[To appear in *The Role of Speech Perception in Phonology*, edited by Elizabeth Hume & Keith Johnson, Academic Press.]

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

Elizabeth Hume Keith Johnson
Ohio State University

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 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

[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 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.

> 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 1996, 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 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., this volume). 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. (this volume) 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.

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

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 apical [s] versus laminal [s**],** 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.' (Aquilina, 1959; Hume, 1996). Since the perceptual cues to word-initial geminates, stops especially, are relatively weak (see, e.g. Abramson, 1987; Muller, 2000), 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, 2000). 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; see Seo & Hume 2000). Hume (1998, 2000) 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

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

interconsonantal position. Winters (2000) also found evidence of a perceptibility gain for patterns of stop/stop metathesis observed cross-linguistically in VCCV sequences (see also Steriade, this volume).

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, 2000). In Seo’s (2000) 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, this volume; Winters 2000.

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 2000 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 (2000) 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 that this process is a case of perceptually tolerated articulatory simplification (Hura et al., 1992; Kohler, 1990; Steriade, this volume). 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).

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

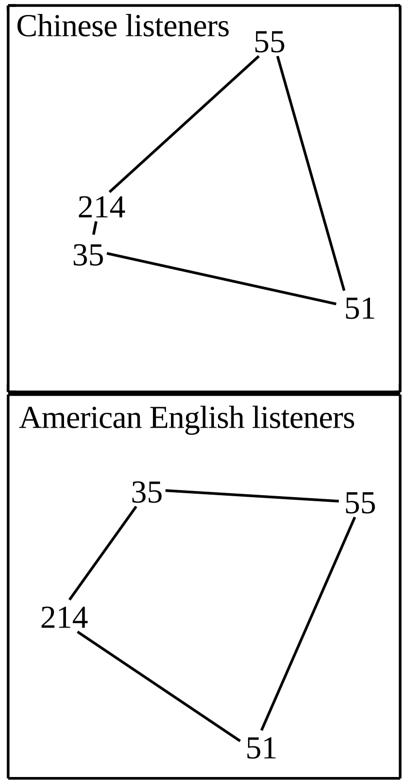

Figure 1. The four tones of Mandarin Chinese in perceptual space for Mandarin Chinese listeners and American English listeners. Multidimensional scaling data from Huang (2000).

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.

