Categorization and discrimination of nonspeech sounds: Differences between steady-state and rapidly-changing acoustic cues[a)]

Daniel Mirman,[b)] Lori L. Holt, and James L. McClelland
Department of Psychology and Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

(Received 13 February 2003; revised 9 April 2004; accepted 6 May 2004)

Different patterns of performance across vowels and consonants in tests of categorization and discrimination indicate that vowels tend to be perceived more continuously, or less categorically, than consonants. The present experiments examined whether analogous differences in perception would arise in nonspeech sounds that share critical transient acoustic cues of consonants and steady-state spectral cues of simplified synthetic vowels. Listeners were trained to categorize novel nonspeech sounds varying along a continuum defined by a steady-state cue, a rapidly-changing cue, or both cues. Listeners' categorization of stimuli varying on the rapidly changing cue showed a sharp category boundary and posttraining discrimination was well predicted from the assumption of categorical perception. Listeners more accurately discriminated but less accurately categorized steady-state nonspeech stimuli. When listeners categorized stimuli defined by both rapidly-changing and steady-state cues, discrimination performance was accurate and the categorization function exhibited a sharp boundary. These data are similar to those found in experiments with dynamic vowels, which are defined by both steady-state and rapidly-changing acoustic cues. A general account for the speech and nonspeech patterns is proposed based on the supposition that the perceptual trace of rapidly-changing sounds decays faster than the trace of steady-state sounds.
 [DOI: 10.1121/1.1766020]

PACS numbers: 43.71.Es, 43.71.Pc, 43.66.Ba [PFA] Pages: 1198–1207

I. INTRODUCTION

Patterns of performance in categorization and discrimination tasks differ across classes of speech sounds. Discrimination of stop consonants is closely predicted by categorization (Liberman *et al.*, 1957), but discrimination of vowels and fricatives exceeds categorization-based predictions (Eimas, 1963; Pisoni, 1973; Healy and Repp, 1982). We hypothesize that the differences in categorization and discrimination patterns arise as a result of differences in the way the auditory system processes the differing acoustic cues that distinguish vowels and consonants. Specifically, we suggest that the rapid transients characteristic of many consonants are processed quite differently than the relatively steady-state frequency information that characterizes steady-state vowel and fricative stimuli. From this hypothesis, we predict that nonspeech sounds that are defined by acoustic cues that reflect these differences will elicit the same patterns of categorization and discrimination performance as stop consonants and synthetic steady-state vowels. The experiments described in this report test this prediction by training listeners to categorize nonspeech sounds that vary along a rapidly-changing cue, a steady-state cue, or both types of cues and then examining categorization and discrimination of the sounds. Before turning to the experiments, we discuss in more detail the evidence for the points motivating our experiments.

A. Categorization and discrimination of different classes of speech sounds

Differences in categorization and discrimination of different classes of speech sounds can be analyzed by comparing observed discrimination performance to discrimination performance predicted from categorization. To predict discrimination from categorization, a discrimination curve is calculated based on the assumption that the listener makes discrimination judgments based entirely on whether the two stimuli are categorized as the same sound or different sounds. Stop consonants elicit sharp categorization functions and discrimination performance is accurately predicted by categorization, a pattern known as categorical perception (Liberman *et al.*, 1957; Wood, 1976; Repp, 1984). The categorization functions elicited by steady-state vowels and fricatives are less sharp and discrimination performance is much more accurate than predicted from categorization (Eimas, 1963; Pisoni, 1973; Healy and Repp, 1982). This result indicates that, at least with steady-state vowels, listeners are not merely using category labels to perform discrimination (e.g., Pisoni, 1971).

Some investigators (Ades, 1977; Healy and Repp, 1982) have attempted to explain differences in patterns of categorization and discrimination performance between steady-state speech sounds (vowels and fricatives) and rapidly-changing speech sounds (stop consonants) in terms of differences in auditory distinctiveness. Distinctiveness is considered a function of perceptual range, which is measured by the sum of the d' between adjacent stimuli (Ades, 1977).

[a)]A preliminary report on this work was presented at the 143rd Meeting of the Acoustical Society of America, June 2002.
[b)]Electronic mail: dmirman@andrew.cmu.edu

This account predicts a direct trade-off between discrimination performance (i.e., auditory distinctiveness) and categorization performance. However, Macmillan *et al.* (1988) controlled for perceptual range and found that differences between vowels and consonants remained. Thus, factors other than perceptual range must contribute to the differences in categorization and discrimination of consonants and vowels.

Steady-state vowels are a simplified approximation of natural vowels, which are additionally specified by dynamic acoustic information (Gottfried and Strange, 1980; Strange *et al.*, 1976). Direct comparisons for 12 vowels of American English indicate that steady-state formants are sufficient for approximately 75% correct vowel identification, but when synthetic formants follow natural formant contours, correct identification is improved to nearly 90% (Hillenbrand and Nearey, 1999). Experiments using stimuli based on natural vowels, which vary along both steady-state and rapidly-changing acoustic cues (Schouten and van Hessen, 1992), have shown a pattern that is not consistent with the predictions of the distinctiveness account of Ades (1977) and Healy and Repp (1982). In these experiments, categorization of dynamic vowels exhibited steep category boundaries (like stop consonants) but discrimination was high and exceeded categorization-based predictions (like steady-state vowels).

The cues that distinguish stop consonants are different from the cues that distinguish steady-state vowels; furthermore, natural, dynamic vowels are defined by a combination of cues (the importance of rapidly-changing cues to vowel identity may vary by vowel; e.g., Hillenbrand and Nearey, 1999). The acoustic patterns of stop consonants can be broadly defined by rapidly-changing acoustic cues. Stop consonants are primarily distinguished by rapid formant transitions and fine temporal distinctions such as voice onset time. In contrast, the acoustic patterns of vowels and fricatives can be broadly defined by steady-state acoustic cues. In particular, the synthetic steady-state vowels that are often used in studies of speech perception are distinguished only by formant center frequencies that remain constant for the duration of the sound. Fricatives, too, are primarily defined by relatively slow-varying acoustic properties (e.g., Jongman *et al.*, 2000). Thus, one possibility is that differences in patterns of categorization and discrimination between stop consonants and steady-state vowels and fricatives arise from general differences between processing rapidly-changing and steady-state acoustic cues.

