

# A rational account of perceptual compensation for coarticulation

Morgan Sonderegger ([morgan@cs.uchicago.edu](mailto:morgan@cs.uchicago.edu))

Department of Computer Science, University of Chicago, Chicago, IL 60637 USA

Alan Yu ([aclyu@uchicago.edu](mailto:aclyu@uchicago.edu))

Phonology Laboratory, Department of Linguistics, University of Chicago, Chicago, IL 60637 USA

## Abstract

A model is presented that explains perceptual compensation for context as a consequence of listeners optimally categorizing speech sounds given contextual variation. In using Bayes' rule to pick the most likely category, listeners' perception of speech sounds, which is biased toward the means of phonetic categories (Feldman & Griffiths, 2007; Feldman, Griffiths, & Morgan, 2009), is conditioned by contextual variation. The effect on the resulting identification curves of varying category frequencies and variances is discussed. A simulation case study of compensation for vowel-to-vowel coarticulation shows the predictions of the model closely correspond to human perceptual data.

**Keywords:** Speech perception; perceptual compensation; rational analysis.

## Introduction

A major challenge for models of speech perception is explaining the effect of context on phonemic identification. Depending on their acoustic, phonological, semantic, syntactic, and even socio-indexical contexts, identical acoustic signals can be labeled differently and different acoustic signals can be labeled identically. One of the most investigated types of contextual effects stems from phonemes' phonetic environments. Because of coarticulation, a phoneme's phonetic realization is heavily context-dependent. To understand speech, the listener must take into account context-induced coarticulatory effects to recover the intended message. The term *perceptual compensation* (PC) has often been used to characterize this type of context-induced adjustment in speech perception. For example, the identification of an ambiguous target syllable as /da/ or /ga/ is shifted by preceding /ar/ or /al/ contexts (Mann, 1980): the same /Ca/ token is *less* likely to be heard as /ga/ in /arCa/ context than in /alCa/ context. This effect has been argued to result from perceptual reduction of the coarticulatory fronting effects of /l/ on a following velar consonant: listeners are compensating for the effect of /l/ on /g/. This paper proposes a simple model in which PC effects emerge as an optimal solution to the problem of categorization in the presence of context-induced variation. In this model, listeners behave as if they are compensating because what is optimal differs by context.

PC effects have been observed in many phonetic settings. The fricative /ʃ/ has lower noise frequencies than /s/, and lip rounding lowers the resonant frequencies of nearby segments. Synthetic fricative noises ranging from /ʃ/ to /s/ are more often identified by English listeners as /s/ when followed by /u/ than by /a/ (Mann & Repp 1980; see also Mitterer 2006), presumably because listeners take into account the lowering effect of lip rounding from /u/ on the noise frequencies of /s/ in

natural coarticulated speech. As another example, the perception of a fundamental frequency ( $f_0$ ) contour can change as a function of vowel height (Hombert, 1978; Silverman, 1987) or consonant voicing (Pardo & Fowler, 1997): /i/ is perceived as lower in pitch relative to an /a/ with the same  $f_0$ , presumably because high vowels typically have higher  $f_0$  than low vowels.

Listeners' language-specific experience crucially affects the degree of perceptual compensation. In a study replicated in part below, Beddor, Harnsberger, & Lindemann (2002) found that English and Shona listeners compensate for the coarticulatory effects of  $V_2$  on  $V_1$  in  $CV_1CV_2$  sequences. That is, listeners identified a continuum of synthesized vowels between /a/ and /e/ more often as /a/ when the following vowel was /i/ than when the following vowel was /a/. Importantly, they observed that Shona listeners compensate more for the vowel contexts that triggered larger acoustic influences in speech production. Compensatory responses can affect listeners' rating judgments as well. English listeners are less accurate in judging vowel nasality in nasal than in non-nasal contexts, with nasal vowels in nasal contexts the most difficult (Beddor & Krakow, 1999; Kawasaki, 1986).

