

Independence and separability in the perception of complex nonspeech sounds

NOAH H. SILBERT, JAMES T. TOWNSEND, AND JENNIFER J. LENTZ
Indiana University, Bloomington, Indiana

All sounds are multidimensional, yet the relationships among auditory dimensions have been studied only infrequently. General recognition theory (GRT; Ashby & Townsend, 1986) is a multidimensional generalization of signal detection theory and, as such, provides powerful tools well suited to the study of the relationships among perceptual dimensions. However, previous uses of GRT have been limited in serious ways. We present methods designed to overcome these limitations, and we use these methods to apply GRT to investigations of the relationships among auditory perceptual dimensions that previous work suggests are independent (frequency, duration) or not (fundamental frequency [f_0], spectral shape). Results from Experiment 1 confirm that frequency and duration do not interact decisionally, and they extend this finding with evidence of perceptual independence. Results from Experiment 2 show that f_0 and spectral shape tend to interact perceptually, decisionally, or both, and that perceptual interactions occur within, but not between, stimuli (i.e., the interactions suggest correlated noise across processing channels corresponding to perceptually separable dimensions). The results are discussed in relation to lower level sensory modeling and higher level cognitive and linguistic issues.

Sound is inherently multidimensional. Even a simple pure tone must have frequency, phase, amplitude, and duration. It may be that these dimensions are processed (e.g., perceived, encoded, accessed in memory, or responded to) independently of each other. For example, it is possible that the frequency of a pure tone can be processed without regard to any particular (positive) value of duration of the tone. Likewise, the duration of a pure tone might be processed without regard to frequency. In such cases, one dimension generalizes perfectly across levels of another, and processing operates independently on separate dimensions.

On the other hand, it may be that different dimensions interact. For example, the frequency of a pure tone may be processed differently for tones of shorter duration than for tones of longer duration, or the duration of a pure tone may be processed differently for high-frequency tones than for low-frequency tones. Particular combinations of values on two (or more) dimensions may be processed in ways that reflect something qualitatively different than the processing of the component dimensions alone would suggest. In such cases, whole items are, in some cognitively important sense, more than the simple combination of their component parts.

Of course, the same issues arise for more complex sounds, which can vary on a number of other dimensions above and beyond frequency, phase, amplitude, and duration. More complex sounds are of particular interest, since many common and important sounds (e.g., speech, environmental sounds, music) vary dynamically on multiple

dimensions. Understanding how these dimensions relate to each other is of fundamental importance in the study of perception and decision making. Two general questions come immediately to mind: Which dimensions of sound, if any, are processed independently? If a set of dimensions interact, how do they interact and to what degree?

These questions are difficult to answer, in part because the multidimensional nature of sound is seldom directly addressed in auditory perception research. Rather, irrelevant dimensions typically are either held constant (e.g., duration [spectral shape] in a spectral shape [duration] categorization task; Smits, Sereno, & Jongman, 2006) or varied randomly (e.g., overall level in a spectral shape discrimination task; Kidd & Mason, 1992). Nonetheless, there is a substantial and growing body of work addressing multidimensional structure in auditory perception.

A number of studies of nonspeech auditory perception have focused on differences in categorization behavior on different dimensions (Mirman, Holt, & McClelland, 2004) and on differential attention weighting across dimensions (Christensen & Humes, 1996, 1997; Holt & Lotto, 2006; Wang & Humes, 2008). In addition, some work in speech perception has been done on the relative contributions of various phonetic dimensions in phoneme categorization (e.g., Massaro & Oden, 1980; Oden & Massaro, 1978) and on the form of category boundaries in phoneme sequence categorization (Nearey, 1992; Smits, 2001).

Dimensional independence and interaction in auditory perception have been addressed directly in only a limited number of studies. Evidence of independence between fre-

N. H. Silbert, nosilber@indiana.edu

quency and duration has been found in at least one categorization experiment (Espinoza-Varas & Jamieson, 1984), evidence of independence between across-frequency level and modulation phase has been found in some, but not all, individual participants (Richards & Lentz, 1998), and it has been argued that changes in frequency and level are processed independently (Zagorski, 1975).

On the other hand, there is some evidence of asymmetric interactions between frequency and amplitude modulation in pure tones that suggests that, whereas frequency perception is independent of the presence or absence of amplitude modulation, perception of amplitude modulation depends on frequency (Corcoran, 1967). Changes in the degree of interaction among processing channels have been observed as a function of masking noise level and phase in binaural perception of pure tones (Pastore & Sorkin, 1972) and as a function of phase in binaural perception of narrowband noise (Sorkin, Pastore, & Pohlmann, 1972). Evidence of interactions between pure tones at different frequencies has been found in both monaural and binaural perception (Sorkin, Pohlmann, & Gilliom, 1973). Finally, interactions have been observed between fundamental frequency (f_0) and spectral shape (pitch and timbre; Melara & Marks, 1990a; Pitt, 1994; Singh & Hirsh, 1992; Warrier & Zatorre, 2002), between loudness and spectral shape (Melara & Marks, 1990a), and between frequency and loudness (Melara & Marks, 1990b).

Failures of independence have also been found in speech perception. Interactions between various spectral properties (i.e., tongue position and nasalization in synthetic speech; Kingston & Macmillan, 1995; Macmillan, Kingston, Thorburn, Walsh-Dickey, & Bartels, 1999) and between spectral and temporal properties (i.e., tongue root position and voice quality, Kingston, Macmillan, Walsh-Dickey, Thorburn, & Bartels, 1997; and spectral and temporal cues to voicing, Kingston, Diehl, Kirk, & Castleman, 2008) have been documented, as have interactions between phonological dimensions (e.g., place and manner of articulation; Eimas, Tartter, Miller, & Keuthen, 1978).

However, all of these sources of evidence of interaction and independence are limited in important ways. First and foremost, the definition of independence is highly variable from study to study. In some cases, the definition of independence is not stated explicitly (e.g., Zagorski, 1975). In other cases, the definition is essentially operational, such as in studies employing the Garner (1974) speeded classification paradigm, in which interaction of processing channels is defined strictly in terms of response time (RT) relationships across baseline and interference experimental conditions (e.g., Melara & Marks, 1990a; Pitt, 1994). Although an operational definition may provide evidence that a failure of independence is present, the lack of theoretical machinery makes it difficult, if not impossible, to pinpoint the locus of interactive effects (e.g., whether the effect is perceptual, decisional, or both).

Significant limitations are found, even in studies in which considerable theoretical machinery is employed. In some cases, the models and analyses employed are not cognitively motivated. For example, the model used by

Corcoran (1967) provides evidence of the presence and magnitude of dependence of modulation perception on frequency perception, but the model conflates distinct perceptual and decisional processes. In other cases, cognitively motivated models are employed in less than general form. For example, evidence for independence between frequency and duration of pure tones is based on the shape of the boundaries separating the high- and low-frequency and long and short duration regions of the stimulus space; insofar as category structure is modeled at all, it is done so implicitly and only in order to provide category boundary structure (Espinoza-Varas & Jamieson, 1984).

The work of Sorkin and colleagues (Pastore & Sorkin, 1972; Sorkin et al., 1972; Sorkin et al., 1973) comes much closer to providing a general model of dimension interaction, but is still limited by an untested and potentially unwarranted assumption and by a conflation of two distinct notions of independence. The model employed by Sorkin and colleagues consists of a two-dimensional perceptual space containing bivariate Gaussian perceptual distributions and simple decision criteria. Simple decision criteria are decision bounds that are parallel to the coordinate axes of the perceptual space, so this assumption denies the possibility that decision making on one dimension may vary as a function of the other dimension(s).

Sorkin and colleagues employed two methods for assessing correlation (or, more generally, interaction) between dimensions, one that measures within-stimulus (i.e., within-distribution) interactions and one (based on Tanner, 1956) that measures between-stimulus (i.e., between-distribution) interactions and relies on an assumption of statistical independence within each distribution. The assumption of within-distribution statistical independence (i.e., zero correlation) also limits the generality of some findings of dimensional interaction in speech perception (Kingston et al., 2008; Kingston & Macmillan, 1995; Kingston et al., 1997; Macmillan et al., 1999).

The present investigation is based on the multidimensional signal detection theory (SDT; Green & Swets, 1966) approach known as general recognition theory (GRT; Ashby & Townsend, 1986; Kadlec & Townsend, 1992a, 1992b; Maddox, 1995; Olzak, 1986; Thomas, 1995, 1996, 1999, 2001a, 2001b, 2003; Wickens, 1992). The following section provides a detailed description of GRT.

The Structure of GRT

As is the case with a number of other models, in GRT, the probability of a particular response when presented with a stimulus is a function of distinct perceptual and decisional processes. It is assumed in GRT that perceptual effects are random, resulting, over the course of multiple stimulus presentations, in a distribution of perceptual effects.¹ Response regions delimited by decision bounds partition the perceptual space exhaustively such that each perceptual effect is mapped onto one and only one response option. In addition to being conceptually appealing, the use of perceptual distributions and decision bounds is mathematically useful, because they provide a straightforward way to relate the model to observed identification and confusion frequencies.

In its most general form, GRT is nonparametric; no assumptions are made about the functional form of the perceptual distributions or decision bounds. This allows for rigorous, general definitions of three logically distinct notions of independence: two concerning perceptual relationships and one concerning decisional relationships. One of the perceptual notions of independence is defined in terms of microlevel, within-stimulus independence, whereas the decisional and other perceptual notions of independence are defined in terms of macrolevel, between-stimuli independence.

Although GRT may be applied to any number of levels on any number of dimensions, the simplest experimental protocol in which all three notions of independence can be addressed consists of identification of four stimuli defined by a factorial combination of two levels on each of two dimensions. For example, suppose we are interested in the relationships between hue and shape in visual perception and that the levels on these dimensions are *red* versus *purple* and *square* versus *rectangle*, respectively. The probability of responding “purple square” when presented with a purple rectangle is then defined as the *volume* (i.e., the double integral) of the bivariate probability distribution corresponding to presentations of purple rectangle stimuli in the region corresponding to “purple square” responses. The same relation holds, with appropriate changes in distribution and response region, for other combinations of stimulus and response.