B. Processing differences between rapidly-changing and steady-state sounds

There is considerable support for the broad distinction between steady-state and rapidly-changing sounds and the supposition that the auditory system processes such sounds differently. Specifically, processing of rapidly-changing sounds is more left-lateralized than processing of steady-state sounds. This result has been found in comparisons of human perception across classes of speech sounds (Cutting, 1974; Allard and Scott, 1975) and in nonhuman primate perception of conspecific calls (Heffner and Heffner, 1984; Hauser and Andersson, 1994). Further, it has been found that processing is more left-lateralized in humans when the formant transition durations are extended in speech sounds (Schwartz and Tallal, 1980) and in nonspeech sounds (Belin *et al.*, 1998). The same result has been demonstrated for nonhuman primates (Hauser *et al.*, 1998). In addition, recent evidence from patterns of correlation in learning to categorize based on different kinds of cues (Golestani *et al.*, 2002) suggests that steady-state and rapidly-changing cues rely on distinct processing mechanisms. The close similarity of lateralization results for speech and nonspeech sounds and for humans and nonhuman primates, as well as the correlations in learning rates, suggest that the auditory system processes rapidly-changing and steady-state sounds differently.

Poeppel (2003; see also Zatorre *et al.*, 2002) has proposed that different temporal integration windows in the left and right nonprimary auditory cortices account for these findings. In particular, Poeppel contends that left nonprimary auditory cortical processing depends on a short temporal integration window (20–40 ms) but right nonprimary auditory cortical processing depends on a longer temporal integration window (150–300 ms). Thus, processing rapidly-changing cues, requiring a shorter temporal integration window, is performed primarily by the left hemisphere. By contrast, analysis of slower-changing cues is performed by the right hemisphere with a longer temporal integration window, thus allowing greater spectral resolution. A similar proposal has been made by Shamma (2000), who argues that acoustic signals are represented at multiple time scales. In particular, rapidly changing sounds, such as plucked instruments and stop consonants, are represented on a fast time scale, but steady-state sounds, such as bowed instruments and vowels, are represented on a slow time scale.

In sum, there is considerable evidence indicating that rapidly-changing and steady-state acoustic cues are processed differently by the auditory system. Furthermore, patterns indicating this difference appear to be quite general, occurring in perception of speech and nonspeech sounds. The left hemisphere advantage emerges for both speech and nonspeech sounds that are defined by rapidly-changing cues, but not for sounds defined by steady-state cues. Similarly, the canonical categorical perception pattern of categorization and discrimination performance (specifically the accurate prediction of discrimination performance from categorization, as discussed above) emerges for stop consonants that are defined by rapidly-changing acoustic cues but not for vowels that are defined by steady-state cues. In the present experiments, we test whether novel nonspeech sounds that are defined by rapidly-changing cues will exhibit a categorical perceptionlike pattern. By comparison, we test whether nonspeech sounds defined by steady-state spectral cues will exhibit the pattern typically observed for synthetic steady-state voweis and fricatives.

II. EXPERIMENTS

The following experiments were designed to test categorization and discrimination of novel nonspeech stimuli. Each of the experiments employed a similar training and testing procedure. The key analyses were posttraining categorization and discrimination performance. Categorization posttest re-

sults were used to generate a ''predicted'' discrimination curve following signal detection theory (Macmillan and Creelman, 1991; Macmillan, 1987). The predicted discrimination curve was computed for each participant based on the hypothesis that participants make same-different discrimination judgments by considering whether the sounds belong to the same category or different categories. Thus, predicted discrimination d' for each pair of stimuli was the difference in (z-transformed) likelihood that the participant would respond that each pair member belongs to the same category. Sharp categorization functions mean that sounds are grouped into discrete categories. Thus, sharp categorization functions predict poor within-category discrimination (since all stimuli within a category consistently receive the same label) and good discrimination across the category boundary (since stimuli across the boundary consistently receive different labels). In contrast, less sharp categorization functions predict moderate discrimination across the entire stimulus series (since all stimuli are partly ambiguous and thus any pair will sometimes receive the same label, and sometimes receive different labels).

The specific research prediction was that rapidly-changing nonspeech sounds would elicit sharp categorization functions and poor within-category discrimination performance compared to discrimination across category boundary. That is, for rapidly-changing sounds categorization performance would accurately predict discrimination performance. In contrast, steady-state nonspeech sounds would elicit less sharp categorization functions, but good discrimination performance at every point on the series. That is, for steady-state sounds discrimination performance would exceed categorization-based predictions.

A. Stimulus space

The stimuli forming categories learned by listeners were drawn from a novel two-dimensional acoustic space. The acoustic space was defined in one dimension by a rapidly-changing amplitude envelope cue and in the other dimension by a steady-state spectral cue to allow independent manipulation of the cues. The non-speech cues were chosen to be generally similar to cues that are manipulated in studies of speech perception. The steady-state spectral cue was held constant throughout the stimulus, analogous to steady-state vowels (e.g., Pisoni, 1973). The rapidly-changing cue was analogous to amplitude rise time, which plays a role in distinctions between classes of consonants (Van Tasell *et al.*, 1987), for example, the stop-glide contrast (e.g., /b/-/w/; Mack and Blumstein, 1983; Walsh and Diehl, 1991). This cue also has been investigated in the context of the nonspeech pluck-bow distinction (Cutting, 1982; Kewley-Port and Pisoni, 1984).

Each stimulus was composed of a 300-ms burst of white noise (10-kHz sample rate) with two 200-Hz bands of energy removed by 50-dB elliptic bandstop filters. This filtering process created two spectral notches characterized by their center frequencies. The center frequencies of the filters used to create the spectral notches remained constant across the entire stimulus duration, but differed from stimulus to stimulus to create a series that varied along a steady-state spectral

