Auditory discontinuities interact with categorization: Implications for speech perception[a)]

Lori L. Holt[b)]
Department of Psychology and the Center for the Neural Basis of Cognition, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213

Andrew J. Lotto
Center for Hearing Research, Boys Town National Research Hospital, Omaha, Nebraska 68131

Randy L. Diehl
Department of Psychology and Center for Perceptual Systems, University of Texas at Austin, Austin, Texas 78712

(Received 5 September 2003; revised 9 June 2004; accepted 14 June 2004)

Behavioral experiments with infants, adults, and nonhuman animals converge with neurophysiological findings to suggest that there is a discontinuity in auditory processing of stimulus components differing in onset time by about 20 ms. This discontinuity has been implicated as a basis for boundaries between speech categories distinguished by voice onset time (VOT). Here, it is investigated how this discontinuity interacts with the learning of novel perceptual categories. Adult listeners were trained to categorize nonspeech stimuli that mimicked certain temporal properties of VOT stimuli. One group of listeners learned categories with a boundary coincident with the perceptual discontinuity. Another group learned categories defined such that the perceptual discontinuity fell within a category. Listeners in the latter group required significantly more experience to reach criterion categorization performance. Evidence of interactions between the perceptual discontinuity and the learned categories extended to generalization tests as well. It has been hypothesized that languages make use of perceptual discontinuities to promote distinctiveness among sounds within a language inventory. The present data suggest that discontinuities interact with category learning. As such, ''learnability'' may play a predictive role in selection of language sound inventories. [DOI: 10.1121/1.1778838]

PACS numbers: 43.71.Es, 43.71.Pc, 43.66.Ba [PFA] Pages: 1763–1773

I. INTRODUCTION

Considering the large set of physically realizable sounds of the human vocal tract, the regularity in phonetic inventories of surveyed languages is remarkable (Maddieson, 1984). There have been a number of hypotheses about factors that shape preferred sound inventories. A common theme running throughout these proposals is that sounds are favored if they are easy to produce and/or easy to distinguish auditorily from other sounds (Diehl *et al*., 2003; Liljencrants and Lindblom, 1972; Stevens, 1972, 1989). For example, according to the quantal theory (Stevens, 1972, 1989), there are regions along certain phonetic dimensions where articulatory changes have large acoustic consequences and other regions where comparable articulatory variation results in much more modest acoustic change. These alternating regions of acoustic instability and stability are assumed to create conditions for a kind of optimization of sound inventories. If sound categories are located in the stable regions, they require less articulatory precision and are thus easier to produce reliably than sound categories in unstable regions. These stable sound categories will also tend to be auditorily distinctive because they are separated by regions of acoustic instability.

Of course, communicative benefits arise not from acoustic distinctiveness among sounds but from distinctiveness within the human auditory system. Mathematical simulations that incorporate distance metrics based on realistic auditory representations have been able to predict some of the regularities witnessed in phonetic inventories (Diehl *et al*., 2003; Lindblom, 1986). Auditory representations differ from acoustic representations because the auditory perceptual space may be ''warped'' in such a way that equal acoustic distinctions need not produce equivalent perceptual changes. Languages may capitalize on regions of perceptual space where sensitivity is enhanced, adopting sounds for which moderate changes in articulatory or acoustic characteristics result in disproportionately large perceptual consequences. For example, there are some regions of acoustic space where sharp peaks in discrimination functions are observed. These ''auditory discontinuities'' [natural boundaries in perceptual space across which discrimination is enhanced (Kuhl and Miller, 1975)][1] would be good candidates for placement of boundaries between phonetic categories because tokens from each category would have enhanced distinctiveness with minimal acoustic change. Enhanced distinctiveness in turn provides more robust communication. Stevens (1989) extended the quantal theory to the relation between acoustic

[a)]Preliminary reports of this work were presented at the 145th Meeting of the Acoustical Society of America, June 2003 and at the 10th Annual Cognitive Neuroscience Society Meeting, March 2003.
[b)]Electronic mail: lholt@andrew.cmu.edu

signals and auditory responses. Certain nonlinearities in the auditory system were assumed to give rise to regions along acoustic dimensions that are auditorily stable and other regions along those same dimensions that are auditorily unstable. Again, it was hypothesized that sound categories tend to be located in the stable regions. An intervening unstable region would then correspond to a kind of auditory discontinuity or natural boundary that enhances the distinctiveness of sound categories.

In addition to enhanced distinctiveness, there is another benefit that may be conferred by utilizing auditory discontinuities as phonetic boundaries. A discontinuity may provide a parsing of the auditory space and delineate an initial ''natural'' boundary between phonetic categories (Kuhl, 1993). For a novice language-learner, these boundaries could facilitate the imposing task of forming phonetic categories from the considerable acoustic variance in speech. Phonetic systems that include boundaries coincident with auditory discontinuities may thus be more *learnable*. If this is the case, learnability may be a characteristic of communication systems that exerts selective pressure on the development of languages (Deacon, 1997; Lotto, 2000). That is, attributes of language systems that promote learning may tend to be favored. This includes aspects of phonetic categories. Evidence from studies of categorization in other domains suggests that some sets of stimuli are more readily categorized than others (Alfonso-Reese *et al*., 2002; Ashby *et al*., 1999). Many factors may contribute to the relative ease of categorization including stimulus dimensionality, covariance among stimulus dimensions, nonlinearity in category boundaries, and overlap in input distribution variability.

We propose here that the placement of category boundaries at preexisting auditory discontinuities may facilitate phonetic category formation and that this increase in learnability may be one of the sources of regularities observed in phonetic inventories. A particular example is the regularity in voicing category distributions seen across languages and the relation of this regularity to a well-documented nonmonotonicity in temporal processing.

A. Auditory discontinuities in temporal processing

Although a host of cues differentiate voiced from voiceless consonants (Lisker, 1986), the temporal interval between oral release and onset of voicing (voice onset time, VOT) is especially predictive of the contrast. Across languages, speakers tend to produce three kinds of voicing: *lead* consonants for which voicing precedes the release of the consonant occlusion (thus having a negative VOT value), *short lag* consonants for which voicing closely follows the oral release (with VOT values around 0–20 ms), and *long lag* consonants for which voicing lags well behind the consonant release (with large positive VOT values). Frequency distributions of VOT productions aggregated across languages reflect three modes, with modes centered at −100, +10, and +75 ms VOT (Cho and Ladefoged, 1999; Keating, 1984; Lisker and Abramson, 1964).[2] These characteristic distributions are cross-linguistically regular not just in their modes, but also in more precise statistical patterning. For example, distributions of VOT productions typically have a gap in the region corresponding to small negative VOT values. These values appear very rarely in speech production cross-linguistically. Keating (1984) suggests that this gap may be a general characteristic of languages regardless of the number of voicing categories. The extent, or range, of the distributions appears to be fairly regular across languages as well. Perhaps partly as a consequence of the absence of VOT productions in the small negative VOT range, short-lag distributions tend to have quite constrained variability across VOT, with values clustered from about 0 to 20 ms VOT (Keating, 1984). Therefore, cross-linguistically, boundaries between the voicing distributions generally occur at roughly −20 and +20 ms VOT.

Most languages make use of only two of the three distributions. English, for example, distinguishes short-lag from long-lag consonants, whereas Spanish contrasts lead with short-lag voicing. Of note, no language reported by Lisker and Abramson (1964) contrasts lead with long-lag voicing. This is of interest because theories predicting sound inventories from acoustic factors might expect to observe two-way voicing contrasts positioned at opposite ends of the VOT continuum, thus enhancing the acoustic distinctiveness of the contrast. However, each of the 51 languages surveyed by Keating *et al*. (1983) employed some type of short-lag stop consonant. Thus, rather than dispersing voicing productions acoustically, languages tend to contrast lead or long-lag voicing with the more intermediate (and less acoustically distinct) short-lag distribution.

