For example, if the subject heard the target [mɪskɚ], decided it was a word, and stated it was *mister*, then it would not be treated as a



[kt^h] [k˺t^h]

Figure 8. For all stop-stop clusters, the amplitude of the signal was reduced to 0dB from after closure of the first stop to before release of the second.

case of perceptual metathesis. But if the subject decided it was a word and said *mixer*, that would be considered an example of metathesis.

The MEL program was used to run the experiment from a PC, collecting reaction time manual responses from a button box. The stimuli were played and the subjects' oral responses were recorded on a Sennheiser HMD410 headphone/microphone. A Quest Electronics Model 155 impulse precision sound level meter measured the stimuli for the comfortable listening level at approximately 75dB SPL with A weighting and F response, ±5dB depending on vowels and consonants. The near-speech reception threshold[9] was at approximately 40dB SPL, same conditions. Oral responses were recorded onto professional audio-tape at half-speed in order to fit an average 45min session on one side of a 60min tape.

Subjects were informed that the first task was to decide whether an English word[10] was spoken or not. They were to press the "YES" button with their right index finger if they thought the token was a word, otherwise they were to press the "NO" button with their left index finger. After making the lexical decision, the subjects were instructed to perform the second task of repeating aloud what they heard, as best they could.

The listeners performed the tasks individually in a sound-attenuated room. After a practice trial using a representative selection of word and non-word foils to ensure the subjects' comprehension of the task, the stimuli were randomly presented in six blocks. Subjects were allowed to pause after each block, and were given a rest break after the third block.

RT analyses of the CLL group were performed on the correct rejection of non-word targets (the metathesized words) as words. To determine what types of perceptual errors listeners made, their audio-tapes were transcribed auditorily and through the examination of spectrograms. Error analyses of the SRT group were performed on the metathesis and non-metathesis errors.

5.3. Results and discussion

RT analysis of CLL correct rejection of non-word targets Overall, CLL subjects were *slower* to identify non-word targets with optimal clusters ([tæski] from *taxi*) than those with non-optimal clusters ([wɪksi] from *whiskey*), as shown in Figure 9. There was a significant effect of optimality on RT (optimal cluster words were slower than non-optimal ones by 79ms, $F = 34.624$, $p < .05$), and obstruent type (stop-stop cluster words were slower than those with fricatives by 43ms, $F = 24.053$, $p < .05$). There was also an interaction between optimality and obstruent type ($F = 7.121$, $p < .05$). Further, there was a significant effect of obstruent ordering on RT (stop-fricatives are faster than fricative-stops by 59ms, and are faster than stop-stops by 72ms, $F = 16.765$, $p < .05$).

---

[9] 40dB SPL is the level that corresponded to SRT for most participants in Winters (2001).

[10] Subjects were instructed to treat compound words like *greenhouse*, *one-way*, and *ice cream* as single words—anything they would expect to find listed in the dictionary.

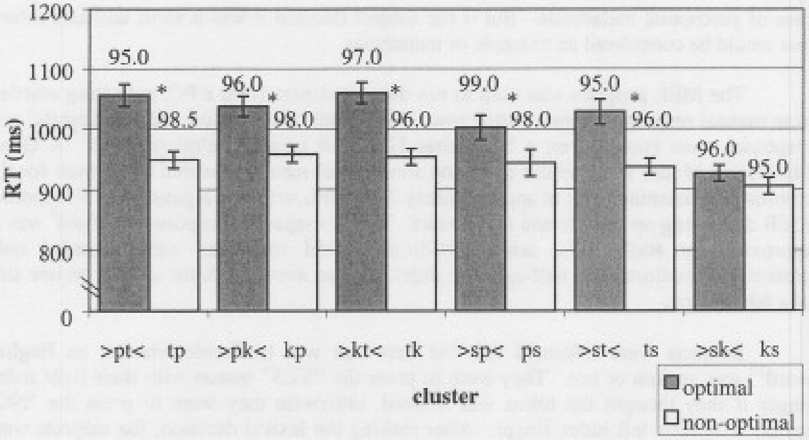


Figure 9. Mean reaction times, standard errors, and percentages correct for the correct rejections of targets in the auditory lexical decision task. Optimal clusters that were significantly slower than their non-optimal pairs are asterisked.

Although the prediction was that the non-optimal clusters would be more confusing, taking a longer time to respond "NO" to, the opposite result was generally found: Optimal clusters have a longer reaction time. This is in keeping with findings by Vitevitch and Luce (1999) in which listeners take longer to reject non-words that are word-like (having a high probability/density of segment sequences, i.e. they are phonetically similar to many words) than non-words that are not word-like. Since the optimal clusters occur in more English words than non-optimal clusters do, they are more word-like, and thus are harder to discount as words.

There is a significant effect of optimality for each pair ($p < .05$) except for >[sk]< and [ks]. Recall that [ks] is the only non-optimal sequence with a higher frequency of occurrence than its optimal pair. This result then could be a cluster frequency effect. However, if that were the case, the prediction would be that [ks] would have a significantly higher RT than >[sk]<. The solution is that there are both optimal perceptual clustering and cluster frequency effects. Since all the other optimal clusters occurred at least as much and usually much more than their non-optimal pairs, their higher frequency of occurrence gave a boost to subjects' performance which was already high based on perceptibility. However, since >[sk]< occurs less frequently than [ks], optimal >[sk]< did not gain this frequency boost.

Error analysis of SRT optimal vs. non-optimal clusters of non-word targets[11] Figure 10 shows the number of metathesis and non-metathesis errors for optimal and non-optimal clusters for listeners in the speech reception threshold condition. Subjects

---

[11] See Appendix D for CLL group errors, and Appendix E for SRT group errors.

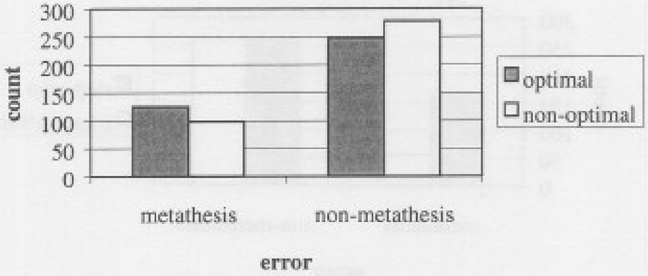


Figure 10. Types of errors for SRT optimal and non-optimal cluster target items. Optimal clusters had significantly more metathesis errors but fewer non-metathesis errors than non-optimal clusters.

made *more* metathesis errors on the non-word targets with optimal clusters (metathesizing them back into English words) than those with non-optimal clusters. Subjects made *fewer* non-metathesis errors on the optimal clusters than the non-optimal ones. This pattern is significant ($p < .05$). The interpretation of why the optimal clusters metathesized is as follows:

> This experiment was attempting to cause listeners to metathesize *to* real words, not *from* real words, as it is attested in language. The results show, in effect, a "metathesis in reverse"—hearing good cues leads the listener back to the underlying form, instead of the underlying form metathesizing to result in good cues that will be preserved. Since the optimal clusters have better cues than non-optimal clusters, there is a higher probability that the listeners heard both obstruents in the optimal clusters correctly. For metathesis to occur, there need to be two obstruents to switch. Because the non-optimal clusters have poorer cues, it is likely that one or both consonants were not heard correctly, and therefore cannot be metathesized—other errors are made instead. If the subjects heard the optimal clusters, and heard enough of the rest of the word to narrow down the word choices, then a temporal change would result in a lexical item. Connine et al (1993) found that changing a few features of a phone can still lead to priming of the base word, so switching features could have similar effects. Since the majority of the real word sources of the targets had zero or one neighbors, if any word was activated during recognition it was more than likely to be one of those.

Error analysis of SRT manner features of non-word targets The number of metathesis and non-metathesis errors for stop-stop and fricative%stop (i.e. fricative-stop and stop-fricative) clusters are shown in Figure 11. Clusters with fricatives and stops were significantly less likely to metathesize than those composed solely of stops ($p < .05$). This supports Steriade's claim that the linear order of adjacent consonants that share manner features is highly non-salient.

Stop-stop clusters and fricative%stop clusters had the same amount of non-metathesis errors, which indicates that fricative%stop clusters are no less salient than

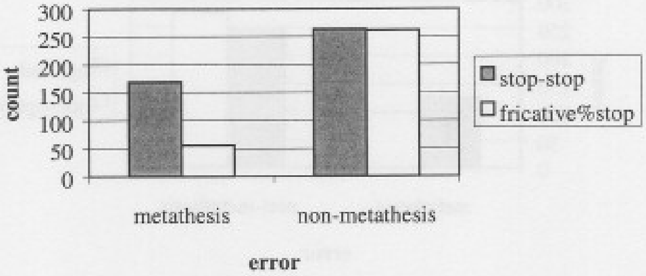


Figure 11. Types of errors for SRT stop-stop and fricative%stop cluster target items. Stop-stop clusters had significantly more metathesis errors than clusters with fricatives, but had roughly the same amount of non-metathesis errors.

stop-stop clusters. However, fricative%stop clusters caused fewer perceptual metathesis errors, indicating that their temporal ordering is more salient than that of stop-stop clusters. Judging from the fact that stops share more manner features than fricatives and stops, the more manner features two consonants share, the fewer cues there are to determine their order.

6. SUMMARY AND CONCLUSION

This study examined acoustic and perceptual cues in obstruent clusters in order to test the hypothesis that metathesis can be a process that maintains identification of the consonants involved. Clusters with poor cues may be susceptible to sound change, but if an obstruent with poor cues can switch to a position that improves its perceptibility, this optimal cluster has a better chance of preservation, as proposed in Hume 1998, 2001 and Steriade 2001. In English, most of the predicted optimal clusters were found to be more prevalent in the lexicon than non-optimal clusters. This could be proof that optimal clusters are more likely to be maintained.

In an auditory lexical decision task, there were effects of both optimality of cues, and frequency of clusters in the lexicon. For the clear listening level group, there was a slow rejection of targets with clusters that occur with high frequency in the lexicon. This usually was in tandem with the slow rejection of optimal clusters, except for >[sk]< and [ks], in which the non-optimal [ks] had a higher lexical frequency. For the speech reception threshold group, targets with optimal clusters were more likely to be perceptually metathesized and realized as the underlying words than targets with non-optimal clusters were because subjects are more likely to hear both consonants in optimal clusters. Clusters with fricatives and stops were less likely to be perceptually metathesized than clusters containing only stops, since the continuity of manner features in a cluster hinders perception of consonant order. Thus good cues indicating the transition between the obstruents in a cluster are important as well as cues into and out of the cluster.

In conclusion, the results of this study suggest that examining the lexicons of languages with metathesis in conjunction with following perceptual principles may provide explanations to some of the patterns observed in language sound systems. Although some of the perceptual findings will need to be adapted for the acoustics of a specific language (such as for languages that do not lenite /t/ postvocalically as in American English), in general, most good perceptual cues are language universal.

REFERENCES

Baayen, R.H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX Lexical Database* (Release 2) [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].

Bailey, C-J. N. (1970). Toward specifying constraints on phonological metathesis. *Linguistic Inquiry*, 1, 347-349.

Blumstein, S. E., and Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *The Journal of the Acoustical Society of America*, 66, 1001-1017.

Bond, Z. S. (1971). Units in speech perception. *Working Papers in Linguistics*, 9, viii-112, Ohio State U.

Bregman, A. S. (1990). *Auditory Scene Analysis*. Cambridge, MA: MIT Press.

Connine, C. M., Blasko, D. G., and Titone, D. (1993). Do the beginnings of spoken words have a special status is auditory word recognition? *Journal of Memory and Language*, 32, 193-210.

Delattre, P. C., Liberman, A. M., and Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *The Journal of the Acoustical Society of America*, 27, 769-773.

Fay, W. H. (1966). Temporal sequence in the perception of speech. *Janua Linguarium*, Series Minor 45.

Frisch, S. A., Large, N. R., and Pisoni, D. B. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language*, 42, 481-496.

Goldinger, S. D. (1996). Auditory lexical decision. *Language and Cognitive Processes*, 11, 559-567.

Hume, E. (1998). The role of perceptibility in consonant/consonant metathesis. *Proceedings of the West Coast Conference on Formal Linguistics*, 17, 293-307.

Hume, E. (2001). Metathesis: Data, Motivation and Phonological Theory. In E. Hume, N. Smith and J. van de Weijer (eds.), *Surface Syllable Structure and Segment Sequencing*. Leiden, NL: HIL.

Hume, E., Johnson, K., Seo, M., and Tserdanelis, G. (1999). A cross-linguistic study of stop place perception. *Proceedings of the XIVth International Congress of Phonetic Sciences*, 2069-2072.

Hock, H. H. (1985). Regular metathesis. *Linguistics*, 23, 529-546.

Kewley-Port, D. (1983). Time-varying features as correlates of place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, 73, 322-335.

Lahiri, A., Gewirth, L., and Blumstein, S. E. (1984). A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross-language study. *The Journal of the Acoustical Society of America*, 76, 391-404.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (ed.), *Cerebral Mechanisms in Behavior: The Hixon Symposium*, 112-146, New York: Wiley.

Luce, P. A. (1986). Neighborhoods of words in the mental lexicon. *Research on Speech Perception* Technical Report No. 6, Indiana U.

Luce, P. A., and Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1-36.

Miller, G. A., and Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America*, 27, 338-352.

Nakao, T. (1986). Metathesis. In D. Kastovsky and A. Szwedek (eds.), *Linguistics across Historical and Geographical Boundaries: In Honour of Jacek Fisiak on the Occasion of His Fiftieth Birthday*, 547-556, New York: Mouton de Gruyter.

Öhman, S. E. G. (1965). Coarticulation in VCV utterances: Spectrographic Measurements. *The Journal of the Acoustical Society of America*, 39 1966 p.151-168.

Pickett, J. M. (1958). Perception of compound consonants. *Language and Speech*, 1(4), 288-304.

Pitt, M. A., and McQueen, J. M. (1998). Is compensation mediated by the lexicon? *Journal of Memory and Language*, 39, 347-370.

Pollack, I., Rubenstein, H., and Decker, L. (1959). Intelligibility of known and unknown message sets. *The Journal of the Acoustical Society of America*, 31, 273-279.

Remez, R. E. (2001). The multimodal nature of speech. In E. Hume and K. Johnson (eds.), *The Role of Speech Perception in Phonology*. New York: Academic Press (in press).

Savin, H. B. (1963). Word-frequency effect and errors in the perception of speech. *The Journal of the Acoustical Society of America*, 35, 200-206.

Seo, M., and Hume, E. (2001). A comparative OT account of metathesis in Faroese and Lithuanian. In E. Hume, N. Smith and J. van de Weijer (eds.), *Surface Syllable Structure and Segment Sequencing*. Leiden, NL: HIL.

Silva, C. M. (1973). Metathesis of obstruent clusters. *Working Papers in Linguistics*, 14, 77-84, Ohio State U.

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*. New York: Academic Press (in press).

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17, 3-45.

Ultan, R. (1978). A typological view of metathesis. In J. H. Greenberg (ed.), *Universals of Human Language*, vol 2., 367-402. Stanford: Stanford UP.

Vitevitch, M. S., and Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40, 374-408.

Wang, M. D., and Bilger, R. C. (1973). Consonant confusion in noise: A study of perceptual features. *The Journal of the Acoustical Society of America*, 54, 1248-1266.

Winters, S. (2001). Putting place in its place: Evaluating place perception in VCCV sequences. In E. Hume, N. Smith and J. van de Weijer (eds.), *Surface Syllable Structure and Segment Sequencing*. Leiden, NL: HIL.

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*. New York: Academic Press (in press).

Yost, W. A. (1994). *Fundamentals of Hearing: An Introduction* (3rd Ed.). New York: AP.

Appendix A. Targets and their source words.

Target	Word	Written and spoken word frequency	Spoken word frequency	Number of neighbors
>pt<				
nɪptɪk	nitpick	0	0	0
hæptɪn	hatpin	8	0	0
fʊptæθ	footpath	50	3	1
aᵘptʊt	output	517	54	1
hiptʌmp	heat pump	0	0	0
ʃɑptʊt	shot-put	0	0	0
fʊptæd	footpad	3	0	0
hɑptɑt	hotpot	4	0	1
swipti	sweet pea	0	0	0
splɪpti	split pea	0	0	0
	Sum	582	57	3
>pk<				
stɑpkaᶦl	stockpile	56	0	1
d͡ʒæpkɑt	jackpot	9	0	1
t͡ʃɪpki	chickpea	6	0	0
kɑpkɪt	cockpit	59	3	0
kræpkɑt	crackpot	10	0	1
stɑpkɑt	stockpot	3	0	0
t͡ʃɛpkɔᶦnt	checkpoint	44	0	0
straᶦpke	strike-pay	0	0	0
sɪpke	sick-pay	12	0	1
t͡ʃɔpkɪt	chalkpit	0	0	0
	Sum	199	3	4
>kt<				
aᵘktʌm	outcome	379	16	0
waᶦktæp	whitecap	6	0	2
suktes	suitcase	334	3	0
naᶦktæp	nightcap	22	0	1
kæktɔl	catcall	6	0	1
aᵘktæst	outcast	39	0	4
fruktek	fruitcake	8	0	0
frektɑr	freight car	0	0	0
striktɑr	streetcar	15	0	0
nʌktes	nutcase	0	0	1
	Sum	809	19	9

Target	Word	Written and spoken word frequency	Spoken word frequency	Number of neighbors
tp				
krɪtpɪk	cryptic	44	0	3
ʌtpaᵘn	uptown	7	0	2
kætpɪv	captive	127	4	3
tɪtpo	tiptoe	104	0	2
pɛtpɔk	pep talk	0	0	1
ʌtpaᶦt	uptight	26	3	3
bætpɪst	Baptist	100	1	1
rɛtpaᶦl	reptile	110	0	0
ʌtpɚnd	upturned	48	0	1
sɛtpɚ	sceptre	25	0	5
	Sum	566	8	16
kp				
nækpɪn	napkin	124	1	0
taᶦkpæst	typecast	1	0	0
wɪkpɔrd	whipcord	6	0	1
ʌkpip	upkeep	30	0	0
kʌkpek	cupcake	0	0	0
tɑkpot	topcoat	9	0	0
slɪkpot͡ʃ	slip-coach	0	0	0
pɑkpɔrn	popcorn	14	0	0
rɪkpɔrd	ripcord	1	0	2
zɪkpod	zip code	0	0	0
	Sum	185	1	3
tk				
vɪtkɚ	victor	178	10	2
ɑtkɪv	octave	35	3	2
spɛtkɚ	spectre	49	0	1
pɛtkɪn	pectin	26	0	1
lætkɪk	lactic	33	0	1
tætkəl	tactile	30	1	0
kɑtkel	cocktail	179	2	0
lɛtkɚn	lectern	40	2	1
trætkɚ	tractor	191	7	1
lætkos	lactose	8	0	0
	Sum	769	25	9

Appendix A. cont.

>sp<				
d͡ʒɪspi	gipsy	96	2	1
ʌspɚd͡ʒ	upsurge	56	3	0
tɪspi	tipsy	18	0	1
ɛspəm	Epsom	8	0	0
tɑspɔᶦl	topsoil	32	0	0
tɑspaᶦd	topside	13	0	0
pɛspɪn	pepsin	0	0	0
stɛspʌn	stepson	6	0	0
næspæk	knapsack	17	0	0
sospʌdz	soapsuds	8	0	0
	Sum	254	5	2

>st<				
pistə	pizza	34	1	0
fʊsti	footsie	1	0	1
fʊstɔr	footsore	4	0	1
aᶷstɛt	outset	100	7	2
kɚsti	curtsy	44	0	3
kæstəp	catsup	8	0	0
nɑsti	Nazi	372	9	1
skɪstɔᶦd	schizoid	7	0	0
d͡ʒɛstəm	jetsam	10	0	0
stestaᶦd	stateside	9	0	1
	Sum	589	17	9

>sk<				
wæski	waxy	24	0	7
tæski	taxi	645	27	7
hæskɔ	hacksaw	5	0	0
pɪski	pixie	1	0	4
dɪski	Dixie	6	0	4
mɪskɚ	mixer	29	1	2
ɛskɪt	exit	253	23	2
tɑskɪk	toxic	106	0	2
flæskən	flaxen	2	0	0
ɛskaᶦz	excise	16	1	3
	Sum	1087	52	31

ps				
d͡ʒæpsɚ	jasper	28	0	0
aᶦpsɪk	icepick	1	0	2
krɪpsi	crispy	5	0	4
æpsən	aspen	15	0	0
tipsun	teaspoon	65	2	1
dɛpsət	despot	21	0	0
prɑpsɚ	prosper	106	5	2
hɑpsɪs	hospice	6	0	0
gripsent	greasepaint	8	0	0
sɛpsul	cesspool	6	1	0
	Sum	261	8	9

ts				
pɑtsə	pasta	36	2	2
itsɚ	Easter	265	19	5
gʌtso	gusto	32	0	2
tetsi	tasty	56	1	5
hetsæk	haystack	21	1	0
kɔtsɪk	caustic	27	0	2
mʌtsɚ	muster	104	0	14
spætsɪk	spastic	11	0	1
d͡ʒɛtsɪŋ	jesting	2	0	1
plætsɚd	plastered	6	0	2
	Sum	560	23	34

ks				
wɪksi	whiskey	623	11	4
haᶷksot	housecoat	21	0	1
hʌksi	husky	45	0	6
pɛksi	pesky	1	0	0
dɪkso	disco	150	2	1
mʌksi	musky	8	0	8
ɛksɔrt	escort	138	1	1
kæksɪt	casket	39	0	1
frɪksi	frisky	11	1	3
vɪksəs	viscous	23	10	1
	Sum	1059	25	26

Appendix B. English word foils.

>pt<	**tp**	**>sp<**	**ps**
riptide	bit part	peace pipe	flip side
styptic	foot-pound	lisping	topsail
striptease	gatepost	crosspiece	dropsy
optic	outpost	waspish	gypsum
sceptic	jetpack	space probe	typeset
uptake	footprint	tailspin	keepsake
raptly	lightproof	misprint	ripsaw
claptrap	hotplate	spoilsport	lapse rate
aptly	waste-pipe	sunspot	upswing
sculptor	dustpan	homespun	campsite

>pk<	**kp**	**>st<**	**ts**
upcast	stickpin	blast-off	jet set
stopcock	neckpiece	mastiff	hot seat
tipcart	backpack	coster	ritzy
slipcase	bookplate	nesting	wet suit
dropkick	spark-plug	taster	pretzel
shopkeep	pickproof	all-star	heartsick
pipeclay	leakproof	shoestring	shirtsleeve
bumpkin	shockproof	brainstorm	pint-sized
pumpkin	inkpad	tombstone	Scotsman
trumpcard	inkpot	limestone	statesman

>kt<	**tk**	**>sk<**	**ks**
backtalk	flatcar	Peace Corps	rock-salt
folktale	oatcake	play-school	hoaxer
ductile	gatecrash	mascot	quicksand
proctor	yacht-club	whiskers	axle
shock troops	shortcake	basket	accent
backtrack	nightclub	icecube	laxly
actress	outcry	bearskin	waxwork
spectral	shift key	dunce cap	locksmith
arctic	test case	briskness	blacksmith
tactful	postcard	task-force	Oxsford

Appendix C. Non-word foils and their source words with no neighbors and zero frequency.

>pt<	
ɛptaᶷnd	egg-bound
brɔpted͡ʒ	broad gauge
smoptɑm	smoke-bomb
neptænd	neckband
lʌptel	lugsail
bɚptɔg	bird dog
straᶦptaᶷnd	strikebound
sɪptɚθ	sick-berth
sɪptɔl	sick call
bɚptid	birdseed

>pk<	
flʌpkaᶦd	flood-tide
rʊpkir	root beer
wepkænd	wave band
tʌpkɑp	tuck-shop
bɛpkost	bedpost
drɪpkaᶦs	drift-ice
drɪpkɪd͡ʒ	driftage
pɑpkaᶷnd	potbound
ropkʊk	road-book
flæpkɑp	flattop

>kt<	
rɪkted͡ʒ	rib cage
wʊstʌlp	wood-pulp
roktɛns	road-sense
diktol	deedpoll
blʌktæŋk	blood bank
ʃɑktɚl	shop-girl
fespe	fete-day
pɑpkaᶷnd	potbound
bɛktɔr	bedsore
hɑktol	hop-pole

tp	
slutpælv	sluice-valve
trʌtpɑrm	truck farm
stɑtpɑr	stockcar
blætpæp	black cap
prɛtpɑks	press-box
pɪtpot	pigboat
fɑtpæŋk	fogbank
brɪtpɪln	brickkiln
gritpʌn	grease-gun
lutpɑks	loosebox

kp	
glokpɪʃ	globefish
braᶦkpek	bridecake
frukpæt	fruit bat
brɛkpɔrd	breadboard
slɑkpol	slop bowl
klɔkpæns	clog-dance
ʃɑkpɛl	shop-bell
stʌkpʊk	stud-book
ʃɑkpɔᶦ	shop-boy
spikpɑp	speed-cop

tk	
tetkɛk	tape deck
aᶦtkit	ice sheet
aᶦtkot	iceboat
haᶷtkɔg	housedog
t͡ʃɔtkɑr	choc-bar
pletkɛt	place-bet
fɑtkaᶷnd	fogbound
rɑtkʌn	rock bun
sætkʌt	sackbut
pɑtkɑp	popshop

Appendix C. cont.