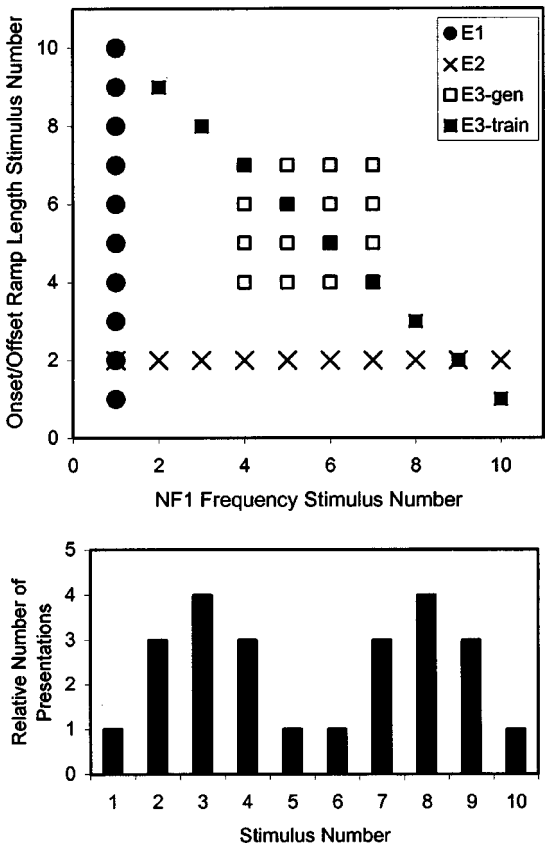

FIG. 1. Top panel: schematic representation of the sampling of the stimulus. The circles are stimuli that vary in ramp length (experiment 1), the crosses are stimuli that vary in NF1 frequency (experiment 2), and the squares are stimuli that vary in both ramp length and frequency (experiment 3; filled squares are standard training and testing stimuli, open squares are generalization testing stimuli). Bottom panel: Relative frequency of presentation of stimuli during categorization training.

dimension. In the experiments presented here, the first notch center frequency (NF1) started at 500 Hz and increased in equal steps (see experiment procedure below). The second notch center frequency (NF2) was fixed for all stimuli at 2500 Hz. This procedure is similar to the typical procedure for manipulating formant frequency to create a steady-state vowel series (e.g., Miller 1953; Hoemeke and Diehl, 1994). Finally, a symmetric linear onset and offset ramp was applied (as in the pluck-bow experiments, e.g., Kewley-Port and Pisoni, 1984). The duration of this ramp was manipulated to create a series distinguished by a rapidly-changing cue. Although the cues that distinguish these stimuli are abstractly similar to cues that distinguish speech sounds, these stimuli were perceived as bursts of noise and not as speech.

Figure 1 (top panel) is a schematic depiction of the sampling of this stimulus space in the following experiments. The axis labels represent generic steps along the series because step size was adjusted based on individual participants' sensitivity to make the steps approximately equally discriminable across listeners (see procedure for details). The horizontal axis represents steps along the NF1 series. The vertical axis represents steps along the ramp length series. Each stimulus series consisted of ten stimuli divided into two equal categories, with the category boundary (defined by ex-

plicit feedback during training) between stimulus 5 and stimulus 6 in the series.

B. Experiment 1

In the first experiment, the ramp length cue was manipulated to create a single-dimension stimulus series varying along a rapidly-changing cue. Following a sensitivity assessment and pretest, the participants were trained to categorize the stimuli and then were tested on categorization and discrimination.

1. Method

a. Participants. Participants were 16 Carnegie Mellon University undergraduates who had not participated in a previous experiment using stimuli based on these cues. Participants received course credit and/or a small payment. All participants reported normal hearing.

b. Stimuli. The stimuli were synthesized as described above using the MATLAB signal processing toolbox. NF1 and NF2 were fixed at 500 and 2500 Hz, respectively. The duration of the linear onset/offset ramp varied in equal steps starting at 5 ms. For example, with a step size of 15 ms, the first stimulus had symmetric onset and offset ramps of 5 ms, the second stimulus had 20 ms ramps, the third had 35 ms ramps, and so on. The size of the steps was determined by sensitivity assessment for each participant (as described below) so that the experimental stimuli would be approximately equally discriminable to each participant.

c. Procedure. Participants completed the experiment while sitting in sound attenuating booths, using labeled electronic button boxes to make responses. Sensitivity of each participant to the ramp length cue was assessed using a same-different discrimination task. An adaptive staircase procedure, in which the step size was increased if discrimination was not accurate enough and decreased if discrimination was too accurate, was used to identify an appropriate ramp step size. Discrimination performance was assessed on the difference between percent hits and percent false alarms[1] with a target range of 30% to 50%. The 32 ''different'' trials (4 repetitions of 8 pairs) consisted of stimulus pairs two steps apart. In addition, there were ten ''same'' trials for which stimulus pair members were identical. The staircase procedure was constrained to seven possible step sizes: 1, 3, 5, 7, 9, 12, and 15 ms. Stimulus pairs were presented with an interstimulus silent interval of 500 ms. After a block of discrimination trials, the participant's performance was assessed. If the participant was not sensitive enough ([%hits −%false alarms]$<$30), then a more discriminable stimulus set with a larger step size was selected. If the participant was too sensitive ([%hits−%false alarms]$>$50), then a less discriminable stimulus set with a smaller step size was selected. This test and step-size adjustment was repeated three times. The starting step size was 7 ms and all participants reached threshold discrimination at 12 or 15 ms steps. At the final step size, listeners participated in the pretest discrimination task consisting of 110 discrimination trials (80 different trials, 30 same trials).

Next, the participants heard 480 categorization training trials in each trial, one of the stimuli (drawn from the set of 10 separated by a step size determined by sensitivity assessment) was presented and the participants categorized it as belonging to one of two categories by pressing one of two buttons on a response box, labeled with arbitrary symbols. After the categorization response, the participants were shown the correct answer by a light above the correct button. Stimuli presented during training followed a bimodal distribution to reflect exposure to two natural categories (e.g., phonetic categories, Lisker and Abramson, 1964) and encourage category formation (Maye *et al.*, 2002; Rosenthal *et al.*, 2001). Feedback was consistent with the distribution-defined categories (category A: stimuli 1–5, category B: stimuli 6–10). Figure 1 (bottom panel) illustrates the relative frequency with which stimuli were presented during training. Categorization training was divided into two equal units (240 trials each) and separated by a discrimination test identical to the pretest.