Provocative early findings in infant speech perception indicated that mechanisms responsible for patterns of voicing perception are functional at a very early age (Eimas *et al*., 1971; Trehub and Rabinovich, 1972). Moreover, experience appears to play a relatively minor role in shaping voicing discrimination early in development (Lasky *et al*., 1975; Streeter, 1976). Lasky *et al*. (1975), for example, found that infants from a Spanish-speaking environment discriminate differences in VOT if stimuli straddle either the Spanish (lead versus short-lag) or the English (short-lag versus long-lag) VOT boundary but show little evidence of discrimination otherwise. This pattern contrasts with Spanish adults, who exhibit a single peak in discriminability across VOT at the lead versus short lag boundary (Abramson and Lisker, 1972; Elman *et al*., 1977; Williams, 1977). Aslin *et al*., (1981) have found comparable patterns of discriminability for infants from English-speaking homes. Moreover, Kikuyu-reared infants exhibit a similar discrimination pattern along a labial VOT continuum despite the fact that Kikuyu adults do not contrast voicing for labial articulations (Streeter, 1976). Thus, infants may exhibit regions of better-than-chance discriminability along a VOT dimension that are not characteristic of native adult perceivers. Moreover, the regions of best discrimination are relatively consistent across groups of infants and align fairly well with the boundaries at −20 and +20 ms VOT that partition the trio of distributions observed in VOT productions across languages.

Research with nonhuman animals has suggested that general characteristics of mammalian auditory processing may be responsible for these patterns of discrimination. Chinchillas and macaque monkeys without significant expe-

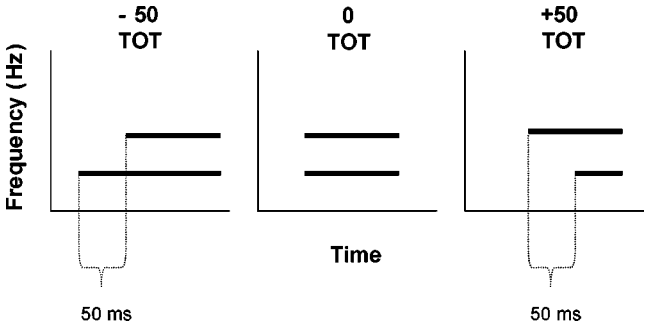

FIG. 1. Schematic spectrograms of TOT stimuli modeled after those of Pisoni (1977). Two tones (500 and 1500 Hz) vary in relative onset to create series of stimuli varying from −60 to +120 ms TOT. Three representative stimuli are shown.

rience with speech discriminate the VOT dimension much like human listeners (Kuhl and Miller, 1975; 1978; Kuhl and Padden, 1982). Kuhl and Miller interpreted these results as indicative of a general psychoacoustic discontinuity in VOT perception. Pisoni (1977) examined the generality of this proposal with human adult listeners by constructing a nonspeech dimension modeling a temporal characteristic of VOT (see also Miller *et al*., 1976). The nonspeech stimuli consisted of coterminous tones of 500 and 1500 Hz. Along a tone-onset-time (TOT) dimension, tones varied in their relative onset times such that the lower-frequency tone preceded the higher-frequency tone in onset, the two tones were simultaneous, or the lower-frequency tone lagged the higher-frequency tone. A schematic of three such stimuli is presented in Fig. 1. In a series of experiments, Pisoni (1977) found two regions of relatively more accurate TOT discrimination. Listeners were excellent at discriminating stimuli straddling approximately −20 and +20 ms TOT, but were poorer at discriminating stimuli spanning the same acoustic distance in other regions along the TOT dimension. Interestingly, the two regions of enhanced nonspeech TOT discrimination align quite closely with the −20 and +20 ms boundaries that partition the three distributions of VOT productions that are observed across languages (Lisker and Abramson, 1964). Figure 2 illustrates this discrimination pattern.[3] Similar patterns of TOT discrimination performance are obtained for 10-week-old infants (Jusczyk *et al*., 1980). Electrophysiological studies have revealed several neurophysiological correlates of discrimination peaks for VOT and TOT stimuli (e.g., Simos and Molfese, 1997; Simos *et al*., 1998; Sinex *et al*., 1991; Steinschneider *et al*., 1999).

Overall, data from adult and infant perception of speech and nonspeech, neurophysiological measures, and animal behavioral experiments indicate that the acoustic VOT and TOT dimensions are not identical to their auditory representations. Rather, there is a discontinuity such that stimulus pairs differentiated by equal acoustic change do not necessarily produce equivalent perceptual change; despite the equivalence of acoustic differences, discriminability may be enhanced for one pair relative to the other. The cross-linguistic regularities in voicing distributions suggest that many languages may exploit this discontinuity by drawing voicing categories from either side of the region of enhanced sensitivity. One benefit of this would be greater perceptual

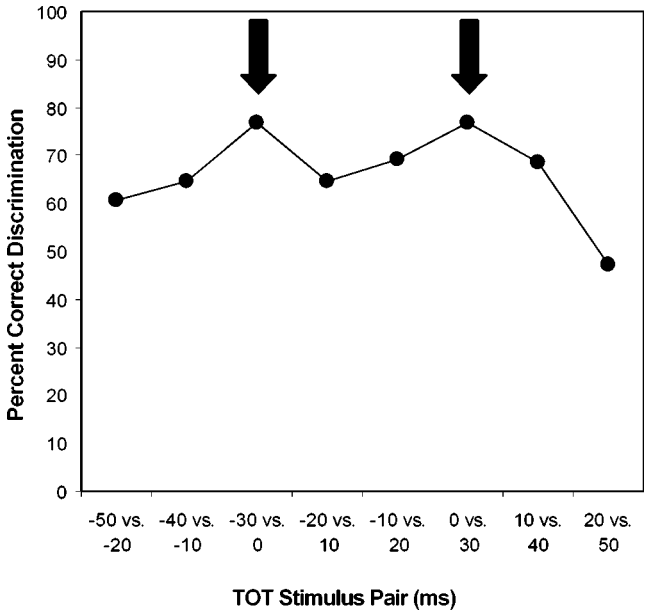

FIG. 2. Discrimination performance of participants in a pilot discrimination test illustrate the auditory discontinuity at approximately −20 ms and +20 ms. Although each stimulus pair spans an equal acoustic distance, discrimination is relatively more accurate when stimulus pairs straddle approximately −20 or +20 ms, as indicated by the arrows.

distinctiveness of tokens with small changes in acoustics (Kuhl and Miller, 1975; Pisoni, 1977). A second possible benefit would be that the discontinuities provide an initial parsing of the VOT space that can facilitate learning voicing categories that are consistent with these natural boundaries. That is, in addition to preferring sound categories that are "easy to say" and "easy to hear," languages may favor categories that are "easy to learn."

The present experiments were designed to test the hypothesis that operating characteristics of the auditory system constrain the learnability of sound categories. Specifically, they test the prediction that categories defined by onset asynchrony are easier to learn if they are consistent with the auditory discontinuity shown to occur at onset asynchronies near 20 ms. In these experiments, the categories are made up of TOT stimuli similar to those described by Pisoni (1977). The use of nonspeech stimuli provides control over stimulus attributes and participants' history of experience with the novel nonspeech TOT stimuli. As with the original work of Pisoni (1977), the results of nonspeech experiments can illuminate some of the general perceptual and cognitive processes that are involved in online speech perception and that may bias the development of linguistic systems.

