Carnegie Mellon University
Research Showcase @ CMU

Department of Psychology | Dietrich College of Humanities and Social Sciences

2-2006

Cue Weighting in Auditory Categorization: Implications for First and Second Language Acquisition

Lori L. Holt
Carnegie Mellon University, lholt@andrew.cmu.edu

Andrew J. Lotto
University of Texas at Austin

Follow this and additional works at: http://repository.cmu.edu/psychology

This Article is brought to you for free and open access by the Dietrich College of Humanities and Social Sciences at Research Showcase @ CMU. It has been accepted for inclusion in Department of Psychology by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

Cue weighting in auditory categorization: Implications for first and second language acquisition[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 Perceptual Systems University of Texas at Austin, Austin, Texas 78712

(Received 22 April 2005; revised 14 February 2006; accepted 28 February 2006)

The ability to integrate and weight information across dimensions is central to perception and is particularly important for speech categorization. The present experiments investigate cue weighting by training participants to categorize sounds drawn from a two-dimensional acoustic space defined by the center frequency (CF) and modulation frequency (MF) of frequency-modulated sine waves. These dimensions were psychophysically matched to be equally discriminable and, in the first experiment, were equally informative for accurate categorization. Nevertheless, listeners' category responses reflected a bias for use of CF. This bias remained even when the informativeness of CF was decreased by shifting distributions to create more overlap in CF. A reversal of weighting (MF over CF) was obtained when distribution variance was increased for CF. These results demonstrate that even when equally informative and discriminable, acoustic cues are not necessarily equally weighted in categorization; listeners exhibit biases when integrating multiple acoustic dimensions. Moreover, changes in weighting strategies can be affected by changes in input distribution parameters. This methodology provides potential insights into acquisition of speech sound categories, particularly second language categories. One implication is that ineffective cue weighting strategies for phonetic categories may be alleviated by manipulating variance of uninformative dimensions in training stimuli. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2188377]

PACS number(s): 43.71.An, 43.66.Ba, 43.71.Hw [ALF] Pages: 3059–3071

I. INTRODUCTION

Outside of the acoustic researcher's laboratory, few sound categories are distinguished by a single acoustic dimension. In the natural world, auditory categories more typically are defined along multiple dimensions. Integration of information across acoustic dimensions, then, must be a central characteristic of auditory processing.

Speech categories provide an excellent illustration of both the complexity with which multiple acoustic dimensions define auditory categories and of the adeptness of the auditory system at integrating information across multiple dimensions. For speech categories, rarely is a single acoustic dimension necessary or sufficient to define category membership; this is the essence of the classic "lack of invariance" issue (Liberman, 1996; Liberman *et al.*, 1967). For example, there are many acoustic dimensions that contribute to voicing, as in the difference between English /ba/ and /pa/; Lisker (1986) cataloged as many as 16 acoustic dimensions that may characterize the English voicing distinction.

Nevertheless, for speech and other auditory categories the existence of multiple acoustic dimensions does not imply their perceptual equivalence. Simply put, some acoustic dimensions play a greater role in determining the perceptual identity of a sound than do others. The fact that acoustic dimensions need not contribute equivalently to category identity has been referred to as *cue weighting*. Acoustic dimensions appear to be perceptually weighted in the sense that some are strongly correlated to categorization responses whereas others, although present, weakly determine perceived category membership. As an example, both spectral and temporal acoustic cues differentiate English tense and lax vowels like /i/ versus /I/. Adult native American-English listeners, however, rely much more on the spectral dimension (formant frequency) than the temporal dimension (vowel duration) in categorizing /i/ and /I/ (e.g., Hillenbrand *et al.*, 2000). Likewise, Francis and colleagues (2000) have demonstrated that although burst cues and formant transition cues co-vary for American English stop consonant categories, listeners rely on the formant cue significantly more than the burst cue in categorization responses (see also, Walley and Carrell, 1983). Even with explicit training with feedback to use the burst cue, the listeners continue to rely on the formant transition (Francis *et al.*, 2000).

