The Influence of Categories on Perception: Explaining the Perceptual Magnet Effect as Optimal Statistical Inference

Naomi H. Feldman
Brown University

Thomas L. Griffiths
University of California, Berkeley

James L. Morgan
Brown University

A variety of studies have demonstrated that organizing stimuli into categories can affect the way the stimuli are perceived. We explore the influence of categories on perception through one such phenomenon, the perceptual magnet effect, in which discriminability between vowels is reduced near prototypical vowel sounds. We present a Bayesian model to explain why this reduced discriminability might occur: It arises as a consequence of optimally solving the statistical problem of perception in noise. In the optimal solution to this problem, listeners' perception is biased toward phonetic category means because they use knowledge of these categories to guide their inferences about speakers' target productions. Simulations show that model predictions closely correspond to previously published human data, and novel experimental results provide evidence for the predicted link between perceptual warping and noise. The model unifies several previous accounts of the perceptual magnet effect and provides a framework for exploring categorical effects in other domains.

Keywords: perceptual magnet effect, categorical perception, speech perception, Bayesian inference, rational analysis

The influence of categories on perception is well known in domains ranging from speech sounds to artificial categories of objects. Liberman, Harris, Hoffman, and Griffith (1957) first described categorical perception of speech sounds, noting that listeners' perception conforms to relatively sharp identification boundaries between categories of stop consonants and that whereas between-category discrimination of these sounds is nearly perfect, within-category discrimination is little better than chance. Similar patterns have been observed in the perception of colors (Davidoff, Davies, & Roberson, 1999), facial expressions (Etcoff & Magee, 1992), and familiar faces (Beale & Keil, 1995), as well

as the representation of objects belonging to artificial categories that are learned over the course of an experiment (Goldstone, 1994; Goldstone, Lippa, & Shiffrin, 2001). All of these categorical effects are characterized by better discrimination of between-category contrasts than within-category contrasts, although the magnitude of the effect varies between domains.

In this article, we develop a computational model of the influence of categories on perception through a detailed investigation of one such phenomenon, the *perceptual magnet effect* (Kuhl, 1991), which has been described primarily in vowels. The perceptual magnet effect involves reduced discriminability of speech sounds near phonetic category prototypes. For several reasons, speech sounds, particularly vowels, provide an excellent starting point for assessing a model of the influence of categories on perception. Vowels are naturally occurring, highly familiar stimuli that all listeners have categorized. As discussed later, a precise two-dimensional psychophysical map of vowel space can be provided, and using well-established techniques, discrimination of pairs of speech sounds can be systematically investigated under well-defined conditions so that perceptual maps of vowel space can be constructed. By comparing perceptual and psychophysical maps, we can measure the extent and nature of perceptual warping and assess such warping with respect to known categories. In addition, the perceptual magnet effect shows several qualitative similarities to categorical effects in perceptual domains outside of language, as vowel perception is continuous rather than sharply categorical (Fry, Abramson, Eimas, & Liberman, 1962) and the degree of category influence can vary substantially across testing conditions (Gerrits & Schouten, 2004). Finally, the perceptual magnet effect has been the object of extensive empirical and computational

research (e.g., Grieser & Kuhl, 1989; Guenther & Gjaja, 1996; Iverson & Kuhl, 1995; Kuhl, 1991; Lacerda, 1995). This previous research has produced a large body of data, which can be used to provide a quantitative evaluation of our approach, as well as several alternative explanations against which our account can be compared.

We take a novel approach to modeling the perceptual magnet effect, complementary to previous models that have explored how the effect might be algorithmically and neurally implemented. In the tradition of rational analysis proposed by Marr (1982) and J. R. Anderson (1990), we consider the abstract computational problem posed by speech perception and show that the perceptual magnet effect emerges as part of the optimal solution to this problem. Specifically, we assume that listeners are optimally solving the problem of perceiving speech sounds in the presence of noise. In this analysis, the listener's goal is to ascertain category membership but also to extract phonetic detail in order to reconstruct coarticulatory and nonlinguistic information. This is a difficult problem for listeners because they cannot hear the speaker's target production directly. Instead, they hear speech sounds that are similar to the speaker's target production but that have been altered through articulatory, acoustic, and perceptual noise. We formalize this problem using Bayesian statistics and show that the optimal solution to this problem produces the perceptual magnet effect.

The resulting rational model formalizes ideas that have been proposed in previous explanations of the perceptual magnet effect but goes beyond these previous proposals to explain why the effect should result from optimal behavior. It also serves as a basis for further empirical research, making predictions about the types of variability that should be seen in the perceptual magnet effect and in other categorical effects more generally. Several of these predictions are in line with previous literature, and one additional prediction is borne out in our own experimental data. Our model parallels models that have been used to describe categorical effects in other areas of cognition (Huttenlocher, Hedges, & Vevea, 2000; Körding & Wolpert, 2004; Roberson, Damjanovic, & Pilling, 2007), suggesting that its principles are broadly applicable to these areas as well.

The article is organized as follows. We begin with an overview of categorical effects across several domains and then focus more closely on evidence for the perceptual magnet effect and explanations that have been proposed to account for this evidence. The ensuing section gives an intuitive overview of our model, followed by a more formal introduction to its mathematics. We present simulations comparing the model to published empirical data and generating novel empirical predictions. An experiment is presented to test the predicted effects of speech signal noise. Finally, we discuss this model in relation to previous models, revisit its assumptions, and suggest directions for future research.

Categorical Effects

Categorical effects are widespread in cognition and perception (Harnad, 1987), and these effects show qualitative similarities across domains. This section provides an overview of basic findings and key issues concerning categorical effects in the perception of speech sounds, colors, faces, and artificial laboratory stimuli.

Speech Sounds

The classic demonstration of categorical perception comes from a study by Liberman et al. (1957), who measured subjects' perception of a synthetic speech sound continuum that ranged from /b/ to /d/ to /g/, spanning three phonetic categories. Results showed sharp transitions between the three categories in an identification task and corresponding peaks in discrimination at category boundaries, indicating that subjects were discriminating stimuli primarily on the basis of their category membership. The authors compared the data to a model in which listeners extracted only category information, and no acoustic information, when perceiving a speech sound. Subject performance exceeded that of the model consistently but only by a small percentage: Discrimination was little better than could be obtained through identification alone. Liberman and colleagues later replicated these results using the voicing dimension in stop consonant perception, with both word-initial and word-medial cues causing discrimination peaks at the identification boundaries (Liberman, Harris, Kinney, & Lane, 1961; Liberman, Harris, Eimas, Lisker, & Bastian, 1961). Other classes of consonants such as fricatives (Fujisaki & Kawashima, 1969), liquids (Miyawaki et al., 1975), and nasals (J. L. Miller & Eimas, 1977) show evidence of categorical perception as well. In all of these studies, listeners show some discrimination of within-category contrasts, and this within-category discrimination is especially evident when more sensitive measures, such as reaction times, are used (e.g., Pisoni & Tash, 1974). Nevertheless, within-category discrimination is consistently poorer than between-category discrimination across a wide variety of consonant contrasts.

A good deal of research has investigated the degree to which categorical perception of consonants results from innate biases or arises through category learning. Evidence supports a role for both factors. Studies with young infants show that discrimination peaks are present in the first few months of life (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Eimas, 1974, 1975), suggesting a role for innate biases. These early patterns may be tied to general patterns of auditory sensitivity, as nonhuman animals show discrimination peaks at category boundaries along the dimensions of voicing (Kuhl, 1981; Kuhl & Padden, 1982) and place (Kuhl & Padden, 1983; Morse & Snowdon, 1975), and humans show similar boundaries in some nonspeech stimuli (J. D. Miller, Wier, Pastore, Kelly, & Dooling, 1976; Pisoni, 1977). Studies have also shown cross-linguistic differences in perception, which indicate that perceptual patterns are influenced by phonetic category learning (Abramson & Lisker, 1970; Miyawaki et al., 1975). The interaction between these two factors remains a subject of current investigation (e.g., Holt, Lotto, & Diehl, 2004).

The role of phonetic categories in vowel perception is more controversial: vowel perception is continuous rather than strictly categorical, without obvious discrimination peaks near category boundaries (Fry et al., 1962). However, there has been some evidence for category boundary effects (Beddor & Strange, 1982) as well as reduced discriminability of vowels specifically near the centers of phonetic categories (Kuhl, 1991), and we return to this debate in more detail in the next section.

Colors

Researchers have argued that color categories are organized around universal focal colors (Berlin & Kay, 1969; Rosch Heider, 1972; Rosch Heider & Oliver, 1972), and these universal tendencies have been supported through more recent statistical modeling results (Kay & Regier, 2007; Regier, Kay, & Khetarpal, 2007). However, color terms show substantial cross-linguistic variation (Berlin & Kay, 1969), and this has led researchers to question whether color categories influence color perception. Experiments have revealed discrimination peaks corresponding to language-specific category boundaries for speakers of English, Russian, Berinmo, and Himba, and perceivers whose native language does not contain a corresponding category boundary have failed to show these discrimination peaks (Davidoff et al., 1999; Roberson, Davidoff, Davies, & Shapiro, 2005; Roberson, Davies, & Davidoff, 2000; Winawer et al., 2007). These results indicate that color categories do influence performance in color discrimination tasks.

More recent research in this domain has asked whether these categorical effects are purely perceptual or whether they are mediated by the active use of linguistic codes in perceptual tasks. Roberson and Davidoff (2000) demonstrated that linguistic interference tasks can eliminate categorical effects in color perception (see also Kay & Kempton, 1984). Investigations have shown activation of the same neural areas in naming tasks as in discrimination tasks (Tan et al., 2008) as well as left-lateralization of categorical color perception in adults (Gilbert, Regier, Kay, & Ivry, 2006). These results suggest a direct role for linguistic codes in discrimination performance, indicating that categorical effects in color perception are mediated largely by language. Nevertheless, categorical effects may play a large role in everyday color perception. Linguistic codes appear to be used in a wide variety of perceptual tasks, including those that do not require memory encoding (Witthoft et al., 2003), and verbal interference tasks fail to completely wipe out verbal coding when the type of interference is unpredictable (Pilling, Wiggett, Özgen, & Davies, 2003).

Faces

Categorical effects in face perception were first shown for facial expressions of emotion in stimuli constructed from line drawings (Etcoff & Magee, 1992) and photograph-quality stimuli (Calder, Young, Perrett, Etcoff, & Rowland, 1996; de Gelder, Teunisse, & Benson, 1997; Young et al., 1997). Stimuli for these experiments were drawn from morphed continua in which the endpoints were prototypical facial expressions (e.g., happiness, fear, anger). With few exceptions, results showed discrimination peaks at the same locations as identification boundaries between these prototypical expressions. Evidence for categorical effects has been found in seven-month-old infants (Kotsoni, de Haan, & Johnson, 2001), nine-year-old children (de Gelder et al., 1997), and older individuals (Kiffel, Campanella, & Bruyer, 2005), indicating that category structure is similar across different age ranges. However, these categories can be affected by early experience as well. Pollak and Kistler (2002) presented data from abused children showing that their category boundaries in continua ranging from fearful to angry and from sad to angry were shifted such that they interpreted a large portion of these continua as angry; discrimination peaks were shifted together with these identification boundaries.

In addition to categorical perception of facial expressions, discrimination patterns show evidence of categorical perception of facial identity, where each category corresponds to a different identity. Beale and Keil (1995) found discrimination peaks along morphed continua between faces of famous individuals, and these results have been replicated with several different stimulus continua constructed from familiar faces (Angeli, Davidoff, & Valentine, 2008; Campanella, Hanoteau, Seron, Joassin, & Bruyer, 2003; Rotshtein, Henson, Treves, Driver, & Dolan, 2005). The categorical effects are stronger for familiar faces than for unfamiliar faces (Angeli et al., 2008; Beale & Keil, 1995), but categorical effects have been demonstrated for continua involving previously unfamiliar faces as well (Levin & Beale, 2000; Stevenage, 1998). The strength of these effects for unfamiliar faces may derive from a combination of learning during the course of the experiment (Viviani, Binda, & Borsato, 2007), the use of labels during training (Kikutani, Roberson, & Hanley, 2008), and the inherent distinctiveness of endpoint stimuli in the continua (Angeli et al., 2008; Campanella et al., 2003).

Learning Artificial Categories

Several studies have demonstrated categorical effects that derive from categories learned in the laboratory, implying that the formation of novel categories can affect perception in laboratory settings. As proposed by Liberman et al. (1957), this learning component might take two forms: Acquired distinctiveness involves enhanced between-category discriminability, whereas acquired equivalence involves reduced within-category discriminability. Evidence for one or both of these processes has been found through categorization training in color perception (Özgen & Davies, 2002) and auditory perception of white noise (Guenther, Husain, Cohen, & Shinn-Cunningham, 1999). These results extend to stimuli that vary along multiple dimensions as well. Categorizing stimuli along two dimensions can lead to acquired distinctiveness (Goldstone, 1994), and similarity ratings for drawings that differ along several dimensions have shown acquired equivalence in response to categorization training (Livingston, Andrews, & Harnad, 1998). Such effects may arise partly from task-specific strategies but likely involve changes in underlying stimulus representations as well (Goldstone et al., 2001).

Additionally, several studies have demonstrated that categories for experimental stimuli are learned quickly over the course of an experiment even without explicit training. Goldstone (1995) found that implicit shape-based categories influenced subjects' perception of hues and that these implicit categories changed depending on the set of stimuli presented in the experiment. A similar explanation has been proposed to account for subjects' categorical treatment of unfamiliar face continua (Levin & Beale, 2000), where learned categories seem to correspond to continuum endpoints. Gureckis and Goldstone (2008) demonstrated that subjects are sensitive to the presence of distinct clusters of stimuli, showing increased discriminability between clusters even when those clusters receive the same label. Furthermore, implicit categories have been used to explain why subjects often bias their perception toward the mean value of a set of stimuli in an experiment. Huttenlocher et al. (2000) argued that subjects form an implicit category that includes the range of stimuli they have seen over the course of an experiment and that they use this implicit category to

correct for memory uncertainty when asked to reproduce a stimulus. Under their assumptions, the optimal way to correct for memory uncertainty using this implicit category is to bias all responses toward the mean value of the category, which in this case is the mean value of the set of stimuli. The authors presented a Bayesian analysis to account for bias in visual stimulus reproduction that is nearly identical to the one-category model derived here in the context of speech perception, reflecting the similar structure of the two problems and the generality of the approach.

Summary

The categorical effects in all of these domains are qualitatively similar, with enhanced between-category discriminability and reduced within-category discriminability. Though there is some evidence that innate biases contribute to these perceptual patterns, the patterns can be influenced by learned categories as well, even by implicit categories that arise from specific distributions of exemplars. Despite widespread interest in these phenomena, the reasons and mechanisms behind the connection between categories and perception remain unclear. In the remainder of this article, we address this issue through a detailed exploration of the perceptual magnet effect, which shares many qualitative features with the categorical effects discussed above.

The Perceptual Magnet Effect

The phenomenon of categorical perception is robust in consonants, but the role of phonetic categories in the perception of vowels has been more controversial. Acoustically, vowels are specified primarily by their first and second formants, F_1 and F_2. *Formants* are bands of frequencies in which acoustic energy is concentrated—peaks in the frequency spectrum—as a result of resonances in the vocal tract. F_1 is inversely correlated with tongue height, whereas F_2 is correlated with the proximity of the most raised portion of the tongue to the front of the mouth. Thus, a front high vowel such as /i/ (as in *beet*) spoken by a male talker typically has center formant frequencies around 270 Hz (F_1) and 2290 Hz (F_2), and a back low vowel such as /a/ (as in *father*) spoken by a male typically has center formant frequencies around 730 Hz and 1090 Hz (Peterson & Barney, 1952). Tokens of vowels are distributed around these central values. A map of vowel space based on data from Hillenbrand, Getty, Clark, and Wheeler (1995) is shown in Figure 1. Though frequencies are typically reported in Hertz, most research on the perceptual magnet effect has used the mel scale to represent psychophysical distance (e.g., Kuhl, 1991). The mel scale can be used to equate distances in psychophysical space because difference limens, the smallest detectable pitch differences, correspond to constant distances along this scale (S. S. Stevens, Volkmann, & Newman, 1937).

Early work suggested that vowel discrimination was not affected by native language categories (K. N. Stevens, Liberman, Studdert-Kennedy, & Öhman, 1969). However, later findings have revealed a relationship between phonetic categories and vowel perception. Although within-category discrimination for vowels is better than for consonants, clear peaks in discrimination functions have been found at vowel category boundaries, especially in tasks that place a high memory load on subjects or that interfere with auditory memory (Beddor & Strange, 1982; Pisoni, 1975; Repp &
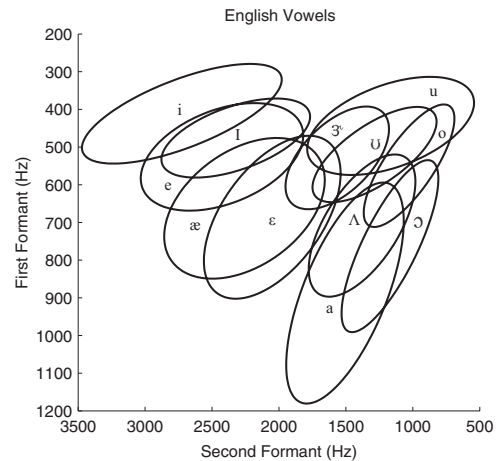
Figure 1. Map of vowel space from Hillenbrand et al.'s (1995) production experiment. Ellipses delimit regions corresponding to approximately 90% of tokens from each vowel category. Adapted from "Acoustic Characteristics of American English Vowels" by J. Hillenbrand, L. A. Getty, M. J. Clark, & K. Wheeler, 1995, *Journal of the Acoustical Society of America, 97,* p. 3103. Copyright 1995 by the Acoustical Society of America. Reprinted with permission.

Crowder, 1990; Repp, Healy, & Crowder, 1979). In addition, between-category differences yield larger neural responses as measured by event-related potentials (Näätänen et al., 1997; Winkler et al., 1999). Viewing phonetic discrimination in spatial terms, Kuhl and colleagues have found evidence of shrunken perceptual space specifically near category prototypes, a phenomenon they have called the perceptual magnet effect (Grieser & Kuhl, 1989; Iverson & Kuhl, 1995; Kuhl, 1991; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992).

Empirical Evidence

The first evidence for the perceptual magnet effect came from experiments with English-learning 6-month-old infants (Grieser & Kuhl, 1989). Using the conditioned head-turn procedure to assess within-category generalization of speech sounds, Grieser and Kuhl found that a prototypical /i/ vowel based on mean formant values in Peterson and Barney's production data was more likely to be generalized to sounds surrounding it than was a nonprototypical /i/ vowel. In addition, they found that infants' rate of generalization correlated with adult goodness ratings of the stimuli, so stimuli that were judged as the best exemplars of the /i/ category were generalized most often to neighboring stimuli. Kuhl (1991) showed that adults, like infants, can discriminate stimuli near a nonprototype of the /i/ category better than stimuli near the prototype. Kuhl et al. (1992) tested English- and Swedish-learning infants on discrimination near prototypical English /i/ (high, front, unrounded) and Swedish /y/ (high, front, rounded) sounds, again using the conditioned headturn procedure; they found that whereas English infants generalized the /i/ sounds more than the /y/ sounds, Swedish-learning infants showed the reverse pattern. On the basis of this evidence, Kuhl et al. described the perceptual magnet effect as a language-specific shrinking of perceptual space near native language phonetic category prototypes, with prototypes acting as

perceptual magnets to exert a pull on neighboring speech sounds (see also Kuhl, 1993). They concluded that these language-specific prototypes are in place as young as 6 months.