Explanations of PC effects have been advanced from several theoretical perspectives. Some emphasize the lexical and phonemic content of the context in determining the identification of the target sound (Elman & McClelland, 1988; Samuel & Pitt, 2003). Gestural theorists, who assume that listeners parse the acoustics in terms of its articulatory sources, argue that listeners attribute the acoustic properties of a target sound to the coarticulatory context rather than to the target (Fowler, 1996, 2006). Auditorists attribute context-induced shifts in category boundaries to general auditory processes such as frequency contrast or spectral contrast (Diehl & Kluender, 1989; Kingston, 1992; Kingston & Diehl, 1995; Lotto & Kluender, 1998). Such auditory explanations are unavailable for compensation effects such as vowel-dependent pitch height compensation (Fowler, 2006; Lotto & Holt, 2006). Motivated by such cases, Lotto & Holt (2006) suggest that the spectral contrast explanation be supplemented with a "general learning" mechanism for category formation from correlations between stimulus parameters.

The generality of PC effects is accentuated by evidence for contextual compensation with speech and non-speech sounds in human and non-humans (Holt, Lotto, & Kluender, 2000; Lotto, 2004). For example, when /da/-/ga/ syllables are preceded by tone glides matching in frequency to the third formant ( $F_3$ ) transition of /al/ or /ar/, listeners' syllable identi-

fication responses shifted in the same direction as when targets were preceded by real speech (/al/ or /ar/). The same effect was observed even when steady-state tones at the offset frequency of /al/ or /ar/  $F_3$  were used (Lotto & Kluender, 1998; cf. Viswanathan, Fowler, & Magnuson, 2009). Lotto, Kluender, & Holt (1997) conditioned four Japanese quails to exemplars of /da/ and /ga/ syllables. Two birds were trained to peck a key when presented with good /da/ exemplars and to not peck when presented with good /ga/ stimuli while two other quails were trained in the reverse condition (/ga/ positive, /da/ negative). After reaching a preset criterion of 10:1 ratio of pecks to positive versus negative stimuli, birds were presented with novel ambiguous CVs preceded by either /al/ or /ar/. All birds displayed a significant shift in peck rates across the change in preceding liquid. The /da/-positive birds pecked substantially more for CVs preceded by /ar/, while /ga/-positive birds pecked more for CVs preceded by /al/. Crucially, both the task and the results were essentially the same as in Mann (1980)'s experiment with human subjects. There is thus strong support for a language-independent, auditory mechanism of compensation.

In this paper, we develop a computational model of PC effects using rational analysis of speech perception and production. Rational analysis (RA; Anderson, 1990; Marr, 1982;) attempts to explain aspects of cognition as adaptive responses to the environment; its central claim is that much of people's behavior when performing some cognitive tasks can be understood as optimal, according to some criterion. RA represents a different type of explanation from existing theories of PC: instead of explaining the behavioral locus (e.g. gestural processing, lexical knowledge, general auditory processes) of PC effects, the model presented here gives an account of *why* PC effects occur, as a consequence of listeners optimally solving the problem of categorization given context-induced variation.

RA accounts have been developed for visual word recognition (Norris, 2006), spoken-word recognition (Norris & McQueen, 2008), perceptual magnet effects (Feldman & Griffiths, 2007; Feldman et al., 2009), and other cognitive domains, such as vision (Marr, 1982; Yuille & Kersten, 2006) and manual movement (Trommershäuser, Gepshtein, Maloney, Landy, & Banks, 2005). Our analysis of PC effects grows out of the rational model of perceptual magnet effects of Feldman et al. (2007, 2009). While "optimal" can be understood in Bayesian (e.g. Tenenbaum & Griffiths, 2001) or maximum likelihood (e.g. Fried & Holyoak, 1984) terms, following Feldman et al. and other recent rational accounts of speech perception (Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Norris & McQueen, 2008), we use Bayesian inference here.

## Model

Our rational model for PC effects assumes a simple scenario where an idealized optimal listener has to categorize some signal as one of two phonetic categories; this is analogous to

the task listeners perform in the two-alternative forced choice (2AFC) paradigm commonly used in PC experiments. The model formalism is adapted from that used by Feldman et al. (2009). We differ in allowing model parameters to change with context, and focus on different aspects of the model's predictions.<sup>1</sup>

The modeled listener hears signal  $S$  in context  $k$ , and must decide whether it belongs to category  $c_1$  or  $c_2$ . Listeners in this model assume  $S$  is normally distributed around a target pronunciation  $T$ , itself normally distributed around a category mean, and categorize based on the likelihood that  $S$  is an instance of the speaker producing an example from  $c_i$  in context  $k$ , with target  $T$ . Formally,

$$T | c_i, k \sim N(\mu_{c_i, k}, \sigma_c), \quad S | T, c_i \sim N(T, \sigma_s)$$

where  $\mu_{c_i, k}$  is the mean of category  $i$  mean in context  $k$ ,  $\sigma_c^2$  is the variance in  $T$  around the category mean, and  $\sigma_s^2$  is the variance in  $S$  around  $T$ . We assume for simplicity that  $\sigma_c$  and  $\sigma_s$  are the same for categories 1 and 2. Although we assume that  $T$  is the variable shifting by context, if it is instead assumed that  $S$  shifts by context in a similar way, all results turn out the same.<sup>2</sup> It thus does not matter under this analysis whether contextual variation is in the target pronunciation,  $T$ , or the acoustic signal itself,  $S$ .