>sp<		**ps**	
blʌspaᶷnt	blood count	kwɑpsɑ	kwacha
kæspɛsk	cash desk	t͡ʃopsæmp	choke-damp
dɛspæks	death tax	ɛpsek	egg-shake
waᶦspaᶦ	wise guy	brɪpsild	brickfield
flæspʌn	flashgun	spopsev	spokeshave
pʌspo	Pashto	siskend͡ʒ	sea change
sispaᶦm	seedtime	bɚptɔg	bird dog
naᶦstɛl	night-bell	stɑpsɪʃ	stockfish
fespe	fete-day	dɪpsɔŋ	diphthong
sæspɔrd	sash-cord	hipsild	heat shield
>st<		**ts**	
blaᶦstəm	blithesome	paᶦtsʊl	pipeful
taᶦstɑrn	tithe-barn	aᶦtsɔl	icefall
pɑstɔᶦ	pot-boy	stɪtsɪft	stick shift
klaᶷstæŋk	cloudbank	aᶦtsild	icefield
skwɑstɑr	squad car	aᶦtso	ice-show
saᶷstɔ	southpaw	witsɔrm	weak form
listʌd	leaf-bud	pʌtsɔg	pug-dog
wʊstʌlp	wood-pulp	dɔtsez	dog-days
trestæp	trade gap	hɑtsild	hop-field
naᶦstɛl	night-bell	rætse	rag-day
>sk<		**ks**	
flaᶦskɛk	flight deck	floksɑrt	flowchart
pruskit	proof sheet	pʊksʊl	pushful
fiskest	fishpaste	slʌksʌnd	slush fund
skɪskæn	skidpan	skaᶦksæk	skyjack
striskɚl	street-girl	slɑksɑp	slop-shop
siskend͡ʒ	sea change	tʌksʊl	tubful
naᶦskɪft	night shift	saᶦkses	side-face
mʌskæk	mudpack	haᶦkser	high chair
piskɪk	peachick	haᶦksʌmp	high jump
pæskɔl	pat-ball	ʃiksɪp	sheepdip

Appendix D. Spoken errors of CLL group.

Error	Count	% of errors	% of targets	
metathesis	24	10.71	1.00	
anticipatory assimilation	107	47.77	4.46	
perseveratory assimilation	7	3.13	0.29	
delete C1	23	10.27	0.96	
delete C2	0	0.00	0.00	
change feature of C1	51	22.77	2.13	
change feature of C2	1	0.45	0.04	
insertion[12]	9	4.02	0.38	
other	2	0.89	0.08	
TOTAL	224		9.33	2400 total targets 20 subjects

CLUSTER	metath	antic	persev	del 1	del 2	ch 1	ch 2	insert	other	TOTAL
>pt<	9	31	1	0	0	6	0	0	0	47
tp	0	8	0	2	0	4	0	0	0	14
>pk<	3	0	1	0	0	3	0	0	0	7
kp	4	12	1	0	0	8	0	1	1	27
>kt<	6	33	0	1	0	5	0	0	1	46
tk	1	23	3	20	0	23	0	0	0	70
>sp<	0	0	0	0	0	0	1	0	0	1
ps	1	0	0	0	0	0	0	3	0	4
>st<	0	0	1	0	0	0	0	3	0	4
ts	0	0	0	0	0	2	0	0	0	2
>sk<	0	0	0	0	0	0	0	1	0	1
ks	0	0	0	0	0	0	0	1	0	1
>TT<	18	64	2	1	0	14	0	0	1	100
TT	5	43	4	22	0	35	0	1	1	111
>sT<	0	0	1	0	0	0	1	4	0	6
Ts	1	0	0	0	0	2	0	4	0	7
optimal	18	64	3	1	0	14	1	4	1	106
non-optimal	6	43	4	22	0	37	0	5	1	118

---

[12] Errors classified as insertions may have other errors besides the insertion.

Appendix E. Spoken errors of SRT group.

Error	Count	% of errors	% of targets
metathesis	225	29.96	18.75
anticipatory assimilation	45	5.99	3.75
perseveratory assimilation	10	1.33	0.83
delete C1	19	2.53	1.58
delete C2	22	2.93	1.83
change feature of C1	101	13.45	8.42
change feature of C2	113	15.05	9.42
insertion	121	16.11	10.08
other	95	12.65	7.92
TOTAL	751		62.58

1200 total targets
10 subjects

CLUSTER	metath	antic	persev	del 1	del 2	ch 1	ch 2	insert	other	TOTAL
>pt<	28	1	2	2	1	4	16	11	10	75
tp	25	11	1	2	1	7	22	1	11	81
>pk<	10	5	2	2	0	10	3	8	2	42
kp	44	6	2	1	1	7	8	8	4	81
>kt<	50	5	1	0	1	9	0	5	11	82
tk	12	14	0	2	1	24	5	9	5	72
>sp<	21	2	1	1	3	2	24	6	15	75
ps	6	1	0	0	1	9	9	15	11	52
>st<	14	0	0	2	8	8	3	14	6	55
ts	9	0	1	3	3	15	12	15	11	69
>sk<	3	0	0	2	1	3	10	20	6	45
ks	3	0	0	2	1	3	1	9	3	22
>TT<	88	11	5	4	2	23	19	24	23	199
TT	81	31	3	5	3	38	35	18	20	234
>sT<	38	2	1	5	12	13	37	40	27	175
Ts	18	1	1	5	5	27	22	39	25	143
optimal	126	13	6	9	14	36	56	64	50	374
non-optimal	99	32	4	10	8	65	57	57	45	377

OSU Working Papers in Linguistics 55, 117-171

Explaining Directional Asymmetry in Turkish [h] Deletion: A Crosslinguistic Study of Perceptibility[1]

Jeff Mielke
mielke@ling.ohio-state.edu

0. Introduction

[h] deletion is a common phenomenon. Some languages have orthographic *h* that is no longer pronounced in certain contexts, and some languages allow /h/ deletion in fast speech. An example of both is English *prohibit/prohibition*. Both words are spelled with *h* but while [h] is always pronounced in the former, it is optional in the latter.

(1)	prohíbət		*proíbət	'prohibit'
	pròhəbíʃən	~	pròəbíʃən	'prohibition'

The difference between the environments where [h] occurs in the two words is that the following vowel is stressed in *prohibit*, and unstressed in *prohibition*. English [h] is generally deleted when it precedes an unstressed vowel, especially in fast speech. This holds true for other pairs such as *inhibit/inhibition* and *vehicular/vehicle*. Thus, it is possible to unify the environments where English [h] is deleted according to whether or not stress is present on the following vowel.

---

[1] This paper builds on previous work by Lena Ovcharova (1999). It has benefited from comments from Elizabeth Hume, Keith Johnson, Donca Steriade, students in the Perception in Phonology seminars, members of the OSU phonetics and phonology discussion group, and audience members at the 2000 Montreal-Ottawa-Toronto Phonology Workshop, the 2000 OSU Colloquium Fest, and the 2000 Mid-Continental Workshop on Phonology. The French and Arabic experiments would not have been possible without the help of Nick Clements and Annie Rialland.

In contrast, Turkish [h] deletion presents an interesting challenge. Turkish [h] is often deleted in fast speech, but only in certain segmental contexts (Lewis 1967, Sezer 1986). While unifying the conditioning environments for English [h] deletion is fairly straightforward, unifying the diverse conditioning environments for Turkish [h] deletion is not. For instance, [h] can be deleted when it is preceded or followed by a fricative. It can also be deleted when it is followed by a sonorant consonant. However, [h] cannot be deleted when it is preceded by a sonorant consonant. The [h] deletion pattern is symmetrical for some contexts (fricatives) and asymmetrical for others (sonorant consonants). Not only is the pattern of [h] deletion asymmetrical, it is asymmetrical in the opposite direction for other contexts. [h] can be deleted when it is preceded by a voiceless stop, but not when it is followed by a voiceless stop. This is the opposite of the pattern of deletion with respect to sonorant consonants. These are just a few of the environments where [h] can be deleted in Turkish, but they are sufficient to show that what unifies the conditioning environments for Turkish [h] deletion is not as transparent as it is for English [h] deletion.

The immediate goal of this paper is to try to understand the seemingly unrelated environments where [h] deletes in Turkish. More generally, this paper is an exploration of the interaction between speech perception and phonology with respect to segment deletion. In the pages that follow, a perceptual account of Turkish [h] deletion is motivated. The findings from this case are then used to try to elucidate the relationship between perception and phonology.

The environments where [h] deletion occurs in Turkish are presented in more detail in section §1. The proposal that Turkish [h] deletion is influenced by perception is introduced in §2, along with the predictions that are made by such a proposal. Perception experiments were performed to test these hypotheses and predictions. The results of an experiment with Turkish-speaking subjects are presented in §3, and the results of a crosslinguistic experiment are presented in §4. §5 deals with issues relevant to aspirated stops, and in §6, the results of an additional experiment are presented to address these issues. The experiment results and their implications for the influence of phonology on perception are discussed in §7. Implications for the influence of perception on phonology are discussed in §8, and a model for predicting sensitivity is proposed, based on the factors found to influence sensitivity in this study. Concluding remarks are in §9.

1. [h] deletion in Turkish

[h] is optionally deleted in fast speech in Turkish, but only in certain segmental contexts. The inventory of Turkish consonants is given in (2), and the environments where [h] is optionally deleted are described in (3-7).

(2) *Turkish consonant phoneme inventory*

	bilabial	labio-dental	alveolar	palato-alveolar	palatal	velar	glottal
stops	p b		t d			k g	
fricatives		f (v)[2]	s z	ʃ ʒ			h
affricates					tʃ dʒ		
nasals	m		n				
liquids			l, r				
glides		v			y		

As mentioned above, [h] is optionally deleted *before* sonorant consonants, but not after them. When [h] is deleted from preconsonantal or final position, compensatory lengthening of the preceding vowel occurs, as in (3a) (Sezer 1986).

(3)	a.	fihrist	~	fiːrist	'index'
		tehlike	~	teːlike	'danger'
		mehmet	~	meːmet	proper name
		köhne	~	köːne	'old'
	b.	merhum		*merum	'the late'
		ilham		*ilam	'inspiration'
		imha		*ima	'destruction'
		tenha		*tena	'deserted'

[h] is optionally deleted *after* voiceless stops but not before them.

(4)	a.	ʃüphe	~	ʃüpe	'suspicion'
		ethem	~	etem	proper name
	b.	kahpe		*kaːpe	'harlot'
		sahte		*saːte	'counterfeit'
		mahkum		*maːkum	'inmate'

[h] is optionally deleted before *and* after voiceless fricatives.

(5)	a.	ishal	~	isal	'diarrhea'
		safha	~	safa	'step'
		meʃhur	~	meʃur	'celebrity'

---

[2] /v/ is realized sometimes as a labiodental fricative and sometimes as a labiodental approximant.

	b.	mahsus	~	ma:sus	'special to'
		tahsil	~	ta:sil	'education'
		ahʃab	~	a:ʃab	'made of brick'

[h] is optionally deleted *after* voiceless affricates, but not before them.

(6)	a.	metʃhul	~	metʃul	'unknown'
	b.	ahtʃi		*a:tʃi	'cook'

[h] is optionally deleted intervocalically, as well as word-finally (perhaps categorically), but not word-initially.

(7)	a.	tohum	~	toum	'seed'
		mühendis	~	müendis	'engineer'
		sahan	~	saan	'copper food dish'
		muhafaza	~	muafaza	'protection'
	b.	timsah	~	timsa:	'crocodile'
	c.	hava		*ava	'air'

The environments where deletion occurs are summarized in (8). There is no evidence that [h] deletes before or after voiced obstruents, in part because it seldom occurs in these environments. In the next section, a perceptual account of Turkish [h] deletion is proposed. See Mielke (to appear a) for a discussion of formal phonological accounts of [h] deletion.

(8) *[h] deletion summary*

Context		Before Context	After Context
voiceless stop	[p, t, k]	no deletion	DELETION
voiceless affricate	[tʃ]	no deletion	DELETION
voiceless fricative	[f, s, ʃ]	DELETION	DELETION
voiced stop	[b, d, g]	no evidence	no evidence
voiced affricate	[dʒ]	no evidence	n/a[3]
voiced fricative	[z, ʒ]	no evidence	no evidence
sonorant consonant	[n, m, l, r]	DELETION	no deletion
intervocalic		DELETION	
word-initial		no deletion	
word-final		DELETION	

---

[3] [h] does not occur after the voiced affricate in Turkish

2. The role of perceptibility in deletion

It has been hypothesized that sounds which are less perceptible are more likely to be altered than more salient sounds (Hura et al. 1992, Kohler 1990, Steriade 2001). The rationale is that the loss of a sound which is already difficult to perceive is not as great as the loss of a more salient sound. The motivation for loss may very well be a non-perceptual factor such as articulatory ease, but the selection of a sound to be deleted may be perceptual (see Hume and Johnson 2001). Perceptibility may be responsible for selecting the environments where [h] deletes in Turkish.

It is reasonable to assume that the demand to minimize articulatory effort is ever-present, and that this demand can be met by segment deletion, but at the expense of the intelligibility of an utterance. Deleting a segment that is very salient will be noticeable to a listener, but deleting a segment that is not very perceptible involves a less significant loss of intelligibility. Ranking the environments in a perceptibility scale from the most salient to the least salient shows the relative cost of [h] deletion by environment.

2.1. Perceptibility scales

In her study of laryngeal neutralization, Steriade (1997) proposes a perceptibility scale for voiced and voiceless consonants, shown in (9), hypothesizing that neutralization of voice contrast will occur in environments where there are fewer cues to voicing. In the environments at the top of the figure, there are many cues to voicing, and these are the environments where voice contrast is most common cross-linguistically. In the environments at the bottom of the figure, there are fewer cues to voicing, and voice contrast in these environments is much more rare.

(9) *Perceptibility scale for laryngeal neutralization (Steriade 1997)*

	Environments	**Examples**
more cues	[+son] __ [+son]	a**b**a vs. a**p**a & a**b**ra vs. a**p**ra
↑	# __ [+son], [-son] __ [+son]	**b**ra vs. **p**ra, **b**a vs. **p**a, as**b**ra vs. as**p**ra, & as**b**a vs. as**p**a
	[+son] __	a**b** vs. a**p**
	[+son] __ [-son]	a**b**sa vs. a**p**sa
	[-son] __ [-son]	as**b**ta vs. as**p**ta
↓	[-son] __ #	as**b** vs. as**p**
fewer cues	# __ [-son]	**b**sa vs. **p**sa

The relevant cues to voicing are closure voicing, closure duration, vowel duration, F0 and F1 value of adjacent vowels, burst duration and amplitude, and VOT value. All of these cues are available for stops between sonorants, the highest category in (9). For the bottom three categories, only closure voicing and closure duration are available to the listener. The availability of non-internal cues (everything but closure voicing and closure duration) depends on context (Steriade 1997).

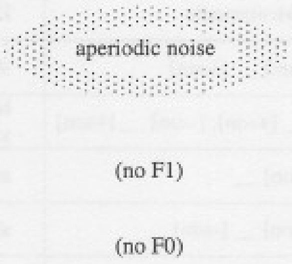
According to Steriade, a language with a voice contrast in a given environment will also have a contrast in other environments which have more cues to voicing. Likewise, a language with voice neutralization in a given environment will also have neutralization in environments with fewer cues to voicing. What unifies the environments with voice neutralization is that they are all perceptually weaker than the environments where contrast is maintained.

Perhaps what unifies the conditioning environments for Turkish [h] deletion is that these are the environments where [h] is the least perceptible, i.e., they are low on the perceptibility scale for [h] environments. The seemingly unrelated environments where [h] can be deleted may be related by being perceptually poor environments for [h].

(10) Hypothesis 1: [h] is less perceptible in environments where it deletes in Turkish than it is in environments where it does not delete.

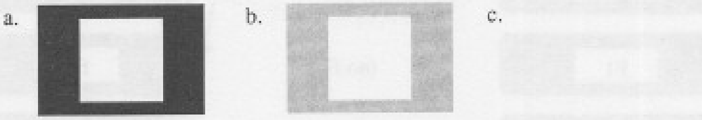
To determine which environments are weak perceptually and which are strong, it is necessary to first look at the cues to the presence of [h] and how neighboring segments can facilitate or detract from their identification. The voiceless glottal fricative has three main segment-internal cues to its presence (see Wright 1996): aperiodic noise in the F2 region, lack of an F1, and lack of an F0.

(11) *The segment-internal cues to [h]*



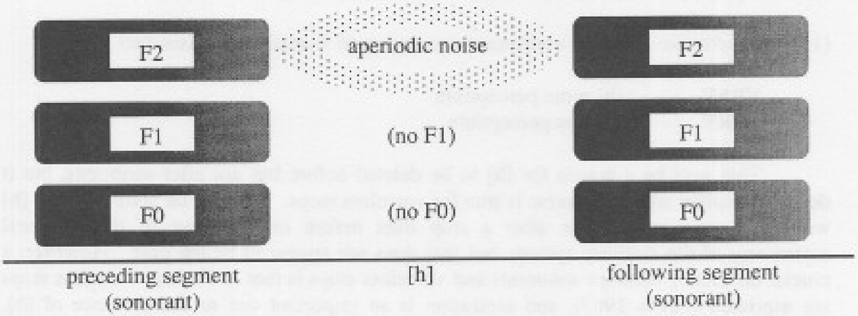
These cues, particularly the two characterized by silence, are very weak. Syntagmatic contrast is important for them to be recognized. A visual metaphor for perceptibility of silence is pictured in (12). There are three white squares, one on a black background (12a) one on a gray background (12b), and one on a white background (12c). Because it contrasts most with its with its surroundings, the square in (12a) is the most visible. The square in (12b) is not as striking because it contrasts less with the gray background, and the square in (12c) is the least salient because it does not contrast at all with its background

(12) *White squares on different backgrounds*



Similarly, [h]'s lack of an F0 is more salient when [h] precedes or follows a voiced segment, which has an F0. Silence does not contrast with silence, but it does contrast with sound. Similarly, [h]'s lack of an F1 is more salient when it is preceded or followed by a sonorant, which has an F1 resonance.

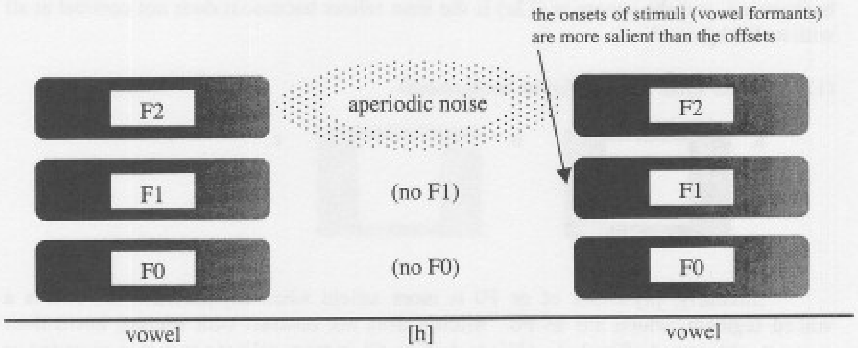
(13) *[h] is most salient between segments that contrast with it*



2.2. Hypotheses with evidence from deletion

The fact that deletion occurs after a sonorant and before a vowel, but not vice versa, can be explained by temporal asymmetries in the auditory system (Bladon 1986, Wright 1996). Auditory nerve fibers exhibit a greater response at the onset of a stimulus signal (such as a vowel) than at the offset. Therefore, all else being equal, consonants before vowels are more perceptible than consonants after vowels, and thus CV transitions provide better cues than VC transitions. This has been shown by Fujimura, Macchi, & Streeter (1978) and Ohala (1992). Fujimura et. al played VCV stimuli with conflicting consonant place cues in the VC and CV transitions to subjects forward and backward, and found that the transitional cues heard by the listeners at the onset of the vowel were more salient, regardless of whether they had been produced as VC or CV transitions.

(14) *CV transitions are more salient than VC transitions*


the onsets of stimuli (vowel formants) are more salient than the offsets

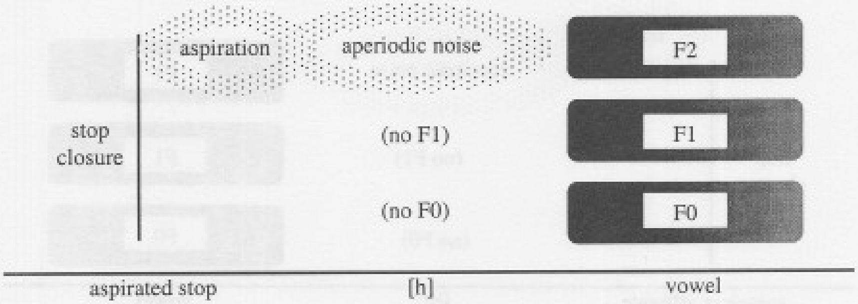
It follows that the contrast between [h] and a following vowel is more salient than the contrast between [h] and a preceding vowel. Therefore, [h] should be more perceptible before a vowel than after one. This could explain why [h] is deleted before a sonorant (and after a vowel) but not after a sonorant (and before a vowel).

(15) *Prediction: sonorant consonant asymmetry (R = sonorant consonant)*

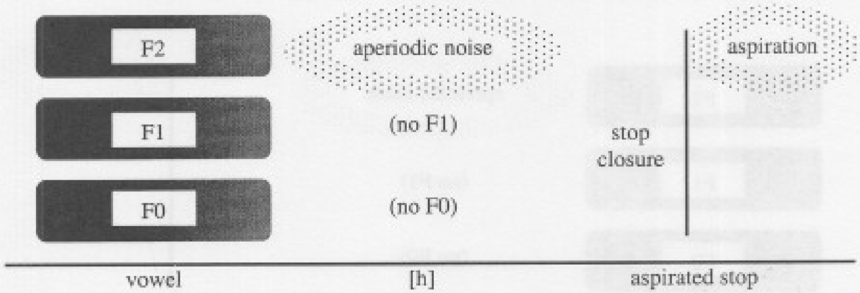
VR**h**V	[h] more perceptible
V**h**RV	[h] less perceptible

This may be a reason for [h] to be deleted before but not after sonorants, but it does not explain why the reverse is true for voiceless stops. It would be predicted that [h] would be more perceptible after a stop than before one because of the temporal asymmetry of the auditory system, but that does not appear to be the case. However, a crucial difference between sonorants and voiceless stops is that in Turkish voiceless stops are aspirated (Lewis 1967), and aspiration is an important cue to the presence of [h]. Hypothetically, if [h] is adjacent to aspiration from another segment as in (16), the aperiodic noise of [h] is less salient than if it is separated from the aspiration by the stop closure, as in (17). In (16) the aperiodic noise of [h] is not separable from the aspiration noise of the voiceless stop – it is essentially an extension of the stop aspiration. In (17) the [h] noise and the stop aspiration noise are separated by the silent stop closure interval.

(16) *[h] is hypothesized to be less salient after voiceless stops*



(17) *[h] is hypothesized to be more salient when the stop follows [h].*



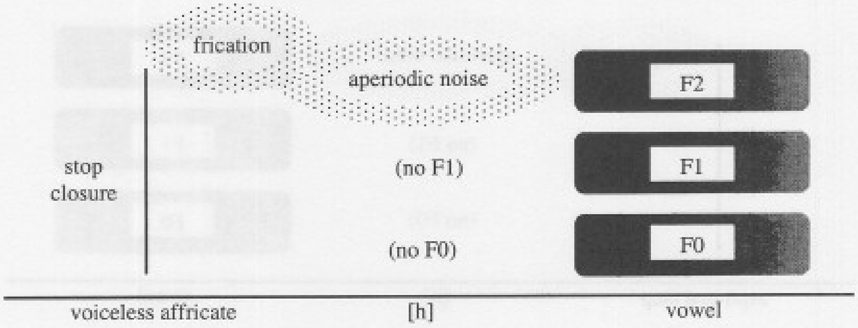
If true, this would help elucidate why [h] is deleted after voiceless stops but not before them. The cues to [h] are more robust before a stop closure than after the aspirated release.

(18) *Prediction: voiceless stop asymmetry (T^h = voiceless stop)*

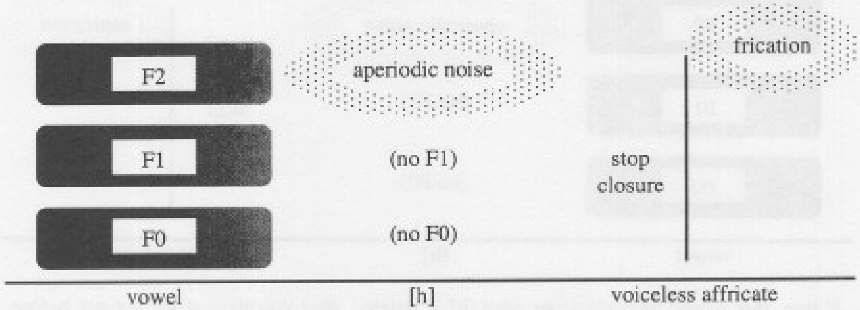
VT^h**h**V [h] less perceptible
V**h**T^hV [h] more perceptible

The fricated release of a voiceless affricate should similarly interfere with [h] perception, so that the cues to [h] are less robust after the fricated release, as in (19), than before the stop closure, as in (20).

(19) *[h] is hypothesized to be less salient after voiceless affricates*



(20) *[h] noise and frication are separated when the affricate follows [h].*



(21) *Prediction: voiceless affricate asymmetry (Tʃ = voiceless affricate)*

VTʃ**h**V	[h] less perceptible
V**h**TʃV	[h] more perceptible

The above figures demonstrate why the directional asymmetry in Turkish [h] deletion patterns is understandable. Unlike voiceless stops, affricates, and sonorant consonants, the deletion pattern for fricatives is symmetrical. Deletion occurs before and after fricatives, and this suggests a general property of fricatives that is detrimental to [h] perception regardless of which side the [h] is on. Specifically, the high-frequency frication noise associated with fricatives is confusable with the high-frequency aspiration noise that is a leading cue to the presence of [h]. Just as [h] is obscured by aspiration when it follows a voiceless stop, [h] is obscured by frication noise when it follows or precedes a voiceless fricative.

Fricatives, stops, and affricates all feature noise (frication or affrication) at the right periphery that can obscure the presence of a following [h]. Compared to [h] after a sonorant consonant, [h] should be less salient after fricatives, aspirated stops, and affricates.

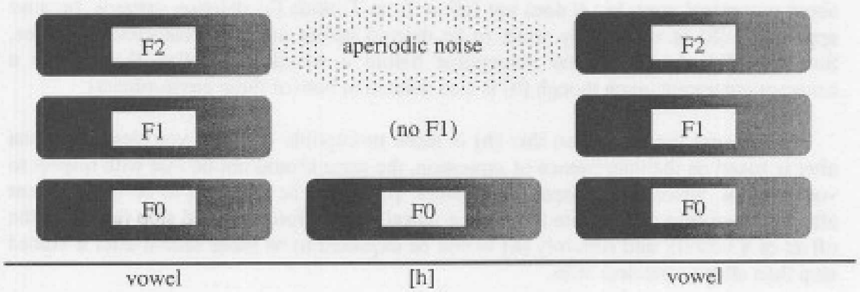
(22) *Prediction: noise affects salience (F = voiceless fricative)*

VR**h**V	[h] more perceptible
VF**h**V, VT^h**h**V, VTʃ**h**V	[h] less perceptible

[h] should also be more perceptible word-initially than word-finally, because the onset of the vowel following initial [h] is more salient than the offset of the vowel preceding final [h]. This is also consistent with the patterns of deletion.