The first type of dimensional independence to be addressed is the within-stimulus perceptual notion, called, in GRT terms, *perceptual independence* (PI). PI holds if and only if statistical independence holds within a given perceptual distribution. Statistical independence holds if and only if the probability of every joint perceptual effect is equal to the product of the probabilities of the corresponding marginal perceptual effects. For example, statistical independence holds if, for every degree of redness and squareness, the probability of a particular degree of redness *and* a particular degree of squareness is equal to the probability of that degree of redness without regard to shape times the probability of that particular degree of squareness without regard to hue. Because PI is a within-stimulus construct, it may, in theory, hold or fail for any subset of perceptual distributions in a given model.

The other perceptual notion of dimensional independence is the macrolevel, between-stimuli notion, called *perceptual separability* (PS) in GRT. Within GRT, hue is said to be perceptually separable from shape if and only if the perceptual effect of hue is not affected by the shape of the stimulus. Note that it is logically possible for PS to hold for one level of one dimension while failing for the other level of the same dimension (e.g., the perceptual effects of red stimuli may be identically distributed for both squares and rectangles, whereas the perceptual effects of purple stimuli may be distributed differently for squares and rectangles). For hue in general to be separable from shape, however, the marginal perceptual effect of hue must be equal across shape for both levels of hue (i.e., the marginal perceptual distributions of red stimuli should

be identical for squares and rectangles, as should the perceptual distributions of purple stimuli).

The last notion of dimensional independence to be addressed is *decisional separability* (DS). If decision making on one dimension is not affected by perceptual effects or by the location of the decision bound on the other dimension, then we say that DS holds.

An illustrative example of a two-dimensional Gaussian GRT model is given in Figure 1. It is convenient to represent the bivariate Gaussian densities by equal likelihood contours (i.e., sets of points at a constant height on the density), which produces a circle or an ellipse for each density; here, the plus signs indicate the means of each density.

The two contours of equal likelihood on the left illustrate PI. The contour on the bottom left corresponds to presentations of red squares. In this case, the variance is equal on each dimension and there is no covariance, so the equal likelihood contour is a circle. The contour on the top left corresponds to presentations of purple squares: The variance on the hue dimension is larger than the variance on the shape dimensions, and PI holds, so the equal likelihood contour is an ellipse with the major axis parallel to the *y*-axis (hue) and the minor axis parallel to the *x*-axis (shape). By way of contrast, the contours on the right illustrate failures of PI. The contour on the bottom right, which corresponds to presentations of red rectangles, has a negative covariance, indicating that the more rectangular a percept is, the more likely it is to appear red, whereas the more square a percept is, the more likely it is to appear purple. In this case, the major axis of the ellipse has a negative slope. The density on the top right illustrates failure of PI due to positive covariance, indicating that the more rectangular a percept is, the more likely it is to

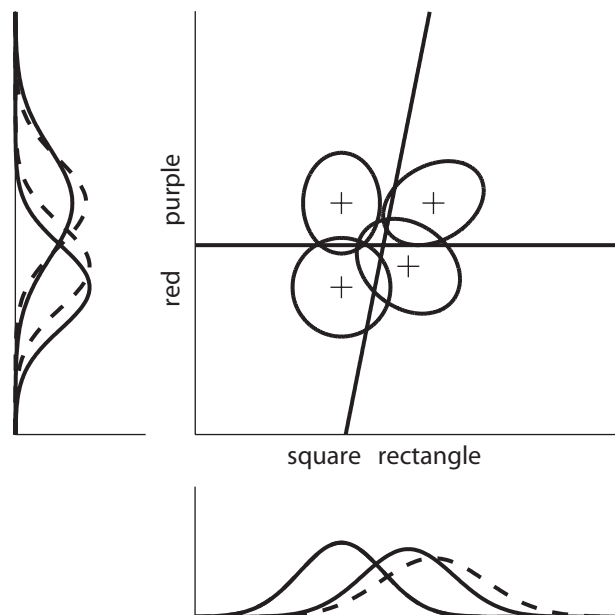


Figure 1. Illustrative two-dimensional Gaussian general recognition theory model. See text for details.

appear purple, and the more square a percept is, the more likely it is to appear red. In this case, the major axis of the ellipse has a positive slope.

The marginal densities illustrate (failures of) PS. The two bivariate densities on the left (red and purple squares) have identical means and variances on the shape dimension (i.e., the x -axis), so their marginal shape densities are identical (i.e., there is a single univariate Gaussian distribution under the shape axis under the “square” label). In this case, we say that shape is perceptually separable from hue at the square level. The bottom two bivariate densities (red squares and rectangles) illustrate failure of PS due to an upward shift (i.e., toward purple) in the hue mean of the red rectangle density relative to the red square density, although the variance on the hue dimension (y -axis) is identical for these two densities. The top two bivariate densities (purple squares and rectangles) illustrate failure of PS due to greater variance on the hue dimension (y -axis) in the purple square density than in the purple rectangle density, whereas the (purple) hue means are identical across levels of the shape dimension. Finally, the two bivariate densities on the right (red and purple rectangles) illustrate failure of PS due to inequalities in the shape means *and* variances across levels on the hue dimension. Here, relative to the red rectangle density, the purple rectangle density has larger shape variance and a rightward-shifted shape mean.

Finally, the decision bound partitioning the hue dimension (i.e., the horizontal bound) illustrates DS, indicating that decisions about hue do not depend on shape. The decision bound partitioning the shape dimension illustrates one way in which DS could fail.² In this case, decisions about shape depend on perception of hue, such that a bias exists to label red percepts as rectangles and to label purple percepts as squares.

Perceptual dependence can be interpreted as correlated noise in separate processing channels. To the extent that frequency channels are determined by (cochlear) auditory filters, then, we might reasonably expect to observe perceptual dependence if we were to present noise-masked sinusoids very close in frequency. Failure of PS can arise, on the other hand, if the salience of the level on one dimension varies as a function of the level of the other dimension. For example, a difference in the frequencies of two sinusoids should be detected more easily if the tones are presented well above threshold than when presented near threshold.

Although PI, PS, and DS are logically distinct, it may be that, in practice, they tend to hold or fail together. For example, failures of both PI and PS can be produced by a (neural) circuit that sums information across processing channels prior to response selection (e.g., in grating orientation judgments across spatial frequencies, as described in Olzak & Wickens, 1997). Failures of DS can arise if an observer is making optimal decisions on the basis of shape and location of perceptual distributions and if PI and/or PS have failed in such a way that the optimal decision bounds are not parallel to the coordinate axes (see, e.g., Ashby & Gott, 1988).

Regardless of when they hold or fail, PI, PS, and DS are defined in terms of unobservable perceptual distribu-

tions and decision bounds. For these notions of independence to be tested, they must be mapped onto observed identification–confusion data. Parameter estimation via model fitting has proven to be a powerful method of evaluating PI, PS, and DS, although seldom simultaneously (Ashby & Lee, 1991; Thomas, 2001b).

Parameter estimation is the process (in Gaussian GRT) of finding means, covariance matrices, and decision bounds that predict, as closely as possible, the observed identification–confusion probabilities. Ashby and Lee (1991) employed parameter estimation in a study of the relationships among identification, similarity judgment, and categorization tasks, as did Thomas (2001a, 2001b), in conjunction with tests of sampling independence and marginal response invariance (Ashby & Townsend, 1986), as well as SDT-based tests (Kadlec & Townsend, 1992a), to investigate interactions between two different pairs of features in realistic line drawings of faces.

The present work is intended, in part, to address limitations to much of the previous work in the GRT framework. The tests of independence developed by Ashby and Townsend (1986) and employed by Thomas (2001b) can test only conjunctions of forms of independence (e.g., PI and DS; PS and DS), and the SDT-based tests developed by Kadlec and Townsend (1992a) can diagnose only a subset of possible types of failure of independence or separability. Previous uses of model fitting with GRT have used parameter estimation while relying on the assumption that DS holds (e.g., Thomas, 2001a, 2001b) or that, when it fails, it fails as a function of decision bounds on the other dimension (Ashby & Lee, 1991).

Although parameter estimation can, in theory, provide powerful tests of PI, PS, and DS, it presents its own set of difficulties. The primary limitation on parameter estimation in the GRT framework is the fact that, in a standard GRT experiment (i.e., an experiment in which four stimuli comprise the combination of two levels on each of two dimensions), the full Gaussian GRT model has more free parameters than the data have degrees of freedom. Assuming linear decision bounds, the full Gaussian GRT model has 21 free parameters (mean vectors and covariance matrices for each distribution, and slopes and intercepts for each decision bound), whereas a 4×4 confusion matrix has only 12 degrees of freedom. Because such data does not constrain a model with so many free parameters, multiple distinct configurations of perceptual distributions and decision bounds may fit the data perfectly.

In order to reduce the number of free parameters in the model, additional assumptions may be introduced. For example, DS can be assumed to hold (and argued to hold on the basis of familiarization with the stimuli, as in Thomas, 2001b, or the decision procedure employed, as in Wickens & Olzak, 1989), and equal variances (but not necessarily covariances) can be assumed, because any change in a variance can be offset by a change in an appropriate mean. Further restrictions on the model’s parameters can be made (e.g., by assuming that PI and PS hold). Restricted models can then be tested against more general models in which these assumptions are relaxed. Unfortunately, the standard GRT task cannot simultaneously test PI, PS,

and DS. Ashby and Lee (1991) were able to test all three notions of independence by employing a larger stimulus set. Identification data for stimuli consisting of a factorial combination of three levels on two dimensions provided them with 72 degrees of freedom, and the general Gaussian GRT model for this situation has 53 free parameters.