What determines the relative weighting of different sources of acoustic information? An adaptive listener would weight dimensions based on experience over time with the acoustic environment. For example, we might be able to pre-

[a)]Portions of this work were presented at the 148th Meeting of the Acoustical Society of America, San Diego, CA, November, 2004 and at the 28th Mid-Winter Meeting of the Association for Research in Otolaryngology, New Orleans, LA, February, 2005.

[b)]Electronic mail: lholt@andrew.cmu.edu

dict weighting functions for speech perception if we knew how acoustic dimensions co-varied with phonetic contrasts in a listener's experience. In principle, the regularity of information along different dimensions could be fully characterized if it were possible to make exhaustive acoustic measurements of the full range of speech productions for a language, an idiolect, an accent, a speaker, or group of speakers. In practice, this has been approached much more modestly by examining regularities in the realization of speech productions of small samples of speakers (e.g., Lisker and Abramson, 1964; Lotto *et al.*, 2004). Nevertheless, even with this more limited approach, it is possible to estimate the distributions of phonetic categories across a subset of acoustic dimensions. Adaptive listeners will tune their weighting of the dimensions based on the characteristics of these distributions to maximize accuracy. However, there are a number of constraints related to sensory processing, cognitive processing, and previous experience that may prevent listeners from achieving ideal performance on any particular categorization task.

Examining this at a finer grain, cue weighting can be thought of as a function of at least four variables. The first two relate to the distributional characteristics of acoustic information and are language specific. First, acoustic dimensions vary in their *informativeness* for category identity; that is, there is a difference in the distinctiveness of the distributions for competing categories along each dimension. For example, voice onset time (VOT) is a robust acoustic signal for voicing across languages. Lisker and Abramson and others (Keating, 1984; Lisker and Abramson, 1964) have examined the distribution of VOT across speech categories in several languages. In American English, for example, VOT values produced by speakers are quite reliably differentiated across speech categories. This is to say that the distributions of the voicing categories do not overlap much along the VOT dimension. As a result, VOT is highly informative for the American English voicing distinction. Because of this informativeness, VOT is likely to be a heavily weighted cue in the perception of voicing categories by American English listeners. Note, as well, that implicit in this description is the hypothesis that listeners are sensitive, not just to the absolute value on an acoustic dimension for a particular stimulus presentation, but also to the distributional characteristics of that cue as it occurs across speech instances within a language.

A second parameter of auditory distributions that could affect perceptual cue weighting is distribution *variance*. Information is carried by variance in the signal and the auditory system seems to be especially sensitive to dimensions that are varying. Studies of nonspeech pattern analysis have established that components with greater relative variance receive greater perceptual weight (Lutfi, 1993; Lutfi and Doherty, 1994). From these findings, one may predict that dimensions on which speech distributions vary the most will receive higher perceptual weights. But in categorization tasks, such as phonetic categorization, the relationship between variance and cue weighting is likely to be more complex. The relationship between within-category and between-category variance is probably an important determinant of perceptual weight. Whereas large overall variance (disregarding category) is indicative of an informative acoustic dimension, large within-category variance can decrease the informativeness of a dimension by creating distribution overlap (in the same way that variance can decrease d' in signal-detection theory). In addition, increased within-category variance across a dimension could mean that the dimension is not robustly related to the category or that there is a great deal of noise in the transmission of this dimension. In either case, the reliability of the dimension is questionable and, correspondingly, the dimension may be weighted less.

It is important to note that each of these distributional aspects of cue weighting must be considered to be language specific, dialect specific, and perhaps even speaker specific. Although the same acoustic dimensions may exist across languages or dialects, the phonetic distributions are clearly not equivalent. For example, across the voiced-voiceless distinction in American English stops, aspiration is correlated with voicing. Voiceless consonants (with longer VOTs) are typically aspirated whereas voiced consonants are much less likely to be produced with aspiration (e.g., Stevens, 1998). Aspiration is thus likely to be a relatively strong cue to voicing in American English. Aspiration also exists for Hindi stop consonant categories, but it does not carry weight as a cue for voicing because it is strongly correlated with an orthogonal category dimension (aspirated versus unaspirated, Benguerel and Bhatia, 1980). As another example, consider the relationship of vowel length to syllable-final consonant voicing. Adult native listeners of languages that do not have syllable-final obstruents or that fail to make a vowel-length distinction as a function of the voicing of the final consonant do not weight vowel duration as a cue to the final-consonant's category identity as much as listeners of languages that make vowel-length distinctions as a function of final-consonant voicing (Crowther and Mann, 1992, 1994; Flege and Wang, 1989). Listeners' weighting of acoustic dimensions, then, arises from experience with the regularities of acoustic realization in the native patterns of speech production, although the learning mechanisms remain, as yet, relatively poorly understood. At a somewhat finer level, perceptual weighting may have speaker-specific characteristics. Weighting of acoustic dimensions across individual speakers may vary somewhat and listeners' familiarity with this variation may contribute to voice recognition.

