The role of probabilistic enhancement in phonologization*

James Kirby

University of Edinburgh

j.kirby@ed.ac.uk

1 Introduction

PHONOLOGIZATION – the process by which intrinsic phonetic variation gives rise to extrinsic phonological encoding – is often invoked to explain the acquisition and transmission of sound patterns (Jakobson 1931; Hyman 1976; Ohala 1981; Blevins 2004). A familiar example is the idea that lexical tone contrasts can trace their origins to the pitch perturbations conditioned by differences in obstruent voicing (Matisoff 1973; Hombert *et al.* 1979). A phonologization account of tonogenesis is sketched in Table 1. First, intrinsic differences in vowel F0 (Stage I) become a perceptual cue to the identity of the initial consonant (Stage II). If other cues to the contrast between initial consonants are lost, the contrast may be maintained solely by differences in F0 (Stage III), setting the stage for a reanalysis of pitch as a contrastive phonological feature.

This process can be observed *in vivo* in Seoul Korean, a language which maintains a 3-way phonological contrast between initial stops (Table 2). While studies of Korean stop acoustics conducted during the 1960s and 1970s found this contrast to be signaled primarily by differences in voice onset time (VOT: Lisker and Abramson 1964; Kim 1965;

*To appear in A. Yu (ed.), *Origins of sound change: Approaches to phonologization* (Oxford University Press). Draft 30/3/2012. Not for quotation or copying.

Stage I	*Stage II*	*Stage III*
pá [‾]	pá [‾]	pá [‾]
bá [/]	bă [⌣]	pă [⌣]

Table 1: Stages of phonologization (after Hyman, 1976). Sparklines show time course of F0 production.

Han and Weizman 1970), subsequent studies have reported that lenis and aspirated stops are no longer distinguished solely by VOT in either production or perception, but rather that F0 has come to play a more central role (Kim *et al.* 2002; Silva 2006*a*,*b*; Wright 2007; Kang and Guion 2008). One way to describe this change is as the phonologization of previously intrinsic, mechanical phonetic variation, conditioned here by initial obstruent voicing.

manner	*Hangul*	*1960s*	*2000s*	*gloss*
fortis	뿔	ppul	púl	'horn'
lenis	불	pul	p^hùl	'fire'
aspirated	풀	p^hul	p^húl	'grass'

Table 2: Phonologization of F0 in Seoul Korean.

While the phonologization model provides a useful descriptive framework for this type of sound change, it also raises several new questions. First, while it is known that multiple acoustic-phonetic cues are available to signal any given phonological contrast (Lisker 1986), there has been relatively little discussion of how and why certain cues are targeted for phonologization. In Seoul Korean, for instance, it has been established that, in addition to VOT and F0, spectral tilt, the amplitude of the release burst, and amplitude of the release burst are relevant perceptual cues to the initial onset contrast (Cho *et al.* 2002; Kim *et al.* 2002; Wright 2007). So why was F0, and not some other cue, phonologized in this case?

A related issue is Hyman's (1976) observation that the phonologization of one cue often entails *de*phonologization of another, a process sometimes referred to as TRANSPHONOLO-

GIZATION (Hagège and Haudricourt 1978). In the case of Seoul Korean, as F0 has become an increasingly important acoustic correlate of the contrast between lenis and aspirated stops, VOT has become correspondingly less informative. Given that contrasts are almost always redundantly cued, this shift is somewhat unexpected. Why might cause an increase in the informativeness of one cue to be accompanied by a decrease in the informativeness of another?

This paper proposes to answer these questions by arguing that phonologization is an emergent consequence of adaptive enhancement in speech (Lindblom 1990; Diehl 2008). In particular, it is proposed that as contrast precision is reduced, cues are enhanced to compensate. The degree of enhancement is argued to be a probabilistic function of contrast precision, while the probability with which a given cue is enhanced is related directly to its informativeness, the degree to which it contributes to accurate identification of a speech sound (what Hume, this volume, refers to as CUE QUALITY). To explore this hypothesis, phonetic categories are modeled as finite mixtures (Nearey and Hogan 1986; Toscano and McMurray 2010), and a case study – the phonologization of F0 in Seoul Korean – is explored in detail through the use of agent-based computational simulations. The results suggest that both probabilistic enhancement and loss of contrast precision interact to drive the process of phonologization.

The remainder of this chapter is structured as follows. Section 2 reviews the roles of the speaker and listener in sound change and motivates an adaptive notion of enhancement. Section 3 discuss the mixture model of phonetic categories, and Section 4 describes the algorithm used to simulate speaker-hearer interaction. These are used to explore the phonologization of F0 in Seoul Korean in Section 5. The results and implications are discussed in Section 6, and Section 7 provides a general conclusion.

2 Bias and enhancement in sound change

Even under relatively ideal conditions, successful speech communication is a challenge. Along with contextual and coarticulatory effects, a range of physiological, social, and cog-

nitive BIAS FACTORS can introduce variability into the acoustic realization, potentially obscuring the speaker's intended message. Garrett and Johnson (this volume) provide a thorough overview of such factors, which include details of motor planning, aerodynamic constraints, and the effects of gestural overlap and perceptual hypercorrection. Moreover, cognitive-selectional biases favoring the transmission of certain sound patterns and speaker-specific social and indexical characteristics may introduce additional asymmetric variability into the speech signal (Wilson 2006; Moreton 2008; Yu, this volume).

What is most important to note for present purposes is that, regardless of their source, different types of bias factors may have a similar effect: namely, they reduce the precision with which the phonetic category intended by the speaker is accurately identified by the listener. In this chapter, the term PRECISION will be used to refer to the accuracy with which a listener can distinguish between members of a phonetic contrast, and then term BIAS will be used specifically to refer to factors which *reduce* this precision. One example of this type of bias is the aerodynamic voicing constraint (Ohala 1997), which conditions a loss of precision between voiced and voiceless stop categories; the neutralization of place cues by high front vowels conditioning asymmetric misperception of [ki] > [ti] is another (Chang *et al.* 2001).

In the context of the case study examined in §5, the bias in question involves hypoarticulation of a phonetic cue, but it is worth abstracting away from details of particular bias factors in order to ask how speakers and hearers might respond to loss of precision more generally. Researchers such as Ohala (1981 *et seq*) often assume, tacitly or otherwise, that speakers produce phonetic targets more or less as they are intended (modulo contextual effects such as coarticulation). The response to a loss of precision may then be a reanalysis on the part of the listener. For example, on this view, phonologization of a cue such as F0 might come about due to listeners' failure to compensate for the intrinsic perturbation effects of an initial consonant on the pitch contour of the following vocalic segment. After these effects have been phonologized, the initial conditioning environment (here, obstruent voicing), now a redundant cue to the contrast, is free to dephonologize.

However, it is not clear what motivates this dephonologization, given that phonetic distinctions are rarely signalled by a single cue. It is also not immediately clear why listeners would fail to compensate for intrinsic variation along one dimension but not another.