II. EXPERIMENT 1

Participants in experiment 1 were trained with explicit feedback to assign TOT stimuli to two categories. One group of listeners learned two categories for which the boundary (as indicated by feedback and the tails of the distributions) was consistent with the reported peak in discriminability at approximately +20 ms. Another group of listeners learned two categories for which the boundary was inconsistent with the peak in discriminability. For this group of listeners, the peak in discrimination at +20 ms occurred within a category.

If learnability is enhanced by aligning category boundaries with perceptual discontinuities, then participants in the consistent condition should learn categories more quickly; that is, they should be faster to reach a criterion level of categorization performance. If learning does not interact with characteristics of the auditory space, then there should be no difference in the time to learn categories in the two conditions.

A. Methods

1. Participants

Twenty-eight listeners participated. One-half of the participants were randomly assigned to the consistent learning condition; the others were assigned to the inconsistent learning condition. All participants reported normal hearing, learned English as a first language, and were Carnegie Mellon University undergraduates or staff. Each participant was paid $15.

2. Stimuli

The stimuli were modeled after those of Pisoni (1977) and were synthesized using Matlab® (Mathworks, Inc.). A TOT dimension was constructed by creating two-tone stimuli with 500- and 1500-Hz tone components. The 1500-Hz tone had a constant duration of 230 ms and was approximately 12 dB lower in amplitude than the 500-Hz tone. Duration of the 500-Hz tone was varied to produce TOT differences. The only deviation from Pisoni's stimulus construction methods was the addition of 5-ms linear amplitude ramps at onset and offset. Stimuli were labeled with negative TOT values when the lower-frequency tone preceded the higher-frequency tone and with positive TOT values when the lower-frequency tone lagged the higher-frequency tone. A TOT value of zero indicated simultaneous onset. The full TOT series consisted of 37 stimuli varying from −60 to +120 ms TOT in 5-ms steps.[4] Figure 1 illustrates representative pseudospectrograms for −50, 0, and +50 ms TOT.

Acoustic presentation was controlled by TDT System II hardware (Tucker-Davis Technologies). Stimuli were converted from digital to analog, low-pass filtered at 4.8 kHz, amplified and presented binaurally over linear headphones (Beyer DT-100) at approximately 70 dB SPL(A).

3. Procedure

Participants first received categorization training with feedback to reach a criterion level of performance. Following training, listeners completed an ABX discrimination task. Next, they received a short categorization refresher. Finally, they again categorized stimuli in a generalization test across the entire TOT series without feedback. Each segment of the experiment was completed while the listener sat comfortably in a sound-attenuating booth wearing headphones and holding an electronic response box. Instructions preceded each segment on an LCD display mounted at eye level in front of the participant. Participants were allowed to take short breaks between segments and the entire session lasted less than 2 h.

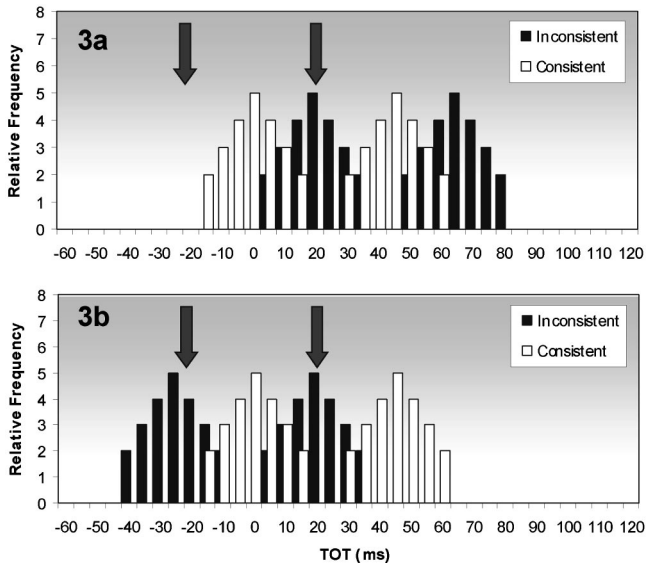

FIG. 3. Stimulus input distributions illustrate the relative frequency of stimulus presentation during training as a function of TOT. (a) Shows stimulus distributions of experiment 1. (b) Shows stimulus distributions for experiment 2. White bars correspond to consistent condition stimuli for which distributions and feedback assignment are coincident with the auditory discontinuity. Black bars show inconsistent condition stimuli where the auditory discontinuity is positioned within a distribution and feedback assignment is inconsistent with the auditory discontinuity.

a. Training. During training, listeners heard a single stimulus over headphones, made a two-alternative forced choice categorization response, and received feedback via a light above the correct response-box button. There were no labels on the response buttons.

Listeners in the consistent condition (for which auditory discontinuity, input distribution boundary, and boundary of feedback assignment were the same) heard stimuli drawn from the distributions illustrated as white bars in Fig. 3(a). The height of the bar illustrates the relative frequency of stimulus presentation for a particular TOT value. The arrows indicate peaks in discrimination observed at −20 and +20 ms TOT (see Fig. 2 pilot-test data). For the consistent condition, stimuli were sampled from two distributions separated along the TOT dimension at approximately +20 ms. Stimuli with TOT values less than +20 ms were assigned to one category and stimuli with TOT greater than +20 ms were assigned to the other category.

The input distributions of the inconsistent condition (for which auditory discontinuity differed from the input distribution boundary and the boundary of feedback assignment) are shown as dark bars in Fig. 3(a). The input distributions were separated along the TOT dimension at approximately +40 ms, a region of the TOT space that has no known nonlinearities in discriminability. The psychoacoustic discontinuity at +20 ms fell at the mode of one of the inconsistent condition distributions.

For each block of trials, listeners heard 46 stimuli that fully sampled the two input distributions. Listeners were instructed that some of the sounds corresponded to the left button whereas others corresponded to the right button. Following each response, the light above the correct button was illuminated as feedback. The mapping between distribution

and response button was counterbalanced across participants. Listeners were encouraged to maximize performance and were told that the length of the experimental session was a function of performance.

After each block of training, listeners' percent correct identification across the 46 categorization trials was calculated. Listeners completed training blocks until reaching a criterion of 90% correct performance (42 of 46 trials correct) or a maximum of 20 blocks (920 trials overall). Categorization of stimuli within a block took less than 5 min to complete.

b. Discrimination. After reaching criterion performance on the training segment or reaching the 20-block maximum, listeners completed an ABX discrimination task. On each trial, a stimulus pair (A and B) separated by five steps (25 ms) along the TOT dimension spanning −25 to +85 ms was presented with a 750-ms interstimulus silent interval. After another 750 ms of silence, a third stimulus (X, identical to either A or to B) was presented. Participants indicated whether the third stimulus was identical to either the first or to the second stimulus by pressing a button on an electronic response box. In all, participants discriminated 23 pairs of stimuli that occurred in four counterbalanced permutations per pair (ABA, ABB, BAA, BAB) for a total of 92 discrimination trials. Participants discriminated this full block of trials two times (making 184 overall discriminations). Collapsed across counterbalanced stimulus orderings, each subject made eight discrimination responses for each pair. Participants received feedback on each trial. Pisoni (1977; Pisoni *et al*., 1982) has argued that providing feedback in the discrimination test is critical because it forces participants to focus on the intended stimulus dimensions and to maintain the same focus from trial to trial.

c. Categorization refresher. To refamiliarize category mappings after the discrimination task, listeners heard two additional blocks, each fully sampling the input distributions. The procedure was identical to that of training. Listeners responded with a categorization response and received feedback on each trial.

d. Generalization test. Finally, listeners completed a generalization test of categorization. This task provided a test of categorization across the entire TOT range including stimuli not experienced during training. From these data, identification boundaries estimated using probit analysis and generalization of the learning were assessed. Stimuli spanned −60 to +120 ms TOT. Each stimulus was equally probable and was presented three times. Listeners were instructed to categorize the stimuli as in previous segments of the experiment. There was no feedback.

