

Generalizing a neuropsychological model of visual categorization to auditory categorization of vowels

W. TODD MADDOX, MICHELLE R. MOLIS, and RANDY L. DIEHL
University of Texas, Austin, Texas

Twelve male listeners categorized 54 synthetic vowel stimuli that varied in second and third formant frequency on a Bark scale into the American English vowel categories /ɪ/, /ʊ/, and /ɜ·/. A neuropsychologically plausible model of categorization in the visual domain, the Striatal Pattern Classifier (SPC; Ashby & Waldron, 1999), is generalized to the auditory domain and applied separately to the data from each observer. Performance of the SPC is compared with that of the successful Normal A Posteriori Probability model (NAPP; Nearey, 1990; Nearey & Hogan, 1986) of auditory categorization. A version of the SPC that assumed piece-wise linear response region partitions provided a better account of the data than the SPC that assumed linear partitions, and was indistinguishable from a version that assumed quadratic response region partitions. A version of the NAPP model that assumed nonlinear response regions was superior to the NAPP model with linear partitions. The best fitting SPC provided a good account of each observer's data but was outperformed by the best fitting NAPP model. Implications for bridging the gap between the domains of visual and auditory categorization are discussed.

Every day we see, hear, taste, touch, and smell objects and events that must be categorized into one of many categories to facilitate understanding or selection of an appropriate behavior. In keeping with the importance of categorization in everyday life, much research has been devoted to an understanding of the perceptual and cognitive operations involved in it. The current thinking is that there are at least two categorization systems and that each is associated with separate brain systems (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Ell, 2001; Pickering, 1997; E. E. Smith, Patalano, & Jonides, 1998). There is general agreement that one system is explicit (i.e., rule or theory based), is available to conscious awareness, and likely involves frontal brain structures. There is less agreement about the nature of the implicit system, but a growing body of work suggests that this system involves a gradual strengthening of stimulus-response associations within the striatum (a region of the basal ganglia that contains the caudate nucleus and the putamen), and is not available to conscious awareness (e.g., Ashby et al., 1998).

Ashby and Waldron (1999; see also Ashby, Waldron, Lee, & Berkman, 2001; Waldron & Ashby, 2001) recently proposed a neuropsychologically plausible model of the implicit categorization system as applied to visually presented stimuli. This model assumes a central role for the striatum and is referred to as the Striatal Pattern Classifier (SPC). Evidence in support of a central role for the striatum in visual categorization and the assumptions of the model will be presented shortly.

In recent years there has been a growing interest in determining whether speech perception can be viewed as a form of pattern recognition. This approach has yielded several important and successful models (e.g., Nearey, 1997; Nossair & Zahorian, 1991). The present study adds to this body of research by extending a successful model of visual categorization (pattern classification) to the auditory domain.

The goals of this study are many. First, we examine the neuropsychological plausibility of the SPC as applied to auditory categorization. Since most real-world auditory categorization problems are likely solved implicitly, this seems like a reasonable starting point. Although the neuropsychological underpinnings of the model are not critical for the present empirical application (i.e., a computational-level interpretation is adequate), we believe that neuropsychological plausibility issues must be emphasized, especially given the explosion of work on the neuroscience of perceptual and cognitive phenomena (e.g., Gazzaniga, 2000). Second, we apply the SPC to data from a vowel categorization study. Third, we compare performance of the SPC with that of a popular model in the speech perception literature, namely the Normal A Posteriori Probability (NAPP)

This research was supported in part by a National Science Foundation Grant SBR-9796206 and by Research Grant 5 R01 MH59196-03 from the National Institute of Mental Health, National Institutes of Health, to W.T.M., and by Research Grant 5 R01 DC00427-11 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health, to R.L.D. We thank Terrance Nearey, Roel Smits, and one anonymous reviewer for helpful comments on an earlier version of this manuscript. We also thank Greg Ashby and John Kingston for discussions that greatly influenced this work. Correspondence should be addressed to W. T. Maddox, Department of Psychology, Mezes Hall 330, Mail Code B3800, University of Texas, Austin, TX 78712 (e-mail: maddox@ps.y.utexas.edu).

model of Nearey and Hogan (1986; Nearey, 1990). The NAPP model is well established in this field and we doubt that the SPC will outperform the NAPP model. In light of this fact, our goal is not to compare the SPC to NAPP in order to reject one model or the other. Rather, our goal is to use the NAPP model as a benchmark against which to compare the SPC. If the SPC provides a reasonable account of the data relative to the NAPP model, then we will deem the SPC a model that warrants additional study. By applying the SPC to auditory categorization, we are taking an important step toward bridging the gap between studies of visual and auditory categorization (see also Kingston & Macmillan, 1995; Macmillan & Kingston, 1995), and toward offering a general neurobiological framework for studying both visual and auditory categorization.

The next section outlines the assumptions of the SPC and generalizes the model to auditory processing. In addition, the NAPP model is described. We then introduce the experiment and detail the experimental methods. We then present the results and theoretical analyses, and finally a summary and discussion.

Striatal Pattern Classifier

The SPC (Ashby & Waldron, 1999) can be viewed as a neuropsychologically plausible implementation of the decision process assumptions of Ashby and Townsend's (1986; Ashby & Perrin, 1988) General Recognition Theory (GRT; also called decision bound theory; Ashby & Maddox, 1993; Maddox & Ashby, 1993). GRT is a generalization of signal detection theory to stimuli that vary along multiple dimensions. In GRT, perceptual processes and decisional processes are characterized by separate and uniquely identifiable parameters.

Perceptual processes. Real-world objects vary along multiple basic stimulus dimensions, each of which is continuous, rather than binary valued (e.g., Ashby, 1992a; Ashby & Maddox, 1998). For example, objects may vary along the dimensions of size, color, location, fundamental frequency, and amplitude. GRT takes as its fundamental axiom that repeated presentations of the same stimulus yield different perceptual effects (i.e., perceptual noise exists) and assumes that a single multidimensional stimulus can be represented perceptually by a multivariate probability distribution (Ashby & Lee, 1993; Ashby & Townsend,

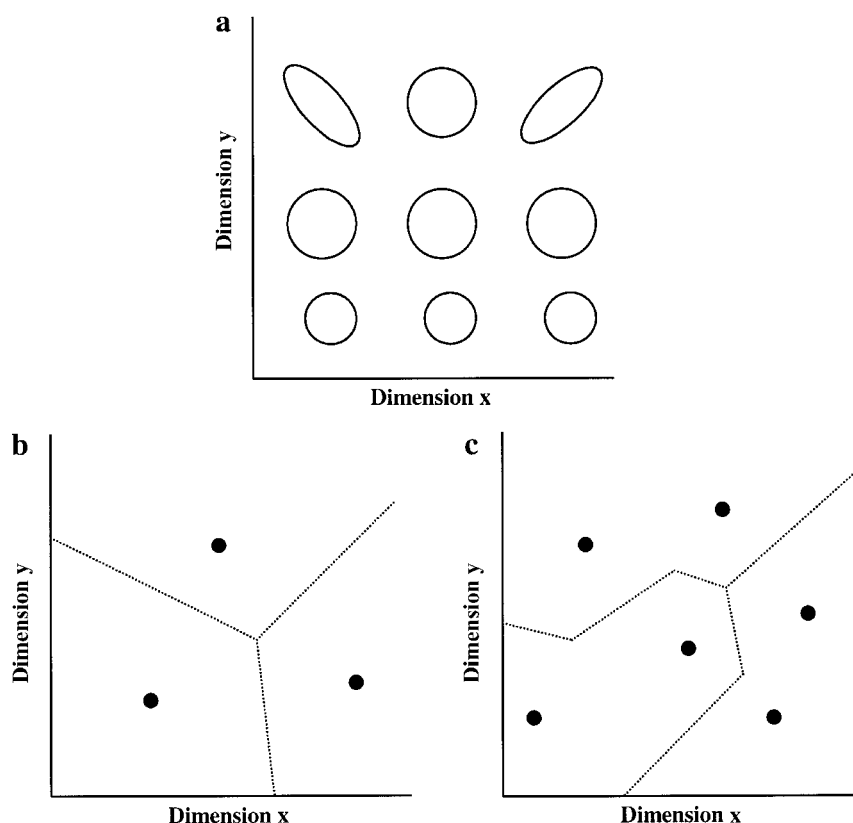


Figure 1. (a) Hypothetical contours of equal likelihood for nine stimuli constructed from the factorial combination of three levels along two stimulus components. (b) Hypothetical minimum-distance response regions and striatal units for three categories assuming one striatal unit per category. (c) Hypothetical minimum-distance response regions and striatal units for three categories assuming two striatal units per category.