A different account is suggested by more broadly functional approaches to sound change, which hypothesize a more active role for the speaker (Liljencrants and Lindblom 1972; Kingston and Diehl 1994; Boersma 1998). A common theme in these treatments is the idea that the acoustic realization of a phonetic target may be modulated both by TALKER-ORIENTED constraints enforcing efficiency in speech communication ('be efficient') as well as LISTENER-ORIENTED constraints requiring speech sounds to be sufficiently distinctive ('be understood'). Talker-oriented constraints are often implemented by penalizing gestures in terms of the energy or precision required for their realization. Listener-oriented constraints are usually implemented in such a way as to maximize distinctiveness between contrasts, although this takes on a variety of forms: combining articulatory gestures which have mutually reinforcing acoustic consequences (Kingston and Diehl 1994), adding redundant features or secondary gestures to reinforce contrast perception (Stevens and Keyser 1989; Keyser and Stevens 2006), encoding a preference for accuracy in the approximation of phonetic targets (Lindblom 1990; Johnson *et al.* 1993; Boersma 1998), or imposing systemic constraints to maximize the distance between contrasts (Liljencrants and Lindblom 1972; Flemming 2002).

A common thread in all of these treatments is the notion of enhancement of phonetic targets. In this chapter, the term ENHANCEMENT will be used specifically to refer to those actions take on the part of the speaker which *increase* the precision of a phonetic contrast. For example, a talker might enhance the contrast between two initial obstruent categories by producing them with hyperarticulated VOT values, or by reducing the variability in their productions of those values. These notions of enhancement and precision will be more rigorously formalized in §3 and §4 below.

Functional approaches predict enhancement to be more likely in situations where it would improve intelligibility for the listener. This suggests at least a partial explanation

for why any particular phonetic property might be phonologized: all else being equal, cues which more reliably signal a difference between categories are more likely to be enhanced. However, it is still not clear why phonologization should be accompanied by dephonologization. Why should the promotion of intrinsic variance to extrinsic indicator of contrast be accompanied by the reverse process?

The answer advanced here is that phonologization is itself a response to loss of contrast precision. If cues are enhanced as a probabilistic function of the current contrast precision (measured as the classification accuracy of the listener) and cue informativeness (measured as a function of their reliability), this means that more informative cues are more likely to be enhanced than less informative cues, and cues will be enhanced to a greater extent when categorization error is high than when it is low (the PROBABILISTIC ENHANCEMENT HYPOTHESIS). Viewed in this way, phonologization can be understood as an emergent consequence of the interaction between bias and enhancement in speech communication.

3 A mixture model of phonetic categories

It order to evaluate this proposal, the notions of precision, informativeness and enhancement must be rigorously quantified. To this end, it is useful to consider a representational scheme for phonetic categories that encodes the multidimensional variability inherent in the speech signal. One formal representation meeting this description is a FINITE MIXTURE MODEL (McLachlan and Peel 2000), which models a statistical distribution as a weighted sum (or mixture) of other distributions. Mixture models have a long history in speech research and have been used in work on speech perception (Lisker and Abramson 1970; Nearey and Hogan 1986; Pierrehumbert 2001; Clayards 2008), the perceptual integration of acoustic cues (McMurray *et al.* 2009; Toscano and McMurray 2010), and the unsupervised induction of phonetic category structure (de Boer and Kuhl 2003; Vallabha *et al.* 2007; Feldman *et al.* 2009).

Following previous researchers, it is assumed that the underlying probability distributions of the mixture components (i.e. the cue dimensions) are normal (Gaussian). In

a GAUSSIAN MIXTURE MODEL (GMM), an observation vector $\mathbf{x} = \{x_1, \ldots, x_D\}$ assumed to be independently generated by an underlying distribution with a probability density function

$$f(\boldsymbol{x}; \theta) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{1}$$

where the structure $\theta = ((\pi_1, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \ldots, (\pi_K, \boldsymbol{\mu}_K, \boldsymbol{\Sigma}_K))$ contains the component weights π, mean vectors $\boldsymbol{\mu}$, and covariance matrices $\boldsymbol{\Sigma}$ of the D-variate component Gaussian densities $\mathcal{N}_1, \ldots, \mathcal{N}_K$. Figure 1A shows how these three parameters describe a given mixture component.

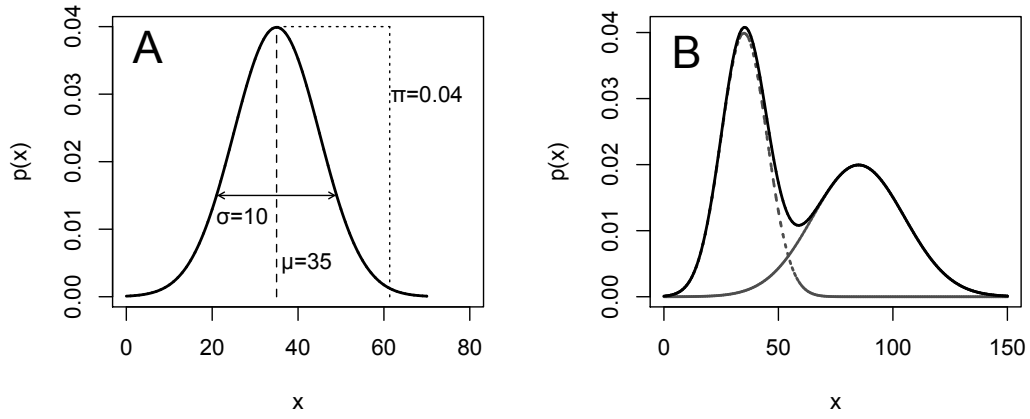

Figure 1: (A) Parameters of a Gaussian distribution for a single component (adapted from McMurray et al., 2009). (B) Two class-conditional Gaussians (dotted grey lines) and their mixture (solid black line).

To make this more concrete, think of $\mathbf{x}$ as a bundle of cue values representing an instance of phonetic category c; of D as representing the number of cue dimensions $(m_1, m_2, \ldots, m_D)$ relevant to the perception of that category; and of K as representing the total number of category labels $(c_1, c_2, \ldots, c_K)$ competing over the region of phonetic space defined by D. For example, for a language like Korean with three initial stops

($K = 3$) cued along five dimensions ($D = 5$), we might have c_1 = /p/, c_2 = /pp/, c_3 = /p$^{\text{h}}$/ and m_1 = VOT, m_2 = burst amplitude, m_3 = F0, m_4 = spectral tilt, and m_5 = following vowel length. A given observation $\mathbf{x}$ will thus consist of five elements, each one providing a value for one of these cues.

Figure 1B illustrates a GMM where $K = 2$ and $D = 1$. The individual component densities are shown in gray, while the mixture density is outlined in black. Although more difficult to visualize, the mixture modeling approach extends straightforwardly to the multivariate case where $D > 1$.

In the GMMs for phonetic categories used in this chapter, experience forms the basis for both production and perception. The speaker's task is to produce an instance of a phonetic category; this may be modeled by sampling cue values from the relevant class-conditional mixture component $\mathcal{N}_k$. The listener's task is to assign this utterance a category label c. If we assume that listeners weight information in the speech signal by its quality (informativeness), we can construct a model of their behavior that would optimize this task. Such models are sometimes referred to as IDEAL OBSERVER models (Geisler 2003; Clayards 2008). The following section provides a brief overview; for a more in-depth treatment, see Clayards (2008) or Kirby (2010).