Iverson and Kuhl (1995) used signal detection theory and multidimensional scaling to produce a detailed perceptual map of acoustic space near the prototypical and nonprototypical /i/ vowels used in previous experiments. They tested adults' discrimination of 13 stimuli along a single vector in F_1–F_2 space, ranging from F_1 of 197 Hz and F_2 of 2489 Hz (classified as /i/) to F_1 of 429 Hz and F_2 of 1925 Hz (classified as /e/, as in *bait*). In both analyses, the authors found shrinkage of perceptual space near the ends of the continuum, especially near the /i/ end. They found a peak in discrimination near the center of the continuum between Stimulus 6 and Stimulus 9. This supported previous analyses, suggesting that perceptual space was shrunk near category centers and expanded near category edges. The effect has since been replicated in the English /i/ category (Sussman & Lauckner-Morano, 1995), and evidence for poor discrimination near category prototypes has been found for the German /i/ category (Diesch, Iverson, Kettermann, & Siebert, 1999). In addition, the effect has been found in the /r/ and /l/ categories in English but not Japanese speakers (Iverson & Kuhl, 1996; Iverson et al., 2003), lending support to the idea of language-specific phonetic category prototypes.

Several studies have found large individual differences between subjects in stimulus goodness ratings and category identification, suggesting that it may be difficult to find vowel tokens that are prototypical across listeners and thus raising methodological questions about experiments that examine the perceptual magnet effect (Frieda, Walley, Flege, & Sloane, 1999; Lively & Pisoni, 1997). However, data collected by Aaltonen, Eerola, Hellström, Uusipaikka, and Lang (1997) on the /i/–/y/ contrast in Finnish adults showed that discrimination performance was less variable than identification performance, and on the basis of these results, the authors argued that discrimination operates at a lower level than overt identification tasks. A more serious challenge has come from studies that question the robustness of the perceptual magnet effect. Lively and Pisoni (1997) found no evidence of a perceptual magnet effect in the English /i/ category, suggesting that listeners' discrimination patterns are sensitive to methodological details or dialect differences, although the authors could not identify the specific factors responsible for these differences. The effect has also been difficult to isolate in vowels other than /i/: Sussman and Gekas (1997) failed to find an effect in the English /ɪ/ (as in *bit*) category, and Thyer, Hickson, and Dodd (2000) found the effect in the /i/ category but found the reverse effect in the /ɔ/ (as in *bought*) category and failed to find any effect in other vowels. Whereas there has been evidence linking changes in vowel perception to differences in interstimulus interval (Pisoni, 1973) and task demands (Gerrits & Schouten, 2004), much of the variability found in vowel perception has not been accounted for.

In summary, vowel perception has been shown to be continuous rather than categorical: Listeners can discriminate two vowels that receive the same category label. However, studies have suggested that even in vowels, perceptual space is shrunk near phonetic category centers and expanded near category edges. In addition, studies have shown substantial variability in the perceptual magnet effect. This variability seems to depend on the phonetic category being tested and also on methodological details. On the basis of the predictions of our rational model, we argue that some of this variability is attributable to differences in category variance between different phonetic categories and to differences in the amount of noise through which stimuli are heard.

Previous Models

Grieser and Kuhl (1989) originally described the perceptual magnet effect in terms of category prototypes, arguing that phonetic category prototypes exert a pull on nearby speech sounds and thus create an inverse correlation between goodness ratings and discriminability. Although this inverse correlation has been examined more closely and has been used to argue that categorical perception and the perceptual magnet effect are separate phenomena (Iverson & Kuhl, 2000), most computational models of the perceptual magnet effect have assumed that it is a categorical effect, parallel to categorical perception.

Lacerda (1995) began by assuming that the warping of perceptual space emerges as a side effect of a classification problem: The goal of listeners is to classify speech sounds into phonetic categories. His model assumes that perception has been trained with labeled exemplars or that labels have been learned using other information in the speech signal. In perceiving a new speech sound, listeners retrieve only the information from the speech signal that is helpful in determining the sound's category, or label, and they categorize and discriminate speech sounds on the basis of this information. Listeners can perceive a contrast only if the two sounds differ in category membership. Implementing this idea in neural models, Damper and Harnad (2000) showed that when trained on two endpoint stimuli, neural networks will treat a voice onset time continuum categorically. One limitation of the model proposed by Lacerda (1995) is that it does not include a mechanism by which listeners can perceive within-category contrasts. As demonstrated by Lotto et al. (1998), this assumption cannot capture the data on the perceptual magnet effect because within-category discriminability is higher than this account would predict.

Other neural network models have argued that the perceptual magnet effect results not from category labels but instead from specific patterns in the distribution of speech sounds. Guenther and Gjaja (1996) suggested that neural firing preferences in a neural map reflect Gaussian distributions of speech sounds in the input and that because more central sounds have stronger neural representations than do more peripheral sounds, the population vector representing a speech sound that is halfway between the center and the periphery of its phonetic category will appear closer to the center of the category than to its periphery. This model implements the idea that the perceptual magnet effect is a direct result of uneven distributions of speech sounds in the input. Similarly, Vallabha and McClelland (2007) have shown that Hebbian learning can produce attractors at the locations of Gaussian input categories and that the resulting neural representation fits human data accurately. The idea that distributions of speech sounds in the input can influence perception is supported by experimental evidence showing that adults and infants show better discrimination of a contrast embedded in a bimodal distribution of speech sounds than of the same contrast embedded in a unimodal distribution (Maye & Gerken, 2000; Maye, Werker, & Gerken, 2002).

These previous models have provided process-level accounts of how the perceptual magnet effect might be implemented algorithmically and neurally, but they leave several questions unanswered.

The prototype model does not give independent justification for the assumption that prototypes should exert a pull on neighboring speech sounds; several models cannot account for better than chance within-category discriminability of vowels. Other models give explanations of how the effect might occur but do not address the question of why it should occur. Our rational model fills these gaps by providing a mathematical formalization of the perceptual magnet effect at Marr's (1982) computational level, considering the goals of the computation and the logic by which these goals can be achieved. It gives independent justification for the optimality of a perceptual bias toward category centers and simultaneously predicts a baseline level of within-category discrimination. Furthermore, our model goes beyond these previous models to make novel predictions about the types of variability that should be seen in the perceptual magnet effect.

Theoretical Overview of the Model

Our model of the perceptual magnet effect focuses on the idea that we can analyze speech perception as a kind of optimal statistical inference. The goal of listeners, in perceiving a speech sound, is to recover the phonetic detail of a speaker's target production. They infer this target production using the information that is available to them from the speech signal and their prior knowledge of phonetic categories. Here we give an intuitive overview of our model in the context of speech perception, followed by a more general mathematical account in the next section.

Phonetic categories are defined in the model as distributions of speech sounds. When speakers produce a speech sound, they choose a phonetic category and then articulate a speech sound from that category. They can use their specific choice of speech sounds within the phonetic category to convey coarticulatory information, affect, and other relevant information. Because there are several factors that speakers might intend to convey, and given that each factor can cause small fluctuations in acoustics, we assume that the combination of these factors approximates a Gaussian, or normal, distribution. Phonetic categories in the model are thus Gaussian distributions of target speech sounds. Categories may differ in the location of their means, or prototypes, and in the amount of variability they allow. In addition, categories may differ in frequency so that some phonetic categories are used more frequently in a language than others. The use of Gaussian phonetic categories in this model does not reflect a belief that speech sounds actually fall into parametric distributions. Rather, the mathematics of the model are easiest to derive in the case of Gaussian categories. As discussed later, the general effects that are predicted in the case of Gaussian categories are similar to those predicted for other types of unimodal distributions.

In the speech sound heard by listeners, the information about the target production is masked by various types of articulatory, acoustic, and perceptual noise. The combination of these noise factors is approximated through Gaussian noise, so that the speech sound heard is normally distributed around the speaker's target production.

Formulated in this way, speech perception becomes a statistical inference problem. When listeners perceive a speech sound, they can assume it was generated by selecting a target production from a phonetic category and then generating a noisy speech sound on the basis of the target production. Listeners hear the speech sound

and know the structure and location of phonetic categories in their native language. Given this information, they need to infer the speaker's target production. They infer phonetic detail in addition to category information in order to recover the gradient coarticulatory and nonlinguistic information that the speaker intended.

With no prior information about phonetic categories, listeners' perception should be unbiased, given that under Gaussian noise, speech sounds are equally likely to be shifted in either direction. In this case, listeners' safest strategy is to guess that the speech sound they heard was the same as the target production. However, experienced listeners know that they are more likely to hear speech sounds near the centers of phonetic categories than speech sounds farther from category centers. The optimal way to use this knowledge of phonetic categories to compensate for a noisy speech signal is to bias perception toward the center of a category, toward the most likely target productions.

In a hypothetical language with a single phonetic category, where listeners are certain that all sounds belong to that category, this perceptual bias toward the category mean causes all of perceptual space to shrink toward the center of the category. The resulting perceptual pattern is shown in Figure 2a. If there is no uncertainty about category membership, perception of distant speech sounds is more biased than perception of proximal speech sounds so that all of perceptual space is shrunk to the same degree.

In order to optimally infer a speaker's target production in the context of multiple phonetic categories, listeners must determine which categories are likely to have generated a speech sound. They can then predict the speaker's target production on the basis of the structure of these categories. If they are certain of a speech sound's category membership, their perception of the speech sound should be biased toward the mean of that category, as was the case in a
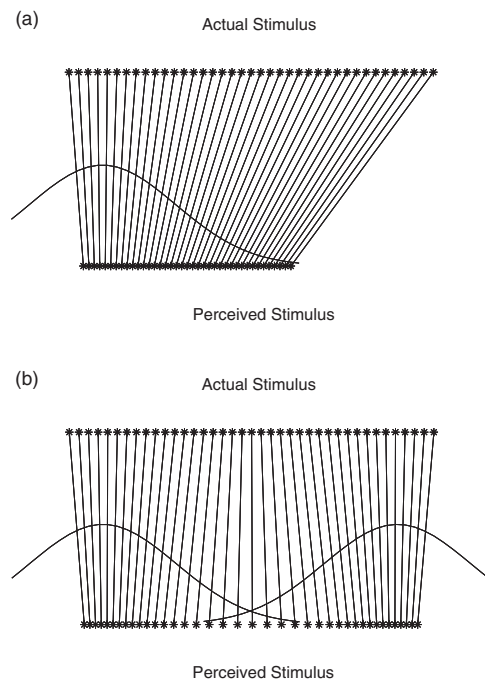

Figure 2. Predicted relationship between acoustic and perceptual space in the case of (a) one category and (b) two categories.

language with one phonetic category. This shrinks perceptual space in areas of unambiguous categorization. If listeners are uncertain about category membership, they should take into account all of the categories that could have generated the speech sound they heard, but they should weight the influence of each category by the probability that the speech sound came from that category. This ensures that under assumptions of equal frequency and variance, nearby categories are weighted more heavily than those farther away. Perception of speech sounds precisely on the border between two categories is pulled simultaneously toward both category means, each canceling out the other's effect. Perception of speech sounds that are near the border between categories is biased toward the most likely category, but the competing category dampens the bias. The resulting pattern for the two-category case is shown in Figure 2b.

The interaction between the categories produces a pattern of perceptual warping that is qualitatively similar to descriptions of the perceptual magnet effect and other categorical effects that have been reported in the literature. Speech sounds near category centers are extremely close together in perceptual space, whereas speech sounds near the edges of a category are much farther apart. This perceptual pattern results from a combination of two factors, both of which were proposed by Liberman et al. (1957) in reference to categorical perception. The first is acquired equivalence within categories due to perceptual bias toward category means; the second is acquired distinctiveness between categories due to the presence of multiple categories. Consistent with these predictions, infants acquiring language have shown both acquired distinctiveness for phonemically distinct sounds and acquired equivalence for members of a single phonemic category over the course of the first year of life (Kuhl et al., 2006).

Mathematical Presentation of the Model

This section formalizes the rational model within the framework of Bayesian inference. The model is potentially applicable to any perceptual problem in which a perceiver needs to recover a target from a noisy stimulus, using knowledge that the target has been sampled from a Gaussian category. We therefore present the mathematics in general terms, referring to a generic stimulus S, target T, category c, category variance σ_c^2, and noise variance σ_S^2. In the specific case of speech perception, S corresponds to the speech sound heard by the listener, T to the phonetic detail of a speaker's intended target production, and c to the language's phonetic categories; the category variance σ_c^2 represents meaningful within-category variability, and the noise variance σ_S^2 represents articulatory, acoustic, and perceptual noise in the speech signal.

The formalization is based on a generative model in which a target T is produced by sampling from a Gaussian category c with mean μ_c and variance σ_c^2. The target T is distributed as

$$T|c \sim N(\mu_c, \sigma_c^2). \tag{1}$$

Perceivers cannot recover T directly, but instead perceive a noisy stimulus S that is normally distributed around the target production with noise variance σ_S^2 such that

$$S|T \sim N(T, \sigma_S^2). \tag{2}$$

Note that integrating over T yields

$$S|c \sim N(\mu_c, \sigma_c^2 + \sigma_S^2), \tag{3}$$

indicating that under these assumptions, the stimuli that perceivers observe are normally distributed around a category mean μ_c, with a variance that is a sum of the category variance and the noise variance.

Given this generative model, perceivers can use Bayesian inference to reconstruct the target from the noisy stimulus. According to Bayes' rule, given a set of hypotheses H and observed data d, the posterior probability of any given hypothesis h is

$$p(h|d) = \frac{p(d|h)p(h)}{\sum_{h \in H} p(d|h)p(h)}, \tag{4}$$

indicating that it is proportional to both the likelihood $p(d|h)$, which is a measure of how well the hypothesis fits the data, and the prior $p(h)$, which gives the probability assigned to the hypothesis before any data were observed. Here, the stimulus S serves as data d; the hypotheses under consideration are all the possible targets T; and the prior $p(h)$, which gives the probability that any particular target will occur, is specified by category structure. In laying out the solution to this statistical problem, we begin with the case in which there is a single category and then move to the more complex case of multiple categories.

One Category

Perceivers are trying to infer the target T given stimulus S and category c, so they must calculate $p(T|S, c)$. They can use Bayes' rule:

$$p(T|S, c) \propto p(S|T)p(T|c). \tag{5}$$

The likelihood $p(S|T)$, given by the noise process (Equation 2), assigns highest probability to stimulus S, and the prior $p(T|c)$, given by category structure (Equation 1), assigns highest probability to the category mean. As described in Appendix A, the right-hand side of this equation can be simplified to yield a Gaussian distribution

$$p(T|S, c) = N\left(\frac{\sigma_c^2 S + \sigma_S^2 \mu_c}{\sigma_c^2 + \sigma_S^2}, \frac{\sigma_c^2 \sigma_S^2}{\sigma_c^2 + \sigma_S^2}\right) \tag{6}$$

whose mean falls between the stimulus S and the category mean μ_c.

This posterior probability distribution can be summarized by its mean (the expectation of T given S and c),

$$E[T|S, c] = \frac{\sigma_c^2 S + \sigma_S^2 \mu_c}{\sigma_c^2 + \sigma_S^2}. \tag{7}$$

The optimal guess at the target, then, is a weighted average of the observed stimulus and the mean of the category that generated the stimulus, where the weighting is determined by the ratio of cate-

gory variance to noise variance.[1] This equation formalizes the idea of a perceptual magnet: The term μ_c pulls the perception of stimuli toward the category center, effectively shrinking perceptual space around the category.

Multiple Categories

The one-category case, while appropriate to explain performance on some perceptual tasks (e.g., Huttenlocher et al., 2000), is inappropriate for describing natural language. In a language with multiple phonetic categories, listeners must consider many possible source categories for a speech sound. We therefore extend the model so that it applies to the case of multiple categories.

Upon observing a stimulus, perceivers can compute the probability that it came from any particular category using Bayes' rule

$$p(c|S) = \frac{p(S|c)p(c)}{\sum_c p(S|c)p(c)}, \tag{8}$$

where $p(S|c)$ is given by Equation 3 and $p(c)$ reflects the prior probability assigned to category c.

To compute the posterior on targets $p(T|S)$, perceivers need to marginalize, or sum, over categories,

$$p(T|S) = \sum_c p(T|S, c)p(c|S). \tag{9}$$

The first term on the right-hand side is given by Equation 6, and the second term can be calculated from Bayes' rule, as given by Equation 8. The posterior has the form of a mixture of Gaussians, where each Gaussian distribution represents the solution for a single category. Restricting our analysis to the case of categories with equal category variance σ_c^2, we find that the mean of this posterior probability distribution is

$$E[T|S] = \sum_c p(c|S)\frac{\sigma_c^2 S + \sigma_S^2 \mu_c}{\sigma_c^2 + \sigma_S^2}, \tag{10}$$

which can be rewritten as

$$E[T|S] = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_S^2} S + \frac{\sigma_S^2}{\sigma_c^2 + \sigma_S^2} \sum_c p(c|S)\mu_c. \tag{11}$$

A full derivation of this expectation is given in Appendix A.

Equation 11 gives the optimal guess for recovering a target in the case of multiple categories. This guess is a weighted average of the stimulus S and the means μ_c of all the categories that might have produced S. When perceivers are certain of a stimulus's category, this equation reduces to Equation 7, and perception of a stimulus S is biased toward the mean of its category. However, when a stimulus is on a border between two categories, the optimal guess at the target is influenced by both category means, and each category weakens the other's effect (Figure 2b). Shrinkage of perceptual space is thus strongest in areas of unambiguous categorization—the centers of categories—and weakest at category boundaries.

This analysis demonstrates that warping of perceptual space that is qualitatively consistent with the perceptual magnet effect emerges as the result of optimal perception of noisy stimuli. In the next two sections, we provide a quantitative investigation of the model's predictions in the context of speech perception. The next section focuses on comparing the predictions of the model with empirical data on the perceptual magnet effect using phonetic category parameters that are estimated from human data. In the subsequent section, we examine the consequences of manipulating these parameters, relating the model's behavior to further results from the literature.

Quantitative Evaluation

In this section, we test the model's predictions quantitatively against the multidimensional scaling results from Experiment 3 in Iverson and Kuhl (1995). These data were selected as a modeling target because they give a clean, precise spatial representation of the warping associated with the perceptual magnet effect, mapping 13 /i/ and /e/ stimuli that are separated by equal psychoacoustic distance onto their corresponding locations in perceptual space. Because these multidimensional scaling data constitute the basis for both this simulation and the experiment reported below, we describe the experimental setup and results in some detail here.

Iverson and Kuhl's (1995) multidimensional scaling experiment was conducted with thirteen vowel stimuli along a single continuum in F_1–F_2 space ranging from /i/ to /e/, whose exact formant values are shown in Table 1. The stimuli were designed to be equally spaced when measured along the mel scale, which equates distances on the basis of difference limens (S. S. Stevens et al., 1937). Subjects performed an AX discrimination task in which they pressed and held a button to begin a trial, releasing the button as quickly as possible if they believed the two stimuli to be different or holding the button for the remainder of the trial (2000 ms) if they heard no difference between the two stimuli. Subjects heard 156 "different" trials, consisting of all possible ordered pairs of nonidentical stimuli, and 52 "same" trials, four of each of the 13 stimuli.