1986). For a two-dimensional stimulus, a bivariate normal distribution is assumed to describe the set of percepts. A bivariate normal distribution is described by a mean and variance along each dimension, as well as a covariance term, $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \text{cov}_{xy}$, where the subscripts x and y denote dimensions x and y . Figure 1a depicts hypothetical equal likelihood contours for nine stimuli constructed from the factorial combination of three levels along two dimensions x and y . With bivariate normal distributions, the equal likelihood contours are always circular or elliptical. The spread of the contour along the x and y dimensions is related to the perceptual variability along each dimension. Note that the three contours associated with Level 1 of dimension y have less variability than the three contours associated with Level 2 of dimension y . One form of perceptual interaction predicted by GRT is especially important to the present application. It is called *perceptual independence*. Perceptual independence holds for a single stimulus if and only if the perceptual effects for dimensions x and y are statistically independent (see Ashby, 1988; Ashby & Maddox, 1991; Ashby & Townsend, 1986; Perrin & Ashby, 1991, for empirical tests of perceptual independence). With bivariate normal distributions, perceptual independence holds when the major and minor axes of the contour are parallel to the coordinate axes (i.e., when the covariance or correlation is zero). Perceptual independence holds for all but the top left and top right contours in Figure 1a. A positive slope for the major axis implies a positive covariance (or positive perceptual dependence, as seen for the top right stimulus), and a negative slope implies a negative covariance (or negative perceptual dependence, as seen for the top left stimulus).¹

Decision processes. In GRT, the experienced observer learns to divide the perceptual space into response regions and assigns a response to each region. The partitions between response regions are called *decision bounds*. On each trial the observer determines the location of the perceptual effect and gives the response associated with that region of the perceptual space. The SPC offers a neuropsychological theory of this process as applied to the categorization of visually presented stimuli. It is well known that stimuli are represented perceptually in higher level visual areas such as inferotemporal cortex (i.e., IT). It is also well established that cells in IT project, in a many-to-one fashion, to striatal cells (J. Wickens, 1993). Because the projections are many to one, it is assumed that a low-resolution map of the perceptual space is represented among the striatal units. Over trials, each striatal unit becomes associated with a particular categorization response. Thus, the striatum can be thought of as associating a categorization response with a cluster of visual cortical cells. On each trial the observer determines which unit is closest to the perceptual effect and gives the associated response. This minimum-distance classification scheme results in response regions that are separated by linear partitions (Ashby & Maddox, 1993). Hypothetical minimum-distance response regions for a three-category

problem for the nine stimuli in Figure 1a under the assumption that one striatal unit is associated with each category are displayed in Figure 1b. The response regions depicted in Figure 1b are not optimized in any sense; that is, they are not constructed in such a way as to maximize long-term accuracy. However, evidence with visually presented stimuli suggests that with experience, observer's response regions, as modeled with the SPC, do become more nearly optimal (Maddox, 2001, 2002). Hypothetical minimum-distance response regions for the same stimuli but where two striatal units are associated with each category are displayed in Figure 1c. In this case the response region partitions are piecewise linear. Importantly, within this theoretical framework, the perceptual representation parameters (Figure 1a) are separate from the response region or striatal unit parameters (Figure 1b).

A large body of neuropsychological, brain imaging, and animal research provides support for the claim that the striatum plays a central role in implicit visual categorization. First, patients with Parkinson's disease and Huntington's disease, whose neuropathology results in a decrement in striatal functioning, demonstrate impaired probabilistic classification learning, a form of implicit categorization (Filoteo, Maddox, & Davis, 2001; Knowlton, Mangels, & Squire, 1996; Maddox & Filoteo, 2001). Second, functional neuroimaging studies with normal individuals suggest that the striatum is activated during a task of probabilistic classification learning (Poldrack, Prabhakaran, Seger, & Gabrieli, 1999; Seger, Poldrack, Prabhakaran, Zhao, Glover, & Gabrieli, 2000). Third, animal studies implicate the striatum in certain aspects of category learning (McDonald & White, 1993; Packard & McGaugh, 1992). Finally, and perhaps most importantly, Waldron, Ell, Ashby, McCormick, and Casale (2000) recently developed a learning model implementation of the SPC with enough biological detail that they were able to test the model by applying it simultaneously to single-cell recording and human behavioral data. Briefly, they modeled the behavior of visual cortical cells, medium spiny cells in the caudate nucleus (a structure within the striatum), and dopamine cells of the substantia nigra. The approach was to calibrate the model by fitting it to single-cell recording data collected within the caudate nucleus and the substantia nigra and then, holding those parameters fixed, to fit the few remaining parameters to human category learning data. The model provided an excellent simultaneous account of the single-cell and human category learning data, and provides the strongest evidence to date for the neurobiological plausibility of the SPC.

A generalization to auditory categorization. Although it is straightforward mathematically to apply the SPC to data from an auditory categorization task, it is important to determine whether there is any empirical support for the neuropsychological plausibility of the SPC as a model of auditory categorization. First, we need to determine whether brain areas in the auditory processing system project to the striatum in much the same way that IT

projects to the striatum. Second, we need to determine whether there is support from the neuropsychological and animal literature for the importance of the striatum in the categorization of auditory stimuli. Although less research has been conducted to address these issues, at least three recent studies are relevant. Two recent animal studies found direct projections from auditory cortex into the striatum (Arnauld, Jeantet, Arsaut, & Desmotes-Mainard, 1996; Chudler, Sugiyama, & Dong, 1995), suggesting that a low-resolution map of the auditory perceptual space is represented among the striatal units. In fact, one of these studies (Chudler et al., 1995) suggested that there was a convergence from multiple sensory systems into the striatum. A third study implicated the striatum in the categorization of auditory stimuli in the rat (Jog, Kubota, Connolly, Hillegaart, & Graybiel, 1999). In that study the rat was trained to run a T-maze. On half the trials a food reward was presented in the left goal box, and on the other half a food reward was presented in the right goal box. An auditory cue (a 1000- or 8000-Hz tone, 75 dB) signaled which box (left or right) contained the reward on each trial. The rats successfully learned the task, and significant striatal involvement was found in this discrimination. To our knowledge no categorization studies with striatal-damaged patients (e.g., Parkinson's or Huntington's disease) have been conducted using auditory stimuli. Clearly this is a ripe area for future research.

To summarize, there does appear to be neuroanatomical, neurophysiological, and behavioral data implicating the striatum in auditory processing and categorization. In light of this fact, our next step will be to examine quantitatively the ability of the SPC to account for data from an auditory categorization task in normal human listeners. Before introducing the vowel categorization experiment and the modeling results, we briefly review the currently popular NAPP model of auditory categorization.

NAPP Model

We only briefly review the NAPP model here since it is described fully in other articles (Nearey, 1990; Nearey & Hogan, 1986). As applied to vowel categorization, the NAPP model assumes that listeners focus on the output of "detectors" or "filters" that are tuned to language-specific vowel categories. These detectors produce normally distributed outputs that correspond roughly to the likelihood of the vowel category given the stimulus input. These likelihoods are then combined using the classic Luce (1963) choice rule (also called the relative goodness rule) to generate predicted response probabilities for each vowel category. From the output of the choice rule one can generate a response surface for the model. The response surface denotes the most likely categorization response for each point in the space. The partitions that separate the various regions of the response surface behave much like the decision bounds in the SPC. Nearey has examined different types of response region partitions, in particular linear and quadratic partitions.

Although NAPP is a model of speech recognition and has many variants depending on the assumptions made about speech processing, the predicted response probabilities from the model can always be derived from logistic regression. In addition, situations exist in which different NAPP model assumptions are mathematically equivalent at the level of the data. In other words, families of NAPP models often map onto the same logistic regression equation. In light of this fact, when fitting the data in the following vowel categorization experiment we will focus on logistic regression models, specifically, linear and nonlinear logistic regression models. Alternative formulations of the model will be outlined in the Discussion section.

Listeners labeled a set of synthetic vowel stimuli that varied in second and third formant frequencies and that ranged perceptually among the categories /ɪ/, /ʊ/, and /ɜː/.

METHOD

Listeners

Listeners were 12 males ranging in age from 18 to 37 years with a mean age of 27.2. An attempt was made to limit listeners to a relatively homogeneous dialect group (Central Texas); all listeners were raised in or around the metropolitan areas of Austin, Houston, or Dallas. They received monetary compensation for their participation.