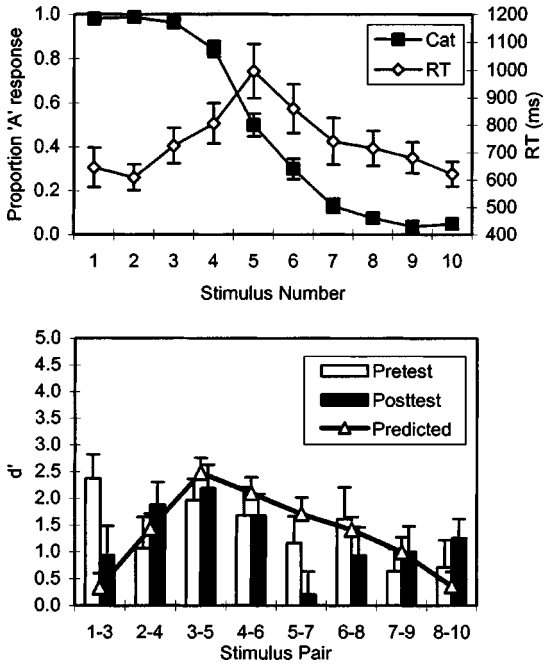

FIG. 2. Experiment 1 results: stimuli varying in ramp length. The top panel shows categorization responses (filled symbols) and reaction times (open symbols). The bottom panel shows discrimination results. Empty bars indicate pretest performance, filled bars indicate posttest performance, and the solid line is performance predicted from categorization according to the signal detection theory model.

After training, the participants completed a discrimination posttest identical to the pretest. Finally, each listener participated in a categorization test consisting of 100 categorization trials (10 trials$\times$10 stimuli) without feedback. Participants completed the experiment during a single 1.5-h session.

2. Results

Figure 2 (top panel) illustrates the average of participants' posttest category responses and corresponding reaction times as a function of stimulus step. Following just 480 training trials, participants learned to assign category labels with high accuracy (87% correct with respect to feedback-defined category labels). Furthermore, reaction times exhib-

ited a pronounced peak at the category boundary confirming that the participants treated the stimuli as belonging to different categories (Pisoni and Tash, 1974; Maddox *et al.*, 1998). Figure 2 (bottom panel) shows the results of the discrimination pretest (empty bars) and posttest (filled bars) as well as the posttest performance predicted from categorization (solid line). There was no change from pretest to posttest and a close correspondence between observed and predicted performance. A repeated measures ANOVA confirmed that there was no overall change from pretest to posttest ($F<1$), a trend towards more accurate discrimination near the center of the series [$F(7,105)=2.029, p=0.058$], and no interaction between location in series and change from pretest to posttest [$F(7,105)=1.492, p=0.178$]. The same test comparing observed and predicted discrimination performance indicated no overall difference between observed and predicted performance ($F<1$), a peak in discrimination accuracy near the center of the series [$F(7,105)=5.169, p<0.001$], and small series-member-specific differences between observed and predicted performance [$F(7,105)=2.114, p=0.048$]. Posthoc pairwise comparisons confirmed that this interaction was produced by deviations between observed and predicted performance at stimulus pairs 5–7 and 8–10.

3. Discussion

The relatively short training procedure used in this experiment was sufficient for participants to learn to categorize stimuli according to onset/offset ramp length. The high categorization accuracy and reaction time peak support this conclusion. In addition, although there was no evidence for a consistent learning-based change in discrimination performance, the posttest performance did fall very close to performance predicted from categorization. That is, for stimuli varying in length of onset/offset ramp, a rapidly-changing acoustic cue, it appears that discrimination and categorization performance are closely matched.

C. Experiment 2

In the second experiment the training and testing procedure of experiment 1 was replicated, but ramp length was held constant and NF1 was manipulated to create a stimulus series varying along a steady-state cue.

1. Method

a. Participants. Participants were 16 Carnegie Mellon University undergraduates who had not participated in a previous experiment using stimuli based on these cues. Participants received course credit and/or a small payment. All participants reported normal hearing.

b. Stimuli. The stimuli for this experiment were synthesized according to the procedure outlined above. However, in this case, the onset/offset ramps were fixed at 10 ms and NF1 was used as the category membership cue. All stimuli had a notch with center frequency of 2500 Hz (NF2) and another spectral notch (NF1) with a lower center frequency. Stimuli were distinguished by NF1, which started at 500 Hz and increased in center frequency in equal steps, the size of which was determined by sensitivity assessment. For example, with a step size of 50 Hz, the first stimulus in the set would have an NF1 center of 500 Hz, the second stimulus would have 550 Hz, the third 600 Hz, etc. The staircase procedure was constrained to 11 possible step sizes: 3, 5, 10, 15, 20, 25, 30, 40, 50, 60, and 75 Hz.

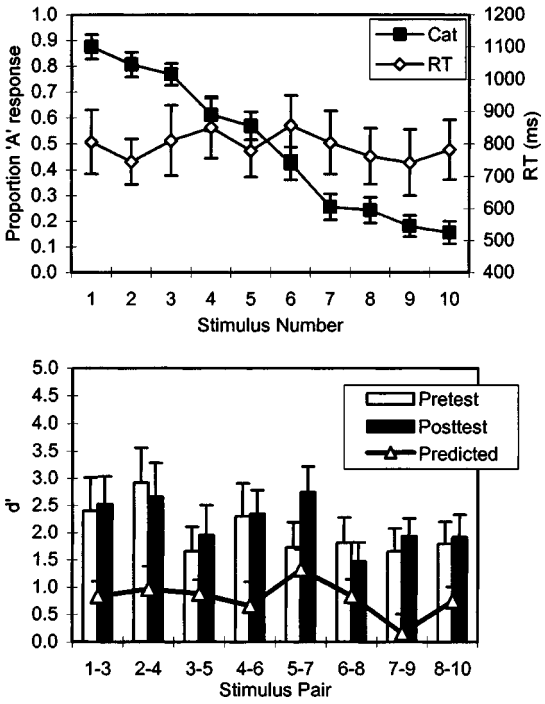

FIG. 3. Experiment 2 results: stimuli varying in NF1. The top panel shows categorization responses (filled symbols) and reaction times (open symbols). The bottom panel shows discrimination results. Empty bars indicate pretest performance, filled bars indicate posttest performance, and the solid line is performance predicted from categorization according to the signal detection theory model.

c. Procedure. The procedure was nearly identical to that of experiment 1. There were two differences in the sensitivity assessment stage. First, pilot studies indicated that sensitivity to this cue was quite variable across participants, thus the number of possible NF1 step sizes was increased to 11 (there were 7 possible ramp step sizes in experiment 1). To accommodate this increase the sensitivity assessment was extended to five blocks (three were used in experiment 1). The initial step size was 25 Hz and listeners' assessed sensitivities included all possible step sizes (3–75 Hz). Second, d' was used as a measure during sensitivity assessment with the target range of 1.5 to 2.5.

