

Syllable structure and integration of voicing and manner of articulation information in labial consonant identification

Noah H. Silbert^{a)}

Center for Advanced Study of Language, University of Maryland, 7005 52nd Avenue, College Park, Maryland 20742

(Received 23 February 2011; revised 13 January 2012; accepted 2 March 2012)

Speech perception requires the integration of information from multiple phonetic and phonological dimensions. A sizable literature exists on the relationships between multiple phonetic dimensions and single phonological dimensions (e.g., spectral and temporal cues to stop consonant voicing). A much smaller body of work addresses relationships between phonological dimensions, and much of this has focused on sequences of phones. However, strong assumptions about the relevant set of acoustic cues and/or the (in)dependence between dimensions limit previous findings in important ways. Recent methodological developments in the general recognition theory framework enable tests of a number of these assumptions and provide a more complete model of distinct perceptual and decisional processes in speech sound identification. A hierarchical Bayesian Gaussian general recognition theory model was fit to data from two experiments investigating identification of English labial stop and fricative consonants in onset (syllable initial) and coda (syllable final) position. The results underscore the importance of distinguishing between conceptually distinct processing levels and indicate that, for individual subjects and at the group level, integration of phonological information is partially independent with respect to perception and that patterns of independence and interaction vary with syllable position. © 2012 Acoustical Society of America. [http://dx.doi.org/10.1121/1.3699209]

PACS number(s): 43.71.An, 43.71.Es, 43.71.Sy [PEI]

Pages: 4076–4086

I. MULTIPLE DIMENSIONS IN SPEECH PERCEPTION

As speech sounds are mapped onto meaningful linguistic representations, information from multiple acoustic-phonetic and phonological dimensions must be combined. Consider, for example, an utterance beginning with the voiceless labiodental fricative [f]. At any given point in time, the signal will be perceived as more or less [f]-like and more or less [v]-, [θ]-, or [p]-like. That is, the percept will be more or less (phonologically) voiced, labial, and fricative-like (as opposed to voiceless, coronal, or stop-like).

In the absence of noise, a clearly articulated [f] will be perceived as more [f]-like than [v]-, [θ]-, or [p]-like, although in many common settings, this will not be the case. In order to accurately perceive a spoken [f], information from voicing, place of articulation, and manner of articulation dimensions (among others) must be combined. It is possible that phonological dimensions are combined independently and that a phoneme is a simple combination of its component parts. Or the opposite may be true and phonological dimensions may interact, producing perceptual effects that are qualitatively different than simple combinations of component parts.

Investigations of phonological information integration are complicated by the fact that the mapping between acoustic-phonetic dimensions and phonological dimensions is many-to-many. For example, in English fricatives, high frequency energy (i.e., above 750 Hz) and spectral shape

(as measured by spectral moments) vary with voicing distinctions, but both also play a role in differentiating place of articulation (see, e.g., Silbert and de Jong, 2008).

Previous work on the perceptual relationships between phonetic cues to distinctive features has been limited by strong assumptions about the nature of the relationships between dimensions, confluences of levels at which dimensions may interact, or both. The general recognition theory (GRT) framework (Ashby and Townsend, 1986; Silbert *et al.*, 2009; Thomas, 2001) enables these limitations to be addressed directly. GRT provides rigorous distinctions between independence of dimensions in decision-making and at multiple levels of perception.

The primary purpose of the present work is to demonstrate of the utility of GRT in modeling distinct perceptual and decisional processes and distinct levels of interaction in speech perception. The secondary purpose of this work is to establish a baseline set of results concerning (a) the relationships between manner and voicing distinctions in English and (b) the modulation of these relationships by syllable structure.

A. Category structure on a single phonological dimension

As noted previously, each phonological contrast may have multiple acoustic cues, and acoustic cues may be “shared” by different phonological contrasts. For example, voicing in stop consonants in syllable onset position can be cued by VOT (voice onset time; see, e.g., Lisker and Abramson, 1964; Volaitis and Miller, 1992; Kessinger and Blumstein, 1997), although consonant release-burst amplitude, aspiration noise amplitude, and the frequency of F0

^{a)}Author to whom correspondence should be addressed. Electronic mail: nsilbert@umd.edu

and F1 at voice onset may also cue stop voicing (Kingston and Diehl, 1994; Oglesbee, 2008).

Place of articulation distinctions also map onto multiple acoustic-phonetic properties. For example, place of articulation in stop consonants can be cued by differences in the spectral structure and amplitude of the release-burst (Blumstein and Stevens, 1979; Forrest *et al.*, 1988; Stevens and Blumstein, 1978); spectral shape and noise amplitude also cue place in fricatives (Jongman *et al.*, 2000; Silbert and de Jong, 2008). The location and movement of vowel formants near the consonant-vowel boundary also vary as a function of place in obstruent consonants (Jongman *et al.*, 2000; Kurowski and Blumstein, 1984).

The perceptual structure of phonological categories seems to depend, at least in part, on the statistical properties of associated acoustic-phonetic cues. For example, categorization rates and category goodness judgments for stop consonants vary as a function of VOT, and this variation depends on speaking rate and place of articulation (Miller and Volaitis, 1989; Volaitis and Miller, 1992). Recent work probing a large number of acoustic-phonetic dimensions indicates that a listener's native language plays an important role in determining the internal structure of phonological categories (Oglesbee, 2008).

In a few cases, research on the internal structure of phonological categories has focused explicitly on interactions between acoustic-phonetic dimensions. Spectral correlates of nasalization and F1 frequency seem to interact perceptually (Kingston and MacMillan, 1995; MacMillan *et al.*, 1999), as do voice quality and F1 frequency (Kingston *et al.*, 1997). More recently, perceptual interactions between a number of spectral and temporal cues to voicing in intervocalic consonants have been reported (Kingston *et al.*, 2008).

B. Category structure on multiple phonological dimensions

Multidimensional relationships between acoustic cues and their mapping onto multiple phonological distinctions have been investigated directly in only a relatively small number of studies. There is evidence of interactions between acoustic cues to place and voicing in categorization (Benkí, 2001; Sawusch and Pisoni, 1974; Volaitis and Miller, 1992), and shifts in categorization functions toward longer VOTs for alveolar stops relative to labial stops have been shown to correspond closely to statistical distributions of produced cues (e.g., VOTs; Nearey and Hogan, 1986). Interactions between a number of segmental and suprasegmental features have also been found in speeded classification tasks (Eimas *et al.*, 1978; Eimas *et al.*, 1981; Miller, 1978). However, not all evidence indicates that such dimensions interact. Evidence of independence between cues has also been found in phonetic categorization (Massaro and Oden, 1980; Miller and Eimas, 1977; Oden and Massaro, 1978).

However, although the findings reported by Volaitis and Miller (1992) and Eimas *et al.* (1978) provide reasonably clear evidence of interactions between acoustic cues to phonological distinctions, a rigorous distinction between perceptual and decisional processes is not maintained in either set

of studies. To the extent that perception and decision-making are distinct cognitive processes, this makes interpretation of evidence of dimensional interactions difficult.