Stimuli

Fifty-six five-formant synthetic vowel stimuli were synthesized using a KLATT88-type cascade resonance synthesizer implemented on a PC. The stimulus space encompassed the American English vowel categories /ɪ/, /ʊ/, and /ɜː/. Stimuli shared a common first formant (F1) frequency but varied in second and third formant frequencies (F2 and F3) on a perceptually motivated frequency scale (Bark). F2 varied between 9.0 and 14.2 Bark (1081–2390 Hz) in equal 0.4-Bark steps. F3 varied from 10.0 to 15.2 Bark (1268–2783 Hz), also in equal 0.4-Bark steps. The Hertz values used as input parameters to the synthesizer were calculated to correspond with their desired Bark value equivalents using an equation from Traunmüller (1990). The two stimuli with the highest F2–F3 frequency combination were subsequently eliminated because preliminary investigations indicated they had an unnatural quality compared with the other stimuli in the set. A plot of the remaining 54 stimuli in the F2, F3 space is displayed in Figure 2a. The frequency values of F1, F4, and F5 were 4.5 Bark (455 Hz), 16.2 Bark (3250 Hz), and 17.0 Bark (3700 Hz), respectively. The default bandwidths of the synthesizer were used: 60 Hz for F1, 90 Hz for F2, 150 Hz for F3, and 200 Hz for F4 and F5. All stimuli were 225 msec in duration (chosen because it is similar to the measured intrinsic duration of these vowels produced in citation form; Hillenbrand, Getty, Clark, & Wheeler, 1995). The fundamental frequency was a constant 132 Hz for the initial 150 msec and thereafter fell linearly to 127 Hz over the final 75 msec. Stimuli were ramped on and off with a 10-msec half-cosine function and were normalized for RMS amplitude.

Procedure

Listeners were seated at separate response stations in a sound-attenuated chamber. Stimuli were presented over Beyer DT-100 headphones at a level of 70 dB SPL. Fourteen randomized blocks of the 54 stimuli were presented (756 trials/listener). Listeners were asked to identify each stimulus by pressing one of three response buttons labeled with the key words *hid*, *hood*, and *heard*, corre-

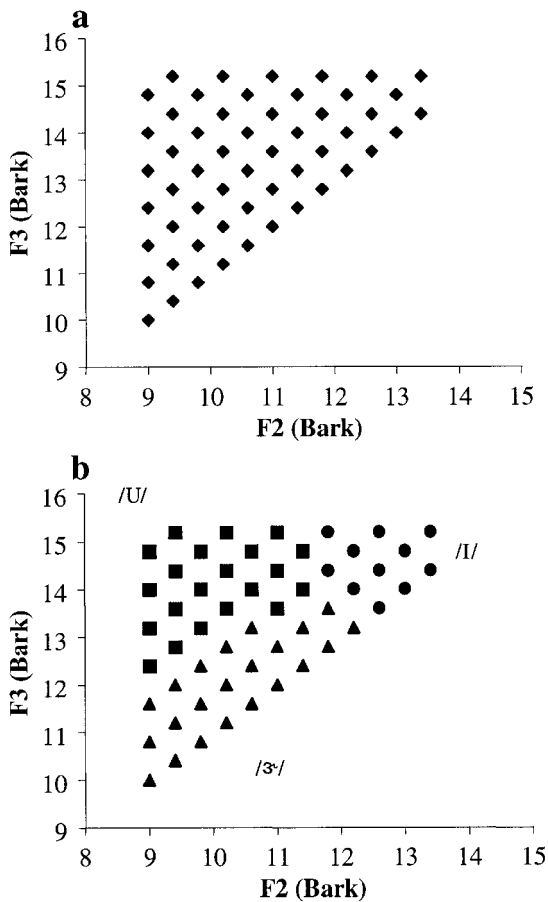


Figure 2. (a) Schematic illustration of the 54 synthetic vowel stimuli in the F2, F3 space. (b) Symbolic representation of the most common vowel category response (/U/, /I/, and /ɜ:/) for each of the 54 stimuli averaged across observers. Circle, /I/; square, /U/; diamond, /ɜ:/.

sponding to the three identifiable American English vowel categories.

RESULTS AND THEORETICAL ANALYSES

Vowel Categorization Response Probabilities

The proportions of /I/, /U/, and /ɜ:/ responses to each stimulus were computed for each listener and used for the model-based analyses. Figure 2b displays the most common response for each of the 54 stimuli, averaged across observers.

Model-Based Analyses

The data to be modeled were 162 observed vowel categorization response probabilities determined from the response probabilities for each of the three vowel categories (*hid*, *hood*, and *heard*) for all 54 stimuli. However, because the response probabilities for the three vowel categories for each stimulus must sum to one, there are only

108 *df* in the data for each listener. All model-based analyses were performed at the level of the individual listener. Data were not collapsed across listeners since averaging often alters the structure of the categorization data in such a way that the correct model of individual performance provides a poor account of the aggregate data (Ashby, Maddox, & Lee, 1994; Estes, 1956; Maddox, 1999; Maddox & Ashby, 1998; J. D. Smith & Minda, 1998).

The parameters of the SPC and logistic regression models were estimated using maximum likelihood procedures by minimizing the negative natural log likelihood ($-\ln L$; Ashby, 1992b; T. D. Wickens, 1982). With logistic regression, closed-form expressions for the best fitting parameters exist and can be obtained from several “canned” programs. Because numerical integration is required to fit the SPC (details provided below), an iterative search routine was used to estimate the model parameters.² Although maximum likelihood estimates can be used to test the empirical validity of a model in isolation, a more useful approach is to compare the fits across models. Three approaches are available for model comparison. First, when two models have the same number of parameters, the maximum likelihood fit values can be compared directly, and the model with the smallest fit value is deemed the better model. Second, when one model is a special case of another model—that is when two models are “nested”—a likelihood ratio (or G^2) test can be used to determine whether the extra parameters of the more general model provide a significant improvement in fit over the more restricted model. Consider two models, M_1 and M_2 , for which M_1 is a special case of M_2 . The statistic $G^2 = 2[\ln L_2 - \ln L_1]$ is asymptotically χ^2 distributed with degrees of freedom equal to the difference in the number of parameters between the models. If G^2 is larger than the χ^2 critical value, then the more general model, M_2 , is deemed superior. If G^2 is smaller than the χ^2 critical value, then the less general model, M_1 , is deemed superior. Finally, situations exist in which two models have different numbers of free parameters and also are not nested. In this case, the statistic $AIC = 2r - 2\ln L$, where r is the number of free parameters and L is the likelihood of the model given the data, can be compared (Akaike, 1974; Takane & Shibayama, 1992). The AIC statistic penalizes a model for extra free parameters in such a way that the smaller the AIC, the closer a model is to the “true model,” regardless of the number of free parameters. Thus, to find the best model among a given set of competitors, one simply computes an AIC value for each model and chooses the model associated with the smallest AIC value. In the present application all three of these model comparison procedures will be utilized.

Striatal pattern classifier. Several versions of the SPC were examined. Following general recognition theory, the SPC assumes that the perceptual effect associated with a single presentation of stimulus i is stochastic and, in the present application, can be represented over trials by a bivariate normal distribution. We denote the perceptual dis-

tribution (i.e., density function) for stimulus i by $f_i(x,y)$, where x and y denote the perceived F2 and F3 values. The SPC also assumes that a low-resolution map of the perceptual space is represented in the striatum. Each striatal unit becomes associated with a particular categorization response. On each trial the observer determines which unit is closest to the perceptual effect and gives the associated response. Thus, the SPC assumes that the low-resolution map of the perceptual space is carved up into separate response regions, with each region having an associated response. We denote the response region for category j by \mathcal{R}_j . Thus, the probability of responding R_j on stimulus S_i trials is equal to the probability that a perceptual effect from the $f_i(x,y)$ distribution falls in the \mathcal{R}_j response region. Computationally, this is equivalent to the proportion of the $f_i(x,y)$ distribution in the \mathcal{R}_j response region. More formally,

$$P(R_j|S_i) = \int_{\mathcal{R}_j} f_i(x,y) dx dy. \quad (1)$$

Because the response regions can take many forms, and because theoretically the mean, variance, and covariance parameters associated with the perceptual distribution $f_i(x,y)$ need not be constrained, evaluating Equation 1 requires numerical integration procedures.³

All models tested made the following two perceptual representation assumptions: (1) The mean perceptual effects (i.e., the mean for each bivariate normal distribution of perceptual effects) were located at the synthesized F2, F3 values (Figure 2a). (2) The perceptual covariance matrices were identical across stimuli. In other words, the F2 perceptual variability was constant across the 54 stimuli, the F3 perceptual variability was constant across the 54 stimuli, and the F2, F3 perceptual dependence was constant across the 54 stimuli. This is referred to as a stimulus invariant (SI) perceptual representation since the perceptual covariance matrix entries are invariant across stimuli (Ashby & Maddox, 1993, 1994; Ashby & Townsend, 1986).

Four sets of perceptual representation assumptions were tested. The four perceptual representations tested differed only in the assumptions made about the F2 and F3 perceptual variabilities, and the F2, F3 perceptual dependence. From most constrained to most general, the versions are as follows:

1. *Perceptual representation Version 1 (PR1)*. This version assumed that the perceptual standard deviation along the F2 dimension was equivalent to the perceptual standard deviation along the F3 dimension (i.e., $\sigma_{F2} = \sigma_{F3}$) and that perceptual independence was satisfied between the perceived F2, F3 values (i.e., $\rho_{F2,F3} = 0$). In other words, the contours of equal likelihood for all 54 stimuli were circles of the same diameter. This model contained one free parameter.