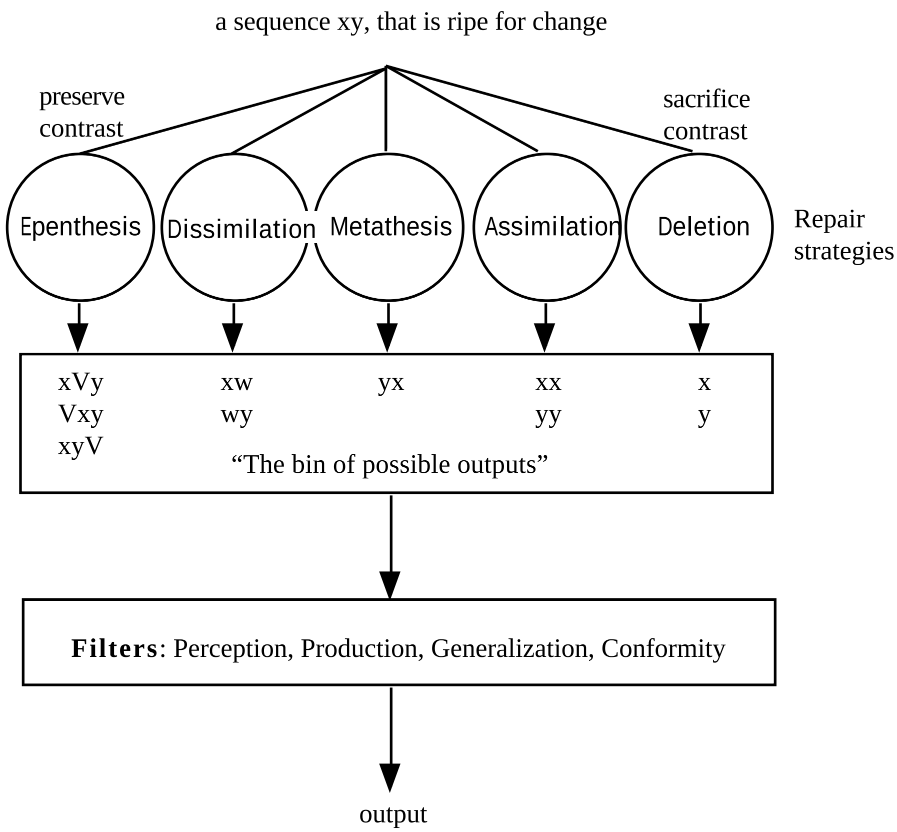

Figure 2. Characterization of phonological repair strategies, and the role of filters in selecting among possible outputs.

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.

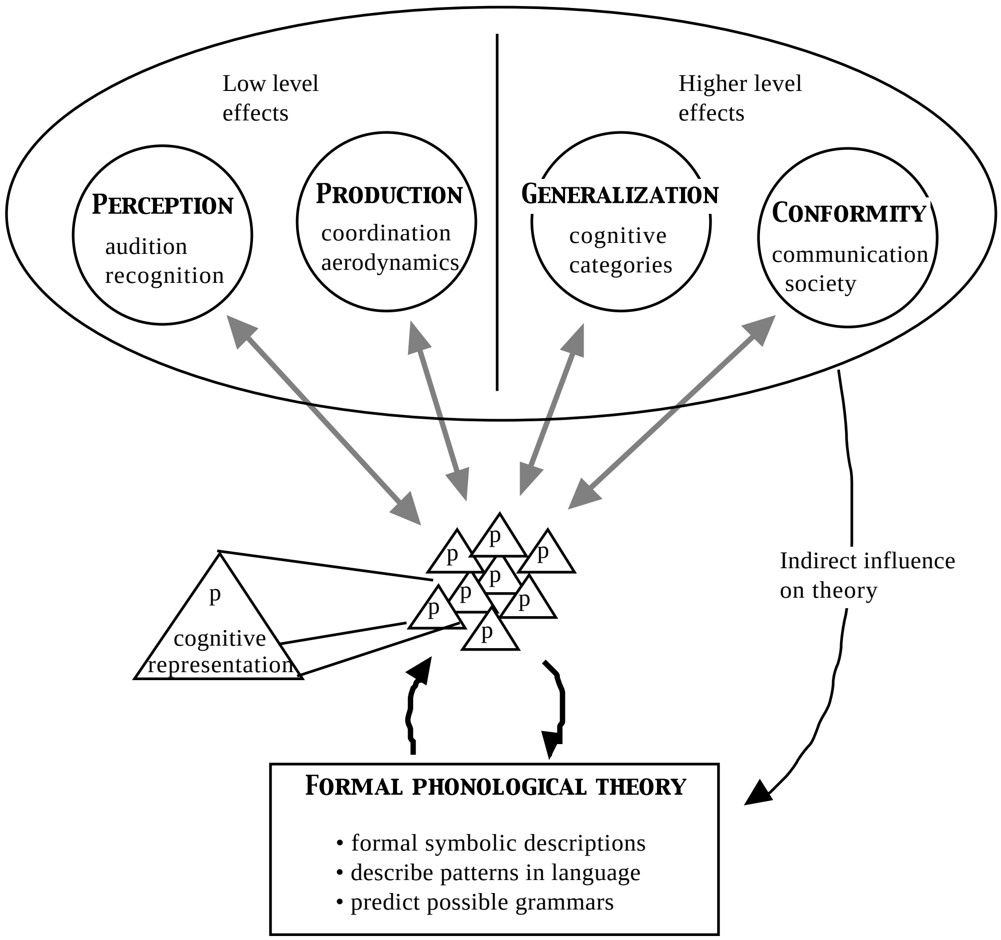

Figure 3. A general model of the interplay of external forces and phonology, broadly defined.

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, this volume). 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 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.