There are some studies in which this distinction is maintained. For example, the normal *a posteriori* probability model (Nearey and Hogan, 1986; Nearey, 1990) models both the structure of perceptual categories and the response selection process, and categorization of sequences of phones has also been modeled using logistic regression models including response bias parameters (Nearey, 1992, 1997, 2001) and as a hierarchical process explicitly developed to model dimensional interactions at multiple levels (Smits, 2001b).

The fuzzy logical model of perception (FLMP) also makes a distinction between perception and decision-making. The FLMP has been used to model integration of information from various acoustic-phonetic dimensions (e.g., VOT, formant values at voicing onset, aspiration noise intensity) in voicing and place of articulation categorization (in nonsense syllable-initial “p,” “b,” “t,” and “d”; Massaro and Oden, 1980; Oden and Massaro, 1978), producing equivocal evidence of perceptual independence and decisional interactions in at least one case (Oden and Massaro, 1978).

C. Limitations of previous work

The work described previously makes it clear that there is rich internal structure to phonological categories. However, a number of theoretical and methodological limitations limit the strength of the evidence of interactions or independence between acoustic exponents of distinctive features. In addition, whether the evidence supports independence or interaction at the acoustic-phonetic level, the complex, many-to-many mapping between acoustics and the more abstract, and lower-dimensional, phonological level make it all but impossible to draw strong inferences about the cognitive relationships between distinctive features.

D. Theoretical limitations

There are two major theoretical limitations to previous work. First, the levels at which dimensions may interact (e.g., perception and decision-making) are often conflated. Second, it is often assumed *a priori* that interactions occur or that independence holds at one or more levels.

1. Conflation of levels of interaction

It is important to maintain the conceptual distinctions between perception and decision-making and between within-stimulus and across-stimuli levels of perceptual interaction (or independence). Decision-making is logically distinct from perception, taking place postperceptually, and is at least partially under the control of a listener. On the other hand, across-stimuli perceptual relationships correspond to generalization of one dimension across levels of another (e.g., VOT across levels of burst amplitude, or voicing across levels of place of articulation), whereas within-stimulus perceptual relationships provide information about feature or cue combinations at a lower level, determining the “shape” of perceptual representations defined by particular

levels of multiple dimensions (e.g., a stimulus with long VOT and large burst amplitude, or a voiced labial).

These levels are frequently conflated. Explicit goodness judgments are the product of both perceptual and decisional factors, and the results reported by Volaitis and Miller (1992) could be due to either, or both in conjunction. Maddox (1992) has shown that interference in speeded classification can be produced by either perceptual or decisional processes (as in, e.g., Eimas *et al.*, 1981). Within- and across-category forms of perceptual interactions are conflated in, e.g., logistic regression models of categorization (see, e.g., Nearey, 1992) and the FLMP (see, e.g., Oden and Massaro, 1978).

2. Assumptions of independence

A general description of the relationships between cognitive dimensions should rely on as few *a priori* assumptions about independence as possible. If it is assumed that one form of independence (e.g., within-stimulus independence) holds, this has only partially understood implications for investigations into other forms of independence (e.g., between-stimuli perceptual and decisional independence).

Previous applications of multidimensional detection theory to speech perception have assumed within-stimulus independence (see, e.g., Kingston and MacMillan, 1995; Kingston *et al.*, 1997; Kingston *et al.*, 2008; MacMillan *et al.*, 1999), and independent combination of information from multiple acoustic dimensions is central to the FLMP (Massaro and Oden, 1980; Oden and Massaro, 1978). On the other hand, the decisional notion of independence is assumed to hold in many applications of GRT to visual perception (see, e.g., Olzak and Wickens, 1997; Thomas, 2001).

E. Methodological limitations

Evidence of interaction between phonological dimensions may also be limited simply because it arises as an unintended consequence of work on other issues. For example, although the results of both Nearey and Hogan (1986) and Volaitis and Miller (1992) indicate that voicing perception and/or decision-making vary with place of articulation, neither study looked at place perception and decision-making as a function of voicing.

As discussed previously, much of the evidence of dimensional interactions in speech perception can be found in studies of the relationships between acoustic-phonetic dimensions in categorization (see, e.g., Nearey, 1992; Oden and Massaro, 1978; Smits, 2001a). Most such studies employ stimuli built on predetermined acoustic-phonetic dimensions (e.g., VOT, formant frequency value at voice onset, etc.). Although this method has clear and proven value, if the goal is to study interactions or independence between *phonological* dimensions, as it is here, strong assumptions about the relevant set of acoustic-phonetic cues should be avoided as much as possible. Many such assumptions can be avoided by using naturally produced, and so naturally variable, stimuli.

II. GENERAL RECOGNITION THEORY

The theoretical and methodological limitations discussed previously may be addressed in the GRT framework.

A brief introduction to the structure of (Gaussian) GRT will provide definitions for three logically distinct notions of independence. This will be followed by a description of the experimental protocols typically employed with GRT along with a discussion of some methods for testing the various forms of independence. Finally, this paper will describe some recent developments that extend these previously established tests substantially.

A. The structure of general recognition theory

GRT is a two-stage model of perception and decision-making. In GRT, two major assumptions are made. First, it is assumed that the presentation of a stimulus produces a random perceptual effect due to internal noise, external noise added to the stimulus, or both. Over the course of many trials, this results in distributions of perceptual effects. Second, it is assumed that perceptual space is exhaustively partitioned into mutually exclusive response regions. So-called decision bounds determine the responses associated with (sets of) perceptual effects.

GRT provides rigorous, general definitions of three logically distinct notions of independence (and interaction) between dimensions. Two of these concern perception, whereas the other concerns decision-making. One of the perceptual notions concerns within-stimulus independence, whereas the other concerns across-stimuli independence. Two additional assumptions will be made here. First, it is assumed that the perceptual distributions are bivariate Gaussian, and second, it is assumed that the decision bounds are linear.

Figure 1 shows an illustrative two-dimensional Gaussian GRT model for the labial obstruents [p], [b], [f], and [v], i.e., consonants consisting of the factorial combination of (stop and fricative) manner of articulation and (voiced and unvoiced) voicing. It is convenient in visualization of

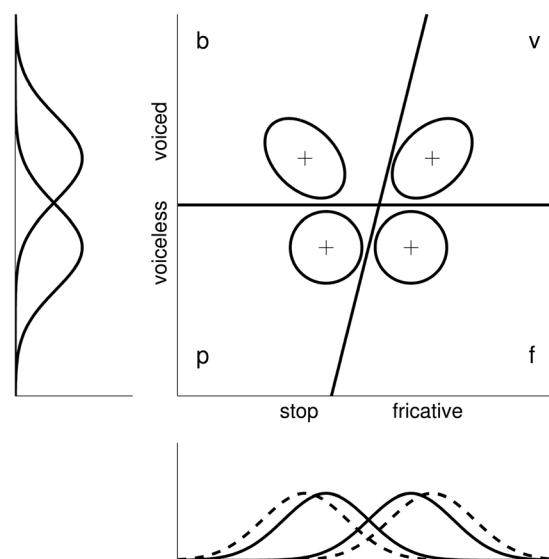


FIG. 1. Illustrative two-dimensional Gaussian GRT model. The ellipses represent contours of equal likelihood; the plus signs indicate the means of the distributions. The solid line marginal densities correspond to the first level on the other dimension (e.g., voiceless or stop); the dashed line marginals correspond to the second level (e.g., voiced or fricative). See the text for details.