2. *PR2*. This version allowed the perceptual standard deviation along the F2 dimension to be *different* from the perceptual standard deviation along the F3 dimension

(i.e., $\sigma_{F2} \neq \sigma_{F3}$) and assumed that perceptual independence was satisfied between the perceived F2, F3 values (i.e., $\rho_{F2,F3} = 0$). In other words, the contours of equal likelihood for all 54 stimuli were ellipsoids of the same size and orientation whose major and minor axes were parallel to the F2, F3 axes. This model contained two free parameters.

3. *PR3*. Like Version 1, Version 3 assumed that the perceptual standard deviation along the F2 dimension was equivalent to the perceptual standard deviation along with the F3 dimension (i.e., $\sigma_{F2} = \sigma_{F3}$), but allowed a perceptual dependence to exist between the perceived F2, F3 values (i.e., $\rho_{F2,F3} \neq 0$). In other words, the contours of equal likelihood for all 54 stimuli were ellipsoids of the same size and orientation whose major and minor axes were *not* required to be parallel to the F2, F3 axes. This model contained two free parameters.

4. *PR4*. This version allowed the perceptual standard deviation along the F2 dimension to be different from the perceptual standard deviation along the F3 dimension (i.e., $\sigma_{F2} \neq \sigma_{F3}$) and allowed a perceptual dependence to exist between the perceived F2, F3 values (i.e., $\rho_{F2,F3} \neq 0$). In other words, the contours of equal likelihood for all 54 stimuli were ellipsoids of the same size and orientation whose major and minor axes were *not* required to be parallel to the F2, F3 axes. This model contained three free parameters.

Response region assumptions. Four different response region assumptions were tested:

1. *SPC1*. This model assumed that a single striatal unit was associated with each of the three vowel categories, for a total of three striatal units. Under these conditions, the decision bounds that partition the F2, F3 space into three vowel categories are linear. An example is displayed in Figure 1b. This model has five free response region parameters.⁴

2. *SPC2*. This model assumed that two striatal units were associated with each of the three vowel categories, for a total of six striatal units. Under these conditions, the decision bounds that partition the F2, F3 space into three vowel categories are piecewise linear. An example is displayed in Figure 1c. This model has 12 free response region parameters.

3. *SPC3*. This model assumed that three striatal units were associated with each of the three vowel categories, for a total of nine striatal units. Under these conditions, the decision bounds that partition the F2, F3 space into three vowel categories are again piecewise linear. This model has 18 free response region parameters.

4. *SPCQC*. The response region assumptions of this model differ somewhat from those of the other three models because technically no striatal units exist. This model assumes that the partitions between response regions are quadratic in nature and that three such quadratic partitions exist. This model can be thought of as a generalization of SPC1, but in which the linear partitions are replaced by

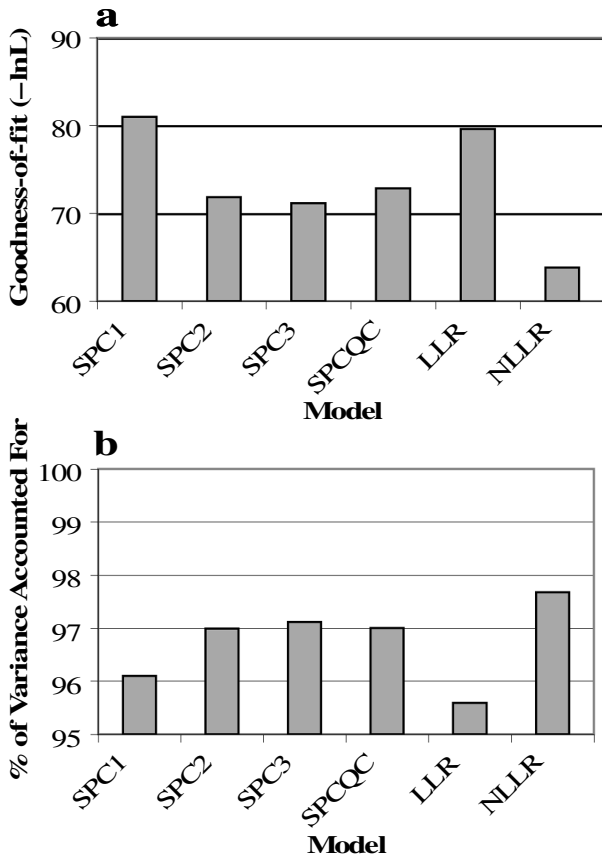


Figure 3. (a) Average goodness-of-fit ($-\ln L$), and (b) average percent of variance accounted for by the SPC and NAPP models.

quadratic partitions. This model has 11 free response region parameters.⁵

The model-based analyses began with a comparison of the four perceptual representation assumptions (PR1–PR4) for the simplest set of response region assumptions (SPC1). Because the models are nested or contain the same number of free parameters, they can be compared using likelihood ratio (G^2) tests. First, we compared the fit of PR1 with the best fitting of PR2 and PR3. For only 2 of 12 observers did the extra free parameters of PR2 or PR3 provide a significant improvement in fit [$G^2(1) = 5.88$, and 38.90 , $p < .05$]. Next, we compared PR1 with PR4 again using likelihood ratio (G^2) tests. For 7 of the 12 observers, the extra free parameters of PR4 provided a significant improvement in fit [$G^2(2)$ ranging from 7.02 to 41.86, $p < .05$]. On the basis of the slight superiority for the PR4 assumptions, and because one aim was to provide the best possible account of the data, all subsequent analyses focus on PR4. Even so, all SPC model were also fit under the PR1 assumptions and these will be discussed briefly.

The goodness-of-fit ($-\ln L$) and percent of variance accounted for by the SPC1, SPC2, SPC3, and SPCQC models (all assuming PR4) for each observer are displayed in

Table 1. The goodness-of-fit ($-\ln L$) and percent of variance accounted for by the same models averaged across observers are displayed pictorially in Figures 3a and 3b, respectively. The results can be summarized as follows. First, on the basis of the average fit, SPC2 provided a significant improvement in fit over SPC1 [$G^2(7) = 18.19$, $p < .05$]. At the individual observer level, SPC2 provided a superior account for 5 of 12 observers [$G^2(7)$ ranging from 20.10 to 52.48, $p < .05$] and a marginally superior account for 2 other observers [$G^2(7) = 12.03$ and 13.66, with a critical value of 12.02, $p < .10$]. In addition, the reduction in unexplained variance for SPC2 relative to SPC1 was nearly 25%. Second, in no case did the extra free parameters of the SPC3 model provide a significant improvement in fit over the SPC2 model at the individual observer level [$G^2(6)$ ranging from 0 to 9.42, $p > .05$] or for the average fit [$G^2(6) = 1.43$, $p > .05$]. Finally, there was little difference in fit between SPC2 and SPCQC. Because the models are not nested, AIC was used to determine which model provided the superior account of the data. For 7 of 12 observers SPCQC provided a superior account of the data, whereas SPC2 provided a better account for the remaining 5 observers. The similarity in the fits of the models is evident from an examination of the averaged AIC values and percent of responses accounted for by each model. The averaged AIC values were 173.71 and 173.63 for SPC2 and SPCQC, respectively. The averaged percent of responses accounted for were 96.99 and 97.00 for SPC2 and SPCQC, respectively.⁶ Clearly more research is needed to tease these models apart. Because the focus of this work is on an application of the SPC, we concentrate the detailed analyses on SPC2.

The piecewise linear response region partitions for the SPC2 model for each of the 12 observers are displayed as the broken lines in Figure 4. Two comments are in order. First, the only strong trend in response region partitions for SPC2 is the general concave-downward shape of the $/\sigma/ - /z^3/$ boundary. Second, the elbows in the piecewise linear partitions often fall outside of the stimulus space. Possibly, these changes in the partitions are affecting mainly the low-probability tails of the distributions, perhaps compensating for inadequate assumptions about nature of “noise” in the system. A few comments are also in order regarding the best fitting perceptual representation parameters. There were individual differences in the resulting perceptual representation parameters, but two general statements can be made. First, it was not the case that the F2 standard deviation was consistently larger or smaller than the F3 standard deviation. Specifically, for 6 of the 12 observers the F2 standard deviation was larger, for 5 of 12 the F3 standard deviation was larger, and for 1 of 12 the standard deviations were equal. Second, for 8 of 12 observers the F2, F3 percepts were negatively correlated, suggesting that perceived values of F2 that were larger than the actual F2 value were associated with perceived values of F3 that were smaller than the actual F3 value (and vice versa). We revisit this finding in the Discussion section.