B. Results

1. Training

The primary hypothesis of experiment 1 was that the auditory discontinuity at approximately +20 ms TOT would interact with category learning. Specifically, it was predicted that category learning should be facilitated in the consistent condition, for which the category boundary coincided with the auditory discontinuity. As can be seen in Fig. 4, the data support this prediction. Listeners in the inconsistent condition required significantly more blocks of categorization training (M=8.21) to learn the TOT categories to a criterion level than did consistent-condition listeners (M=3.36), $t(13)=2.28$, $p<0.02$.

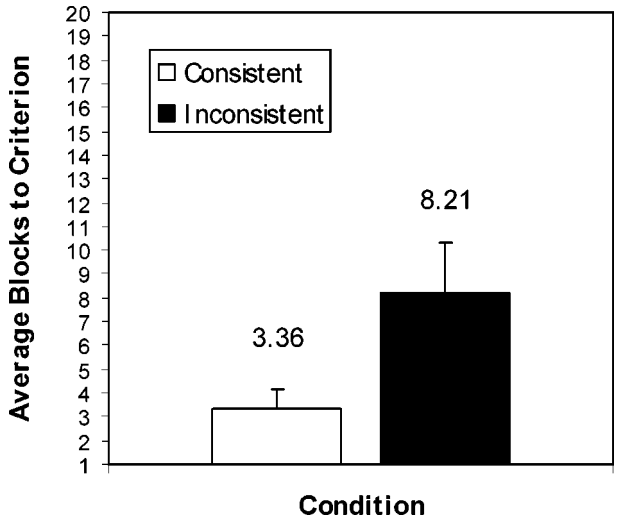

FIG. 4. Mean number of blocks required to reach criterion categorization performance during training for consistent and inconsistent condition listeners in experiment 1. Error bars represent standard error of the mean.

2. Generalization

Listeners labeled TOT stimuli varying from −60 to +120 ms TOT in the generalization segment of the experiment, thus responding to stimuli present in categorization training and to novel stimuli. These categorization results (presented as a function of an arbitrary category "A" indicating the input distribution with the lowest mean value along the TOT dimension) are shown with input distributions from training in Fig. 5. A 2×37 (condition × TOT stimulus) ANOVA revealed a significant main effect of condition, $F(1,36)=18.77$, $p<0.001$. Inconsistent condition participants categorized more of the TOT series stimuli as members of the "A" category than consistent condition listeners. There was also a significant main effect of TOT Stimulus, $F(1,36)=81.57$, $p<0.001$, indicating that categorization varied as a function of the TOT stimulus series as would be expected if listeners learned to accurately categorize the stimuli. There was also a significant condition × TOT stimulus interaction, $F(1,36)=2.28$, $p<0.001$, indicating that the identification functions differed in their shape across conditions. This pattern most likely arises from consistent condition listeners' categorization of TOT stimuli on the far negative end of the stimulus series (see Fig. 5). Consistent condition listeners imperfectly generalized to stimuli at this end of the series, instead rather consistently mimicking the frequency distribution of the trained categories in their generalization responses. Identification boundaries, as estimated from probit boundary analyses computed for individual subjects, support these analyses. There was a significant boundary shift across conditions, $t(13)=5.42$, $p<0.001$ (33.17 ms TOT for consistent listeners versus 51.62 ms TOT for inconsistent listeners). There was no significant difference in the

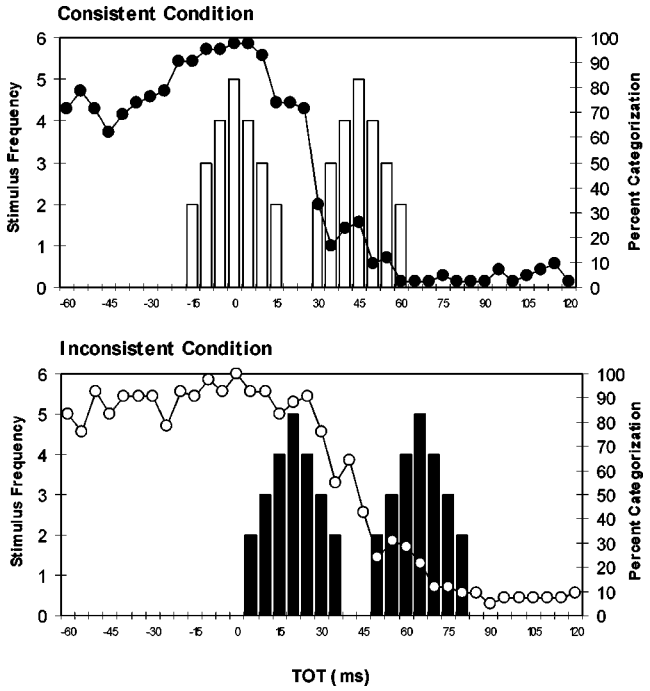

FIG. 5. Categorization responses (presented on the right-most *y*-axis as a function of an arbitrary category "A" indicating the input distribution with the lowest mean value along the TOT dimension) to stimuli presented in the generalization task are shown for consistent and inconsistent condition listeners in experiment 1. The histogram bars denote the position of the categories learned during training along the TOT dimension as a function of the relative frequency of stimulus presentation plotted on the left-most *y*-axis.

probit slopes, $t(13)=0.63$, $p=0.54$; the steepness of the identification functions was statistically equivalent across conditions.

3. Discrimination

Listeners also discriminated TOT stimuli in an ABX task. Whereas categorization and generalization analyses demonstrated significant effects across groups, there was little evidence that patterns of discrimination shifted with categorization training. A 2×23 (condition $\times$ discrimination pair) ANOVA revealed that consistent group listeners were somewhat better at all discriminations although not significantly so, $F(1,22)=3.22$, $p<0.09$. Even this slight trend is surprising, considering that consistent condition listeners had significantly less experience listening to the stimuli because they were faster to reach criterion categorization performance. Pairs were not equally discriminable, $F(1,22)=4.80$, $p<0.0001$, but the patterns did not relate cleanly to categorization experience. There was no condition $\times$ discrimination pair interaction, $F(1,22)=1.28$, n.s. Overall, discrimination does not appear to be a reliable index of listeners' categorization competence in the present paradigm.

4. Summary

The main prediction of the learnability hypothesis was supported by the results of experiment 1. Participants trained with a category boundary that was consistent with the auditory discontinuity learned to categorize the sounds faster than participants in the inconsistent condition. The data also make it clear that the presence of a single discontinuity does not determine possible learnable categories. Participants in the inconsistent condition *did* learn the categories on which they were trained. The generalization test shows a shift in identification boundaries and no obvious change in the identification function related to the discontinuity region. Thus, characteristics of general auditory representations appear to affect the ease with which categories are learned, but the general learning system is capable of overcoming the initial parsing of the perceptual space.