2. Results

Figure 3 (top panel) shows the categorization data, which indicate that participants learned to assign category labels with moderate accuracy (73.7% correct) although they did not exhibit a sharp category boundary, nor did they show a reaction time peak at the boundary. Furthermore, the pattern of discrimination results in Fig. 3 (bottom panel) shows that discrimination performance was higher than would be predicted from categorization performance, a pattern that is quite different from the results of experiment 1 (Fig. 2). Repeated measures ANOVA results indicated no change from pretest to posttest ($F<1$), a trend suggesting minor differ-

ences in discriminability across the series [$F(7,105) = 2.019, p = 0.059$], and no series member-specific change from pretest to posttest ($F < 1$). The same test comparing observed and predicted discrimination performance showed an overall difference between observed and predicted performance [$F(1,15) = 17.324, p < 0.001$], no significant differences in performance across the stimulus series [$F(7,105) = 1.845, p = 0.086$], and no stimulus pair-specific differences between observed and predicted performance ($F < 1$).

3. Discussion

Participants learning categories defined by NF1 (experiment 2) did not achieve the same level of accuracy in categorization as the participants learning categories cued by ramp length (experiment 1), despite the categorization training procedures being identical across experiments 1 and 2. By contrast, discrimination performance on stimuli defined by NF1 was quite high. In fact, participants' discrimination performance far exceeded the level that would be predicted from their categorization performance.

One possible explanation for this difference is that varying ramp length allows stimuli to be described as ''gradual'' and ''abrupt,'' but varying NF1 does not lend itself to verbal labels derived from experience outside the experiment. That is, it was easier to label the ramp length stimuli because they were consistent with labels that participants already know, but the NF1 stimuli require learning new labels. However, during postexperiment debriefing participants did not use ''gradual'' and ''abrupt'' to describe the variation in ramp length-based stimuli (there was no consistent response, participants provided such disparate descriptions as masculine/feminine and ''coming towards''/''going away''). Conversely, most participants described the NF1 variation as being ''pitch-like.'' Thus, if participants were using labels other than those specified by the experiment, categorization in experiment 2 should be more accurate (because NF1 variation was consistently heard as variation of a familiar cue, i.e., pitch), but the opposite pattern was observed.

The differences between experiments 1 and 2 could also be explained if the experiment 2 categorization task were more difficult than the experiment 1 task. If this were the case, 480 learning trials may not have been sufficient for listeners to learn categories in experiment 2. Sharper categorization functions and more accurate discrimination predictions may have emerged with more training. To test this possibility, an extended version of experiment 2 was conducted. This experiment used the same basic paradigm, but greatly extended categorization training. Listeners completed seven 1-h sessions (session 1: pretests and initial categorization training, sessions 2–6: categorization training, session 7: final categorization training and posttests). After 6720 categorization training trials (14 times more than experiment 2) listeners ($N = 10$) exhibited identical results: less sharp categorization (71% correct), no reaction time peak, and high discrimination performance exceeding categorization-based prediction. This replication indicates that the differences between results of experiments 1 and 2 are not due to a simple difficulty of learning category labels for the steady-state stimuli.

The pattern of data in experiment 2 is quite different from the sharp categorization and close correspondence between categorization and discrimination performance observed in experiment 1. The main difference between experiments 1 and 2 was that in the latter experiment, the cue that differentiated stimuli and defined their category membership was NF1, a steady-state spectral cue, but in experiment 1 the cue was onset/offset ramp length, a rapidly-changing cue. This pattern is similar to the reported differences in categorization and discrimination of stop consonants compared to steady-state vowels and fricatives and corresponds with studies indicating that the auditory system may process rapidly-changing and steady-state acoustic cues differently. Cue differences may interact with the cognitive processes that underlie categorization and discrimination.

As discussed in the Introduction, direct comparisons of categorization of dynamic and steady-state synthetic vowels have shown that rapidly-changing cues improve vowel identification (Hillenbrand *et al.*, 2001; Hillenbrand and Nearey, 1999). Importantly, this improvement in identification comes without a decrease in discrimination performance (Kewley-Port and Watson, 1994; Kewley-Port, 1995; Kewley-Port and Zheng, 1999). That is, speech sounds that are defined by both steady-state and rapidly-changing cues are categorized according to a sharp boundary and discriminated at levels that exceed categorization-based predictions (Schouten and van Hessen, 1992). In the preceding experiments, we have demonstrated that for nonspeech sounds that are distinguished by a rapidly-changing cue, categorization is sharp and accurately predicts discrimination, but for nonspeech sounds distinguished by a steady-state cue, categorization is less sharp and discrimination exceeds categorization-based prediction. If these results are driven, at leat in part, by differences in the way steady-state and rapidly-changing cues interact with the cognitive processes of categorization and discrimination, then the same sharpening of the categorization function and better-than-predicted discrimination performance should be observed when the nonspeech steady-state and rapidly-changing cues used in the previous experiments are combined. To test this prediction, the procedure used in experiments 1 and 2 was repeated, but the ramp length and NF1 cues were combined such that both the rapidly-changing cue and the steady-state spectral cue were available to participants performing the categorization and discrimination tasks.

D. Experiment 3

1. Method

a. Participants. Participants were 17 Carnegie Mellon University undergraduates who had not participated in a previous experiment using stimuli based on these cues. Participants received course credit and/or a small payment. All participants reported normal hearing.

b. Stimuli. The stimuli for this experiment were generated by combining NF1 and ramp cues. Stimuli differed along both cues such that either cue was sufficient for categorization. Filled square symbols in Fig. 1 (top panel) show an abstract representation of the stimulus space sampling. The ramp step size was fixed at 15 ms, but the NF1 step size