Two other variables that can influence cue weighting can be thought to be non-distributional and not language specific or speaker specific. First, the basic auditory representation of an acoustic dimension can influence its perceptual weight. Some acoustic dimensions are more robustly encoded by the auditory system than others. For example, there may be discontinuities in auditory processing along particular acoustic dimensions such that equal physical steps do not produce equivalent changes in percept (e.g., Kuhl and Miller, 1975; Stevens, 1989). VOT, again, serves as a fine example because its perception is thought to be influenced by a general auditory discontinuity in processing the onsets of simultaneous acoustic events (Kuhl and Miller, 1975; Pisoni, 1977). The acoustic dimension distinguishing VOT can be thought of as a relative onset asynchrony between two component frequencies, one higher and one lower in frequency. The

lower-frequency component may either lead or lag the higher-frequency component, or it may occur simultaneously with it. Research from human adult, infant, and animal behavior as well as electrophysiology from these populations has suggested that the mammalian auditory system is poor at resolving onset asynchrony differences less than about 20 ms, whether they are speech or nonspeech stimuli (e.g., Jusczyk *et al.*, 1980; Kuhl and Miller, 1975; Pisoni, 1977; Simos and Molfese, 1997; Simos *et al.*, 1998: Sinex *et al.*, 1991; Steinschneider *et al.*, 1999). Sinex and colleagues, for example, relate this auditory discontinuity to changes in the variance of the neural representation at the auditory nerve; there appears to be a relatively more robust neural code in the representation of stimuli with onset asynchronies greater than about 20 ms (Sinex and McDonald, 1989; Sinex *et al.*, 1991). It has been suggested that languages may exploit this discontinuity in auditory processing by developing voicing categories that straddle this sensitivity and thus exaggerate the perceived difference between categories (e.g., Pisoni, 1977). Recent research has indicated that in addition to the perceptual benefit of exaggerated discriminability, such placement may also facilitate category learning (Holt *et al.*, 2004). By these means, the relatively quantal representation for VOT (and consequent high discriminability of the category distinction) may contribute to a stronger cue weight. Of course, the opposite circumstances may also influence perceptual weighting; acoustic dimensions may be less heavily weighted as a consequence of a weak auditory representation, for example, if they are more susceptible to noise or are represented with more variability by the auditory system than other acoustic dimensions. Moreover, if the auditory system does not have sufficient resolving power or if there is too much noise in the representation of an acoustic dimension then the covariance between the dimension and the category label may be underestimated by the perceiver. In general, more discriminable acoustic dimensions will tend to be more informative for a listener and, correspondingly, should be expected to have a greater perceptual weight.

Cue weighting may also be influenced by task. The acoustic dimensions heavily weighted for phonetic categorization may be much less informative in making a lexical decision, identifying a talker, evaluating the emotional content of speech, or making other perceptual decisions. The same acoustic dimension present in the same signal may be heavily weighted for one perceptual task, but less heavily weighted for a different task. Likewise, changes in the task that affect the auditory representation (for example, the addition of low-frequency noise that selectively masks a particular acoustic dimension) or the informativeness of a cue (for example, manipulation of the distribution characteristics on an acoustic dimension in the laboratory) would also be expected to bring about shifts in listeners' perceptual cue weighting. This expectation of adaptive plasticity is particularly important for our understanding of how to alter cue weighting in aid of acquiring new categories, for example in the acquisition of second-language (L2) phonetic categories. The present experiments exploit these possibilities.