III. EXPERIMENT 2

In experiment 1, the two conditions differed in the range of TOT values that were categorized as well as the consistency of the category boundary with the auditory discontinuity. Although the overall range of TOT was equivalent across conditions, the inconsistent distributions included more large TOT values than did consistent distributions. It is possible that the perceptual space spanned by the acoustic TOT stimuli is compressively nonlinear (Fechner, 1860), rendering perceived stimulus differences of the inconsistent condition less salient and thereby increasing the difficulty of the inconsistent categorization task. This, too, would be an interaction of general operating characteristics of the auditory system with categorization, but not the sort under investigation in experiment 1.

Experiment 2 controls for this possibility. For this experiment, the two input distributions of the experiment 1 inconsistent condition were modified such that the input distribution with a mode at +65 ms TOT [see Fig. 3(a)] was reflected about +20 ms TOT to create a new input distribution with negative TOT values. The other three input distributions of experiment 1 remained unchanged, as shown in Fig. 3(b). Note that, in experiment 2, both of the inconsistent-condition input distributions spanned a region in auditory space defined by high discriminability. The peak in discriminability at approximately +20 ms fell within one input distribution and the peak in discriminability at about −20 ms fell within the opposite distribution. As a result, experiment 2 provides an opportunity to observe whether categories inconsistent with both of these auditory discontinuities are even more difficult for listeners to learn.

A. Methods

1. Participants

Participants were recruited from the Carnegie Mellon University undergraduate psychology participant pool. Each participant earned a credit toward a psychology research requirement and $5 for participation. All participants had normal hearing and learned English as the first language. Fourteen listeners participated in each condition for a total of 28 participants.

2. Stimuli

The TOT series stimuli were identical to those of experiment 1.

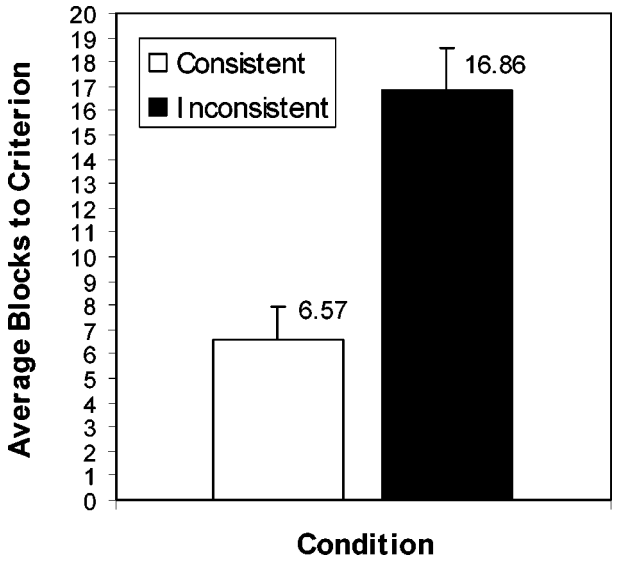

FIG. 6. Mean number of blocks required to reach criterion categorization performance during training for consistent and inconsistent condition listeners in experiment 2. Error bars represent standard error of the mean.

3. Procedure

The procedure was nearly identical to that of experiment 1. The only difference between experiments 1 and 2 was the position of one of the inconsistent condition distributions along the TOT dimension. Figure 3(b) illustrates the input distributions for experiment 2. Three of the four input distributions are the same as those of experiment 1. The TOT input distribution from experiment 1 centered on +65-ms TOT was replaced with an inconsistent condition distribution centered at −25-ms TOT, creating a boundary at approximately −5-ms TOT.

As a consequence of the reflected input distribution, inconsistent condition listeners in experiment 2 were required to categorize input distributions that crossed both −20 and +20 ms discontinuities. As a result, the inconsistent condition learning task of experiment 2 may have been even more difficult than that of experiment 1, for which only one input distribution crossed a discontinuity. Stimulus distributions for the consistent condition were equivalent to those of experiment 1 and thus provide a replication.

Tasks and experimental apparatus for this experiment were identical to those of experiment 1.

B. Results

1. Training

Figure 6 illustrates the mean number of blocks to criterion categorization performance across conditions. The results of experiment 1 do not appear to be merely a result of compressive nonlinearity in perception of large TOT values. Listeners in the inconsistent condition required significantly more blocks of categorization training (M=16.86) to learn the TOT categories to a criterion level than did consistent-condition listeners (M=6.57), $t(13)=4.71$, $p<0.0001$. Thus, the results of experiment 1 are replicated in experiment 2.

The consistent condition was identical to that of experiment 1. As would be expected, there was no significant difference in the number of blocks listeners required to reach criterion in the consistent condition across experiments, $t(13)=1.908$, $p=0.79$. However, it took listeners significantly longer to learn the inconsistent categories of experiment 2 than those of experiment 1 [16.86 vs 8.21 blocks, $t(13)=2.99$, $p<0.01$]. Note that one of the distributions for the inconsistent condition was identical in both experiments and the distance between distributions was constant. Despite these similarities, participants in experiment 2 performed more poorly in the categorization training task.

In fact, listeners in experiment 2 inconsistent condition had difficulty reaching 90% accuracy within the arbitrarily set maximum 20 blocks (920 trials) of training. In experiment 1, categorization performance in the last block of training was equivalent across conditions, $t(13)=1.84$, $p=0.09$, suggesting that listeners ultimately learned their respective categories equally well. However, in experiment 2, consistent condition listeners were significantly more accurate in the last block of training than were inconsistent condition listeners, $t(13)=3.866$, $p=0.002$. This is primarily due to the high proportion of inconsistent condition listeners ($N=10$, 71%) that did not reach 90% criterion within 20 training blocks. Categorization accuracy in the last block was only 72.05% for this subset of inconsistent condition listeners. Nevertheless, the task of inconsistent listeners was not impossible. One listener achieved criterion categorization performance in a single block of trials.[5]

It should also be noted, however, that there was a difference in the listener samples between experiments 1 and 2 such that across conditions, listeners in experiment 2 were slower (M=11.71 blocks) to reach criterion than experiment 1 listeners (M=5.78 blocks), $t(27)=3.44$, $p=0.002$, and the performance of experiment 1 Inconsistent condition listeners was not statistically distinguishable from the consistent condition listeners of experiment 2, $t(13)=0.583$, $p=0.57$. This latter piece of evidence should be considered in light of the fact that the arbitrarily 20-block maximum for the training session differentially affected inconsistent condition listeners, creating the possibility of a ceiling effect for the inconsistent condition and artificially reducing the hypothesized effect between the inconsistent and consistent conditions. Thus, interpretation of between-experiment comparisons should be interpreted with some caution. Nevertheless, whatever the source of listener differences between the experiments, each group's performance supports the primary hypothesis that the operating characteristics of the auditory system interact with learning to make inconsistent condition categories more difficult to learn than consistent condition categories.