3.1 The ideal observer

In order to come to a decision about whether or not a given utterance $\mathbf{x} = \{x_1, \ldots, x_D\}$ is a member of category c, the ideal observer requires access to two sources of information: $p(c)$ (the prior probability of the category c) and $p(\mathbf{x}|c)$ (the probability of the observation, given that it is a member of category c). These probabilities may be estimated from the statistical distributions of speech cues (Maye *et al.* 2002; Clayards *et al.* 2008). The probability that the speaker intended an instance of category c given the evidence that cue m_d takes on value x can then be evaluated using Bayes' rule, as shown in (2).

$$p(c|x_d) = \frac{p(x_d|c)p(c)}{\sum_{k=1}^{K} p(x_d|c_k)p(c_k)} \tag{2}$$

If contrasts are represented in a high-dimensional space, posterior probabilities can still be computed using (2), but are instead conditional on the entire utterance vector, i.e. $p(c|\mathbf{x})$. As D increases, however, the number of observations required to obtain robust parameter estimates begins to grow quickly. Under the assumption that cues are conditionally independent (Clayards 2008; Toscano and McMurray 2010), the probability that an utterance $\mathbf{x}$ bears category label c is simply the product of the conditional probabilities $p(x_1|c), p(x_2|c), \ldots, p(x_D|c)$ normalized over all K categories competing over the D-dimensional phonetic space, as shown in (3).

$$P(c|x_1, \ldots, x_D) = \frac{p(x_1|c)p(x_2|c), \ldots, p(x_D|c)p(c)}{\sum_{k=1}^{K} p(x_1|c_k)p(x_2|c_i), \ldots, p(x_D|c_k)p(c_k)} \tag{3}$$

3.2 Cue informativeness

The ideal observer model predicts that listeners should make use of the probability distribution of all cues when attempting to identify a speaker's intended utterance. The existence of multiple cues to phonetic categories does not, however, imply their equivalence: some cues provide more information about the perceptual identity of a sound than do others. The informativeness of a cue can be approximated as its statistical reliability, although other factors may also contribute (Holt and Lotto 2006). Intuitively, the less distributional overlap between two categories, the more informative the cue in determining the perceptual identity of an input.

Figure 2 illustrates this concept along a single cue dimension. The solid lines in Figure 2A show the distribution for two categories with little overlap along cue m, while the dotted lines show the distribution for two categories with more overlap. The categorization functions in Figure 2B show the probability of categorizing a stimulus as c_1 given the value of m, computed using Equation 3. Note that while the value of m for which the probability of the stimulus belonging to either category c_1 or c_2 is the same (i.e. the point on the y-axis where the function crosses 0.5), the slope of the functions differs, reflecting increased uncertainty in the case of the dotted distributions in Figure 2A. In other words,

cue m is more informative in distinguishing between the solid distributions than it is in distinguishing between the dotted distributions.

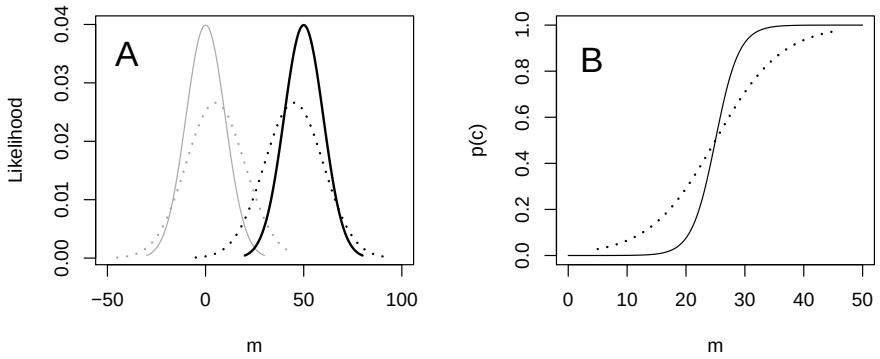

Figure 2: (A) Probability distributions of a cue dimension m for two categories c_1 (dark lines) and c_2 (light lines). Solid lines show a mixture where there is little overlap between the components, dashed lines a mixture with more overlap. (B) Optimal categorization functions given the distributions in (A). (Adapted from Clayards et al., 2008.)

While reliability of a cue can be expressed as an identification function, it is also useful to have an index of a cue's informativeness relative to other cues. One way to accomplish this is based on the detection-theoretic d' statistic (Green and Swets 1966), the absolute value of the difference in category means divided by the average variance:

$$d'(m) = \frac{(\mu_m|c_1 - \mu_m|c_2)^2}{(\sigma_m^2|c_1 + \sigma_m^2|c_2)/2} \tag{4}$$

The informativeness ω_m for an individual cue can then be expressed as

$$\omega_m = \frac{d'(m)}{\sum_{m \in M} d'(m)} \tag{5}$$

3.3 Categorization and contrast precision

Equation 3 allows the listener to compute the probability of category membership, but it does not determine how such information should be used to assign a category label. The approach taken here is to assign utterances a category label proportional to their relative strength of group membership (Nearey and Hogan 1986). For example, an utterance which has probability 0.9 of belonging to category c_1 and probability 0.1 of belonging to category c_2 will be assigned label c_1 90% of the time, and label c_2 10% of the time. However, the statistically optimal classifier – the model which maximizes classification accuracy – assigns the category label with the highest maximum *a posteriori* probability. To continue with the previous example, an utterance which has probability 0.9 of belonging to category c_1 and probability 0.1 of belonging to category c_2 will always be assigned label c_1 by the optimal classifier.

Although optimal classifiers make strong assumptions and their predictions are not always in line with human classification behavior (Ashby and Maddox 1993), they provide a lower bound on the error rate that can be obtained for a given classification problem. In this work, contrast precision ϵ is defined as the current error rate of the optimal classifier for that contrast, i.e.

$$\epsilon = 1 - \sum_{k=1}^{K} \int p(\mathbf{x}|k)p(k)d\mathbf{x} \tag{6}$$

4 Modeling probabilistic enhancement

The previous section has provided an overview of how speech production and perception can be modeled in a probabilistic mixture model framework, allowing for the quantification of the notions of contrast precision and cue informativeness. This section explores how the hypothesis of probabilistic enhancement can be tested using computational simulation.

In §2, enhancement was informally described as any action taken by the speaker to increase contrast precision. In light of the previous discussion, we can now begin to give a more precise definition: if contrast precision is defined in terms of statistical reliability, enhancing a cue means affecting an *increase in informativeness* along that cue dimension. If the probabilistic enhancement hypothesis is correct, then the targeting of cues for enhancement should be to some extent predictable based on their informativeness.