Iverson and Kuhl (1995) reported a total accuracy rate of 77% on different trials and a false alarm rate of 31% on same trials, but they did not further explore direct accuracy measures. Instead, they created a full similarity matrix consisting of log reaction times of different responses for each pair of stimuli. To avoid sparse data in the cells where most participants incorrectly responded that two stimuli were identical, the authors replaced all same responses with the trial length, 2,000 ms, effectively making them into different responses with long reaction times. This similarity matrix was used for multidimensional scaling, which finds a perceptual map that is most consistent with a given similarity matrix. In this case, the authors constrained the solution to be in one dimension and assumed a linear relation between similarity values and distance in perceptual space. The interstimulus distances obtained from this analysis are shown in Figure 3. The perceptual map obtained through multidimensional scaling showed that neighboring stimuli near the ends of the stimulus vector were separated by less perceptual distance than neighboring stimuli near the center of the vector. These results agreed qualitatively with data obtained in Experiment 2 of the same article (Iverson & Kuhl, 1995), which used d' as an unbiased estimate of perceptual distance. We chose the multidimensional scaling data as our modeling target because they are more extensive than the d' data, encompassing the entire range of stimuli.

---

[1] The expectation is optimal if the penalty for misidentifying a stimulus increases with squared distance from the target.

We tested a two-category version of the rational model to determine whether parameters could be found that would reproduce these empirical data. Equal variance was assumed for the two categories, and parameters in the model were based as much as possible on empirical measures in order to reduce the number of free parameters. The simulation was constrained to a single dimension along the direction of the stimulus vector. The parameters that needed to be specified were as follows:

$\mu_{/i/}$: /i/ category mean

$\mu_{/e/}$: /e/ category mean

σ_c^2: category variance

σ_S^2: uncertainty in the speech signal.

Subject goodness ratings from Iverson and Kuhl (1995) were first used to specify the mean of the /i/ category, $\mu_{/i/}$. These goodness ratings indicated that the best exemplars of the /i/ category were Stimuli 2 and 3, so the mean of the /i/ category was set halfway between these two stimuli.[2]

The mean of the /e/ category, $\mu_{/e/}$, and the sum of the variances, $\sigma_c^2 + \sigma_S^2$, were calculated as described in Appendix B on the basis of phoneme identification curves from Lotto et al. (1998). These identification curves were produced through an experiment in which subjects were played pairs of stimuli from the 13-stimulus vector and asked to identify either the first or the second stimulus in the pair as /i/ or /e/. The other stimulus in the pair was one of two reference stimuli, either Stimulus 5 or Stimulus 9. Lotto et al. obtained two distinct curves in these two conditions, showing that the phoneme boundary shifted depending on the identity of the reference stimulus. Because the task used for multidimensional scaling involved presentation of all possible pairings of the 13 stimuli, the phoneme boundary in the model was assumed to be halfway between the boundaries that appeared in these two referent conditions. In order to identify this boundary, we fit two logistic curves to the prototype and nonprototype identification
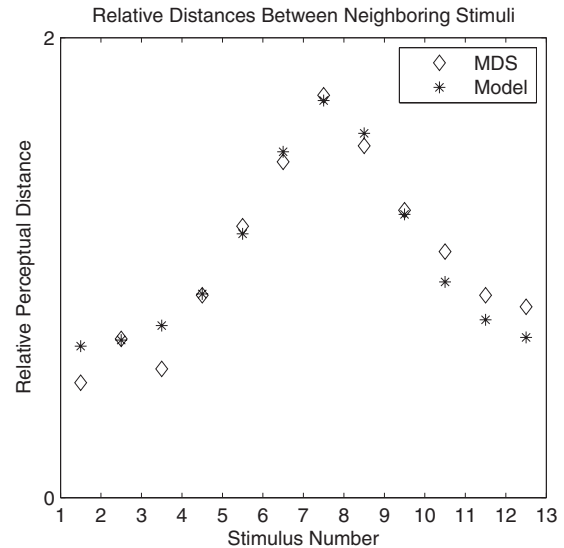

Figure 3. Relative distances between neighboring stimuli in Iverson and Kuhl's (1995) multidimensional scaling (MDS) analysis and in the model. Adapted from P. Iverson & P. K. Kuhl, "Mapping the Perceptual Magnet Effect for Speech Using Signal Detection Theory and Multidimensional Scaling, 1995, *Journal of the Acoustical Society of America, 97,* p. 559. Copyright 1995 by the Acoustical Society of America. Reprinted with permission.

curves. The two curves were constrained to have the same gain, and the biases of the two curves were averaged to obtain a single bias term. On the basis of Equation 34, these values indicated that $\mu_{/e/}$ should be placed just to the left of Stimulus 13; Equation 35 yielded a value of 10,316 for $\sigma_c^2 + \sigma_S^2$. The resulting discriminative boundary is shown together with the data from Lotto et al. (1998) in Figure 4.

The ratio between the category variance σ_c^2 and the speech signal noise σ_S^2 was the only remaining free parameter, and we chose its value so as to maximize the fit to Iverson and Kuhl's (1995) multidimensional scaling data. This direct comparison was made by calculating the expectation $E[T|S]$ for each of the 13 stimuli according to Equation 11 and then determining the distance in mels between the expected values of neighboring stimuli. These distances were compared with the distances between stimuli in the multidimensional scaling solution. Because multidimensional scaling gives relative, and not absolute, distances between stimuli, we evaluated this comparison on the basis of whether mel distances in the model were proportional to distances found through multidimensional scaling. As shown in Figure 3, the model yielded an extremely close fit to the empirical data, yielding interstimulus distances that were proportional to those found in multidimensional scaling ($r = .97$). This simulation used the following parameters:

$\mu_{/i/}$: $F_1 = 224$ Hz, $F_2 = 2413$ Hz

$\mu_{/e/}$: $F_1 = 423$ Hz, $F_2 = 1936$ Hz

Table 1

Formant Values for Stimuli Used in the Multidimensional Scaling Experiment, Reported in Iverson and Kuhl (2000)

Stimulus no.	F_1 (Hz)	F_2 (Hz)
1	197	2489
2	215	2438
3	233	2388
4	251	2339
5	270	2290
6	289	2242
7	308	2195
8	327	2148
9	347	2102
10	367	2057
11	387	2012
12	408	1968
13	429	1925

Note. F_1 and F_2 represent the first and second formants, respectively. Reprinted from P. Iverson & P. K. Kuhl, "Perceptual Magnet and Phoneme Boundary Effects in Speech Perception: Do They Arise From a Common Mechanism?" 2000, *Perception & Psychophysics, 62,* p. 879.

---

[2] Note that this is more extreme than the mean value of the /i/ category produced by male speakers in Peterson and Barney (1952), which would instead correspond to Stimulus 5.
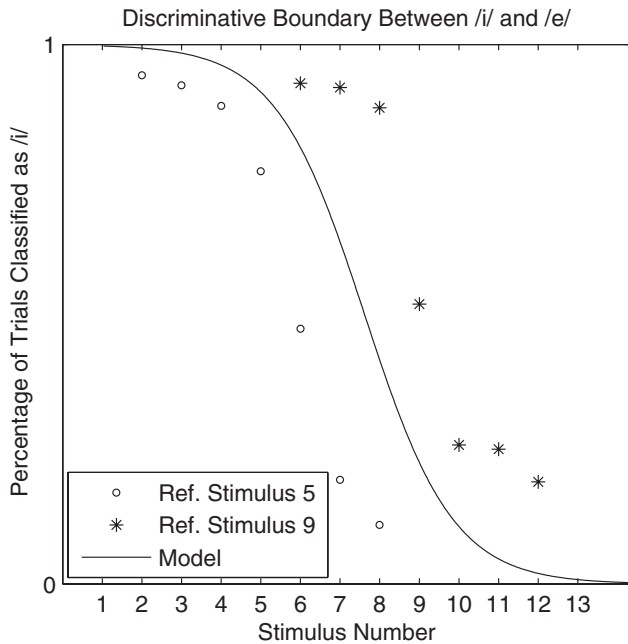
Figure 4. Identification percentages obtained by Lotto et al. (1998) with Reference (Ref.) Stimuli 5 and 9 were averaged to produce a single intermediate identification curve in the model (solid line). Adapted from A. J. Lotto, K. R. Kluender, & L. L. Holt, "Depolarizing the Perceptual Magnet Effect," 1998, *Journal of the Acoustical Society of America*, *103*, p. 3650. Copyright 1998 by the Acoustical Society of America. Reprinted with permission.

$$\sigma_c^2: 5{,}873 \ (\sigma_c = 77 \text{ mels})$$

$$\sigma_S^2: 4{,}443 \ (\sigma_S = 67 \text{ mels})$$

The fit obtained between the simulation and the empirical data is extremely close; however, the model parameters derived in this simulation are meant to serve only as a first approximation of the actual parameters in vowel perception. Because of the variability that has been found in subjects' goodness ratings of speech stimuli, it is likely that these parameters are somewhat off from their actual values, and it is also possible that the parameters vary between subjects. Instead, the simulation is a concrete demonstration that the model can reproduce empirical data on the perceptual magnet effect quantitatively as well as qualitatively using a reasonable set of parameters, supporting the viability of this rational account.

Effects of Frequency, Variability, and Noise

The previous section has shown a direct quantitative correspondence between model predictions and empirical data. In this section we explore the behavior of the rational model under various parameter combinations, using the parameters derived in the previous section as a baseline for comparison. These simulations serve a dual purpose: They establish the robustness of the qualitative behavior of the model under a range of parameters, and they make predictions about the types of variability that should occur when category frequency, category variance, and speech signal noise are varied. We first introduce several quantitative measures that can be used to visualize the extent of perceptual warping and subsequently use these measures to illustrate the effects of parameter manipulations.

Characterizing Perceptual Warping

Our statistical analysis establishes a simple function mapping a stimulus, *S*, to a percept of the intended target, given by $E[T|S]$. This is a linear mapping in the one-category case (Equation 7), but it becomes nonlinear in the case of multiple categories (Equation 11). Figure 5 illustrates the form of this mapping in the cases of one category and two categories with equal variance. Note that this function is not an identification function: The vertical axis represents the exact location of a stimulus in a continuous perceptual space, $E[T|S]$, not the probability with which that stimulus receives a particular label. Slopes that are more horizontal indicate that stimuli are closer in perceptual space than in acoustic space. In the two-category case, stimuli that are equally spaced in acoustic space are nevertheless clumped near category centers in perceptual space, as shown by the two nearly horizontal portions of the curve near the category means. In order to analyze this behavior more closely, we examine the relationship among three measures: *identification*, the posterior probability of category membership; *displacement*, the difference between the actual and perceived stimulus; and *warping*, the degree of shrinkage or expansion of perceptual space.

The identification function $p(c|S)$ gives the probability of a stimulus having been generated by a particular category, as calculated in Equation 8. This function is then used to compute the posterior on targets, summing over categories. In the case of two categories with equal variance, the identification function takes the form of a logistic function. Specifically, the posterior probability of category membership can be written as
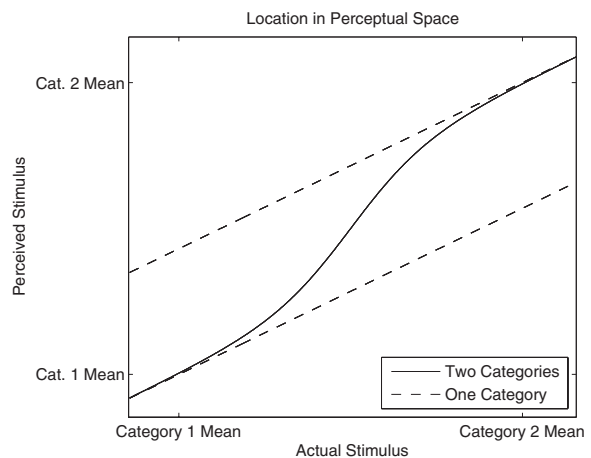

Figure 5. Model predictions for location of stimuli in perceptual space relative to acoustic space. Dashed lines indicate patterns corresponding to a single category; solid lines indicate patterns corresponding to two categories of equal variance. Cat. = Category.

$$p(c_1|S) = \frac{1}{1 + e^{-gS+b}}, \qquad (12)$$

where the gain and bias of the logistic are given by

$$g = \frac{\mu_1 - \mu_2}{\sigma_c^2 + \sigma_S^2}$$

and

$$b = \frac{\mu_1^2 - \mu_2^2}{2(\sigma_c^2 + \sigma_S^2)}.$$

An identification function of this form is illustrated in Figure 6a. In areas of certain categorization, the identification function is at either 1 or 0; a value of 0.5 indicates maximum uncertainty about category membership.

Displacement involves a comparison between the location of a stimulus in perceptual space $E[T|S]$ and its location in acoustic space S. It corresponds to the amount of bias in perceiving a stimulus. We can calculate this quantity as

$$E[T|S] - S = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_S^2}S + \frac{\sigma_S^2}{\sigma_c^2 + \sigma_S^2}\sum_c p(c|S)\mu_c - S$$

$$= \frac{\sigma_S^2}{\sigma_c^2 + \sigma_S^2}\left(\sum_c p(c|S)\mu_c - S\right). \qquad (13)$$

In the one-category case, this means that the amount of displacement is proportional to the distance between the stimulus S and the mean μ_c of the category. As stimuli get farther away from the category mean, they are pulled proportionately farther toward the center of the category. The dashed lines in Figure 6b show two cases of this. In the case of multiple categories, the amount of displacement is proportional to the distance between S and a weighted average of the means μ_c of more than one category. This is shown in the solid line, where ambiguous stimuli are displaced less than would be predicted in the one-category case because of the competing influence of a second category mean.

Finally, perceptual warping can be characterized by the distance between two neighboring points in perceptual space that are separated by a fixed step ΔS in acoustic space. This quantity is reflected in the distance between neighboring points on the bottom layer of each diagram in Figure 2. By the standard definition of the derivative as a limit, as ΔS approaches zero this measure of perceptual warping corresponds to the derivative of $E[T|S]$ with respect to S. This derivative is

$$\frac{dE[T|S]}{dS} = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_S^2} + \frac{\sigma_S^2}{\sigma_c^2 + \sigma_S^2}\sum_c \mu_c \frac{dp(c|S)}{dS}, \qquad (14)$$

where the last term is the derivative of the logistic function given in Equation 12. This equation demonstrates that distance between two neighboring points in perceptual space is a linear function of the rate of change of $p(c|S)$, which measures category membership of stimulus S. Probabilities of category assignments are changing most rapidly near category boundaries, resulting in greater perceptual distances between neighboring stimuli near the edges of categories. This is shown in Figure 6c,

(a) Category Identification

(b) Stimulus Displacement

(c) Warping of Perceptual Space
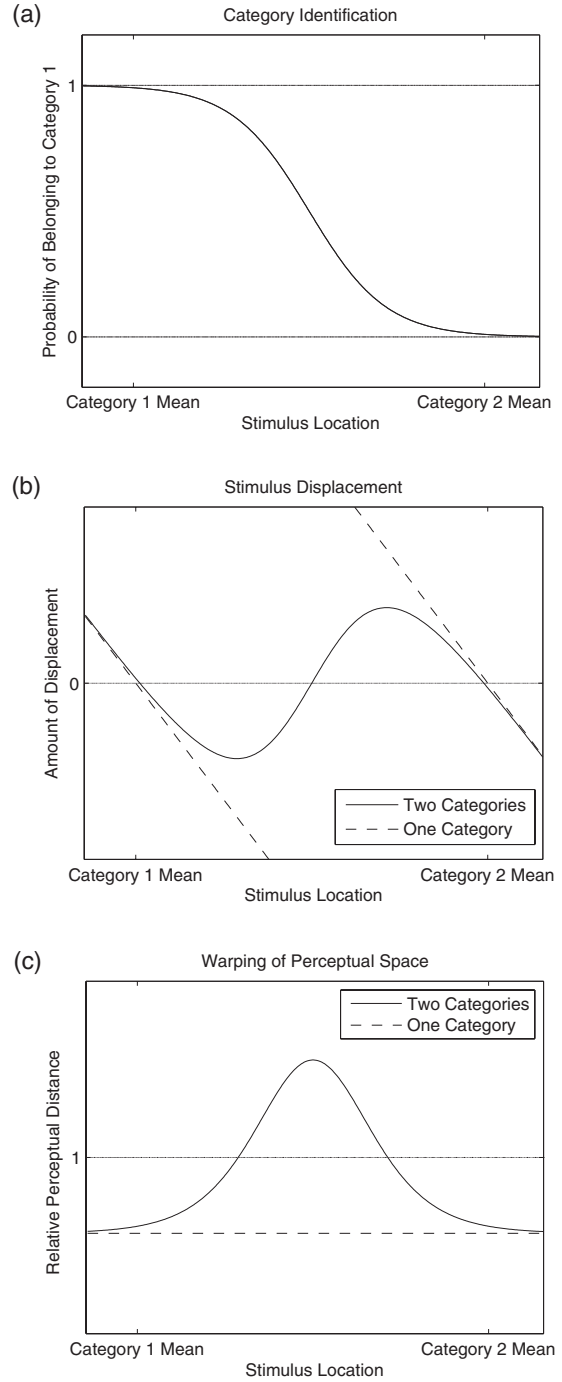
Figure 6. Model predictions for (a) identification, (b) displacement, and (c) warping. Dashed lines indicate patterns corresponding to a single category; solid lines indicate patterns corresponding to two categories of equal variance.

and the form of the derivative is described in more detail in Appendix C.

In summary, the identification function (Equation 12) shows a sharp decrease at the location of the category boundary, going from a value near one (assignment to Category 1) to a value

near zero (assignment to Category 2). Perceptual bias, or displacement (Equation 13), is a linear function of distance from the mean in the one-category case but is more complex in the two-category case; it is positive when stimuli are displaced in a positive direction and negative when stimuli are displaced in a negative direction. Finally, warping of perceptual space (Equation 14), which has a value greater than one in areas where perceptual space is expanded and a value less than one in areas where perceptual space is shrunk, shows that all of perceptual space is shrunk in the one-category case but that there is an area of expanded perceptual space between categories in the two-category case. Qualitatively, we note that displacement is always in the direction of the most probable category mean and that the highest perceptual distance between stimuli occurs near category boundaries. This is compatible with the idea that categories function like perceptual magnets and also with the observation that perceptual space is shrunk most in the centers of phonetic categories. In the remainder of this section, we use these measures to explore the model's behavior under various parameter manipulations that simulate changes in phonetic category frequency, within-category variability, and speech signal noise.

Frequency

Manipulating the frequency of phonetic categories corresponds in our model to manipulating their prior probability. This manipulation causes a shift in the discriminative boundary between two categories, as described in Appendix B. In Figure 7a, the boundary is shifted toward the category with lower prior probability so that a larger region of acoustic space between the two categories is classified as belonging to the category with higher prior probability. Figure 7b shows that when the prior probability of Category 1 is increased, most stimuli between the two categories are shifted in the negative direction toward the mean of that category. This occurs because more sounds are classified as being part of Category 1. Decreasing the prior probability of Category 1 yields a similar shift in the opposite direction. Figure 7c shows that the location of the expansion of perceptual space follows the shift in the category boundary.