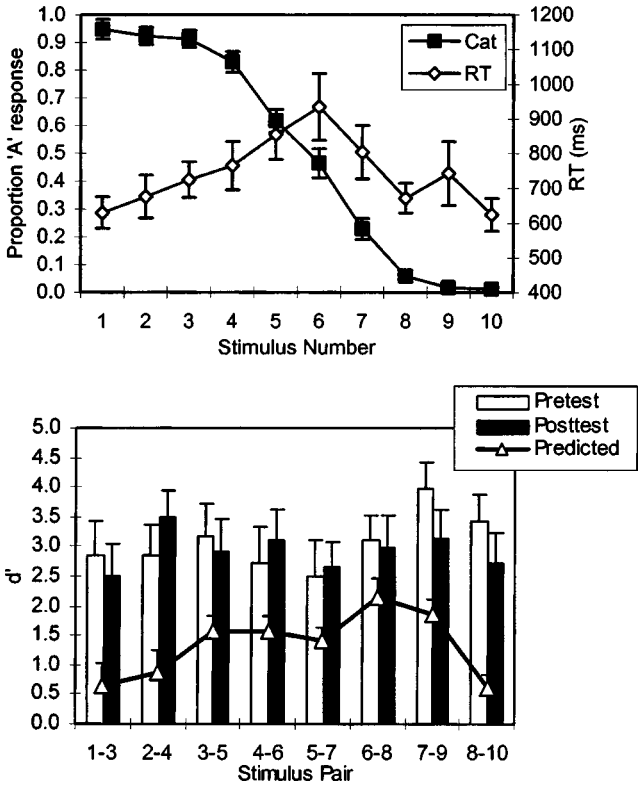

FIG. 4. Experiment 3 results: stimuli varying in both ramp length and NF1. The top panel shows categorization responses (filled symbols) and reaction times (open symbols). The bottom panel shows discrimination results. Empty bars indicate pretest performance, filled bars indicate posttest performance, and the solid line is performance predicted from categorization according to the signal detection theory model.

was determined by sensitivity assessment similar to experiment 2. The sensitivity assessment in experiment 1 and pilot studies resulted in nearly all participants having a 15-ms step size (with a few participants having a 12-ms step size). Therefore, it was assumed that sensitivity to ramp length was sufficiently constant across listeners to make independent sensitivity assessment for each cue unnecessary. In addition to the 10 training stimuli, 12 generalization stimuli from the region near the category boundary were synthesized in order to examine the shape of each participant's category boundary.

c. Procedure. The procedure was nearly identical to that of experiment 1. There were two changes made to the procedure to accommodate the change in stimuli. First, sensitivity assessment consisted of five blocks as in experiment 2 (although the performance criterion was based on the difference between percent hits and percent false alarms, as in experiment 1). Second, a generalization test was added to the end of the experiment. During the generalization test, the 16 stimuli (12 novel stimuli plus the 4 stimuli from the training set that are closest to the boundary) surrounding the boundary area (see Fig. 1, top panel, empty squares) were presented 20 times each without feedback (as in the categorization posttest).

2. Results

Figure 4 (top panel) shows that categorization posttest results were similar to those observed in experiment 1. The categorization function was sharp with high accuracy (84.5% correct) and reaction time exhibited a moderate peak at the category boundary. Regression analysis of the generalization responses with respect to each of the cues revealed significant effects of both cues [NF1: $t(14)=11.943, p<0.001$; ramp: $t(14)=14.011, p<0.001$]. That is, during generalization participants used both cues to make category assignments. The discrimination results shown in Fig. 4 (bottom panel) followed qualitatively the same pattern as observed in experiment 2. Comparison of pretest and posttest discrimination showed no significant differences (all F's <1). As in experiment 2, a comparison of predicted and observed discrimination posttest data showed an overall difference between observed and categorization-predicted performance [$F(1,16)=19.074, p<0.001$], some differences in discriminability across the stimulus series [$F(7,112)=2.178, p=0.041$], and no series member-specific differences between observed and predicted performance [$F(7,112)=1.268, p=0.272$]. Thus, for stimuli defined by both the ramp cue and the NF1 cue, categorization performance was similar to categorization of stimuli defined by just the ramp cue but discrimination performance was similar to discrimination of stimuli defined by just the NF1 cue.

3. Discussion

The sharp categorization function and above-predicted discrimination performance observed in this experiment reflected a "best of both worlds" of the patterns observed in experiments 1 and 2. That is, the combination of both acoustic cues elicited a maximally accurate combination of categorization and discrimination performance. The sharp categorization function and reaction time peak at the boundary were qualitatively similar to the result from experiment 1, in which the ramp length cue was the category-distinguishing cue. The high discrimination performance relative to the prediction from categorization was similar to the results of experiment 2, in which the steady-state NF1 cue was the category-distinguishing cue. These results are similar to the findings of researchers studying vowels with rapidly-changing cues (Schouten and van Hessen, 1992). In both speech and nonspeech contexts, when both steady-state and rapidly-changing cues are available, categorization is sharp (as with just the rapidly-changing cue) and discrimination exceeds categorization-based predictions (as with just the steady-state cue). Figure 5 summarizes the observed and categorization-based predicted discrimination performance for the three experiments and makes clear the difference in the patterns of performance. For stimuli that have a steady-state cue (experiments 2 and 3) to stimulus identity, discrimination performance is higher than predicted from categorization performance. But for stimuli that are defined only by a rapidly-changing cue (experiment 1), discrimination performance is accurately predicted by categorization performance.

III. GENERAL DISCUSSION

The present research examined differences in categorization and discrimination of sounds varying in rapidly-changing and steady-state acoustic cues. A novel stimulus

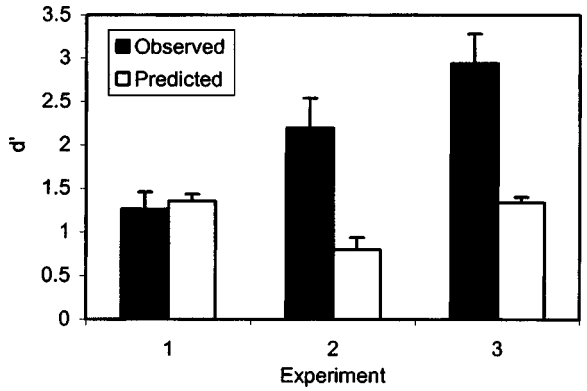

FIG. 5. Overall observed and predicted posttest discrimination performance by experiment. When a steady-state cue is available (experiments 2 and 3), discrimination performance exceeds categorization-based predictions.

space was created by applying bandstop filters and onset/offset ramps to bursts of white noise. Participants were taught to categorize stimuli varying along one of the dimensions of this space in blocks of categorization trials with feedback. Following training, participants were tested on categorization and discrimination of the stimuli. Results indicated that participants could categorize stimuli that varied along the rapidly-changing ramp length cue very effectively, with discrimination performance approximately at the level predicted by categorization performance (experiment 1). However, the same training procedure resulted in much less sharp categorization of stimuli that varied along the steady-state NF1 cue and produced discrimination that exceeded the level predicted from categorization responses (experiment 2). A study of categorization and discrimination along the NF1 cue following extensive categorization training (14 times more than experiment 2) found the same result. Thus, the difference in results between experiments 1 and 2 is not caused by a difference in rate of learning categories defined by these cues. When both cues were available (experiment 3), performance was ''the best of both worlds,'' combining the sharp categorization observed in the ramp cue experiment with discrimination performance that exceeded predictions from categorization, as observed in experiments using the steady-state NF1 cue.