One way to explore the predictions of this hypothesis is through the use of computational simulation. The framework described here is broadly exemplar-based, in that it tracks the production and perception of individual utterances, but it differs from previous models in several ways. In treatments such as Pierrehumbert (2001) or Wedel (2006), agents map speech tokens onto a granular similarity space based on the token's similarity to a stored exemplar prototype; exemplars which fall between the cracks of this space are then encoded as identical. Thus, a stored exemplar need not correspond to a unique perceptual experience *per se*, but rather to an 'equivalence class' of perceptual experiences. The present implementation differs slightly in that exemplars are used to estimate the parameters of the cue distributions relevant for some phonetic contrast. Instead of being mapped to prototypes, experienced tokens are stored together with decay weights, which are used to determine when an exemplar should be deleted from the list of tokens associated with a given category label. Once the decay weight of a token falls below a user-defined threshold, it is deleted from the list and is no longer referenced during parameter estimation. When simulating speech production, values for each cue are simply sampled from each conditional density in the usual fashion. In this way, the same exemplar list may be referenced in both the production and perception of phonetic categories. A more detailed discussion of the framework described below can be found in Kirby (2010).

4.1 Architecture

Simulations are run for a fixed number of iterations. Each agent is characterized by a lexicon, a set of exemplar lists $\mathcal{E}_1, \ldots, \mathcal{E}_K$ corresponding to their experience with phonetic

categories $c_1, \ldots, c_K$. Before the simulation begins, these lists are populated by sampling from the conditional densities of a GMM representation of each category. For simplicity, here we consider agents with lexica containing just two categories.

Subsequently, each iteration consists of a single interaction between two agents, one acting as speaker and the other as listener (the framework can also be extended to accommodate more than two agents). Each iteration contains four steps: production, enhancement, bias, and categorization. All agents use the same production and categorization strategies described in §3. However, the strength of bias and the degree of enhancement can be altered by manipulating two tuning parameters:

1. a vector $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_D\}$, encoding the strength of the phonetic bias affecting each cue dimension; and

2. a constant $\beta \in [0, 1]$ representing the functional load or system-wide importance of the contrast (Martinet 1952; Hockett 1955).

Each iteration then proceeds through the following steps:

1. **Production.** In the production phase, the talker agent selects a target category c_k based on the mixture weights π_k, and samples a series of values $x_1, \ldots, x_D$ from the conditional densities $\mathcal{N}_d(x|k, \theta)$ to form a PRODUCTION TARGET $\mathbf{x} = (x_1, \ldots, x_D)^T$.

2. **Enhancement.** Enhancement contains two sub-steps: first, determining *if* a cue will be enhanced, and second, determining *which* cue is enhanced. The probability that any particular dimension m_d will be enhanced is an exponential function of the current contrast ϵ and the functional load constant $\beta \in [0, 1]$. The likelihood of enhancement at any iteration is inversely proportional to the contrast precision (ϵ) scaled by the importance of the contrast (β), i.e. $P(\text{enhance}) = \epsilon^{\beta}$.

   In the event that an utterance is selected for enhancement during a given iteration, each cue has its distributionally-defined informativeness ω_d chance of being enhanced in that iteration (see §3.2). Once a specific cue has been targeted for enhancement, its production target value x_d is modified by sampling from a modified

distribution with an exaggerated mean and a reduced variance, thereby potentially increasing the statistical reliability of the dimension. The *degree* to which the mean value is increased and variance reduced is attenuated by the precision and functional load of the contrast as well as by the informativeness of the cue dimension selected (see Kirby 2010 for details). The end result is that more reliable cues are more likely to be produced with extreme (hyperarticulated) values than less reliable cues, and cues will be enhanced to a greater extent when error (ϵ) is high and β is low (i.e., functional load is high).

3. **Bias.** Next, the (potentially enhanced) production target is modified along one or more cue dimensions by adding the bias vector $\boldsymbol{\lambda}$. In order to ensure that cue values stay within a well-defined range, each bias term λ_d may be scaled relative to the distance between category means before being applied, approaching 0 when the means become identical (i.e. when the dimension is no longer informative in distinguishing the contrast).

4. **Categorization.** Finally, the modified production target $\mathbf{x}'$ is presented to the listener agent for classification, who assigns it a category label as described in §3.1. Once labeled, $\mathbf{x}'$ is added to the appropriate exemplar list. Both agents then recompute the memory decay weights for each exemplar in their lexicon, and delete exemplars whose weights have fallen below the decay threshold. In the next iteration, the role of speaker is assumed by the listener agent and vice versa.

In summary, the architecture provides two tuneable parameters ($\boldsymbol{\lambda}$ and β) corresponding to phonetic bias and functional load, respectively. Varying these parameters allows us to explore the effects of probabilistic enhancement in different scenarios, and to see what parameter values best approximate observed data patterns. In the following section, the probabilistic enhancement hypothesis is explored in this framework using empirical data from the phonologization of F0 in Seoul Korean.

5 Transphonologization in Seoul Korean

Armed with the computational framework described above, it is now possible to test the probabilistic enhancement hypothesis using empirical language data. Here, we consider the case of the phonologization of F0 in Seoul Korean described in §1. Apparent time studies suggest that while the distinction between lenis and aspirated stops in Seoul Korean of the 1960s was mainly cued by VOT, this distinction is now cued chiefly by F0 at the onset of the following vowel and has been accompanied by a loss of contrast along the VOT dimension (Silva 2006*a*,*b*; Kang and Guion 2008). This is a classic instance of transphonologization, where reduction of informativeness along one cue dimension is accompanied by enhancement of a previously redundant dimension. The goal of these simulations was to determine if these shifts in the distribution of cues could be replicated without making specific reference to F0 as a target of enhancement.

The proposal advanced here holds that phonologization is driven by loss of contrast precision, and there exists considerable evidence for a systemic production bias affecting VOT in Seoul Korean (Silva 1992, 1993, 2006*a*). In particular, lenis /p t k/ and aspirated /p^h t^h k^h/ stops tend to be produced with similar VOT in initial position. On Silva's analysis, fortis stops would not be subject to this same bias, since they are phonologically geminate (2006a:303). Since this proposed bias factor would not have affected the production of fortis stops, the following discussion is limited to the contrast between lenis and aspirated stops for expository clarity.

The simulations described here considered five cues which have been argued to be relevant for the perception of the Korean stop contrast: voice onset time (VOT), F0 and duration of the following vowel (VLEN), the difference in amplitude between the first two formants of the vowel (H_1-H_2), and the amplitude of the burst (BA). Data on each of these cues reported in Cho *et al.* (2002), Kim *et al.* (2002), Silva (2006*b*), and Kang and Guion (2008) were used to seed the initial exemplar lists of two ideal observer agents with a simple lexicon consisting of just two syllables, lenis /pa/ and aspirated /p^ha/. This state corresponds to the cue distributions reported for Seoul Korean speakers in the

1960s. The initial parameters and their corresponding informativeness values are shown in Table 3; two-dimensional scatterplots showing the joint distributions of VOT and each of the cues are shown in the first row of Figure 3. The second row of Figure 3 shows distributions based on the parameters shown in the second half Table 3, estimated based on the speech of younger speakers gathered in the 2000s. It is to these distributions that the state of the agents will be compared at the end of each simulation run. In other words, we want to see under what circumstances the agents' states will evolve from the top row of Figure 3 to the bottom row.