Another way to generate data with degrees of freedom sufficient to constrain the general Gaussian GRT model is to manipulate stimulus presentation base rates (and/or payoff schemes; see, e.g., Maddox, 1995; Maddox & Bohil, 1998a, 1998b, 2003) to produce, for example, 4×4 confusion matrices for multiple experimental conditions. It is worth noting that, although base-rate manipulations can produce data sufficiently rich to constrain the GRT model, an additional assumption about how (or whether) perceptual distributions and decision bounds change across base-rate conditions is required. Because this method is employed in the present project, next we describe the base-rate conditions employed and these additional assumptions required by their use.

Base-Rate Manipulations and Degrees of Freedom

Base-rate and payoff manipulations have long been employed in studies employing (unidimensional) SDT, often to enable the use of isosensitivity curves in data analysis. It is typically assumed that shifts in base rate induce shifts in decision criteria while having no effect on the location or shape of the perceptual distributions. A number of multidimensional categorization experiments provide compelling evidence that perceptual distributions are stationary, whereas decision bounds shift in response to changes in stimulus presentation base-rate and payoff schemes (Maddox, 1995; Maddox & Bohil, 1998a, 1998b, 2003). In many cases, asymmetries in base rates between two categories induce larger-than-optimal shifts in decision bounds, whereas decision bound shifts due to payoff manipulations tend to be overly conservative (e.g., Maddox & Bohil, 1998b).

We employ base-rate manipulations here with multidimensional stimulus-response sets to enable simultaneous tests of all three GRT notions of dimensional interaction. Under the assumption that perceptual distributions are unaffected whereas decision bounds shift in response to changes in base rate, the number of degrees of freedom in the data grows more quickly than the number of free parameters needed to account for decision bound shifts. We employ five base-rate conditions to produce data with 60 degrees of freedom; the additional decision bound parameters add only 8 free parameters to the model (for a total of 29).

EXPERIMENT 1

As noted above, there is some fairly strong evidence that frequency and duration are processed at least partly independently (i.e., in decision making)—namely, the observation that frequency and duration category boundaries depend only on frequency and duration, respectively (Espinoza-Varas & Jamieson, 1984). There is substantial indirect evidence of (perceptual) independence between

these dimensions. Duration discrimination does not appear to depend on the frequency bandwidth of noise signals (Abel, 1972), and work on detection of brief (i.e., <100 msec) pure tones (i.e., work on temporal integration) suggests that frequency and duration are independent in that the same basic pattern of duration and intensity trade-offs occur at multiple frequencies (e.g., Florentine, Fastl, & Buus, 1988; Plomp & Bouman, 1959).

Consideration of models of frequency and duration encoding also leads us to expect PI and separability between these dimensions. Insofar as frequency is encoded by location on the basilar membrane (and subsequent tonotopic neural activity), and insofar as duration is encoded by a counting process (e.g., of neural pulses, as in Creelman, 1962), we do not expect either to influence the other.

It is, perhaps, important to note that work on (silent or partly filled) gap detection suggests that frequency and duration may interact (e.g., Eddins, Hall, & Grose, 1992; Green & Forrest, 1989; Shailer & Moore, 1983, 1985, 1987). However, because the duration of interest here is that of noise signals (i.e., filled intervals, not silent gaps), we do not expect such interactions.

The present experiment seeks to confirm and extend previous results by employing the full machinery of GRT to probe PI, PS, and DS of frequency and duration in broadband noise stimuli. Broadband noise stimuli were used because they are more complex and speech-like than, yet still comparable to, the pure tone stimuli employed in the most closely related previous work on frequency and duration (Espinoza-Varas & Jamieson, 1984).

Method

Participants. Seven adults (3 male, 4 female; 1 participant was the first author) were recruited from the university community. All participants were screened to ensure normal hearing. The average age of participants was 24.8 years (range, 22–33 years). All but 1 were right-handed, and all but 1 were from the Midwestern United States. All were native speakers of English and had, on average, 4 years of foreign language study. Two had 4–6 years' experience as musicians.

Stimuli. The stimuli were 1000-Hz wide broadband noises. Each stimulus took one of two values on each of two dimensions: duration and frequency range (or, equivalently, center frequency). Long stimuli were 300 msec; short stimuli were 250 msec. All had 10-msec (squared [cos]sine) rise/fall times. High-frequency-range stimuli extended from 510 to 1510 Hz (center frequency = 1010 Hz), and low-frequency-range stimuli extended from 490 to 1490 Hz (center frequency = 990 Hz). All were embedded in background white noise (10–12010 Hz) and were presented monaurally to each participant's right ear at a +9-dB signal-to-noise ratio at 60 dB SPL (approximately the level of conversational speech). Stimulus sounds always began 200 msec after the onset of the background noise, and the background noise always lasted 800 msec. The stimuli were generated at the beginning of each experimental block, and each stimulus and each segment of background noise were unique waveforms (i.e., participants never heard exactly the same stimulus or background noise within or across experimental blocks). A representation of these stimuli is given in Figure 2.

Procedure. Each participant was seated alone in a double-walled sound-attenuating booth in front of a computer terminal. Stimuli were presented via a Tucker-Davis Technologies real-time processor (TDT RP2.1; sampling rate = 24414 Hz), programmable attenuator (TDT PA5), and headphone buffer (TDT HB6), and Sennheiser HD250 Linear II headphones. Before the first session (familiariza-

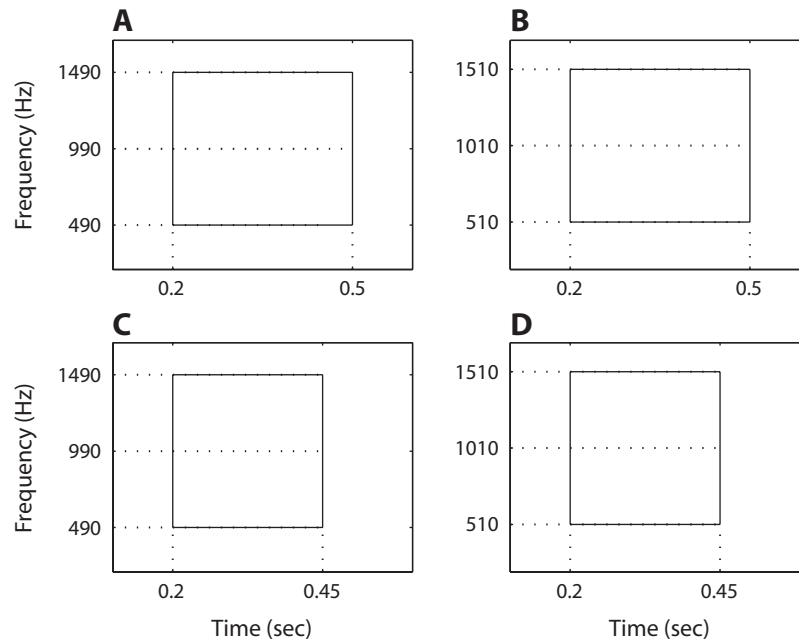


Figure 2. Schematic illustration of the frequency \times duration stimuli. The x-axis on each panel represents time; the y-axis represents frequency. Solid-line rectangles represent the broadband noise stimuli. Dashed lines indicate the spectral and temporal limits within any given trial, as well as the center frequency of each stimulus type. Panels A and B represent long-duration stimuli; panels C and D represent short-duration stimuli. Panels A and C represent low-frequency stimuli; panels B and D represent high-frequency stimuli.

tion and training), participants read an instruction sheet, were given verbal instructions, and were prompted for questions about the procedure. Sessions lasted from one to four experimental blocks, each lasting approximately 35 min (after completion of the data collection discussed here, the same participants also participated in related experimental blocks, with a subset of the stimuli lasting 15–20 min; the results of these shorter blocks are not presented here).

Each experimental block began with brief written instructions reminding participants to respond as accurately and as quickly as possible, indicating the relative base rates of the stimuli and providing explicit guessing advice for trials on which the participant was uncertain of the stimulus identity. After the instructions were cleared from the screen, four button icons corresponding to the buttons on the button box became visible. The levels and dimensions were labeled to the left of and below the on-screen buttons, and on each on-screen button, the level combination (i.e., “low, short”; “high, short”; “low, long”; and “high, long”) appeared in black text. Button-response assignments were held constant; the x-axis always corresponded to duration (with “short” on the left and “long” on the right), and the y-axis always corresponded to center frequency (with “low” on the bottom and “high” on the top).

Each trial consisted of the following steps: (1) a visual signal (the word “listen”) presented on the computer monitor, (2) 0.5 sec of silence, (3) stimulus presentation, (4) response, (5) feedback, and (6) 2 sec of silence. Responses were collected via a button box, with buttons arranged to correspond to the structure of the stimulus space (i.e., the corners of a square corresponding to the two levels on two dimensions of the stimuli). Feedback was given visually via color-coded text above the on-screen buttons and via changes in the on-screen buttons. The word “Correct” or the word “Incorrect” appeared along with brief descriptions of the presented stimulus and the response chosen. On each trial, text corresponding to correct responses appeared in green, and the color of the text on the correct on-screen button changed to green. When incorrect responses were

given, text corresponding to incorrect responses appeared in red, and the color of the text on the incorrect response on-screen button changed to red. Before each successive trial, the feedback text disappeared and the button text color was reset to black.

The relative base rates of the stimuli for each experimental condition are given in Table 1. Rows and columns indicate duration and frequency range, respectively. Each box represents a base-rate condition, and the numbers in the boxes indicate the relative base rate of each stimulus within each condition. The *standard* condition, in which each stimulus type is presented an equal number of times, is represented in the middle box. In each of the four other conditions, stimulus base rates were shifted such that, in one of the four shifted base-rate conditions, stimuli at one level of a single dimension were presented four times more often than were stimuli at the other value on the same dimension, whereas pairs of stimuli sharing a value on a manipulated dimension were presented equally often (i.e., base rate on the nonmanipulated dimension was held constant across levels).