Gaussian GRT models to make use of equal likelihood contours, or sets of points the same height above the (x, y) plane; here, (x, y) correspond to (*manner, voicing*). Given the presentation of a particular stimulus, the points inside the corresponding ellipse are more likely to occur than are the points outside the ellipse.

1. Three forms of independence

The within-stimulus notion of independence in the GRT framework is called *perceptual independence*. Perceptual independence holds for a given perceptual distribution if, and only if, stochastic independence holds. In Gaussian GRT, perceptual independence holds within a given distribution if, and only if, correlation between the perceptual effects on each dimension is zero.

The two bivariate distributions on the bottom of the main panel of Fig. 1 illustrate perceptual independence. In this (fictitious) example the correlation between manner and voicing is zero in the perceptual distributions for [p] and [f]. On the other hand, perceptual independence fails for the other two distributions. In the [b] distribution, there is negative correlation between manner and voicing (i.e., the more “fricative-like” a perceptual effect is, the more likely it is to also be “voiceless”), and there is a positive correlation between the perceptual effects in the [v] distribution. The sign of the correlation is defined with respect to the dimensions and levels, and the effect of perceptual correlations is to change the relative proportions of distributions that fall in the different response regions.

The across-stimuli notion of independence in GRT is called *perceptual separability*. We say that, for example, voicing is perceptually separable from manner if the marginal perceptual effect of voicing is identical across levels of manner; this is illustrated in the main and left panels Fig. 1. On the voicing dimension, the perceptual distribution for [b] is identical to the perceptual distribution for [v]; the same relationship holds between [p] and [f]. On the other hand, in this example, manner is *not* perceptually separable from voicing. Perceptual separability has failed due to shifts in the marginal means for the voiced stops such that [b] tends to be perceived as “more stop-like” (i.e., is closer to the stop end of the dimension) than [p], and [v] tends to be perceived as “more fricative-like” (i.e., is closer to the fricative end of the dimension) than [f].

Finally, the decision-related notion of independence in GRT is called *decisional separability*. Decisional separability holds if, and only if, a decision bound is parallel to the (appropriate) coordinate axis. In Fig. 1, decisional separability holds for the bound separating the voiced and voiceless response regions; decisions about the voicing of a stimulus do not depend in any way on a stimulus’ manner. Decisional separability fails here on the manner dimension; there is a bias toward responding “stop” in the voiced region of perceptual space and a bias toward responding “fricative” in the voiceless region.

This Gaussian GRT model predicts identification-confusion probabilities as the double integrals of the bivariate perceptual distributions in the appropriate response regions.

For example, the predicted probability of responding [f] when presented with [p] would be the double integral of the [p] distribution in the [f] response region.

It is important to keep in mind that, in the GRT framework, perceptual independence, perceptual separability, and decisional separability are defined with respect to an unobservable perceptual space. This is important, in part, because the Gaussian GRT model can be applied to experiments probing very different levels of speech perception. For example, it can be applied to probe perceptual relationships between acoustic cues within phonological categories (as in, e.g., [Kingston et al., 2008](#)), or it can be applied to study relationships between distinctive features more directly (as it is here). Hence, failures of perceptual *separability* in the former are approximately comparable to failures of perceptual *independence* in the latter. The comparison is only approximate because interactions in a higher-dimensional acoustic-phonetic space (e.g., the four-dimensional space defined by f_0 , F1, closure duration, and voicing continuation, as in [Kingston et al., 2008](#)) are likely to have incompletely understood effects on the structure of lower-dimensional feature space (e.g., two-dimensional voicing by manner space).

2. Testing independence

Independence and separability can be tested by fitting and comparing models of perception and decision-making (see, e.g., [Silbert et al., 2009](#); [Thomas, 2001](#); [Wickens, 1992](#)). However, this approach suffers from some important limitations, due, at least in part, to the fact that the most general Gaussian GRT model (with linear bounds) may have more free parameters than the data have degrees of freedom. In such cases, multiple configurations of perceptual distributions and decision bounds may account for the data equally well. Failures of one kind of independence may also produce empirical patterns that are essentially identical to those produced by the failure of a different kind of independence (e.g., certain kinds of failure of perceptual separability and decisional separability are impossible to disambiguate in identification data; see, e.g., [Silbert 2010](#), Chap. 3, for a discussion of this kind of “model mimicry” issue in GRT). A common solution to this has been to assume that decisional separability holds while testing perceptual independence and perceptual separability ([Olzak, 1986](#); [Thomas, 2001](#); [Wickens, 1992](#)). This assumption is made in the present work, as well. Justification for this assumption may be found in the fact that decision-making is at least partially under the control of the listener, whereas perception is not.

A major limitation of almost all previous work in the GRT framework is the fact that each individual subject’s data are analyzed separately. Whereas detailed quantitative information about each subject’s perceptual and decisional space is readily derived from his or her data, until recently, group-level statistical properties have been neglected. The hierarchical Gaussian GRT model described in the following addresses this limitation directly by estimating perceptual distribution and decision-bound parameters at the individual subject level, simultaneously estimating

parameters for parent distributions governing the individual subject parameters.

B. A hierarchical Gaussian GRT model

Each individual subject's data are modeled with a two-dimensional space containing four perceptual distributions and partitioned into four response regions by two decision criteria. A number of the model's parameters must be fixed *a priori* so that unique estimates of the other parameters may be derived. Thus, the mean of one perceptual distribution is fixed at (0, 0), and all marginal variances are fixed at unity.

Each subject produces K data vectors \mathbf{d}_{ik} consisting of the counts of the i th subject's R responses to the k th stimulus. These counts are modeled as a multinomial random variable parameterized by the vector of probabilities θ_{ik} and N_k , the number of presentations of stimulus k , denoted as follows:

$$\mathbf{d}_{ik} \sim \text{Multinomial}(\theta_{ik}, N_k).$$

The probability of the i th subject giving the r th response to the k th stimulus (θ_{irk}) is, as described previously, the double integral of the k th perceptual distribution over the r th response region, denoted Φ_2^r . For a given stimulus, the four response probabilities are determined by the perceptual distribution mean μ , the correlation ρ , and the decision criteria κ ,

$$\theta_{irk} = \Phi_2^r(\kappa, \mu, \rho).$$

Across subjects, the k th perceptual mean μ and j th decision criterion κ are modeled as normal random variables with means η_k and ψ_j and precisions (i.e., the reciprocal of variance) τ and χ , respectively, and each correlation ρ is modeled as a truncated (at ± 0.975) normal random variable with mean ν_k and precision π ,

$$\mu_{ik} \sim \text{Gaussian}(\eta_k, \tau),$$

$$\kappa_{ij} \sim \text{Gaussian}(\psi_j, \chi),$$

$$\rho_{ik} \sim \text{Gaussian}(\nu_k, \pi), \quad \rho_{ik} \in [-0.975, 0.975].$$

Finally, the group-level mean and precision parameters are modeled as normal and gamma random variables, respectively. The three¹ group-level stimulus means η_k are distributed with means of (0, 2), (2, 0), and (2, 2) and variances of 2. The shape and rate parameters governing all of the group-level precision parameters were set to 5 and 1, respectively, emphasizing standard deviations near and below 1, although allowing any positive value.