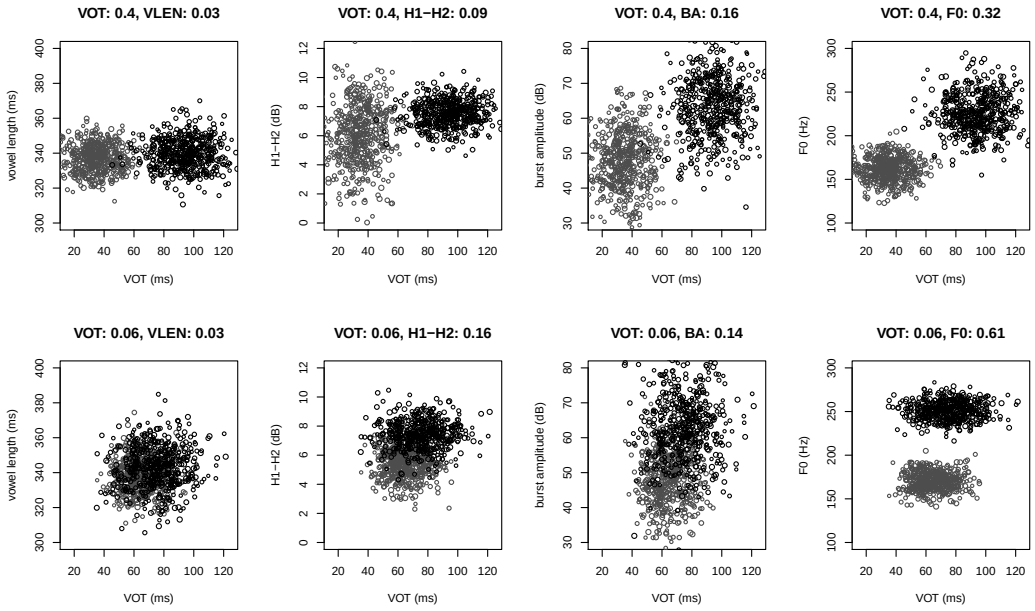

Figure 3: Row 1: distribution of five cues to the laryngeal contrast in Korean (gray = lenis /pa/, black = aspirated /p^ha/) used to seed the simulations, based on the speech recorded in the 1960s. Row 2: distribution of the same cues based on the speech recorded in the 2000s. Data estimated from Cho *et al.* (2002); Kim *et al.* (2002); Silva (2006*b*); Kang and Guion (2008). Captions give cue informativeness as computed by Equation (5). VOT = voice onset time (in ms); VLEN = vowel length (in ms); H_1-H_2 = spectral tilt (in dB); BA = burst amplitude (in dB); F0 (in Hz).

Three series of simulations are reported, each seeded with the same initial configuration. The first round of simulations considered the effects of applying probabilistic enhancement in the absence of phonetic bias (§5.1); the second considered the effect of applying phonetic bias to the production of a single cue, but without enhancement (§5.2); and the third explored the effects of applying both enhancement and bias (§5.3).

The simulations reported here are representative runs of 25,000 iterations, at which point the statistical reliability of the cue targeted by the bias factor and/or the probability of enhancement approached zero. Goodness of fit between the target distributions and the results of the various simulations was quantified by the KULLBACK-LEIBLER (KL) DIVERGENCE (Kullback and Leibler 1951) between each target and simulated cue dimension. A non-symmetric measure of the dissimilarity between two distributions, KL divergence equals zero when two distributions are identical, and grows with the dissimilarity between them.

5.1 Enhancement without bias

As can be seen in the top rows of Table 3 and Figure 3, it would appear that a contrast along the F0 dimension already existed in Seoul Korean of the 1960s, albeit covertly. One interpretation of the phonologization model is that active enhancement of cues on the part of speakers itself conditions the transition of a cue from covert to overt indicator of

	Category	VOT	VLEN	H_1-H_2	BA	F0
1960s	lenis	35 (11)	337 (8)	6 (2)	48 (8)	162 (14)
	aspirated	93 (15)	340 (15)	7.5 (1)	64 (9)	227 (21)
	ω	0.4	0.03	0.09	0.16	0.32
2000s	lenis	65 (11)	338 (10)	5.5 (1)	48 (8)	170 (10)
	aspirated	73 (15)	343 (12)	7.5 (1)	64 (9)	250 (11)
	ω	0.06	0.03	0.16	0.14	0.61

Table 3: Mean (s.d.) and informativeness ω of cues to Korean stops, estimated from data in Cho *et al.* (2002); Kim *et al.* (2002); Silva (2006*b*); Kang and Guion (2008). VOT = voice onset time (in ms); VLEN = vowel length (in ms); H_1-H_2 = spectral tilt (in dB); BA = burst amplitude (in dB); F0 (in Hz).

contrast. This interpretation may be tested by considering the application of probabilistic enhancement in the absence of any external bias. In this set of simulations, the β constant was set to 0, meaning that some cue was always enhanced at each timestep. Each element in the $\boldsymbol{\lambda}$ vector was also set to 0, meaning that no phonetic bias factors were applied.

The first row of Figure 4 shows the results of a representative simulation run using these parameter settings. In each case, the most informative cue at initialization (here, VOT) maintained its relative dominance throughout the simulation. The overall degree of enhancement was extremely small, reflecting the fact that the precision of the contrast is never in jeopardy, although as shown in Figure 5, the error rate does fluctuate somewhat over time. In short, these parameter settings result in few or no changes to the cue structure of the categories over time, demonstrating that probabilistic enhancement alone is insufficient to induce phonologization of a phonetic dimension along which categories may be only weakly separated. Furthermore, it shows that enhancement along one cue dimension does not in and of itself entail loss of contrast along another. This suggests that some other mechanism is necessary to drive the process of phonologization.

5.2 Bias without enhancement

The second set of simulations considered the inverse of the above interpretation. If two categories are redundantly (if perhaps weakly) distinguished along some cue dimension, it is possible that this cue will become more informative simply as a result of continuous application of systemic bias to a highly informative cue. To test this hypothesis, simulations were run in which the VOT element of the bias vector $\boldsymbol{\lambda}$ was computed dynamically as $|\log(\mu_{c_1} - \mu_{c_2})|$, a range of 0 to about 4ms. This had the effect that VOT values for category c_2 words (/p$^{\text{h}}$a/) were produced with slightly shorter VOTs at each timestep, while values for category c_1 words (/pa/) were produced with slightly longer VOTs. No cues were enhanced in these simulations, i.e. $P(\text{enhance})$ was set to 0.

The results of a representative simulation run are shown in the second row of Figure 4. As evidence both by the scatterplots as well as the ω values, VOT has ceased to

be informative in distinguishing this contrast; to the extent that a contrast between the two categories still exists, it is supported chiefly by a difference in F0 (row 2, panel 4). This differs slightly from the attested modern Korean situation (row 4) in that the actual parameters characterizing the distributions of F0 have not changed for either category: F0 has become the most informative cue simply because all other cues have become less informative. However, the empirical Korean data indicate that the F0 means for aspirated and lenis obstruents have shifted slightly away from one another, suggesting that they have been enhanced both in terms of a shift in means as well as a reduction in variance (compare rows 1 and 2 of Figure 3).