Table 1
Base-Rate Conditions (%)

Stimulus Duration	Stimulus Frequency					
	Low	High	Low	High	Low	High
Long			10	10		
Short			40	40		
Long	10	40	25	25	40	10
Short	10	40	25	25	40	10
Long			40	40		
Short			10	10		

Note—Solid lines enclose experimental conditions. Base rates are indicated in relative terms (i.e., 40% in a low–short cell and 10% in a low–long cell indicate four times as many low–short stimuli as low–long stimuli in that condition).

For example, in the topmost box, short stimuli are presented four times more frequently than are long stimuli; low- and high-frequency stimuli are presented equally frequently. The explicit guessing advice given at the beginning of each block indicated, for example, that short stimuli would be heard four times as often as long stimuli, so that, if unsure, a participant was four times more likely to be correct during that block by guessing “short” rather than “long.”

Each participant received 2 short (approximately 15 min) and 2 regular-length equal base-rate blocks to familiarize them with the sounds and to ensure that performance was consistently above chance. Data collection took place over 10 experimental blocks (2 for each base-rate condition), and each block consisted of 500 trials, for a total of 5,000 trials of collected data. Participants were paid \$8/h, with a \$2/h bonus for completion of the experiment. The participant with the highest accuracy received a \$10 bonus, as did the participant with the fastest overall RTs.

Analyses. Parameter estimation was carried out as follows. For each model, fit was evaluated via the Bayesian information criterion (BIC; Kass & Raftery, 1995), defined as $BIC = -2\log(L) + k\log(N)$, where $\log(L)$ is the natural logarithm of the (multinomial) likelihood for a given set of parameters, k is the number of free parameters in the fitted model, and N is the number of observations (in both experiments, $N = 5,000$). The BIC has two desirable properties: It weighs goodness of fit against the complexity of the model, which can help reduce overfitting (i.e., fitting unknown random effects rather than the perceptual and decisional structures of interest), and it provides an estimate of an intuitive measure of relative goodness of fit between models (i.e., the Bayes factor). The lower the BIC, the better the fit of the model.

Predicted identification confusion probabilities (to be employed in the likelihood function) were calculated by numerical approximation to the double integrals of bivariate normal densities over response regions delineated by linear decision bounds. After each model was fit, a small amount of noise was added to each parameter, then the fit of the new set of parameters was evaluated. If the new set of parameters improved the fit, the new set was kept. Otherwise, an independent sample of noise was added to the old set of parameters. If only worse fits were found for 10 consecutive attempts, the magnitude of the noise added to the parameters was decreased, in order to more fully explore any (local) minima in the parameter space. If only worse fits were found for another 10 consecutive attempts, the magnitude of the noise added to the parameters was increased, in order to attempt to escape from any local minima in the parameter space. The procedure ended after 500 improvements in fit were found.

In order to simplify the exploration of high-dimensional parameter spaces, following Ashby and Lee (1991), the first model to be fit allowed only the means of the perceptual distributions and decision bound intercepts to vary (all marginal variances were set equal to 1, all covariances were set equal to 0). Next, intercepts, means, and variances were allowed to vary. Then intercepts, means, variances, and covariances were allowed to vary. Finally, the slopes of the decision bounds were also allowed to vary. Multiple random initial parameter sets all resulted in similar parameter configurations after the *means-only* step, suggesting that the procedure consistently converges on the same fitted model.

Each participant's data was fit to a set of Gaussian GRT models with linear decision bounds. Figure 3 illustrates the hierarchical relationships among the fitted models. The most general model is at the top of the graph; the most restricted model is at the bottom. Lines connecting two models indicate a hierarchical relationship in which the lower model is obtained by restricting the values of a subset of the more general model's parameters. In the most general model, PI, PS, and DS may all fail (i.e., the dimensions may interact within each stimulus, across stimuli, and with regard to decision making). The most general model can be restricted such that PI or PS or DS is forced to hold; this is represented in the second row from the top. Each of these models can be further restricted by forcing a second form of dimensional independence to hold; this is represented in the third row from the top. Finally, the most restricted model is obtained

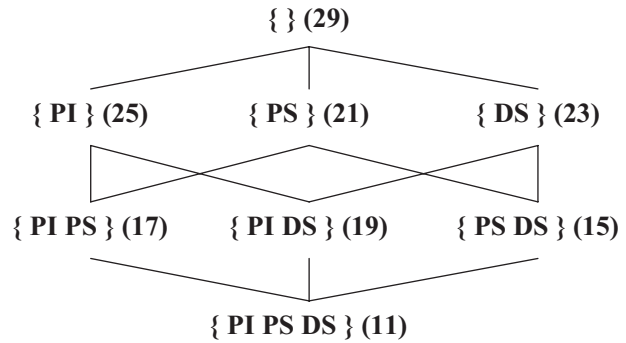


Figure 3. Fitted-model hierarchy. PI, perceptual independence; PS, perceptual separability; DS, decisional separability. See text for details.

by forcing PI, PS, and DS to hold. The number of free parameters for each model appears in parentheses.

The graph in Figure 3 is incomplete in three respects. First, the models displayed were chosen because the focus of this work is on dimensional interactions. A large number of parameter restrictions not considered here could readily be implemented (e.g., a single correlation parameter for all distributions). Second, PS was tested separately for each dimension as well as jointly for both (i.e., for each PS model in Figure 3, three models were fit to the data: one in which frequency is separable from duration, one in which duration is separable from frequency, and one in which frequency range and duration are mutually separable). Third, for each model that allows failure of DS, two models were fit: one in which decision bounds corresponding to different base-rate conditions could have different slopes and one in which all the decision bounds for a given dimension had the same slope (i.e., in which a change in base rate merely shifts the intercept of a decision bound).

Results

Table 2 provides each participant's proportion of correct identifications (p_{correct}), their mean RT (in milliseconds, measured from stimulus onset), their best-fitting model, and two measures of overall goodness of fit of the model. Because there is no strict criterion for evaluating overall goodness of fit, a number of measures are employed here. The two rightmost columns, respectively, provide the mean and median absolute values of the differences between the observed and predicted identification confusion probabilities; smaller values indicate better fit.

Accuracy rates were well above chance for all participants (column 2), and mean RTs were similar across participants; there is no obvious relationship between accuracy and mean RT. All but 2 participants had the same best-fitting models—namely, the model simultaneously exhibiting DS, PS, and PI. The best-fitting models of the other 2 participants exhibited PI, partial PS (Participant 2's best-fitting model, labeled PS, suggests that frequency is separable from duration, but not vice versa), and failure of DS. Mean and median deviations of predicted from observed identification confusion probabilities (columns 7 and 8) indicate that overall model fits are quite good.

Figure 4 depicts the best-fitting models for Participants 2, 3, and 5. For each participant, the top right portion of each panel shows decision bounds and equal likelihood contours for the modeled representation of

Table 2
Model Fit Statistics

Participant	Accuracy		Model			Goodness of Fit	
	(p_{correct})	μ_{RT}				$M_{ \text{predicted} - \text{observed} }$	$M_{ \text{predicted} - \text{observed} }$
1	.597	1,201	DS	PS	PI	.021	.013
2	.490	1,029		PS _{FR}	PI	.039	.034
3	.722	1,380	DS	PS	PI	.028	.026
4	.633	1,179	DS	PS	PI	.029	.031
5	.603	1,283		PS	PI	.087	.092
6	.643	932	DS	PS	PI	.031	.029
7	.701	995	DS	PS	PI	.057	.059

Note— μ , mean; M , median; PI, perceptual independence; PS, perceptual separability; DS, decisional separability; FR, frequency.

two-dimensional perceptual space. The solid decision bounds correspond to the equal base-rate condition, and the dotted decision bounds correspond to the four unequal base-rate conditions. In every case, observed shifts in decision bounds increased the bias toward (i.e., the size of the response region corresponding to) the more frequently presented stimuli.

The bottom right and top left portions of each panel depict the marginal densities for frequency and duration, respectively. Only two marginal densities are visible, because PS holds in the best-fitting models for all participants but one (Participant 2, for whom frequency appears to be perceptually separable from duration, but not vice versa). Finally, the bottom left portion of each panel shows a scatterplot of predicted (y -axis) and observed (x -axis) identification confusion probabilities to aid in assessment of the overall fit of the model. The better the fit, the closer the points should be to the diagonal dotted line connecting coordinates (0,0) and (1,1).

Figure 4A depicts the best-fitting model for Participant 2, the only participant whose best-fitting model includes failure of PS, and 1 of 2 participants whose best-fitting models include failures of DS. The large degree of overlap among the four perceptual distributions reflects the fact that Participant 2 had relatively low accuracy (although still well above chance). Figure 4B depicts the best-fitting model for Participant 3. Although there is some between-participant variation in the magnitude of decision bound shifts and the ratios of marginal variances, Participant 3's best-fitting model is reasonably typical of the best-fitting models, including all three forms of dimensional independence. Figure 4C shows Participant 5's best-fitting model, which includes, along with PS and PI, a low-magnitude failure of DS.

Discussion

The model-fitting and comparison results provide strong evidence that frequency (range) and duration are independent and separable at the particular stimulus values investigated here, with the exceptions that Participant 2 exhibited at least partial failures of each kind of dimensional independence, and Participant 5 exhibited failures of DS.

A possible explanation for Participant 2's differences from the others is that the nature of the relationships among dimensions changes as a function of overall accuracy. Par-

ticipant 2 had the lowest accuracy in this experiment, although it is worth noting that Participant 2's average RT was not excessively fast or slow, either of which might be expected to correspond to low accuracy. It may be that frequency and duration are more likely to be independent and separable only when the levels on each dimension are sufficiently salient. Investigation of such a possibility is, of course, beyond the scope of the present article.

The widespread presence of DS is consistent with Espinoza-Varas and Jamieson's (1984) work on these dimensions, and the apparent presence of PI and PS in most of our participants extends these earlier findings substantially.