It is important to note that, because the model has as many free parameters as the data has degrees of freedom at the individual subject level, the model is expected to (and does) fit the data very well. The parameter estimates for a given subject are constrained by that subject's data, on the one hand, and by the group-level and other subjects' parameters, on the other. The purpose of fitting this model to the data is to simultaneously estimate values of and uncertainty

about individual and group-level perceptual and decisional parameters.

III. INTERIM SUMMARY

As noted previously, this work has two goals. The primary goal is to demonstrate, through the application of GRT to speech perception, the importance of maintaining rigorous distinctions between processing levels when studying relationships between cognitive dimensions. The secondary goal is to establish a set of baseline results concerning the relationships between distinctive manner and voicing features in English obstruents.

IV. EXPERIMENT 1: MANNER AND VOICING IN ONSET POSITION

Experiment 1 is an investigation of perceptual independence and separability between manners of articulation and voicing in syllable-initial labial consonants [p], [b], [f], and [v]. The two experiments presented here are companion projects to similar work on place and voicing. Although there has been some previous work on voicing and place (see, e.g., Eimas *et al.*, 1981; Oden and Massaro, 1978; Sawusch and Pisoni, 1974), and on place and manner (see, e.g., Eimas *et al.*, 1978), it seems that possible interactions between voicing and manner have not been similarly studied, at least not in labial consonants. These consonants were chosen here in part because they overlap partially with and have a factorial feature structure similar to the consonants used in the companion experiments on [p], [b], [t], and [d].

A. Stimuli

In order to avoid strong assumptions about the relevant acoustic-phonetic dimensions, naturally produced nonsense syllables were used as stimuli. In order to ensure that the subjects did not simply attend to some irrelevant acoustic feature of a particular token of a particular category, a small degree of within-category variability was introduced by using four tokens of each stimulus type—[pa], [ba], [fa], and [va]—all produced by the author (a mid-30 s midwestern, male phonetician). Multiple acoustic measurements (e.g., VOT, F0 at vowel onset, F1 and F2 at vowel onset and midpoint, spectral moments of release burst²) were analyzed and extensive pilot experimentation was carried out to ensure both that no particular token was overly acoustically distinct and that the stimuli were within the normal range of values for these consonants. The stimuli for both experiments were recorded during a single session in a quiet room via an Electrovoice RE50 microphone and a Marantz PMD560 solid-state digital recorder at 44.1 kHz sampling rate with 16-bit depth

Naturally produced [i.e., not (re)synthesized] tokens can be very acoustically distinct, however, and identification data with very high accuracy is not particularly informative with respect to perceptual interactions. Thus, stimuli were embedded in "speech-shaped" noise (i.e., white noise filtered such that higher frequencies had relatively lower amplitude than lower frequencies).

B. Procedure

Each participant was seated in a double-walled sound attenuating booth with four “cubicle” partitions. One, two, or three participants could run simultaneously, each in front of his or her own computer terminal. Stimuli were presented at -3 dB signal-to-noise ratio at ~ 60 dB sound pressure level via Tucker-Davis-Technologies Real-Time processor (TDT RP2.1; sampling rate 24 414 Hz), programmable attenuator (TDT PA5), headphone buffer (TDT HB6), and Sennheiser HD250 II Linear headphones. Before the first session (familiarization and training), participants read a written instruction sheet, were given verbal instructions, and were prompted for questions about the procedure. Sessions lasted from one to four experimental blocks. Experimental blocks lasted ~ 25 min.

Each experimental block began with brief written instructions reminding participants to respond as accurately and as quickly as possible. Explicit guessing advice was provided for trials on which the participant was uncertain of the stimulus identity.³ After the instructions were cleared from the screen, four “buttons” corresponding to the buttons on a hand-held button box became visible. On the on-screen buttons the letters p, b, f, and v appeared in black text. Button-response assignments were randomly assigned for each block with the constraint that the basic dimensional structure was always maintained (e.g., p and f always appeared as neighbors on a single dimension, never on opposite corners).

Each trial consisted of the following steps: (1) A visual signal (the word “listen”) presented on the computer monitor; (2) 0.5 s of silence; (3) stimulus presentation; (4) response; (5) feedback; and (6) 1 s of silence. Responses were collected via a button box with buttons arranged to correspond to the structure of the stimulus space (i.e., two levels on each of two dimensions). Feedback was given visually via color-coded (green for correct, red for incorrect) text above and on the on-screen buttons. Either the word “Correct” or the word “Incorrect” appeared along with brief descriptions of the presented stimulus and the response chosen. The feedback text disappeared and the button text color was reset to black before each successive trial.

Each participant received two short (~ 15 min) and two regular length blocks to familiarize them with the stimuli and ensure that performance was consistently above chance. The data analyzed here consist of 800 trials completed in two blocks of 400 trials each. Participants were paid \$6/h with a \$4/h bonus for completion of the experiment. The participants with the highest accuracy and fastest mean response times each received \$20 bonuses.

1. Subjects

Eight adults (three male, five female) were recruited from the university community. The average age of participants was 22 (18–27). All were native speakers of English with, on average, 5.75 (2–7) years of second language study. All but two were right handed, and all but two were from the Midwest (the other two were from the East). All participants were screened to ensure normal hearing.

2. Analysis

The hierarchical Gaussian GRT model described previously was fit to the eight subjects’ data. Response counts were tallied by stimulus category, not by individual stimuli. The first 5000 samples from the posterior distribution were discarded, and each of three chains of sampled values was thinned such that only every 400th sample was retained in order to ensure independence between the samples. For the final analysis, 300 samples were retained.

Autocorrelation functions were examined to ensure independence within chains, and the value of the \hat{R} statistic was checked for each parameter to ensure that the chains were well mixed; $\hat{R} \leq 1.1$ can be taken to indicate that the chains are well-mixed, a necessary condition for convergence to the true posterior distribution of the parameters (Gelman *et al.*, 2004, pp. 296–297).

C. Results

Table I shows each subject’s overall accuracy for experiment 1. Overall accuracy was high, ranging from 70% to 88% correct.