---

[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.

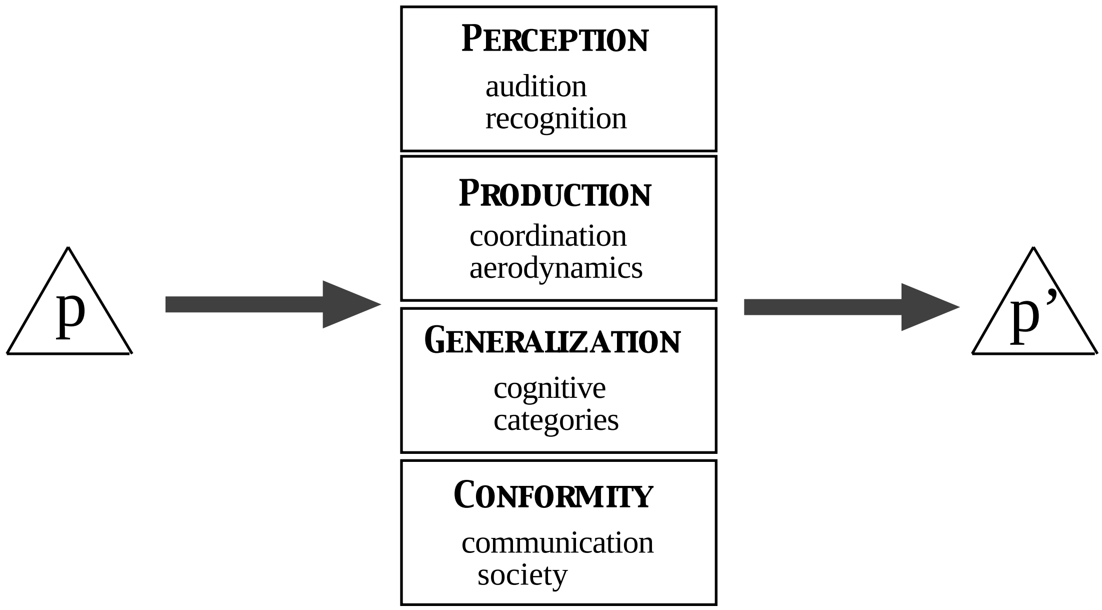

Figure 4. The mapping of *p* onto *p'* can be decomposed into a set of filters. Each component of the mapping process independently influences the relationship between *p* and *p'* and, hence, the structure of *p'*.

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.

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 apply over centuries are at work in daily language use as well.

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.

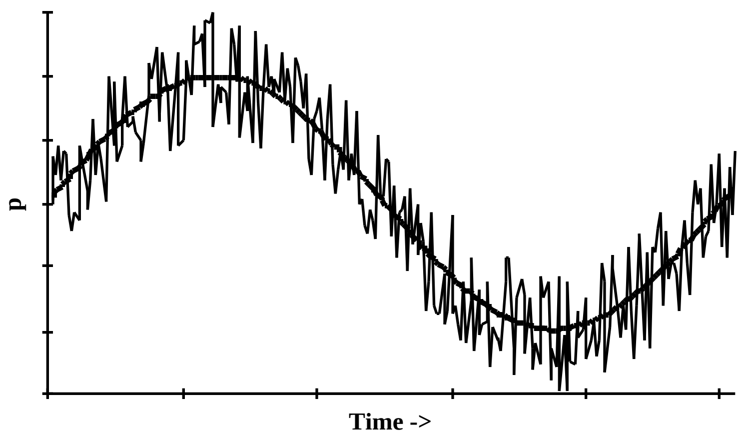

Figure 5. A coarse-grained time scale shows general tendencies, illustrated by the slowly changing line, while a fine-grained time scale shows rapidly fluctuating change. Time in this illustration is on the horizontal axis, and the vertical axis is meant to show, in an abstract one dimensional projection,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

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 (2000) 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.