This shift qualitatively resembles the boundary shift that has been documented on the basis of lexical context (Ganong, 1980). In contexts where one phoneme would form a lexical item and the other would not, phoneme boundaries are shifted toward the phoneme that makes the nonword, so that more of the sounds between categories are classified as the phoneme that would yield a word. Similar effects have also been found for lexical frequency (Connine, Titone, & Wang, 1993) and phonotactic probability (Massaro & Cohen, 1983; Pitt & McQueen, 1998). To model such a shift using the rational model, information about a specific lexical or phonological context needs to be encoded in the prior $p(c)$. The prior distribution would thus reflect the information about the frequency of occurrence of a phonetic category in a specific context. The rational model then predicts that the boundary shift can be modeled by a bias term of magnitude $\log \frac{p(c_2)}{p(c_1)}$ and that the peak in discrimination should shift together with the category boundary.
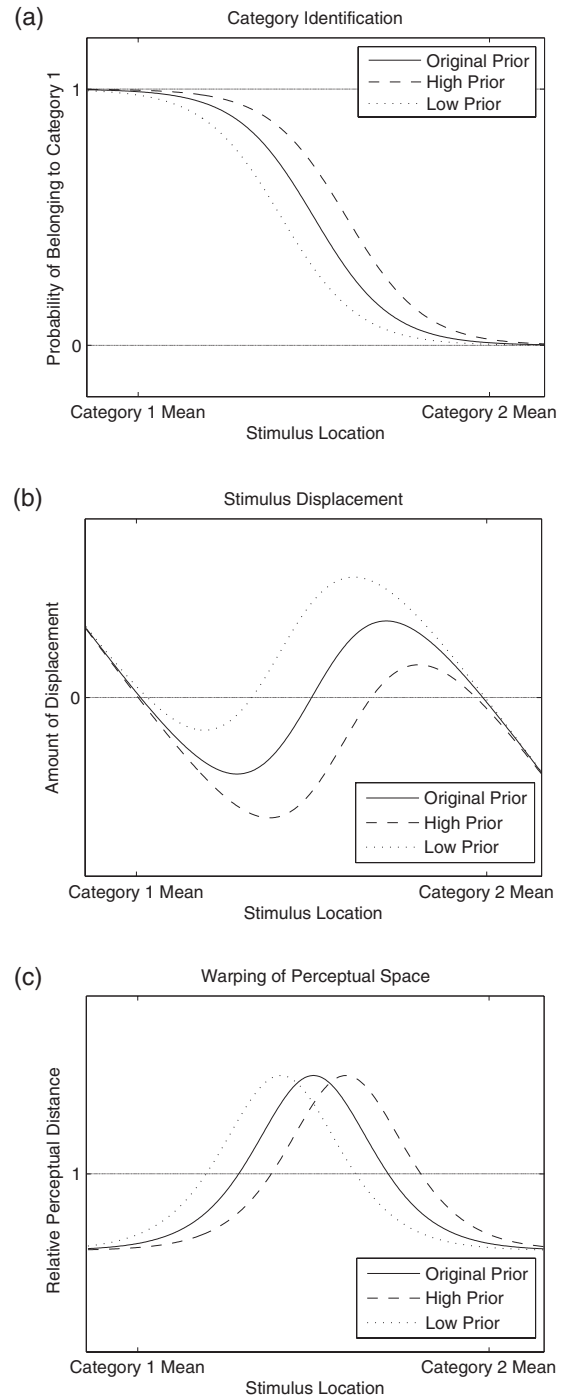
Figure 7. Effects of prior probability manipulation on (a) identification, (b) displacement, and (c) warping. The prior probability of category 1, $p(c_1)$, was either increased or decreased while all other model parameters were held constant.

Variability

The category variance parameter indicates the amount of meaningful variability that is allowed within a phonetic category. One correlate of this might be the amount of coarticulation that a

category allows: Categories that undergo strong coarticulatory effects have high variance, whereas categories that are resistant to coarticulation have lower variance.[3] In the model, categories with high variability should differ from categories with low variability in two ways. First, the discriminative boundary between the categories should be either shallow, in the case of high variability, or sharp, in the case of low variability (Figure 8a). This means that listeners should be nearly deterministic in inferring which category produced a sound in the case of low variability, whereas they should be more willing to consider both categories if the categories have high variability. This pattern has been demonstrated empirically by Clayards, Tanenhaus, Aslin, and Jacobs (2008), who showed that the steepness of subjects' identification functions along a /p/–/b/ continuum depends on the amount of category variability in the experimental stimuli.

In addition to this change in boundary shape, the rational model predicts that the amount of variability should affect the weight given to the category means relative to the stimulus S when perceiving acoustic detail. Less variability within a category implies a stronger constraint on the sounds that the listener expects to hear, and this gives more weight to the category means. This should cause more extreme shrinkage of perceptual space in categories with low variance.

These two factors should combine to yield extremely categorical perception in categories with low variability and perception that is less categorical in categories with high variability. Figure 8b shows that displacement has a higher magnitude than baseline for stimuli both within and between categories when category variance is decreased. Displacement is reduced with higher category variance. Figure 8c shows the increased expansion of perceptual space between categories and the increased shrinkage within categories that result from low category variance. In contrast, categories with high variance yield more veridical perception.

Differences in category variance might explain why it is easier to find perceptual magnet effects in some phonetic categories than in others. According to vowel production data from Hillenbrand et al. (1995), reproduced here in Figure 1, the /i/ category has low variance along the dimension tested by Iverson and Kuhl (1995). The difficulty in reproducing the effect in other vowel categories might be partly attributable to the fact that listeners have weaker prior expectations about which vowel sounds speakers might produce within these categories.

This parameter manipulation can also be used to explore the limits on category variance: The rational model places an implicit upper limit on category variance if one is to observe enhanced discrimination between categories. This limit occurs when categories are separated by less than two standard deviations, that is, when the standard deviation increases to half the distance to the neighboring category. When the category variance reaches this point, the distribution of speech sounds in the two categories becomes unimodal, and the acquired distinctiveness between categories disappears. Instead of causing enhanced discrimination at the category boundary, noise now causes all speech sounds to be pulled inward toward the space between the two category means, as illustrated in Figure 9. Shrinkage of perceptual space may be slightly less between categories than within categories, but all of perceptual space is pulled toward the center of the distribution. This perceptual pattern resembles the pattern that would be predicted if these speech sounds all derived from a single category,
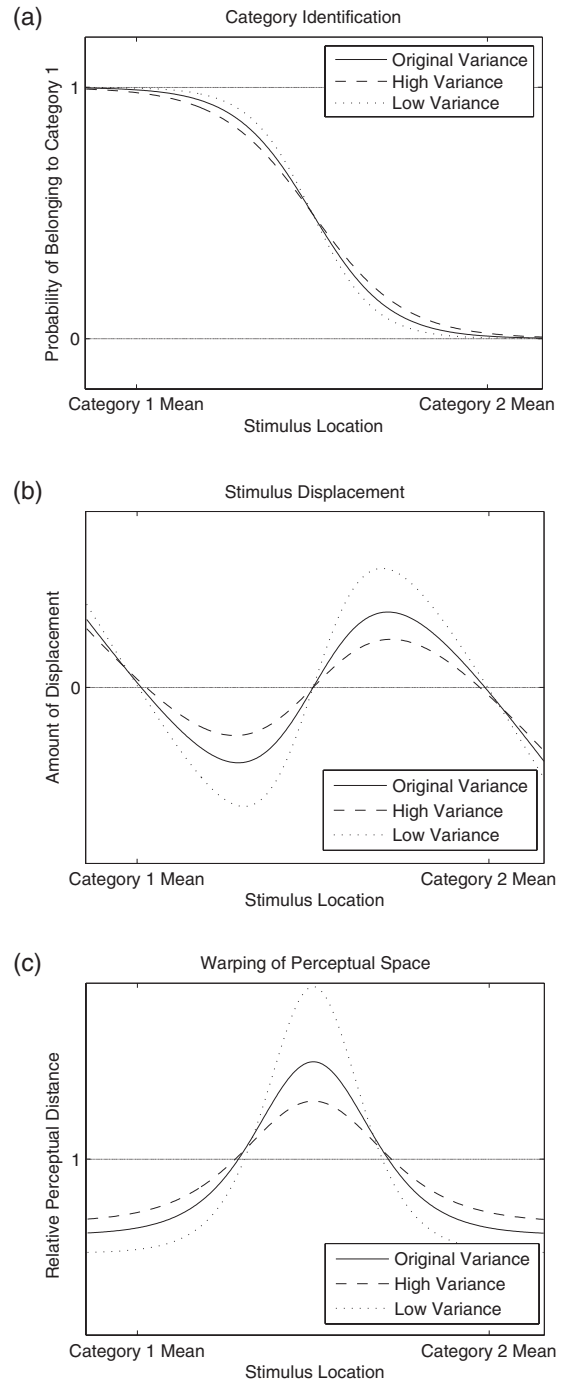
Figure 8. Effects of category variance on (a) identification, (b) displacement, and (c) warping. The category variance parameter σ_c^2 was either increased or decreased while all other model parameters were held constant.

---

[3] Coarticulatory effects are context dependent rather than being an inherent property of specific phonetic categories. However, listeners should be able to estimate the typical range of coarticulation that occurs within specific contexts and thus obtain a context-specific estimate of category variance.

Actual Stimulus

Perceived Stimulus
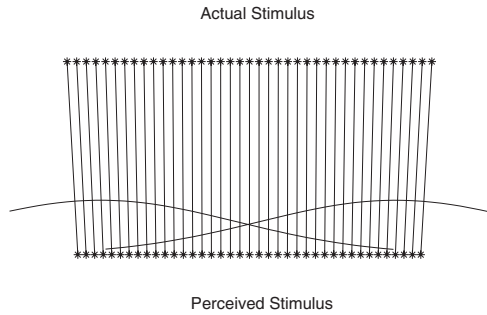
Figure 9. Categories that overlap to form a single unimodal distribution act perceptually like a single category: Speech sounds are pulled toward a point between the two categories.

indicating that it is the distribution of speech sounds in the input, rather than the explicit category structure, that produces perceptual warping in the model.

Noise

Manipulating the speech signal noise also affects the optimal solution in two different ways. More noise means that listeners should be relying more on prior category information and less on the speech sound they hear, yielding more extreme shrinkage of perceptual space within categories. However, adding noise to the speech signal also makes the boundary between categories less sharp so that in high-noise environments, listeners are uncertain of speech sounds' category membership (Figure 10a). This combination of factors produces a complex effect: Whereas adding low levels of noise makes perception more categorical, there comes a point where noise is too high to determine which category produced a speech sound, blurring the boundary between categories.

With very low levels of speech signal noise, perception is only slightly biased (Figure 10b), and there is a very low degree of shrinkage and expansion of perceptual space (Figure 10c). This occurs because the model relies primarily on the speech sound in low-noise conditions, with only a small influence from category information. As noise levels increase to those used in the simulation in the previous section, the amount of perceptual bias and warping both increase. With further increases in speech signal noise, however, the shallow identification function begins to interfere with the availability of category information. For unambiguous speech sounds, displacement and shrinkage are both increased, as shown at the edges of the graphs in Figure 10. However, this does not simultaneously expand perceptual space between the categories. Instead, the high uncertainty about category membership causes reduced expansion at points between categories, dampening the difference between between-category and within-category discriminability.

The complex interaction between perceptual warping and speech signal noise suggests that there is some level of noise for which one would measure between-category discriminability as much higher than within-category discriminability. However, for very low levels of noise and for very high levels of noise, this difference would be much less noticeable. This suggests a possible explanation for variability that has been found in perceptual warping even among studies that have examined the English /i/ cate-

gory (e.g., Lively & Pisoni, 1997). Extremely low levels of ambient noise should dampen the perceptual magnet effect, whereas the effect should be more prominent at higher levels of ambient noise.

A further prediction regarding speech signal noise concerns its effect on boundary shifts. As discussed above, the rational model
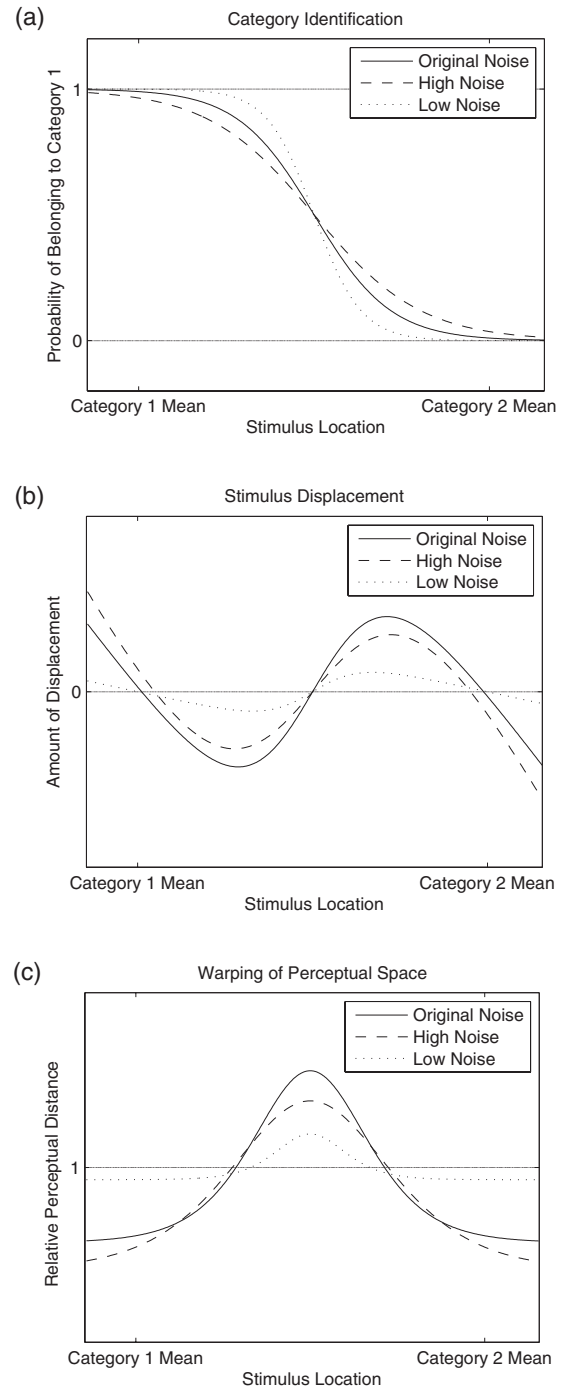

Figure 10. Effects of speech signal noise on (a) identification, (b) displacement, and (c) warping. The speech signal noise parameter σ_s^2 was either increased or decreased while all other model parameters were held constant.

predicts that when prior probabilities $p(c)$ are different between two categories, there should be a boundary shift caused by a bias term of $\log\dfrac{p(c_2)}{p(c_1)}$. This bias term produces the largest boundary shift for small values of the gain parameter, which correspond to a shallow category boundary (see Appendix B). High noise variance produces this type of shallow category boundary, giving the bias term a large effect. This is illustrated in Figure 11, where for constant changes in prior probability, larger boundary shifts occur at higher noise levels. This prediction qualitatively resembles data on lexically driven boundary shifts: Larger shifts occur when stimuli are low-pass filtered (McQueen, 1991) or presented in white noise (Burton & Blumstein, 1995).

Summary

Simulations in this section have shown that the qualitative perceptual patterns predicted by the rational model are the same under nearly all parameter combinations. The exceptions to this are the case of no noise, in which perception should be veridical, and the case of extremely high category variance or extremely high noise, in which listeners cannot distinguish between the two categories and effectively treat them as a single, larger category. In addition, these simulations have examined three types of variability in perceptual patterns. Shifts in boundary location occur in the model as a result of changes in the prior probability of a phonetic category, and these shifts mirror lexical effects that have been found empirically (Ganong, 1980). Differences in the degree of categorical perception in the model depend on the amount of meaningful variability in a category, and these predictions are consistent with the observation that the /i/ category has low variance along the relevant dimension. Finally, the model predicts effects of ambient noise on the degree of perceptual warping, a methodological detail that might explain the variability of perceptual patterns under different experimental conditions.
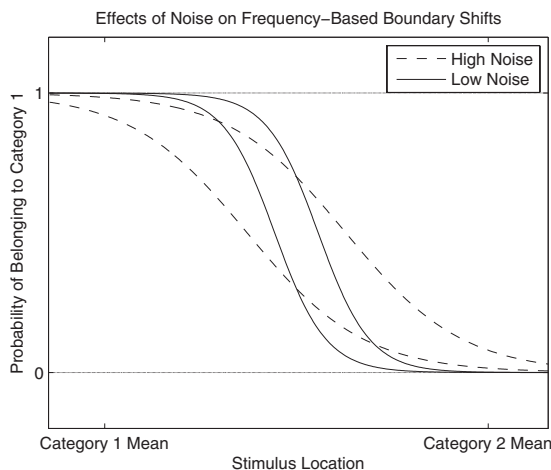

Figure 11. Effects of speech signal noise on the magnitude of a boundary shift. Simulations at both noise levels used prior probability values for c_1 of 0.3 (left boundary) and 0.7 (right boundary). The boundary shift is nevertheless larger for higher levels of speech signal noise.

Testing the Predicted Effects of Noise

Simulations in the previous section suggested that ambient noise levels might be partially responsible for the contradictory evidence that has been found in previous empirical studies of the perceptual magnet effect. In this section, we present an experiment to test the model's predictions with respect to changes in speech signal noise. The rational model makes two predictions about the effects of noise. The first prediction is that noise should yield a shallower category boundary, making it difficult at high noise levels to determine which category produced a speech sound. This effect should lower the discrimination peak between categories at very high levels of noise and is predicted by any model in which noise increases the variance of speech sounds from a phonetic category. The second prediction is that listeners should weight acoustic and category information differentially depending on the amount of speech signal noise. As noise levels increase, they should rely more on category information, and perception should become more categorical. This effect is predicted by the rational model but not by other models of the perceptual magnet effect, as discussed in detail later in the article. Although this effect is overshadowed by the shallow category boundary at very high noise levels, examining low and intermediate levels of noise allows us to test this second prediction.

Previous research into effects of uncertainty on speech perception has focused on the role of memory uncertainty. Pisoni (1973) found evidence that within-category discrimination, in comparison with between-category discrimination, shows a larger decrease in accuracy with longer interstimulus intervals. Pisoni interpreted these results as evidence that within-category discrimination relies on acoustic (rather than phonetic) memory and that acoustic memory traces decay with longer interstimulus intervals. Iverson and Kuhl (1995) also investigated the perceptual magnet effect at three different interstimulus intervals; though they did not explicitly discuss changes in warping related to interstimulus interval, within-category clusters appear to be tighter in their 2,500-ms condition than in their 250-ms condition. These results are consistent with the idea that memory uncertainty increases with longer interstimulus intervals.

Several studies have also studied asymmetries in discrimination, under the assumption that memory decay will have a greater effect on the stimulus that is presented first. However, many of these studies have produced contradictory results, making the effects of memory uncertainty difficult to interpret (Polka & Bohn, 2003; Repp & Crowder, 1990). Furthermore, data from Pisoni (1973) indicate that longer interstimulus intervals do not necessarily increase uncertainty: Discrimination performance was worse with a 0-ms interstimulus interval than with a 250-ms interstimulus interval.