Figure 2 shows the fitted model. Within each panel, the x -axis indicates manner of articulation, with stops to the left and fricatives to the right, and the y -axis indicates voicing, with voiceless at the bottom and voiced at the top. Hence, the perceptual distribution(s) and response region(s) for [p] are located at the bottom left, for [b] at the top left, for [f] at the bottom right, and for [v] at top right.

The smaller panels show, for individual subjects 1–8 (moving counterclockwise from top left), the decision bounds and perceptual distribution equal likelihood contours for the median posterior decision criterion, mean, and correlation parameter estimates. Uncertainty of parameter estimates is shown by indicating the 95% highest (probability) density interval [(HDI); i.e., the range of parameter estimates delimiting the middle 95% of the posterior distribution], which are akin to confidence intervals. For each subject, the decision bounds are plotted at the median decision criterion value, and the perpendicular lines near the bottom and right of each panel indicate the 95% HDIs for these estimates. The perceptual distribution mean for [p] was fixed at (0, 0), and for the other perceptual distributions, the median estimate of the distribution mean is indicated by the intersection of the horizontal and vertical lines within the equal likelihood contour; the length of these lines indicate the 95% HDIs for the distribution mean estimates. The median correlation estimate for each perceptual distribution is indicated by the shape of the equal likelihood contour, and the 95% HDI is indicated by the vertical line plotted in the corner nearest the contour; the small circle indicates zero correlation (i.e., perceptual independence).

TABLE I. Accuracy results, experiment 1.

Subject	1	2	3	4	5	6	7	8
$p(C)^a$	0.87	0.78	0.88	0.85	0.80	0.70	0.80	0.81

^aProportion correct.

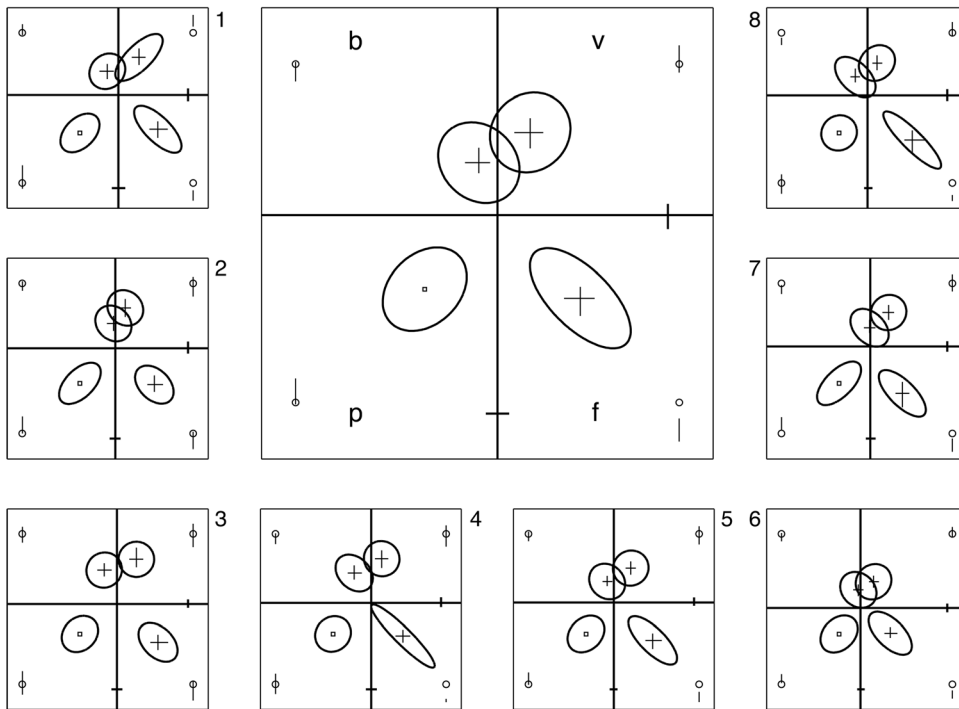


FIG. 2. Fitted hierarchical Bayesian Gaussian GRT model, onset manner, and voicing. See the text for details.

The large panel in the middle shows the fitted group-level model. The decision bounds, perceptual distribution means, and equal likelihood contours indicate the median group level ψ , η , and ν parameters, respectively. The 95% HDIs for these parameters are indicated as in the smaller panels.

For every individual subject and at the group level, which is to say in every panel of Fig. 2, it is clear that perceptual separability fails between voicing and manner on both dimensions. Voicing is not perceptually separable from manner because the salience between voiceless [p] and [f] is much higher than the salience between voiced [b] and [v]. The effect causing manner to not be perceptually separable from voicing is smaller, although still consistent across subjects; the perceptual effects of [v] tend to be “more voiced” than the perceptual effects of [b] (i.e., the [v] distribution is higher up on the voicing dimension than the [b] distribution), and, for some subjects, the perceptual effects for [f] are also slightly more voiceless than those for [p].

For subjects 1 and 4–8, and at the group level, perceptual independence fails for [f] by virtue of a negative correlation with zero outside the 95% HDI; the median correlation is negative in the [f] distribution for subjects 2 and 3, as well, but zero is within their respective HDIs. There is a weaker tendency toward positive correlation in the [p] distribution. For subjects 2 and 5–7, zero is outside the HDI, but for the other subjects, and at the group level, perceptual independence cannot be ruled out. There is an even weaker tendency toward negative correlations in the [b] distribution and positive correlations in the [v] distribution, although zero is outside the HDIs for two subjects (for [b]) and one subject (for [v]).

Although the assumption of decisional separability obviates consideration of decisional interactions, the response criterion parameters do allow inferences to be drawn with respect to response bias. Whereas the location of the manner criterion indicates no consistent bias toward either stop or

fricative labels, to varying degrees across subjects, there does seem to be a small bias toward voiceless (rather than voiced) responses (i.e., the horizontal bound is shifted such that the voiceless response regions tend to be bigger than the voiced).

D. Experiment 1 discussion

The fact that voicing is not perceptually separable from manner has a number of possible causes. Silbert and de Jong (2008) report that voiced labial fricatives tend to be shorter, have less noise energy, and have more voicing energy than do their voiceless counterparts. Jongman *et al.* (2000) report similar duration results, although their voiced labial fricatives had higher noise amplitude than did their voiceless labial fricatives (they do not report voicing energy results). Differences in VOT and release burst amplitude in voiced and voiceless stops are analogous to these differences in fricatives (Kessinger and Blumstein, 1997; Silbert, 2010; Volaitis and Miller, 1992). It is plausible, although we cannot confirm it here, that the observed failure of perceptual separability on the manner dimension is due to a larger acoustic (and/or auditory) difference between long, relatively loud voiceless frication in the [f] stimuli and the long VOT, relatively quiet aspiration noise in the [p] stimuli than between the short, relatively low energy noise in the [b] and [v] stimuli. Possible causes of the apparent failure of separability on the voicing dimension are less clear, although differences in high and low frequency energy (i.e., frication noise and voicing energy) may play a role. Higher-level information (e.g., lexical neighborhood structure) may also influence perceptual relationships between features (although, see Norris *et al.*, 2000).