The probability  $S$  comes from category  $c_1$  can be calculated with Bayes' rule:

$$P(c_1 | S, k) = \frac{P(c_1 | k)P(S | c_1, k)}{P(c_2 | k)P(S | c_2, k) + P(c_1 | k)P(S | c_1, k)} \quad (1)$$

$P(c_i | k)$  is the probability of category  $i$  occurring in context  $k$ , i.e. in the lexicon as a whole. The  $P(S | c_i, k)$  are calculated by integrating over all possible  $T$ , giving a logistic function:

$$P(c_1 | S, k) = \left( 1 + \frac{f_2}{f_1} e^{b - Sg} \right)^{-1} \quad (2)$$

where

$$b = \frac{1}{2} \frac{\mu_{c_1, k}^2 - \mu_{c_2, k}^2}{\sigma_s^2 + \sigma_c^2}, \quad g = \frac{\mu_{c_1, k} - \mu_{c_2, k}}{\sigma_s^2 + \sigma_c^2}$$

and  $f_i = P(c_i | k)$  is the frequency of category  $i$  in context  $k$ .

Studies of PC generally focus on locating the *crossover point*, where  $S$  is maximally ambiguous between categories, i.e.  $S'$  (see Fig. 1) such that  $P(c_1 | S', k) = P(c_2 | S', k) = 0.5$ . Solving from (2) gives

$$S' = \frac{\mu_{c_1, k} + \mu_{c_2, k}}{2} + \frac{\sigma_s^2 + \sigma_c^2}{\mu_{c_1, k} - \mu_{c_2, k}} \ln\left(\frac{f_2}{f_1}\right) \quad (3)$$

<sup>1</sup>Space constraints prevent us from giving detailed derivations below; these are given by (Feldman et al., 2009).

<sup>2</sup>Specifically, if we assume compensation is in  $S$ , of the form

$$T | c_i, k \sim N(\mu_{c_i}, \sigma_c), \quad S | T, c_i, k \sim N(T + \Delta_{i, k}, \sigma_s).$$

That is, the distribution of  $T$  varies by category, but is not affected by context. Given  $T$ , the distribution of  $S$  has a mean offset from  $T$  by an amount  $\Delta_{i, k}$ , which depends on the context.

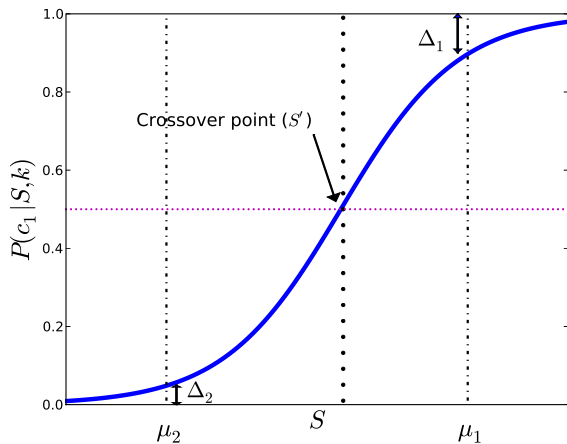


Figure 1: Schematic of a modeled identification curve.  $\mu_1, \mu_2$  are category means,  $S'$  is the crossover point, and  $\Delta_1, \Delta_2$  are miscategorization probabilities.

Perceptual compensation is thus captured in this model in terms of a shift in the crossover point as a function of the context. Note that if it is assumed that  $f_1 = f_2$ ,  $S'$  is simply halfway between the category means, while if category frequencies are not equal ( $f_1 \neq f_2$ ),  $S'$  is shifted.

The shape of the identification curve also changes as system parameters are changed. Two important properties of the curve, schematized in Fig. 1, are the slope at the crossover point and the misclassification probabilities at the category means.