Initial and intervocalic [h] are both followed by vowels, and so both benefit from CV transitions. However, [h] is voiced intervocalically in Turkish (Ovcharova 1999), and this should render it less distinct from a following vowel than when it is not voiced. Voicelessness is one of the cues to [h], and when it is lost, an important part of syntagmatic contrast is lost as well. Therefore, [h] would be expected to be less perceptible intervocalically than it is initially. This is consistent with data on deletion.

(23) *Because intervocalic [h] is voiced, it is hypothesized to be less salient than initial [h].*



This final example differs from the previous examples in that the difference in perceptibility between initial and intervocalic [h] is the result of an allophonic change in [h] that is conditioned by the environment, rather than a difference in perceptibility directly attributable to the environment alone. In addition to the intervocalic environment, [h] is voiced between a vowel and a glide in Turkish. For this reason, glides are not included in the predictions referring to sonorant consonants above and elsewhere in this paper, but would rather be expected to behave more similarly to vowels with respect to [h] perceptibility.

2.3. Hypotheses without evidence from deletion

To this point two kinds of reasons for claiming low perceptibility have been presented. One follows from the hypothesis that [h] is deleted in environments of low perceptual salience (Hypothesis 1). If [h] deletes in a certain environment and the hypothesis is true, then [h] must not be very perceptible in that environment. The second reason draws on what is known about acoustic cues. If an acoustic feature of an environment is likely to interfere with a cue that is important for perception of [h], then [h] must not be very perceptible in that environment. All of the above claims are based on both of these reasons for positing low [h] perceptibility.

The second reason for claiming low perceptibility makes a prediction, stated in (24). If the environments where [h] deletes are truly perceptually weak environments because of the perceptual cues being assumed, then it should be possible to predict other environments where [h] should be more or less salient based on these cues, even if there is no evidence from deletion.

(24) Hypothesis 2: [h] is less perceptible in environments where there are poorer cues to its presence.

One of the predictions this makes is that [h] should be less perceptible before voiceless fricatives than after them, for the same reason [h] is hypothesized to be less perceptible before sonorants than after them. This follows from the assumptions made about perceptual cues, but it does not follow from Turkish [h] deletion patterns, because apparently, [h] is sufficiently weak to be deleted before and after voiceless fricatives. Similarly, [h] should be less perceptible before a voiceless fricative than before a sonorant consonant, even though [h] is also deleted in both of these environments.

Because the prediction that [h] is more perceptible before a voiceless stop than after is based on the interference of aspiration, the same should not be true with respect to voiced stops, which are unaspirated. Rather, [h] would be expected to be more salient after a voiced stop (and before the onset a vowel) than before a voiced stop (and after the offset of a vowel), and certainly [h] would be expected to be more salient after a voiced stop than after a voiceless stop.

In addition to the confusability of [h] with its absence in the same environment, (e.g. Vt^hhV vs. Vt^hV), it is conceivable for [h] to be confusable with its absence in a different environment, i.e., a voiceless aspirated stop is confusable with a voiced stop followed by [h] (e.g. VdhV vs. Vt^hV). Aspiration is an important cue distinguishing voiced and voiceless stops in languages in which voiceless stops are aspirated. So a sequence of an unaspirated stop and [h] could be interpreted as an aspirated stop, and vice versa. This possibility is discussed in §5 and §6.

2.4. Summary

The predictions made up to this point about the relative salience of [h] are summarized in (25), along with the evidence and rationale for each one.

(25) *Summary of predictions:*

	Prediction	Deletion Evidence	Phonetic Rationale
a.	[h] / son __ V > [h] / V __ son	merhum, *merum fihrist ~ fi:rist	All else being equal, prevocalic consonants are more salient than postvocalic consonants.
b.	[h] / V __ vls stop > [h] / vls stop __ V	kahpe, *ka:pe ʃüphe ~ ʃüpe	Aspiration of voiceless stops and affricates interferes with [h] perception.
c.	[h] / V __ vls aff > [h] / vls aff __ V	ahtʃi, *a:tʃi metʃhul – metʃul	The fricated release of affricates interferes with [h] perception.
d.	[h] / son __ V > [h] / vls stop __ V	merhum, *merum ʃüphe ~ ʃüpe	Frication and aspiration both interfere with the perception of a following [h]. Sonorants have neither frication nor aspiration, and so do not interfere with [h] perception this way.
e.	[h] / son __ V > [h] / vls aff__ V	merhum, *merum metʃhul ~ metʃul	
f.	[h] / son __ V > [h] / vls fric __ V	merhum, *merum mahsus ~ ma:sus	
g.	[h] / # __ V > [h] / V __ V	hava, *ava tohum ~ toum	Prevocalic consonants are more perceptible than postvocalic consonants, and lenition of intervocalic [h] may render it less perceptible than word-initial [h].
h.	[h] / # __ V > [h] / V __ #	hava, *ava timsah ~ timsa:	
i.	[h] / vcd stop __V > [h] / vls stop __ V	n/a	[h] is predicted to be less perceptible after voiceless stops because of aspiration. Voiced stops, lacking aspiration, are not expected to show this effect.
j.	[h] / V __ son > [h] / V __ vls fric	n/a	Frication should interfere with [h] perception before voiceless fricatives.
k.	[h] / vls fric __ V > [h] / V __ vls fric	n/a	Prevocalic consonants are more perceptible than postvocalic consonants

> = "is more salient than"

In (26), the phonetic environments are listed according to how many of the cues in (11) they allow. Listing the environments this way gives a rough approximation of relative salience of [h]. Weighting the cues would allow a more accurate approximation (see §8). The three main cues to the presence of [h] are its lack of F0 resonance (except when it is voiced), lack of F1 resonance, and aperiodic noise in the F2 region. For each cue, there are two points in time at which they can contrast with neighboring sounds.

(26) *Phonetic environments listed by the number of cues to [h]:*

Environment	Cues					
	F0		F1		noise	
	offset	onset	offset	onset	onset	offset
sonorant __ V	X	X	X	X	X	X
V __ sonorant	X	X	X	X	X	X
vcd stop __ V	X	X		X	X	X
V __ vcd aff	X	X	X		X	X
V __ vcd stop	X	X	X		X	X
# __ V		X		X	X	X
V __ V			X	X	X	X
glide __ V			X	X	X	X
vcd aff __ V	X	X		X		X
vcd fric __ V	X	X		X		X
V __ glide			X	X	X	X
V __ vcd fric	X	X	X		X	
V __ vls aff	X		X		X	X
V __ vls stop	X		X		X	X
V __ #	X		X		X	X
V __ vls fric	X		X		X	
vls aff __ V		X		X		X
vls fric __ V		X		X		X
vls stop __ V		X		X		X

[h]'s voicelessness is apparent at the offset of a preceding voiced segment, and again at the onset of a following voiced segment. When [h] is voiceless and preceded and followed by voiced segments, both the F0 offset and the F0 onset are cues to the presence of [h] ("offset" and "onset" refer to the stopping and starting of F0, not [h]). For example, both the F0 offset and the F0 onset are available cues when [h] is preceded by a liquid and followed by a vowel. But when [h] appears between a voiceless segment and a vowel or between a word boundary and a vowel, only the F0 offset or F0 onset is an available cue. If the vowel precedes [h], only F0 offset is a cue. If the vowel follows [h], only F0 onset is a cue. If [h] appears intervocalically or between a vowel and a glide, [h] is voiced, and in that case neither F0 offset nor F0 onset is a cue to the presence of [h], because there is no interruption of voicing.

F1 offset is a cue to [h]'s presence when [h], which has no F1 resonance, follows a sonorant, which does have an F1 resonance. When [h] precedes a sonorant, F1 onset is a cue to the presence of [h]. When [h] is between sonorants, both F1 offset and F1 onset are cues to [h]'s presence. When [h] follows an obstruent, F1 offset is not a cue to [h] because there is no F1 resonance that stops when [h] starts. Similarly, when [h] precedes an obstruent, F1 onset is not a cue to [h].

While F0 and F1 resonances are properties of surrounding segments which may facilitate the perception of [h], aperiodic noise is a property of [h] that facilitates its own perception. This means that the onset and offset occur in the opposite order, compared to F0 and F1. The noise onset can occur at the beginning of [h] and the noise offset can occur at the end of [h]. Both noise onset and noise offset are cues to [h] when [h] is between vowels or between any other segments which lack high-frequency noise. When [h] is preceded by a fricative, only noise offset is a cue to [h]. Noise does not begin at the beginning of [h], because the preceding segment is already noisy. When [h] is followed by a fricative, only noise onset is a cue to the presence of [h]. Affricates and aspirated stops end with noise, so noise onset is not an available cue when [h] is preceded by an affricate or an aspirated stop. When [h] is followed by an aspirated stop, noise offset is still an available cue because noise stops at the end of [h] and does not start again until the release of the stop closure.

If the conditioning environments for Turkish [h] deletion truly are the environments where [h] is less salient, a perception experiment should show that the predictions about relative salience are correct.

3. Experiment 1: Turkish listeners

A perception experiment was designed in order to test the relative salience of [h] in various phonetic environments. In the experiment, subjects listened to nonsense words one at a time. Half of the stimuli contained [h] and half did not. Subjects were shown each word in Turkish orthography (minus any "h"s) on a computer screen and responded by clicking a mouse on where in the word they heard an [h] sound, if they heard one at all.

This experiment is similar to the experiment conducted by Ovcharova (1999). One crucial difference is that in Ovcharova's experiment, subjects indicated whether or not they heard an [h] in each word, but did not indicate where it was in the word. One advantage of this approach was that it was not necessary to show the subjects the word, and thus the subjects were not given any extra information (partial transcription) that they would not normally have. Because subjects were not provided with any of the segments in the stimuli, the possibility of confusing voiced stops and [h] with aspirated stops was still present. When the subject is provided with the other segments in the stimuli, this confusion is not possible. This concern is addressed below in §5 and §6.

A significant drawback to not giving a partial transcription (as in Ovcharova 1999), however, was that the approach made it impossible to study some of the types of errors made by subjects. Because there were two possible conditions ("[h]" or "no [h]") and only two possible responses, "yes" or "no", there were four possible scenarios. First, if the stimulus did not contain [h], and the subject answered "no", the subject was correct in not hearing [h]. Second, if the stimulus did contain [h] and the subject answered "no", it was clear that the subject failed to hear [h] in the environment where it occurred. Third, if the stimulus did not contain [h] and the subject answered "yes", the subject had incorrectly heard an [h], but it was impossible to tell where the false alarm occurred, e.g., whether the subject incorrectly heard an [h] before or after a consonant or at the beginning or end of the word, and this information is crucial for determining which environment is more confusing for the identification of [h]. Fourth, if the stimulus did contain [h] and the subject answered "yes", this was interpreted as a correct identification. However, it is not necessarily the case that the [h] the subject heard was in the right place. It is possible for a subject to fail to hear the actual [h] but believe there was an [h] somewhere else, and this should not be counted as a correct identification.

In the present study, by forcing subjects to indicate where they heard the [h], it is possible to determine in cases of the third type where the false alarms occurred and in cases of the fourth type whether the subjects were correctly identifying [h] or hearing it where it was not.

3.1. Methods

3.1.1. Subjects

Stimuli were produced by a male native speaker of Turkish. Six female and 15 male native speakers of Turkish, in Columbus, Ohio, aged 19-33, participated in the experiment as subjects. The results of one German-Turkish bilingual were not included in calculations, and the results of another subject were omitted because of experimenter error.

3.1.2. Procedures

160 target nonwords containing [h], 80 foil nonwords not containing [h], and 80 nontarget nonwords not containing [h] were recorded 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.

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. This was to simulate compensatory lengthening that occurs in Turkish when [h] is deleted from preconsonantal or word-final position. In Turkish orthography this is indicated by a "ğ" following the vowel. This character was not included in the on-screen transcription because transcribing it would indicate that vowel length is not attributable to [h] deletion.[4]

---

[4] This should result in an increase in false alarms for postvocalic environments, as compared to a similar experiment with "ğ" in the transcription.

(27) *Stimuli in consonant environments*

Context		Target Stimuli		Foil Stimuli
		Before	After	
voiceless stop	[p, t, k]	8	8	8
voiceless affricate	[tʃ]	8	8	8
voiceless fricative	[f, s, ʃ]	8	8	8
voiced stop	[b, d, g]	8	8	8
voiced affricate	[dʒ]	8	8	8
voiced fricative	[v, z, ʒ]	8	8	8
nasal	[n, m]	8	8	8
liquid	[l, r]	8	8	8
glide	[j]	4	4	4
TOTAL		68	68	68

(28) *Stimuli in vowel environments*

Context	Target Stimuli	Foil Stimuli
intervocalic	8	4
word-initial	8	4
word-final	8	4
TOTAL	24	12

(29) *Total stimuli*

Total Target Stimuli (with [h])	Total Foil Stimuli (without [h])	Nontarget Stimuli (VCCVs without [h])
160	80	80

The stimuli were randomized and played to subjects over Sennheiser HD 420 headphones in a sound booth. As subjects heard each nonword they were presented with the segments in the word other than [h] on a computer screen and instructed to click on the point in the nonword where they heard [h] or to click on button representing no [h] if they heard no [h] in the word. An "h" appeared on screen at the point in the word where the subject clicked. See Appendix E for a sample screen view.

3.1.3. Data analysis

Sensitivity (d') (Green & Swets 1966, Winer 1971, MacMillan & Creelman 1991) was computed for each subject for each of the 21 environments. The d's for each environment were averaged. d' is a measure of sensitivity based on correct identification and false alarm rates. A d' of zero indicates that correct identification and false alarm rates were the same, that subjects had no sensitivity to the presence or absence of [h]. A positive d' indicates that subjects reported hearing [h] more often when it was present than when it was not. A very high d', such as 3, indicates a very high correct identification rate and a very low false alarm rate.

3.2. Results and discussion

The average sensitivity for each environment is given in (30). Sensitivity varied according to what type of segment was adjacent to the [h] (rows), and according to whether the [h] was before or after the segment (columns). The lowest measured sensitivity was in the word-final environment, and much higher sensitivity was measured in various preconsonantal and postconsonantal environments, as well as word-initially and intervocalically.

(30) *Sensitivity (d') by environment for Turkish subjects*

Context		Before Context (VhX)	After Context (XhV)
voiceless stop	[p, t, k]	2.583	2.233
voiceless affricate	[tʃ]	2.558	2.274
voiceless fricative	[f, s, ʃ]	2.423	2.144
voiced stop	[b, d, g]	2.861	2.707
voiced affricate	[dʒ]	2.769	2.838
voiced fricative	[v, z, ʒ]	2.841	2.426
nasal	[n, m]	2.838	2.964
liquid	[l, r]	2.841	3.028
glide	[j]	2.155	1.777
intervocalic		2.248	
word-initial		2.376	
word-final		0.734	

A repeated-measures analysis of variance (ANOVA) with [h] location (the 21 locations in (30)) as an independent variable showed a main effect for the location of [h] within the stimuli (df = 1,18; F = 19.828, $p < 0.001$).

The differences in salience are consistent with the hypothesis for most environments. In (31), the results of this experiment are given alongside the predictions about salience that were made in the previous section. Glides have not been included here with liquids and nasals because they pattern with vowels in terms of intervocalic voicing. Although nasals and liquids were not predicted to differ in their influence on [h] perceptibility, multiple p values are given for nasals and liquids, respectively, when the two p values are different.

(31) *Sensitivity (d') in terms of predictions*

	Prediction			Result			
a.	[h] / son __ V	>	[h] / V __ son	2.996	>	2.840	p = .285, .058
b.	[h] / V __ vls stop	>	[h] / vls stop __ V	2.583	>	2.233	p = .053
c.	[h] / V __ vls aff	>	[h] / vls aff __ V	2.558	?	2.274	p = .282
d.	[h] / son __ V	>	[h] / vls stop __ V	2.996	>	2.233	p < .001[5]
e.	[h] / son __ V	>	[h] / vls aff__ V	2.996	>	2.274	p < .001
f.	[h] / son __ V	>	[h] / vls fric __ V	2.996	>	2.144	p < .001
g.	[h] / # __ V	>	[h] / V __ V	2.376	?	2.248	p = .548
h.	[h] / # __ V	>	[h] / V __ #	2.376	>	0.734	p < .001
i.	[h] / vcd stop__V	>	[h] / vls stop __ V	2.707	>	2.233	p = .005
j.	[h] / V __ son	>	[h] / V __ vls fric	2.840	>	2.423	p = .001, .002
k.	[h] / vls fric __ V	>	[h] / V __ vls fric	2.144	?	2.423	p = .185

Prevocalic [h] is more perceptible than preconsonantal or prepausal [h]. In the case of [h]s which were adjacent to sonorant consonants (31a), sensitivity is marginally higher when [h] follows the consonant (i.e., is prevocalic) than when [h] precedes the consonant, thought not significantly for nasals, but nearly significant for liquids (p = .058). This is consistent with the prediction that [h] is more salient before a vowel, due to the heightened auditory response.

As shown in (31b), [h] is also more perceptible before voiceless stops than after them (p = .053). This is consistent with the hypothesis that the aspiration noise involved in these sounds interferes with the perception of a following [h] enough to overcome the prevocalic/postvocalic asymmetry found with [h] before and after sonorants. [h] is not significantly more perceptible before voiceless affricates than after them (p = .282), as shown in (31c).

[h] is significantly more perceptible after sonorant consonants than after voiceless stops (31d), affricates (31e), or fricatives (31f) (p < .001 in all three cases). This is consistent with the hypothesis that [h] contrasts with F0- and F1-bearing sonorants more than with voiceless obstruents which lack both.

---

[5] In (31d-f), p < .001 for both nasals and liquids.

Sensitivity to word-initial [h] is not significantly higher (p = .548) than for intervocalic [h]) (31g). This is not inconsistent with the prediction that because intervocalic [h] is voiced, syntagmatic contrast with vowels is reduced, as compared with unvoiced initial [h]. Sensitivity to word-final [h] is far lower (p < .001) than initial [h], consistent with the temporal asymmetry also borne out in the results for sonorant consonants (31h).

[h] is significantly more perceptible (p = .005) after voiced stops than before them (31i), consistent with the prediction that without aspiration, perceptibility of [h] before and after stops should match the perceptibility of [h] before and after other unaspirated consonants.

[h] is significantly more perceptible (p = .007) before sonorants than before voiceless fricatives (31j). This is consistent with the hypothesis that [h] contrasts with sonorants more than with voiceless fricatives.

The one area where the results are inconsistent with predictions is [h] before and after voiceless fricatives (31k). The context after a voiceless fricative was predicted to be a more salient environment due to the fact that [h] is prevocalic when it follows a fricative, but the opposite pattern emerges, though it is not statistically significant (p = .185).

3.3. Summary

With some exceptions, these results show a relationship between perception and phonology. However, the nature of this relationship is not clear in these results. Perception and phonology could be related because perception influences phonology, i.e., processes such as deletion occur according to universal patterns of perception. Alternatively, perception and phonology could be related because phonology influences perception, i.e., a process such as deletion influences the way speakers perceive sounds. Or perception and phonology could be related in both ways. The two possibilities are not mutually exclusive.

It is impossible to tell which of these is occurring without looking at more than one language. Each relationship makes predictions which can be tested in a cross-linguistic perception experiment or a cross-linguistic survey of deletion phenomena. A perception experiment on one language can show correlation between perception and phonology, but a cross-linguistic experiment is necessary to show causation.

If perception influences phonology, then languages with [h] deletion should delete [h] in environments which are perceptually weak universally, not just perceptually weak for languages with deletion. If speakers of languages which lack [h] deletion have more difficulty perceiving [h] in environments where it is frequently deleted in other languages than in environments where it is seldom deleted, this can be interpreted as evidence that the conditioning environments for deletion are the product of phonetic universals. If the relationship were strictly one in which phonology influences perception, speakers of languages without deletion would not show a difference.

Also, if perception influences phonology, languages that delete [h] where it is perceptually salient would be expected also to delete [h] in environments where it is less salient, and conversely, languages that preserve [h] where it is not very salient would be expected also to maintain [h] in environments where it is more salient. Testing these predictions requires a cross-linguistic typological survey which is beyond the scope of this paper (see Mielke, to appear b).

4. Experiment 2: Crosslinguistic [h] perception

If it is true that phonology influences perception, and listeners become more sensitive to contrasts based on their native phonology, then speakers should be more sensitive to the presence or absence of [h] in environments where it is phonologically significant (i.e., contrastive or at least present) in their own language, as compared to a language without a contrast. Whether or not speakers are good at perceiving [h] would depend on whether or not [h] is present in the language, on what environments it is allowed in, and on whether or not it is contrastive in those environments.

Additional languages were selected according to the distribution of [h] in each language, so that a variety of distributions would be represented among the listeners in this experiment. An ideal set of languages would include a language that allows [h] in many environments, a language that allows [h] only in certain environments, and a language that has no [h] sound at all.

Arabic, which allows [h] before and after nearly all consonants, was selected as a language with [h] in many environments. English, which has [h] in all of the prevocalic positions in the study, was selected as a language with [h] in fewer environments than in Turkish or Arabic. French was selected as a language with no [h] sound. If perception of [h] is influenced by the phonology, then speakers of these three languages should perform differently in the perception experiment, being less able to perceive [h] in environments that are unfamiliar.

The distribution of [h] in the four languages of this study is shown in (32). See Appendix A for lists of words in these languages with [h] in these environments.

(32) *Distribution of [h] in the four languages of this study*

Context	Turkish	Arabic	English	French
vls stop __ vowel	YES	YES	YES	--
vls affricate __ vowel	YES	--	YES	--
vls fricative __ vowel	YES	YES	YES	--
vcd stop __ vowel	YES	YES	YES	--
vcd affricate __ vowel	--	--	YES	--
vcd fricative __ vowel	YES	YES	YES	--
sonorant __ vowel	YES	YES	YES	--
glide __ vowel	YES	YES	--	--
# __ vowel	YES	YES	YES	--
vowel __ vowel	YES	YES	YES	--
vowel __ vls stop	YES	YES	--	--
vowel __ vls affricate	YES	--	--	--
vowel __ vls fricative	YES	YES	--	--
vowel __ vcd stop	YES	YES	--	--
vowel __ vcd affricate	YES	--	--	--
vowel __ vcd fricative	YES	YES	--	--
vowel __ sonorant	YES	YES	--	--
vowel __ glide	YES	YES	--	--
vowel __ #	YES	YES	--	--

Sources: Harrell (1966), Kornrumpf (1979), Oflazer (1994), M. Alaoui (p.c.)

To determine whether the above results are universal or specific to Turkish and to tease apart the influence of perception from the influence of phonology, the perception experiment was repeated with subjects from the three additional languages.

4.1. Methods

4.1.1. Subjects

The English speaking subjects consisted of 17 female and ten male Ohio State University undergraduates, all native speakers of American English. The results of a Farsi-English bilingual subject were not included in calculations. The results of five other subjects were also not included because the subjects misunderstood the instructions and exhibited "spelling behavior", i.e., they indicated where words would be spelled with "h" in English rather than where they heard [h]. For example, these subjects placed "h" after "a" whenever they heard a long [a], even if there was no [h]. The French speaking subjects consisted of one male and twenty-four female native speakers of French in Paris, France, aged 18-28. The results of a German-French bilingual and a Polish-French bilingual were excluded, as well as the results of two others who misunderstood the instructions and exhibited English "spelling behavior". The Arabic speaking subjects consisted of two female and ten male native speakers of Arabic in Paris, France,[6] aged 20-36. Of the twelve, seven were from Morocco, three were from Algeria, one was from Mauritania, and one was from Jordan. The varieties of Arabic represented in the study maintain [h] in the contexts given in (32) (Zawadowski 1978).

4.1.2. Procedures

Procedures for English, French, and Arabic subjects were identical to procedures for Turkish subjects, except that French and Arabic subjects received instructions in French rather than English. Stimuli and other procedures were unchanged.

4.1.3. Data analysis

Sensitivity (d') was again computed for each subject for each of the 21 environments. The d's for each environment were averaged.

4.2. Results and discussion

The results from Turkish listeners were included with the results from the crosslinguistic experiment. A repeated-measures analysis of variance (ANOVA) with [h] location and language as independent variables showed main effects for language and for the location of [h] within the stimuli, and a significant interaction between language and location.

(33) *ANOVA results*

Source of Variance	**DF**	**F**	**P**
Between listeners			
Language	1,69	60.253	<0.001
Within listeners			
Location	1,69	41.855	<0.001
Location * Language	1,69	5.168	<0.001

The correlation of the results for the four languages was computed based on the entire set of d' values. An R square value close to one indicates a high degree of correlation between two languages, and an R square value close to zero indicates very little correlation between two languages.

---

[6] Arabic/French bilingualism is not viewed as a problem for the Arabic subjects, because French has no /h/ sound, and a speaker's language background with respect to /h/ should be the same as for a monolingual Arabic speaker (but very different from a monolingual French speaker).