These failures of perceptual separability can be interpreted as imperfect generalization of features across one another, and the failure of perceptual independence for [f] (and for a subset of subjects and to a lesser degree [p] and

[b]) may be interpreted as the presence of cross-talk between the cognitive channels processing voicing and manner (see, e.g., Ashby, 1989). The importance of maintaining the conceptual distinction between within-category and between-category notions of independence is highlighted by the fact that perceptual separability fails across the board, whereas there is much more variability across subjects with respect to failure of perceptual independence.

Finally, it is worth reiterating that previous work applying GRT to speech perception has relied on the strong, and possibly incorrect, *a priori* assumption that perceptual independence holds (see, e.g., Kingston and MacMillan, 1995; Kingston *et al.*, 1997; Kingston *et al.*, 2008; MacMillan *et al.*, 1999), and that essentially all previous work in the GRT framework has conducted tests of perceptual separability and independence only at the individual subjects level (see, e.g., Silbert *et al.*, 2009; Thomas, 2001). In general, GRT allows the assumption of perceptual independence to be relaxed and tested, and the hierarchical (Bayesian Gaussian) GRT model enables simultaneous statistical treatment of the individual-subject and group levels.

V. EXPERIMENT 2: MANNER AND VOICING IN CODA POSITION

Experiment 2 is an investigation of perceptual independence and perceptual separability between manner of articulation and voicing in syllable-final labial obstruents [p], [b], [f], and [v]. These are the same consonants as those investigated in Experiment 1, at least at an abstract level, but, not surprisingly, the acoustic cues to the phonological voicing distinctions differ between onset and coda positions. In the stops, for example, VOT is not a cue to voicing in coda position, whereas the ratio of vowel and consonant duration is (Port and Dalby, 1982). In fricatives, on the other hand, although there are similar voicing-related differences in duration in onset and coda position, the difference between voiced and voiceless fricatives is reduced with respect to noise and voicing energy (Silbert and de Jong, 2008). The acoustic cues to manner are not expected to differ substantially between onset and coda position, although this does not appear to be well-documented.

A. Stimuli

Four tokens of each stimulus type—[ap], [ab], [af], and [av]—were produced by the author. The stimuli for both experiments (as well as two others) were recorded on the same equipment during the same session.

B. Procedure

The procedure was identical to that employed in experiment 1, with appropriate changes made to instructions and button labels.

1. Subjects

Eight adults (three male, five female) were recruited from the university community. The average age of participants was 21.8 (18–27). All were native speakers of English

with, on average, 4.75 (2–9) years of second language study. All but two were right handed, and all but one were from the midwest (the other was from the east coast). All participants were screened to ensure normal hearing. Six of the subjects that participated in this experiment also participated in the previous experiment (i.e., manner and voicing in onset position).

2. Analysis

Analyses were carried out in the same manner as those in experiment 1.

C. Results

Table II provides a summary of the results from experiment 2. As in experiment 1, overall accuracy was high in experiment 2, ranging from 70% to 88% correct.

The fitted model for experiment 2 is displayed in Fig. 3. As discussed previously, the eight individual subject-level perceptual and decisional spaces are depicted in the small panels, moving counterclockwise from the top left, and the group-level model is shown in the large panel in the middle.

As in the onset position, manner is not perceptually separable from voicing due to a large difference in salience between voiceless [p] and [f] and between voiced [b] and [v]. Note, however, that this difference is more extreme in the coda position. For a number of subjects, [b] and [v] seem to be completely indistinguishable. Although there is some variability across subjects, voicing seems to be at least approximately perceptually separable from manner.

Also as in the onset position, there is a tendency toward negative correlations in the [f] distribution and positive correlation in the [p] distribution. However, perceptual independence only fails for [f] for subjects 1 and 3, and it only fails for [p] for subject 2. Perceptual independence holds at the group level for all four consonants. Finally, there is no clear response bias on either dimension.

D. Experiment 2 discussion

The similarity between the onset and coda manner by voicing spaces are clear. However, despite these similarities, there (at least) four interesting differences, as well. First, although the salience between [b] and [v] in the onset position was quite low, these two sounds are essentially indistinguishable in coda position; their perceptual distributions overlap substantially, for some individual subjects almost completely. Second, the salience of the voicing dimension relative to manner is higher in the coda position than it is in the onset position. Third, voicing seems to be (very close to) perceptually separable from manner in the coda position, whereas the small but consistent shift in the [v] distribution in the onset position caused perceptual separability of

TABLE II. Accuracy results, experiment 2.

Subject	1	2	3	4	5	6	7	8
$p(C)^a$	0.87	0.88	0.78	0.85	0.80	0.70	0.80	0.81

^aProportion correct.

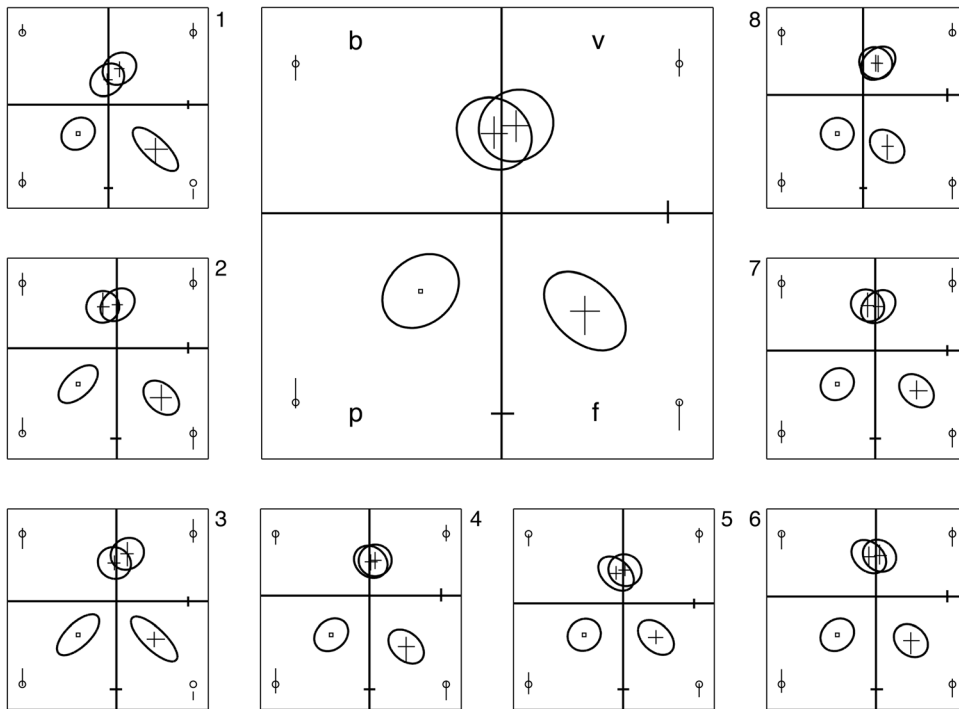


FIG. 3. Fitted hierarchical Bayesian Gaussian GRT model, coda manner, and voicing. See the text for details.

voicing from manner to fail. Fourth, although there are similar tendencies with respect to correlations within distributions in both syllable positions, perceptual independence cannot be ruled out in any consonant in coda position.