The identification curve's slope at the crossover point is a rough measure of the "degree of uncertainty" (Clayards et al., 2008) of the category boundary:

$$\text{slope at } S' = \left. \frac{dP(c_1 | S, k)}{dS} \right|_{S=S'} = \frac{\mu_{c_1, k} - \mu_{c_2, k}}{4(\sigma_S^2 + \sigma_c^2)}$$

The shallower the slope, the greater the uncertainty. The slope is steeper when the difference in category means is larger relative to category variances. Unlike the crossover point's location, the slope does not change depending on whether  $f_1 = f_2$ .

Categorization uncertainty can also be quantified as the *misclassification probabilities*  $\Delta_1$  and  $\Delta_2$ , defined as the probability a signal  $S$  produced at the mean of category  $i$  — a "perfect" exemplar from  $c_i$  — is misclassified. We find

$$\Delta_1 = \left(1 + \frac{f_1}{f_2} e^{\frac{(\mu_1 - \mu_2)^2}{2V}}\right)^{-1} \quad \Delta_2 = \left(1 + \frac{f_2}{f_1} e^{\frac{(\mu_1 - \mu_2)^2}{2V}}\right)^{-1}$$

where  $V = \sigma_c^2 + \sigma_S^2$ . The misclassification probabilities decrease as the ratio of the difference in category means to the variance increases. When  $f_1 > f_2$ ,  $\Delta_1$  decreases and  $\Delta_2$  increases (and vice versa for  $f_1 < f_2$ ).

To illustrate the adequacy of the proposed model and its treatment of perceptual compensation, the next section re-

ports the results of a simulation study of PC for anticipatory vowel-to-vowel coarticulation in English.

## A Simulation Study

A modified replication study of Beddor et al. (2002)'s seminal perception and production study of vowel-to-vowel coarticulation in English was conducted. The perceptual results serve as the observed PC responses. These were compared to responses predicted by the rational model, using parameter values obtained from two production studies.

### Observed perceptual responses

Eighteen native English speakers at the University of Chicago participated in a perception experiment, consisting of a training phase followed by a test phase. The training phase was intended to expose subjects to speech in which each of  $V_1=/a/$  and  $V_1=/e/$  was equally likely to occur in the context of following  $V_2=/a/$  or  $V_2=/i/$ , corresponding to  $f_1=f_2$  in our model. The test phase asked listeners to classify an ambiguous vowel  $V_1$  as  $/a/$  or  $/e/$ , in the context of  $V_2=/a/$  or  $/i/$ .

In the training phase, listeners heard  $CV_1CV_2$  tokens ( $C=/p/, /t/, \text{ or } /k/, V=/e/ \text{ or } /a/, V_2=/a/ \text{ or } /i/$ ). Tokens were constructed by splicing together CV syllables produced in isolation by an adult male speaker of English. A total of thirty-six tokens were constructed ( $=3C \times 2V_1 \times 3C \times 2V_2$ ). Each  $CV_1CV_2$  token was heard ten times, for a total of 360 tokens, presented in random order. To encourage attention to the training stimuli, listeners performed a phoneme monitoring task where they were asked to identify whether or not each token contained a medial  $/t/$ .

In the test phase, listeners performed a 2AFC categorization task on  $V_1$  in  $bV_1bV_2$  context, with  $V_1$  varying in  $F_1$ - $F_3$  along an 9-step  $/a-e/$  continuum, and  $V_2=/a/$  or  $/i/$ . The nine-step continuum was generated using Akustyk (Plichta & Preston, 2004), an add-on program for vowel analysis in Praat (Boersma & Weenink, 2001), by interpolating the formant values between two syllables ( $/ba/$  and  $/be/$ ) produced in isolation.<sup>3</sup> The test tokens were then created by splicing together each individual continuum syllable with either a  $/bi/$  or a  $/ba/$  syllable, also produced in isolation. The same speaker produced the speech stimuli used in both the training and test phases. Each subject heard each test stimulus ten times, for a total of 180 tokens, presented in random order. Subjects were paid a nominal fee to participate in the studies.

Fig. 2 shows empirical curves of the proportion of  $V_1=/a/$  responses in  $V_2=/a/$  and  $V_2=/i/$  contexts, as a function of position on the  $V_1$  continuum. Error bars correspond to 95% confidence intervals over individual-subject proportions.