Table 1
Goodness-of-Fit ($-\ln L$) and Percent of Variance Accounted for by the SPC
and NAPP Models as Applied to the Vowel Categorization Data

Observer	SPC Models				NAPP Models	
	SPC1 (9)	SPC2 (15)	SPC3 (21)	SPCQC (18)	LLR (6)	NLLR (12)
Goodness-of-Fit ($-\ln L$)						
1	69.73	69.71	69.67	69.23	71.94	63.15
2	47.25	41.24	39.65	41.74	48.78	39.20
3	104.70	97.87	93.16	92.83	95.10	79.79
4	108.01	82.08	81.83	76.15	97.04	67.77
5	63.81	62.53	62.45	61.79	71.54	54.90
6	107.41	106.06	106.06	107.35	100.66	82.33
7	69.11	59.06	58.56	67.48	67.97	61.91
8	38.26	37.61	37.57	38.15	39.84	38.14
9	85.93	59.69	59.46	63.52	84.35	58.99
10	93.63	83.33	82.44	84.23	102.75	75.79
11	81.13	75.24	75.16	80.76	85.72	76.52
12	102.38	87.81	87.62	90.54	89.49	66.83
Average	80.95	71.85	71.14	72.81	79.60	63.78
Percent of Variance						
1	98.49	98.50	98.67	98.64	98.60	98.66
2	97.47	97.50	98.14	97.99	97.99	98.65
3	92.80	93.67	94.04	94.02	94.00	95.92
4	95.35	97.50	97.56	97.96	95.47	98.36
5	97.66	97.75	97.84	98.27	97.17	98.94
6	93.32	93.52	93.52	93.34	94.22	95.72
7	97.01	98.32	98.35	96.80	97.02	97.87
8	98.66	98.78	98.79	98.68	98.28	98.53
9	92.56	96.61	96.77	97.09	93.73	97.18
10	94.46	95.60	95.54	95.54	93.77	96.84
11	97.62	97.69	97.71	97.68	96.48	96.91
12	97.77	98.47	98.49	97.94	90.37	98.53
Average	96.10	96.99	97.12	97.00	95.59	97.68

Note—Number of parameters in parentheses.

NAPP model. It is straightforward to show that the linear version of the NAPP model is equivalent to a system of three linear logistic equations (Lachenbruch, 1975; Nearey, 1990) and thus can be fit using readily available “canned” statistical software. In addition, the quadratic version of the NAPP model is equivalent to nonlinear logistic models and can also be fit using statistical software. Although we will be fitting logistic regression models to the data, these applications were motivated by the psychologically meaningful NAPP model, so the models will be referred to as NAPP-LLR and NAPP-NLLR for applications of linear and nonlinear logistic regression, respectively.

The number of free parameters, goodness-of-fit ($-\ln L$), and percent of variance accounted for by the NAPP-LLR and NAPP-NLLR models for each observer are presented in Table 1. The averaged fit and percent of variance accounted for are displayed in Figure 3. The results were clear. A G^2 test comparing the average fit values for the NAPP-NLLR and NAPP-LLR models was significant [$G^2(6) = 31.64, p < .001$]. In addition, for 10 of 12 observers, the NAPP-NLLR model provided a significant improvement in fit over the NAPP-LLR model [$G^2(6)$ ranging from 17.57 to 58.53, $p < .05$]. Finally, the reduction in unexplained variance for NAPP-NLLR relative to

NAPP-LLR was nearly 50%. Taken together these results provide strong support for the use of nonlinear response region partitions within the framework of the NAPP model. The nonlinear response region partitions for the NAPP-NLLR model for each of the 12 observers are displayed as the solid curves in Figure 4.

SPC and NAPP comparisons. NAPP was developed in the speech perception domain and has been quite successful; thus we did not expect the SPC to provide as good an account of the data. Even so, some comparisons with NAPP are in order. First, we compared the SPC model that assumed linear response region partitions (i.e., SPC1) with the linear version of NAPP (i.e., NAPP-LLR). Because the models are not nested, we used AIC to compare the models. As expected, the NAPP-LLR model was superior to SPC1. On the basis of the average AIC, for 9 of 12 observers the NAPP-LLR model provided a better account of the data. Even so, the percent of variance accounted for was high, and was essentially identical, with a slight advantage for SPC1 (percent of variance accounted for: SPC1 = 96.10%; NAPP-LLR = 95.59%). In addition, the response region partitions were similar across models. Second, we compared the piecewise linear SPC2 with the NAPP-NLLR. Again, the NAPP model

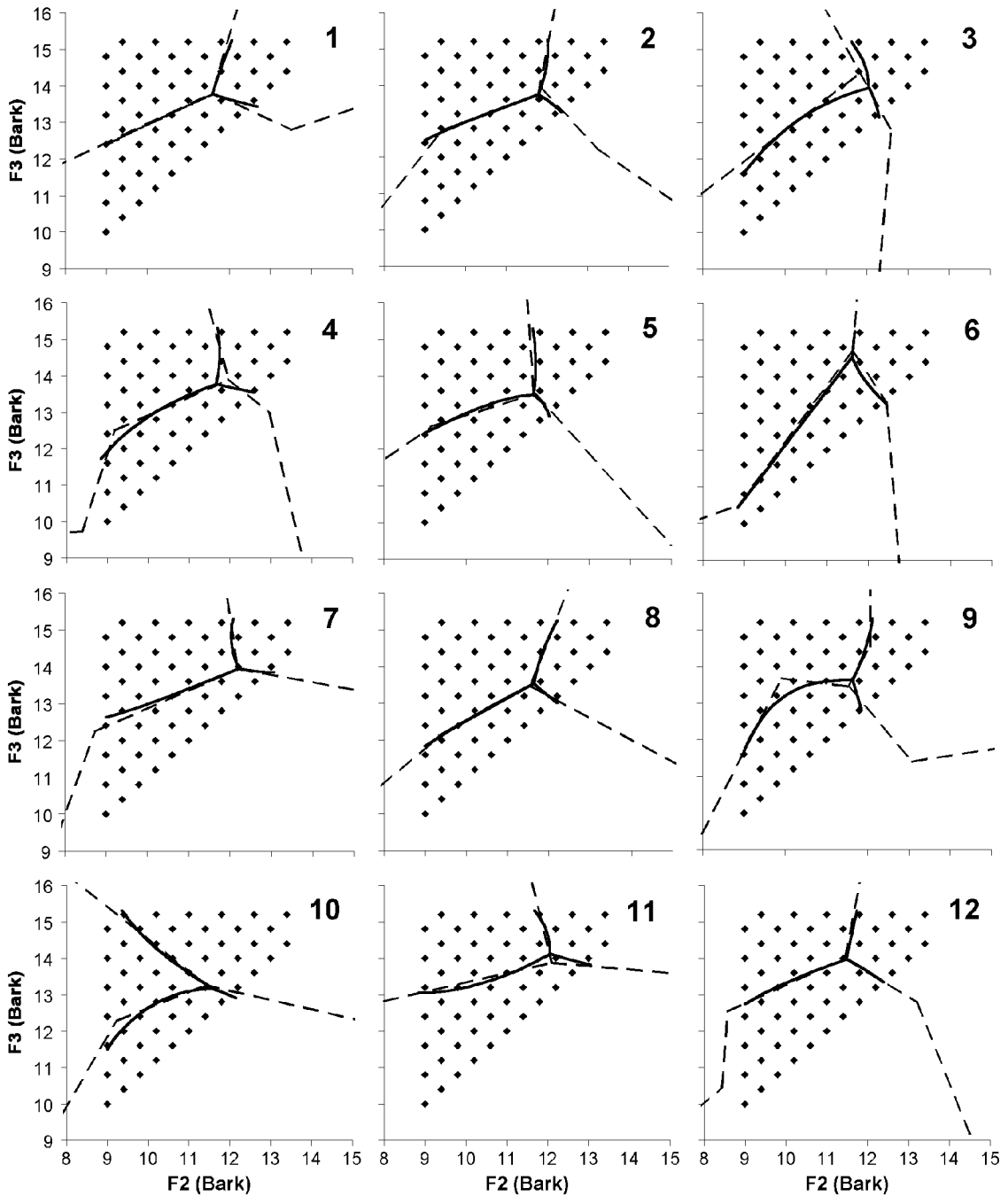


Figure 4. Response region partitions for the piecewise linear SPC2 model (broken lines) and for the NAPP-NLLR model (solid curves) for each of the 12 listeners.

was superior but in this case the superiority was larger. The AIC advantage was larger for the average fit, and was larger for all 12 observers. Even so, the percent of variance accounted for showed less than a percentage point improvement (96.99% and 97.68% for the SPC2 and NAPP-NLLR models, respectively), although importantly this

does denote nearly a 25% reduction in unexplained variance. The fact that versions of both models that assumed linear response region partitions were rejected suggests that the additional flexibility afforded by nonlinear decision rules was necessary to capture the listener's vowel categorization behavior. As Figure 4 suggests, both mod-

els partitioned the perceptual space similarly, although notable exceptions are evident.