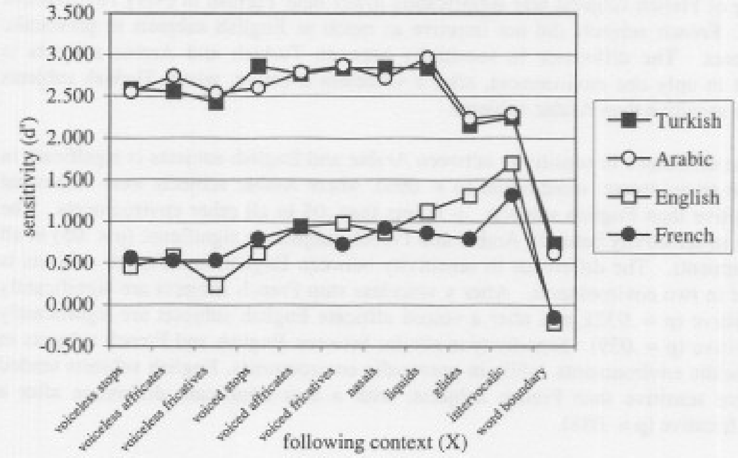
(34) *Correlation (R square)*

Languages	R square	DF	F	P
Turkish & Arabic	0.857	1,19	113.545	<0.001
Turkish & English	0.236	1,19	5.868	0.026
Turkish & French	0.301	1,19	8.180	0.010
Arabic & English	0.296	1,19	7.995	0.011
Arabic & French	0.405	1,19	12.936	0.002
English & French	0.733	1,19	52.288	<0.001

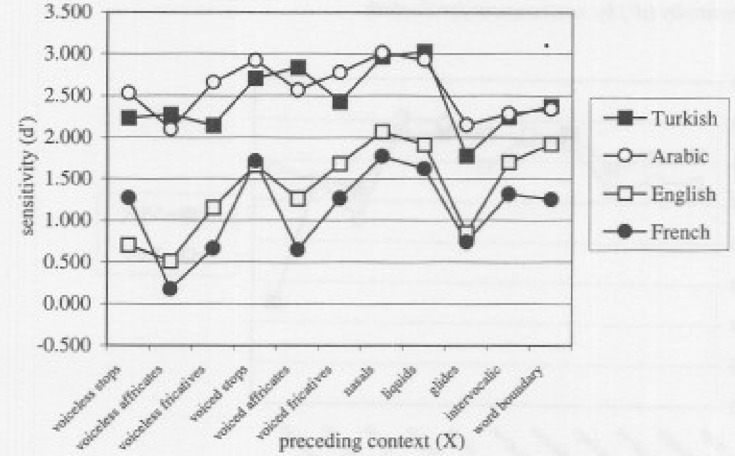
The results for Arabic are strongly correlated with the results for Turkish. Both groups of subjects showed very high sensitivity, as compared with English and French, which are also strongly correlated with each other. In fact there is no environment in which English or French subjects had higher sensitivity than either Turkish or Arabic subjects. This grouping of Arabic with Turkish in terms of sensitivity coincides with the grouping of Arabic and Turkish as languages that permit [h] in many environments, particularly preconsonantal environments. English and French have lower sensitivity, and similarly, both languages permit [h] in fewer environments than Turkish and Arabic.

Charts (35) and (36) show the results separately as "VhX" (before various contexts, postvocalic) environments in (35) and "XhV" (after various contexts, prevocalic) environments in (36). Displaying the results in this manner allows for comparison of all four languages on the same environments. The data used for these charts is located in Appendix C. Although nasals and liquids are grouped together elsewhere in this paper because they behave identically in conditioning Turkish [h] deletion and [h] perceptibility is expected to be very similar with nasals as with liquids, the results for nasals and liquids were calculated separately in order to test this prediction, and are presented separately in this section.

(35) *Sensitivity (d') to [h] before context (VhX)*



(36) *Sensitivity (d') to [h] after context (XhV)*

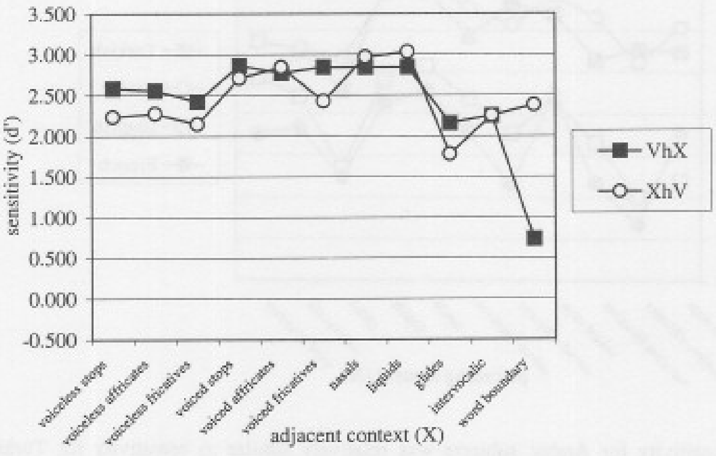


Sensitivity for Arabic subjects was relatively similar to sensitivity for Turkish subjects, while English and French subjects showed lower perceptibility than Turkish subjects in every environment. The difference between English and Turkish is significant (p < .05) in every environment. English subjects performed more similarly to Turkish subjects in prevocalic environments, especially in intervocalic and initial environments. Sensitivity of French subjects was significantly lower than Turkish in every environment (p < .05). French subjects did not improve as much as English subjects in prevocalic environments. The difference in sensitivity between Turkish and Arabic subjects is significant in only one environment, after a voiceless fricative, where Turkish subjects were more sensitive than Arabic subjects.

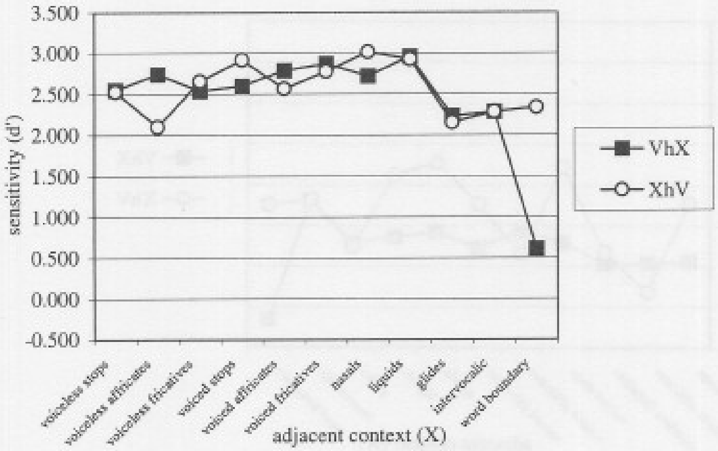
The difference in sensitivity between Arabic and English subjects is significant in all but one environment, word-initial (p = .096), where Arabic subjects were somewhat more sensitive than English subjects. p is less than .05 in all other environments. The difference in sensitivity between Arabic and French subjects is significant (p < .05) in all 21 environments. The difference in sensitivity between English and French subjects is significant in two environments. After a voiceless stop French subjects are significantly more sensitive (p = .032), and after a voiced affricate English subjects are significantly more sensitive (p = .039). Sensitivity is similar between English and French subjects in non-prevocalic environments, while in prevocalic environments, English subjects tended to be more sensitive than French subjects, with a near-significant difference after a voiceless fricative (p = .058).

The next four charts (37-41) show the results for all environments, with a different chart for each of the four languages, allowing the comparison of different environments within each language.

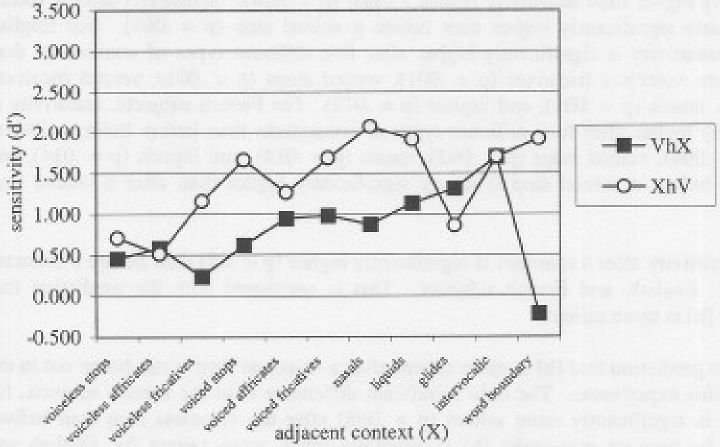
(37) *Sensitivity (d') by environment for Turkish*



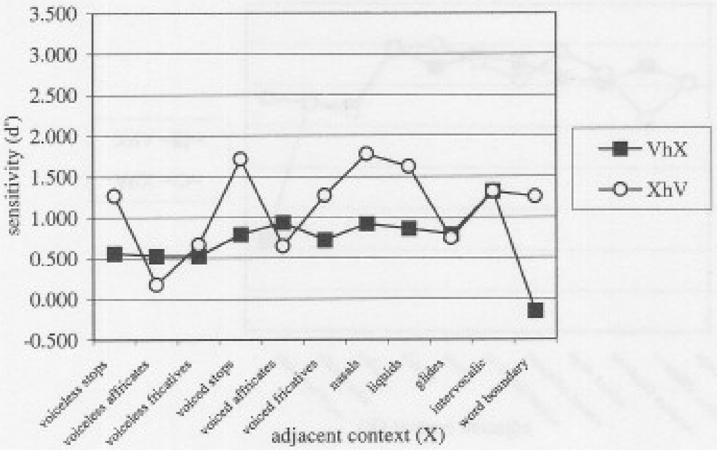
(38) *Sensitivity (d') by environment for Arabic*



(39) *Sensitivity (d') by environment for English*



(40) *Sensitivity (d') by environment for French*



Sensitivity is significant for word-initial vs. word-final environments for all four groups of speakers. For Arabic, sensitivity before a voiceless affricate is significantly higher than sensitivity after a voiceless affricate (p = .003). Sensitivity after a nasal is significantly higher than sensitivity before a nasal (p = .024). Sensitivity after a voiced stop is nearly significantly higher than before a voiced stop (p = .065). For English subjects, sensitivity is significantly higher after five different types of consonants than before them: voiceless fricatives (p < .001), voiced stops (p < .001), voiced fricatives (p = .017), nasals (p < .001), and liquids (p = .013). For French subjects, sensitivity is significantly higher after four different types of consonants than before them: voiceless stops (p = .008), voiced stops (p = .002), nasals (p = .014), and liquids (p = .014), and sensitivity before a voiced stop is nearly significantly higher than after a voiced stop (p = .077).

Sensitivity after a sonorant is significantly higher (p < .05) than before a sonorant for Arabic, English, and French subjects. This is consistent with the prediction that prevocalic [h] is more salient.

The prediction that [h] is more salient after a voiceless stop is not borne out in the results of this experiment. The only significant difference is in the French subjects, for whom [h] is significantly more salient (p = .008) after the voiceless stop than before. This may be because prevocalic [h] is generally much more salient for English and French subjects. Particularly in the English results, while there is a great difference between before and after most consonants, the difference is much smaller for voiceless stops and voiceless affricates. This may be an effect of aspiration that is not large enough to overcome the effect of prevocalicity.

Sensitivity is significantly higher before a voiceless affricate than after only for Arabic subjects (p = .003). However, since there is no similar difference for voiceless stops, this may not be entirely attributable to aspiration.

Sensitivity after liquids and nasals is significantly higher than after voiceless stops for English subjects (p < .001), but not for French subjects. Only sensitivity after liquids is higher for Arabic subjects (p = .005), although sensitivity after nasals is nearly significantly higher than after voiceless stops (p = .060). Sensitivity after liquids and nasals is significantly higher (p < .05) than after voiceless affricates for all groups of subjects, and significantly higher than after voiceless fricatives for all groups of subjects with one exception. Sensitivity after nasals is only marginally higher than after voiceless fricatives for Arabic subjects (p = .107).

Word-initial [h] is not significantly more salient than intervocalic [h] for any group of subjects, but it is significantly more salient (p < .001) than word-final [h] for all groups of subjects. The effect of intervocalic voicing may not be enough to significantly impede perception, but word-final [h] suffers for being prepausal.

Sensitivity after a voiced stop is significantly higher than sensitivity after a voiceless stop for English subjects (p = .001), but not quite significant for French (p = .138) or Arabic subjects (p = .061). Aspiration does have an effect, but the effect is different for different groups of subjects. Aspiration is discussed further in §5 and §6.

Sensitivity before a sonorant is significantly higher (p < .05) than before a voiceless fricative for all groups of subjects, except that sensitivity before a nasal is not significantly higher for Arabic subjects. The effect of fricative noise appears to be present universally.

Sensitivity after a voiceless fricative is significantly higher than before a voiceless fricative for English subjects (p < .001), but not for any other group of subjects. This may be a result of the English prevocalic [h] rather than the salience of prevocalic [h], because the difference is only significant for English subjects.

4.3. Summary

The results show a striking difference between different groups of subjects. Turkish and Arabic subjects were very sensitive to [h] in most environments, and English and French subjects were considerably less sensitive, particularly in non-prevocalic environments, and English subjects were more sensitive than French subjects in prevocalic environments. (35) and (36) show that there is no environment where English or French subjects were more sensitive to [h] than Turkish or Arabic subjects. In fact, the difference in sensitivity between each of the two "[h]-sensitive" groups (Turkish and Arabic) and each of the two "non-sensitive" groups (English and French) is significant for all of the 21 environments.

Another salient aspect of these results is that despite differences in overall sensitivity between the different groups, the patterns of relative perceptibility are very similar across languages, and the lines tend to parallel each other. This is especially apparent in (36) and in the charts for English and French (39 & 40).

5. Considerations for aspirated stops

Aspirated stops are of special importance in this study because of the perceptual similarity between [h] and stop aspiration. Examining how subjects responded to aspirated stops is informative, and also raises some questions about experiment design.

Arabic subjects had higher sensitivity to [h] after voiceless stops than Turkish subjects, and French subjects had higher sensitivity than English subjects. Apparently the two pairs of languages differ in this respect for different reasons. Recall that the sensitivity measure d' is a function of correct identification rate and false alarm rate. Therefore there are two ways a d' can be lowered, either by lowering the correct identification rate, or by raising the false alarm rate.

In the case of Turkish and Arabic, the Turkish subjects had a slightly higher false alarm rate and a slightly lower correct identification rate. This is consistent with the previous explanation that the sensitivity of Turkish speakers in this environment is lowered by the fact that [h] is not contrastive in this environment.

Compared to Arabic and Turkish, the differences between English and French can be accounted for in another way. French does not have aspirated stops anywhere except word-finally (Valdman 1976). The French subjects had a higher false alarm rate after aspirated stops, which is understandable given that aspirated stops can be perceived as unaspirated stops followed by [h]. The difference in sensitivity is due to the fact the French subjects had a much higher correct identification rate (see Appendix D). The correct identification rate of the English subjects (30.36%) was lower than after any other consonant. The reason for this may be that the Turkish aspirated stops in the stimuli are not as heavily aspirated as English aspirated stops (Lewis 1967). To an English speaker, a stop with a comparatively short voice onset time followed by an [h] is not as likely to be perceived as having an [h], if the combined duration of the aspiration and the [h] is short enough to be simply the aspiration of a stop. (41) shows the VOT for all of the voiceless stop foil stimuli (intervocalic voiceless stops with no [h]) and the voiceless stop + [h] stimuli, along with the [h] identification rate for each group of stimuli. ([h] identification rate for the voiceless stop + [h] stimuli is the correct identification rate and [h] identification rate for the foil stimuli is the false alarm rate)

(41) *[h] identification rate for voiceless stops with and without [h]*

	VOT		[h] identification rate			
	average	range	Arabic	Turkish	French	English
voiceless stop + [h]	86 ms	51-115 ms	85.42%	83.55%	53.57%	30.36%
voiceless stop (foil)	44 ms	19-62 ms	10.42%	9.87%	11.90%	7.76%

In a study of noncoronal stop perception, Volaitis and Miller (1992) found that for a fast speech rate, English-speaking subjects found labial stops produced with VOTs up to 87.15 ms and velar stops with VOTs up to 92.10 ms to be "normal" voiceless stops, and stops with higher VOTs were "exaggerated" voiceless stops. Many of the voiceless stop + [h] stimuli in this study, which were produced by a speaker of Turkish, fall within the range that Volaitis and Miller found to be acceptable voiceless stops for English listeners

6. Experiment 3: English listeners

As mentioned in the previous section, providing subjects with a partial transcription removes the possibility of confusing an aspirated stop with an unaspirated stop followed by [h]. The English subjects in Experiment 2 may have had a low [h] identification rate after voiceless stops because they were aware that the stops in those stimuli were [p], [t], and [k], not [b], [d], and [g]. A long VOT was allowable without alarm because the voiceless stop could account for the VOT. If there were a possibility that the stops could be voiced, aspiration would be an indication that there was an [h]. In Ovcharova's (1999) study, subjects were not given a partial transcription, and this problem did not arise. To check the results of the previous experiment, another experiment was run, with a task more similar to Ovcharova's.

6.1. Methods

6.1.1. Subjects

The subjects in this experiment were 17 female and five male Ohio State University undergraduates, aged 18-27, all native speakers of American English who did not participate in the first experiment. The results of one Greek-English bilingual speaker were not included.

6.1.2. Procedures

Procedures were similar to the previous experiment. However, instead of seeing the partial transcription and clicking on the screen where [h] was heard, subjects were asked to choose between two responses: "h" if there was an [h] in the stimulus, and "Ø" if there was not. Stimuli and other procedures were unchanged. See Appendix E for a sample screen view.

6.1.3. Data analysis

Because d' is not measurable without false alarm rates, only the correct identification rate was calculated. The correct identification rates (C/I) for each environment were averaged. They were then compared to two different rates from the English subjects in Experiment 2: the C/I rate, where [h] was reported in the correct position, and the total identification (T/I) rate, where [h] was reported in a stimulus containing [h], even if it was reported in the wrong place. This second measure is more similar to the C/I rate for the present experiment, where subjects simply reported whether or not they heard [h].

6.2. Results and discussion

In all environments, the correct identification rates from Experiment 3 were lower than or similar to the other two rates, as expected, since the responses contributing to the C/I rate for each environment are a subset of the responses contributing to the T/I rate. The T/I rate for the partial transcription task was virtually the same as or higher than the C/I rate for the task without transcription in all but four environments: after voiceless stops, after voiceless affricates, after liquids, and after glides. In the case of the stops and affricates, this means that subjects were less likely to report hearing an [h] if they knew a voiceless stop or affricate was in the stimulus and accounted for at least part of what they heard.

(42) *Identification rates after obstruents for both experiments*

Context		Experiment 2		Experiment 2	Experiment 3
		C/I Rate		**T/I Rate**	**C/I rate**
voiceless stop	[p, t, k]	51.79%	>	46.43%	30.36%
voiceless affricate	[tʃ]	51.79%	>	47.02%	35.71%
voiceless fricative	[f, s, ʃ]	67.86%	<	70.24%	62.50%
voiced stop	[b, d, g]	64.29%	<	74.40%	63.10%
voiced affricate	[dʒ]	58.93%	<	64.29%	57.14%
voiced fricative	[v, z, ʒ]	66.67%	<	73.81%	63.10%

6.3. Summary

For Experiment 2, this means that the identification rates and in particular the false alarm rates for stimuli containing an [h] following a voiceless stop or affricate were likely lower than they would have been if the voiceless stop and affricate were not transcribed. As a result, true sensitivity to [h] following a voiceless stop or affricate is probably lower than the results indicate, and aspiration may have more of an effect on [h] perception than Experiment 2 was able to show.

7. Discussion: the influence of phonology on perception

If phonology had no influence on perception, the results for all four groups of subjects should be the same. They are clearly not the same, and they are not the same for a number of reasons.

In general, Turkish and Arabic sensitivity is very high, in fact nearly perfect, reaching a ceiling level and thus making comparison of the two languages and the different environments difficult. Several environments had zero false alarms,[7] and for Turkish subjects, five environments (before nasals, liquids, and voiced fricatives and affricates, and after liquids) had correct identification rates above 98%. Arabic subjects had correct identification rates of 100% for before liquids and glides. For Turkish and Arabic subjects, detecting [h] in these environments is too easy for the results to show anything more than that detection is easy. Noise could be added to the signal to make the task more difficult, but noise can affect stimuli in unexpected ways, and it is important for the experiments in this study to be run without noise. However, replicating the experiments with noise may prove to be informative as well.

In contrast to the high sensitivity of Turkish and Arabic subjects, the opposite problem presents itself in the results of the English and French subjects. The level of sensitivity is very low for many environments, particularly postvocalic environments, so that comparison between environments is difficult because they are all about as low as they can be. Zero d' is chance performance, and the results for English and French are very near zero for several environments.

One approach to analyzing these results is to recognize that in general, Turkish and Arabic subjects perform near the ceiling and English and French subjects perform near the floor. Where there is a deviation from these low and high patterns, there may be a more specific effect to explain. Four factors appear to determine how the phonology of a language affects sensitivity to [h]. They are the presence of [h] in the language, familiarity with [h] in specific environments, the presence of non-prevocalic [h] in the language, and the contrastiveness of [h] in specific environments.

7.1. Presence of [h]

Of the four languages in the study, the one language which does not have [h], i.e., French, is the language whose subjects showed the least sensitivity to [h]. Thus, whether or not a language has [h] as a possible sound is a factor that contributes to sensitivity of [h].

---

[7] d' is undefined when the false alarm rate or correct identification rate is zero or one. In the event that the false alarm rate or the correct identification rate for any subject was zero or one, the total was adjusted by an amount equal to half of one error or correct identification. For example, 8/8 becomes 7.5/8 and 0/8 becomes 0.5/8.

7.2. Familiarity with [h] in specific environments

One factor that appears to determine [h] sensitivity is familiarity. Turkish and Arabic are both languages with [h] in many environments, as compared to English and French, and Turkish and Arabic subjects had higher sensitivity to [h] than English and French subjects in every environment. Being exposed to a language with [h] in many environments causes a listener to be more sensitive to the presence of [h] in those environments.

7.3. Non-prevocalic [h]

The dichotomy between the two groups of languages (Turkish and Arabic as opposed to English and French) is most striking in the non-prevocalic environments in (35). Generally, the rift is between the languages with [h] in many environments and languages with [h] in fewer environments, but in this case the rift is between languages that have [h] before consonants (Turkish and Arabic) and languages that do not allow [h] before consonants (English and French). The split is not so severe in (36), which shows prevocalic environments, because English has [h] in these environments. English subjects were significantly more sensitive to [h] in a number of prevocalic environments than French subjects, because English has prevocalic [h], and thus English speakers are more familiar with it.

Similarly, Turkish, Arabic, and English all have prevocalic [h], but Turkish and Arabic subjects were nevertheless more sensitive to [h] in prevocalic environments, although English subjects were more sensitive than French subjects. Apparently the skill of perceiving non-prevocalic [h] is transferable to prevocalic [h] (and not available to English speakers). Being able to perceive [h] when it is not followed by a vowel makes listeners even more able to perceive it when it is followed by a vowel. Listeners must have the ability to recognize [h] using a smaller number of cues, and their increased ability to utilize these cues benefits their [h] perception even in environments where more cues are present.

The fact that French subjects were more sensitive to prevocalic [h] than to postvocalic [h] is supportive of the hypothesis that [h] should be generally more perceptible before vowels than after, because the onset of the vowel is more salient than the offset. This may in part explain the difference in sensitivity of prevocalic and postvocalic [h] for English subjects, although the lack of non-prevocalic [h] in English is likely responsible. Nevertheless, this is a likely reason for the smaller difference between lowest and highest sensitivity in (38). The unfamiliarity of French subjects in particular is partially compensated for by the acoustical advantage claimed by prevocalic consonants. As measured in this experiment, Arabic and Turkish subjects are nearly maximally sensitive to [h] in many prevocalic and postvocalic environments, and perhaps do not need the auditory advantage afforded by prevocalic [h].

7.4. Contrast

Another factor contributing to sensitivity is contrast. While the sensitivity of Turkish and Arabic listeners is virtually the same in nearly all environments, it does differ in three environments: after voiceless stops, after voiceless fricatives, and after voiced fricatives. Two of these are environments where [h] can be deleted in Turkish. This is not the case with all of the environments where [h] deletes in Turkish, but perhaps these are the weakest in terms of acoustic cues, as opposed to other environments where the cues may be robust enough to overcome the lack of native language contrast.

When there is optional [h] deletion, the contrast between [h] and the lack of [h] is not meaningful. In Arabic, where this contrast is typically maintained, listeners are more sensitive to its presence or absence. In Turkish, where this contrast is often neutralized, it is less necessary for listeners to be sensitive to [h] in these environments in order to recognize words. Thus, a lack of contrast leads to a lack of sensitivity.

The four factors will be addressed further in the next section, where they are important in the model for predicting d'.

8. Discussion: the influence of perception on phonology

If perception influences phonology in Turkish [h] deletion, the environments where deletion is observed would be expected to be perceptually weak universally. The first experiment showed that for Turkish subjects, sensitivity before and after each type of consonant is consistent with deletion patterns. The results for Arabic, English, and French subjects are supportive.

Constructing a universal perceptibility scale (uninfluenced by phonology) for [h] environments based on the results from the four languages is difficult because the relative salience of [h] environments is different for each group of subjects. Developing a model of sensitivity may help to isolate the universal and language-specific factors and indicate what might be universal.[8]

In Experiments 1 and 2, sensitivity (d') was computed as a function of normalized correct identification and false alarm rates. In this section a model for predicting d' will be proposed, based on what are hypothesized to be the factors that determine the sensitivity of a speaker of a given language to the presence of [h] in a given environment. In this study, subjects speaking four different languages were tested in 21 different phonetic environments. This gives 84 possible combinations of environments and languages. For each of the 84 cases, phonetic and language-specific factors are relevant.

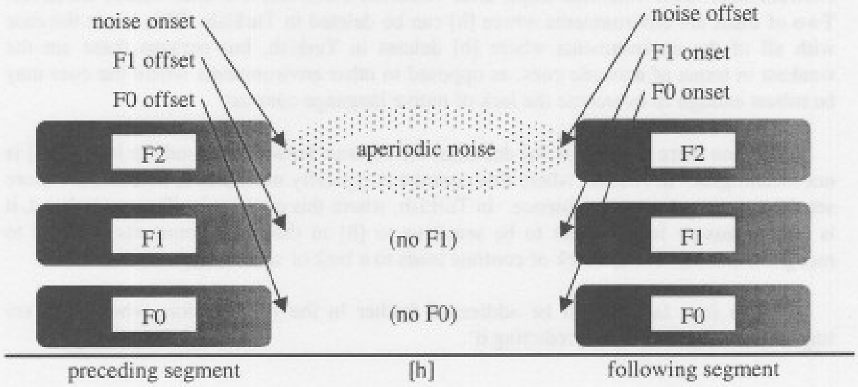
Ten variables were considered. Six variables are based on phonetic cues proposed in §2, noise onset, noise offset, F1 offset, F1 onset, F0 offset, and F0 onset.

---

[8] See Mielke (to appear a) for a more detailed discussion of the influence of perception on phonology.

These cues are illustrated in (44). Four language-specific variables are based on the factors discussed in §7: presence of [h], non-prevocalic [h], familiarity, and contrast.