The  $V_1$  categorization responses ( $1=/a/$ ) were modeled using a mixed-effects logistic regression (Baayen, 2008; Jaeger,

<sup>3</sup>The  $F_1$  values of the nine steps along the  $/a-e/$  continuum were 713Hz, 682Hz, 635Hz, 606Hz, 592Hz, 563Hz, 522Hz, 500Hz, and 483 Hz. Values for the higher formants were adjusted as well to create a more natural-sounding continuum. For simplicity, we focus on the coarticulatory effect on  $F_1$  since the context vowels only vary in height and not in backness.

2008) with VOWEL CONTEXT (/a/ or /i/) and CONTINUUM (1–9) as fixed effects, and random effects of SUBJECT and BLOCK (test token number) on the intercept. As a measure of model quality, Nagelkerke’s pseudo- $R^2$  was 0.64, relative to a model with only the intercept. There were significant effects of CONTINUUM and VOWEL CONTEXT ( $p < 0.001$ ), as well as their interaction ( $p < 0.05$ ). The effect of VOWEL CONTEXT was an increase in  $V_1=/a/$  responses for  $V_2=/i/$  compared to  $V_2=/a/$ , in agreement with the results of Beddor et al. (2002): native English listeners appear to perceptually compensate for the coarticulatory effects of a following vowel.

### Model-predicted perceptual responses

To predict expected identification curves using Eqn. 2, we need the category means of /a/ and /e/ ( $V_1$ ) in the context of following /a/ or /i/ ( $V_2$ ), and category variances for  $V_1$  in  $V_2=/a/$  and  $V_2=/i/$  contexts.<sup>4</sup> (Recall that we are assuming equal variances of  $V_1=/a/$  and  $V_1=/i/$ , given the following context.) Eqn. 2 also includes the relative probability ( $f_1/f_2$ ) of  $V_1=/a/$  and  $V_1=/i/$  in each  $V_2$  context. We assume that  $f_1/f_2 = 1$  following the training phase.

The category mean and variance parameters were estimated from two production studies. Category means were based on 40 productions of the form  $bV_1bV_2$  (10 for each combination of  $V_1 \in \{a, e\}$  and  $V_2 \in \{a, i\}$ ) by the speaker whose speech was the basis of the training and test tokens. Category variances were calculated from productions of initial stressed /adV<sub>1</sub>CV<sub>2</sub>/ sequences ( $V_{1\&2}=/a/, /e/,$  or  $/i/$  and  $C=/p/$  or  $/b/$ ), each repeated ten times in random order, by four male, phonetically-trained native English speakers. No subjects who participated in the perception experiment participated in the production studies as well.

We thus assumed that during the experiment, subjects adjusted their expectation of category means to match the speaker they were hearing, but that their category variances reflected variation across speakers.<sup>5</sup>

For all production data, formant values were measured at the midpoint of the target  $V_1$ . Means and variances were calculated over Bark-transformed  $F_1$  values for  $V_1$ . Variances for  $V_1$  when  $V_2=/a/$  were taken to be the mean of the variances for /aCa/ stimuli and for /eCa/ stimuli. Variances for  $V_1$  when  $V_2=/i/$  were calculated similarly. The resulting model parameters are listed in Table 1.

The predicted identification curves for  $V_2=/a/$  and  $V_2=/i/$  contexts are given in Fig. 2. For comparison with the experimental results, Step 1 was taken to be the mean of  $\mu_{c_2}$  (where  $c_2$  is “ $V_1=/e/$ ”) in  $V_2=/a/$  and  $V_2=/i/$  contexts, and Step 9 was

Table 1: Model parameters obtained from the production study, where  $c_1$  is “ $V_1=/a/$ ”,  $c_2$  is “ $V_1=/e/$ .” B=Bark.

$V_2$	$\mu_{c_1}$	$\mu_{c_2}$	$\sigma_C^2 + \sigma_S^2$
/a/	6.69 B	4.67 B	0.568 B <sup>2</sup>
/i/	6.76 B	4.26 B	0.619 B <sup>2</sup>

taken to be the mean of  $\mu_{c_1}$  (where  $c_1$  is “ $V_1=/a/$ ”) in  $V_2=/a/$  and  $V_2=/i/$  contexts.