2. Generalization

In the generalization segment of the experiment, listeners labeled stimuli varying from −60 to +120 ms TOT with no feedback. Average categorization functions (expressed in terms of an arbitrary category "A") for consistent and inconsistent condition listeners are shown in Fig. 7. A 2×37 (condition × TOT stimulus) ANOVA of these results revealed a significant main effect of TOT Stimulus, $F(1,36)=83.04$, $p<0.0001$, corresponding to the sigmoid shape of

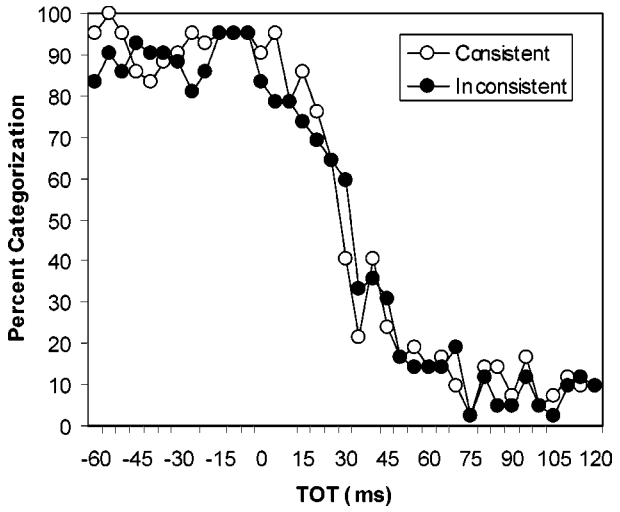

FIG. 7. Categorization responses during generalization task for consistent and inconsistent condition listeners in experiment 2.

the categorization function and indicating quite consistent categorization by listeners. Contrary to the results of experiment 1, there was no significant influence of condition upon generalization responses, $F(1,36)=1.06$, $p=0.3$. As indicated in Fig. 7, there was no evidence of a category boundary shift. There was no significant difference in probit boundaries, $t(12)=0.272$, $p=0.79$ ($M_{Con}=34.4$, $M_{Incon}=37.2$), and no difference in the slopes of the functions, $t(12)=1.84$, $p=0.09$.

3. Discrimination

In a 2×23 (condition × stimulus pair) ANOVA, there was a main effect of stimulus pair, $F(1,22)=5.283$, $p<0.0001$, indicating differences in discrimination accuracy as a function of stimulus pair. There was no condition main effect ($F<1$) and no condition by stimulus pair interaction ($F<1$).

4. Summary

The effect observed in experiment 1 was replicated in experiment 2. As in experiment 1, consistent condition listeners were faster to reach criterion categorization performance than listeners in the inconsistent condition. This result supports the original contention that it is the placement of the category boundary in relation to the auditory discontinuity and not the range over which TOT stimuli are categorized that makes the inconsistent distributions more difficult to learn.

It is of potential interest that experiment 2 inconsistent condition listeners were significantly slower to learn the categories than were experiment 1 inconsistent condition listeners. This appears to be due to the placement of experiment 2 inconsistent condition distributions on auditory discontinuities at both −20 and +20 ms TOT. The difficulty of learning the categories in the inconsistent condition of experiment 2 may be responsible for the absence of a category boundary shift relative to the consistent condition in the generalization test.

IV. DISCUSSION

The present findings indicate that auditory categories are not equally learnable. Basic auditory sensitivities interact with auditory categorization, causing some distributions of sounds to be more easily categorized than others. Specifically, categories separated by regions of high discriminability are more easily learned than those that span such regions. Nevertheless, despite the advantage that they confer to learnability, psychoacoustic sensitivities do not necessarily dictate categorization. Listeners in the inconsistent conditions required more experience, but did ultimately learn to categorize the sounds with high accuracy in experiment 1.

These results may have implications for learning phonetic categories. TOT stimuli were originally created to mimic the onset asynchrony between lower and higher frequency energy that corresponds to VOT in speech (Pisoni, 1977). If this analogy holds, then considering the present results it may be hypothesized that the placement of phonetic category distributions on either side of the discontinuity around 20 ms may confer enhanced learnability of speech categories distinguished by VOT. If linguistic systems are selected to be learnable, then phonetic inventories would be expected to include categories separated by the +20 and −20 ms VOT boundaries. This pattern is, in fact, widely attested among the world's languages (Keating, 1984; Lisker and Abramson, 1964). For example, the English labial voicing distinction (/b/ vs /p/) is realized with distributions of VOT values on either side of the 20-ms boundary. The fact that inconsistent listeners eventually reached criterion performance suggests that contrasts that do not follow this pattern are not unachievable, but may be more difficult to learn.

Previous attempts to account for regularities in phonetic inventories have focused on the communicative benefits arising from ease of production and ease of discrimination (e.g., Diehl *et al.*, 2003; Liljencrants and Lindblom, 1972; Lindblom, 1986; Stevens, 1972). In the case of VOT, the "easy to say"/"easy to hear" approach proposes that aligning category boundaries with auditory discontinuities enhances discriminability of tokens from different categories without imposing substantial articulatory costs. The learnability account proposed here should be seen as augmenting, not replacing, such accounts. Phonetic systems should be easy to say, easy to hear, *and* easy to learn. In fact, if assignment of speech sounds to equivalence classes (e.g., phonetic categories) serves a communicative function, then the ease with which categories can be formed and the accuracy with which sounds are assigned to categories may be an essential value on which linguistic systems are selected.

The proposal that auditory discontinuities can provide an initial parsing of the perceptual space for categorization is similar to Kuhl's description of "natural auditory boundaries" in her native language magnet theory of phonetic acquisition (Kuhl, 1991, 1993, 2000). According to Kuhl (1993), the initial discriminative abilities of infants reflect the presence of these boundaries. As infants are exposed to native language distributions of speech sounds, these boundaries are modified and sometimes deleted by the process of categorization. The results of the current experiments lend some support to this proposal. The onset asynchrony discon-

tinuity provided a natural boundary for the consistent condition categories but the negative effects of this discontinuity in the inconsistent condition of experiment 1 were ameliorated by learning. However, we must be cautious in generalizing from this one example to all phonetic categories. The voicing contrast is a fairly clear case of alignment between linguistic boundaries and natural auditory boundaries. There is also evidence for natural boundaries between some place-of-articulation contrasts for syllable-initial consonants (Eimas, 1974; Kuhl and Padden, 1983). However, there is no empirical evidence that auditory discontinuities are pervasive and underlie all or even most phonetic distinctions (e.g., Macmillan, 1987; Rosen and Howell, 1987). For example, recent studies have not found strong support for proposed auditory discontinuities in the vowel space (Fahey and Diehl, 1996; Fahey *et al*., 1996; Hoemeke and Diehl, 1994; Molis *et al*., 1998).

Inasmuch as auditory discontinuities do not dictate the categories that can be learned (recall that inconsistent condition listeners could learn the present categories with more experience), it is also important to acknowledge that auditory discontinuities themselves are malleable. Auditory discontinuities appear not to be immutable boundaries independent of other stimulus characteristics, but rather are a function of the spectral and temporal makeup of the stimulus and its surrounding context. For example, the boundary for TOT identification shifts with changes in the frequency separation between the two tones (Parker, 1988). Similarly, animal studies of VOT identification demonstrate boundary shifts with increased frequency separation of F1 (at energy onset) and higher formants (Kluender and Lotto, 1994; Kuhl and Miller, 1978). Recent intracranial electrophysiological recording experiments in humans and monkeys provide physiological evidence that the psychophysical discontinuity typically described as the ''20-ms boundary'' is dependent on the spectral characteristics of the stimuli (Steinschneider *et al*., 2004). Neurons' response properties across a VOT series shift as the F1 frequency of the speech syllables varies. Thus, although we have spoken of a 20-ms auditory discontinuity, it is the case that this boundary, in fact, is dependent upon the spectral characteristics of the acoustic components that define the onset asynchrony. As a result, we would expect the relative impact of the placement of categorization training boundaries to vary with changes in stimulus characteristics.