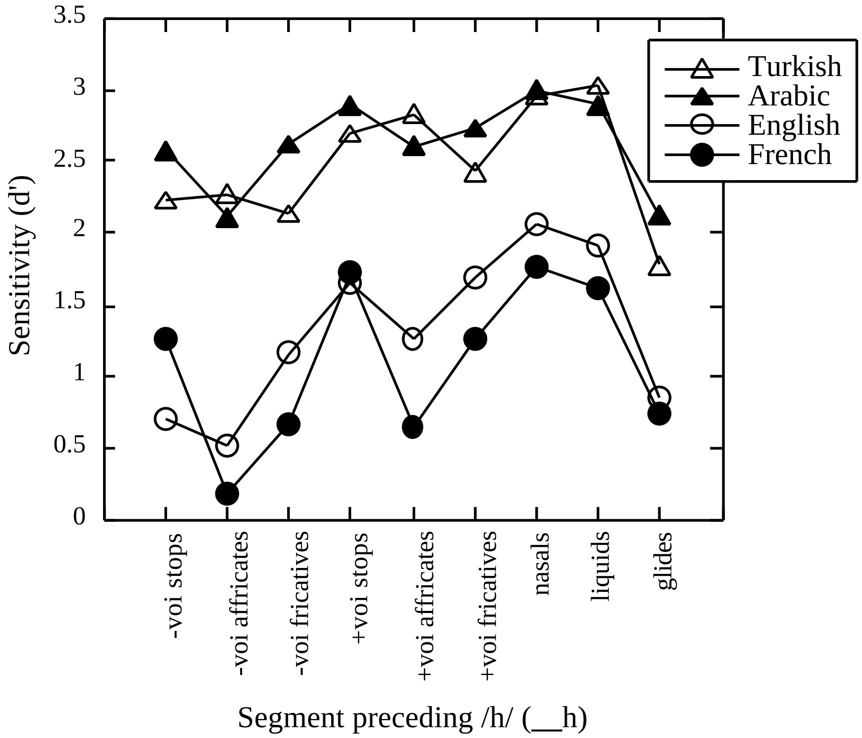

Figure 6. Perceptual sensitivity to /h/ in different segmental contexts by listeners of American English, Turkish, Arabic, & French. Data from Mielke (2000).

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.
Aquilina, Joseph. (1955) *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. (2000) Patterns of perceptual compensation and their phonological consequences. In E. Hume & K. Johnson (eds.), *The Role of Speech Perception in Phonology*. 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.
Chomsky, N. & Halle, M. (1968) *The Sound Pattern of English*. New York: Harper & Row.
Clements, 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*. 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*. 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. (1996) Phonetically driven phonology: The role of optimality theory and inductive grounding. *Proceedings of the 1996 Milwaukee Conference on Formalism and Functionalism in Linguistics*.
Huang. T. (2000) 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. (2000) Metathesis: Data, motivation & phonological theory. In E. Hume, N. Smith & J. van de Weijer (2000).
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.) (2000) *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 Midwest*. 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. (2000) A Perceptual Account of Turkish h-Deletion. *The Interplay of Speech Perception and Phonology*, OSUWPL, vol. 55.
Muller, J. (2000) The Phonetics and Phonology of Initial Geminates. PhD dissertation, Ohio State University (in progress).
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. Connel & A. Arvaniti (eds), *Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence*. 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. (2000) The realization of l/n sequences in Korean. *The Interplay of Speech Perception and Phonology*, OSUWPL, vol. 55.
Seo, M. & Hume, E. (2000) 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. (2000) Directional asymmetries in assimilation: A directional account. In E. Hume & Johnson, K. (eds.), *The Role of Speech Perception in Phonology*. New York: Academic Press.
Sweet, H. (1877) *A Handbook of Phonetics*. Oxford: Clarendon Press.
Tserdanelis, G. (2000) Manner dissimilation directionality in Greek. *The Interplay of Speech Perception and Phonology*, OSUWPL, vol. 55.

Winters, S. (2000) VCCV perception: Putting place in its place. *The Interplay of Speech Perception and Phonology*, OSUWPL, vol. 55. (Also published in E. Hume, N. Smith & J. van de Weijer (2000))

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