DISCUSSION

This article reports the results of an auditory vowel categorization experiment in which listeners classified 54 synthetic vowel stimuli that varied along the F2 and F3 dimensions into one of three vowel categories, /i/, /u/, and /ɜ/. A successful, neuropsychologically plausible model of categorization in the visual domain, the SPC, was generalized to the auditory domain and was applied separately to each listener's data from the auditory vowel categorization task. A version of the SPC that assumed two striatal units per category, and thus piecewise linear response region partitions, provided a good description of the data, accounting on average for 97% of the variance in the data. This finding is important because it suggests that a model with a reasonable neurobiological architecture can be applied in both the visual and auditory domains. This provides an important step toward bridging the gap between visual and auditory categorization and toward a neurobiological understanding of the systems involved in these two different, but related, forms of categorization.

A version of logistic regression that assumed nonlinear response region partitions (NAPP-NLLR) provided a better account of the data than a version that assumed linear partitions (NAPP-LLR). The linear versions of the SPC and NAPP provided approximately equal accounts of the data, although there was a slight but consistent advantage for the NAPP-LLR model. The nonlinear version of logistic regression (NAPP-NLLR) on the other hand provide a larger and consistent performance improvement in AIC fit over the piecewise linear version of the SPC. Despite the large AIC difference, the predictive power of the models was approximately equal. Specifically, we computed the absolute value of the deviation between predicted and observed response probabilities for each stimulus and then averaged these values. Averaged across observers, the average absolute deviation between predicted and observed response probabilities was .036 for the SPC2 model and .034 for the NAPP-NLLR model, essentially the same value. Interestingly, at the individual observer level, the average absolute deviation was *smaller* for the SPC2 model for 5 of the 12 observers, even though the AIC value was larger. One possibility is that the SPC2 is yielding one or two large mispredictions and that these mispredictions are elevating the AIC value. This possibility is supported by the fact that the maximum absolute deviation was larger for the SPC2 model for 10 of the 12 observers. Across these 10 observers the average of the maximum absolute deviation was .30 for the SPC2 model and .25 for the NAPP-NLLR model, a fairly large difference.

Response Region Nonlinearity

The superiority of the nonlinear version of the NAPP and of the piecewise linear version of the SPC over their corresponding linear versions is noteworthy in view of cer-

tain earlier findings: Vowel category boundaries in Swedish (Carlson, Granström, & Fant, 1970), Russian (Karnickaya, Mushikov, Slepokurova, & Zhukov, 1975), and German (Hose, Langner, & Scheich, 1983) have been reported to be approximately linear when formant frequencies are scaled in units similar to Bark (e.g., Mel or log units). However, in these earlier studies no comparisons were made between linear and nonlinear models of the data, so it is unclear whether the results are actually at odds with the present findings. Given the fact that the perceptual distributions (i.e., the perceptual covariance matrices) likely differ substantially across vowels, it is not surprising that complex response region partitions are necessary to account for vowel categorization data. A series of cross-language studies of vowel perception are now being planned in our laboratory to explore further the issue of boundary linearity/nonlinearity.

Relations Between the SPC and NAPP Models

In this section we compare and contrast the architectures of the different models. One important distinction between the SPC and the NAPP-based logistic regression models is that the SPC makes specific assumptions about perceptual processing of each individual stimulus. In short, the SPC is in accordance with multidimensional signal detection theory (Ashby & Townsend, 1986; Green & Swets, 1967; Macmillan & Creelman, 1991) and assumes that repeated presentations of a single stimulus yield different perceptual effects (i.e., perceptual noise exists). The NAPP model makes no strong claim about the perceptual processing of individual stimuli, although in its present form it assumes implicitly no perceptual noise. Thus, the SPC assumes a probabilistic perceptual representation, whereas the NAPP model assumes a deterministic perceptual representation.

The SPC assumes that the observer determines the location of the perceptual effect on each trial and gives the categorization response associated with that region of the perceptual space. Thus, each location (or perceptual effect) in the space is deterministically associated with one of the categorization responses. The NAPP model, on the other hand, assumes that each stimulus is processed by a set of "detectors" or "filters" that are tuned to language-specific vowel categories. These detectors produce normally distributed outputs that correspond roughly to the likelihood of the vowel category given the stimulus input. These category likelihoods are then combined using the relative goodness rule to determine a categorization response probability for each category. Thus, each location in the space is probabilistically associated with one of the categorization responses.

Although NAPP is described as using the Luce choice rule (Nearey, 1990), a probabilistic response rule, to generate predicted response probabilities, there is an alternate, mathematically equivalent, interpretation that assumes a deterministic response rule. This is referred to as the "extremal Thurstonian model" (Albert & Chib, 1993; Bock, 1975).⁷ In this interpretation the "filter" outputs are from a Type 1 extreme value distribution, and the filter

with the highest output is chosen on each trial. Under this interpretation the outputs are noisy, but the response rule is deterministic. Thus, both the SPC and extremal Thurstonian interpretation of NAPP assume a deterministic response rule. The models differ only in the “noise” assumptions with the SPC assuming noise in the perceptual representation or input and the NAPP model assuming noise in the output.

The fact that the NAPP models were superior to the SPC models might lead one to conclude that the “noise” assumptions of NAPP are supported over the noise assumptions of the SPC. From a purely goodness-of-fit perspective this is a reasonable conclusion. However, there is good reason to withhold judgment. First, it is well established from physiology and psychophysics that noise exists in all sensory systems (see Geisler, 1989, for an excellent review regarding vision; see also Ashby & Lee, 1993). Thus it seems likely that the “correct” model should assume a probabilistic perceptual representation. One possibility is that the normality assumption of the SPC is incorrect. Although this assumption fares well with visual stimuli, it is possible that some other distribution better captures perceptual noise in audition. Another possibility is that the assumption that the perceptual variance-covariance matrix is identical across stimuli (i.e., the stimulus invariance assumption) is incorrect. Our intuition is that this assumption is generally incorrect and that the necessity to relax it depends most on the procedural aspects of the experiment. For example, with highly confusable stimuli, short exposure durations, or masked displays, the influence of perceptual noise on performance is increased, and the stimulus invariance assumption will more likely need to be relaxed (e.g., Maddox, 2001, 2002). In the present study, though, the stimuli are highly discriminable, are not masked, and are of reasonable exposure durations. Thus, it is highly likely that the additional perceptual flexibility afforded the SPC (like violations of stimulus invariance) is not necessary to account for the present data. Despite these important issues, the pure predictive power of the logistic regression models should not be undervalued.

SPC Perceptual Representation Parameters

The best fitting parameters from the most parsimonious SPC model suggested that high perceived values of F2 were associated with low perceived values of F3; that is, there was a negative perceptual dependence between perceived F2 and F3 values. The latter effect may reflect a type of auditory integration whereby formants located within 3 Bark (or critical band) units are averaged into a single spectral prominence with a “center of gravity” located between the nominal values of the two formants (Chistovich & Lublinskaya, 1979). By this account, raising the frequency of the second formant so that it enters the region of spectral integration with the third formant would, in effect, lower the perceived value of F3. Unfortunately, the model-based analyses to this point do not provide a rigorous test of this hypothesis since the perceptual dependence parameter was constant across all stimuli, even those that

differ substantially in F2 and F3. If this spectral integration hypothesis is correct, then the perceptual dependence should emerge only for those stimuli with closely spaced formants. As a better test of this hypothesis we separated the stimuli into those that differed by fewer than 3 Bark units and those that differed by more than 3 Bark units. We then refit the SPC2 model with the PR4 perceptual representation assumptions, but with one caveat; we assumed perceptual independence for the stimuli separated by more than 3 Bark units and allowed a perceptual dependence to hold for the stimuli separated by fewer than 3 Bark units. Notice that this model has the same number of free parameters as the SPC4 model displayed in Table 1 and Figure 3. The results were mixed. For 7 of the 12 observers the 3-Bark model provided a superior account of the data. Restricting attention to these 7 observers, in 6 of the 7 cases the correlation was negative, although the absolute magnitude was small (average correlation $-.08$). Even though these results hint at a possible effect of spectral integration, the small size of this effect and the variability across listeners appear to cast doubt on the role of integration as a significant factor in vowel perception. (See Diehl, 2000, for additional evidence against the spectral integration hypothesis.)

The SPC as a Model of Category Learning

The SPC is a neurobiologically plausible model of implicit category learning. Although it is likely that vowel categories are learned implicitly, this was not a study of vowel category *learning*; rather, this study focused on categorization of highly learned categories. The current thinking is that highly learned categories are represented in cortex and that this cortical learning is mediated by subcortical learning in structures such as the striatum (Ashby et al., 1998). Thus, it seems reasonable to assume that the cortical representation is analogous in some respect to the subcortical representation that emerges through extensive practice. In other words, the subcortical representation after extensive learning provides a window onto the cortical representation. It is in this light that we use the SPC in this study.