This more stimulus-bound way of viewing auditory discontinuities aligns with behavioral evidence of context-sensitivity of perception along VOT series (e.g., Miller and Liberman, 1979; Miller *et al*., 1986; Summerfield, 1981). That is, the present data do not argue that auditory discontinuities determine category boundaries absolutely. Categories with boundaries violating an auditory discontinuity (as in the inconsistent conditions) may be learned with more experience and, likewise, category boundaries may be shifted with adjacent context or changes in stimulus characteristics. What the present data suggest is that learnability of categories is enhanced when the boundaries of category distributions are aligned with the operating characteristics of auditory processing. When the discontinuities shift because of stimulus characteristics, so, too, should category boundaries. This is exactly what is seen for place of articulation and VOT.

Despite converging evidence regarding the 20-ms asynchrony discontinuity, there has been considerable debate over its existence (e.g., Kewley-Port *et al*., 1988; Pastore, 1988; Pastore and Farrington, 1996). One reason is that the behavioral correlates of the discontinuity are not consistently observed across measurement procedures and changes in spectra and may sometimes be impacted by differences in the relative degree of experience across participants (Watson and Kewley-Port, 1988). However, the present results are difficult to reconcile without suggesting that the auditory system may differentially process stimuli along the TOT dimension. A possible experience-bound alternative interpretation of the present findings is that English-speaking listeners' category boundary for voicing in speech influenced their perception and category learning of nonspeech stimuli along the TOT dimension. Consistent with this proposal, recent results have demonstrated an influence of native-language experience upon nonspeech perception (Bent *et al*., 2003). However, in experiment 2 inconsistent categories spanned the natural boundaries at both +20 ms (close to the native English voicing boundary) and −20 ms (in a region of space not used contrastively in English voicing). Native-English listeners were significantly poorer at learning the categories in this condition than in the inconsistent condition of experiment 1 that spanned only the +20 ms boundary. Therefore, the possibility that native-language-specific categories along the VOT dimension interact with TOT learning does not account fully for the present findings.

If learnability is a determinant of the structure of phonetic systems, then empirical work on general perceptual categorization is relevant for comprehensive theories of speech perception. Whereas there is a rich tradition of categorization work in vision, there is a dearth of information on the factors that may facilitate the formation of complex auditory categories. Some of these factors may be low dimensionality of the stimulus space, covariance among stimulus dimensions, and linear category boundaries. With respect to the last factor, Diehl *et al*. (2001) report that boundaries for vowel categories in psychoacoustically scaled (e.g., Mel or Bark) formant space appear to be linear. They point to results from experiments in which participants are asked to identify tokens broadly sampled across the formant space (e.g., Molis, 1999). In these studies, the boundaries that provide the best discriminative separation of the identification scores were often (although not exclusively) simple linear functions. Many boundaries were horizontal or vertical lines reflecting unidimensional distinctions between vowel categories, whereas most other boundaries were horizontal lines that reflected distinctions based on formant frequency averaging or difference scores. It is intuitively (and computationally) clear that categories defined by linear boundaries would be easier to learn than categories with complex nonlinear boundaries. However, in order to make strong claims about learnability, the ease of learning nonspeech auditory categories with different boundary configurations must be examined.

Factors affecting phonetic acquisition extend beyond the structure of the categories to the details of the learning environment. Kuhl *et al*. (2003) recently demonstrated that in-

fants exposed to live adult models picked up phonetic information better than infants exposed to recordings of speakers. This result suggests the importance of social interaction in the learning situation, perhaps to direct the attention of the learner to important aspects of the sensory input. In addition, the characteristics of speech directed at children may be tailored to enhance learnability of phonetic categories. Child-directed speech has led to more efficient learning of phonetic categories by a computer model than adult-directed speech (de Boer and Kuhl, 2003).

The degree to which these particular factors of input distributions and learning situations enhance category formation can be tested using methods similar to those of the current study. With nonspeech stimuli, the researcher has control over the training set and the training conditions. Data can be collected from identification, discrimination, and category goodness rating tasks to provide a clear picture of how these factors interact with learning and the structure of the resulting categories. As exemplified by the experiments described here, these results have implications for explaining regularities in phonetic systems specifically and linguistic systems more generally.

ACKNOWLEDGMENTS

This work was supported by a National Science Foundation grant (NSF BCS-0078768) to LLH and AJL, by a James S. McDonnell Foundation award for Bridging Mind, Brain, and Behavior to LLH, and by a research grant (5 R01 DC00427) from NIDCD to RLD. The authors thank Christi Adams for her help in conducting the experiments.

[1]A better term for this region may be a "non-monotonicity." The present arguments apply not only to regions of perceptual space that result in a qualitative change in perception or a step-function in responding, but more generally to regions of perceptual space warped in relation to physical acoustic space.

[2]See Cho and Ladefoged (1999) for language-specific nuances that supplement this broad description.

[3]The data depicted in Fig. 2 were collected in a pilot ABX discrimination study using the same TOT stimuli as in the categorization studies described in this report.

[4]Pisoni (1977) investigated a range of TOT spanning −50 to +50 ms.

[5]It is interesting to note that this listener had more than 12 years of musical training and studied composition as a major in the School of Fine Arts.

Abramson, A. S., and Lisker, L. (**1972**). "Voice-timing perception in Spanish word-initial stops," J. Phonetics **1**, 1–8.

Alfonso-Reese, L. A., Ashby, F. G., and Brainard, D. H. (**2002**). "What makes a categorization task difficult?" Percept. Psychophys. **64**, 570–583.

Ashby, F. G., Queller, S., and Berretty, P. M. (**1999**). "On the dominance of unidimensional rules in unsupervised categorization," Percept. Psychophys. **61**, 1178–1199.

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., and Perey, A. J. (**1981**). "Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience," Child Dev. **52**, 1135–1145.

Bent, T., Bradlow, A. R., and Wright, B. A. (**2003**). "The influence of linguistic experience on pitch perception in speech and nonspeech sounds," J. Acoust. Soc. Am. **113**, 2256.

Cho, T., and Ladefoged, P. (**1999**). "Variation and universals in VOT: Evidence from 18 languages," J. Phonetics **27**, 207–229.

de Boer, B., and Kuhl, P. K. (**2003**). "Investigating the role of infant-directed speech with a computer model," ARLO **4**, 129–134.

Deacon, T. W. (**1997**). *The Symbolic Species: The Co-evolution of Language and the Brain* (Norton, New York).

Diehl, R. L., Lindblom, B., and Creeger, C. P. (**2003**). "Increasing realism of auditory representations yields further insights into vowel phonetics," in *Proceedings of the 15th International Congress of Phonetic Sciences, Vol. 2* (Causal, Adelaide), pp. 1381–1384.

Diehl, R. L., Molis, M. R., and Castleman, W. A. (**2001**). "Adaptive design of sound systems: Some auditory considerations," in *The Role of Perceptual Phenomena in Phonological Theory*, edited by K. Johnson and E. Hume (Academic, San Diego), pp. 123–139.

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

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

Elman, J. L., Diehl, R. L., and Buchwald, S. E. (**1977**). "Perceptual switching in bilinguals," J. Acoust. Soc. Am. **62**, 971–974.

Fahey, R., and Diehl, R. (**1996**). "The missing fundamental in vowel height perception," Percept. Psychophys. **58**, 725–733.

Fahey, R. P., Diehl, R. L., and Traunmüller, H. (**1996**). "Perception of back vowels: Effects of varying F1—F0 Bark distance," J. Acoust. Soc. Am. **99**, 2350–2357.

Fechner, G. T. (**1860**). *Elemente der Psychophysik* (Breitkopf und Härtel, Leipzig).

Hoemeke, K. A., and Diehl, R. L. (**1994**). "Perception of vowel height: The role of F1-F0 distance," J. Acoust. Soc. Am. **96**, 661–674.