Theories of frequency encoding employ mechanisms such as location on the basilar membrane and phase locking of neural impulses to signal frequency, neither of which plausibly plays a role in duration encoding. Similarly, there is no obvious reason to expect duration encoding to rely on either of these frequency analysis mechanisms. For these reasons, we expect that other spectral properties of sound (e.g., spectral shape) may also be independent and separable from duration. Of course, arguments based on low- or even high-level physiology have often proven faulty in the face of behavioral data, requiring that appropriate psychophysical experiments be used for assessing such expectations.

The use of broadband noise in the present study (as opposed to pure tones, as used in Espinoza-Varas & Jamieson, 1984) was motivated in large part by a desire to make the stimuli relevantly speech-like, while avoiding the confounds produced by the extensive experience that adult native speakers have with speech. To the extent that frequency range and duration are the primary cues to different phonological contrasts (e.g., as in place of articulation and voicing in fricatives; Silbert & de Jong, 2008; Stevens, 1998), the present results suggest that these contrasts (i.e., phonological dimensions) are independent and separable, as well. Of course, such issues must be examined directly.

EXPERIMENT 2

In contrast to frequency and duration, there is evidence that f_0 and spectral shape are *not* processed independently. For example, speeded classification of sounds on either pitch (i.e., f_0) or timbre (i.e., spectral shape) consistently

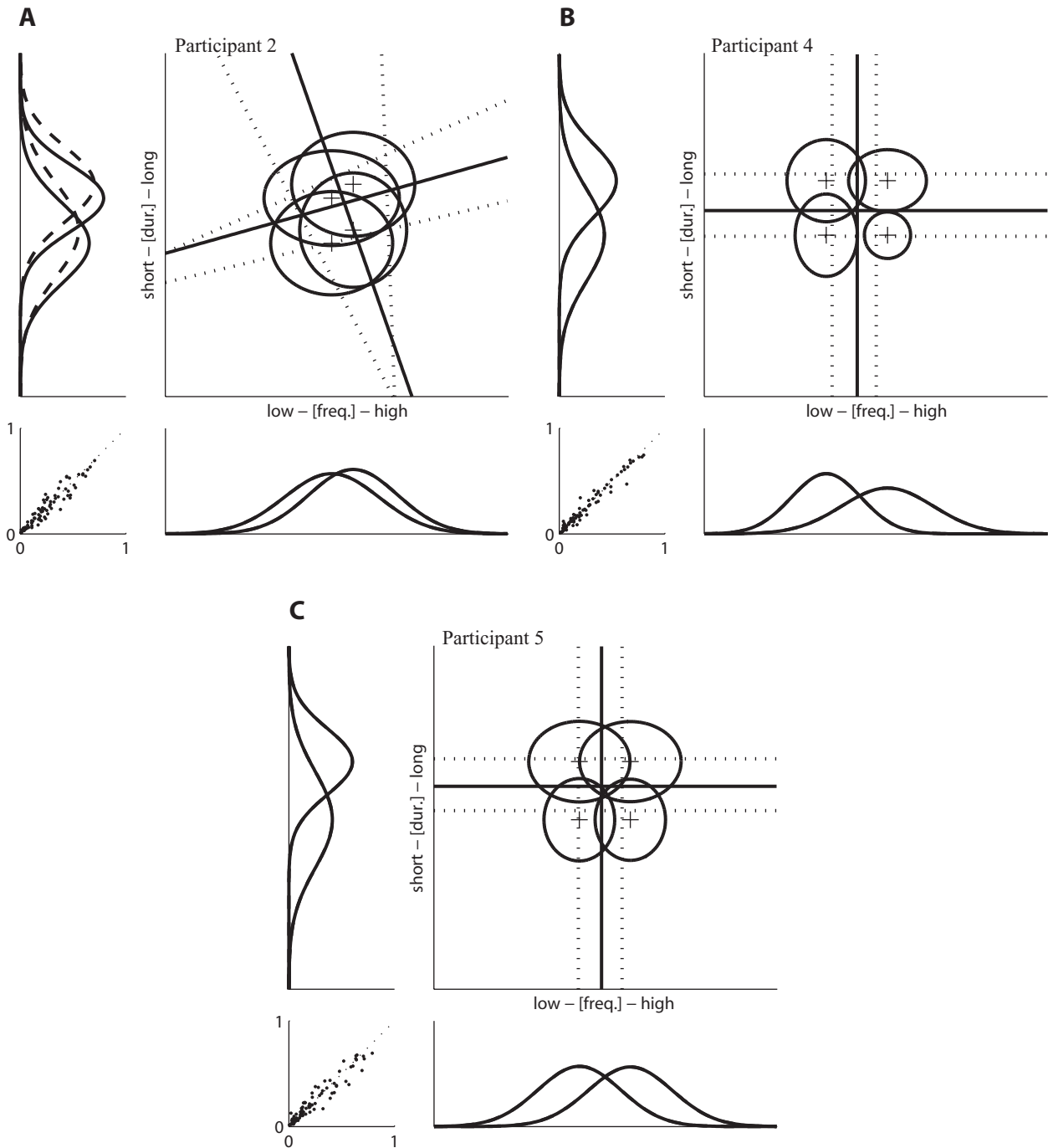


Figure 4. Panels A, B, and C depict the best-fitting Gaussian GRT models for Participants 2, 4, and 5, respectively. The top right portion of each panel shows decision bounds and equal likelihood contours of fitted perceptual distributions. Plus signs (+) indicate distribution means, the x -axis represents the perceptual effect of frequency, and the y -axis represents the perceptual effect of duration. The top left and bottom right portions of each panel show marginal perceptual distributions. Solid lines indicate marginal distributions for Level 1 (low frequency or short duration) on the other dimension; dashed lines indicate marginal distributions for Level 2 (high frequency or long duration) on the other dimension. The bottom left portion of each panel shows predicted (y -axis) \times observed (x -axis) identification confusion probabilities.

produces interference between these two dimensions (e.g., Krumhansl & Iverson, 1992; Melara & Marks, 1990a). Similarly, patterns of identification of changes in f_0 or spectral locus (i.e., the range of frequencies over which

components of a harmonic complex are present in a stimulus) indicate that small changes in f_0 (about 4 Hz or less) can be overridden by changes in spectral locus, such that changes in spectral shape are perceived as changes in both

f_0 and spectral shape, suggesting partial conflation of these two acoustic dimensions onto a single perceptual dimension (Singh & Hirsh, 1992). Although the bulk of evidence indicates that these dimensions interact, it is worth noting that there is also evidence of independence between f_0 and spectral shape in data from tasks probing short-term memory (Semal & Demany, 1991).

Although there are multiple sources of evidence of interaction between f_0 and spectral shape, the locus of interaction of these two dimensions has not been directly investigated. Interference in speeded classification tasks (e.g., Melara & Marks, 1990a) cannot disambiguate between perceptual and decisional sources of interaction (Maddox, 1992). Similarly, although the discrimination task employed by Singh and Hirsh (1992) provides strong evidence of perceptual interactions, their data are compatible with interactions at multiple levels; a failure of PI, PS, or both can correspond to perceptual conflation of distinct physical dimensions (see, e.g., Olzak & Wickens, 1997).

In Experiment 2, we sought to confirm and extend earlier results by employing the techniques described in Experiment 1 to probe the relationships between f_0 and spectral shape in harmonic complexes. We chose the stimuli used in Experiment 2 because they are comparable to the stimuli employed in previous work on these dimensions, yet still provide reasonably speech-like properties.

Method

Participants. Seven adults (1 male, 6 female; 2 of the females also participated in Experiment 1) were recruited from the university community. Screening ensured that all had normal hearing. The average age was 26.7 years (range, 20–39 years). All but 1 were right-handed, and all but 1 were from the Midwestern United States. All but 1 were native speakers of English (the other was a native speaker of Serbian) and had, on average, 6 years of foreign language study (this includes years studying English for the Serbian native speaker). Four had some experience as musicians (average, 15 years).

Stimuli. The stimuli were 13-component harmonic complexes (i.e., the sum of 13 sinusoids, all of which were consecutive integer multiples of the frequency of the lowest component). Each stimulus took one of two values on each of two dimensions: f_0 (the frequency of the lowest component) and spectral shape. The low f_0 was 150 Hz, the high f_0 was 152 Hz. Spectral shape was manipulated by multiplying flat spectra by a raised Gaussian curve centered at either 850 Hz (low spectral prominence) or 1050 Hz (high spectral prominence); the Gaussian curve had a standard deviation of 150 Hz, which caused three or four components to have substantially higher amplitude than the rest of the components (the largest difference between a prominent and nonprominent component was approximately 7 dB). Spectral shaping was carried out on power spectra, followed by inverse Fourier transformation (i.e., the spectra were not decibel transformed). All stimuli were normalized, so that raw differences in energy between stimuli would not cue spectral shape differences. All stimuli were embedded in white background noise (10–12010 Hz) presented monaurally to each participant's right ear at 60 dB SPL, with a +5.5-dB signal-to-noise ratio. Stimulus sounds always began 200 msec after the onset of the background noise, and the background noise always lasted 800 msec. The stimuli were generated at the beginning of each experimental block, and each stimulus and each segment of background noise were unique waveforms, because the beginning phase of each component was randomized when the stimuli were generated (i.e., participants never heard exactly the same stimulus or background noise within or across experimental blocks). A schematic representation of these stimuli is given in Figure 5.

Procedure. The procedure for Experiment 2 was identical to that for Experiment 1, except for the labels of the levels and dimensions. In Experiment 2, the f_0 dimension levels were described as “low” and “high” (pitch), and the spectral shape dimension levels were described as “dull” (low spectral prominence) and “sharp” (high spectral prominence).

Results

The proportion correct, mean RT, best-fitting models, and mean and median deviations between predicted and observed identification confusion probabilities are given in Table 3.

As is indicated in Table 3, there is considerable variation among participants. Accuracy was well above chance, and mean RT was similar for all participants. The best-fitting models indicated that mutual PS between f_0 and spectral shape held for all but 1 participant; Participant 3's best-fitting model had f_0 separable from spectral shape, but not vice versa. As was indicated by the mean and median predicted–observed response probability deviations, overall model fits are quite good.