Adding white noise is a more direct method of introducing speech signal uncertainty, and its addition to speech stimuli has consistently been shown to decrease subjects' ability to identify stimuli accurately. Subjects make more identification errors (G. A. Miller & Nicely, 1955) and display a shallower identification function (Formby, Childers, & Lalwani, 1996) with increased noise, consistent with the rational model's predictions. Although it is known that subjects rely to some extent on both temporal and spectral cues in noisy conditions (Xu & Zheng, 2007), it is not known how reliance on these acoustic cues compares with reliance on prior information about category structure. To test whether reliance on category information is greater in higher noise condi-

tions than in lower noise conditions, we replicated Experiment 3 of Iverson and Kuhl (1995)—their multidimensional scaling experiment—with and without the presence of background white noise.

The rational model predicts that perceptual space should be distorted to different degrees in the noise and no-noise conditions. At moderate levels of noise, we should observe more perceptual warping than with no noise because of higher reliance on category information. At very high noise levels, however, if subjects are unable to make reliable category assignments, warping should decrease; as noted, this decrease is predicted by any model in which subjects are using category membership to guide their judgments. Thus, whereas the model is compatible with changes in both directions for different noise levels, our aim is to find levels of noise for which warping is higher with increased speech signal noise. Moreover, manipulating the noise parameter in the rational model should account for behavioral differences due to changing noise levels.

Method

Subjects. Forty adult participants were recruited from the Brown University community. All were native English speakers with no known hearing impairments. Participants were compensated at a rate of $8 per hour. Data from two additional participants were excluded, one because of equipment failure and one because of failure to understand the task instructions.

Apparatus. Stimuli were presented through noise cancellation headphones, Bose Aviation Headset model AHX-02, from a computer at comfortable listening levels. Subjects' responses were entered and recorded using the computer that presented the stimuli. The presentation of the stimuli was controlled through Bliss software (Mertus, 2004), developed at Brown University for use in speech perception research.

Stimuli. Thirteen /i/ and /e/ stimuli, modeled after the stimuli in Iverson and Kuhl (1995), were created with the KlattWorks software (McMurray, 2009). Stimuli varied along a single F_1–F_2 vector that ranged from an F_1 of 197 Hz and an F_2 of 2489 Hz to an F_1 of 429 Hz and an F_2 of 1925 Hz. The stimuli were spaced at equal intervals of 30 mels; exact formant values are shown in Table 1. For all stimuli, F_3 was set at 3010 Hz, F_4 at 3300 Hz, and F_5 at 3850 Hz. The bandwidths for the five formants were 53, 77, 111, 175, and 281 Hz. Each stimulus was 435 ms long. Pitch rose from 112 to 130 Hz over the first 100 ms and dropped to 92 Hz over the remainder of the stimulus. Stimuli were normalized in Praat (Boersma, 2001) to have a mean intensity of 70 dB.

For stimuli in the noise condition, we created 435 ms of white noise using Praat by sampling randomly from a uniform [−0.5, 0.5] distribution at a sampling rate of 11025 Hz. The mean intensity of this waveform was then scaled to 70 dB. The white noise was added to each of the 13 stimuli, creating a set of stimuli with a zero signal-to-noise ratio.

Procedure. Subjects were assigned to either the no-noise or the noise condition. After reading and signing a consent form, they completed ten practice trials designed to familiarize them with the task and stimuli and subsequently completed a single block of 208 trials. This block included 52 "same" trials, four trials for each of 13 stimuli, and 156 "different" trials in which all possible ordered pairs of nonidentical stimuli were presented once each. In each trial, participants heard two stimuli sequen-

tially with a 250-ms interstimulus interval. They were instructed to respond as quickly as possible, pressing one button if the two stimuli were identical and another button if they could hear a difference between the two stimuli. Responses and reaction times were recorded.

This procedure was nearly identical to that used by Iverson and Kuhl (1995), though the response method differed slightly in order to provide reaction times for same responses in addition to different responses. We also eliminated the response deadline of 2000 ms and instead recorded subjects' full reaction times for each contrast, up to 10,000 ms.

Results and Discussion

Of the 8,320 responses, 14 were excluded from the analysis because subjects either responded before hearing the second stimulus or failed to respond altogether within the 10–s response period. Table 2 shows the percentage of the remaining trials on which subjects responded same for each contrast. As expected, the percentage of same responses was extremely high for one-step discriminations and got successively lower as the psychoacoustic distance between stimuli increased. This correlation was significant in a by-item analysis for both the no-noise ($r = -0.85$, $p < .01$) and the noise ($r = -0.87$, $p < .01$) condition.[4]

Figure 12a shows these confusion data schematically, where darker squares indicate a higher percentage of same responses. This schematic representation highlights three differences between the conditions. First, the overall percentage of same responses was higher in the noise condition than in the no-noise condition, as evidenced by the higher number of dark squares. Second, the percentage of same responses declined more slowly in the noise condition than in the no-noise condition with increasing psychophysical distance, as reflected by a more gradual change from dark squares to light squares in the noise condition. Third, the difference between within-category and between-category contrasts was greater in the noise condition than in the no-noise condition. Whereas the no-noise condition showed fairly constant performance along any given diagonal, with only a small dip in the percentage of same responses toward the center of the stimulus continuum, the noise condition showed a much larger difference along the diagonal, with a strong decrease in same responses near the between-category contrasts at the center of the stimulus continuum. This third difference suggests that there is a larger degree of within-category shrinkage and between-category expansion of perceptual space in the noise condition, consistent with the predictions of the rational model.

Same–different model. We used the rational model to simulate these confusion data, assuming that participants perceive speech sounds by sampling a target production from the posterior distribution on target productions, $p(T|S)$. We extended the model to account directly for same–different responses by assuming that participants respond same if the sampled target productions for the two speech sounds are within a threshold distance ϵ of each other; otherwise they respond different. The parameter ϵ thus played a similar role to the response criterion

---

[4] All statistical significance tests reported in this article are two-tailed.

Table 2
Percentage of Trials on Which Subjects Responded "Same" for Each Pair of Stimuli in the No-Noise and Noise Conditions

Stimulus no.	1	2	3	4	5	6	7	8	9	10	11	12	13
					No-noise condition								
1	98.8	82.5	82.5	40.0	22.5	7.5	5.0	5.0	0.0	0.0	2.5	0.0	2.5
2		97.5	95.0	70.0	52.5	10.0	5.0	0.0	2.5	2.5	0.0	0.0	0.0
3			91.3	97.5	75.0	32.5	12.5	5.0	2.5	0.0	2.5	2.5	0.0
4				97.5	87.5	40.0	12.5	5.0	2.5	0.0	2.5	0.0	0.0
5					97.5	77.5	27.5	12.5	5.0	2.5	0.0	0.0	0.0
6						92.5	75.0	30.0	15.0	2.5	2.5	2.6	0.0
7							91.3	75.0	42.5	17.5	5.0	5.0	0.0
8								95.0	80.0	50.0	32.5	7.5	5.0
9									93.8	87.5	67.5	27.5	22.5
10										92.5	87.5	76.9	37.5
11											97.5	87.5	65.0
12												96.3	97.5
13													100.0
					Noise condition								
1	95.0	95.0	87.5	80.0	82.5	57.5	25.0	7.0	5.0	0.0	0.0	5.0	2.5
2		96.3	97.5	97.5	87.5	80.0	42.5	15.0	5.0	5.0	0.0	2.5	2.5
3			95.0	97.5	90.0	80.0	42.5	30.0	7.5	0.0	2.5	2.5	0.0
4				92.5	95.0	90.0	55.0	20.0	10.0	7.5	2.5	2.5	7.5
5					95.0	90.0	67.5	27.5	5.0	12.5	2.5	15.0	10.0
6						96.3	87.5	50.0	25.0	7.5	2.5	10.0	2.5
7							91.3	75.0	42.5	20.0	12.5	12.5	10.0
8								87.5	72.5	52.5	40.0	20.0	10.0
9									90.0	92.5	72.5	47.5	37.5
10										93.8	95.0	85.0	52.5
11											95.0	100.0	85.0
12												95.0	97.5
13													95.0

of the observer in signal detection theory (Green & Swets, 1966), determining the magnitude of a difference that will yield a positive response. Under this model, the number of same responses to a given contrast is predicted to follow a binomial distribution $B(n, p)$ where n is the number of trials in which a given contrast was presented and p is the probability that the two sampled target productions for that contrast are within a distance ϵ of each other, $p(|T_A - T_B| \leq \epsilon|S_A, S_B)$. This probability can be computed as described in Appendix D.

The simulation used the same category means $\mu_{/i/}$ and $\mu_{/e/}$ and category variance σ_c^2 as the simulation of the Iverson and Kuhl (1995) data. The noise variance was a free parameter that could vary between conditions to capture differences in perceptual warping; in addition, the decision threshold ϵ was a free parameter that could vary between the two conditions, allowing the model to capture the overall greater number of same responses in the noise condition. These free parameters were chosen to maximize the likelihood of the same–different data. The best fitting model used parameters of $\epsilon = 76$ mels and $\sigma_S^2 = 878$ ($\sigma_S = 30$ mels) for the no-noise condition and $\epsilon = 111$ mels and $\sigma_S^2 = 2129$ ($\sigma_S = 46$ mels) for the noise condition. Using these parameters, we found that the percentage of same responses predicted by the model for each contrast was highly correlated with that found empirically ($r = .98$ for the no-noise condition; $r = .97$ for the noise condition), and these correlations remained high even after controlling for acoustic distance ($r = .94$ and $r = .87$ for the no-noise and noise conditions, respectively). Model performance is shown schematically in Figure 12b.

The key prediction for this experiment was that the noise variance parameter σ_S^2 could account for differences in performance between the no-noise and noise conditions. However, in the above simulation, ϵ was an additional free parameter that could vary between conditions. To demonstrate quantitatively that the noise parameter accounted for differences above and beyond those accounted for simply by varying the decision threshold, we used a generalized likelihood ratio test (e.g., Rice, 1995) to compare the full model described above with a restricted model (Figure 12c) in which the noise parameter was constant across conditions. Like the full model, the restricted model used category means and the category variance from the previous simulations, and the decision threshold was a free parameter that could vary between the two conditions.[5] The models differed only in their assumptions about the noise parameter. These two models thus constitute a nested hierarchy, and we can determine whether the additional noise parameter makes a statistically significant difference by examining the difference between the log likelihoods of the models, computed using the maximum likelihood estimates of the parameters. Under the null hypothesis that the data were generated from the restricted model, twice this difference has a $\chi^2(1)$ distribution. The log

---

[5] Constraining ϵ to be the same between the two conditions significantly decreases the likelihood of the data; however, even under the assumption of a constant threshold, allowing the speech signal noise parameter to vary between conditions makes a statistically significant difference.
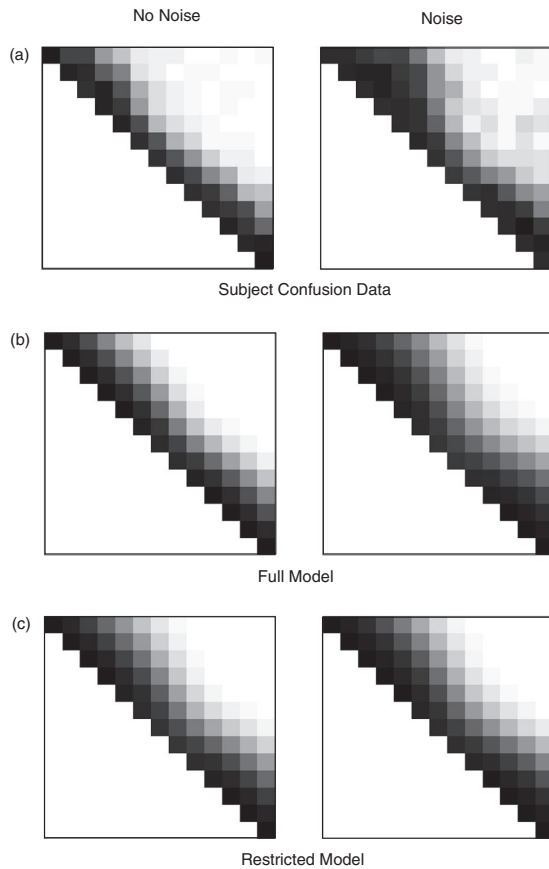
No Noise

Noise

(a)

Subject Confusion Data

(b)

Full Model

(c)

Restricted Model

Figure 12. Confusion matrices showing the percentage of same responses to each contrast in (a) subject data, (b) the full model in which the noise parameter varied between conditions, and (c) the restricted model in which the noise parameter was constrained to be the same between conditions. The left-hand plots show the no-noise condition, and the right-hand plots show the noise condition. Darker cells indicate a higher percentage of same responses.

likelihood of the data was -676 under the restricted model[6] and -568 under the full model. The full model therefore accounted for these data significantly better than did the restricted model, $\chi^2(1) = 216, p < .001$; allowing the noise parameter to change between the noise and no-noise conditions resulted in a statistically significant improvement in fit.

This comparison indicates that the rational model accounts for additional differences between conditions beyond the overall increase in same responses. As noted earlier, there are two such differences apparent in the data. First, the decrease in same responses with psychophysical distance is more gradual in the noise condition than in the no-noise condition. In the rational model, this occurs because listeners in the noise condition assume that the speech sound might have come from a wider range of target productions, leading to higher variability in the posterior distribution (Equation 25). Higher posterior variance leads to a shallower decline in same responses. Second, the responses are more categorical in the noise condition than in the no-noise condition, as evidenced by response patterns along each diagonal. This occurs in the rational model because of increased weighting of category information in higher noise conditions (Equation 7).

Although both of these aspects of the data are compatible with the rational model, a straightforward alternative explanation is available for the first. In modeling these data we have made the assumption that the stimulus heard by experimental participants is identical to the stimulus played. This assumption allows the use of known stimulus values S when computing listeners' optimal percepts. However, in reality there is likely to be some variability in the stimuli heard by listeners, and this variability should be higher in the noise condition than in the no-noise condition. The shallow decrease in same responses in the noise condition might then be a simple result of higher stimulus variability. Taking into account experimental noise might improve the performance of the restricted model by providing a mechanism to account for this shallower decrease in same responses in the noise condition.

To investigate this possibility, we simulated experimental noise in the restricted model by drawing values of S, the speech sound heard by listeners, from a Gaussian distribution centered around each stimulus value. The probability of a same response for a given contrast was approximated by drawing 100 samples of each speech sound in the pair and computing the probability of a same response for each pair of samples. These probabilities were then averaged to obtain the expected probability of a same response for each contrast, and a binomial model was used to compute the likelihood of the data. The experimental noise variance was a free parameter that varied between the two conditions, under the assumption that listeners in the two conditions heard the stimuli through different amounts of noise. A third noise parameter that governed listeners' inferences was held constant between the two conditions, as in the restricted model, implementing an assumption that listeners weight category information equally in the two conditions. This model yielded a log likelihood of -618, significantly higher than the restricted model described above, $\chi^2(2) = 116, p < .001$ but lower than the full model despite having one more free parameter.[7] The remaining difference in likelihood between this model and the full model reflects listeners' increased reliance on category information in higher noise conditions, as captured by our rational model.

Multidimensional scaling. The two noise parameters used in the simulation of our confusion data were both lower than the noise variance estimated on the basis of the Iverson and Kuhl (1995) data. However, the ambient noise level in Iverson and Kuhl's experiment should have been comparable to that of our no-noise condition and was almost certainly lower than the zero signal-to-noise ratio in our noise condition. This discrepancy may reflect a difference in analysis methods. Whereas Iverson and Kuhl used multidimensional scaling to analyze their results, we based

---

[6] The maximum likelihood parameters for the restricted model were $\epsilon = 85$ mels and $\epsilon = 103$ mels for the no-noise and noise conditions, respectively, and $\sigma_S^2 = 1447$ ($\sigma_S = 38$ mels).

[7] This model cannot be compared with the full model in a generalized likelihood ratio test because the two models are not nested. To make a nested variant of the full model, we augmented it with the same two free parameters for experimental noise. This augmented full model had a log likelihood of -568. It therefore accounted for the data significantly better than did the augmented restricted model, $\chi^2(1) = 100, p < .001$, though it did not yield any improvement over the original full model. This again indicates that allowing the inference-related noise parameter to differ between the two conditions results in a statistically significant improvement in fit.

our analysis directly on subject confusion data. To draw a closer comparison to the results from Iverson and Kuhl and to further help visualize the difference between the noise and no-noise conditions, we used multidimensional scaling to create perceptual maps from the behavioral data.

Our multidimensional scaling analysis incorporated information from both reaction times and same–different responses. Reaction time data were normalized across subjects by first taking the log transform to ensure normal distributions and then converting these to z-scores for each subject. Psychoacoustic distance had a significant positive correlation with these normalized reaction times for same responses ($r = .45$, $p < .01$, for the no-noise condition; $r = .27$, $p < .02$, for the noise condition),[8] reflecting the predicted result that subjects who responded same were slower when the stimuli were separated by a greater psychoacoustic distance. Conversely, the data showed a significant negative correlation between psychoacoustic distance and normalized reaction times on different responses ($r = -.69$, $p < .01$, for the no-noise condition; $r = -.56$, $p < .01$, for the noise condition), indicating that subjects were faster to respond different when the stimuli were farther apart in psychoacoustic space. Both same and different reaction times were therefore included as measures of perceptual distance in our multidimensional scaling analysis.

The intuition behind our multidimensional scaling analysis, which is supported by the correlations presented above, is that reaction times and same–different responses are consistent with a subject's perceptual map of the stimuli. Different responses with short reaction times indicate that stimuli are far apart in this perceptual map; different responses with long reaction times indicate that stimuli are closer together; same responses with long reaction times indicate that stimuli are even closer, and same responses with short reaction times indicate that stimuli are extremely close together in the perceptual map. Nonmetric multidimensional scaling (Shepard, 1980) is an optimization method that aims to minimize violations of distance rankings in a perceptual map. It assumes a monotonic relation between reaction times and perceptual distance but does not assume any parametric form for this relation.[9]

We constructed a similarity matrix for each condition that mirrored these intuitions. This was implemented computationally by subtracting z-scores for same responses from a z-score of six,[10] effectively transforming same responses into different responses with extremely long reaction times, such that shorter reaction times on a same response mapped onto longer reaction times on a different response. This is similar to the procedure used by Iverson and Kuhl, who substituted a reaction time of 2000 ms (the trial length in their experiment) for any same response. The median score across subjects for each contrast was then entered into the similarity matrix, and scores were normalized to fall between zero and one.

Nonmetric multidimensional scaling solutions based on these similarity matrices are shown in Figure 13. The plots are modeled after Figure 5: The horizontal axis shows acoustic space, and the vertical axis shows perceptual space. A linear function would indicate a linear mapping between acoustic and perceptual space, whereas nonlinearities suggest that perceptual space is warped relative to acoustic space. Areas that are more nearly horizontal indicate greater shrinkage of perceptual space. These multidimensional scaling solutions suggest that there is a difference in sub-

jects' perceptual maps between the two conditions. Consistent with results from Iverson and Kuhl, there is some evidence of perceptual warping in the no-noise condition, but here interstimulus distances are relatively constant. As predicted by our model, perceptual space is more warped in the noise condition than in the no-noise condition. Unambiguous stimuli near category centers are very close together in perceptual space, whereas stimuli near the category boundary are much farther apart. The precise stimulus locations in these multidimensional scaling solutions are not compatible with the parameters used for the simulation of raw confusion data, suggesting that multidimensional scaling yields an imperfect perceptual map of the stimuli. It is possible that Iverson and Kuhl's (1995) multidimensional scaling analysis produced a parallel exaggeration of the degree of warping, yielding the discrepancy in noise parameters discussed above. However, the multidimensional scaling solution illustrates the same qualitative difference between the conditions as is seen in the raw confusion data: Subjects in the noise condition relied more on category information than subjects in the no-noise condition.