Qualitatively, the fit between the experimental and model-predicted curves in Fig. 2 is very good, without fitting any free model parameters to the production data. Both experimental and model curves show a rightward shift for  $V_2=/a/$  context, and the predicted slope at the crossover point for both pairs of curves are approximately the same.<sup>6</sup> However, the quality of the fit depends on how rational model parameters are derived from the production study, and should be interpreted with caution. For example, category variances ( $\sigma_C^2 + \sigma_S^2$ ) would be smaller if based on tokens from a single speaker rather than several speakers, making the slope of the rational model curves steeper.

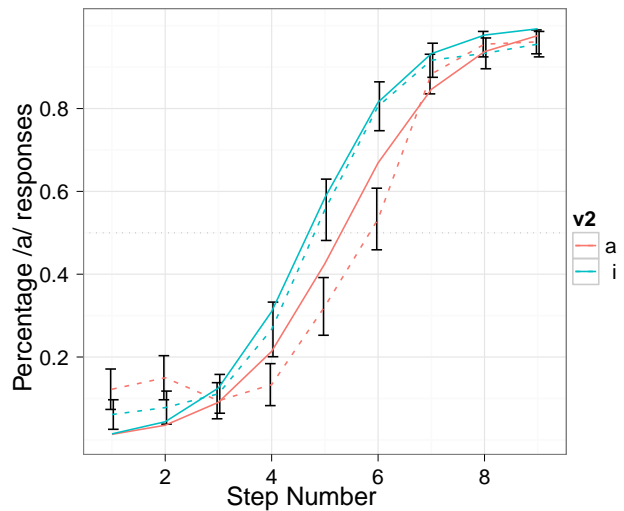


Figure 2: Dashed lines: Proportion of /a/ responses for  $V_2=/a/$  (right curve) and  $V_2=/i/$  (left curve) contexts, across all subjects. Error bars are 95% confidence intervals, based on individual-subject proportions. Solid lines: Predicted identification curves, based on production data. Dotted line: Crossover point (rate=0.5).

### Discussion

We have illustrated a rational model of perceptual compensation effects and shown that, given a simple probabilistic model for the observed values of an acoustic-phonetic cue

<sup>4</sup>Nearey & Hogan (1986) propose two models for deriving identification curves from production data. Their ‘NAPP’ model is similar to the present model, but is not derived from an RA viewpoint. We also map production data to model parameters differently.

<sup>5</sup>Another interpretation of these category variances, suggested by a reviewer, is that subjects assume the tokens have category variances typical of a single speaker, but also account for some “noise” in perception, beyond the variance observed in the production data of an individual speaker.

<sup>6</sup>The correlation between the two sets of curves is very high ( $r = 0.987$ ,  $p < 0.001$ ), indicating good qualitative agreement.

(here,  $F_1$  values) associated with a speech sound, it is possible to understand perceptual compensation as an idealized rational listener arriving at an optimal solution based on evidence from prior experience. In this model, by choosing the most probable categorization response given the context, based on their knowledge of the probability distribution of the relevant cue in that context, listeners appear to ‘undo’ the effect of coarticulation. Different contexts are associated with different cue distributions, and hence difference categorization responses.

Rational models provide a very general expression of the computational problem being solved when performing some cognitive task, and are largely orthogonal to proposed mechanisms by which the task is performed. Our model proposes an abstract explanation for why PC occurs, but is compatible with a role for different proposed mechanisms for PC effects via “prior knowledge” encoded in the cue distributions and category frequencies. The model assumes that listeners have different cue distributions for different contexts, but does not specify the source of the distributions; it could be that knowledge about gestures or general auditory capabilities generate or underly the distributions. The category frequencies could reflect knowledge of lexical or phonotactic probabilities, as pointed out by Feldman et al. (2009).

The model is able to accommodate two types of PC effects — language-dependent and domain-general — usually emphasized in opposing accounts of PC. That PC effects are language-dependent is expected because many coarticulatory effects are language-specific. Since language-specific coarticulatory effects are reflected in acoustic-phonetic cues, listeners’ categorization responses should mirror the (language-specific) probability distributions of the relevant cues. The model is general in that it is not restricted to linguistically-relevant acoustic cues. As long as a non-linguistic acoustic cue has a probability distribution, the idealized rational listener (human or non-human) would seem to compensate in the same way as she would if the acoustic cue were linguistic.