Figure 6 depicts the best-fitting models of Participants 3, 5, and 7, chosen to illustrate a number of interesting patterns evident in the model fits. All but 1 participant (Participant 5) exhibited failure of DS, and only Participant 7's best-fitting model included nonparallel decision bounds across base-rate conditions. The spectral shape (i.e., horizontal) bounds for Participants 2, 4, 6, and 7 have negative slopes (i.e., decreased the size of the regions for low f_0 –high spectral prominence and high f_0 –low spectral prominence responses). A positive spectral shape bound slope can be seen in Figure 6A (Participant 3), whereas a large negative spectral shape bound slope can be seen in Figure 6C (Participant 7). The f_0 (i.e., vertical) bounds for Participants 1, 2, 4, 6, and 7 had positive slopes (i.e., decreased the size of the regions for low f_0 –high spectral prominence and high f_0 –low spectral prominence responses), fairly large positive f_0 bound slopes can be seen in Figure 6C (Participant 7), and no participant's best-fitting model included f_0 bounds with negative slope.

The best-fitting models of Participants 1, 4, 5, 6, and 7 exhibited failure of PI, and in each case, this took the form of negative covariance, often of large magnitude. For 4 participants (Participants 3, 5, 6, and 7), failure of PI, failure of DS, or both appear to conspire to increase confusions between the low f_0 –high spectral prominence and high f_0 –low spectral prominence stimuli than between the low f_0 –low spectral prominence and high f_0 –high spectral prominence stimuli. The patterns of failure of PI counterbalance the decrease in the size of the corresponding response regions due to failure of DS for Participants 1, 4, 6, and 7.

As in Experiment 1, there is no obvious correspondence between mean RT, accuracy, and the observed patterns of interaction.

Discussion

The present results are largely consistent with previous work on pitch and timbre interactions, keeping in mind the differences in methodologies employed. Speeded classifi-

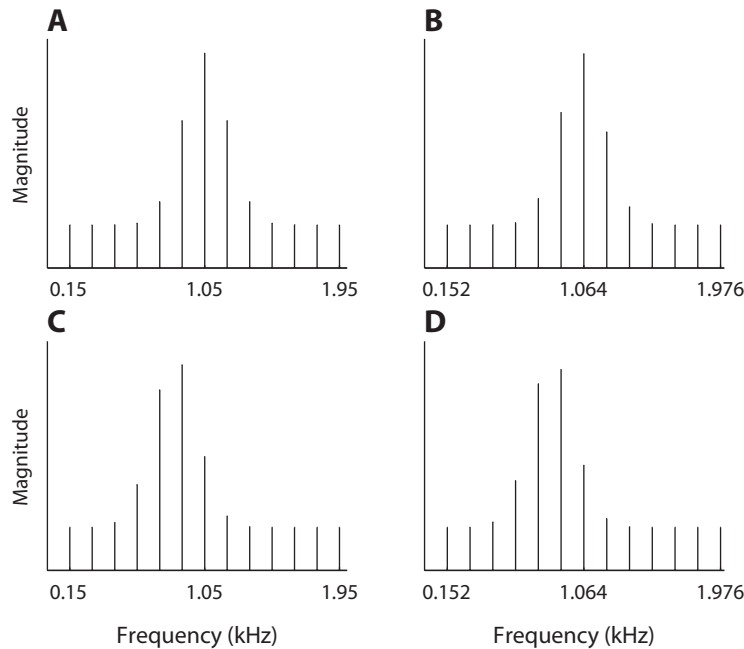


Figure 5. Schematic illustration of the $f_0 \times$ spectral shape stimuli. The x -axis on each panel represents frequency (in kHz), the y -axis (relative) power (i.e., each panel represents the power spectrum of the stimuli prior to inverse Fourier transformation). Panels A and B represent high spectral prominence stimuli; panels C and D represent low spectral prominence stimuli. Panels A and C represent low f_0 stimuli; panels B and D represent high f_0 stimuli. The frequencies (in kHz) of the 1st, 7th, and 13th frequency components are given on the x -axis of each panel.

cation of sounds varying in pitch and timbre in the Garner paradigm (Melara & Marks, 1990a; Pitt, 1994) has consistently shown that these dimensions interact. However, the Garner (1974) paradigm often relies on unstated assumptions and cannot be used to identify the locus of interactive effects (Maddox, 1992). Furthermore, the Garner paradigm is based on interference (or its absence) in a selective attention task, which is quite different from the identification task employed here, in which participants must allocate attention to all relevant dimensions. The task employed by Singh and Hirsh (1992) also required attention to both f_0 and spectral properties of the stimuli, although their analyses conflate within-stimulus and between-stimuli perceptual levels of analysis. Our GRT-based analyses at the individual participant level suggest that the interactions

between these dimensions can vary somewhat across participants and that, when present, these interactions take the form of failures of PI, of DS, or both.

For example, the best-fitting model for Participant 5 exhibits negative correlation (i.e., failure of PI) in the low f_0 –high spectral prominence and high f_0 –low spectral prominence distributions. The best-fitting model for Participant 7 also exhibits negative correlations as well as failures of DS, such that the response regions for the high f_0 –high spectral prominence and low f_0 –low spectral prominence are enlarged relative to the low f_0 –high spectral prominence and high f_0 –low spectral prominence response regions. The large relative confusability of low f_0 –high spectral prominence and high f_0 –low spectral prominence observed here is consistent with Singh and

Table 3
Model Fit Statistics

Participant	Accuracy (p_{correct})	μ_{RT}	Model		$\mu_{ \text{predicted} - \text{observed} }$	$M_{ \text{predicted} - \text{observed} }$
1	.574	1,072	PS		.020	.020
2	.819	1,083	PS	PI	.028	.018
3	.756	963	PS	f_0 PI	.034	.035
4	.707	1,094	PS		.062	.039
5	.855	774	DS	PS	.015	.010
6	.679	1,225		PS		.067
7	.772	932	PS		.031	.016

Note— μ , mean; M , median; PI, perceptual independence; PS, perceptual separability; DS, decisional separability.

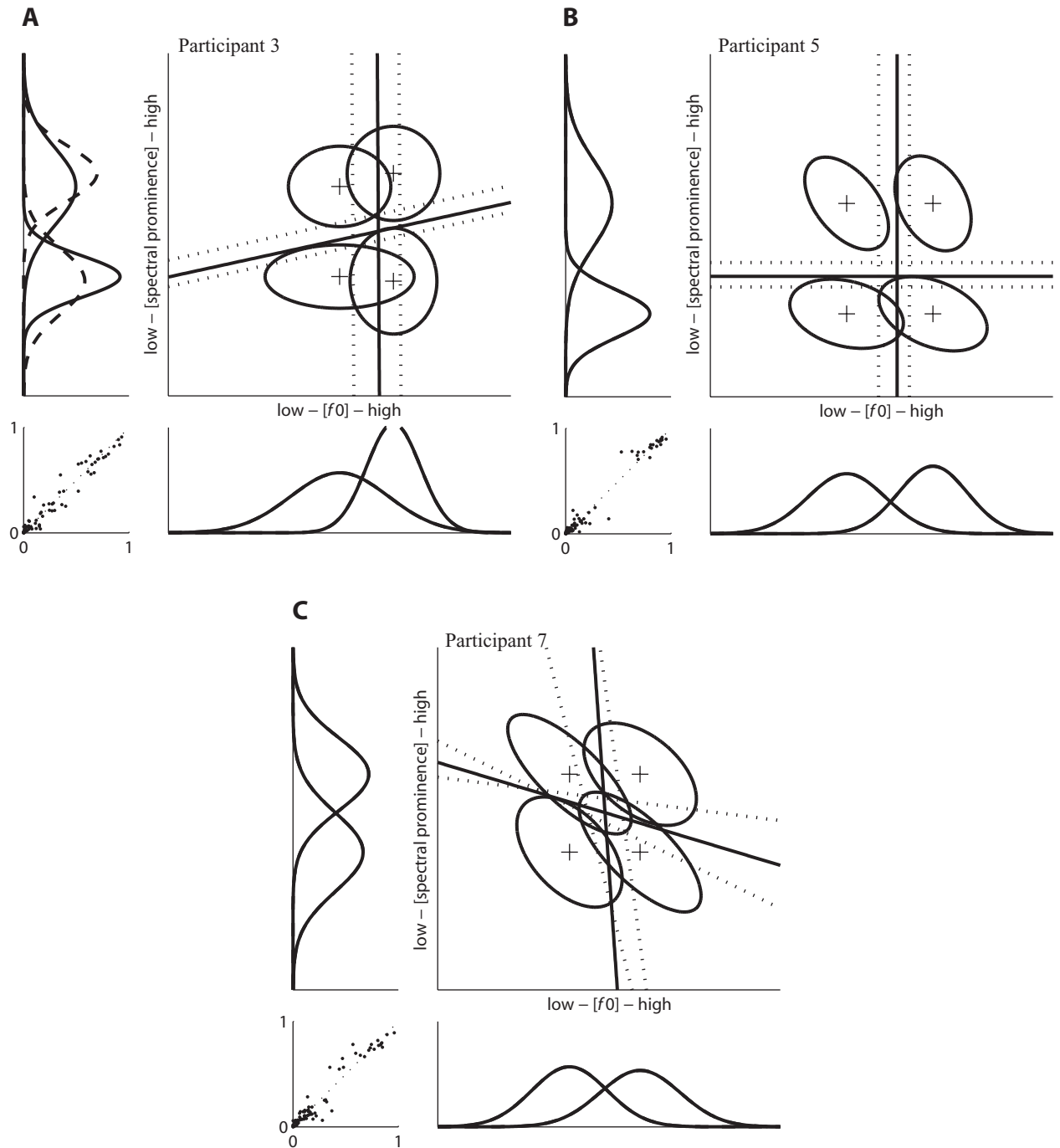


Figure 6. Panels A, B, and C depict the best-fitting Gaussian GRT models for Participants 3, 5, and 7, respectively. The upper right portion of each panel shows decision bounds and equal likelihood contours of fitted perceptual distributions. Plus signs (+) indicate distribution means, the x -axis represents the perceptual effect of f_0 , and the y -axis represents the perceptual effect of spectral shape. The top left and bottom right portions of each panel show marginal perceptual distributions. Solid lines indicate marginal distributions for Level 1 (low frequency or short duration) on the other dimension; dashed lines indicate marginal distributions for Level 2 (high frequency or long duration) on the other dimension. The bottom left portion of each panel shows predicted (y -axis) \times observed (x -axis) identification confusion probabilities.