(43) *Universal (phonetic) variables*



The values of the phonetic variables are determined by environment regardless of language. Each environment receives a value of zero or one for each of the six variables, depending on whether that cue to [h] is present in the environment.

(44) *Universal (phonetic) variables*

Environment	F0 offset	F0 onset	F1 offset	F1 onset	noise onset	noise offset
vls stop __ vowel	0	1	0	1	0	1
vls affricate __ vowel	0	1	0	1	0	1
vls fricative __ vowel	0	1	0	1	0	1
vcd stop __ vowel	1	1	0	1	1	1
vcd affricate __ vowel	1	1	0	1	0	1
vcd fricative __ vowel	1	1	0	1	0	1
sonorant __ vowel	1	1	1	1	1	1
glide __ vowel	0	0	1	1	1	1
# __ vowel	0	1	0	1	1	1
vowel __ vowel	0	0	1	1	1	1
vowel __ vls stop	1	0	1	0	1	1
vowel __ vls affricate	1	0	1	0	1	1
vowel __ vls fricative	1	0	1	0	1	0
vowel __ vcd stop	1	1	1	0	1	1
vowel __ vcd aff	1	1	1	0	1	1
vowel __ vcd fricative	1	1	1	0	1	0
vowel __ sonorant	1	1	1	1	1	1
vowel __ glide	0	0	1	1	1	1
vowel __ #	1	0	1	0	1	1

The language-specific variables receive a value of zero or one depending on the phonology of each language. "Presence of [h]" is 1 for Turkish, Arabic, and English, which have /h/ in their inventories, and 0 for French which does not. "Non-prevocalic [h]" is 1 for Turkish and Arabic, which permit [h] in non-prevocalic environments, and 0 for English and French, which do not. "Familiarity" is 1 for those environments where [h] is allowed in a language, so it is 1 for most environments in Turkish and Arabic, and for prevocalic environments in English, and 0 for all environments in French. "Contrast" is 1 for the subset of environments with a value of 1 for "Familiarity" that have a meaningful contrast between [h] and its absence in those environments.

(45) *Language specific variables*

Variable	Turkish	Arabic	English	French
Presence of [h]	1	1	1	0
Non-prevocalic [h]	1	1	0	0
Familiarity[9]	0 or 1	0 or 1	0 or 1	0 or 1
Contrast	0 or 1	0 or 1	0 or 1	0 or 1

The ten variables are summarized in (46)

---

[9] Values of familiarity and contrast variables depend on environment.

(46) *Variables hypothesized to be factors in determining sensitivity*

Variable		Description	Value
Universal (phonetic) variables	F0 onset	The transition from [h] to the following segment is marked by the onset of voicing.	0 or 1
	F0 offset	The transition from the preceding segment into [h] is marked by the offset of voicing.	0 or 1
	F1 onset	The transition from [h] to the following segment is marked by the onset of the F1 resonance.	0 or 1
	F1 offset	The transition from the preceding segment into [h] is marked by the offset of the F1 resonance.	0 or 1
	Noise onset	The transition from the preceding segment into [h] is marked by the onset of aperiodic noise.	0 or 1
	Noise offset	The transition from [h] to the following segment is marked by the offset of aperiodic noise.	0 or 1
Language-specific variables	Presence of [h]	The language has [h] in its consonant inventory	0 or 1
	Non-prevocalic [h]	The language has [h] in non-prevocalic position.	0 or 1
	Familiarity	The language has [h] in the environment in question.	0 or 1
	Contrast	The presence of [h] is contrastive in the environment in question.	0 or 1

A stepwise linear regression was performed, and six of the ten variables were found to have a significant contribution to d'. Whether or not a language has [h] in non-prevocalic position was found to be the largest contributing factor to sensitivity in general, as this is the difference between the two languages with very high sensitivity and the two languages with very low sensitivity. Four of the other five significant variables are phonetic. In all three sets of cues, the onset of the resonance or noise is significant, which is consistent with the predictions based on Wright (1996) and Fujimura et. al (1978) that the onset of a stimulus would be more important for [h] perception than the offset of the same stimulus. This holds true even though the onset of noise occurs at the beginning of [h] and the onset of F0 and F1 occur at the end of [h]. The offset of F0 is also significant, as well as whether or not [h] is contrastive in the environment.

(47) *Stepwise linear regression: variables with a significant contribution to d'*

	Variable	Coefficient	T	P
a.	Non-prevocalic [h]	1.363	14.627	<.001
b.	F0 onset	.413	3.554	.001
c.	Contrast	.349	3.609	.001
d.	Noise onset	.323	3.648	.010
e.	F1 onset	.453	3.717	<.001
f.	F0 offset	.256	2.120	.037

The formula for predicting sensitivity based on these variables is given in (48).

(48) *Formula for predicting d'*

$$d' = 1.363a + .413b + .349c + .323d + .453e + .256f$$

where: $a = 1$ if the language has non-prevocalic [h]; otherwise $a = 0$
$b = 1$ if the onset of voicing is at the end of [h]; otherwise $b = 0$
$c = 1$ if [h] is contrastive in the environment; otherwise $c = 0$
$d = 1$ if aperiodic noise begins at the beginning of [h]; otherwise $d = 0$
$e = 1$ if the F1 resonance begins at the end of [h]; otherwise $e = 0$
$f = 1$ if the offset of voicing is at the beginning of [h]; otherwise $f = 0$

Based on the four phonetic cues that have been found to be significant variables, the ranking in (49) is a basic universal perceptibility scale. The coefficients found in (48) are multiplied by the universal phonetic variables for each environment, giving a predicted d' for each environment. These environments can then be ranked by predicted d', giving a universal perceptibility scale. In reality, language-specific factors would influence d' as well. In this model, that would not affect the ranking.

(49) *Predicted d' (excluding language-specific variables, i.e., for a = 0 & c = 0)*

Environment	Cues								Predicted d'
	F0 onset		noise onset		F1 onset		F0 offset		
sonorant __ V	.413	+	.323	+	.453	+	.256	=	1.445
V __ sonorant	.413	+	.323	+	.453	+	.256	=	1.445
vcd stop __ V	.413	+	.323	+	.453	+	.256	=	1.445
# __ V	.413	+	.323	+	.453	+	0	=	1.189
vcd aff __ V	.413	+	0	+	.453	+	.256	=	1.122
vcd fric __ V	.413	+	0	+	.453	+	.256	=	1.122
V __ vcd aff	.413	+	.323	+	0	+	.256	=	0.992
V __ vcd stop	.413	+	.323	+	0	+	.256	=	0.992
V __ vcd fric	.413	+	.323	+	0	+	.256	=	0.992
vls aff __ V	.413	+	0	+	.453	+	0	=	0.866
vls fric __ V	.413	+	0	+	.453	+	0	=	0.866
vls stop __ V	.413	+	0	+	.453	+	0	=	0.866
V __ V	0	+	.323	+	.453	+	0	=	0.776
glide __ V	0	+	.323	+	.453	+	0	=	0.776
V __ glide	0	+	.323	+	.453	+	0	=	0.776
V __ vls stop	0	+	.323	+	0	+	.256	=	0.579
V __ vls aff	0	+	.323	+	0	+	.256	=	0.579
V __ vls fric	0	+	.323	+	0	+	.256	=	0.579
V __ #	0	+	.323	+	0	+	.256	=	0.579

With this hypothetical universal scale, it is possible to compare these environments to the deletion environments, and evaluate Hypothesis 1, that [h] is less perceptible in the environments were it deletes in Turkish.

(50) *Comparison of Turkish [h] deletion environments to perceptibility scale for Turkish and hypothesized universal scale*

Perceptibility scale based on Turkish (Experiment 1)	
Environment	**Deletion**
liquid __ V	
nasal __ V	
V __ vcd stop	
V __ vcd fric	
V __ liquid	YES
vls aff __ V	YES
V __ nasal	YES
V __ vcd aff	
vcd stop __ V	
V __ vls stop	
V __ vls aff	
vcd fric __ V	
V __ vls fric	YES
# __ V	
V __ vls aff	
V __ V	YES
vls stop __ V	YES
V __ glide	
V __ vls fric	YES
glide __ V	
V __ #	YES

more perceptible

less perceptible

Perceptibility scale based on six phonetic factors	
Environment	**Turkish Deletion**
liquid __ V	
nasal __ V	
V __ liquid	YES
V __ nasal	YES
vcd stop __ V	
# __ V	
vcd aff __ V	
vcd fric __ V	
V __ vcd aff	
V __ vcd stop	
V __ vcd fric	
vls aff __ V	YES
vls fric __ V	YES
vls stop __ V	YES
V __ V	YES
glide __ V	
V __ glide	
V __ vls stop	
V __ vls aff	
V __ vls fric	YES
V __ #	YES

In both rankings, there is a tendency for the deletion environments to be the environments where [h] is less perceptible, but based on these rankings it is not the case that all of the deletion environments are less salient than all of the non-deletion environments. The fact that some of the deletion environments are the environments where [h] is less salient, but that deletion environments do not seem to be exclusively the least salient environments, can be explained in a number of ways.

First, the way sensitivity was measured in the experiments did not duplicate the way sensitivity is used in everyday conversation. No noise was added to the stimuli, and the quiet environment that was created in the sound booth is not very similar to real-world listening conditions. For the Turkish and Arabic subjects, detecting [h] was very easy in most environments, particularly for the ones found to be most salient, and therefore the environments found to be least confusable in the controlled environment of the sound booth may not be particularly relevant in conversation.

Second, two listeners who differ in their sensitivity to [h] may attend to different cues. The model of sensitivity advanced in this paper assumes that all subjects are using the cues in the same way. For example, the coefficients for F0 offset, F1 offset, and noise onset may need to be higher for speakers of languages with non-prevocalic [h] than they are for speakers of languages without non-prevocalic [h], because these speakers have more experience utilizing these cues exclusively to recognize [h]. The above model of sensitivity does not allow this. See Mielke (to appear a) for a more sophisticated model of sensitivity.

Third, there may be non-perceptual factors involved. Environments for deletion may be generalized in ways that may not match the results of the experiments (see also Cole and Iskarous 2001). Perception *influences* phonology, after all. It does not replace phonology.

9. Conclusion

These experiments have demonstrated a bi-directional relationship between perception and phonology. The influence of phonology on perception is seen in the widely varying performances of subjects with different language backgrounds. The ability of listeners to detect [h] depends on where [h] is allowed in the native language and how it is used.

The influence of perception on phonology is seen in the asymmetrical pattern of [h] deletion in Turkish. Not only does the asymmetry match the patterns of perceptibility in the majority of environments Turkish, it matches patterns of perceptibility in Arabic, English and French, which are not influenced by Turkish phonology.

The goal of this paper has been to show that perception and phonology are related. The example of Turkish [h] deletion makes this relationship quite clear, and the fact that perception is important in this phonological phenomenon shows that this relationship indeed exists. This is not to claim that perception can explain everything in phonology, but that along with other factors, the influence of perception on phonology is not to be overlooked.

Appendix A - Examples of the distribution of [h]

The following lists of words show the distribution of [h] as claimed in (39).

A.1 - Turkish

Context	Example	Context	Example
vls stop __ vowel	şüphe	vowel __ vls stop	kahpe
vls aff __ vowel	meçhul	vowel __ vls aff	ahçı
vls fric __ vowel	ishak	vowel __ vls fric	ahşap
vcd stop __ vowel	tedhiş	vowel __ vcd stop	ahbap
vcd aff __ vowel	--	vowel __ vcd aff	mahcup
vcd fric __ vowel	mazhar	vowel __ vcd fric	mahzur
nasal __ vowel	imha	vowel __ nasal	köhne
liquid __ vowel	merhum	vowel __ liquid	ihlal
glide __ vowel	meyhane	vowel __ glide	ihya
# __ vowel	hayır	vowel __ #	sabah
vowel __ vowel	şahin		

Source: Kornrumpf (1979), Oflazer (1994)

A.2 - Arabic

Context	Example	Context	Example
vls stop __ vowel	məthaf	vowel __ vls stop	muhtaram
vls aff __ vowel	--	vowel __ vls aff	--
vls fric __ vowel	shur	vowel __ vls fric	wæhʃ
vcd stop __ vowel	rəbha	vowel __ vcd stop	rəhba
vcd aff __ vowel	--	vowel __ vcd aff	--
vcd fric __ vowel	mežhud	vowel __ vcd fric	lehža
nasal __ vowel	minha	vowel __ nasal	məhna
liquid __ vowel	marhaba	vowel __ liquid	məlul
glide __ vowel	tˤajha	vowel __ glide	jəhja
# __ vowel	həʒʒæm	vowel __ #	tˤah
vowel __ vowel	bæhit		

Source: Harrell (1966), M. Alaoui (p.c.)

A.3 - English

Context	Example	Context	Example
vls stop __ vowel	knighthood	vowel __ vls stop	--
vls aff __ vowel	beachhead	vowel __ vls aff	--
vls fric __ vowel	fishhook	vowel __ vls fric	--
vcd stop __ vowel	bloodhound	vowel __ vcd stop	--
vcd aff __ vowel	hedgehog	vowel __ vcd aff	--
vcd fric __ vowel	hogshead	vowel __ vcd fric	--
nasal __ vowel	inherit	vowel __ nasal	--
liquid __ vowel	forehand	vowel __ liquid	--
glide __ vowel	keyhole	vowel __ glide	--
# __ vowel	help	vowel __ #	--
vowel __ vowel	vehicular		

Appendix B – Stimuli for experiments 1-3[10]

Context	Target Stimuli				Foil Stimuli	
	Before Context		After Context			
voiceless stops	ühpe[11]	yhpɛ[12]	üphe	yphɛ	opa	o:pa
	ahte	ahtɛ	athe	athɛ	ita	i:ta
	ahkum	ahkum	akhum	akhum	eka	ɛ:ka
	ohpa	ohpa	opha	opha	üta	y:ta
	ahtı	ahtɯ	athı	athɯ	üpe	ypɛ
	ühkü	yhky	ükhü	ykhy	ekü	ɛky
	ehpe	ɛhpɛ	ephe	ɛphɛ	epe	ɛpɛ
	ihta	ihta	itha	itha	atı	atɯ
voiced stops	ühbe	yhbɛ	übhe	ybhɛ	oba	o:ba
	ahde	ahdɛ	adhe	adhɛ	ida	i:da
	ahgum	ahgum	aghum	aghum	ega	ɛ:ga
	ohba	ohba	obha	obha	üda	y:da
	ahdı	ahdɯ	adhı	adhɯ	übe	ybɛ
	ühgü	yhgy	üghü	yghy	egü	ɛgy
	ehbe	ɛhbɛ	ebhe	ɛbhɛ	ebe	ɛbɛ
	ihda	ihda	idha	idha	adı	adɯ

---

[10] The majority of these nonwords are from Ovcharova (1999). All stimuli were recorded new for this study, and additional nonwords were added for environments not included in Ovcharova's study.

[11] The first column is Turkish orthography, as the stimuli were presented on screen (without "h"s).

[12] The second column is IPA transcription.

voiceless fricatives	ihsa	ihsa	isha	isha	isa	isa
	ahşe	ahʃɛ	aşhe	aʃhɛ	öşe	œ:ʃɛ
	ahfa	ahfa	afha	afha	afa	afa
	ehfe	ɛfhɛ	efhe	ɛfhɛ	ıse	ɯsɛ
	öhşüt	œhʃyt	öşhüt	œʃhyt	aşe	aʃɛ
	ahşı	ahsɯ	aşhı	aʃhɯ	afa	a:fa
	ıhse	ɯhsɛ	ıshe	ɯshɛ	ofe	o:fɛ
	ühfe	yhfɛ	üfhe	yfhɛ	öse	œ:ʃɛ
voiced fricatives	ihza	ihza	izha	izha	ova	ova
	ahje	ahʒɛ	ajhe	aʒhɛ	izo	izo
	ahva	ahva	avha	avha	aze	azɛ
	ehve	ɛhvɛ	evhe	ɛvhɛ	aza	a:za
	öhjüt	œhʒyt	öjhüt	œʒhyt	eje	ɛʒɛ
	ahjı	ahʒɯ	ajhı	aʒhɯ	aja	a:ʒa
	ıhze	ɯhzɛ	ızhe	ɯzhɛ	ive	i:vɛ
	ühve	yhvɛ	üvhe	yvhɛ	öze	œ:zɛ
voiceless affricates	ahça	ahtʃa	açha	atʃha	içi	itʃi
	uhçu	uhtʃu	uçhu	utʃhu	eçi	ɛ:tʃi
	ihçi	ihtʃi	içhi	itʃhi	ıçı	ɯ:tʃɯ
	ehçi	ɛhtʃi	açhı	atʃhɯ	içe	i:tʃɛ
	ühçü	yhtʃy	üçhü	ytʃhy	aça	a:tʃa
	ehçe	ɛhtʃɛ	eçhe	ɛtʃhɛ	uçu	utʃu
	ihça	ihtʃa	içha	itʃha	eçe	ɛtʃɛ
	öhçe	yhtʃɛ	öçhe	œtʃhɛ	öçe	œtʃɛ
voiced affricates	ahca	ahdʒa	acha	adʒa	ici	idʒi
	uhcu	uhdʒu	uchu	udʒu	eci	ɛ:dʒi
	ihci	ihdʒi	ichi	idʒi	ıcı	ɯ:dʒɯ
	ehci	ɛhdʒi	achı	adʒɯ	ice	i:dʒɛ
	ühcü	yhdʒy	üchü	ydʒy	aca	a:dʒa
	ehce	ɛhdʒɛ	eche	ɛdʒɛ	ucu	udʒu
	ihca	ihdʒa	icha	idʒa	ece	ɛdʒɛ
	öhce	œhdʒɛ	öche	œdʒɛ	öce	œdʒɛ

nasals	ahmı	ahmɯ	omha	omha	ena	ɛːna
	öhmü	œhmy	ömhü	œmhy	eme	ɛːme
	ehna	ɛhna	enha	ɛnha	ömü	œmy
	ehme	ɛhmɛ	emhe	ɛmhɛ	anu	anu
	ohnu	ohnu	anhe	anhɛ	emi	ɛmi
	ihma	ihma	emho	ɛmho	emi	ɛːmi
	ehne	ɛhnɛ	inha	inha	ame	aːmɛ
	ahme	ahmɛ	onhu	onhu	ina	ina
liquids	ihla	ihla	ilha	ilha	iri	iːri
	ihri	ihri	irhi	irhi	ile	iːlɛ
	ihle	ihlɛ	ilhe	ilhɛ	ara	ara
	ahra	ahra	arha	arha	ila	ila
	ihra	ihra	ilhi	ilhi	ere	ɛrɛ
	ühlü	yhly	arha	arha	oru	oːru
	ihru	ihru	urha	urha	ira	iːra
	ohlü	ohly	alhi	alhi	ela	ɛla
glides	uhya	uhja	uyha	ujha	ıya	ɯːja
	ehye	ɛhjɛ	eyhe	ɛjhɛ	öyu	œːju
	uhye	uhjɛ	uyhe	ujhɛ	ayu	aːju
	öhya	œhja	öyha	œjha	öyö	œːjœ
intervocalic	tahan	tahan	köhen	kœhen	taan	taːan
	rohum	rohum	keher	kɛhɛr	loum	loːum
	muhan	muhan	lohum	lohum	muan	muːan
	tihir	tihir	sahal	sahal	köen	kœːɛn
word-initial	halam	halam	hemon	hɛmon	alam	alam
	hürin	hyrin	helir	hɛlir	ürin	yrin
	holan	holan	holar	holar	olan	olan
	helor	hɛlor	honen	honen	elir	ɛlir
word-final	rulah	rulah	ralah	ralah	rala	ralaː
	nulah	nulah	nelih	nɛlih	nula	nulaː
	maloh	maloh	mulih	mulih	luna	lunaː
	amah	amah	ralih	ralih	muna	munaː

Nontarget Stimuli							
apte	aptɛ	ökne	œ	ımsa	ɯmsa	ırtı	ɯrtɯ
atke	atkɛ	öfte	tfe	ufta	ufta	umku	umku
opta	opta	itka	itka	örme	œrmɛ	opsa	opsa
esta	ɛsta	anka	anka	ilka	ilka	üptü	ypty
ibra	ibra	ümke	ymkɛ	üske	yskɛ	amsa	amsa
ekme	ɛkmɛ	iska	iska	amba	amba	ekse	ɛksɛ
utpe	utpɛ	onra	onra	ölte	œltɛ	olga	olga
arte	artɛ	atra	atra	ustu	ustu	ufsa	ufsa
avpı	avpɯ	elke	ɛlkɛ	arsa	arsa	adra	adra
ayka	ayka	onle	onlɛ	ıftı	ɯftɯ	afka	afka
ekle	ɛklɛ	üktü	ykty	arte	artɛ	alda	alda
ente	ɛntɛ	üspü	yspy	opke	opkɛ	upsu	upsu
ıvza	ɯvza	armı	armɯ	imdi	imdi	erbe	ɛrbɛ
omde	omdɛ	asre	asrɛ	ekme	ɛkmɛ	ilne	ilnɛ
olne	olnɛ	ölne	œlnɛ	ipli	ipli	itke	itkɛ
ikti	ikti	önke	œnkɛ	isti	isti	urnu	urnu
laban	laban	begin	bɛgin	poter	potɛr	atke	atkɛ
rapan	rapan	tüküs	tykys	falat	falat	utpe	utpɛ
kulun	kulun	seten	sɛtɛn	apte	aptɛ	arte	artɛ

Appendix C – Table of sensitivity results (Experiments 1 and 2)

Context		Turkish	Arabic	English	French
before voiceless stop	V __ [p, t, k]	2.583	2.500	0.455	0.561
before voiceless affricate	V __ [tʃ]	2.558	2.698	0.581	0.530
before voiceless fricative	V __ [f, s, ʃ]	2.423	2.449	0.236	0.529
before voiced stop	V __ [b, d, g]	2.861	2.656	0.621	0.793
before voiced affricate	V __ [dʒ]	2.769	2.773	0.945	0.938
before voiced fricative	V __ [v, z, ʒ]	2.841	2.837	0.977	0.721
before nasal	V __ [n, m]	2.838	2.696	0.867	0.923
before liquid	V __ [l, r]	2.841	2.972	1.127	0.860
before glide	V __ [j]	2.155	2.221	1.311	0.790

after voiceless stop	[p, t, k] __ V	2.233	2.573	0.705	1.270
after voiceless affricate	[tʃ] __ V	2.274	2.109	0.513	0.175
after voiceless fricative	[f, s, ʃ] __ V	2.144	2.621	1.155	0.666
after voiced stop	[b, d, g] __ V	2.707	2.893	1.664	1.715
after voiced affricate	[dʒ] __ V	2.838	2.613	1.262	0.650
after voiced fricative	[v, z, ʒ] __ V	2.426	2.728	1.683	1.268
after nasal	[n, m] __ V	2.964	3.004	2.072	1.773
after liquid	[l, r] __ V	3.028	2.903	1.911	1.620
after glide	[j] __ V	1.777	2.126	0.853	0.740
intervocalic	V __ V	2.248	2.256	1.699	1.317
word-initial	# __ V	2.376	2.310	1.919	1.252
word-final	V __ #	0.734	0.550	-0.220	-0.156

Appendix D - Response rates

The following tables show the average response rates for each type of stimulus in experiments 1 and 2. Correct responses are in boldface.