Our model predicts that compensation effects could be ameliorated or even reversed via adjustments to the model parameters. In general, an observed PC effect corresponds to different values of  $S'$  (the crossover point) in different contexts, say  $k_1$  and  $k_2$ . The second term of (3) predicts that  $S'$  in  $k_1$  and  $S'$  in  $k_2$  depend on the relative frequencies of  $c_1$  and  $c_2$  in these contexts. Thus, if  $f_2/f_1$  differs significantly by context, the context-dependent PC effect can be exaggerated, diminished, canceled, or even reversed. Failure to compensate could therefore occur for sudden change in  $f_2/f_1$  for  $k_1$  but not  $k_2$ . Since this proposed effect depends on the second term of (3), compensation could also be undone by changes in variances ( $\sigma_C^2 + \sigma_S^2$ ) or category mean differences ( $\mu_{c_1,k} - \mu_{c_2,k}$ ) for  $k_1$  versus  $k_2$ . We are currently running experiments to test the predicted effects of category frequency on compensation.

This understanding of PC failure has serious implications for current theories of sound change. Many researchers,

most notably Ohala (1993), argue that articulatory and perceptual factors shape phonological systems through listener misperception-induced sound changes, and that the synchronic typology of sound patterns is a consequence of the phonologization of such phonetic “precursors” (Barnes, 2006; Blevins, 2004; Blevins & Garrett, 1998, 2004; Kavitskaya, 2001; Yu, 2004). That is, sound change occurs when listeners mistake as representative of the speaker’s target pronunciation the effects of the speakers’ production system, the listeners’ own perceptual system, or ambient distortion of the acoustic stream. However, this account assumes that errors in perception (i.e. failure to compensate for contextual variation) lead to adjustments in perceptual and production norms. The fact that perceptual compensation is observed so robustly in speech raises questions about the feasibility of this type of model of sound change. Earlier work has assumed that failure to compensate for contextually-induced variation occurs when listeners do not detect the conditioning context. Our model suggests that the relative magnitude of compensation can be mediated by properties of the language’s lexicon (e.g. the relative frequencies of phones) as well as speakers’ prior experience with the language (e.g. pronunciation variation). That is, given certain lexical or contextual conditions, a change in compensatory response may take place even when the conditioning contextual information is accurately perceived.

## Conclusion

The model proposed here allows the incorporation of both speech-specific and general auditory factors. It proposes that perceptual compensation effects emerge as a consequence of an optimal response to the problem of categorization in the presence of context-induced variation. To be sure, the present model is simplistic, and only a first step toward modeling compensatory phenomena in general. Future work will develop more general models, e.g. with unequal category variances and multiple ( $>2$ ) categories, and explore their effects on predicted categorization behavior. Nonetheless, the present model contributes to the growing number of studies that attempt to understand speech perception from a rationalist point of view (Clayards et al., 2008; Feldman & Griffiths, 2007; Feldman et al., 2009; Norris & McQueen, 2008).

**Acknowledgments** We thank Matt Goldrick and James Kirby for comments on an earlier version of this paper, Max Bane for statistics discussion, and Ed King for setting up the experiment. Part of this work was presented at the 2009 Linguistic Society of America meeting. MS is supported by a Department of Education GAANN fellowship.

## References

- Anderson, J. (1990). *The Adaptive Character of Thought*. Hillsdale, N.J.: Erlbaum.
- Baayen, R. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: CUP.