Jusczyk, P. W., Pisoni, D. B., Walley, A., and Murray, J. (**1980**). "Discrimination of relative onset time of two-component tones by infants," J. Acoust. Soc. Am. **67**, 262–270.

Keating, P. A. (**1984**). "Phonetic and phonological representation of stop consonant voicing," Language **60**, 286–319.

Keating, P., Linker, W. L., and Huffman, M. (**1983**). "Patterns of allophone distribution for voiced and voiceless stops," J. Phonetics **11**, 277–290.

Kewley-Port, D., Watson, C. S., and Foyle, D. C. (**1988**). "Auditory temporal acuity in relation to category boundaries: Speech and nonspeech stimuli," J. Acoust. Soc. Am. **83**, 1133–1145.

Kluender, K. R., and Lotto, A. J. (**1994**). "Effects of first formant onset frequency on [-voice] judgments result from general auditory processes not specific to humans," J. Acoust. Soc. Am. **95**, 1044–1052.

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

Kuhl, P. K. (**1993**). "Innate predispositions and the effects of experience in speech perception: The Native Language Magnet theory," in *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, edited by B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, and J. Morton (Kluwer Academic, Norwell, MA), pp. 259–274.

Kuhl, P. K. (**2000**). "Language, mind, and the brain: Experience alters perception," in *The New Cognitive Neurosciences*, edited by M. S. Gazzaniga (MIT, Cambridge), pp. 99–115.

Kuhl, P. K., and Miller, J. D. (**1975**). "Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants," Science **190**, 69–72.

Kuhl, P. K., and Miller, J. D. (**1978**). "Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli," J. Acoust. Soc. Am. **63**, 905–917.

Kuhl, P. K., and Padden, D. M. (**1982**). "Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques," Percept. Psychophys. **32**, 542–550.

Kuhl, P. K., and Padden, D. M. (**1983**). "Enhanced discriminability at the phonetic boundaries for the place feature in macaques," J. Acoust. Soc. Am. **73**, 1003–1010.

Kuhl, P. K., Tsao, F.-M., and Liu, H.-M. (**2003**). "Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning," Proc. Natl. Acad. Sci. U.S.A. **100**, 9096–9101.

Lasky, R. E., Syrdal-Lasky, A., and Klein, R. E. (**1975**). "VOT discrimination by four to six and a half month old infants from Spanish environments," J. Exp. Child Psychol. **20**, 215–225.

Liljencrants, J., and Lindblom, B. (**1972**). "Numerical simulation of vowel quality systems: The role of perceptual contrast," Language **48**, 839–862.

Lindblom, B. (**1986**). "Phonetic universals in vowel systems," in *Experimental Phonology*, edited by J. J. Ohala and J. J. Jaeger (Academic, Orlando, FL), pp. 13–44.

Lisker, L. (**1986**). "'Voicing' in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees," Lang Speech **29**, 3–11.

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

Lotto, A. J. (**2000**). ''Language acquisition as complex category formation,'' Phonetica **57**, 189–196.
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.
Maddieson, I. (**1984**). *Patterns of Sound* (Cambridge U. P., Cambridge).
Miller, J. D., Wier, C. C., Pastore, R. E., Kelly, W. J., and Dooling, R. J. (**1976**). ''Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception,'' J. Acoust. Soc. Am. **60**, 410–417.
Miller, J. L., and Liberman, A. M. (**1979**). ''Some effects of later-occurring information on the perception of stop consonant and semivowel,'' Percept. Psychophys. **25**, 457–465.
Miller, J. L., Green, K. P., and Reeves, A. (**1986**). ''Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast,'' Phonetica **43**, 106–115.
Molis, M. R. (**1999**). ''Perception of vowel quality in the F2/F3 plane,'' *Proceedings of the 14th International Congress of Phonetic Sciences*, pp. 191–194.
Molis, M. R., Diehl, R. L., and Jacks, A. (**1998**). ''Phonological boundaries and the spectral center of gravity,'' J. Acoust. Soc. Am. **103**, 2981.
Parker, E. M. (**1988**). ''Auditory constraints on the perception of voice-onset time: The influence of lower tone frequency on judgments of tone-onset simultaneity,'' J. Acoust. Soc. Am. **83**, 1597–1607.
Pastore, R. E. (**1988**). ''Burying straw men in imaginary graves: A reply to Kewley-Port, Watson, and Foyle (1988),'' J. Acoust. Soc. Am. **84**, 2262–2266.
Pastore, R. E., and Farrington, S. M. (**1996**). ''Measuring the difference limen for identification of order of onset for complex auditory stimuli,'' Percept. Psychophys. **58**, 510–526.
Pisoni, D. B. (**1977**). ''Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops,'' J. Acoust. Soc. Am. **61**, 1352–1361.
Pisoni, D. B., Aslin, R. N., Perey, A. J., and Hennessy, B. L. (**1982**). ''Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants,'' J. Exp. Psychol. Hum. Percept. Perform. **8**, 297–314.
Rosen, S., and Howell, P. (**1987**). ''Auditory, articulatory and learning explanations of categorical perception in speech,'' in *Categorical Perception*, edited by S. Harnad (Cambridge U. P., Cambridge), pp. 113–160.
Simos, P. G., and Molfese, D. L. (**1997**). ''Electrophysiological responses from a temporal order continuum in the newborn infant,'' Neuropsychologia **35**, 89–98.
Simos, P. G., Diehl, R. L., Breier, J. I., Molis, M. R., Zouridakis, G., and Papanicolaou, A. C. (**1998**). ''MEG correlates of categorical perception of a voice onset time continuum in humans,'' Cognit. Brain. Res. **7**, 215–219.
Sinex, D., McDonald, L., and Mott, J. (**1991**). ''Neural correlates of nonmonotonic temporal acuity for voice onset time,'' J. Acoust. Soc. Am. **90**, 2441–2449.
Steinschneider, M., Volkov, I., Noh, M., Garell, P., and Howard, M. (**1999**). ''Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex,'' J. Neurophysiol. **82**, 2346–2357.
Steinschneider, M., Fishman, Y. I., Volkov, I. O., and Howard, M. A. (**2004**). ''Spectral modulation of temporal responses in human and monkey primary auditory cortex: Relevance for voice onset time (VOT) encoding,'' in *Abstracts of the Twenty-seventh Annual Midwinter Research Meeting of the Association for Research in Otolaryngology*, p. 296.
Stevens, K. N. (**1972**). ''The quantal nature of speech: Evidence from articulatory-acoustic data,'' in *Human Communication: A Unified View*, edited by E. E. J. David and P. B. Denes (McGraw–Hill, New York), pp. 51–66.
Stevens, K. N. (**1989**). ''On the quantal nature of speech,'' J. Phonetics **17**, 3–45.
Streeter, L. A. (**1976**). ''Language perception of 2-month-old infants shows effects of both innate mechanisms and experience,'' Nature (London) **259**, 39–41.
Summerfield, Q. (**1981**). ''On articulatory rate and perceptual constancy in phonetic perception,'' J. Exp. Psychol. Hum. Percept. Perform. **7**, 1074–1095.
Trehub, S. E., and Rabinovich, M. S. (**1972**). ''Auditory-linguistic sensitivity in early infancy,'' Dev. Psychol. **6**, 74–77.
Watson, C. S., and Kewley-Port, D. (**1988**). ''Some remarks on Pastore (1988),'' J. Acoust. Soc. Am. **84**, 2266–2270.
Williams, L. (**1977**). ''The perception of stop consonant voicing by Spanish-English bilinguals,'' Percept. Psychophys. **21**, 289–297.