To summarize the above, not all available information is equivalent in perceptual processing and the differential contribution of acoustic dimensions to categorization arises as a result of experience with regularities in the input, the robustness or variability of the perceptual coding of an acoustic dimension, and its informativeness to category identity as a function of task. Cue weighting is a quantitative description of how auditory information (and information from other modalities as well) is integrated in perceptual categorization. An appreciation of perceptual cue weighting leads to the perspective that speech categorization is not just a matter of detecting available auditory cues along various acoustic dimensions, but also applying some weighting function that is, at least in part, dependent on experience with phonetic distributions.

This hypothesis has important implications for language acquisition. There has been considerable interest in the perceptual difficulties of L2 learners in acquiring second-language speech categories. One underlying cause of these difficulties may be that well-established native-language weighting functions are inappropriate for the L2. The language-specific nature of cue weighting may create perceptual difficulties in integrating acoustic information for nonnative speech categories. Understanding the mechanisms of cue weighting and investigating the means by which it may be possible to change these functions are therefore of considerable importance for understanding L2 category acquisition. As an example, take the classic case of English syllable-initial /l/ versus /r/ perception by native Japanese listeners (Miyawaki *et al.*, 1975). For native American English listeners, the starting frequency of the third formant (F3) carries most of the perceptual weight, but Japanese listeners appear to heavily weight F2 starting frequency in categorizing English /l/ and /r/ (Iverson *et al.*, 2003; Yamada and Tohkura, 1990). This weighting is nonoptimal as F2 is not varied contrastively for syllable-initial liquids by English speakers (Lotto *et al.*, 2004). An understanding of the basic processes by which acoustic dimensions are perceptually weighted and determination of the means by which these weights may be altered therefore are likely to have straightforward applications to understanding this classic L2 learning problem, and others.

Cue weighting also appears to be important for a full understanding of first language (L1) acquisition. Although infants begin to become perceptually tuned to the characteristics of L1 already in the first year (Jusczyk, 1997; Kuhl *et al.*, 1992; Werker and Tees, 1984), this process ultimately takes significant time to develop to adultlike perception (e.g., Morrongiello *et al.*, 1984; Nittrouer *et al.*, 1998; Nittrouer *et al.*, 2000; Nittrouer, 2004; Parnell and Amerman, 1978). Specifically, children appear to apply different weightings to acoustic dimensions in perceiving L1 speech categories (Hazan and Barrett, 2000; Mayo and Turk, 2004; Nittrouer, 2004). For example, in languages in which adults weight vowel duration as a strong cue in categorization of following final stop consonants, children tend to favor the consonant-influenced vowel-formant transitions over vowel duration (Nittrouer, 2004).

The purpose of the present experiments is to address issues of cue weighting in an effort to begin to understand the means by which the auditory system weights acoustic

dimensions and, also, to investigate methods by which listeners' cue weighting might be shifted. This latter aim is of particular importance in understanding the link between cue weighting and speech category acquisition in L1 and L2. It is presumed that cue weighting functions are determined in large part by the characteristics of distributions of experienced sounds. However, uncovering the role of experience in shaping speech categorization is quite difficult because it is impossible to know the precise history of speech experience that a listener brings to the laboratory. Therefore, as a starting point, we have taken the approach of investigating novel nonspeech auditory category learning. This allows us to maintain total control over listeners' histories of experience with the acoustic exemplars, thus ensuring that we have a precise characterization of the distributional characteristics. We can then describe the mapping between the informativeness of each dimension and the correlation of that dimension with categorization responses (i.e., its weighting). In the experiments that follow, we examine how changing the characteristics of these acoustic distributions or the manner of presentation can affect weighting functions. In each experiment, listeners are trained with feedback to categorize nonspeech sounds as coming from one of two predefined distributions. These microgenetic experiments were designed to examine auditory category formation with application to phonetic category acquisition. In fact, the characteristics of the distributions that were used were very similar to previous studies examining speech sound categorization in human adults, infants and non-human animals (Grieser and Kuhl, 1989; Kluender *et al.*, 1998; Kuhl, 1991).