As predicted for moderate noise levels, we observed increased perceptual warping with increased speech signal noise. These results provide evidence that listeners are sensitive to the level of speech signal noise and that their perception reflects these differing noise levels in a way that is compatible with the optimal behavior predicted by the rational model. This effect of noise is not directly predicted by previous models, though it may be compatible with some of them, as discussed in the next section.

Comparison With Previous Models

Our rational model has taken a new approach to explaining the perceptual magnet effect, framing it as the optimal solution to the inference problem of perceiving speech sounds in the presence of noise. However, the solution derived in this analysis shares elements with several previous computational models, which have implicitly incorporated mechanisms that implement reliance on prior information and optimal inference of category membership. These parallels allow the various approaches to be seen as complementary descriptions of the same system that we describe here, articulated at different levels of analysis (Marr, 1982). Previous models provide process-level accounts showing how a system like the one we propose might be implemented, whereas the rational model uses analysis of the computational-level problem to explain why the mechanisms proposed by previous models should work.

Exemplar Model

A direct mathematical connection occurs with Lacerda's (1995) model, in which listeners' discrimination abilities are the side

---

[8] These correlations are relatively low on account of sparse data in cells where most participants responded different. The correlations increase to $r = .72$ and $r = .52$ (both $ps < .01$) for the no-noise and noise conditions, respectively, when the analysis is limited to zero-, one-, two-, and three-step contrasts.

[9] This differs from Iverson and Kuhl's (1995) assumption of a linear relationship between log reaction times and perceptual distance.

[10] The exact value did not affect the analysis, as long as the value was high enough that z-scores for different responses and z-scores for same responses did not overlap substantially.
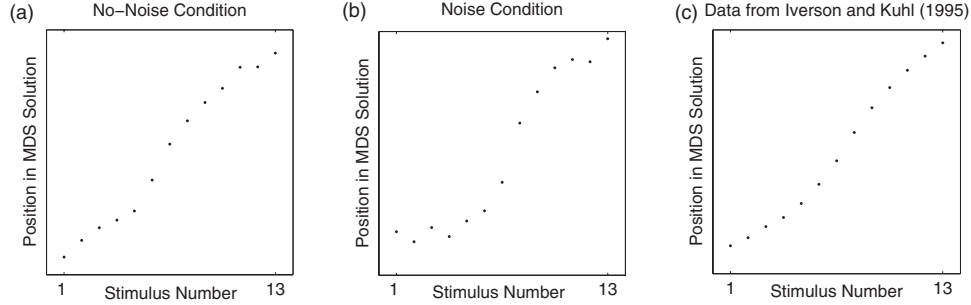
Figure 13. Perceptual maps for the (a) no-noise and (b) noise conditions obtained through multidimensional scaling (MDS). Data from (c) Iverson and Kuhl's (1995) multidimensional scaling experiment are shown for comparison.

effect of an exemplar-based categorization problem. Lacerda's model rests on the assumption that phonetic categories have approximate Gaussian distributions and that listeners store labeled exemplars from these categories. Perception requires listeners to determine the category membership of a new speech sound. Lacerda defines a speech sound's similarity to a category as the proportion of stored exemplars within some distance ϵ from the speech sound that belong to the category. Listeners' discrimination of two speech sounds then depends on the difference between the two speech sounds' similarity values.

In a system with two categories A and B, the similarity of a speech sound x to Category A (sA) is defined in the exemplar model as

$$sA(x) = \frac{NeighbA(x, \epsilon)}{NeighbA(x, \epsilon) + NeighbB(x, \epsilon)}, \quad (15)$$

where $NeighbA(x, \epsilon)$ indicates the number of neighbors within range ϵ of speech sound x. The discrimination function depends on the difference in similarity between neighboring speech sounds; as the distance between neighboring speech sounds approaches zero, this corresponds to the derivative of the similarity function. The discrimination function is therefore defined as

$$discr(x) = \frac{\left| \frac{dsA(x)}{dx} \right| + \left| \frac{dsB(x)}{dx} \right|}{k}, \quad (16)$$

where k is an arbitrary constant. This indicates that the discriminability at a point in perceptual space depends on the rate of change of category membership.

The mathematics underlying this exemplar model have a direct connection to our rational model. The first point of connection is that the similarity function in the exemplar model approximates the posterior probability of category membership in the rational model. This can be seen by noting that the exemplars are generated from a Gaussian distribution so that listeners who have heard N_A exemplars from Category A have heard approximately $\int_{x-\epsilon}^{x+\epsilon} p(S|A)N_A dS$ exemplars from Category A within a range ϵ from speech sound x. As epsilon approaches zero, the number of neighbors is proportional to $p(x|A)N_A$. The similarity metric then becomes

$$sA(x) = \frac{p(x|A)N_A}{p(x|A)N_A + p(x|B)N_B}, \quad (17)$$

which is equivalent to Bayes' rule as long as the number of stored exemplars in each category, N_A and N_B, is proportional to the prior probabilities of the categories, $p(A)$ and $p(B)$. This calculation yields the posterior probability $p(A|x)$, indicating that the similarity metric used in the exemplar model approximates the posterior probability of category membership.

Furthermore, the discrimination function defined in the exemplar model is a component of the measure of warping defined in the rational model. This can be shown by substituting $p(A|x)$ and its analogue $p(B|x)$ into the discrimination function, yielding

$$discr(x) = \frac{\left| \frac{dp(A|x)}{dx} \right| + \left| \frac{dp(B|x)}{dx} \right|}{k}. \quad (18)$$

Recall that Equation 14, which defined perceptual warping in the rational model, included the term

$$\sum_c \mu_c \frac{dp(c|S)}{dS}.$$

There is a direct correspondence between the derivative terms in the two equations: Both indicate that the discriminability at a particular point in perceptual space is a linear function of the rate of change in the identification function. The constant k in the exemplar model corresponds in our model to a number that is based on the speech signal noise, category variance, and distance between the two category means, as discussed in Appendix C. Unlike in the exemplar model, discriminability in the rational model includes an additional component that is not based on category membership: Listeners can discriminate speech sounds that differ acoustically to the extent that they rely on acoustic information from the speech sounds.

This analysis shows that the rational model incorporates the idea from Lacerda's exemplar model that discrimination peaks occur near category boundaries as a result of the distributions of exemplars in phonetic categories. Our model goes beyond the exemplar model to account for better than chance within-category discriminability and to provide independent justification for why discrimination should be best near those speech sounds where category uncertainty is highest. This maximum discriminability occurs because of the attractors that form at each phonetic category through optimal compensation for speech signal noise. The attractors pull equally on speech sounds that are on the boundary between pho-

netic categories, but as soon as a speech sound is to one side or the other of the boundary, perception is influenced more by the mean of the more probable category.

Despite their similarities, the two models differ in the goal they assign to the listener. Whereas Lacerda argues that listeners perceive only similarity to phonetic categories, shown here to be a measure of category membership, the rational model is based on the assumption that listeners are trying to extract acoustic detail from the speech signal. Because of this theoretical difference, the two models yield differing predictions on the role of speech signal noise in speech perception: Lacerda's model does not predict the experimental result that reliance on category information should increase as a result of increased speech signal noise.

Neural Network Models

Additional links can be drawn between our rational model and several neural network models that have been proposed to account for categorical effects in speech perception. Guenther and Gjaja (1996) focused specifically on the perceptual magnet effect, proposing that Gaussian distributions of speech sounds can create a bias in neural firing preferences that favors category centers. In their model, most neurons preferentially respond to speech sounds near category centers, whereas few neurons favor speech sounds near category edges. This is a direct result of their unsupervised learning mechanism, which causes the distribution of neural firing preferences to mirror the distribution of speech sounds in the input. With such a distribution in place, a population vector computed over the entire population of neurons will include disproportionately many responses from neurons that detect sounds near category centers, biasing perception toward prototypical speech sounds.

While learning is not addressed in our model, the perceptual mechanism used in the neural model has a direct link to the model proposed here. Shi, Feldman, and Griffiths (2008) demonstrated that one can perform approximate Bayesian inference using an exemplar model by storing samples from the prior distribution, weighting each sample by its likelihood, and averaging over the values of these weighted samples. The neural model proposed by Guenther and Gjaja can be interpreted as implementing this type of approximate inference. In their model, the neural firing preferences come to mirror the distribution of speech sounds in the input so that the firing preference of each neuron represents a possible target production sampled from the prior distribution. The activation of each neuron in the model then depends on the similarity of its firing preference to the speech sound heard. Specifically, the similarity is given by the dot product of the two unit vectors representing formant values, which has its maximum when the two formant values are equal. Though this differs from the Gaussian likelihood function we propose, it implements the idea that formant values most similar to the speech sound are given the highest weight. Finally, the percept of a sound is given by the population vector, which is a weighted average of neural firing preferences in which the weight assigned to each neuron is equal to its activation. Perception through the neural map therefore implements approximate Bayesian inference: The prior is given by neural firing preferences, and the likelihood function is given by the activation rule. Although this neural implementation itself makes no predictions about the dependence of perceptual warping on speech signal noise, our analysis indicates that the dependence can be implemented in this framework through a mechanism that changes the neural activation rule, parallel to changing the likelihood function, based on noise levels.

Vallabha and McClelland (2007) presented a neural model of the /r/ and /l/ categories in which learning is also based on Gaussian distributions of speech sounds. This model has three layers of representation: an acoustic layer determined entirely by the input, a middle layer that represents perceptual space, and a final layer that represents category information. The category layer contains bidirectional connections with the perceptual layer such that the perception of a speech sound can help determine its category, but the category identification then exerts a bias on perception, moving the perceptual representation closer to the mean of a phonetic category. This is similar to the account of categorical perception provided by the TRACE model (McClelland & Elman, 1986). The model shares several theoretical components with the rational model, as it allows both category information and acoustic information to influence perception. However, we know of no explicit mathematical connections between the two models, and the authors did not address the neural model's dependence on noise.

Several models of categorical perception are presented and reviewed by Damper and Harnad (2000). These models have in common that they are trained, in a supervised or unsupervised manner, on endpoint stimuli comprising voiced and voiceless tokens and tested on a voice onset time continuum between these endpoints. Results indicate that two types of neural networks, a perceptron and a brain-state-in-a-box model (following J. A. Anderson, Silverstein, Ritz, & Jones, 1977) can reproduce the sharp category boundary between voiced and voiceless stops. In the perceptron, this categorization behavior likely results from the sigmoid activation function of the output unit, which resembles the logistic categorization function given in Equation 12. The brain-state-in-a-box model does not include this logistic categorization function but does include a mechanism mapping each input to its nearest attractor, creating a sharp change in behavior near the category boundary. These models therefore capture the idea that the discrimination function is dependent on categorization, but they fail to capture the within-category discriminability that has been shown for vowels. Because they only model categorization behavior, these models also fail to predict increased reliance on category information under noisy conditions.

These neural network models all implement some of the ideas contained in the rational model: either the idea that prior probability favors speech sounds near the center of a category or the idea that discrimination is best near category boundaries. Models that implement the idea of bias toward category centers could theoretically be extended to account for increased bias under noisy conditions. However, the rational model goes further than this to explain why the dependence on noise should occur at all.

Acoustic and Phonetic Memory

Finally, the idea that both acoustic information from the speech sound and phonetic information from the category mean contribute to a listener's percept has been suggested previously by Pisoni (1973) and others, who argued that the differences between vowel and consonant perception stem from the fact that vowels rely more

on acoustic memory, whereas consonants rely more on phonetic memory. Like the Bayesian model, this account of acoustic and phonetic memory predicts that as the acoustic uncertainty increases, listeners should rely increasingly on phonetic memory, making perception more categorical. This idea has been tested in empirical studies that interfered with acoustic memory to obtain more categorical perception of vowels (Repp et al., 1979) or encouraged use of acoustic memory to obtain less categorical perception of consonants (Pisoni & Lazarus, 1974). In addition, tasks that required less memory load were found to increase especially the within-category discriminability of vowels (Pisoni, 1975).

This model is compatible with our Bayesian analysis, given some assumptions about the interaction between acoustic and phonetic memory and the degree to which each is used. The perception of speech sounds in the Bayesian model is a weighted average of the speech sound S and the means μ_c of a set of phonetic categories. One possible mechanism for implementing this approach would be to store the speech sound in acoustic memory and activate the phonetic category mean in phonetic memory. Under this assumption, the Bayesian model complements the process-level memory model by predicting the extent to which each mode of memory is used: For categories with high variability and in lower noise conditions, listeners should rely more on acoustic memory, whereas for categories with low variability and in higher noise conditions, listeners should rely more on phonetic memory.

It is worth noting that the closed-form solution given in Equation 11 holds only in the case of Gaussian phonetic categories and Gaussian noise. Qualitatively similar effects are predicted for any unimodal distribution of speech sounds, but these cases generally do not yield a quantitative solution that takes the form of a weighted average between acoustic and phonetic components. However, the weighted average may provide a close approximation to optimal behavior even in these cases.

Summary

In this section, we have shown that direct links can be drawn between the rational model and several process-level models that have been proposed to account for the perceptual magnet effect and categorical perception more generally. Any of these mechanisms might be consistent with the computational-level account we propose, and our analysis does not provide evidence for one particular implementation over another. Instead, our model contributes by providing a higher level explanation of the principles that underlie the behavior of many of these models and by identifying phenomena such as the importance of speech signal noise that have not been predicted by previous accounts.

General Discussion

This article has described a Bayesian model of speech perception in which listeners infer the acoustic detail of a speaker's target production on the basis of the speech sound they hear and their prior knowledge of phonetic categories. Uncertainty in the speech signal causes listeners to infer a sound that is closer to the mean of the phonetic category than the speech sound they actually heard. Assuming that a language has multiple phonetic categories, listen-

ers use the probability with which different categories might have generated a speech sound to guide their inference of the acoustic detail. Simulations indicate that this model accurately predicts interstimulus distances in the detailed perceptual map from Iverson and Kuhl's (1995) multidimensional scaling experiment as well as discrimination data from a novel experiment investigating the effect of noise on listeners' use of category information. The remainder of the article revisits the model's assumptions and qualitative predictions in the context of previous research on the perceptual magnet effect, phonetic category acquisition, spoken word recognition, and categorical effects in other domains.

The Perceptual Magnet Effect

The rational model predicts that three factors are key in determining the nature of perceptual warping: category frequency, category variance, and speech signal noise. Nearly all values of these parameters imply the same pattern of perception, though to differing degrees. Speech sounds are pulled toward the means of nearby categories, yielding reduced discriminability near the centers of phonetic categories and increased discriminability near category edges. This is qualitatively in line with previous descriptions of the perceptual magnet effect. However, research on the perceptual magnet effect has found seemingly conflicting empirical data: Several studies have found better discrimination near category boundaries than near the prototype, consistent with the idea of a perceptual magnet effect (Diesch et al., 1999; Grieser & Kuhl, 1989; Iverson & Kuhl, 1995, 1996; Iverson et al., 2003; Kuhl, 1991), whereas other studies have found that the effect does not extend to other vowel categories (Sussman & Gekas, 1997; Thyer et al., 2000) or that methodological details affect the degree to which categorical effects are observed (Lively & Pisoni, 1997; Pisoni, 1975). The model's predictions concerning differences in category variance and noise conditions suggest some avenues by which this debate might be resolved.

The predicted influence of category variance on perceptual warping may provide a reason why some categories show a higher degree of categorical perception than others. Data from Hillenbrand et al. (1995) suggest that the /i/ category has lower variance than other vowel categories in the direction tested by Iverson and Kuhl (1995), and it may be because of these higher levels of variability that the perceptual magnet effect has been difficult to find in other categories. Clayards et al. (2008) have demonstrated that adults are sensitive to the degree of within-category variability in an identification task, and our model predicts that this sensitivity carries over to discrimination tasks and makes perception less categorical in categories with high variability.

A second factor that should affect perceptual warping is the amount of speech signal noise, and the results of our experiment demonstrate that the perceptual magnet effect in the English /i/ and /e/ categories can be modulated by adding white noise. One immediate implication of this is that details of stimulus presentation are critical in speech perception experiments. Poor stimulus quality might actually yield better categorical perception results, and similar manipulations of memory uncertainty should also have this effect. This idea is consistent with results that show more pronounced discrimination peaks at category boundaries with longer interstimulus intervals, where memory uncertainty should be highest (Pisoni, 1973). Further research is necessary to determine the

extent to which these factors can explain the variability in empirical results.

Another debate in the literature discusses the extent to which the perceptual magnet effect is a between-category or within-category phenomenon, and the rational model provides a way of reconciling these two characterizations. The within-category account involves speech sound prototypes that act as perceptual magnets, pulling the perception of speech sounds toward them (Kuhl, 1991). The idea of a perceptual magnet is formalized in Equation 7, where speech sounds are perceived based on the mean of the category that produced them. The between-category account ties the perception of speech sounds to the task of inferring category membership (Lacerda, 1995). In line with this, the Bayesian solution to the problem of speech perception with multiple categories (Equation 11) is consistent with the idea that listeners calculate the probability of each phonetic category having generated a speech sound. However, in contrast to Lacerda's model, which assumes that listeners are perceiving only category membership, the present model predicts that listeners perceive speech sounds in terms of speakers' intended target productions, a continuous variable that depends only partly on category membership. The rational model therefore synthesizes these two previous proposals into a single framework in which the perceptual magnet effect arises through the interaction between shrinkage of perceptual space toward category centers and enhanced discrimination between categories through optimal inference of category membership.

Similar to probabilistic models in visual perception (e.g., Yuille & Kersten, 2006), the use of the term *inference* here is not meant to imply that listeners are performing explicit computations, and the model does not attempt to distinguish between inference and perception. Likewise, in determining which categories might have generated a speech sound, listeners need not be making explicit categorization judgments. This computation may involve nothing more than implicit and automatic activation of the relevant phonetic categories, or even simple retrieval of stored exemplars (Shi et al., 2008). The argument presented here is that the perceptual magnet effect results from a process that approximates the mathematics of optimal inference and that this process is advantageous to listeners because it allows them to perceive speech sounds accurately.

Phonetic Category Acquisition

The rational model assumes that listeners have prior knowledge of phonetic categories in their language. Although this is true of adult listeners, it poses an acquisition problem because infants need to learn which categories are present in their native language. This acquisition problem has been studied in the context of several computational models that use a mixture of Gaussians approach to recover Gaussian categories from unlabeled input. In de Boer and Kuhl's (2003) model, the expectation maximization algorithm (Dempster, Laird, & Rubin, 1977) was used to find an appropriate set of three vowel categories from a batch of stored exemplars. More recently, McMurray, Aslin, and Toscano (2009) used an incremental algorithm to learn the category parameters for a voicing contrast, and Vallabha, McClelland, Pons, Werker, and Amano (2007) applied this incremental algorithm to vowel formant and duration data from English and Japanese infant-directed speech. Incremental algorithms lend psychological plausibility to this ac-

count, allowing infants to learn from each speech sound as it is heard. The Gaussian categories learned by this type of algorithm would provide the necessary prior information assumed in our Bayesian model.