Hirsh's (1992) finding that upward (downward) shifts in spectral locus could override downward (upward) shifts in f_0 . It is also interesting that Singh and Hirsh found some evidence of individual differences in the interactions of

pitch and timbre. However, because of differences in task demands and the structure of the stimuli, care should be taken in drawing any direct parallels between the interlistener variability in their study and ours.

It is clear that f_0 and spectral shape tend to interact, although it is not entirely clear what mechanisms underlie these interactions. Fundamental frequency determines the spacing between the components of the harmonic complexes, the period of the signal, and, in the stimuli employed here, the frequency of the first (i.e., lowest) component tone. Insofar as perceived pitch is based on f_0 , then, it should be based on something like place on the basilar membrane, phase locking of neural impulses to the waveform, or both. Spectral shape, on the other hand, is defined by a shift in the relative amplitude of a subset of the pure tone components. Although a change in spectral shape affects the shape of the waveform and the relative excitation of different locations on the basilar membrane, it cannot change the periodicity of the signal or the spacing between harmonic components. Our findings are also consistent with work showing that individual components of complex tones can play a significant role in the perception of pitch (e.g., Moore, Glasberg, & Peters, 1985; Plomp, 1967).

It seems likely that the locus of interactive effects between f_0 and spectral shape is, in some respect, postsensory. Clearly, this is a reasonable interpretation of failures of DS. On the other hand, failures of PI can be interpreted as correlated noise between processing channels (here, channels for f_0 and spectral shape). It is less obvious that this should be anything other than a sensory effect; the present results suggest that it is. Furthermore, the presence of some individual differences in Experiment 2 suggests the possibility of differing strategies across participants, also likely a postsensory phenomenon.

Finally, as with the results of Experiment 1, we might expect that linguistic sounds varying on similar dimensions should produce similar behavioral effects. Vowels are distinguished, in part, by their spectral qualities, and languages make varying use of f_0 . In addition, low f_0 should be at least weakly correlated with formant frequencies, since vocal tract size, which delimits the range of possible formant frequencies, is related to vocal fold size, which largely determines f_0 .

GENERAL DISCUSSION

Although all sounds are multidimensional, research investigating relationships among auditory dimensions is relatively rare. Most such research suffers from methodological limitations and the absence of an appropriate theoretical framework for analyzing dimensional interactions. GRT overcomes a number of these limitations by providing a framework within which interactions between dimensions in perception and decision making can be investigated rigorously and quantitatively. Manipulation of stimulus presentation base rates, parameter estimation, and model comparison has enabled the present work to extend the GRT framework substantially.

The frequency \times duration results obtained here are consistent with related research. Espinoza-Varas and Jamieson (1984) found that category boundaries for both duration and frequency were independent of the level on the other, irrelevant dimension. Similarly, our findings of

PI, PS, and DS all point to similar conclusions. The model-fitting analyses also provide a more detailed picture than previously was available. In addition to finding that decision bounds for each dimension were independent of the other dimension, we found strong support for both PS and PI. The perceptual effects of frequency and duration do not, in general, appear to influence one another in perception or decision making.

These findings are also consistent with lower level models of sensory processing. In such models, frequency perception is often modeled as a function of location on the basilar membrane, phase locking of neural firing with the signal, or both, neither of which plausibly play a role in duration encoding, which has been modeled with some success as a counting process (e.g., neural pulses; Creelman, 1962).

The $f_0 \times$ spectral shape results obtained here were in general agreement with related research, although there are some intriguing differences between the present research and previous work. In general, it appears that f_0 and spectral shape interact, although the degree and manner of interaction may vary somewhat across individual participants. Although previous work has employed analytic frameworks that conflate logically distinct levels at which interaction may occur (e.g., the Garner paradigm; see, e.g., Melara & Marks, 1990a), the present findings provide, again, a more detailed picture.

In some cases, the interactions suggest the presence of correlated noise in separate processing channels; 3 participants' best-fitting models indicate failure of PI. In each case, this failure appears to be due to (negative) covariance within perceptual distributions. Six participants' best-fitting models indicate failures of DS, most of which enlarged the regions for low f_0 –high spectral prominence and high f_0 –low spectral prominence responses.

It is possible that perceptual interactions arise due to a cochlear interaction between component spacing and relative amplitude of components, although this seems unlikely given that, for the relevant range of frequencies, the space between components (150 or 152 Hz) is greater than the bandwidth of the auditory filter around the frequency of the spectral prominence (~ 132 Hz at 1000 Hz). It may be that stimuli with components that are closer together would produce a different pattern of results. However, models of sensory processing of f_0 and spectral shape (e.g., Meddis & O'Mard, 1997) suggest that the interactive effects must be caused by higher level (i.e., postsensory) processing, since f_0 is determined by the spacing between components and the periodicity of the waveform in a harmonic complex, whereas the spectral shape is determined by the relative amplitude of the components.

It is important to keep in mind that GRT is intended as a model of asymptotic perceptual and decisional behavior. Although the participants in each experiment received some training prior to data collection, it may be that behavior had not yet completely stabilized during the earliest experiment blocks. If so, we would expect additional perceptual and decisional (i.e., criterial) noise. Such perceptual noise would reduce accuracy and be

indistinguishable from the other sources of noise in the experiments (e.g., externally added white noise and internal neural noise), here modeled as bivariate Gaussian distributions. Decisional noise, if present, would likely be subsumed by perceptual noise in the GRT framework; Ashby (2000) argued that, in general, the perceptual noise in GRT is contaminated by decisional noise. Given that the overall model fits were quite good, however, we think that unstable behavior was a minor problem at worst, if, in fact, it was present at all. In any case, it is not obvious how, or if, behavioral instability would change patterns of interaction and independence, as opposed to simply increasing or reducing accuracy. It would be fairly straightforward to probe the effects of training on these patterns in future work.

The results of both experiments inform our expectations about auditory processing in speech perception. To the extent that the dimensions probed here play a role in distinguishing linguistically relevant speech sounds (e.g., the spectral properties and duration of noise in fricatives; f_0 and spectral shape of vowels) from one another, we should expect to see similar patterns in identification confusion data generated in speech perception experiments.

However, two factors limit the applicability of the present findings to speech perception. First, speech sounds often vary on many more than two dimensions, so the relationship between the present findings and speech perception is not completely straightforward. Second, the differences across the levels of frequency range, duration, and f_0 employed here are much smaller than analogous differences in spectral shape and duration in fricatives, although the difference between the two spectral shape levels is on par with at least some analogous differences between vowels. It is possible that small differences along a given dimension are processed in a qualitatively different manner than are larger differences along the same dimension, although it is an empirical matter as to whether this is the case. Current work in our research group is employing the experimental and analytic tools developed here to the auditory perception of (English) consonants.

In conclusion, we have presented a powerful tool for analyzing relationships between perceptual and decisional dimensions. These tools address a variety of methodological and theoretical limitations common in multidimensional perception research, and our application of these tools to auditory perception extends work on each of two sets of dimensions in complex nonspeech sound. First, we have found strong evidence that frequency and duration are processed independently, both decisionally and, at multiple levels, perceptually. Second, we have found evidence that f_0 and spectral shape (often referred to as “pitch” and “timbre” in related literature) may interact in perception, decision making, or both, and that the nature and degree of interaction varies somewhat from participant to participant. Although, by necessity, our approach has its own limitations, we believe that it points the way to productive avenues of future research.

AUTHOR NOTE

This research was supported by NIH Grants R01-MH0577-17-07A1 and T32 MH019879-15. We thank Lynn Nygaard, Jerry Balakrishnan, and two anonymous reviewers for their comments and criticisms. Address correspondence to N. H. Silbert, Indiana University, 1101 E. 10th Street, Bloomington, IN 47405 (e-mail: nosilber@indiana.edu).

REFERENCES

- ABEL, S. M. (1972). Duration discrimination of noise and tone bursts. *Journal of the Acoustical Society of America*, **51**, 1219-1223.
- ASHBY, F. G. (2000). A stochastic version of general recognition theory. *Journal of Mathematical Psychology*, **44**, 310-329.
- ASHBY, F. G., & GOTT, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 33-53.
- ASHBY, F. G., & LEE, W. W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General*, **120**, 150-172.
- ASHBY, F. G., & TOWNSEND, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, **93**, 154-179.
- CHRISTENSEN, L. A., & HUMES, L. E. (1996). Identification of multidimensional complex sounds having parallel dimension structure. *Journal of the Acoustical Society of America*, **99**, 2307-2315.
- CHRISTENSEN, L. A., & HUMES, L. E. (1997). Identification of multidimensional stimuli containing speech cues and the effects of training. *Journal of the Acoustical Society of America*, **102**, 2297-2310.
- CORCORAN, D. W. (1967). Perceptual independence and recognition of two-dimensional auditory stimuli. *Journal of the Acoustical Society of America*, **42**, 139-142.
- CREELMAN, C. D. (1962). Human discrimination of auditory duration. *Journal of the Acoustical Society of America*, **34**, 582-593.
- DECARLO, L. T. (2002). Signal detection theory with finite mixture distributions: Theoretical developments with applications to recognition memory. *Psychological Review*, **109**, 710-721.
- EDDINS, D. A., HALL, J. W., III, & GROSE, J. H. (1992). The detection of temporal gaps as a function of frequency region and absolute noise bandwidth. *Journal of the Acoustical Society of America*, **91**, 1069-1077.
- EIMAS, P. D., TARTTER, V. C., MILLER, J. L., & KEUTHEN, N. J. (1978). Asymmetric dependencies in processing phonetic features. *Perception & Psychophysics*, **23**, 12-20.
- ESPINOZA-VARAS, B., & JAMIESON, D. G. (1984). Integration of spectral and temporal cues separated in time and frequency. *Journal of the Acoustical Society of America*, **76**, 732-738.
- FLORENTINE, M., FASTL, H., & BUUS, S. (1988). Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking. *Journal of the Acoustical Society of America*, **84**, 195-203.
- GARNER, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.
- GREEN, D. M., & FORREST, T. G. (1989). Temporal gaps in noise and sinusoids. *Journal of the Acoustical Society of America*, **86**, 961-970.
- GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- HOLT, L. L., & LOTTO, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America*, **119**, 3059-3071.
- KADLEC, H., & TOWNSEND, J. T. (1992a). Implications of marginal and conditional detection parameters for the separabilities and independence of perceptual dimensions. *Journal of Mathematical Psychology*, **36**, 325-374.
- KADLEC, H., & TOWNSEND, J. T. (1992b). Signal detection analyses of dimensional interactions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 181-227). Hillsdale, NJ: Erlbaum.
- KASS, R. E., & RAFTERY, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, **90**, 773-795.
- KIDD, G., JR., & MASON, C. R. (1992). A new technique for measuring spectral shape discrimination. *Journal of the Acoustical Society of America*, **91**, 2855-2864.
- KINGSTON, J., DIEHL, R. L., KIRK, C. J., & CASTLEMAN, W. A. (2008).