II. EXPERIMENT 1

In Experiment 1, listeners were trained to categorize unfamiliar sounds drawn from two distributions situated in a two-dimensional acoustic space. The two acoustic dimensions were equated for perceptual discriminability in pilot psychophysical experiments and possessed no known *a priori* relationship. Moreover, the input distributions were created so that the two dimensions were equally informative for the categorization task. The questions under investigation were whether listeners would integrate information from both dimensions in making category decisions and, if so, whether the resulting cue weighting functions were predictable from the (in this case equal) informativeness of the dimensions.

A. Methods

1. Participants

Fourteen volunteers recruited from Carnegie Mellon University, Pittsburgh, PA, participated for course credit or a small payment. All listeners reported normal hearing.

2. Stimuli

The experiment used methods from auditory category-learning experiments (e.g., Holt *et al.*, 2004; Mirman *et al.*, 2004) whereby distributions of sounds drawn from a two-dimensional acoustic space were sampled in stimulus presentation. Listeners labeled the sound and received feedback to learn to assign category labels during training. In a subsequent test, novel stimuli drawn from the same two-dimensional acoustic space were presented to assess the relative weighting of the two acoustic dimensions as a consequence of categorization training. This design allows response patterns to be directly compared to distribution characteristics (e.g., means, variances, overlap).

The nonspeech stimuli used in all four experiments were frequency-modulated tones. The two-dimensional acoustic space from which stimuli were sampled was defined by center, or carrier, frequency (CF) and modulation frequency (MF). Within this acoustic space, two distributions corresponding to the to-be-learned auditory categories were created; they are illustrated in Fig. 1(a) as open symbols. Forty-eight unique stimulus exemplars defined each of the categories; each category exemplar is represented by an open symbol in Fig. 1(a), with circles marking one category and diamonds the other. Each stimulus was created from a sine wave tone with a particular CF modulated with a depth of 100 Hz at the corresponding MF. For example, if the CF was 760 Hz and the MF was 203 Hz, the tone was modulated from 710 to 810 Hz at a rate of 203 Hz. Each stimulus was 300 ms long and was sampled at 10 kHz with 16-bit resolution. Stimuli were created using the Cool Edit (Syntrillium, Inc., Phoenix, AZ) software.

It was desirable that sampling of the distributions along the two acoustic dimensions produce perceptual changes that were roughly equivalent. To this end, estimates of the just-noticeable-difference along the CF and MF dimensions were determined by pilot psychophysical studies in which one dimension (CF or MF) was varied while the other was held constant at an intermediate value. Based on informal listening, these tests began with minimum steps of 15 Hz along the CF dimension and 3 Hz along the MF dimension. Young, normal-hearing listeners ($N=20$) recruited from the same population as those participating in the reported categorization studies responded to pairs of stimuli varying in either CF or MF in an AX discrimination test with stimulus pairs separated by 1, 2, 3, 4, or 5 steps. This series of pilot discrimination studies revealed that the CF and MF dimensions were approximately equally discriminable (70% accuracy) when step size was adjusted to be 30 Hz along the CF dimension and 18 Hz along the MF dimension. The results of these studies also confirmed that discrimination was flat across each of the dimensions, indicating that there were no auditory discontinuities in discriminability across the region of the two-dimensional stimulus space under observation.

For Experiment 1, the positioning of the input distributions in the acoustic space resulted in the CF and MF acoustic cues being equally informative for the categorization task. Distribution overlap was equal on both dimensions. One way to quantify "informativeness" is to calculate the relative increase in accuracy over chance performance that an ideal observer could attain by using the dimension. For the current study, we used a simple criterion-bound model in which the ideal observer would place a criterion at the optimal (in terms of accuracy) position along the dimension. All stimuli on one side of the criterion are designated as category "A" and all on the other side are designated category "B." Any