Learning explicit Gaussian categories yields a prior that is consistent with this model, but it is also possible to relax the assumptions of normality and of discrete categories so that the perceptual magnet effect arises simply as a result of listeners' estimating the distribution of speech sounds in their language. Formal analyses of models of categorization have shown that simply storing exemplars can provide an alternative method for estimating the distribution associated with a category (Ashby & Alfonso-Reese, 1995). If it is assumed that probabilities are assigned to stimuli in a way that is determined by their similarity to previously observed exemplars, and that the distribution associated with a category results from summing the probabilities produced by each exemplar from that category, the result is a kernel density estimator, a nonparametric method for estimating probability distributions (Silverman, 1986). Given sufficiently many exemplars, the distribution estimated in this fashion will approximate the distribution associated with the category. If the category distribution is Gaussian, the result will be approximately Gaussian. However, listeners do not need explicit knowledge of this larger structure. Rather, they can obtain the same perceptual effect by treating each exemplar as its own category. In this scenario, listeners need to take many small overlapping categories, or kernels, into account using Equation 11. In our discussion of limits on category variance, we showed that if two Gaussian categories produce a collective unimodal distribution, all of perceptual space is biased inward toward a point between the categories. Here, kernels that represent speech sounds from a Gaussian phonetic category will combine to produce a unimodal Gaussian distribution. The mathematics of this case reduce to the mathematics of the case of a single discrete category, with the weight on speech sound S equal to the sum of the kernel width and the variance in the locations of kernels.

This method of learning distributions based on individual speech sounds removes the need for listeners to have knowledge of explicit categories, reducing the severity of the learnability problem. It suggests that the perceptual magnet effect requires prior knowledge of the distributions of speech sounds in the input but does not require knowledge of the discrete categories that these distributions represent. The mere presence of the perceptual magnet effect does not necessarily imply knowledge of discrete phonetic categories. Furthermore, this analysis can be used to relax the assumptions of Gaussian phonetic categories. Any unimodal distribution in the locations of exemplars should produce a qualitatively similar effect to that obtained with Gaussians, since as soon as the kernels representing exemplars are close enough together to yield a combined unimodal distribution, perception will be biased inward to a point between those exemplars.

Multiple Dimensions

In this article, we have examined a simplified problem in speech perception, involving stimuli that lie along a single psychoacoustic dimension. Real speech input contains multiple dimensions that are relevant for categorizing and discriminating stimuli, and in

future work it will be interesting to examine discrimination patterns in categories that vary along multiple dimensions (e.g., Iverson et al., 2003) as well as patterns of trading relations in phoneme identification (e.g., Repp, 1982). Both of these problems require the use of more complex representations, such as multidimensional Gaussians, to represent phonetic categories and noise processes.

Preliminary simulations of the two-dimensional /r/–/l/ data from Iverson and Kuhl (1996) using multidimensional Gaussians with diagonal covariance matrices suggest that our rational model captures some aspects of these data but that the model would need to be extended to fully capture human data in multiple dimensions. These /r/–/l/ data show two basic effects. First, there is shrinkage toward category means along the F_3 dimension, the dimension that separates the two categories. This shrinkage is weakest near the boundary between the categories, as predicted by the rational model. Second, the data show shrinkage in the F_2 dimension, and this F_2 shrinkage is strongest at F_3 values that are near the category means. Although the rational model predicts shrinkage in the F_2 dimension, it predicts the same amount of F_2 shrinkage at any value of F_3.

This issue can potentially be addressed in two ways within the framework of the rational model. First, one can relax the assumption of Gaussian categories and Gaussian noise, an assumption that we have adopted only for computational simplicity. The neural map proposed by Guenther and Gjaja (1996) provides evidence that relaxing the Gaussian assumption will allow the model to capture human performance. As discussed above, Guenther and Gjaja's model implements an approximate form of optimal Bayesian inference (Shi et al., 2008). The likelihood is given by their activation function, which is non-Gaussian, and the prior distribution is given by neural firing preferences in their neural map, which may be non-Gaussian as a result of their learning algorithm. This neural model therefore implements an approximation of our rational model that relaxes the Gaussian assumption. Their model obtains a close fit to the two-dimensional /r/–/l/ data, suggesting that, in principle, the rational model is capable of capturing this pattern.

A second potential extension to the rational model would allow sounds to be generated from nonspeech categories. Currently, all sounds are assumed to belong to the /r/ and /l/ categories, but incorporating a nonspeech category would allow sounds that are different from native language categories to be classified as nonspeech. In the data from Iverson and Kuhl (1996), sounds that are farthest from phonetic category centers are biased less than predicted by our current model. Consistent with this, a nonspeech category with a uniform distribution over acoustic space would weaken the perceptual bias for sounds that are very different from native language categories. This would accord with suggestions from the speech perception literature that sounds dissimilar to native language phonetic categories remain perceptually unassimilated (e.g., Best, McRoberts, & Sithole, 1988). It would also parallel the suggestion by Huttenlocher et al. (2000) that participants performing a visual stimulus reproduction task are less likely to treat extreme stimulus values as belonging to the category of experimental stimuli, weakening the bias toward the edge of the category.

Phoneme Identification and Spoken Word Recognition

Speech perception involves recognizing not only speech sounds but also words, and our framework is potentially compatible with several models of spoken word recognition. Shortlist B (Norris & McQueen, 2008) uses a Bayesian framework to characterize word recognition in fluent speech at a computational level, and a potential connection to this model comes through the quantity $p(c|S)$, which is used as a primitive in Shortlist B to compute word and path probabilities for spoken utterances. On an implementational level, our model is potentially compatible with either interactive (McClelland & Elman, 1986) or feed-forward (Norris, McQueen, & Cutler, 2000) architectures, which give different accounts as to how acoustic and lexical information are combined during phoneme recognition. Any computation that ultimately yields the posterior on target productions $p(T|S)$ is compatible with our model. Under a feed-forward account, acoustic and lexical information would combine at a decision level to generate the posterior distribution, whereas in an interactive account, an initial guess at the distribution on target productions might be recursively updated by lexical feedback until it settles on the correct posterior distribution. The model is also potentially compatible with either an episodic lexicon (e.g., Bybee, 2001) or a more abstract lexicon (e.g., McClelland & Elman, 1986) that nevertheless includes phonetic detail. As discussed above, groups of exemplars can produce perceptual patterns similar to those obtained using abstract categories. The presence of a perceptual magnet effect for isolated phonemes suggests that some prior information is available at the level of the phoneme (see also McQueen, Cutler, & Norris, 2006), but this might be achieved either through abstraction or through analogy with stored lexical items.

At the level of phoneme perception, the rational model is aimed primarily at explaining discrimination performance, but the quantity $p(c|S)$ can potentially account for performance in explicit phoneme identification tasks as well. Consistent with our model's predictions, Clayards et al. (2008) have demonstrated that listeners are sensitive to the degree of category variance when performing explicit categorization tasks. Nevertheless, we acknowledge the possibility that the quantity $p(c|S)$ used for identification tasks is different from that used for discrimination tasks. Such divergence might be due to incorporation of additional information (e.g., lexical information) into explicit categorization tasks or to loss of information through imperfect approximations of the target production T before explicit categorization occurs. These possibilities remain open to further investigation.

Central to the rational model is the assumption that listeners have knowledge of phonetic categories but are trying to infer phonetic detail. This contrasts with previous models that have assumed listeners recover only category information about phonemes. Phonemes do distinguish words from one another; however, it is not clear that listeners abstract away from phonetic detail when storing and recognizing words (Goldinger, 1996; Ju & Luce, 2006; McMurray, Tanenhaus, & Aslin, 2002). Evidence has shown that listeners are sensitive to subphonemic detail at both neural and behavioral levels (Andruski, Blumstein, & Burton, 1994; Blumstein, Myers, & Rissman, 2005; Joanisse, Robertson, & Newman, 2007; Pisoni & Tash, 1974). Phonetic detail provides coarticula-

tory information that can help listeners identify upcoming words and word boundaries, and data from priming studies have suggested that listeners use this coarticulatory information online in lexical recognition tasks (Gow, 2001). This implies that listeners not only infer a speech sound's category but also attend to the phonetic detail within that category in order to gain information about upcoming phonemes and words. Though one could contend that listeners ultimately categorize speech sounds into discrete phonemes, their more direct goal must be to extract all relevant acoustic information from the speech signal. Because of its core assumption that listeners recover the phonetic detail of speech sounds they hear, the rational model is in accord with these behavioral results showing the use of phonetic detail in spoken word recognition.

Categorical Effects in Other Domains

The assumptions underlying the rational model are not specific to the structure of speech, and this makes the modeling results potentially applicable beyond the specific case of vowel perception. The extent to which this model can account for phenomena such as categorical perception of consonants, colors, or faces is an exciting question for future research. A generalization of these results to consonant perception seems to be the most straightforward, and results that are qualitatively compatible with the rational model's predictions have been found in stop consonant perception as measured by identification tasks (Burton & Blumstein, 1995; Clayards et al., 2008; Ganong, 1980). To the extent that consonants can be modeled as distributions of speech sounds along acoustic dimensions, the same principles that apply to vowel perception should yield insight into consonant perception. However, additional factors may need to be taken into account when modeling perception of consonants, especially stop consonants. Discrimination peaks have been found near stop consonant boundaries in animals (Kuhl & Padden, 1982, 1983) and very young infants (Eimas et al., 1971), suggesting that patterns in stop consonant perception are not solely the result of estimating distributions of speech sounds in the input but also involve auditory discontinuities. Auditory discontinuities are found in nonspeech stimuli as well (J. D. Miller et al., 1976; Pisoni, 1977) and might result from differential perceptual uncertainty depending on the stimulus value (Pastore et al., 1977). Influences of auditory discontinuities on category learning have been shown in adults (Holt et al., 2004), and future research might investigate how these discontinuities interact with learned categories in speech perception and whether they continue to influence perception after phonetic categories are acquired.

The rational model suggests that cross-linguistic differences in speech perception result from differences in the distributions of speech sounds heard by listeners, where perception is biased toward peaks in these distributions. A key issue in applying these results to color and face perception therefore involves examining the extent to which categories in these domains can be characterized as clusters of exemplars. This seems plausible for both facial expressions and facial identities; however, the distribution of colors in the world is unlikely to depend on linguistic experience. Categorical perception of color appears instead to be mediated by linguistic codes, and effects of verbal interference on categorical perception of facial expressions parallel those in color perception (Roberson & Davidoff, 2000; Tan et al., 2008). The model presented here does not incorporate the notion of linguistic codes, and it may need to be extended to account for these results. Nevertheless, direct behavioral parallels have been drawn between color perception and speech perception (e.g., Bornstein & Korda, 1984). In the domain of face perception, stronger categorical effects in familiar faces than in unfamiliar faces (Beale & Keil, 1995) and shifts in the discrimination peak based on shifted category boundaries (Pollak & Kistler, 2002) are consistent with the rational model's predictions. Indeed, categorical perception of facial expressions has been argued to be more in line with prototype bias accounts than with labeling accounts (Roberson et al., 2007). These qualitative similarities may indicate that categories based on exemplar distributions and those based on linguistic codes are processed in a similar manner, but further investigation is required to determine the extent of these parallels.

Finally, evidence that our results are applicable beyond the specific case of speech perception comes from nonlinguistic domains in which versions of this model have previously been proposed. Huttenlocher et al. (2000) used the same one-category model to explain category bias in visual stimulus reproduction, and this has been followed by demonstrations of similar effects with other types of visual stimuli (Crawford, Huttenlocher, & Hedges, 2006; Huttenlocher, Hedges, Corrigan, & Crawford, 2004). Körding and Wolpert (2004) explained subjects' behavior in motor tasks using the same analysis. Similar ideas have also been used to describe optimal visual cue integration (Landy, Maloney, Johnston, & Young, 1995) and audiovisual integration (Battaglia, Jacobs, & Aslin, 2003). Although this does not mean that the mechanisms being used in these domains are equivalent, it at least implies that several low-level systems use the same optimal strategy when combining sources of information under uncertainty, explaining why categories should influence perception in each of these cases.

References

Aaltonen, O., Eerola, O., Hellström, Å., Uusipaikka, E., & Lang, A. H. (1997). Perceptual magnet effect in the light of behavioral and psychophysiological data. *Journal of the Acoustical Society of America, 101,* 1090–1105.

Abramson, A. S., & Lisker, L. (1970). Discriminability along the voicing continuum: Cross language tests. In *Proceedings of the 6th International Congress of Phonetic Sciences* (pp. 569–573). Prague, Czech Republic: Academia.

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review, 84,* 413–451.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonemic differences on lexical access. *Cognition, 52,* 163–187.

Angeli, A., Davidoff, J., & Valentine, T. (2008). Face familiarity, distinctiveness, and categorical perception. *The Quarterly Journal of Experimental Psychology, 61,* 690–707.

Ashby, F. G., & Alfonso-Reese, L. A. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology, 39,* 216–233.

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America, 20,* 1391–1397.

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition, 57,* 217–239.

Beddor, P. S., & Strange, W. (1982). Cross-language study of perception of the oral–nasal distinction. *Journal of the Acoustical Society of America, 71,* 1551–1561.

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution.* Berkeley: University of California Press.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 345–360.

Blumstein, S. E., Myers, E. B., & Rissman, J. (2005). The perception of voice onset time: An fMRI investigation of phonetic category structure. *Journal of Cognitive Neuroscience, 17,* 1353–1366.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341–345.

Bornstein, M. H., & Korda, N. O. (1984). Discrimination and matching within and between hues measured by reaction times: Some implications for categorical perception and levels of information processing. *Psychological Research, 46,* 207–222.

Burton, M. W., & Blumstein, S. E. (1995). Lexical effects on phonetic categorization: The role of stimulus naturalness and stimulus quality. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1230–1235.

Bybee, J. (2001). *Phonology and language use.* Port Chester, NY: Cambridge University Press.

Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L., & Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition, 3,* 81–117.

Campanella, S., Hanoteau, C., Seron, X., Joassin, F., & Bruyer, R. (2003). Categorical perception of unfamiliar facial identities, the face–space metaphor, and the morphing technique. *Visual Cognition, 10,* 129–156.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition, 108,* 804–809.

Connine, C. M., Titone, D., & Wang, J. (1993). Auditory word recognition: Extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 81–94.

Crawford, L. E., Huttenlocher, J., & Hedges, L. V. (2006). Within-category feature correlations and Bayesian adjustment strategies. *Psychonomic Bulletin and Review, 13,* 245–250.

Damper, R. I., & Harnad, S. R. (2000). Neural network models of categorical perception. *Perception & Psychophysics, 62,* 843–867.

Davidoff, J., Davies, I., & Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature, 398,* 203–204.

de Boer, B., & Kuhl, P. K. (2003). Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters Online, 4,* 129–134.

de Gelder, B., Teunisse, J.-P., & Benson, P. J. (1997). Categorical perception of facial expressions: Categories and their internal structure. *Cognition and Emotion, 11,* 1–23.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, 39*B, 1–38.

Diesch, E., Iverson, P., Kettermann, A., & Siebert, C. (1999). Measuring the perceptual magnet effect in the perception of /i/ by German listeners. *Psychological Research, 62,* 1–19.

Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics, 16,* 513–521.

Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the r–l distinction by young infants. *Perception & Psychophysics, 18,* 341–347.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*(3968), 303–306.

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition, 44,* 227–240.

Formby, C., Childers, D. G., & Lalwani, A. L. (1996). Labelling and discrimination of a synthetic fricative continuum in noise: A study of absolute duration and relative onset time cues. *Journal of Speech and Hearing Research, 39,* 4–18.

Frieda, E. M., Walley, A. C., Flege, J. E., & Sloane, M. E. (1999). Adults' perception of native and nonnative vowels: Implications for the perceptual magnet effect. *Perception & Psychophysics, 61,* 561–577.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech, 5,* 171–189.

Fujisaki, H., & Kawashima, T. (1969). On the modes and mechanisms of speech perception. *Annual Report of the Engineering Research Institute, 28,* 67–72.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 110–125.

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception and Psychophysics, 66,* 363–376.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences, 103,* 489–494.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1166–1183.

Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123,* 178–200.

Goldstone, R. L. (1995). Effects of categorization on color perception. *Psychological Science, 6,* 298–304.

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition, 78,* 27–43.

Gow, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language, 45,* 133–159.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Grieser, D., & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech–sound prototypes. *Developmental Psychology, 25,* 577–588.

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

Guenther, F. H., Husain, F. T., Cohen, M. A., & Shinn-Cunningham, B. G. (1999). Effects of categorization and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America, 106,* 2900–2912.

Gureckis, T. M., & Goldstone, R. L. (2008). The effect of the internal structure of categories on perception. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1876–1881). Austin, TX: Cognitive Science Society.

Harnad, S. (1987). Introduction: Psychophysical and cognitive aspects of categorical perception: A critical overview. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 1–25). New York: Cambridge University Press.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97,* 3099–3111.

Holt, L. L., Lotto, A. J., & Diehl, R. L. (2004). Auditory discontinuities

interact with categorization: Implications for speech perception. *Journal of the Acoustical Society of America, 116,* 1763–1773.

Huttenlocher, J., Hedges, L. V., Corrigan, B., & Crawford, L. E. (2004). Spatial categories and the estimation of location. *Cognition, 93,* 75–97.

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General, 129,* 220–241.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America, 97,* 553–562.

Iverson, P., & Kuhl, P. K. (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *Journal of the Acoustical Society of America, 99,* 1130–1140.

Iverson, P., & Kuhl, P. K. (2000). Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism? *Perception & Psychophysics, 62,* 874–886.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87,* B47–B57.

Joanisse, M. F., Robertson, E. K., & Newman, R. L. (2007). Mismatch negativity reflects sensory and phonetic speech processing. *NeuroReport, 18,* 901–905.

Ju, M., & Luce, P. A. (2006). Representational specificity of within-category phonetic variation in the long-term mental lexicon. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 120–138.

Kay, P., & Kempton, W. (1984). What is the Sapir–Whorf hypothesis? *American Anthropologist, 86,* 65–79.

Kay, P., & Regier, T. (2007). Color naming universals: The case of Berinmo. *Cognition, 102,* 289–298.

Kiffel, C., Campanella, S., & Bruyer, R. (2005). Categorical perception of faces and facial expressions: The age factor. *Experimental Aging Research, 31,* 119–147.

Kikutani, M., Roberson, D., & Hanley, J. R. (2008). What's in the name? Categorical perception for unfamiliar faces can occur through labeling. *Psychonomic Bulletin and Review, 15,* 787–794.

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature, 427,* 244–247.

Kotsoni, E., de Haan, M., & Johnson, M. H. (2001). Categorical perception of facial expressions by 7-month-old infants. *Perception, 30,* 1115–1125.

Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech–sound categories. *Journal of the Acoustical Society of America, 70,* 340–349.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50,* 93–107.

Kuhl, P. K. (1993). Early linguistic experience and phonetic perception: Implications for theories of developmental speech perception. *Journal of Phonetics, 21,* 125–139.

Kuhl, P. K., & Padden, D. M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception & Psychophysics, 32,* 542–550.