D.1 - Turkish

Context	[h] Location	Response			
voiceless stop		initial	before	after	final
	before	0.00%	**92.48%**	1.50%	1.50%
	foil	0.00%	6.58%	9.87%	10.53%
	after	0.66%	0.00%	**83.55%**	1.32%
voiced stop		initial	before	after	final
	before	0.00%	**95.39%**	1.97%	0.00%
	foil	0.00%	1.97%	0.66%	3.95%
	after	0.00%	1.32%	**89.47%**	1.32%
voiceless fricative		initial	before	after	final
	before	0.66%	**88.82%**	2.63%	1.32%
	foil	0.00%	9.21%	5.92%	0.66%
	after	0.00%	4.61%	**76.97%**	1.32%
voiced fricative		initial	before	after	final
	before	0.00%	**98.03%**	1.32%	0.00%
	foil	0.00%	5.26%	1.97%	0.00%
	after	0.66%	7.89%	**81.58%**	0.66%
voiceless affricate		initial	before	after	final
	before	0.66%	**99.34%**	0.00%	0.00%
	foil	0.00%	14.29%	3.01%	1.50%
	after	0.00%	1.32%	**80.26%**	1.32%

voiced affricate		initial	before	after	final
	before	0.00%	**99.34%**	0.00%	0.00%
	foil	0.00%	7.24%	0.66%	5.92%
	after	0.00%	0.66%	**95.39%**	0.00%
nasal		initial	before	after	final
	before	0.00%	**98.68%**	0.00%	0.00%
	foil	0.00%	5.92%	0.00%	2.63%
	after	0.00%	0.00%	**96.71%**	0.00%
liquid		initial	before	after	final
	before	0.00%	**99.34%**	0.66%	0.00%
	foil	0.00%	6.58%	0.00%	3.95%
	after	0.00%	1.32%	**98.68%**	0.00%
glide		initial	before	after	final
	before	0.00%	**98.68%**	1.32%	0.00%
	foil	0.00%	5.26%	1.32%	3.95%
	after	0.00%	14.47%	**77.63%**	0.00%
intervocalic		initial	intervoc.	post	final
	intervocalic	0.00%	**97.74%**	0.00%	0.00%
	foil	0.00%	15.79%	0.00%	0.00%
word-initial		initial	before	after	final
	initial	**92.11%**	0.00%	0.00%	0.00%
	foil	2.63%	0.00%	0.00%	0.00%
word-final		initial	before	after	final
	final	0.00%	0.66%	0.66%	**42.11%**
	foil	0.00%	1.32%	0.00%	9.21%

D.2 - Arabic

Context	[h] Location	Response			
voiceless stop		initial	before	after	final
	before	1.19%	**92.86%**	0.00%	1.19%
	foil	4.17%	9.38%	1.04%	10.42%
	after	2.08%	0.00%	**85.42%**	4.17%
voiced stop		initial	before	after	final
	before	2.08%	**93.75%**	0.00%	0.00%
	foil	6.25%	6.25%	1.04%	1.04%
	after	0.00%	0.00%	**95.83%**	3.13%
voiceless fricative		initial	before	after	final
	before	2.08%	**89.58%**	1.04%	2.08%
	foil	0.00%	8.33%	2.08%	16.67%
	after	3.13%	2.08%	**87.50%**	1.04%

voiced fricative		initial	before	after	final
	before	2.08%	**95.83%**	1.04%	0.00%
	foil	3.13%	3.13%	0.00%	13.54%
	after	1.04%	1.04%	**89.58%**	2.08%
voiceless affricate		initial	before	after	final
	before	1.04%	**95.83%**	0.00%	0.00%
	foil	2.38%	5.95%	1.19%	11.90%
	after	1.04%	0.00%	**72.92%**	6.25%
voiced affricate		initial	before	after	final
	before	3.13%	**95.83%**	0.00%	0.00%
	foil	0.00%	5.21%	1.04%	11.46%
	after	1.04%	0.00%	**86.46%**	4.17%
nasal		initial	before	after	final
	before	3.13%	**91.67%**	0.00%	0.00%
	foil	2.08%	3.13%	0.00%	4.17%
	after	1.04%	0.00%	**97.92%**	1.04%
liquid		initial	before	after	final
	before	0.00%	**100.00%**	0.00%	0.00%
	foil	2.08%	3.13%	0.00%	3.13%
	after	0.00%	5.21%	**94.79%**	0.00%
glide		initial	before	after	final
	before	0.00%	**100.00%**	0.00%	0.00%
	foil	0.00%	4.17%	0.00%	0.00%
	after	0.00%	4.17%	**91.67%**	2.08%
intervocalic		initial	intervoc.	post	final
	intervocalic	0.00%	**94.05%**	4.76%	0.00%
	foil	4.17%	10.42%	0.00%	0.00%
word-initial		initial	before	after	final
	initial	**93.75%**	1.04%	2.08%	0.00%
	foil	8.33%	2.08%	0.00%	0.00%
word-final		initial	before	after	final
	final	0.00%	1.04%	1.04%	**46.88%**
	foil	0.00%	2.08%	0.00%	22.92%

D.3 - English

Context	[h] Location	Response			
voiceless stop		initial	before	after	final
	before	5.44%	**24.49%**	14.29%	6.12%
	foil	3.57%	10.71%	7.74%	9.52%
	after	2.98%	7.74%	**30.36%**	5.36%

voiced stop		initial	before	after	final
	before	7.74%	**28.57%**	11.31%	2.38%
	foil	5.95%	8.33%	8.33%	9.52%
	after	2.38%	6.55%	**63.10%**	2.38%
voiceless fricative		initial	before	after	final
	before	2.98%	**11.90%**	29.17%	5.36%
	foil	2.98%	4.17%	23.81%	2.98%
	after	0.60%	1.79%	**62.50%**	5.36%
voiced fricative		initial	before	after	final
	before	7.14%	**35.71%**	8.93%	2.38%
	foil	5.36%	4.17%	7.74%	7.14%
	after	2.98%	2.38%	**63.10%**	5.36%
voiceless affricate		initial	before	after	final
	before	5.95%	**31.55%**	25.00%	4.17%
	foil	1.36%	10.88%	17.69%	2.72%
	after	3.57%	4.17%	**35.71%**	3.57%
voiced affricate		initial	before	after	final
	before	5.36%	**41.07%**	22.62%	3.57%
	foil	3.57%	8.93%	15.48%	1.19%
	after	1.79%	2.38%	**57.14%**	2.98%
nasal		initial	before	after	final
	before	7.14%	**45.83%**	2.98%	3.57%
	foil	9.52%	16.67%	2.38%	2.98%
	after	7.14%	10.71%	**70.83%**	0.00%
liquid		initial	before	after	final
	before	10.12%	**39.88%**	10.71%	4.17%
	foil	4.76%	3.57%	5.36%	5.95%
	after	4.17%	10.12%	**68.45%**	2.98%
glide		initial	before	after	final
	before	3.57%	**60.71%**	9.52%	7.14%
	foil	3.57%	9.52%	1.19%	0.00%
	after	4.76%	23.81%	**38.10%**	4.76%
intervocalic		initial	intervoc.	post	final
	intervocalic	0.68%	**82.99%**	1.36%	0.00%
	foil	5.95%	20.24%	1.19%	0.00%
word-initial		initial	before	after	final
	initial	**78.57%**	2.98%	1.19%	0.00%
	foil	4.76%	4.76%	3.57%	8.33%
word-final		initial	before	after	final
	final	8.33%	9.52%	5.36%	**14.88%**
	foil	5.95%	8.33%	0.00%	13.10%

D.4 - French

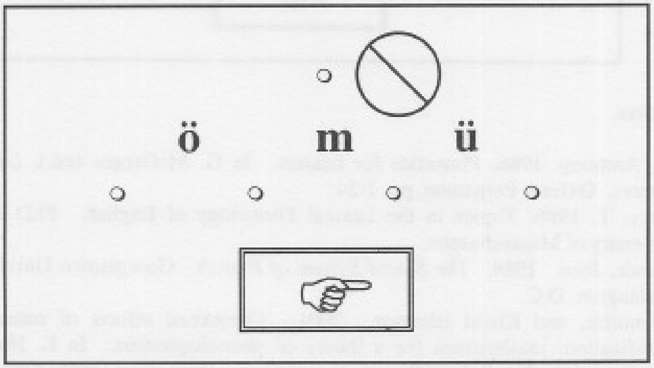
Context	[h] Location	Response			
voiceless stop		initial	before	after	final
	before	23.13%	**28.57%**	14.97%	4.08%
	foil	18.45%	8.93%	11.90%	3.57%
	after	12.50%	3.57%	**53.57%**	8.33%
voiced stop		initial	before	after	final
	before	28.57%	**35.71%**	14.88%	3.57%
	foil	16.07%	9.52%	8.93%	3.57%
	after	10.71%	5.36%	**65.48%**	5.36%
voiceless fricative		initial	before	after	final
	before	25.60%	**26.79%**	11.90%	1.19%
	foil	16.07%	8.93%	11.90%	5.95%
	after	15.48%	5.36%	**33.93%**	7.14%
voiced fricative		initial	before	after	final
	before	30.36%	**30.95%**	8.33%	2.98%
	foil	10.12%	7.74%	4.76%	4.76%
	after	13.10%	5.36%	**46.43%**	9.52%
voiceless affricate		initial	before	after	final
	before	24.40%	**36.31%**	22.62%	2.98%
	foil	14.29%	18.37%	19.73%	4.08%
	after	12.50%	6.55%	**26.79%**	4.17%
voiced affricate		initial	before	after	final
	before	23.81%	**41.07%**	20.24%	1.19%
	foil	8.93%	10.71%	23.21%	1.19%
	after	8.93%	4.17%	**44.64%**	4.76%
nasal		initial	before	after	final
	before	29.17%	**41.07%**	7.74%	2.98%
	foil	12.50%	10.71%	4.17%	4.76%
	after	18.45%	8.33%	**61.90%**	0.60%
liquid		initial	before	after	final
	before	24.40%	**31.55%**	19.05%	4.76%
	foil	5.95%	4.17%	2.98%	4.76%
	after	13.69%	8.93%	**55.95%**	4.17%
glide		initial	before	after	final
	before	29.76%	**51.19%**	11.90%	1.19%
	foil	16.67%	16.67%	2.38%	2.38%
	after	15.48%	16.67%	**33.33%**	5.95%
intervocalic		initial	intervoc.	post	final
	intervocalic	3.40%	**82.99%**	4.76%	2.04%
	foil	3.57%	36.90%	1.19%	0.00%

word-initial		initial	before	after	final
	initial	**56.55%**	2.38%	7.74%	8.33%
	foil	4.76%	2.38%	9.52%	5.95%
word-final		initial	before	after	final
	final	5.95%	5.95%	9.52%	**14.88%**
	foil	8.33%	3.57%	3.57%	10.71%

Appendix E - Sample screen views

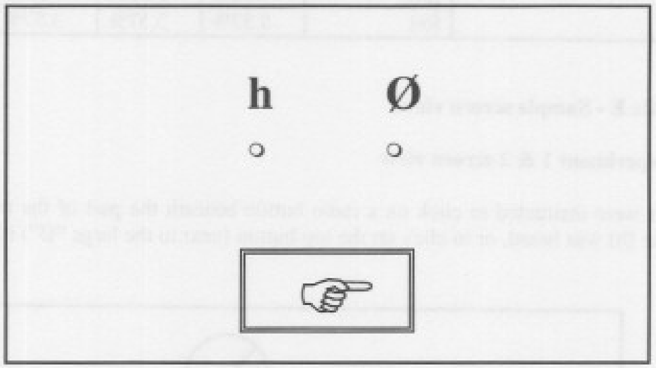
E.1 - Experiment 1 & 2 screen view

Subjects were instructed to click on a radio button beneath the part of the transcription where an [h] was heard, or to click on the top button (next to the large "Ø") if no [h] was heard.



E.2 - Experiment 3 screen view

Subjects were instructed to click on a radio button beneath the "h" if an [h] was heard, or to click on the button beneath the "Ø" if no [h] was heard.



References

Bladon, Anthony. 1986. Phonetics for hearers. In G. McGregor (ed.), *Language for Hearers*. Oxford: Pergamon, pp. 1-24.

Borowsky, T. 1986. Topics in the Lexical Phonology of English. PhD dissertation, University of Massachusetts.

Casagrande, Jean. 1984. *The Sound System of French*. Georgetown University Press. Washington, D.C.

Cole, Jennifer, and Khalil Iskarous. 2001. Contextual effects of consonant place identification: implications for a theory of phonologization. In E. Hume and K. Johnson (eds.) *The Role of Speech Perception in Phonology*. New York: Academic Press.

Fujimura, O., M. J. Macchi & L. A. Streeter, 1978. Perception of Stop Consonants with Conflicting Transitional Cues: A Cross-linguistic Study, *Language and Speech*, Vol 21, Part 4, 1978.

Green, David M. & John A. Swets, 1966. Signal detection theory and psychophysics. New York: Wiley.

Hammond, Michael 1999. The Phonology of English: A Prosodic Optimality-Theoretic Approach. Oxford University Press.

Harrell, Richard S. 1966. *A Dictionary of Moroccan Arabic: Arabic-English*. Georgetown University Press, Washington, D.C.

Heath, Jeffrey. 1987. *Ablaut and Ambiguity: Phonology of a Moroccan Arabic Dialect*. SUNY Press.

Hume, Elizabeth & Keith Johnson. 2001. A Model for the Interplay of Speech Perception and Phonology. In E. Hume and K. Johnson (eds.) *The Role of Speech Perception in Phonology*. New York: Academic Press.

Hura, S.L., Lindblom, B. & R. Diehl. 1992. On the role of perception in shaping phonological assimilation rules. *Language & Speech* 35, 59-72.

Jensen, John T. 1993. *English Phonology*. John Benjamins Publishing Company. Amsteram/Philadelphia.

Johnson, Keith. 1997. *Acoustic & Auditory Phonetics*. Blackwell. Oxford.

Keating, Patricia A. 1988. Underspecification in phonetics. *Phonology* 5. 275-292.

Kohler, K. 1990. Segmental reduction in connected speech: Phonological facts and phonetic explanations. In W.J. Hardcastle & A. Marchal (eds.) *Speech Production and Speech Modeling* (pp. 69-92). Dordrecht: Kluwer Academic Publishers.

Kornrumpf, H.-J. 1979. *Langenscheidt's Universal Dictionary: Turkish-English, English-Turkish*. Kangenscheidt. Berlin and München.

Lewis, G. L. 1967. *Turkish Grammar*. Clarendon Press. Oxford.

MacMillan, Neil A. and Douglas Creelman, 1991. *Detection Theory: A User's Guide*. Cambridge. New York.

Mielke, Jeff. 2000. A perceptual account of Turkish [h] deletion. Paper presented at the Montreal-Ottawa-Toronto Workshop on Phonology, York University.

Mielke, Jeff. to appear a. Directional asymmetry in Turkish [h] deletion: a perceptual account.

Mielke, Jeff. to appear b. Crosslinguistic patterning of laryngeal consonants.

Oflazer, Kemal. 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*. Chicago: CLS. 319-338.

Ovcharova, Olena 1999. A perception-based study of consonant deletion in Turkish. Poster presented at the ICPhS satellite meeting "The Role of Speech Perception Phenomena in Phonology"

Sezer, Engin. 1986. An autosegmental analysis of compensatory lengthening in Turkish. in L. Wetzels & E. Sezer, eds. *Studies in Compensatory Lengthening*. Dordrecht. Foris Publications.

Steriade, Donca. 1997. Phonetics in Phonology: The Case of Laryngeal Neutralization. ms. UCLA.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: a perceptual account. In E. Hume and K. Johnson (eds.) *The Role of Speech Perception in Phonology*. New York: Academic Press.

Valdman, Albert. 1976. *Introduction to French Phonology and Morphology*. Newbury House Publishers. Rowley, Massachusetts.

Volaitis, Lydia E., and Joanne L. Miller. 1992. Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *JASA* 92 (2), pt. 1.

Winer, B. J. 1971. *Statistical Principles in Experimental Design*. McGraw-Hill: New York.

Wright, R. 1996. Consonant Clusters and Cue Preservation in Tsou. PhD dissertation. UCLA.

Zawadowski, Y. N. (1978) *The Maghrib Arabic Dialects*. Central Department of Oriental Literature. Moscow.

OSU Working Papers in Linguistics 55: 172-199

A perceptual Account of Manner Dissimilation in Greek.

Georgios Tserdanelis
gdanelis@ling.ohio-state.edu

1. Introduction

In this paper, some phonological data from Greek are brought forth, namely a process of optional consonant cluster manner dissimilation found, with some variability, in all Modern Greek dialects (Newton, 1972; B. Joseph and I. Philippaki-Warburton, 1987). Greek consonant clusters consisting of two voiceless stops (e.g. /pt/) or two voiceless fricatives (e.g. /fθ/) can optionally dissimilate into a fricative plus stop [ft], but not into a stop plus fricative *[pθ]. The voiceless fricative /θ/ can change to a stop [t] when it is preceded by another voiceless fricative, so that an /fθ/ cluster can be realized as [ft]. However, when the second fricative in a fricative cluster is the sibilant [s], the cluster optionally dissimilates into stop plus [s]. Thus, the cluster /fs/ can be realized as [ps], the /f/ changing into a [p] in this case leading to an asymmetrical pattern of dissimilation when the sibilant /s/ is a member of such clusters. The introduction of these data is used as a starting point for the formulation of some general hypotheses about the perceptibility of such dissimilated clusters.

Manner dissimilation in Greek can be viewed as the result of what, in traditional phonological terms, would be a classic case of an output rule: two stops or two fricatives in the input are changed to a fricative and stop or a stop and fricative in the output. Why would users of a specific language favor one configuration of continuancy specification over another? To what extent is the resolution towards one particular configuration language specific and to what extent does unbiased, "universal" perceptibility influence the direction of such a resolution? In other words, given a specific phonological system, with a number of contrastive elements and processes, what is the extent to which cross-linguistic perceptual salience shapes and governs local (language specific) processes and contrasts? The bi-directional nature of such considerations is not hard to detect: For a given phonological system to come into existence it has to be subject to specific articulatory constraints (as there is a limit to what a human vocal tract can produce) and auditory constraints (as there is a limit to what the human ear can hear) from the very start. Yet, sound systems are subject to both universal, biologically determined articulatory and perceptual constraints, and at the same time language specific cognitive constraints, that must have evolved along side the observable sound systems. If we want to extract and formalize these constraints in a comprehensive formal theory of phonology, we have to look at existing phonological systems in order to guide and keep our analyses and predictions within the possible universe of natural human languages. The interplay then arises by acknowledging the fact that perception shapes phonology but at the same time phonology can shape speech perception and production.

This interplay of universal perceptual salience and language specific perceptibility is the goal of much recent research (for an overview see Hume and Johnson, 2001 this volume). The goal of this paper is to take a specific phonological process in Greek, namely consonant manner dissimilation and attempt to extract what is idiosyncratic to the particular phonological system that this process is attested in, and what falls under broader salience factors that have to do with independent perceptual considerations in a cross-linguistic framework. Dissimilation as a phonological process is of special interest to a research program that has as one of its goals to understand the role of perception in the evolution and structure of phonological systems across the world's observable languages.

The first part of this paper (sections 2-4) describes the phonological process of dissimilation in Greek in more detail, and the second part (sections 5-8) reports on the results of a discrimination experiment that was designed to gauge the perceptibility of dissimilated consonant clusters (e.g. [ft]) vs. non-dissimilated ones (e.g. [pt]). In order to achieve this, native speakers with two different phonological systems (English and Greek) were asked to evaluate stimuli derived from local contrasts in an alien, for the most part, sound system for the English listeners, and in a familiar, for the most part, sound system for the Greek listeners.

2. Phonotactics, syllable structure and lexical contrasts.

Standard Modern Greek has the consonant inventory given in table 1:

Table 1. Phonemic consonant inventory of Greek.

Place Manner	Bilabial	Labio-dental	Inter-dental	Alveolar	Velar
stops	p b			t d	k g
fricatives		f v	θ ð	s z	x ɣ
affricates				t^s d^z	
nasals	m			n	
laterals				l	
flap				ɾ	

Voiceless stops in Greek are unaspirated and contrast at three places of articulation: labial, coronal and dorsal. They also have three fricative counterparts at these same places of articulation [f, θ, x]. Voiceless fricatives and stops can combine to form biconsonantal clusters word-initially, word-medially and more rarely word-finally.

In Greek, voiceless bi-consonantal clusters can be found in a large number of words, and they seem to behave as complex syllable onsets since they can be found in absolute word-initial position as in [ktinɔs] 'beast', [xθεs] 'yesterday' [ksεnɔs] 'stranger'. Inter-vocalic clusters can be found in words such as [aptos] 'tangible', [efθis] 'straight' and [kɐfsis] 'burning'. Clusters also occur word finally, though they are not frequent in this position and always contain /s/ as in [vlɐks] 'idiot' or [miɔps] 'myopic person'. The possible combinatorics of voiceless C_1C_2 clusters in Greek, found in various positions in the word, are shown in table 2.

Table 2. Possible biconsonantal voiceless clusters attested in the Greek lexicon.

	p	t	k	f	θ	x	s
p	▨	**pt**εriɣɐ 'wing'		sa**pf**ɔ 'Sappho'			**ps**içi 'soul'
t		▨			a**tθ**is 'Atthis'		t^s (affricate)
k	ε**kp**iisi 'sale'	**kt**inos 'beast'	▨	ε**kf**ilɔs 'pervert'	ε**kθ**εsi 'proposition'	ε**kx**ɔrɔ 'assign'	**ks**εnɔs 'foreigner'
f	ε**fp**ɔrɔs 'affluent'	**ft**εrɔ 'feather'	ε**fk**ɔlɔs 'easy'	▨	**fθ**inɔs 'cheap'	**fx**ɐristɔ 'thank'	rε**fs**i 'flow'
θ					▨		
x	t^sɐ**xp**inis 'dapper'	**xt**εnɐ 'comb'			**xθ**εs 'yesterday'	▨	
s	**sp**iti 'house'	**st**εnɔs 'narrow'	**sk**εpsis 'thought'	**sf**irizɔ 'whistle'	**sθ**εnɔs 'strength'	**sx**ɐrɐ 'grid'	▨

In this consonant matrix some clusters are more common, that is exemplified by more words, than others ([pt] or [kt] for stops and [fθ, xθ] for fricatives especially) and dissimilation is more likely to apply in words with such clusters[1]. Clusters that appear word initially can also appear word medially and more rarely word finally. Clusters that are only given word medially cannot appear word initially. [ts] clusters are best analyzed as affricates in Greek even though they can occur across word boundaries as can some of the clusters that would fill many of the gaps (empty boxes) in the above table (no geminates are allowed in Standard Modern Greek, not even across word boundaries, as indicated by the dark shaded boxes). Clusters in the light shaded boxes are the least frequent. The velar fricative /x/ and the vowel /i/ have a palatal allophone [ç] which can be found after all three voiceless stops as in the word /fotia/ 'fire' pronounced as [fɔ'tçɐ]

---

[1] Some words are given in their version with a dissimilated cluster (e.g. **xt**εnɐ 'comb'). These words can also be found more rarely non-dissimilated (ktenɐ). Similarly, **kt**inos 'beast' can also be found as **xt**inos. Details about their distribution will be discussed in section 4.

but they are not of immediate interest to this study since, as it will be shown in the next section, are not the result of dissimilation. The same holds for the voiced bi-consantal clusters such as the ones found in words like [ɐvɣɔ] 'egg', or [ɐbdul] 'Abdul' which also seem to pattern differently than voiceless clusters and are not subject to dissimilation in Standard Modern Greek[2].

3. The dissimilation process

In Modern Greek, as noted in the introduction, voiceless consonant clusters of the form Stop+Stop and Fricative+Fricative (when neither of the fricatives is the sibilant /s/) optionally dissimilate to Stop+Fricative. For example, clusters of the form /pt/ (two stops) or /fθ/ (two non-sibilant fricatives) optionally dissimilate to a fricative plus stop, e.g. [ft].

(1) Two consecutive non-sibilant fricatives or stops[3] can change to fricative+stop.

a.	ptero	~	ftero	'feather'
	ktena	~	xtena	'comb'
	epta	~	efta	'seven'
	okto	~	oxto	'eight'
	ekpiisi	~	expiisi	'sale'
b.	xθes	~	xtes	'yesterday'
	fθinos	~	ftinos	'cheap'
	skefθika	~	skeftika	'I thought'
	anixθika	~	anixtika	'I was opened'
	fxaristo	~	fkaristo	'I thank'

When the cluster is formed by two stops the first member of the cluster changes into a fricative, whereas in the case of two fricatives it is the second member that changes. This process is schematized in (2) :

(2) Dissimilation output convergence:

Input Cluster		*Output*		*Input Cluster*	
[**p**t]	------>	[ft]	<------	[f**θ**]	(sounds in **bold** get dissimilated).
[**k**t]	------>	[xt]	<------	[x**θ**]	

epta 'seven' changes to *efta*	*fθinos* 'cheap' changes to *ftinos*
okto 'eight' changes to *oxto*	*xθes* 'yesterday' changes to *xtes*

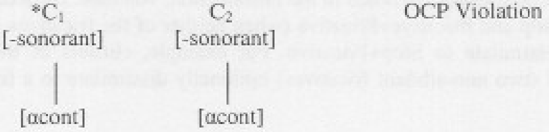
---

[2] In other Greek dialects, especially in Cypriot Greek, voiced clusters seem to dissimilate, for example /avγo/ turning into /afko/ 'egg' where we have both devoicing and dissimilation the /γ/ turning to a /k/ after a /v/ and then the /v/ assimilating for voicing to the /g/ turning into a /f/. A detailed description of such clusters in Cypriot Greek is found in Newton, B. 1972.

[3] Notice that almost all clusters have either /t/ or /θ/ as their second member, a tendency in the phonotactics of Greek which is not discussed in this paper.

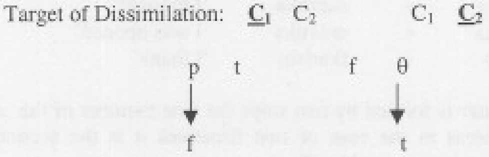
In (2) we see that there is convergence toward a preferred output of a Fricative+Stop cluster from both directions; that is, both Stop+Stop and Fricative+Fricative clusters, when dissimilation is applied to them, yield an identical output of Stop+Fricative. This process could be analyzed in formal phonology by invoking the Obligatory Contour Principle[4] proposed originally for other dissimilation phenomena (McCarthy 1986, Odden, 1987, Yip, 1988).

(3)



As schematized in (3) two adjacent obstruents that share the same value for the feature [continuant] violate the OCP. Thus, the optional dissimilation process in Greek can be viewed as a strategy used to avoid an OCP violation. The question then arises as to which consonant has to change its specification to satisfy the OCP, since based on the schema in (2) a change in the value of [continuant] in either C_1 or C_2 would suffice. And this is exactly what happens in Greek depending on whether the C1C2 cluster is formed by two stops or two fricatives. Recall that if it is formed by two stops, it is C_1 that undergoes dissimilation. But if the cluster is formed by two fricatives, it is C_2 that dissimilates. The differential target of manner dissimilation is shown in (4).

(4)



This observation then raises the issue of directionality in the dissimilation process in Greek. Why is it that a stop would undergo dissimilation when it is followed by another stop (a /p/ in a /pt/ cluster for example) whereas a fricative would undergo dissimilation only when it is preceded by another fricative (a /θ/ in a /fθ/ cluster)?

To complicate matters even further the sibilant fricative /s/ seems to enjoy a special status when it is found in fricative clusters as seen in the examples in (5).

---

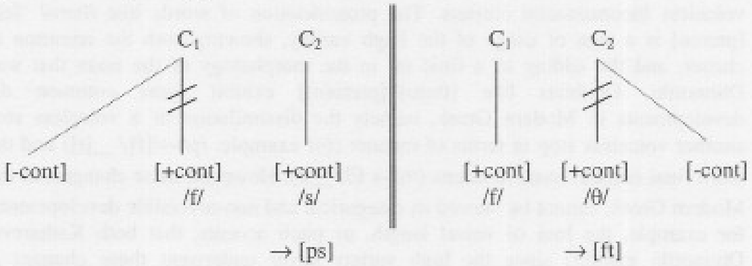
[4] Abbreviated as the OCP this principle is defined (McCarthy, 1986) thus: at the melodic level, adjacent identical elements are prohibited.

(5) Clusters with the sibilant /s/ in first and second position.

a.	sxini	~	skini	'rope'
	pisθika	~	pistika	'I was convinced'
	sfoŋgos	~	spogos[5]	'sponge'
b.	trex-o		e-trek-sa	'I run' (present ~ past)
	kafsimo	~	kapsimo	'burning'
	kaθ-izo		e-kat-sa[6]	'I sit' (present ~ past)

Thus, clusters with /s/ do not follow the generalization in (4). If /s/ is the first member of a cluster as in (5a) then the generalization holds, it is C2 that dissimilates. However, if /s/ is in a C2 position, as in the examples in (5b), it is C1 that dissimilates. This special case that results in a directionality asymmetry in the dissimilation output with regard to clusters with /s/ is schematized in (6).