The findings of these experiments mirror findings in speech perception. The pattern of categorization and discrimination of stop consonants is similar to the pattern of categorization and discrimination of the ramp-cued stimuli, whereas the pattern for steady-state vowels and fricatives is similar to the pattern for NF1-cued stimuli (Eimas, 1963; Pisoni, 1973; Healy and Repp, 1982; Repp, 1984). The pattern of categorization and discrimination of nonspeech sounds defined by both the ramp and NF1 cues is similar to the pattern for vowels defined by both steady-state spectral cues and rapidly-changing cues (Schouten and van Hessen, 1992). These data suggest that differences in patterns of categorization and discrimination performance reflect differences in processing of acoustic properties that speech and nonspeech sounds share. As reviewed in the Introduction, converging evidence from lateralization studies, individual difference studies, and studies in nonhuman animals all support the hypothesis that the auditory systems process rapidly-changing and steady-state cues differently and that these differences give rise to performance differences between vowels and consonants and between classes of nonspeech sounds.

To account for the cue-task interaction described in this report, it is useful to consider the demands of the discrimination task. Evidence suggests that decay of the perceptual trace is one factor limiting discrimination performance (e.g., Pisoni, 1973; Cowan and Morse, 1986). The perceptual trace decays relatively quickly, but if a category label has been assigned, the label may be available after the perceptual trace has decayed. In support of these hypotheses, researchers have demonstrated that discrimination performance falls closer to categorization-based predictions when the interstimulus-interval (ISI) between the sounds to be discriminated is extended (e.g., Pisoni, 1973; Cowan and Morse, 1986). Some researchers have suggested that the perceptual trace of stop consonants is less available for discrimination than the perceptual trace of steady-state vowels (Macmillan *et al.*, 1988). If so, this difference may explain differences in discrimination performance between the two types of speech sounds. Generalizing this idea to encompass nonspeech sounds, suppose that the perceptual trace of rapidly-changing sounds decays faster than that of steady-state sounds. Rapid decay of the perceptual trace would encourage reliance on category labels because they can be maintained in memory for a longer time. On the other hand, suppose steady-state sounds leave a longer-lasting perceptual trace. If the perceptual trace decays slowly, discrimination performance can exceed category label-based performance. In the context of the present experiments, this account claims that the perceptual trace of stimuli defined by the ramp cue decays more quickly than the perceptual trace of the stimuli defined by the NF1 cue. As the perceptual trace decays, listeners are forced to rely more on assigning category labels (learned during the categorization training phase), therefore discrimination performance is more closely predicted by categorization performance for the ramp stimuli (experiment 1) than for the NF1 stimuli (experiment 2). Thus, if one assumes that transient cues leave more transient perceptual traces, the memory demands of the discrimination task explain improved discrimination performance when steady-state cues are available. This account predicts that discrimination of nonspeech sounds defined by steady-state cues will fall closer to categorization-based predictions if longer interstimulus-intervals are used.

In summary, in the present experiments, listeners categorized and discriminated novel nonspeech sounds defined by a rapidly-changing acoustic cue, a steady-state cue, or both types of cues. The results showed three things. First, nonspeech sounds defined by a rapidly-changing acoustic cue elicited sharp categorization performance and discrimination performance that was accurately predicted by the assumption that the discrimination is performed solely on the basis of category labels. This pattern of results has been reported for stop consonants, which are distinguished by rapidly-changing acoustic cues such as formant transitions. Second, nonspeech sounds defined by a steady-state acoustic cue elicited less sharp categorization performance and discrimina-

tion performance exceeded predictions based on categorization performance. This pattern of results has been reported for synthetic steady-state vowels and fricatives, which are distinguished by steady-state acoustic cues such as formant frequency. Third, nonspeech sounds defined by both a rapidly-changing cue and a steady-state cue elicited both sharp categorization functions and discrimination performance that exceeded predictions based on categorization performance. This pattern of results has been reported for natural vowels that are distinguished by both steady-state and rapidly-changing acoustic cues. These similarities in data patterns for speech and nonspeech sounds suggest that categorization and discrimination performance are influenced by differences between auditory processing of rapidly-changing and steady-state acoustic cues for both types of sounds.

ACKNOWLEDGMENTS

This work was supported by a National Science Foundation grant (NSF BCS-0078768), by a James S. McDonnell Foundation award for Bridging Mind, Brain, and Behavior to LLH and Andrew Lotto, and by the Center for the Neural Basis of Cognition. DM was supported by NIH Training Grant No. T32N507433-03. JLM was supported by Grant No. MH64445 from the National Institute of Mental Health. The authors thank Andrew Lotto for stimulating discussion and Christi Adams, Siobhan Cooney, Seth Liber, Camilla Kydland, and Monica Datta for their help in conducting the experiments. The authors also thank Michael Owren and John Kingston for their helpful comments on an early draft.

[1]The relatively small number of trials during each block of sensitivity assessment causes traditional d' measurements to be somewhat unstable. This measure was used to keep this segment of the experiment relatively short.

Ades, A. E. (**1977**). ''Vowels, consonants, speech, and nonspeech,'' Psychol. Rev. **84**, 524–530.

Allard, F., and Scott, B. L. (**1975**). ''Burst cues, transition cues, and hemispheric specialization with real speech sounds,'' Q. J. Exp. Psychol. **27**, 487–497.

Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M.-C., and Samson, Y. (**1998**). ''Lateralization of speech and auditory temporal processing,'' J. Cogn Neurosci. **10**, 536–540.

Cowan, N., and Morse, P. A. (**1986**). ''The use of auditory and phonetic memory in vowel discrimination,'' J. Acoust. Soc. Am. **79**, 500–507.

Cutting, J. E. (**1974**). ''Two left-hemisphere mechanisms in speech perception,'' Percept. Psychophys. **16**, 601–612.

Cutting, J. E. (**1982**). ''Plucks and bows are categorically perceived, sometimes,'' Percept. Psychophys. **31**, 462–476.