If the SPC mediates cortical learning, and given the truly remarkable ability that humans have to perceive and understand speech sounds, it is likely that the subcortical, striatal representation is quite flexible. To put it more concretely, it is likely that a large number of striatal units exist and that the number necessary to adequately solve a categorization problem is determined by the complexity of the problem to be solved. This conjecture is supported by the finding that piecewise linear and quadratic response region partitions were necessary to adequately account for the data.

Future Directions

There are many directions for future research. One weakness of the present study is that the F2 \times F3 stimulus space was fairly sparsely sampled. Under these conditions, responding is very consistent, and the same vowel

category is generally chosen for repeated presentations of the same stimulus. This yields response probabilities near 0 or 1 for the vowel categories and thus makes it difficult to compare rigorously the performance of each model to other models. In short, the best test of a model results when response probabilities are intermediate between 0 and 1. Future research should use stimuli that more densely sample the space. Data of this kind will provide a more fertile testing ground for the various models.

Another fruitful approach will be to have listeners perform categorization and identification using the same fixed set of auditory stimuli. In categorization the listeners would be required to categorize each stimulus into one of the three vowel categories. In identification, on the other hand, each vowel stimulus would have its own unique response label. SPC models of identification (Ashby et al., 2001; Maddox, 2001, 2002) could then be used to provide an estimate of the underlying perceptual representations for each vowel stimulus. These could then be held fixed in an attempt to model the vowel categorization performance. The advantages of this approach are many. First, a better estimate of the true perceptual representation can be garnered from the identification data. Second, the complex interplay between perceptual and decisional (response region) processes can be investigated by comparing and contrasting the performance of models that make different assumptions about processing at each level. In particular, the effects of learning can be examined (Maddox, 2002). Finally, a more complete understanding of auditory stimulus processing can be obtained by modeling performance simultaneously across these two different, but related, tasks (Ashby & Lee, 1991; Maddox & Ashby, 1996; see also Maddox & Bogdanov, 2000).

Speech perception research and modeling is also ripe for cognitive neuroscience applications. For example, if the SPC has merit, it should be difficult for patients with striatal lesions, such as patients with Parkinson's or Huntington's disease, to learn speech or vowel categories. In addition, brain-imaging techniques could be brought to bear. Cognitive neuroscience applications in the visual categorization domain are growing rapidly, but could be informed by similar studies with auditory stimuli.

To summarize, the present study generalized and applied a successful model of visual categorization to the categorization of vowel sounds. The model provided a good account of the data and offers an important bridge between visual and auditory categorization, and toward a neurobiological understanding of the systems involved in these two forms of categorization. Although competitive, the model was consistently outperformed by the NAPP model of speech recognition.

REFERENCES

- AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716-723.
- ALBERT, J. H., & CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, **88**, 669-679.
- ARNAULD, E., JEANTET, Y., ARSAUT, J., & DESMOTES-MAINARD, J. (1996). Involvement of the caudal striatum in auditory processing: *c-fos* response to cortical application of picrotoxin and to auditory stimulation. *Molecular Brain Research*, **41**, 27-35.
- ASHBY, F. G. (1988). Estimating the parameters of multidimensional signal detection theory from simultaneous ratings on separate stimulus components. *Perception & Psychophysics*, **44**, 195-204.
- ASHBY, F. G. (1992a). Multidimensional models of categorization. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 449-484). Hillsdale, NJ: Erlbaum.
- ASHBY, F. G. (1992b). Multivariate probability distributions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 1-34). Hillsdale, NJ: Erlbaum.
- ASHBY, F. G., & ALFONSO-REESE, L. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology*, **39**, 216-233.
- ASHBY, F. G., ALFONSO-REESE, L. A., TURKEN, A. U., & WALDRON, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, **105**, 442-481.
- ASHBY, F. G., & ELL, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences*, **5**, 204-210.
- ASHBY, F. G., & LEE, W. W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General*, **120**, 150-172.
- ASHBY, F. G., & LEE, W. W. (1993). Perceptual variability as a fundamental axiom of perceptual science. In S. C. Masin (Ed.), *Foundations of perceptual theory* (pp. 369-399). Amsterdam: North-Holland.
- ASHBY, F. G., & MADDOX, W. T. (1991). A response time theory of perceptual independence. In J. P. Doignon & J. C. Falmagne (Eds.), *Mathematical psychology: Current developments* (pp. 389-414). New York: Springer-Verlag.
- ASHBY, F. G., & MADDOX, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, **37**, 372-400.
- ASHBY, F. G., & MADDOX, W. T. (1994). A response time theory of separability and integrality in speeded classification. *Journal of Mathematical Psychology*, **38**, 423-466.
- ASHBY, F. G., & MADDOX, W. T. (1998). Stimulus categorization. In M. H. Birnbaum (Ed.), *Measurement, judgment, and decision making: Handbook of perception and cognition* (pp. 251-301). San Diego: Academic Press.
- ASHBY, F. G., MADDOX, W. T., & LEE, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. *Psychological Science*, **5**, 144-150.
- ASHBY, F. G., & PERRIN, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review*, **95**, 124-150.
- ASHBY, F. G., & TOWNSEND, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, **93**, 154-179.
- ASHBY, F. G., & WALDRON, E. M. (1999). On the nature of implicit classification. *Psychonomic Bulletin & Review*, **6**, 363-378.
- ASHBY, F. G., WALDRON, E. M., LEE, W. W., & BERKMAN, A. (2001). Suboptimality in categorization and identification. *Journal of Experimental Psychology: General*, **130**, 77-96.
- BOCK, R. D. (1975). *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill.
- CARLSON, R., GRANSTRÖM, B., & FANT, G. (1970). Some studies concerning perception of isolated vowels. *Speech Transmission Laboratory: Quarterly Progress & Status Report*, **2/3**, 19-35.
- CHISTOVICH, L. A., & LUBLINSKAYA, V. V. (1979). The "center of gravity" effect in vowel spectra and critical distance between the formants. *Hearing Research*, **1**, 185-195.
- CHUDLER, E. H., SUGIYAMA, K., & DONG, W. K. (1995). Multisensory convergence and integration in the neostriatum and globus pallidus of the rat. *Brain Research*, **674**, 33-45.
- DIEHL, R. L. (2000). Searching for an auditory description of vowel categories. *Phonetica*, **57**, 267-274.
- ESTES, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, **53**, 134-140.
- FILOTEU, J. V., MADDOX, W. T., & DAVIS, J. D. (2001). Quantitative modeling of category learning in patients with amnesia. *Journal of the International Neuropsychological Society*, **7**, 1-19.