Because many of the acoustic cues to manner and, within fricatives, to voicing are the same in onset and coda position, it should not be surprising that similar patterns of failure of perceptual separability between voicing and manner are observed in both syllable positions. To the extent that the acoustics differ, it is reasonable to hypothesize that these caused the observed differences. This is beyond the scope of the current project, although, and so must be left as an open question for now.

VI. CONCLUSION AND GENERAL DISCUSSION

A. Recapitulation and summary

Although much work has been done on the relationships between acoustic cues to phonological distinctions (e.g., voicing and place; Oden and Massaro, 1978; Sawusch and Pisoni, 1974), little, if any, work has focused directly on the relationships between phonological dimensions. The present work represents the first documentation of perceptual interactions between manner of articulation and voicing in English obstruents.

The present work also demonstrates the utility of GRT in overcoming a number of methodological and theoretical limitations to work on the relationships between acoustic-phonetic dimensions. In much previous work, strong assumptions about the relevant acoustic cues to distinctive features and conflation of conceptually distinct processing levels make it difficult to draw robust inferences.

A new hierarchical Bayesian Gaussian GRT model was fit to data from two identification tasks shows that. Results show that, in the onset position, perceptual separability fails

between voicing and manner. The salience between [p] and [f] is much greater than the salience between [b] and [v], and the salience between [v] and [f] is greater than the salience between [p] and [b]. Perceptual independence tends to fail for [f], but not for the other three consonants. The results for coda position are similar, but not identical. The salience of the manner distinction varies as a function of voicing in much the same way, but perceptual independence holds more consistently, voicing is closer to being perceptually separable from manner, and the salience of voicing relative to manner is greater in the coda position than in the onset position.

These results underscore the importance of maintaining the distinction between the within-stimulus and between-stimuli perceptual levels. They also indicate that one may draw flawed inferences if one simply assumes that voicing and manner are independent.

B. Limitations to the present work

It may be argued that these stops and fricatives are not distinguished by a single phonological feature. That is, the manner distinction between [p]/[b] and [f]/[v] is also a place distinction; the fricatives are labiodental and the stops bilabial. There are at least two reasons why this is irrelevant. First, English does not have bilabial fricatives or labiodental stops, so in the context of the English obstruents, this distinction is as minimal as it can be. Second, even if the number of features is the primary determinant of the salience of a distinction (as argued, e.g., by Bailey and Hahn, 2005), this cannot explain the results reported here. The same set of features distinguishes [b] and [v], on the one hand, and [p] and [f], on the other, yet the salience varies substantially between these two pairs.

It is important to keep the scope of the present findings in proper perspective. Although they provide a rigorous baseline

to which future work can be compared, the present results are of only limited generality. Only a few stimuli were used in each experiment, and these were all produced by a single talker. Although measures were taken to ensure that the stimuli were suitable, to the extent that they deviate from typical productions of the same categories in the larger speech community, the perceptual results reported here may be idiosyncratic. The speech-shaped noise used to mask the stimuli may also be responsible for some portion of the observed results. Both of these concerns are, ultimately, empirical matters. Forthcoming results from a parallel set of experiments using multitalker babble to mask the same stimuli should provide valuable information with respect to masker-specific effects, and further experimentation using a larger number of more variable stimuli is currently in development.

C. Implications for related work and future directions

A number of directions for future research are apparent. First, there are many possible factors driving these patterns of interaction. As noted previously, the mapping between acoustics and phonological structure is many-to-many and complex. A thorough statistical model of multidimensional phonetic and phonological spaces in speech production is a prerequisite to a full understanding of the extent to which the acoustic structure of phonological categories drive perceptual relationships between dimensions. The present work is currently being extended in an effort to model the phonological and phonetic production space of a large subset of English obstruent consonants. This work should, in the long run, and in conjunction with studies like [Kingston et al. \(2008\)](#), provide a more complete picture of structure of phonological categories in production and perception.

Higher-level linguistic factors may also play a role in determining the nature of perceptual and decisional space. Nonwords were used here in order to minimize the possible influence of lexical factors. Future work in the GRT framework will manipulate higher-level (e.g., lexical) factors in order to probe the architecture of the cognitive systems underlying speech perception (see, e.g., [Norris et al., 2000](#)).

Finally, models such as GRT also hold promise in the study of non-native speech perception. GRT may be linked to a quantitative model of production categories (roughly as described in [Smits, 2001b](#)), providing a detailed quantitative implementation of the perceptual assimilation model (PAM; see, e.g., [Best et al., 2001](#)). Previous work has shown how multidimensional perceptual distributions and response bounds can be used to provide a measure of perceptual similarity (see, e.g., [Ashby and Perrin, 1988](#)), a concept at the heart of both the PAM. With careful quantitative modeling of production and perception in each of two languages, for example, it would be possible to model the relationships between native and non-native categories in a more fine-grained manner than has thus far been accomplished. The present work represents a (modest) step in toward this goal.

ACKNOWLEDGMENTS

This work benefited enormously from the knowledge, advice, and patience of James Townsend, Kenneth de Jong,

and Jennifer Lentz. The work was funded by NIH Grant No. 2-R01-MH0577-17-07A1.

¹The mean of the perceptual distribution fixed at (0, 0) is not modeled as a random variable at the group level.

²See supplementary material at <http://dx.doi.org/10.1121/1.3699209> for spectrograms and scatter plots of acoustic measures for the stimuli used here. These are available by request from the author.

³Each subject completed ten blocks, two in each of five stimulus presentation base rate conditions. Only the condition in which each stimulus was presented equally frequently will be analyzed and discussed here.