Eimas, P. D. (**1963**). ''The relation between identification and discrimination along speech and nonspeech continua,'' Lang Speech **6**, 206–217.

Golestani, N., Paus, T., and Zatorre, R. J. (**2002**). ''Anatomical correlates of learning novel speech sounds,'' Neuron **35**, 997–1010.

Gottfried, T. L., and Strange, W. (**1980**). ''Identification of coarticulated vowels,'' J. Acoust. Soc. Am. **68**, 1626–1635.

Hauser, M. D., and Andersson, K. (**1994**). ''Left hemisphere dominance for processing vocalizations in adult, but not infant, rhesus monkeys: Field experiments,'' Proc. Natl. Acad. Sci. U.S.A. **91**, 3946–3948.

Hauser, M. D., Agnetta, B., and Perez, C. (**1998**). ''Orienting asymmetries in rhesus monkeys: The effect of time-domain changes on acoustic perception,'' Anim. Behav. **56**, 41–47.

Healy, A. F., and Repp, B. H. (**1982**). ''Context independence and phonetic mediation in categorical perception,'' J. Exp. Psychol. Hum. Percept. Perform. **8**, 68–80.

Heffner, H. E., and Heffner, R. S. (**1984**). ''Temporal lobe lesions and perception of species-specific vocalizations by macaques,'' Science **226**, 75–76.

Hillenbrand, J. M., and Nearey, T. M. (**1999**). ''Identification of resynthesized hVd/utterances: Effects of formant contour,'' J. Acoust. Soc. Am. **105**, 3509–3523.

Hillenbrand, J. M., Clark, M. J., and Nearey, T. M. (**2001**). ''Effect of consonant environment on vowel formant patterns,'' J. Acoust. Soc. Am. **109**, 748–763.

Hoemeke, K. A., and Diehl, R. L. (**1994**). ''Perception of vowel height: The role of F_1-F_0 distance,'' J. Acoust. Soc. Am. **96**, 661–674.

Jongman, A., Wayland, R., and Wong, S. (**2000**). ''Acoustic characteristics of English fricatives,'' J. Acoust. Soc. Am. **108**, 1252–1263.

Kewley-Port, D. (**1995**). ''Thresholds for formant-frequency discrimination of vowels in consonantal context,'' J. Acoust. Soc. Am. **97**, 3139–3146.

Kewley-Port, D., and Pisoni, D. B. (**1984**). ''Identification and discrimination of rise time: Is it categorical or noncategorical?'' J. Acoust. Soc. Am. **75**, 1168–1176.

Kewley-Port, D., and Watson, C. S. (**1994**). ''Formant-frequency discrimination for isolated English vowels,'' J. Acoust. Soc. Am. **95**, 485–496.

Kewley-Port, D., and Zheng, Y. (**1999**). ''Vowel formant discrimination: Towards more ordinary listening conditions,'' J. Acoust. Soc. Am. **106**, 2945–2958.

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (**1957**). ''The discrimination of speech sounds within and across phoneme boundaries,'' J. Exp. Psychol. **54**, 358–368.

Lisker, L., and Abramson, A. S. (**1964**). ''A cross-language study of voicing in initial stops: Acoustical measurements,'' Word **20**, 384–422.

Mack, M., and Blumstein, S. E. (**1983**). ''Further evidence of acoustic invariance in speech production: The stop-glide contrast,'' J. Acoust. Soc. Am. **73**, 1739–1750.

Macmillan, N. A. (**1987**). ''Beyond the categorical/continuous distinction: A psychophysical approach to processing modes,'' in *Categorical Perception: The Groundwork of Cognition*, edited by S. Harnad (Cambridge U.P., New York), pp. 53–85.

Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge U.P., New York).

Macmillan, N. A., Goldberg, R. F., and Braida, L. D. (**1988**). ''Resolution of speech sounds. Basic sensitivity and context memory on vowel and consonant continua,'' J. Acoust. Soc. Am. **84**, 1262–1280.

Maddox, W. T., Ashby, F. G., and Gottlob, L. R. (**1998**). ''Response time distributions in multidimensional perceptual categorization,'' Percept. Psychophys. **60**, 620–637.

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

Miller, R. L. (**1953**). ''Auditory tests with synthetic vowels,'' J. Acoust. Soc. Am. **25**, 114–121.

Pisoni, D. (**1971**). ''On the nature of categorical perception of speech sounds,'' Ph.D. thesis, University of Michigan, Ann Arbor.

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

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

Poeppel, D. (**2003**). ''The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time,' '' Speech Commun. **41**, 245–255.

Repp, B. H. (**1984**). ''Categorical perception: Issues, methods, findings,'' Speech Lang. **10**, 243–335.

Rosenthal, O., Fusi, S., and Hochstein, S. (**2001**). ''Forming classes by stimulus frequency: Behavior and theory,'' Proc. Natl. Acad. Sci. U.S.A. **98**, 4265–4270.

Schouten, M. E., and Van Hessen, A. J. (**1992**). ''Modeling phoneme perception: I Categorical perception,'' J. Acoust. Soc. Am. **92**, 1841–1855.

Schwartz, J., and Tallal, P. (**1980**). ''Rate of acoustic change may underlie hemispheric specialization for speech perception,'' Science **207**, 1380–1381.

Shamma, S. A. (**2000**). ''Physiological basis of timbre perception,'' in *The New Cognitive Neurosciences, 2nd Edition*, edited by M. S. Gazzaniga. (MIT, Cambridge, MA), pp. 411–423.

Strange, W., Verbrugge, R. R., Shankweiler, D. P., and Edman, T. R. (**1976**). ''Consonant environment specifies vowel identity,'' J. Acoust. Soc. Am. **60**, 213–224.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., and Widin, G. P. (**1987**). ''Speech

waveform envelope cues for consonant recognition,'' J. Acoust. Soc. Am. **82**, 1152–1161.
Walsh, M. A., and Diehl, R. L. (**1991**). ''Formant transition duration and amplitude rise time as cues to the stop/glide distinction,'' Q. J. Exp. Psychol. A **43A**, 603–620.
Wood, C. C. (**1976**). ''Discriminability, response bias, and phoneme categories in discrimination of voice onset time,'' J. Acoust. Soc. Am. **60**, 1381–1389.
Zatorre, R. J., Belin, P., and Penhune, V. B. (**2002**). ''Structure and function of auditory cortex: Music and speech,'' Trends Cogn. Sci. **6**, 37–46.