- On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics*, **36**, 28-54.
- KINGSTON, J., & 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. *Journal of the Acoustical Society of America*, **97**, 1261-1285.
- KINGSTON, J., MACMILLAN, N. A., WALSH-DICKEY, L. W., THORBURN, R., & BARTELS, C. (1997). Integrality in the perception of tongue root position and voice quality in vowels. *Journal of the Acoustical Society of America*, **101**, 1696-1709.
- KRUMHANSL, C. L., & IVERSON, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 739-751.
- MACMILLAN, N. A., KINGSTON, J., THORBURN, R., WALSH-DICKEY, L. W., & BARTELS, C. (1999). Integrality of nasalization and F1: II. Basic sensitivity and phonetic labeling measure distinct sensory and decision-related interactions. *Journal of the Acoustical Society of America*, **106**, 2913-2932.
- 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. (1995). Base-rate effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 288-301.
- MADDOX, W. T., & BOHIL, C. J. (1998a). Base-rate and payoff effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1459-1482.
- MADDOX, W. T., & BOHIL, C. J. (1998b). Overestimation of base-rate differences in complex perceptual categories. *Perception & Psychophysics*, **60**, 575-592.
- MADDOX, W. T., & BOHIL, C. J. (2003). A theoretical framework for understanding the effects of simultaneous base-rate and payoff manipulations on decision criterion learning in perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 307-320.
- MASSARO, D. W., & ODEN, G. C. (1980). Evaluation and integration of acoustic features in speech perception. *Journal of the Acoustical Society of America*, **67**, 996-1013.
- MEDDIS, R., & O'MARD, L. (1997). A unitary model of pitch perception. *Journal of the Acoustical Society of America*, **102**, 1811-1820.
- MELARA, R. D., & MARKS, L. E. (1990a). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception & Psychophysics*, **48**, 169-178.
- MELARA, R. D., & MARKS, L. E. (1990b). Perceptual primacy of dimensions: Support for a model of dimensional interaction. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 398-414.
- MIRMAN, D., HOLT, L. L., & MCCLELLAND, J. L. (2004). Categorization and discrimination of nonspeech sounds: Differences between steady-state and rapidly changing acoustic cues. *Journal of the Acoustical Society of America*, **116**, 1198-1207.
- MOORE, B. C. J., GLASBERG, B. R., & PETERS, R. W. (1985). Relative dominance of individual partials in determining the pitch of complex tones. *Journal of the Acoustical Society of America*, **77**, 1853-1860.
- NEAREY, T. (1992). Context effects in a double-weak theory of speech perception. *Language & Speech*, **35**, 153-171.
- ODEN, G. C., & MASSARO, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, **85**, 172-191.
- OLZAK, L. A. (1986). Widely separated spatial frequencies: Mechanism interactions. *Vision Research*, **26**, 1143-1153.
- OLZAK, L. A., & WICKENS, T. D. (1997). Discrimination of complex patterns: Orientation information is integrated across spatial scale; spatial-frequency and contrast information are not. *Perception*, **26**, 1101-1120.
- PASTORE, R. E., & SORKIN, R. D. (1972). Simultaneous two-channel signal detection: I. Simple binaural stimuli. *Journal of the Acoustical Society of America*, **51**, 544-551.
- PITT, M. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 976-986.
- PLOMP, R. (1967). Pitch of complex tones. *Journal of the Acoustical Society of America*, **41**, 1526-1533.
- PLOMP, R., & BOUMAN, M. A. (1959). Relation between hearing threshold and duration for tone pulses. *Journal of the Acoustical Society of America*, **31**, 749-758.
- RICHARDS, V. M., & LENTZ, J. J. (1998). Sensitivity to changes in level and envelope patterns across frequency. *Journal of the Acoustical Society of America*, **104**, 3019-3029.
- SEMAL, C., & DEMANY, L. (1991). Dissociation of pitch from timbre in auditory short-term memory. *Journal of the Acoustical Society of America*, **89**, 2404-2410.
- SHAILER, M. J., & MOORE, B. C. J. (1983). Gap detection as a function of frequency, bandwidth, and level. *Journal of the Acoustical Society of America*, **74**, 467-473.
- SHAILER, M. J., & MOORE, B. C. J. (1985). Detection of temporal gaps in bandlimited noise: Effects of variations in bandwidth and signal-to-masker ratio. *Journal of the Acoustical Society of America*, **77**, 635-639.
- SHAILER, M. J., & MOORE, B. C. J. (1987). Gap detection and the auditory filter: Phase effects using sinusoidal stimuli. *Journal of the Acoustical Society of America*, **81**, 1110-1117.
- SILBERT, N., & DE JONG, K. (2008). Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production. *Journal of the Acoustical Society of America*, **123**, 2769-2779.
- SINGH, P. G., & HIRSH, I. J. (1992). Influence of spectral locus and f_0 changes on the pitch and timbre of complex tones. *Journal of the Acoustical Society of America*, **92**, 2650-2661.
- SMITS, R. (2001). Evidence for hierarchical categorization of coarticulated phonemes. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 1145-1162.
- SMITS, R., SERENO, J., & JONGMAN, A. (2006). Categorization of sounds. *Journal of Experimental Psychology: Human Perception & Performance*, **32**, 733-754.
- SORKIN, R. D., PASTORE, R. E., & POHLMANN, L. D. (1972). Simultaneous two-channel signal detection: II. Correlated and uncorrelated signals. *Journal of the Acoustical Society of America*, **51**, 1960-1965.
- SORKIN, R. D., POHLMANN, L. D., & GILLIOM, J. D. (1973). Simultaneous two-channel signal detection: III. 630- and 1400-Hz signals. *Journal of the Acoustical Society of America*, **53**, 1045-1050.
- STEVENS, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.
- TANNER, W. P. (1956). Theory of recognition. *Journal of the Acoustical Society of America*, **28**, 882-888.
- THOMAS, R. D. (1995). Gaussian general recognition theory and perceptual independence. *Psychological Review*, **102**, 192-200.
- THOMAS, R. D. (1996). Separability and independence of dimensions within the same-different judgment task. *Journal of Mathematical Psychology*, **40**, 318-341.
- THOMAS, R. D. (1999). Assessing sensitivity in a multidimensional space: Some problems and a definition of a general d' . *Psychonomic Bulletin & Review*, **6**, 224-238.
- THOMAS, R. D. (2001a). Characterizing perceptual interactions in face identification using multidimensional signal detection theory. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges* (pp. 193-227). Mahwah, NJ: Erlbaum.
- THOMAS, R. D. (2001b). Perceptual interactions of facial dimensions in speeded classification and identification. *Perception & Psychophysics*, **63**, 625-650.
- THOMAS, R. D. (2003). Further considerations of a general d' in multidimensional space. *Journal of Mathematical Psychology*, **47**, 220-224.
- WANG, X., & HUMES, L. E. (2008). Classification and cue weighting of multidimensional stimuli with speech-like cues for young normal hearing and elderly hearing-impaired listeners. *Ear & Hearing*, **29**, 725-745. doi:10.1097/AUD.0b013e31817bdd42
- WARRIER, C. M., & ZATORRE, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception & Psychophysics*, **64**, 198-207.
- WICKENS, T. D. (1992). Maximum-likelihood estimation of a multivariate Gaussian rating model with excluded data. *Journal of Mathematical Psychology*, **36**, 213-234.
- WICKENS, T. D., & OLZAK, L. A. (1989). The statistical analysis of concurrent detection ratings. *Perception & Psychophysics*, **45**, 514-528.
- ZAGORSKI, M. (1975). Perceptual independence of pitch and loudness in

a signal detection experiment: A processing model for 2ATFC (21FC) experiments. *Perception & Psychophysics*, **17**, 525-531.

NOTES

1. The distributions in SDT and GRT need not, in general, be considered perceptual (e.g., SDT has been profitably applied to memory research in which no plausible role for perceptual distributions exists; e.g., DeCarlo, 2002). Throughout the present work, however, distributions are interpreted as perceptual in nature.

2. By definition, if DS holds for a given dimension, the decision bound is a line parallel to a coordinate axis (e.g., for the bound partitioning the y -axis, the bound is a line parallel to the x -axis). There are a number of ways that DS could fail. Here, a linear bound with nonzero slope illustrates one possible way. Only linear bounds are considered here.

(Manuscript received September 26, 2008;
revision accepted for publication July 12, 2009.)