- Barnes, J. (2006). *Strength and weakness at the interface: Positional neutralization in phonetics and phonology*. Berlin: Mouton de Gruyter.
- Beddor, P., Harnsberger, J., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics*, 30(4), 591–627.
- Beddor, P., & Krakow, R. (1999). Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *Journal of the Acoustical Society of America*, 106, 2868–2887.
- Blevins, J. (2004). *Evolutionary Phonology*. Cambridge: CUP.
- Blevins, J., & Garrett, A. (1998). The origins of consonant-vowel metathesis. *Language*, 74(3), 508–56.
- Blevins, J., & Garrett, A. (2004). The evolution of metathesis. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically-based phonology*. Cambridge: CUP.
- Boersma, P., & Weenink, D. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5(9/10), 341–345.
- Clayards, M., Tanenhaus, M., Aslin, R., & Jacobs, R. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, 108, 804–809.
- Diehl, R., & Kluender, K. (1989). On the objects of speech perception. *Ecological Psychology*, 1(2), 121–144.
- Elman, J., & McClelland, J. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory & Language*, 27(2), 143–165.
- Feldman, N., & Griffiths, T. (2007). A rational account of the perceptual magnet effect. In D. McNamara & J. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society*. Austin, TX: Cognitive Science Society.
- Feldman, N., Griffiths, T., & Morgan, J. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, 116(4), 752–782.
- Fowler, C. (1996). Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America*, 99, 1730–1741.
- Fowler, C. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics*, 68(2), 161–177.
- Fried, L., & Holyoak, K. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 10(2), 234–257.
- Holt, L., Lotto, A., & Kluender, K. (2000). Neighboring spectral content influences vowel identification. *Journal of the Acoustical Society of America*, 108, 710–722.
- Hombert, J.-M. (1978). Consonant types, vowel quality, and tone. In V. Fromkin (Ed.), *Tone: a linguistic survey*. New York: Academic Press.
- Jaeger, T. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446.
- Kavitskaya, D. (2001). *Compensatory Lengthening: Phonetics, phonology, diachrony*. Unpublished doctoral dissertation, University of California at Berkeley.
- Kawasaki, H. (1986). Phonetic explanation for phonological universals: The case of distinctive vowel nasalization. In J. Ohala & J. Jaeger (Eds.), *Experimental phonology*. Orlando: Academic Press.
- Kingston, J. (1992). The phonetic and phonology of perceptually motivated articulatory covariation. *Language & Speech*, 35, 99–114.
- Kingston, J., & Diehl, R. (1995). Intermediate properties in the perception of distinctive feature values. In B. Connell & A. Arvaniti (Eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*. Cambridge: CUP.
- Lotto, A. (2004). Perceptual compensation for coarticulation as a general auditory process. In A. Agwuale, W. Warren, & S.-H. Park (Eds.), *Proceedings of the 2003 Texas Linguistics Society Conference*. Somerville, MA: Cascadilla.
- Lotto, A., & Holt, L. (2006). Putting phonetic context effects into context: A commentary on Fowler (2006). *Perception & Psychophysics*, 68(2), 178–183.
- Lotto, A., & Kluender, K. (1998). General contrast effects in speech perception: effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, 60(4), 602–19.
- Lotto, A., Kluender, K., & Holt, L. (1997). Perceptual compensation for coarticulation by Japanese Quail (*Coturnix coturnix japonica*). *Journal of the Acoustical Society of America*, 102, 1134–1140.
- Mann, V. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, 28(5), 407–12.
- Mann, V., & Repp, B. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, 28(3), 213–28.
- Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman.
- Mitterer, H. (2006). On the causes of compensation for coarticulation: Evidence for phonological mediation. *Perception & Psychophysics*, 68(7), 1227–1240.
- Nearey, T., & Hogan, J. (1986). Phonological contrast in experimental phonetics: Relating distributions of measurements production data to perceptual categorization curves. In J. Ohala & J. Jaeger (Eds.), *Experimental phonology*. Orlando: Academic Press.
- Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, 113(2), 327–57.
- Norris, D., & McQueen, J. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115(2), 357–395.
- Ohala, J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical linguistics: Problems and perspectives*. London: Longman.
- Pardo, J., & Fowler, C. (1997). Perceiving the causes of coarticulatory acoustic variation: consonant voicing and vowel pitch. *Perception & Psychophysics*, 59(7), 1141–52.
- Plichta, B., & Preston, D. (2004). Akustyk for Praat (Version 1.7.2) [Computer software manual]. East Lansing, MI.
- Samuel, A., & Pitt, M. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory & Language*, 48(2), 416–434.
- Silverman, K. (1987). *The structure and processing of fundamental frequency contours*. Unpublished doctoral dissertation, Cambridge University.
- Tenenbaum, J., & Griffiths, T. (2001). Generalization, similarity, and Bayesian inference. *Behavioral & Brain Sciences*, 24, 629–640.
- Trommershäuser, J., Gepshtein, S., Maloney, L., Landy, M., & Banks, M. (2005). Optimal compensation for changes in task-relevant movement variability. *Journal of Neuroscience*, 25(31), 7169–7178.
- Viswanathan, N., Fowler, C., & Magnuson, J. (2009). A critical examination of the spectral contrast account of compensation for coarticulation. *Psychonomic Bulletin & Review*, 16(1), 74–79.
- Yu, A. (2004). Explaining final obstruent voicing in Lezgian: Phonetics and history. *Language*, 80(1), 73–97.
- Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: Analysis by synthesis? *Trends in Cognitive Sciences*, 10(7), 301–308.