As shown in panel 2 of Figure 5, in the absence of any kind of enhancement, the precision of the contrast degrades steadily over time as bias is applied. These simulation results indicate that while a redundant or covert contrast may become exposed by a systemic production bias, at least in the present case, bias alone cannot account for the shifts in cue distributions that are empirically observed.

5.3 Bias and enhancement

The third and final series of simulations considered the effect of applying VOT bias while allowing for probabilistic enhancement of cues. Here, the β constant was arbitrarily fixed at 0.5, and the same dynamic VOT bias described in §5.2 was applied. Thus, while bias was applied at each iteration, the likelihood of enhancement covaried with contrast precision.

A representative agent state after 25,000 iterations is shown in the third row of Fig. 4. Of the three types of simulations run, these results most closely resemble the empirical data, as evidenced by the small KL divergences shown in Table 4 and the high ω value for F0 (compare rows 3 and 4 of Figure 4). While both spectral tilt and burst amplitude are somewhat more informative relative to their initial values, F0 is the most informative cue to the contrast. Crucially, the phonologization of F0 was an adaptive, probabilistic response to the continued application of a bias in the production of VOT, resulting in an

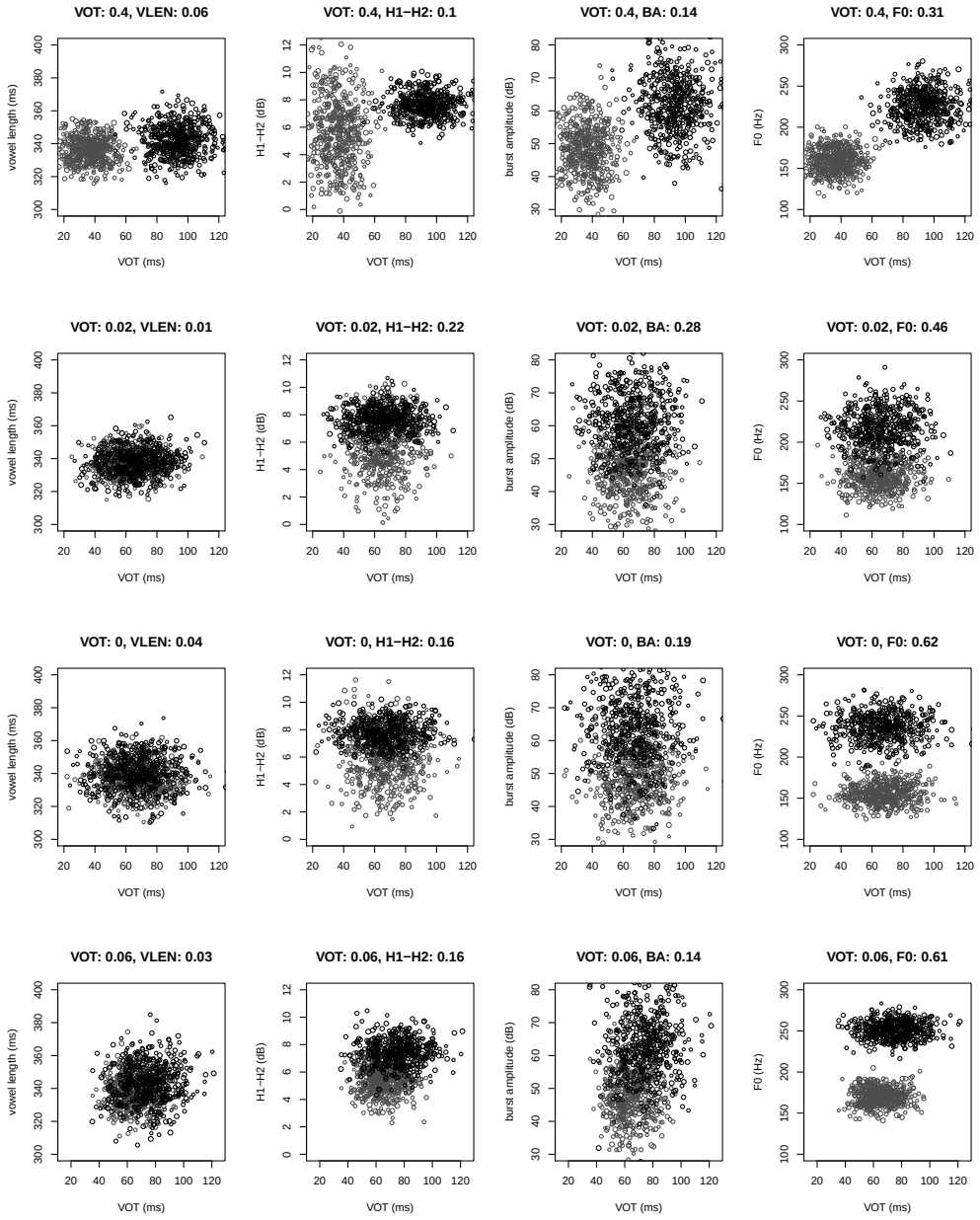

Figure 4: Cue distributions (gray = lenis /pa/, black = aspirated /p^ha/) after 25,000 iterations. Row 1: enhancement without bias. Row 2: bias without enhancement. Row 3: bias and enhancement. Row 4: empirical targets. Captions give cue informativeness as computed by Eq. (5). VOT = voice onset time (in ms); VLEN = vowel length (in ms); H_1-H_2 = spectral tilt (in dB); BA = burst amplitude (in dB).

Source	Category	VOT	VLEN	H_1-H_2	BA	F0
	lenis	36 (10)	336 (8)	5.6 (2.4)	48 (7.4)	159 (15)
enhancement	aspirated	92 (13)	342 (10)	7.6 (0.9)	62 (8.7)	225 (20)
only	ω	0.4	0.06	0.1	0.14	0.31
	KL	0.2	0.002	0.27	0.05	0.01
	lenis	65 (11)	340 (8)	6.3 (1.8)	48 (7)	162 (12)
bias only	aspirated	65 (16)	340 (9)	7.7 (0.9)	64 (8)	227 (20)
	ω	0	0	0.13	0.29	0.57
	KL	0.09	0.002	0.16	0.05	0.01
	lenis	66 (12)	338 (7)	4.7 (2.5)	49 (7.6)	152 (12)
bias +	aspirated	67 (19)	341 (10)	7.3 (0.9)	65 (9.6)	248 (17)
enhancement	ω	0	0.04	0.16	0.19	0.62
	KL	0.09	0.002	0.09	0.06	0.008
	lenis	65 (11)	338 (10)	5.5 (1)	48 (8)	170 (10)
target	aspirated	73 (15)	343 (12)	7.5 (1)	64 (9)	250 (11)
	ω	0.06	0.03	0.16	0.14	0.61
(cf. initial)	KL	0.16	0.002	0.12	0.06	0.008

Table 4: Comparison of mean (s.d.), cue informativeness, and KL divergence (in bits) for three simulation scenarios. VOT = voice onset time (in ms); VLEN = vowel length (in ms); BA = burst amplitude (in dB); H_1-H_2 = spectral tilt (in dB); F0 (in Hz).