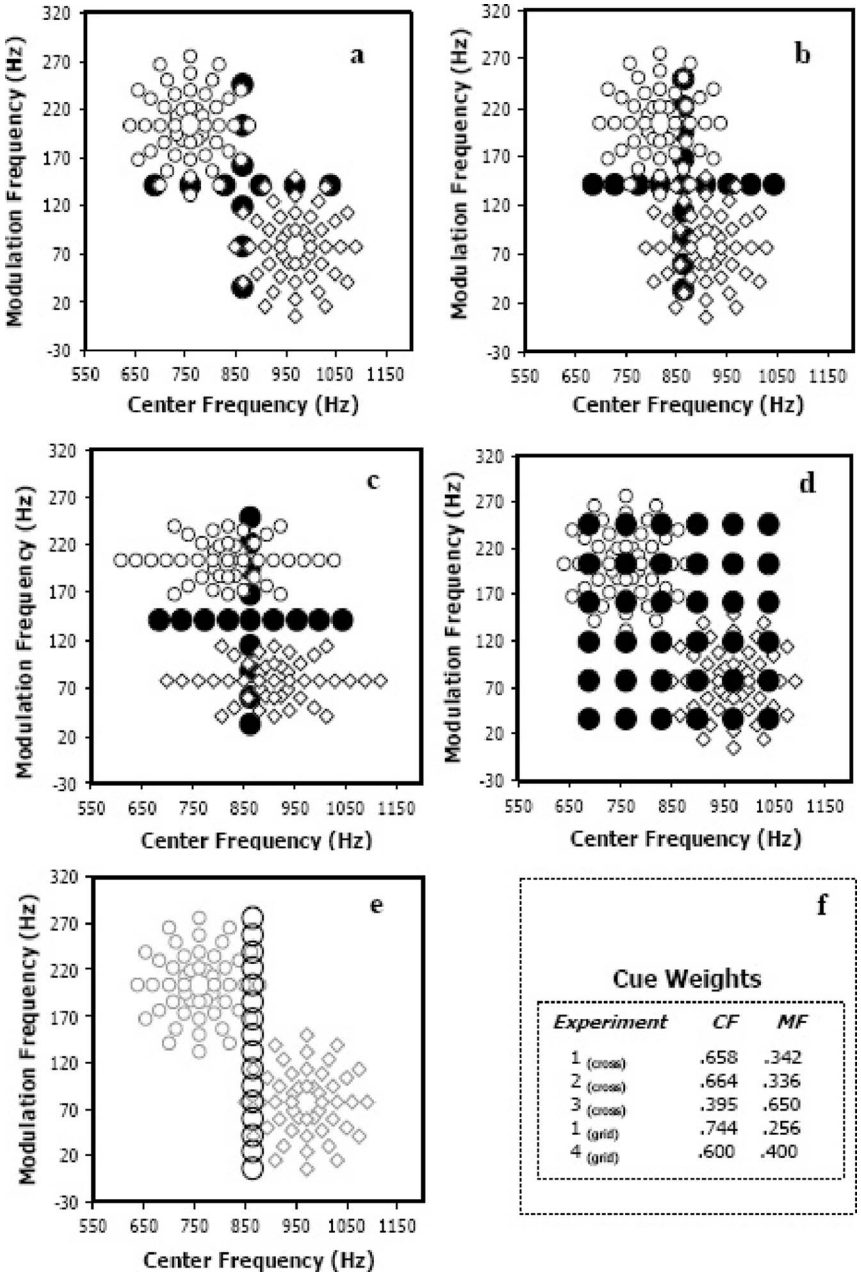

FIG. 1. Schematic representations of the stimulus input distributions in the center frequency (CF) by modulation frequency (MF) acoustic space. Open circles and open diamonds illustrate the individual stimuli from the two categories, respectively. The large filled circles show novel stimuli withheld from training from which cue weights were calculated. (a) Stimulus distributions for Experiment 1. (b) Stimulus distributions for Experiment 2. (c) Stimulus distributions for Experiment 3. (d) Stimulus distributions for Experiment 4. (e) Training distributions for Experiment 4 are shown as small light symbols for reference; they are identical to those of Experiment 1. The large open circles illustrate the stimuli presented during passive listening prior to categorization training. (f) Summary of the relative cue weights for the CF and MF dimensions across experiments.