(6) Sibilant Asymmetry



e.g. *θa γraf + so* [γrɐpsɔ] "I will write"
γraf + θi + ka [γreftikɐ] "I wrote myself"

When the second member of the cluster is the sibilant /s/, as is the case on the left in (6), then it is C_1 that dissimilates in a fricative cluster and not C_2 as expected in the non-sibilant fricative cluster to the right. If /s/ were to follow the non-sibilant fricative dissimilation pattern, /s/ might be expected to dissimilate to a homorganic stop [t] but this process is not attested in Greek, further supporting the special status of the sibilant fricative /s/ when it is found in consonant clusters[7].

---

[5] This particular example is a strictly diachronic one but it completes an otherwise incomplete pattern of dissimilation and it is used for sake of expository completeness.

[6] A /ts/ cluster is probably realized as a $[t^s]$ affricate by most speakers, but its derivation seems to follow the general schema.

[7] Finally, there are clusters that are underlying sequences of stop plus fricative or fricative plus stop (e.g. [ɛkθɛsi] 'essay', [ɛftixɔs] 'fortunately', [ɐftɔs] 'he' [skɛpsi] 'thought' etc.). These contour clusters already satisfy the OCP and show no alternations (except for some rare cases of hypercorrection, for example [skɛpsi] – [sxɛfsi] 'thought') and are not the result of dissimilation.

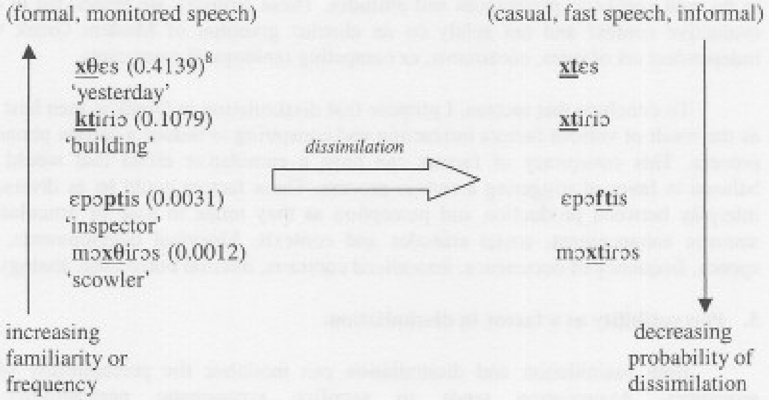
4. Sociolinguistic factors and dissimilation.

As noted above, manner dissimilation in Greek is optional. It is, however, a very robust and active process in the language. Its application is exemplified by L2 phonology patterns of Greeks learning English as a second language such as the pronunciation of the English word *fact* as [fɐxt] or *McDonalds* as [mɐxtɔnɐlts]. This optionality is subject to sociolinguistic influences that stem from a long history of persistent diglossia in Greek speaking communities that have traditionally employed two varieties with unequal status: the high variety called Katharevousa 'purifying' and the low variety called Dhimotiki 'popular' (Ferguson, 1959). In Katherevousa the dissimilation process was resisted, whereas in an "idealized" Dhimotiki phonological system, dissimilation would apply to all candidate clusters. However, the disentanglement of the two phonological systems is problematic since all speakers could code-switch at any given moment and the strong influence of the high variety on the low resulted in the present day situation where the varieties have merged into a highly variable system that exhibits both patterns. One of the most salient differences between the high variety and the low was the realization of voiceless biconsonantal clusters. The pronunciation of words like /ftero/ 'feather', as [pteron] is a sign of usage of the High variety, showing both the retention of a stop cluster, and the adding of a final /n/ in the morphology of the noun that was lost in Dhimotiki. Doublets like [ftero]~[ptero(n)] exhibit some common diachronic developments in Modern Greek, namely the dissimilation of a voiceless stop before another voiceless stop in terms of manner (for example: /p/→/[f]/ __[t]) and the loss of word final coronal nasal in nouns (/n/→ Ø/ __#). However, these changes, in the case of Modern Greek, cannot be viewed as categorical and non-reversible developments (unlike, for example, the loss of vowel length, or pitch accents, that both Katharevousa and Dhimotiki exhibit) since the high variety never underwent these changes and thus provided speakers with a constant source of variability that was the result of both diachronic sound changes and synchronic sociolinguistic bi-dialectism. This prolonged bi-dialectism can be best viewed as stylistic co-variation piggybacking on the normal development of Dhimotiki dialects alongside of the artificially archaizing Katharevousa style of speech that everybody was sooner or later exposed to (e.g. through the church).

The issues of the diachronic development of dissimilation in Greek are quite complex and beyond the scope of this paper, but dissimilation can also be observed synchronically in alternations that are found even in Katharevousa, in clusters with sibilants. For example, the past tense (aorist) of the ancient Greek verb [grɐp^{h}-ɔː] (1st person, singular present), which through regular sound changes, gave Modern Greek [ɣrɐf-ɔ] 'I write", is found in both varieties with a dissimilated /ps/ cluster [ɛɣrɐp-sɐ] (Dhimotiki), [ɛɣrɐp-sɔn] (Katharevousa) 'I wrote', having the /f/ of the stem turning into a [p] before the /s/ found in the past tense morpheme. Here there is no variability, except for paradigm internal alternations in the verb itself, and this is the result of inherited patterns found in ancient Greek that applied de-aspiration in clusters like [p^{h}s] blocking the regular development of [p^{h}] to [f] in the environment before a sibilant. Indeed the very existence of a cluster such as [fs] in modern Greek can be attributed to the influence of the learned tradition as codified in the high variety that, even though it "carried out" regular sound changes in one level of the phonology, namely the unconditioned change of [p^h] to [f], in another level preserved, or created new phonotactic conventions namely [fs]. Thus Dhimotiki, as it went through regular phonological developments, was never cut off from the conservatism of Katharevousa; at no point were these phonological changes fully realized, or categorical, due to the constant symbiosis of the two varieties (except perhaps in extremely isolated mono-dialectal communities).

The main point of this section is to show the intricate history behind dissimilation in Greek that seems to be the result of complicated diachronic, synchronic and sociolinguistic patterns. Dhimotiki (the low variety) has been declared the official language of Greece since 1974, but the present day language can be best viewed as the fusion of the two varieties more so than a pure form of one or another (Macridge, 1985). In many respects, and especially in the phonotactics found in learned vocabulary, Katharevousa patterns are very common to the point of being obligatory. In others, such as morphological patterns, Dhimotiki patterns are also practically obligatory. However, the main characteristic of this natural merging of the two patterns is the resulting optionality in the application or not of certain phonological processes such as dissimilation. Speakers then can exploit this optionality as a social marker, denotative of stylistic preferences (Kazazis, 1992). It can also be conditioned by other factors such as speech rate, word frequency, and ultimate source (high or low variety) of the lexical item containing such a candidate cluster for dissimilation. An idealized schematization of the different interactions between these factors is given in (7).

(7) Application of Dissimilation



---

[8] These are frequencies per thousand words extracted from the Hellenic National Corpus™ (HNC). This 13 million word corpus is developed by the ILSP (Institute of Speech and Language Processing) in Athens, Greece and described in Hatzigeorgiu et. al. (2000).

As shown in (7), speakers would be more likely to dissimilate the very common word for 'yesterday', a traditional Dhimotiki vocabulary item, less likely the Katharevousa, yet very frequent word for 'building', and even less likely the less frequent Katharevousa word for 'inspector'. Finally, the dissimilation of the /xθ/ cluster to [xt] in the word for 'scowler', which is a very infrequent Katharevousa word, would be highly improbable. The decreasing probability in the application of dissimilation is marked by the downward pointing arrow to the right of the dissimilated words and the increasing familiarity/frequency by the upward pointing arrow to the left.

This probabilistic approach in the application of certain phonological processes with sociolingusitic significance is best captured by Guy and Boberg (1997) in their discussion of the Obligatory Contour Principle. By examining the inherent phonological variability found in languages, they propose a new association of the initials OCP as Optional Contour Preference, treating the OCP family of constraints as probabilities rather than violable conditions. In the case of consonant manner dissimilation as found Modern Greek this approach seems to be highly relevant since any formal attempt to describe such a system would fail to capture the multitude of factors that could influence the process at hand. In particular, speech rate seems to be a very important factor in altering the probability of dissimilation. A small pilot study conducted with two native speakers of Greek who were asked to read a text with target clusters, at two different speech rates (one slow and careful, and one fast and casual) showed that speakers are more likely to dissimilate clusters when they read the text faster than slower (Tserdanelis, 2000). Of course, this finding is only relevant to reading style but a tenuous extrapolation to other speech events does not seem very far-fetched. If speakers choose when to apply the OCP on an item-by-item basis then any attempt to constrain this variability and to identify certain patterns in those choices would have to be based also on criteria that have to do with speakers' preferences and attitudes. These attitudes are embedded in a social evaluative context and not solely on an abstract grammar of Modern Greek with an independent set of rules, constraints, or competing rankings of constraints.

To conclude this section, I propose that dissimilation in Greek is then best viewed as the result of various factors interacting and conspiring to induce a certain phonological process. This conspiracy of factors can have a cumulative effect that would tip the balance in favor of triggering a certain process. These factors could be as diverse as the interplay between production and perception as they relate to ease of articulation and acoustic enhancement, social attitudes and contexts, historical developments, rate of speech, frequency of occurrence, lexicalized contrasts, internal borrowing, analogy, etc.

5. Perceptibility as a factor in dissimilation.

Both assimilation and dissimilation can modulate the perceptibility of sound sequences. Assimilation tends to sacrifice syntagmatic perceptibility while accommodating ease of articulation (Steriade, 2001) whereas dissimilation has been interpreted as enhancing perceptibility while demanding more complex articulations than in sequences of non-dissimilated segments (Suzuki, 1998). Gauging and understanding the relevant perceptibility of such segmental sequences then, can prove to be a fertile field of inquiry in order to understand in their totality the mechanics of processes such as manner dissimilation in Modern Greek. By isolating perceptibility, or any other factor that seems to have a bearing on a given phonological pattern, and treating that factor as a potentially independently motivated phenomenon, some interesting generalizations might be arrived at about the structure of sound systems cross-linguistically. Thus, a systematic decomposition of a process affecting a sound system, may lead to a deeper understanding of the phenomenology of the process itself and of possible language sound patterns in general. The interplay and the various interactions between production and perception, or between history (faithfulness) and function (optimization) in the structure and development of languages, are perhaps more important than any one of these factors studied in isolation. But in order to study these interactions, relevant data about some of the interacting components could be very informative. These components can take the form of social, grammatical, phonological and phonetic influences on a given sound system.

In this study, the role of perception in the evolution of dissimilation as a synchronic process in Modern Greek is investigated. Many have noted language preferences for particular sequential contrasts. Greenberg (1978), for example, gives some cross-linguistic generalizations about obstruent consonant clusters summarized in (8). In these implicational universals, a statement "x > y" means that the presence of x in a language implies that y also exists in that language.

(8) Contrast in continuancy is favored over its absence:

a) TTV > FTV, SFV
b) VTT > VFT, VTF
c) FFV > FTV, TFV
d) VFF > VFT, VTF

In (8) T stands for a stop consonant, V for any vowel and F for a fricative consonant. (8a) Then should be read as a stop + stop (TT) consonant cluster before a vowel (V) is less common cross-linguistically than a fricative + stop (FT) or a stop + fricative (TF) before a vowel (V). The same applies for postvocalic stop clusters as well (8b) and clusters with two fricatives (8c-d). According to these cross-linguistic observations, contrast in continuancy between adjacent prevocalic obstruent segments is more common than not. Notice, that Greenberg does not talk about the perceptibility of such clusters, but only about how they pattern quantitatively in the world's languages[9]. Functional models based on salient acoustic modulations in segment sequences have also been proposed to account for the type of generalizations given in (8) (Ohala 1992, Wright 1996). These approaches emphasize both the inherent qualities of segments and their syntagmatic optimization in terms of acoustic salience. For example, [s] is more perceptible than [θ], and a [t] before a vowel is more perceptible than a [t] before another stop. Thus, preference for certain

---

[9] In formal phonological theory, segmental sequential constraints have been proposed for observed cross-linguistic tendencies in the realization of consonant clusters, such as sonority-based models (Clements, 1988).

sequential contrasts that can result in a modulated acoustic signal such as in Greek dissimilation may have a perceptual basis.

With this as a basis, some general hypotheses about the role of perception in dissimilation in Greek are proposed. First, with regard to the preference for contour clusters, in terms of manner of articulation, the motivation for alternating between stops and fricatives in biconsonantal clusters may be due to perceptual enhancement (increased contrast) at the cost of coordinating two distinct manners of articulation next to each other (Ohala, 1990). In other words, place perception is enhanced in dissimilated clusters in Greek; a manner contrast between fricatives and stops increases the probability that both of the segments in a biconsonantal cluster can be perceived. Second, with regard to the directionality of the dissimilation.

In the case of the dissimilation process in Greek, recall that the preferred outcome of dissimilation is (FTV), fricative + stop before a vowel, unless the (F) is the sibilant [s] in which case the preferred outcome is (TFV), stop + sibilant before a vowel, always preserving [s] even when it is in second position in a fricative cluster. To understand this directionality, we hypothesize that listeners are be better at differentiating between one configuration over another, if indeed there is some degree of difference in the perceptibility of dissimilated clusters depending on what comes first and what comes second (given two choices: fricative or stop) in biconsonantal clusters. Taking into account the cues of different segments in various contexts (Wright, 1996), optimal arrangements can be predicted. For example, since fricatives as opposed to stops have internal cues (Johnson, 1997) by creating a cluster of a fricative followed by a stop an optimization of cues is achieved especially in absolute word initial, prevocalic position: [ftV] is better than [pθV] because both the stop burst and the vowel onset transitions associated with the prevocalic stop are preserved (Fujimura et al. 1978). Some of these cues are realized better in particular contexts, for example a stop burst and aspiration before a vowel rather than before a fricative, where fricative noise can mask them. Similarly, fricatives, even though they lack bursts, have some internal formant structure that can be used by listeners to identify place of articulation, whereas stops rely solely on the preceding and following vowels for place information extracted from formant transitions. As shown in table 3 the optimal configuration, in word initial, prevocalic position when cues are taken into account is for a fricative to precede a stop.

Table 3. Cues for stops and fricatives in clusters after silence.

	#pθV	#ftV	#ptV	#fθV
Burst for C1	NO	N/A	NO	N/A
Burst for C2	N/A	YES	YES	N/A
Formants for C1	NO	YES	NO	YES
Formants for C2	YES	YES	YES	YES

This optimization can be extended to other phonotactic environments as well, having perhaps as a starting point the absolute word initial position where the perceptual gain from the fricative+stop configuration is maximized. For example, both in absolute word initial position and inter-vocalically, a preference for prevocalic stops could also have a perceptual basis because CV formant transitions have been experimentally found to be more perceptible than VC transitions for stop place (Fujimura et al. 1978).

Finally, The sibilant /s/ when it is found in a cluster, never gets altered in Greek no matter what position (C1 or C2) it is found in. Sibilants, unlike other fricatives, are characterized by aperiodic high frequency energy and spectral peaks above 4 kHz, unique acoustic characteristics that perhaps render them more distinct perceptually. This could be a factor in Greek dissimilation where a stop occurs before a fricative (e.g /fsV/→[psV]) , unlike the optimal configuration, in terms of cue preservation, shown in table 3. Thus the inherent acoustic salience of sibilants may be an overriding factor in their syntagmatic perceptibility. Following Kohler 1990, Steriade 2000, we hypothesize that changing the /s/ would lead to a more noticeable change (cf.,) and thus is avoided by speakers. With regard to the dissimilation pattern in Greek then, we hypothesize that listeners will be better at discriminating between clusters that have as one of their members the sibilant /s/ than between clusters that do not. The presence of the salient acoustic cues of the sibilant would enhance the perceptibility of both members of a cluster, preserving the percept of two segments as opposed to one. The distinctiveness of /s/ then can be used by listeners to maintain robust perceptual contrasts with all other segments in the system both syntagmatically and paradigmatically (Hura et al, 1992).

The perceptibility of segment sequences that differ in their specification for continuancy vs. those sequences that do not can be directly tested experimentally, by having listeners discriminate between such clusters and then record the time it took them to arrive at such discriminatory decisions. These results then can be indirectly correlated with the perceptibility of such clusters: the longer it took for listeners to discriminate between two contrastive stimuli the harder those stimuli would be to perceive (Shepard, 1987). We can predict then that Greek listeners should be faster and more accurate at discriminating between contour clusters. Similarly they should also be faster and more accurate at discriminating between (FT) rather than (TF) clusters and between clusters with [s] rather than between clusters with non-sibilant fricatives.

To summarize this section, perceptual salience is introduced as one of the factors that could play a role in the realization of dissimilation. It is posited that a perceptual account might shed light on dissimilation and sequential contrasts in general. In order to test the above hypotheses empirical perceptual data were collected by means of an AX, reaction time discrimination experiment that tests the perceptibility of such clusters by Greek and English listeners, as discussed below.

6. Experimental design and data collection.

6.1 Method

6.1.1 Stimuli

An adult male, phonetically trained native speaker of Modern Greek, recorded nonsense VCCV stimuli containing all possible combinations of the relevant consonants in Greek [p,t,k,f,θ,x,s] that participate in the process of manner dissimilation. The consonants were flanked by the vowel [ɐ], with stress on the first syllable. There were two restrictions in the possible combinations in the consonant clusters constructed: there were no geminates, that is no [tt] or [ff] clusters, and no homorganic clusters such as [kx] or [fp], except in the case of clusters with the sibilant [s] where homorganic clusters such as [st] or [sθ] were constructed and included in the stimuli. The possible stop+stop, fricative+fricative, stop+fricative and fricative+stop, clusters including those with the sibilant [s], are given in (9):

(9) Cluster types (T=Stop, F=Fricative, S=sibilant):

a. TT:	tp, tk, pt, pk, kp, kt	6
b. FF:	θf, θx, xf, xθ, fx, fθ	6
c. FS/SF:	fs, xs, θs / sf, sx, sθ	6
e. TF/FT:	pθ, px, tf, tx, kf, kθ / θp, θk, ft, fk, xt, xp	12
f. TS/ST:	ps, ks, ts[10] / sp, sk, st	6

	Total Number of clusters:	36

The recordings were edited to ensure uniformity of segmental length and recording amplitude. Furthermore, if any of the clusters had a burst in the release of the first stop, in a stop+stop cluster such as in [tp] in [ˈɐtpɐ], the burst was deleted and replaced by silence to ensure again the uniformity of the stimuli, since there was some variability in the strength of the burst when a burst was present at all (most stop+stop clusters had very weak bursts or no bursts at all after the first stop[11]). The stimuli were organized into pairs of two nonsense words containing the clusters in (9) of the form [ˈ**ɐtpɐ**] ~ [ˈ**ɐtkɐ**] where the first member of the C_1C_2 cluster is kept constant [t] and the second C_2 was varied [p~k]. The opposite types were also included, that is pairs of the form [ˈ**ɐtpɐ**] ~ [ˈ**ɐkpɐ**] in which C_2 is kept constant [p] and C_1 varies [t~k]. All the pairs constructed in this way are given in appendix 1. Because of this variation in C_1 vs. C_2 constancy, there were a total of (36 x 2) = 72 pairs per block, and every listener was presented with 6 blocks of these pairs randomized, resulting in (72x6) = 432 stimuli pairs per session. The 36 sound files were recorded on analog TEAC cassette tape recorder (normal tape, Dolby NR_ON, mono-Left, recording level 6/10) in a sound booth at OSU phonetics lab

---

[10] [ts] clusters were constructed by splicing together a final [t] and an initial [s], in order to differentiate the cluster [ts] from the affricate [tˢ].

[11] In *pt/kt* sequences in Greek the first stop can optionally be released both word initially and word medially something that is less likely in English for example in the pronunciation of words like *apt* or *act*. (Pagoni-Tetlow, 1998).

using a HMD head mounted microphone and a SX202 Dual Mic pre-amp (gain 40). The stimuli were read in pairs 4 times by the speaker without a carrier phrase and with a short pause between each stimulus (~500ms) and a longer pause (>2000ms) between each pair. The best tokens were selected and digitized on an IBM PC running PCquirer at 22KHz, and saved in .WAV format.

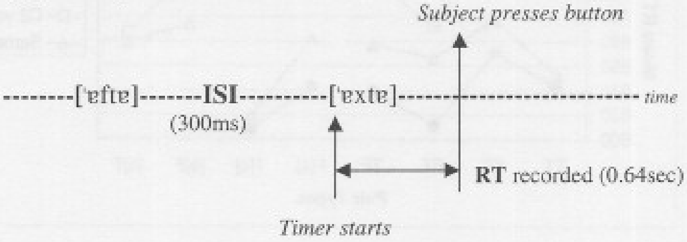
6.1.2 Listeners

The listeners for this experiment were 17 native speakers of Modern Greek (9 males and 8 females), ranging in age from 21 to 63, and 20 native speakers of American English (12 females and 8 males) ranging in age from 18 to 31. Both the Greek and English listeners were from various dialect/accent backgrounds.

6.1.3 Procedure

Pairs were created using MEL® (Microcomputer Experimental Laboratory software) and 6 randomized blocks were generated containing all the pairs found in appendix 1. In each block, the pairs were constructed using the identical .wav sound files to eliminate any production variability. The participants were given an instruction sheet before the experiment informing them about the task and the fact that they were going to be listening to nonsense words and not real words. Listeners were presented the pairs over headphones (Nova 16, 8 ohms, stereo headphones), in a sound attenuated booth, at a comfortable listening level (~70dB) in the fashion schematized in (10):

(10) Stimuli presentation



In (10) the interrupted line represents time (moving from left to right), and the stimuli are presented serially, ['ɐfta] then ['ɐxtɐ] with an inter-stimulus interval (ISI) of 300 ms (note that time measurements are iconic in this diagram) and a Reaction Time (RT) recorded at 640 ms after the start of the second stimulus. Listeners were instructed to listen carefully to the stimuli and decide as accurately and as quickly as possible whether the two nonsense words that they heard were the same or different. They had a choice of two buttons, labeled SAME or DIFFERENT. If they thought that the second word was the same as the first word of the pair presented they pressed SAME. If they thought the first word was different, as is the case in (10), they were supposed to press DIFFERENT. Listeners could only hear the pairs once, and they needed to make their decision within 4 seconds after the presentation of the second word of the stimulus pair. After 4 seconds had passed, the program timed out and the next pair was presented with no reaction time recorded. If the listener identified the pair correctly, feedback was given on a computer screen located in the booth[12]. The feedback had the form: CORRECT ANSWER and the RT was shown in seconds, such as 0.64secs. If the listeners made a mistake, e.g. pressed SAME when the pairs were different, or DIFFERENT when the pairs were the same, a computer message WRONG ANSWER was flashed on the screen, no RT feedback was given and the program proceeded with the next pair.

7. Results.

7.1 RT measurements

Average RTs for the different types of clusters for the Greek listeners are shown in figure 1. These results are calculated over the correct identifications, i.e. when the pairs were different the listeners responded by pressing the DIFFERENT button, and the correct rejections, i.e. when the pairs were the same the SAME button was pressed.

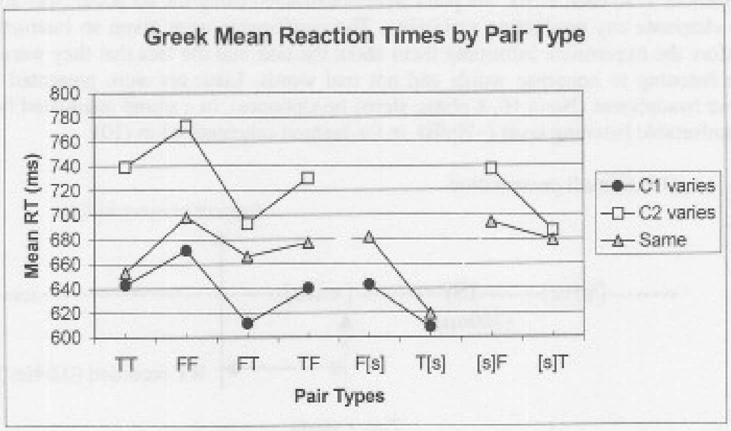


Figure 1. Mean RTs of correct identifications and correct rejections for Greek listeners.

The different cluster types (T=stop, F=fricative, [s] sibilant) are shown on the abscissa, and the average reaction times are shown in the ordinate. Grey triangles represent the "same" trials, for example [ˈɐftɐ]~[ˈɐftɐ], solid circles represent pairs that varied in C_1 having C_2 constant, for example [ˈɐftɐ]~[ˈɐxtɐ], and clear squares represent pairs that varied in C_2 having C_1 constant, for example [ˈɐftɐ]~[ˈɐfpɐ]. Standard deviations and percentages of errors are given in appendix 2. The results for the English listeners are shown in figure 2.

---

[12] Feedback was given in order to keep the alertness level of the listeners high during a rather long and repetitious discrimination task.

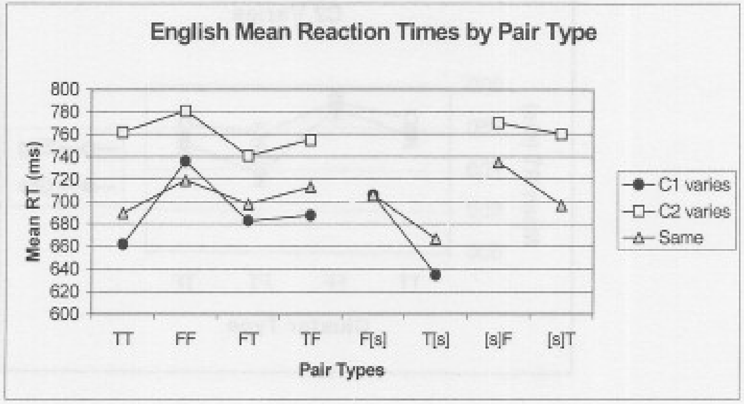


Figure 2. Mean RTs of correct identifications and correct rejections for English listeners.

If we compare the performance between the two groups of listeners within similar cluster types, that is, when C1 varied, as opposed to when C2 varied or when both pairs were the same we get the patterns shown in figures 3 through 5.