- GAZZANIGA, M. S. (2000). *The cognitive neurosciences* (2nd ed.). Cambridge, MA: MIT Press.
- GEISLER, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, **96**, 267-341.
- GREEN, D. M., & SWETS, J. A. (1967). *Signal detection theory and psychophysics*. New York: Wiley.
- HILLENBRAND, J. M., 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.
- HOSE, B., LANGNER, G., & SCHEICH, H. (1983). Linear phoneme boundaries for German synthetic two-formant vowels. *Hearing Research*, **9**, 13-25.
- JOG, M. S., KUBOTA, Y., CONNOLLY, C. I., HILLEGAART, V., & GRAYBIEL, A. M. (1999). Building neural representations of habits. *Science*, **286**, 1745-1749.
- KARNICKAYA, E. G., MUSHIKOV, V. N., SLEPOKUROVA, N. A., & ZHUKOV, S. J. (1975). Auditory processing of steady-state vowels. In G. Fant & M. A. A. Tathan (Eds.), *Auditory analysis and the perception of speech* (pp. 37-53). New York: Academic Press.
- KINGSTON, J., & MACMILLAN, N. A. (1995). Integrality of nasalization and F_1 in vowels in isolation and before oral and nasal consonants: A detection-theoretic application of the Garner paradigm. *Journal of the Acoustical Society of America*, **97**, 1261-1285.
- KNOWLTON, B. J., MANGELS, J. A., & SQUIRE, L. R. (1996). A neostriatal habit learning system in humans. *Science*, **273**, 245-254.
- LACHENBRUCH, P. A. (1975). *Discriminant analysis*. New York: Hafner.
- LUCE, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 103-189). New York: Wiley.
- MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
- MACMILLAN, N. A., & KINGSTON, J. (1995). Integrality, correspondence, and configularity represent different degrees of perceptual interaction, not different types. In C. A. Possamai (Ed.), *Fechner Day 95* (pp. 243-248). Cassis, France: International Society for Psychophysics.
- MADDOX, W. T. (1992). Perceptual and decisional separability. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 147-180). Hillsdale, NJ: Erlbaum.
- MADDOX, W. T. (1999). On the dangers of averaging across observers when comparing decision bound and generalized context models of categorization. *Perception & Psychophysics*, **61**, 354-374.
- MADDOX, W. T. (2001). Separating perceptual processes from decisional processes in identification and categorization. *Perception & Psychophysics*, **63**, 1183-1200.
- MADDOX, W. T. (2002). Learning and attention in multidimensional identification, and categorization: Separating low-level perceptual processes and high-level decisional processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 99-115.
- MADDOX, W. T., & ASHBY, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, **53**, 49-70.
- MADDOX, W. T., & ASHBY, F. G. (1996). Perceptual separability, decisional separability, and the identification-speeded classification relationship. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 795-817.
- MADDOX, W. T., & ASHBY, F. G. (1998). Selective attention and the formation of linear decision boundaries: Comment on McKinley and Nosofsky (1996). *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 301-321.
- MADDOX, W. T., & BOGDANOV, S. V. (2000). On the relation between decision rules and perceptual representation in multidimensional perceptual categorization. *Perception & Psychophysics*, **62**, 984-997.
- MADDOX, W. T., & FILOTEO, J. V. (2001). Striatal contributions to category learning: Quantitative modeling of simple linear and complex non-linear rule learning in patients with Parkinson's Disease. *Journal of the International Neuropsychological Society*, **7**, 710-727.
- MCDONALD, R. J., & WHITE, N. M. (1993). A triple dissociation of memory systems: Hippocampus, amygdala, and dorsal striatum. *Behavioral Neuroscience*, **107**, 3-22.
- NEAREY, T. M. (1990). The segment as a unit of speech perception. *Journal of Phonetics*, **18**, 347-373.
- NEAREY, T. M. (1997). Speech perception as pattern recognition. *Journal of the Acoustical Society of America*, **101**, 3241-3254.
- NEAREY, T. M., & HOGAN, J. (1986). Phonological contrast in experimental phonetics: Relating distributions of measurements production data to perceptual categorization curves. In J. Ohala & J. Jaeger (Eds.), *Experimental phonology* (pp. 141-161). New York: Academic Press.
- NOSSAIR, Z. B., & ZAHORIAN, S. A. (1991). Dynamic spectral shape features as acoustic correlates for initial stop consonants. *Journal of the Acoustical Society of America*, **89**, 2978.
- PACKARD, M. G., & MCGAUGH, J. L. (1992). Double dissociation of fornix and caudate nucleus lesions on acquisition of two water maze tasks: Further evidence for multiple memory systems. *Behavioral Neuroscience*, **106**, 439-446.
- PERRIN, N. A., & ASHBY, F. G. (1991). A test for perceptual independence with dissimilarity data. *Applied Psychological Measurement*, **15**, 79-93.
- PICKERING, A. D. (1997). New approaches to study of amnesic patients: What can a neurofunctional philosophy and neural network methods offer? *Memory*, **5**, 255-300.
- POLDRACK, R. A., PRABHAKARAN, V., SEGER, C. A., & GABRIELI, J. D. E. (1999). Striatal activation during acquisition of a cognitive skill. *Neuropsychology*, **13**, 564-574.
- SEGER, C. A., POLDRACK, R. A., PRABHAKARAN, V., ZHAO, Z., GLOVER, G. H., & GABRIELI, J. D. E. (2000). Hemispheric asymmetries and individual differences in visual concept learning as measured by functional MRI. *Neuropsychologia*, **38**, 1316-1324.
- SMITH, E. E., PATALANO, A. L., & JONIDES, J. (1998). Alternative strategies of categorization. *Cognition*, **65**, 167-196.
- SMITH, J. D., & MINDA, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1411-1436.
- TAKANE, Y., & SHIBAYAMA, T. (1992). Structures in stimulus identification data. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 335-362). Hillsdale, NJ: Erlbaum.
- TRAUNMÜLLER, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America*, **88**, 97-100.
- WALDRON, E. M., & ASHBY, F. G. (2001). The effects of concurrent task interference in category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin & Review*, **8**, 168-176.
- WALDRON, E. M., ELL, S. W., ASHBY, F. G., MCCORMICK, M., & CASALE, M. (2000, March). *A striatal-based model of implicit category learning*. Poster presented at the annual meeting of the Cognitive Neuroscience Society, San Francisco.
- WICKENS, J. (1993). *A theory of the striatum*. New York: Pergamon.
- WICKENS, T. D. (1982). *Models for behavior: Stochastic processes in psychology*. San Francisco: Freeman.

NOTES

1. A second form of perceptual interaction predicted by GRT is called *perceptual separability*. Perceptual separability holds when the perceptual effects of a particular component are unaffected by the level of the other component—that is, if the distribution of percepts associated with one component is unaffected by the level of the other component (see Ashby & Townsend, 1986; Maddox, 1992 for details). In Figure 1a, dimension y is perceptually separable from dimension x , but dimension x is not perceptually separable from dimension y . To manage the number of free parameters in the model and to provide a straightforward comparison with the NAPP model, all SPC models tested in this article assumed that perceptual separability was satisfied along both stimulus dimensions.

2. A major problem in model fitting is to avoid “local minima.” Local minima result when the parameter estimation algorithm identifies a set of parameters that are “locally” superior (i.e., provide the best fit) but not “globally” superior. Although one can never be certain that they have obtained the global minima, procedures should be followed to minimize this possibility. Here, multiple starting parameter values were utilized. In addition, since the models were nested, the parameters from a more restricted model were used as starting values for a more general model. These procedures likely improve the chances of identifying the global minima.

3. The details of the numerical integration procedure can be found in Ashby (1992b) and Ashby et al. (2001), but will be briefly outlined here.

In short, one needs to evaluate Equation 1 for an arbitrary bivariate normal distribution, $f_i(x,y)$, and an arbitrary response region, \mathcal{R}_i . Although a number of procedures are available, the approach taken in our lab is to apply numerical integration using Cholesky factorization (see Ashby, 1992b, pp. 24–26). The steps can be summarized as follows. First, using the Cholesky factorization matrix, one transforms the bivariate normal distribution $f_i(x,y)$ into a bivariate Z distribution. Second, the bivariate Z distribution is subdivided into $n \times n$ equi-probability regions (in the present case, $n = 10$). Third, the centroid for region 1×1 of the bivariate Z distribution is transformed back into the x,y space (via the inverse Cholesky matrix). Fourth, the response region \mathcal{R}_i that contains the transformed centroid is determined, and the probability (under the bivariate Z distribution) associated with that centroid is assigned to response \mathcal{R}_i . This process is repeated for all $n \times n$ regions and for all i stimuli. If the resulting probability was zero, we arbitrarily set it to .0001. This was necessary to fit the model.

4. Since each striatal unit is defined by its location in the F2, F3 space, one might conclude that this model requires six free parameters. In fact, in fitting the model, six parameters were freely estimated. However, a five-parameter representation exists that yields exactly the same response region partitions. To see this, notice that by connecting the three striatal units, one can form a triangle. The perpendicular bisectors of the sides of the triangle determine the response region partitions, and the bisectors all converge at the center of the triangle. Thus, the three response region partitions can be defined by three slopes (one per bisector) and the location of the center of the triangle in F2, F3 space (we thank Terrance Nearey for suggesting this useful geometric representation).

5. Since each quadratic partition is defined by five coefficients one might conclude that this model requires 15 free parameters. In fact, in fitting the model, 15 parameters were freely estimated. Three bivariate normal “category response region” distributions were defined. Each of these

has 5 free parameters (an F2 mean, an F3 mean, an F2 variance, an F3 variance, and an F2, F3 covariance). For each perceptual effect, the likelihood under each of the three “category response region” distributions was computed, and the perceptual effect was assigned to the category with the highest likelihood. Under these conditions the response region partitions will be quadratic and will meet at a fixed location in F2, F3 space. However, an 11-parameter representation exists that yields exactly the same response region partitions, and thus this model has only 11 effective parameters. Unfortunately, no straightforward geometric explanation is available.

6. The same model comparisons were also made for the PR1 assumptions. The results mirrored those for PR4. Specifically, SPC2 provided a significant improvement in fit over SPC1 on average [$G^2(7) = 19.41$, $p < .05$], and for 6 of 12 observers [$G^2(7)$ ranging from 14.92 to 58.98, $p < .05$]. SPC3 did not provide a significant improvement in fit for the average or for any observer. Finally, for 5 of 12 observers SPCQC provided a superior account of the data (based on AIC), whereas SPC2 provided a better account for the remaining 7 observers (averaged AIC = 183.88 and 187.89 for SPC2 and SPCQC, respectively; averaged percent of variance accounted for = 96.28% and 96.32% for SPC2 and SPCQC, respectively).

7. Mathematical equivalencies like those described between the Luce choice version of NAPP and the extremal Thurstonian version of NAPP are not uncommon. In fact, Ashby and Maddox (1993; see also Ashby & Alfonso-Reese, 1995) derived similar relations among a wide variety of categorization models that differed substantially in their assumptions about the perceptual representation and about response selection.

(Manuscript received November 20, 2000;
revision accepted for publication September 12, 2001.)