Kuhl, P. K., & Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America, 73,* 1003–1010.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9,* F13–F21.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*(5044), 606–608.

Lacerda, F. (1995). The perceptual-magnet effect: An emergent consequence of exemplar-based phonetic memory. In K. Elenius & P. Branderud (Eds.), *Proceedings of the 13th International Congress of Phonetic Sciences* (Vol. 2, pp. 140–147). Stockholm: KTH and Stockholm University.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research, 35,* 389–412.

Levin, D. T., & Beale, J. M. (2000). Categorical perception occurs in newly learned faces, other-race faces, and inverted faces. *Perception & Psychophysics, 62,* 386–401.

Liberman, A., Harris, K. S., Eimas, P. D., Lisker, L., & Bastian, J. (1961). An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. *Language and Speech, 4,* 175–195.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54,* 358–368.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology, 61,* 379–388.

Lively, S. E., & Pisoni, D. B. (1997). On prototypes and phonetic categories: A critical assessment of the perceptual magnet effect in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1665–1679.

Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 732–753.

Lotto, A. J., Kluender, K. R., & Holt, L. L. (1998). Depolarizing the perceptual magnet effect. *Journal of the Acoustical Society of America, 103,* 3648–3655.

Marr, D. (1982). *Vision.* San Francisco: W. H. Freeman.

Massaro, D. W., & Cohen, M. M. (1983). Phonological context in speech perception. *Perception & Psychophysics, 34*(4), 338–348.

Maye, J., & Gerken, L. (2000). Learning phonemes without minimal pairs. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th Annual Boston University Conference on Language Development* (pp. 522–533). Somerville, MA: Cascadilla Press.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82,* B101–B111.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86.

McMurray, B. (2009). *Klattworks: A [somewhat] new systematic approach to formant-based speech synthesis for empirical research.* Unpublished manuscript.

McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: Insights from a computational approach. *Developmental Science, 12,* 369–378.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86,* B33–B42.

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance, 17*(2), 433–443.

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science, 30,* 1113–1126.

Mertus, J. (2004). *Bliss user's manual.* Providence, RI: Brown University.

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

Miller, J. D., Wier, C. C., Pastore, R. E., Kelly, W. J., & Dooling, R. J.

(1976). Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America, 60,* 410–417.

Miller, J. L., & Eimas, P. D. (1977). Studies on the perception of place and manner of articulation: A comparison of the labial-alveolar and nasal-stop distinctions. *Journal of the Acoustical Society of America, 61,* 835–845.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics, 18,* 331–340.

Morse, P. A., & Snowdon, C. T. (1975). An investigation of categorical speech discrimination by rhesus monkeys. *Perception & Psychophysics, 17,* 9–16.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385,* 432–434.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115,* 357–395.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23,* 299–325.

Özgen, E., & Davies, I. R. L. (2002). Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General, 131,* 477–493.

Pastore, R. E., Ahroon, W. A., Baffuto, K. J., Friedman, C., Puleo, J. S., & Fink, E. A. (1977). Common-factor model of categorical perception. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 686–696.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24,* 175–184.

Pilling, M., Wiggett, A., Özgen, E., & Davies, I. R. L. (2003). Is color "categorical perception" really perceptual? *Memory & Cognition, 31,* 538–551.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13,* 253–260.

Pisoni, D. B. (1975). Auditory short-term memory and vowel perception. *Memory and Cognition, 3,* 7–18.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception. *Journal of the Acoustical Society of America, 61,* 1352–1361.

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America, 55,* 328–333.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics, 15,* 285–290.

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

Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication, 41,* 221–231.

Pollak, S. D., & Kistler, D. J. (2002). Early experience is associated with the development of categorical representations for facial expressions of emotion. *Proceedings of the National Academy of Sciences, 99,* 9072–9076.

Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences, 104,* 1436–1441.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92,* 81–110.

Repp, B. H., & Crowder, R. G. (1990). Stimulus order effects in vowel discrimination. *Journal of the Acoustical Society of America, 88,* 2080–2090.

Repp, B. H., Healy, A. F., & Crowder, R. G. (1979). Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 129–145.

Rice, J. A. (1995). *Mathematical statistics and data analysis* (2nd ed.). Belmont, CA: Duxbury.

Roberson, D., Damjanovic, L., & Pilling, M. (2007). Categorical perception of facial expressions: Evidence for a "category adjustment" model. *Memory & Cognition, 35,* 1814–1829.

Roberson, D., & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition, 28,* 977–986.

Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology, 50,* 378–411.

Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General, 129,* 369–398.

Rosch Heider, E. (1972). Universals in color naming and memory. *Journal of Experimental Psychology, 93,* 10–20.

Rosch Heider, E., & Oliver, D. C. (1972). The structure of the color space in naming and memory for two languages. *Cognitive Psychology, 3,* 337–354.

Rotshtein, P., Henson, R. N. A., Treves, A., Driver, J., & Dolan, R. J. (2005). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nature Neuroscience, 8,* 107–113.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science, 210,* 390–398.

Shi, L., Feldman, N. H., & Griffiths, T. L. (2008). Performing Bayesian inference with exemplar models. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 745–750). Austin, TX: Cognitive Science Society.

Silverman, B. W. (1986). *Density estimation.* London: Chapman & Hall.

Stevenage, S. V. (1998). Which twin are you? A demonstration of induced categorical perception of identical twin faces. *British Journal of Psychology, 89,* 39–57.

Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M., & Öhman, S. E. G. (1969). Cross-language study of vowel perception. *Language and Speech, 12,* 1–23.

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America, 8,* 185–190.

Sussman, J. E., & Gekas, B. (1997). Phonetic category structure of [ɪ]: Extent, best exemplars, and organization. *Journal of Speech, Language, and Hearing Research, 40,* 1406–1424.

Sussman, J. E., & Lauckner-Morano, V. J. (1995). Further tests of the "perceptual magnet effect" in the perception of [i]: Identification and change/no-change discrimination. *Journal of the Acoustical Society of America, 97,* 539–552.

Tan, L. H., Chan, A. H. D., Kay, P., Khong, P.-L., Yip, L. K. C., & Luke, K.-K. (2008). Language affects patterns of brain activation associated with perceptual decision. *Proceedings of the National Academy of Sciences, 105,* 4004–4009.

Thyer, N., Hickson, L., & Dodd, B. (2000). The perceptual magnet effect in Australian English vowels. *Perception & Psychophysics, 62,* 1–20.

Vallabha, G. K., & McClelland, J. L. (2007). Success and failure of new speech category learning in adulthood: Consequences of learned Heb-

bian attractors in topographic maps. *Cognitive, Affective, and Behavioral Neuroscience, 7,* 53–73.

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences, 104,* 13273–13278.

Viviani, P., Binda, P., & Borsato, T. (2007). Categorical perception of newly learned faces. *Visual Cognition, 15,* 420–467.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences, 104,* 7780–7785.

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V., et al. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research, 7,* 357–369.

Witthoft, N., Winawer, J., Wu, L., Frank, M., Wade, A., & Boroditsky, L. (2003). Effects of language on color discriminability. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 1247–1252). Mahwah, NJ: Erlbaum.

Xu, L., & Zheng, Y. (2007). Spectral and temporal cues for phoneme recognition in noise. *Journal of the Acoustical Society of America, 122,* 1758–1764.

Young, A. W., Rowland, D., Calder, A. J., Etcoff, N. L., Seth, A., & Perrett, D. I. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition, 63,* 271–313.

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: Analysis by synthesis? *Trends in Cognitive Sciences, 10,* 301–308.

Appendix A

Computing Expected Target Productions

Given a generative model where $p(T|c) = N(\mu_c, \sigma_c^2)$ and $p(S|T) = N(T, \sigma_S^2)$, we can use Bayes' rule for the one-category case, $p(T|S, c) \propto p(S|T, c)p(T|c)$, to express the posterior on targets as

$$p(T|S, c) \propto \frac{1}{\sqrt{2\pi\sigma_c^2}} \exp\left(-\frac{(T - \mu_c)^2}{2\sigma_c^2}\right)$$
$$\times \frac{1}{\sqrt{2\pi\sigma_S^2}} \exp\left(-\frac{(S - T)^2}{2\sigma_S^2}\right). \quad (19)$$

The normalizing constants can be eliminated while still retaining proportionality, so this expression becomes

$$p(T|S, c) \propto \exp\left(-\frac{(T - \mu_c)^2}{2\sigma_c^2} - \frac{(S - T)^2}{2\sigma_S^2}\right). \quad (20)$$

Expanding the terms in the exponent and eliminating those terms that do not depend on T, we obtain

$$p(T|S, c) \propto \exp\left(-\frac{T^2}{2\sigma_c^2} + \frac{2T\mu_c}{2\sigma_c^2} + \frac{2ST}{2\sigma_S^2} - \frac{T^2}{2\sigma_S^2}\right). \quad (21)$$

The expression in the exponent can be simplified into one term that depends on T^2 and a second term that depends on T, so that

$$p(T|S, c) \propto \exp\left(-\frac{\sigma_c^2 + \sigma_S^2}{2(\sigma_c^2\sigma_S^2)}T^2 + \frac{2(\sigma_c^2 S + \sigma_S^2\mu_c)}{2(\sigma_c^2\sigma_S^2)}T\right). \quad (22)$$

We make the form more similar to a Gaussian distribution,

$$p(T|S, c) \propto \exp\left(-\frac{T^2 - 2\frac{\sigma_c^2 S + \sigma_S^2\mu_c}{\sigma_c^2 + \sigma_S^2}T}{2\frac{\sigma_c^2\sigma_S^2}{\sigma_c^2 + \sigma_S^2}}\right). \quad (23)$$

and multiply by the constant

$$\exp\left(-\frac{\frac{(\sigma_c^2 S + \sigma_S^2\mu_c)^2}{(\sigma_c^2 + \sigma_S^2)^2}}{2\frac{\sigma_c^2\sigma_S^2}{\sigma_c^2 + \sigma_S^2}}\right)$$

to complete the square, preserving proportionality because this new term does not depend on T. The expression

$$p(T|S, c) \propto \exp\left(-\frac{\left(T - \frac{\sigma_c^2 S + \sigma_S^2\mu_c}{\sigma_c^2 + \sigma_S^2}\right)^2}{2\frac{\sigma_c^2\sigma_S^2}{\sigma_c^2 + \sigma_S^2}}\right) \quad (24)$$

now has the form of a Gaussian distribution with mean

$$\frac{\sigma_c^2 S + \sigma_S^2\mu_c}{\sigma_c^2 + \sigma_S^2}$$

and variance

$$\frac{\sigma_c^2\sigma_S^2}{\sigma_c^2 + \sigma_S^2}.$$

The posterior distribution in the one-category case is therefore

$$p(T|S, c) = N\left(\frac{\sigma_c^2 S + \sigma_S^2\mu_c}{\sigma_c^2 + \sigma_S^2}, \frac{\sigma_c^2\sigma_S^2}{\sigma_c^2 + \sigma_S^2}\right) \quad (25)$$

and the expected value of T is the mean of this Gaussian distribution,

$$E[T|S, c] = \frac{\sigma_c^2 S + \sigma_S^2\mu_c}{\sigma_c^2 + \sigma_S^2}. \quad (26)$$

To compute the expectation $E[T|S]$ in the case of multiple categories, we use the formula $E[T|S] = \int Tp(T|S)dT$, where $p(T|S)$ is computed by marginalizing over categories, $p(T|S) = \sum_c p(T|S, c)p(c|S)$. The expression for the expectation becomes

$$E[T|S] = \int T\sum_c p(T|S, c)p(c|S)dT. \quad (27)$$

Bringing T inside the sum and then exchanging the sum and the integral yields

$$E[T|S] = \sum_c \int Tp(T|S, c)p(c|S)dT. \quad (28)$$

Because $p(c|S)$ does not depend on T, this is equal to

$$E[T|S] = \sum_c p(c|S) \int Tp(T|S, c)dT, \quad (29)$$

where $\int Tp(T|S, c)dT$ denotes $E[T|S, c]$, the expectation in the one-category case (Equation 26). The expectation in the case of multiple categories is therefore

$$E[T|S] = \sum_c p(c|S)\frac{\sigma_c^2 S + \sigma_S^2\mu_c}{\sigma_c^2 + \sigma_S^2}, \quad (30)$$

which is the same as the expression given in Equation 10.

Appendix B

Calculating Category Parameters From Identification Curves

Given a logistic identification curve for the percentage of participants that identified each stimulus as belonging to Category 1 in a two-category forced-choice identification task, one can derive the category means and common variance by noting that the curve is an empirical measure of $p(c_1|S)$, which in a two-category forced-choice task is defined according to Bayes' rule (Equation 4) as

$$p(c_1|S) = \frac{p(S|c_1)p(c_1)}{p(S|c_1)p(c_1) + p(S|c_2)p(c_2)}. \quad (31)$$

Each part of the fraction can be divided by the quantity in the numerator. Two inverse functions are applied to the last term, the exponential power and the natural logarithm, yielding

$$p(c_1|S) = \frac{1}{1 + e^{\log\frac{p(S|c2)\,p(c2)}{p(S|c1)\,p(c1)}}}. \quad (32)$$

Assuming that the two categories c_1 and c_2 have equal prior probability, and using the distribution for $p(S|c)$ given in Equation 3, we can simplify Equation 32 to a logistic equation of the form

$$p(c_1|S) = \frac{1}{1 + e^{-gS+b}}, \quad (33)$$

where

$$g = \frac{\mu_1 - \mu_2}{\sigma_c^2 + \sigma_S^2} \quad \text{and} \quad b = \frac{\mu_1^2 - \mu_2^2}{2(\sigma_c^2 + \sigma_S^2)}.$$

Thus, given values for g, b, and μ_1, one can calculate the value of μ_2 and the sum $\sigma_c^2 + \sigma_S^2$ as follows:

$$\mu_2 = \frac{2b}{g} - \mu_1 \quad \text{and} \quad (34)$$

$$\sigma_c^2 + \sigma_S^2 = \frac{\mu_1 - \mu_2}{g}. \quad (35)$$

Without the assumption of equal prior probability, the bias term instead becomes

$$b = \frac{\mu_1^2 - \mu_2^2}{2(\sigma_c^2 + \sigma_S^2)} + \log\frac{p(c_2)}{p(c_1)},$$

which produces a shift of the logistic toward the mean of the less probable category. Because the category boundary occurs where $p(c_1|S) = 1/2$ or $S = b/g$, this bias term produces a shift of magnitude

$$\frac{\log\frac{p(c_2)}{p(c_1)}}{g},$$

where g is the gain of the logistic. The extra bias term therefore creates a larger shift in boundary locations for small values of the gain parameter, which can arise through high category variance σ_c^2, high noise variance σ_S^2, or small separation between category means $\mu_1 - \mu_2$.

(Appendixes continue)

Appendix C

Measure of Warping

Perceptual warping, which is a measure of the degree of shrinkage or expansion of perceptual space, corresponds mathematically to the derivative of the expected target $E[T|S]$ with respect to S. We begin with the expectation from Equation 11

$$E[T|S] = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_S^2} S + \frac{\sigma_S^2}{\sigma_c^2 + \sigma_S^2} \sum_c \mu_c p(c|S) \quad (36)$$

and compute its derivative, using the chain rule to compute the derivative of the second term,

$$\frac{dE[T|S]}{dS} = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_S^2} + \frac{\sigma_S^2}{\sigma_c^2 + \sigma_S^2} \sum_c \mu_c \frac{dp(c|S)}{dS}. \quad (37)$$

This is the expression given in Equation 14. However, this derivative includes a term that corresponds to the derivative of the identification function. In the two-category case, the identification function has the form of a logistic function

$$p(c_1|S) = \frac{1}{1 + e^{-gS + b}}$$

whose derivative is given by

$$\frac{dp(c_1|S)}{dS} = gp(c_1|S)[1 - p(c_1|S)], \quad (38)$$

where

$$g = \frac{\mu_1 - \mu_2}{\sigma_c^2 + \sigma_S^2}.$$

Because $p(c_2|S) = 1 - p(c_1|S)$ in the two-category case, the derivative of the logistic for $p(c_2|S)$ is identical to Equation 38 except that the gain has the opposite sign. Substituting this into Equation 37 and expanding the sum yields

$$\frac{dE[T|S]}{dS} = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_S^2} + \frac{\sigma_S^2}{\sigma_c^2 + \sigma_S^2} \{\mu_1 gp(c_1|S)[1 - p(c_1|S)]$$
$$- \mu_2 gp(c_1|S)[1 - p(c_1|S)]\}, \quad (39)$$

which can be simplified to

$$\frac{dE[T|S]}{dS} = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_S^2} + \frac{\sigma_S^2}{\sigma_c^2 + \sigma_S^2} g(\mu_1 - \mu_2)\{p(c_1|S)[1 - p(c_1|S)]\}$$
$$(40)$$

or, substituting in the expression for the gain of the logistic,

$$\frac{dE[T|S]}{dS} = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_S^2} + \frac{\sigma_S^2(\mu_1 - \mu_2)^2}{(\sigma_c^2 + \sigma_S^2)^2} \{p(c_1|S)[1 - p(c_1|S)]\}.$$
$$(41)$$

Appendix D

Same–Different Task

Given two stimuli S_A and S_B, the posterior probability that the targets T_A and T_B are within range ϵ of each other is $p(|T_A - T_B| \leq \epsilon|S_A, S_B)$, which is equivalent to $p(-\epsilon \leq T_A - T_B \leq \epsilon|S_A, S_B)$. This probability can be computed analytically by marginalizing over category assignments for the two stimuli,

$$\sum_{c_A} \sum_{c_B} p(-\epsilon \leq T_A - T_B \leq \epsilon|c_A, c_B, S_A, S_B)p(c_A|S_A)p(c_B|S_B) \quad (42)$$

under the assumption that the two stimuli are generated independently (c_A and S_A are independent of c_B and S_B). To compute the first term, note that the distributions $p(T_A|c_A, S_A)$ and $p(T_B|c_B, S_B)$ are both Gaussians as given by Equation 6. Their difference therefore follows a Gaussian distribution, with its mean equal to the difference between the two means and its variance equal to the sum of the two variances,

$$T_A - T_B|c_A, c_B, S_A, S_B \sim$$

$$N\left(\frac{\sigma_c^2(S_A - S_B) + \sigma_S^2(\mu_A - \mu_B)}{\sigma_c^2 + \sigma_S^2}, \frac{2\sigma_c^2\sigma_S^2}{\sigma_c^2 + \sigma_S^2}\right). \quad (43)$$

Given this density, the probability of falling within a range between $-\epsilon$ and ϵ can be expressed in terms of the standard cumulative normal distribution Φ,

$$p(-\epsilon \leq T_A - T_B \leq \epsilon|c_A, c_B, S_A, S_B) = \Phi\left(\frac{\epsilon - \mu_d}{\sigma_d}\right)$$
$$- \Phi\left(\frac{-\epsilon - \mu_d}{\sigma_d}\right), \quad (44)$$

where

$$\mu_d = \frac{\sigma_c^2(S_A - S_B) + \sigma_S^2(\mu_A - \mu_B)}{\sigma_c^2 + \sigma_S^2}$$

and

$$\sigma_d = \sqrt{\frac{2\sigma_c^2\sigma_S^2}{\sigma_c^2 + \sigma_S^2}}.$$

The second and third terms in Equation 42 can then be computed independently for stimuli S_A and S_B from Equation 12.