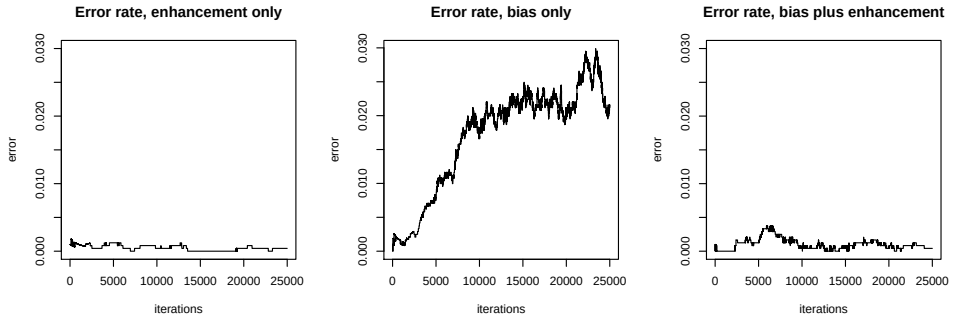

Figure 5: Comparison of contrast precision as measured by classification error rate at each simulation timestep for simulations reported in 5.1–5.3.

increasing loss of informativeness along that dimension. At no point was F0, or any other cue, specifically targeted for enhancement. As seen in panel 3 of Figure 5, while the error rate increased slightly in the early iterations of this simulation, it was quickly reduced by the countervailing force of probabilistic enhancement.

6 General discussion

The simulation results presented above demonstrate how phonologization may be predicted in a model where probabilistic enhancement is an adaptive response to a loss of contrast precision. This is not to say that phonologization must always be driven exclusively by loss of contrast precision, or that loss of precision will invariably result in phonologization; to be sure, there are cases in which bias leads to contrast neutralization (Kirby 2011). Nevertheless, these results indicate that at least some cases of phonologization may be the result of enhancement in response to a systemic production bias, and that both the presence of a redundant or covert contrast and the reduction of primary cues need to be present simultaneously in order for phonologization to take place.

As measured by KL divergence, the distributions resulting from the application of both enhancement and bias were most similar to the target Korean distributions compared with those resulting from the application of only enhancement or only bias. While the KL divergences reported in Table 4 are generally quite small, it is worth noting that the KL divergences between the initial and final (target) distributions are quite small as well. The KL divergences for various dimensions should thus not be interpreted in an absolute sense, but instead relative to other values for the same cue dimension.

It is important to note that it is not simply the presence of both bias and probabilistic enhancement that allow for accurate modeling of phonologization, but also in understanding how different parameter settings can give rise to different outcomes for simulations of differing lengths. This is precisely the strength of the the present account, which provides a framework in which to map out under what circumstances phonologization is more or less likely, given an empirical characterization of language-specific biases and cue distri-

butions. This model goes beyond the observation that a system biased against one cue will choose another by arguing that precisely *which* cue takes over can be predicted with some accuracy. In this formulation, the speaker plays an important role in sound change, enhancing phonetic cues in a fashion optimally suited to accommodate the communicative needs of listeners. In other words, the present model provides a principled explanation for why F0, and not H_1-H_2 or burst amplitude, was the cue which transphonologized in Seoul Korean.

However, depending on the distributional patterns and bias factors involved, the outcome could well be different for another contrast or another language. The results obtained in §5 are dependent on the initial state of the agents when the simulation begins, and similar results may not necessarily obtain for other initial states. In particular, if all cues are equally balanced in terms of their informativeness at the start of a simulation, then all will maintain their relative informativeness on this scheme if a constant bias is applied. Similarly, a strong bias (or low β) can overwhelm the probabilistic enhancement strategy, leading to neutralization even in cases where both bias and enhancement are applied.

The present model makes two assumptions which deserve further mention. The first is that all cues are conditionally independent in perception. While structure of the acoustic cues available to listener may be consistent with a linear model (Clayards 2008), this does not necessarily mean that they treated as such by listeners, as other factors such as task and saliency may play a role in determining how these dimensions are ultimately weighted (Holt and Lotto 2006; Toscano and McMurray 2010). To a certain extent, this assumption is orthogonal to the issues discussed in the present chapter, as probabilistic enhancement could just as easily be applied regardless of whether cue perception is represented by a linear or a multivariate model. However, the range of potential outcomes in a model which does not make this assumption has yet to be fully explored.

The second assumption is that any acoustic-phonetic dimension serving as a perceptual cue is amenable to enhancement in speech production. This is a somewhat stronger

version of the phonetic knowledge hypothesis than that originally proposed by Kingston and Diehl (1994), who argued that cues are enhanced based on the degree to which they contribute to the perception of an INTEGRATED PERCEPTUAL PROPERTY (IPP) which reinforces an existing phonological contrast. In the case of a voicing contrast for initial stops, for example, Kingston & Diehl would predict that cues with similar auditory properties, such as F1 and F0, would integrate, while cues such as closure duration and F0 would not, because they do not both contribute to the amount of low-frequency energy present near a stop consonant (Kingston *et al.* 2008). If cues are enhanced based on the degree to which they contribute to IPPs, this predicts that certain cues might not be enhanced regardless of their distributional informativeness in signaling a contrast. In contrast, the probabilistic enhancement predicts cues will be targeted based on informativeness and contrast precision, regardless of their relationship to IPPs. The different predictions made by these two theories awaits further experimental investigation.

7 Conclusion

This chapter has argued for the role of probabilistic enhancement in phonologization through computational simulation of an ongoing sound change in Seoul Korean. Two challenges faced by a phonologization model of sound change were addressed: determining how cues are selected, and explaining why phonologization is often accompanied by dephonologization. It was proposed that cues are targeted for enhancement as a probabilistic function of their informativeness, so a cue which may be targeted for enhancement in one language may be ignored in another. Simulation results using empirically derived cue values were presented, providing strong support for the idea that loss of contrast precision may drive the phonologization process. Depending on the distribution of cues, the interaction of phonetic bias and probabilistic enhancement can set the stage for a reorganization of the system of phonological contrasts.

Acknowledgements

Portions of this work have appeared previously in Kirby (2010). I would like to thank Bob Ladd, Bob McMurray, Morgan Sonderegger, Alan Yu, and Yuan Zhao for helpful comments and suggestions on previous versions of this manuscript.

References

Ashby, F. Gregory and Maddox, W. Todd (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, **37**, 372–400.

Blevins, Juliette (2004). *Evolutionary phonology*. Cambridge University Press, Cambridge.

Boersma, Paul (1998). *Functional phonology*. Ph.D. thesis, University of Amsterdam.

Chang, Steve S., Plauché, Madeline C., and Ohala, John J. (2001). Markendess and consonant confusion asymmetries. In *The role of perceptual phenomenona in phonological theory* (ed. K. Johnson and E. Hume), pp. 79–101. Academic Press, San Diego, CA.

Cho, Taehong, Jun, Sun-Ah, and Ladefoged, Peter (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics*, **30**, 193–228.

Clayards, Meghan (2008). *The ideal listener: Making optimal use of acoustic-phonetic cues for word recognition*. Ph.D. thesis, University of Rochester.