- Ashby, F. G. (1989). "Stochastic general recognition theory," in *Human Information Processing: Measures, Mechanisms, and Models*, edited by D. Vickers and P. L. Smith (Elsevier, Amsterdam), pp. 435–457.
- Ashby, F. G., and Perrin, N. A. (1988). "Toward a unified theory of similarity and recognition," *Psychol. Rev.* **95**, 124–150.
- Ashby, F. G., and Townsend, J. T. (1986). "Varieties of perceptual independence," *Psycho. Rev.* **91**, 154–179.
- Bailey, T. M., and Hahn, U. (2005). "Phoneme similarity and confusability," *J. Mem. Lang.* **52**, 339–362.
- Benkí, J. R. (2001). "Place of articulation and first formant transition pattern both affect perception of voicing in English," *J. Phonet.* **29**, 1–22.
- Best, C. T., McRoberts, G. W., and Goodell, E. (2001). "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system," *J. Acoust. Soc. Am.* **109**, 775–794.
- Blumstein, S. E., and Stevens, K. N. (1979). "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants," *J. Acoust. Soc. Am.* **66**, 1001–1017.
- Eimas, P. D., Tartter, V. C., and Miller, J. L. (1981). "Dependency relations during the processing of speech," in *Perspectives on the Study of Speech*, edited by P. D. Eimas and J. L. Miller (Lawrence Erlbaum Associates, Hillsdale, NJ), Chap. 8, pp. 283–309.
- Eimas, P. D., Tartter, V. C., Miller, J. L., and Keuthen, N. J. (1978). "Asymmetric dependencies in processing phonetic features," *Percept. Psychophys.* **23**, 12–20.
- Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (1988). "Statistical analysis of word-initial voiceless obstruents: Preliminary data," *J. Acoust. Soc. Am.* **84**, 115–123.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis*, 2nd ed., Chapman & Hall/CRC Texts in Statistical Science Series (Chapman & Hall/CRC, London), pp. 296–297.
- Jongman, A., Wayland, R., and Wong, S. (2000). "Acoustic characteristics of english fricatives," *J. Acoust. Soc. Am.* **108**, 1252–1263.
- Kessinger, R. H., and Blumstein, S. E. (1997). "Effects of speaking rate on voice-onset time in Thai, French, and English," *J. Phonet.* **25**, 143–168.
- Kingston, J., and Diehl, R. L. (1994). "Phonetic knowledge," *Language* **70**, 419–454.
- Kingston, J., Diehl, R. L., Kirk, C. J., and Castleman, W. A. (2008). "On the internal perceptual structure of distinctive features: The [voice] contrast," *J. Phonet.* **36**, 28–54.
- Kingston, J., and MacMillan, N. A. (1995). "Integrality of nasalization and f1 in vowels in isolation and before oral and nasal consonants: A detection-theoretic application of the Garner paradigm," *J. Acoust. Soc. Am.* **97**, 1261–1285.
- Kingston, J., MacMillan, N. A., Dickey, L. W., Thorburn, R., and Bartels, C. (1997). "Integrality in the perception of tongue root position and voice quality in vowels," *J. Acoust. Soc. Am.* **101**, 1696–1709.
- Kurowski, K., and Blumstein, S. E. (1984). "Perceptual integration of the murmur and formant transitions for place of articulation in nasal consonants," *J. Acoust. Soc. Am.* **76**, 383–390.
- Lisker, L., and Abramson, A. (1964). "A cross-language study of voicing in initial stops: Acoustical measurements," *Word* **20**, 384–422.
- MacMillan, N. A., Kingston, J., Thorburn, R., Dickey, L. W., and Bartels, C. (1999). "Integrality of nasalization and F1. II. Basic sensitivity and phonetic labeling measure distinct sensory and decision-related interactions," *J. Acoust. Soc. Am.* **106**, 2913–2932.
- Maddox, W. T. (1992). "Perceptual and decisional separability," in *Multidimensional Models of Perception and Cognition*, edited by F. G. Ashby (Lawrence Erlbaum Associates, Hillsdale, NJ), pp. 147–180.

- Massaro, D. W., and Oden, G. C. (1980). "Evaluation and integration of acoustic features in speech perception," *J. Acoust. Soc. Am.* **67**, 996–1013.
- Miller, J. L. (1978). "Interactions in processing segmental and suprasegmental features of speech," *Percept. Psychophys* **24**, 175–180.
- Miller, J. L., and Eimas, P. D. (1977). "Studies on the perception of place and manner of articulation: A comparison of the labial-alveolar and nasal-stop distinctions," *J. Acoust. Soc. Am.* **61**, 835–845.
- Miller, J. L., and Volaitis, L. E. (1989). "Effect of speaking rate on the perceptual structure of a phonetic category," *Percept. Psychophys.* **46**, 505–512.
- Nearey, T. M. (1990). "The segment as a unit of speech perception," *J. Phonet.* **18**, 347–373.
- Nearey, T. M. (1992). "Context effects in a double-weak theory of speech perception," *Lang. Speech* **35**, 153–171.
- Nearey, T. M. (1997). "Speech perception as pattern recognition," *J. Acoust. Soc. Am.* **101**, 3241–3254.
- Nearey, T. M. (2001). "Phoneme-like units and speech perception," *Lang. Cogn. Process.* **16**, 673–681.
- Nearey, T. M., and Hogan, J. T. (1986). "Phonological contrast in experimental phonetics: Relating distributions of production data to perceptual categorization curves," in *Experimental Phonology*, edited by J. J. Ohala and J. J. Jaeger (Academic, Orlando, FL), Chap. 8, pp. 141–161.
- Norris, D., McQueen, J. M., and Cutler, A. (2000). "Merging information in speech recognition: Feedback is never necessary," *Behav. Brain Sci.* **23**, 299–325.
- Oden, G. C., and Massaro, D. W. (1978). "Integration of featural information in speech perception," *Psycholog. Rev.* **85**, 172–191.
- Oglesbee, E. N. (2008). "Multidimensional stop categorization in English, Spanish, Korean, Japanese, and Canadian French," Ph.D. thesis, Indiana University, Bloomington, IN.
- Olzak, L. A. (1986). "Widely separated spatial frequencies: Mechanism interactions," *Vision Res.* **26**, 1143–1153.
- Olzak, L. A., and Wickens, T. D. (1997). "Discrimination of complex patterns: Orientation information is integration across spatial scale; spatial-frequency and contrast information are not," *Perception* **26**, 1101–1120.
- Port, R. F., and Dalby, J. (1982). "Consonant/vowel ratio is a cue for voicing in English," *Percept. Psychophys.* **32**, 141–152.
- Sawusch, J. R., and Pisoni, D. B. (1974). "On the identification of place and voicing features in synthetic stop consonants," *J. Phonet.* **2**, 181–194.
- Silbert, N., and de Jong, K. (2008). "Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production," *J. Acoust. Soc. Am.* **123**, 2769–2779.
- Silbert, N., Townsend, J. T., and Lentz, J. J. (2009). "Independence and separability in the perception of complex nonspeech sounds," *Attention Percept. Psychophys.* **71**, 1900–1915.
- Silbert, N. H. (2010). "Integration of phonological information in obstruent consonant identification," Ph.D. thesis, Indiana University, Bloomington, IN.
- Smits, R. (2001a). "Evidence for hierarchical categorization of coarticulated phonemes," *J. Exp. Psychol. Hum. Percept. Perform.* **27**, 1145–1162.
- Smits, R. (2001b). "Hierarchical categorization of coarticulated phonemes: A theoretical analysis," *Percept. Psychophys.* **63**, 1109–1139.
- Stevens, K. N., and Blumstein, S. E. (1978). "Invariant cues for place of articulation in stop consonants," *J. Acoust. Soc. Am.* **64**, 1358–1368.
- Thomas, R. D. (2001). "Perceptual interactions of facial dimensions in speeded classification and identification," *Percept. Psychophys.* **63**, 625–650.
- Volaitis, L. E., and Miller, J. L. (1992). "Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories," *J. Acoust. Soc. Am.* **92**, 723–735.
- Wickens, T. D. (1992). "Maximum-likelihood estimation of a multivariate Gaussian rating model with excluded data," *J. Math. Psychol.* **36**, 213–234.