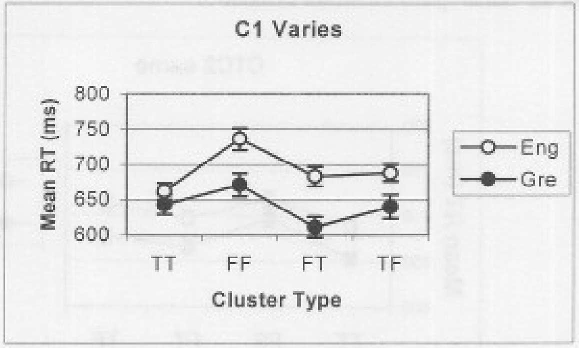


Figure 3. Mean RTs of correct identifications when C1 varied (eg. apta–akta) (T=stop F=Fricative).

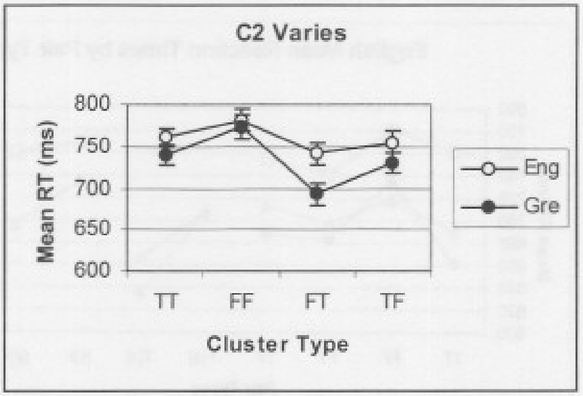


Figure 4. Mean RTs of correct identifications when C2 varied (eg. apta~apka).

In the graphs above we see that both English and Greek listeners pattern similarly showing the slowest reaction times for clusters with two fricatives (FF), for example [ɐfθɐ]~[ɐxθɐ] and the fastest for fricative-stop sequences (FT). However, English listeners showed a preference for (TT) clusters when C1 varied, but when C2 varied, as shown in figure 4, both groups of listeners pattern exactly the same in all pair types[13]. Data for the "same" pairs are shown in figure 5.

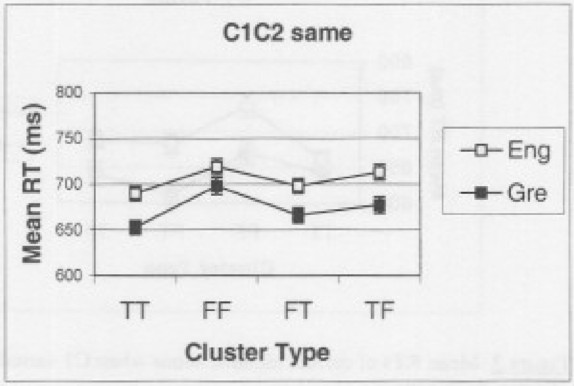


Figure 5. Mean RTs for correct identifications when the clusters were identical (e.g. apta~apta)

---

[13] Also in figure 4 (when C2 varied), RTs are overall slower than when C1 varied (figure 3) since listeners had to wait for a uniqueness point in the cluster that came later in the signal (C2 position).

Again, we see that both Greek and English listeners pattern similarly. The slowest reaction times are recorded for (FF) clusters and the fastest for (TT) clusters. However, it is interesting that both groups maintained better performance when discriminating between fricative plus stop (FT) clusters rather than stop plus fricative (TF) clusters.

7.2 Error Rates

If we look at the error rates we also see a similar overall pattern, between the two groups of listeners but also some interesting differences:

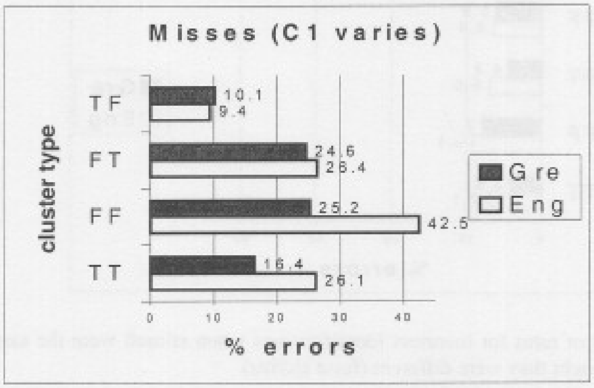


Figure 6. Error rates for incorrect identifications when stimuli were different (varied in C1 position) and listeners thought they were the same (misses).

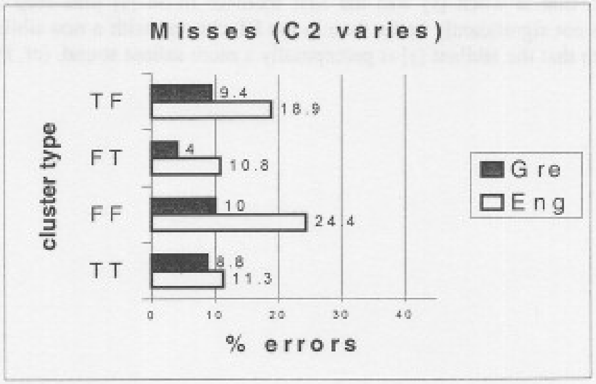


Figure 7. Error rates for incorrect identifications when stimuli were different (varied in C2 position) and listeners thought they were the same (misses).

Both Greek and English listeners made the most mistakes when discriminating between fricative plus fricative clusters. It is also noticeable how error rates decreased overall when it was the second consonant C2 that varied in the cluster. When C1C2 were identical, listeners had the fewest errors as shown in figure 8. Generally, faster reaction times correspond to fewer errors (higher accuracy).

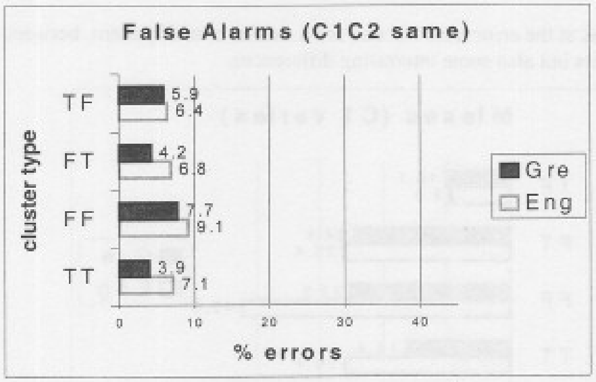


Figure 8. Error rates for incorrect identifications when stimuli were the same but listeners thought they were different (false alarms).

If we turn our attention now to the sibilant [s], we saw in figures 1 and 2 that reaction times were overall faster than in the non sibilant clusters and so the data in figure 9 show that accuracy with sibilant clusters was also better. However, it is interesting to note one exception, that is when [s] was the first member in an [s] plus stop cluster reaction times were not significantly better than in the FT clusters with a non sibilant [s] contrary to the claim that the sibilant [s] is perceptually a more salient sound. (cf. Figures 1, 2 and 9 below).

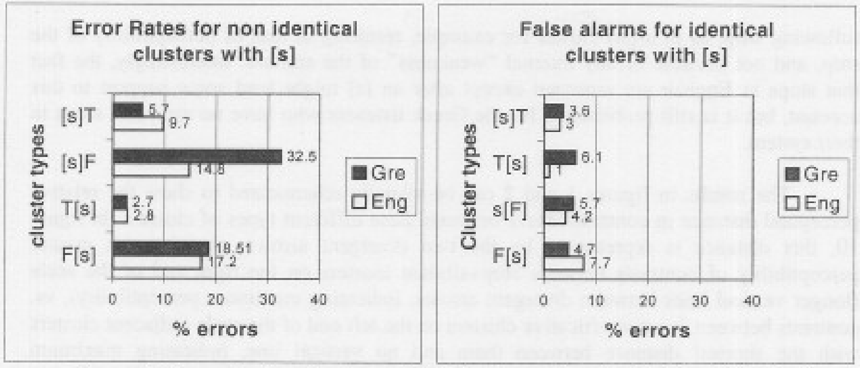


Figure 9. Error rates for clusters with [s]. In the graph to the left F and T varied, either in second C2 or first C1 position (e.g. aspa~asta or apsa~atsa). In the graph on the right the clusters were identical (e.g asta~asta or apsa~apsa).

8. Discussion

The two groups of listeners show similar trends in their average reaction times but also some very interesting differences. Most notably, the Greek listeners were faster in every category [F=14.1, p<.05], verifying the native language effect. Also, they exhibited their fastest reaction times in discriminating between dissimilated pairs of the form Fricative + Stop for non sibilant clusters (showing a significant effect of cluster type on RT, [F=32.4, p<.05]) and Stop + Fricative for clusters with the sibilant [s] [F=6.1, p<.05], exactly as predicted by the directionality preference of the dissimilation process described in section 2, and thus verifying the prediction that fricative plus stop clusters are more perceptible than stop plus fricative. When compared with the results for the English speakers, the Greek results seem to show a very strong preference for dissimilated clusters, something that is not very clear for the English listeners, who seem to be better at discriminating stop+stop clusters when C_1 varies, but show the same pattern with the Greeks (yet not as robust) when C_2 varies. In the case of the clusters with the sibilant [s], both groups of listeners seem to be noticeably better in discriminating stop + [s] clusters than fricative + [s] clusters, but again the pattern is not as robust when C_2 varies in [s]+fricative vs. [s]+stop clusters, even though the trend towards preferring the dissimilated clusters is maintained.

Another unexpected pattern exhibited by both listeners is that of a rather sluggish reaction time average in discriminating between [s]+stop clusters (687ms for the Greeks, 760ms for the Americans) vs. the fricative+stop clusters (693ms for the Greeks, 741ms for the Americans), given the fact that [s]+stop clusters are frequent in both languages. If the directional asymmetry exhibited by clusters containing [s] in Greek as shown in (6) earlier is due to the intrinsic salience of [s] (which as discussed earlier has a distinct high-pitched turbulent noise) resisting modification because of it being more perceptible than [f, θ, x], then this result is problematic for such an account. Perhaps this pattern is due to some masking effect that this high-pitched turbulent noise can have on the cues of the following stop, as in a [st] cluster for example, resulting in loss of perceptibility of the stop, and not because of any internal "weakness" of the sibilant. Interestingly, the fact that stops in English are aspirated except after an [s] might lend some support to this account, but it is still problematic for the Greek listeners who have no aspirated stops in their system.

The results in figures 1 and 2 can be visually schematized to show the relative perceptual distance in contrastiveness between these different types of clusters. In figure 10, this distance is represented by the two divergent arrows, showing the greater perceptibility of contrasts between stop+sibilant clusters on the right end of the scale (longer vertical lines between divergent arrows, indicating maximum perceptibility), vs. contrasts between fricative+fricative clusters on the left end of the scale (adjacent clusters with the shortest distance between them and no vertical line, indicating maximum confusability). The mean reaction times in figure 3 are averaged across the different directionality conditions, to best show the dissimilatory end points, even though in the space in between, some ambiguities, even reversals of the perceptibility rankings were found in the data, espcially TT>FT but not TF for the English listeners.

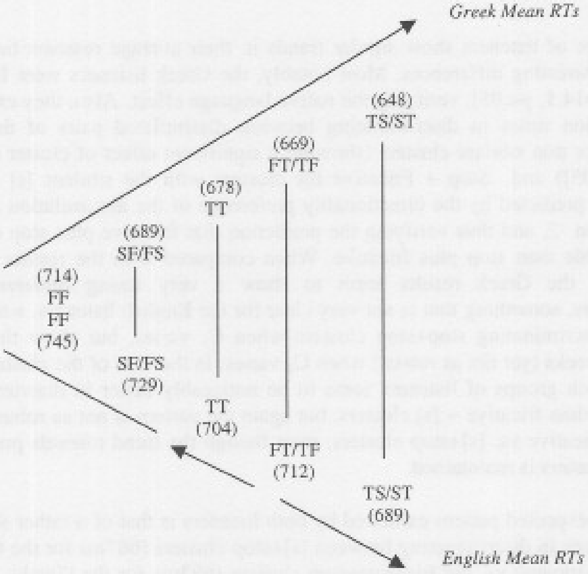


Figure 10. Perceptual salience schematized in terms of abstract distance.

If we were to rank the relevant clusters in terms of perceptibility based on reaction times, and maintain the different directionality data obtained by this experiment, we increase the resolution of the perceptibility hierarchies in (8) provided by Greenberg, so as to reflect the favored segmental sequences in terms of perceptibility by both Greek and English listeners:

(11) revised rankings (T= Stop, F=Fricative, S=Sibilant) – here ">" means "is more easily perceived than".

Greeks: FT > TT > TF > FF cluster type
(656) (678) (682) (714) RT

English: TT > FT > TF > FF cluster type
(704) (707) (718) (745) RT

If we include the special consideration of the clusters containing the sibilant [s] we then get the paired rankings shown in (12):

(12) Sibilant cluster perceptibility rankings:

- Greeks: TS > TF (608) (640) ST > FT (687) (693) FS > FF (643) (671) SF > FF (737) (773)
- English TS > TF (635) (640) **ST < FT** **(760) (741)** FS > FF (706) (736) SF > FF (770) (781)

Notice that the unexpected ranking in ST < FT (bold type) for English listeners is paralleled by the marginal ST > FT for Greek listeners (only 6ms difference in average RT).

In terms of the acoustics of these clusters, the fact that Greek listeners showed a preference for prevocalic stops in the form of FT vs. TF clusters, can be attributed to the fact that a stop before a vowel as in [ˈɐftɐ] is acoustically more robust, having more cues before a vowel (burst and vowel transitions in the onset CV of the vowel) than before a fricative, as is the case in [ˈɐtfɐ]. The worst position then would be before a stop as in [ˈɐtpɐ], when the [t] may lack a burst all together and have offset (VC) transitions in the vowel, something that has been found to be perceptually worse (Fujimura et al, 1978). This phonetic enhancement that takes place by having a voiceless stop prevocalically rather than preconsonantally could also be a factor in the course of development of dissimilatory processes in a language. In addition, the fact that fricatives have internal cues, but stops have bursts could be another factor influencing the formation of sequences such as [ft] where, as discussed earlier in section 5, an optimization takes place, fricative and stops cues maximized, vs. [tf] where a stop cue (burst) is lost in fricative noise. Also the fact that the sibilant [s] seems to pattern differently, resisting dissimilation to a [-continuant] when found in clusters with other fricatives, also lends credence to a claim of better perceptibility of highpitched turbulent noise.

In terms of the phonology of these clusters in Greek and English, we can see the language effect, in the average reaction times. Greek listeners were a lot faster in all categories even in the ones that were preferred by English listeners such as the stop+stop clusters when C_1 varied. The fact that the English phoneme inventory lacks a contrast in the fricative series, namely the velar fricative [x], could also be a relevant factor in hindering the English listeners' ability to discriminate between clusters containing fricatives. This effect was also found in another experiment (Hume et al. 1999) with English listeners that were asked to identify Korean stop bursts, and did worse in the task than Korean listeners, who have an additional manner contrast in the stop series (aspirated, unaspirated and tense) vs. the two way contrast of English stops (unaspirated, and aspirated). Also the fact that both groups of listeners had the slowest RTs for fricative plus fricative clusters, may reflect both the rarity of such clusters in both languages as compared to those of stops and stops plus fricatives, with the Greek speakers actively decreasing their occurrence even more, by dissimilating [fθ] to [ft], and [xs] to [ks] to take two random examples. Furthermore, the finding of a preferred dissimilation output, namely fricative+stop over stop+fricative highlights the necessity of specifying directionality in OCP accounts of dissimilation which would otherwise fail to be informative as to which configuration is more favored cross-linguistically. Also the fact that in Greek these clusters can be found both intervocalically and in absolute word initial positions (complex onsets) as in /ptero/ 'feather' or /xtima/ 'land field' could also be a source of greater ability to discriminate for the Greek listeners as opposed to the English ones who have experience with more restricted phonotactics in their language.

Finally, in terms of the sociolinguistic ramifications of dissimilation, evidence from this experiment suggests that a post-diglossic phonological fusion, and dialect/style mixture in general, can be the source of a great variability in the speech signal. Dissimilation in Greek is an active process only probabilistically and it is severely constrained by as yet formally undefined factors. The perception results obtained from this experiment, can then be used as a starting point for better generalizations in terms of which factors seem to be more relevant in the description and analysis of phonological phenomena as they are paralleled by gross acoustic salience (contextual and inherent), and are manipulated by speakers and perceived by listeners in regular but not monotonic patterns, something that was shown by the limited yet identifiable differences in perception of consonant clusters between these two groups of listeners. Inherent variability in a phonological system, then, can be directly correlated with social attitudes towards such variability, that could sustain it or eliminate it, with perceptibility being just one parameter that can be overridden at any given point, either in historical development as it unfolds over time, or in highly idiosyncratic and evanescent individual speech events.

9. Conclusion

In this paper, the process of manner dissimilation in Greek was investigated, and empirical results from a perception experiment were presented in support of the thesis that perceptual considerations, in addition to other factors, can influence the phonology of a language. Dissimilation can be used as a diagnostic process for understanding the limits and mechanics of the role of perception in phonology. The experimental results in particular showed that listeners belonging to two distinct speech communities exhibited similar, but not identical discriminatory capabilities when presented with an identical number of contrasts. The differential perception of these contrasts as found here, is by no means claimed to be what causes the dissimilation process in Greek, only that it influences it in directions that take overall perceptual salience into consideration; the different parameters and influential factors are weighted and evaluated, resulting only in tendencies that can potentially be verified cross-linguistically. The results presented here constitute such an attempt, and are perhaps a small contribution to some possible answers to the questions generated by trying to understand and describe the complex structure of phonological systems and of language in general. Further laboratory studies of perceptual salience are needed to determine the limits of interaction between universal tendencies and language specific phonological constructs. In particular, future studies varying the linguistic source of contrasts used as stimuli, Greek in this case, and applying the same experimental design to different linguistic populations that possess separate and distinct phonological systems and especially segmental sequencing constraints, are needed. Empirical evidence like this then could help formulate more informed theories and possibly even predictions about the possible avenues of language change and language variation. By investigating the actual language users as they are going into various discriminative states, even in a laboratory setting, I believe is a good way to arrive at some understanding of the interactions between learned habits of varying complexity, and innate, species specific, predispositions.

Acknowledgements

I would like to thank Beth Hume, Keith Johnson and Brian Joseph for their helpful comments and suggestions. I would also like to thank the members of the Ohio State Hellenic Student Association who volunteered as listeners for the discrimination experiment. Many thanks for helpful comments also go to my colleagues in Phonies and Psychoids, as well as to the participants of the sixth Mid-Continental Workshop on Phonology held in Columbus, OH, October 20-22, 2000 where an early version of this paper was presented.

REFERENCES

Clements, G. N. 1988. The role of the sonority cycle in Core syllabification. In J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the grammar and physics of speech.* Cambridge University Press (pp. 283-333).

Ferguson, C. 1959. Diglossia. *Word* 15: 325-40.

Fujimura, O., Macchi, M. J. and Streeter, L. A. 1978. Perception of stop consonants with conflicting transitional cues: a cross-linguistic study. *Language and Speech* 21:337-346.

Greenberg, J.H. 1978. Some generalizations concerning initial and final consonant clusters. In Greenberg, J.H. (ed.) *Universals in Human languages*, vol 2: *Phonology*. Stanford, CA: Stanford University Press (pp. 243-280).

Guy, G. and Boberg, C. 1997. Inherent variability and the obligatory contour principle. *Language Variation and Change* 9: 149-164.

Hatzigeorgiu N., Gavrilidou M., Piperidis S., Carayannis G., Papakostopoulou A., Spiliotopoulou A., Vacalopoulou A., Labropoulou P., Mantzari E., Papageorgiou H., Demiros I. (2000) Design and implementation of the online ILSP Greek Corpus. In *Proceedings of the LREC 2000 Conference, Athens 2000*, pp. 1737-1742.

Hume, E. and Johnson, K. 2001. A model of the interplay of speech perception in phonology. In E. Hume and K. Johnson (eds.) *The Role of Speech Perception in Phonology*. New York: Academic Press (pp. 3-26).

Hume, E., Johnson, K., Seo, M., Tserdanelis, G., Swinters, S. 1999. A cross linguistics study of stop place perception. *Proceedings of the XIVth International Congress of Phonetic Sciences*. 2069-2072.

Hura, S.L., Lindblom B. & Diehl R. L. 1992. On the role of perception in shaping phonological assimilation rules. *Language and Speech* 35: 59-72.

Johnson, K. 1997. *Acoustic and Auditory Phonetics*. Oxford. Blackwell Publishers.

Joseph B. and Philippaki-Warburton, I., 1987 *Modern Greek*, London, Croom Helm

Kazazis, K. 1992. Sunday Greek Revisited. *Journal of Modern Greek Studies*. Vol 10. No 1. 57-69

Kohler, K. 1990. Segmental Reduction in connected speech: Phonological facts and phonetic explanations. In W. J. Hardcastle & A. Marchal (eds.) *Speech Production and Speech Modeling*. Dordrecht: Kluwer Academic Publishers (pp. 69-92).

Macridge, 1985 *The Modern Greek Language*. Clarendon Press. Oxford

McCarthy, J. 1986 OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17: 207-263.

Newton, B., 1972. *The Generative Interpretation of a Dialect*, Cambridge, Cambridge University Press.

Odden, D., 1987. Dissimilation as Deletion in Chukchi. In A. Miller and J. Power (eds.), *Proceedings from ESCOL 3* (pp. 235-246). Columbus, OH: Ohio State University.

Ohala, J. 1990. The phonetics and Phonology of aspects of assimilation. In J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology*, vol. 1 Cambridge: Cambridge University Press (pp 258-275).

Ohala, J. 1992. Alternatives to the Sonority Hierarchy for explaining Segmental sequential constraints. *Chicago Linguistics Society: Papers from the Parasession on the Syllable*. Chicago: CLS. 319-338.

Pagoni-Tetlow, S. 1998. Breaking up is (not) hard to do: the case of the modern Greek *pt/kt* sequences. In *Themes in Greek Linguistics*. Joseph B., Horrocks, G. & Philippaki-Warburton I. (eds.). John Benjamins Publishing Company. Philadelphia.

Shepard, R. N. 1987. Toward a universal law of generalization for psychological science. *Science* 237. 1317-1323.

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*. New York: Academic Press (pp. 219-250).

Suzuki, K., 1998. *A Typological Investigation of Dissimilation*. Doctoral dissertation, University of Arizona.

Tserdanelis, G., 2000. A quantitative approach to manner dissimilation in Greek biconsonantal cluster. ms. The Ohio State University.

Wright, R. (1996) *Consonant Clusters and Cue Preservation in Tsou*. PhD dissertation. UCLA.

Yip, M. (1988). The Obligatory Contour Principle and Phonological Rules: A loss of identity. *Linguistic Inquiry* 19: 65-100

APPENDIX 1: Constructed pairs of stimuli.

	TT	*FF*	*FS/SF*	*TF/FT*	*ST/TS*
2nd member constant in pair	pt_kt tk_pk tp_kp	θf_xf fθ_xθ θx_fx	fs_xs θs_xs θs_fs	ft_xt θp_xp θk_fk px_tx kθ_pθ tf_kf	ks_ps ts_ks ts_ps
foils	pt_pt kt_kt tk_tk	θf_θf θx_θx fx_fx	fs_fs xs_xs θs_θs	ft_ft xt_xt xp_xp θp_θp θk_θk fk_fk	ks_ks ps_ps ts_ts
1st member constant in pair	pk_pt tk_tp kp_kt	θx_θf xθ_xf fθ_fx	sf_sx sθ_sf sθ_sx	tf_tx px_pθ kf_kθ θp_θk xt_xp ft_fk	sk_sp st_sp st_sk
foils	pk_pk tp_tp kp_kp	xθ_xθ fθ_fθ xf_xf	sf_sf sx_sx sθ_sθ	tf_tf tx_tx px_px pθ_pθ kθ_kθ kf_kf	sk_sk sp_sp st_st

APPENDIX 2: Mean reaction times, standard deviations and percentages of errors.

NON-DISSIMILATED CLUSTERS

	Stops		
	Tt–Tt	tT–tT	TT–TT
	Greek		
Mean RT	738	643	653
SD	175	220	179
% Errors	8.8	16.4	3.9
	English		
Mean RT	761	662	690
SD	263	193	204
% Errors	11.3	26.1	7.1

	Fricatives		
	Ff–Ff	fF–fF	FF–FF
	Greek		
Mean RT	773	671	698
SD	225	244	232
% Errors	10	25.2	7.7
	English		
Mean RT	**781**	736	719
SD	214	221	223
% Errors	24.4	42.5	9.1

T=Stop, F=Fricative
S=Sibilant

A Capital T indicates that the stop was constant in C_1C_2 cluster. A small case t indicates that the stop varied. for example a pair like Tt–Tt stands for a pair of actual stimuli: [ˈɐpcɐ]–[ˈɐpkɐ].

DISSIMILATED CLUSTERS

	Fricative + Stop		
	Ft–Ft	fT–fT	FT–FT
	Greek		
Mean RT	693	611	666
SD	219	220	195
% Errors	4.0	24.6	4.2
	English		
Mean RT	741	683	698
SD	220	222	205
% Errors	10.8	26.4	6.8

	Stop + Fricative		
	Tf–Tf	tF–tF	TF–TF
	Greek		
Mean RT	730	640	677
SD	205	253	211
% Errors	9.4	10.1	5.9
	English		
Mean RT	755	688	713
SD	219	214	224
% Errors	18.9	9.4	6.4

SIBILANT CLUSTERS

	Fricative + /s/	
	fS~fS	FS~FS
	Greek	
Mean RT	643	682
SD	220	196
% Errors	18.51	4.7
	English	
Mean RT	706	706
SD	222	209
% Errors	17.2	7.7

	/s/ + Fricative	
	Sf~Sf	SF~SF
	Greek	
Mean RT	737	694
SD	185	241
% Errors	14.8	5.7
	English	
Mean RT	770	735
SD	174	239
% Errors	32.5	4.2

	Stop+ /s/	
	tS~tS	TS~TS
	Greek	
Mean RT	**608**	619
SD	198	179
% Errors	2.7	1
	English	
Mean RT	635	667
SD	199	200
% Errors	2.8	6.1

	/s/ + Stop	
	St~St	ST~ST
	Greek	
Mean RT	687	679
SD	164	236
% Errors	5.7	3
	English	
Mean RT	760	697
SD	203	258
% Errors	9.7	3.6