Clayards, Meghan, Tanenhaus, Michael K., Aslin, Richard, and Jacobs, Robert A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, **108**, 804–809.

de Boer, Bart and Kuhl, Patricia (2003). Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters On-line*, **4**(4), 129–134.

Diehl, Randy L. (2008). Acoustic and auditory phonetics: The adaptive design of speech sound systems. *Philosophical Transactions of the Royal Society*, **363**, 965–978.

Feldman, Naomi H., Griffiths, Thomas L., and Morgan, James L. (2009). Learning phonetic categories by learning a lexicon. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (ed. N. Taatgen and H. van Rijn), pp. 2208–2213. Cognitive Science Society, Austin, TX.

Flemming, Edward (2002). *Auditory representations in phonology*. Routledge, New York.

Geisler, Wilson S. (2003). Ideal observer analysis. In *The Visual Neurosciences* (ed. L. M. Chalupa and J. S. Werner), Volume 1, pp. 825–837. The MIT Press, Boston.

Green, David M. and Swets, John A. (1966). *Signal detection theory and psychophysics*. Wiley, New York.

Hagège, Claude and Haudricourt, André-Georges (1978). *La phonologie panchronique*. Presses Universitaires de France, Paris.

Han, Mieko S. and Weizman, Raymond S. (1970). Acoustic features of Korean /P, T, K/, /p, t, k/ and /ph, th, kh/. *Phonetica*, **22**, 112–128.

Hockett, Charles F. (1955). *A Manual of Phonology*. Volume 21-4, International Journal of American Linguistics. Indiana University Publications, Bloomington.

Holt, Lori L. and Lotto, Andrew J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquistion. *Journal of the Acoustical Society of America*, **119**, 3059–3071.

Hombert, Jean-Marie, Ohala, John J., and Ewan, William G. (1979). Phonetic explanations for the development of tones. *Language*, **55**(1), 37–58.

Hyman, Larry M. (1976). Phonologization. In *Linguistic studies presented to Joseph H. Greenberg* (ed. A. Juilland), pp. 407–418. Anma Libri, Saratoga.

Jakobson, Roman (1931). Prinzipien der historischen Phonologie. *Travaux du Cercle Linguistique de Prague*, **4**, 247–267.

Johnson, Keith, Flemming, Edward, and Wright, Richard (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language*, **69**, 505–528.

Kang, Kyoung-Ho and Guion, Susan G. (2008). Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *Journal of the Acoustical Society of America*, **124**(6), 3909–3917.

Keyser, Samuel J. and Stevens, Kenneth N. (2006). Enhancement and overlap in the speech chain. *Language*, **82**(1), 33–63.

Kim, Chin-Wu (1965). On the autonomy of the tensity feature in stop classification (with special reference to Korean stops). *Word*, **21**, 339–359.

Kim, Mi-Ryoung, Beddor, Patrice S., and Horrocks, Julie (2002). The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics*, **30**, 77–100.

Kingston, John and Diehl, Randy L. (1994). Phonetic knowledge. *Language*, **70**, 419–454.

Kingston, John, Diehl, Randy L., Kirk, Cecilia J., and Castleman, Wendy A. (2008). On the internal perceptual structure of distinctive features. *Journal of Phonetics*, **36**, 28–54.

Kirby, James P. (2010). *Cue selection and category restructuring in sound change*. Ph.D. thesis, University of Chicago.

Kirby, James P. (2011). Modeling the acquisition of covert contrast. In *Proceedings of the Seventeenth International Conference of the Phonetic Sciences*, Hong Kong.

Kullback, Solomon and Leibler, Richard A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, **22**(1), 79–86.

Liljencrants, Jonas and Lindblom, Björn (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, **48**, 839–862.

Lindblom, Björn (1990). Explaining phonetic variation: A sketch of the H & H theory. In *Speech production and speech modeling*, pp. 403–439. Kluwer, Dordrecht.

Lisker, Leigh (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech*, **29**, 3–11.

Lisker, Leigh and Abramson, Arthur (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, **20**, 384–422.

Lisker, Leigh and Abramson, Arthur (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the Sixth International Congress of Phonetic Sciences*, Prague, pp. 563–567. Academia Publishing House of the Czechoslovak Academy of Sciences.

Martinet, André (1952). Function, structure, and sound change. *Word*, **8**(1), 1–32.

Matisoff, James A. (1973). Tonogenesis in Southeast Asia. In *Consonant types and tone* (ed. L. Hyman), Southern California Occasional Papers in Linguistics, pp. 71–95. University of Southern California, Los Angeles.

Maye, Jessica, Werker, Janet F., and Gerken, LouAnn (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, **82**(3), B101–B111.

McLachlan, Geoffry J. and Peel, David (2000). *Finite mixture models*. Wiley, New York.

McMurray, Bob, Aslin, Richard N., and Toscano, Joseph C. (2009). Statistical learning of phonetic categories: Insights from a computational approach. *Developmental Science*, **12**(3), 369–378.

Moreton, Elliott (2008). Analytic bias and phonological typology. *Phonology*, **25**, 83–127.

Nearey, Terrance and Hogan, John T. (1986). Phonological contrast in experimental phonetics: Relating distributions of production data to perceptual categorization curves.

In *Experimental phonology* (ed. J. J. Ohala and J. J. Jaeger), pp. 141–162. Academic Press, Orlando.

Ohala, John J. (1981). The listener as a source of sound change. In *CLS 17-2: Papers from the parasession on language and behavior* (ed. C. Masek, R. Hendrick, and M. Miller), pp. 178–203. Chicago Linguistic Society, Chicago.

Ohala, John J. (1997). Aerodynamics of phonology. In *Proc. 4th Seoul International Conference on Lingusitics [SICOL]*, Seoul, pp. 92–97.

Pierrehumbert, Janet (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In *Frequency effects and the emergence of linguistic structure* (ed. J. Bybee and P. Hopper), pp. 137–157. John Benjamins, Amsterdam.

Silva, David J. (1992). *The phonetics and phonology of stop lenition in Korean.* Ph.D. thesis, Cornell University.

Silva, David J. (1993). A phonetically based analysis of [voice] and [fortis] in Korean. In *Japanese/Korean Linguistics* (ed. P. M. Clancy), Volume 2, pp. 164–174. CSLI, Stanford.

Silva, David J. (2006*a*). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology*, **23**, 287–308.

Silva, David J. (2006*b*). Variation in voice onset time for Korean stops: A case for recent sound change. *Korean Linguistics*, **13**, 1–16.

Stevens, Kenneth N. and Keyser, Samuel J. (1989). Primary features and their enhancement in consonants. *Language*, **65**, 81–106.

Toscano, Joseph C. and McMurray, Bob (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, **34**, 434–464.

Vallabha, Gautam K., McClelland, James L., Pons, Ferran, Werker, Janet F., and Amano, Shigeaki (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences*, **104**(33), 13273–13278.

Wedel, Andrew B. (2006). Exemplar models, evolution and language change. *The Linguistic Review*, **23**, 247–274.

Wilson, Colin (2006). Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science*, **30**, 945–982.

Wright, Jonathan (2007). *Laryngeal contrasts in Seoul Korean*. Ph.D. thesis, University of Pennsylvania, Philadelphia, PA.