stimuli landing on the criterion are considered "A" 50% of the time. Unlike more sophisticated ideal observer models, there is no noise in the encoding of the value on the dimension and the response is not probabilistic. With this decision model, one can calculate the informativeness of a dimension as

$$I_{\text{dimension}} = (\%\text{Correct Ideal Observer} - \text{Chance})/(\text{Perfect Performance} - \text{Chance}). \quad (1)$$

Equation (1) gives informativeness as the increase in accuracy over chance due to information in a dimension relative to the total amount of possible improvement. This function is equal to 1 when perfect performance can be achieved using that dimension and 0 when the dimension provides no information above chance. For Experiment 1, there are two possible responses and the prior probabilities of each response are equal; therefore, perfect performance is 100% and chance is 50%. By using either acoustic dimension alone, a perceiver could attain 95.8% correct with an optimal linear boundary [as can be seen in Fig. 1(a), the two categories are nearly perfectly linearly separable along either acoustic dimension with optimal boundaries at CF=865 Hz and MF =140 Hz]. Since the use of either acoustic dimension yields the same optimal percent correct, these acoustic dimensions are equally informative to the categorization task $I_{\text{CF}}=I_{\text{MF}}(95.8-50/100-50)=0.916$. However, integrating information across the two acoustic dimensions is more informative to the categorization response than either dimension individually; the categorization task can be performed perfectly if an ideal observer integrates across both acoustic dimensions (i.e., $I_{\text{CF+MF}}=1.0$).

In addition to the 96 stimulus exemplars (open symbols) comprising the categorization training stimuli, 12 novel test

stimuli (filled circles) were created to probe listeners' use of the CF and MF dimensions following training. These novel stimuli were reserved from categorization training to be used in generalization tests. They varied orthogonally in the acoustic space such that the value along one acoustic dimension was held constant at an intermediate value (CF =865 Hz or MF=140 Hz) while the values along the other acoustic dimension varied.[1]

Acoustic presentation was controlled by TDT System II hardware (Tucker-Davis Technologies, Alachua, FL). Stimuli were converted from digital to analog, low-pass filtered at 4.8 kHz, amplified and presented diotically over linear headphones (Beyer DT-150) at approximately 70 dB SPL(A).

3. Procedure

a. Categorization training. Listeners were tested individually in sound-attenuating booths. On each trial, listeners heard a single stimulus, pressed one of two unlabeled buttons to record a categorization response and received feedback as to the correct response via illumination of a light above the correct response button.

Listeners were instructed that the task was to determine whether each sound belonged to the left-most or the right-most response button. The experimenter also informed listeners that, although they would need-to guess at first, they should attempt to use the feedback given on each trial to guide later responses. Participants were encouraged to attempt to get as many responses correct as possible. Feedback assignment followed the symbols of Fig. 1; stimuli represented as open circles were assigned to one category (one response button) whereas those illustrated as open diamonds were assigned to the opposite category (and button). Left versus right button assignment to stimulus distributions was counterbalanced across participants.

Each block of training included a single presentation of each of the 96 training stimuli. There were ten blocks of training overall and the order of stimulus presentation was randomized within each block. After each block, listeners were able to take a brief break.

b. Test. Following training, listeners completed five more blocks of the categorization task. Within a block, stimulus presentation was randomized. These blocks were identical to the previous training blocks, except that in addition to the original 96 training stimuli, novel generalization stimuli were introduced. On these trials, listeners did not receive informative feedback; lights above all of the buttons lit following the response. This approach was chosen so that listeners' categorization of the generalization stimuli could be observed without encouraging participants to learn labels for the generalization stimuli across the course of testing. Listeners continued to receive feedback on trials for which familiar training stimuli were presented to assure that they continued to assign category labels in a manner that respected the distribution characteristics and feedback assignments present in training. This model of feedback assignment and mixing of novel and familiar stimuli is identical to paradigms successfully used in studying nonhuman animal category learning (see, e.g., Kluender *et al.*, 1998). Continuing to provide feedback for familiar stimuli, but not for novel stimuli, provides a safeguard against responses that drift away from learned categories as a consequence of the distribution differences that novel generalization stimuli introduce in the input. It also encourages participants to remain engaged in the task.

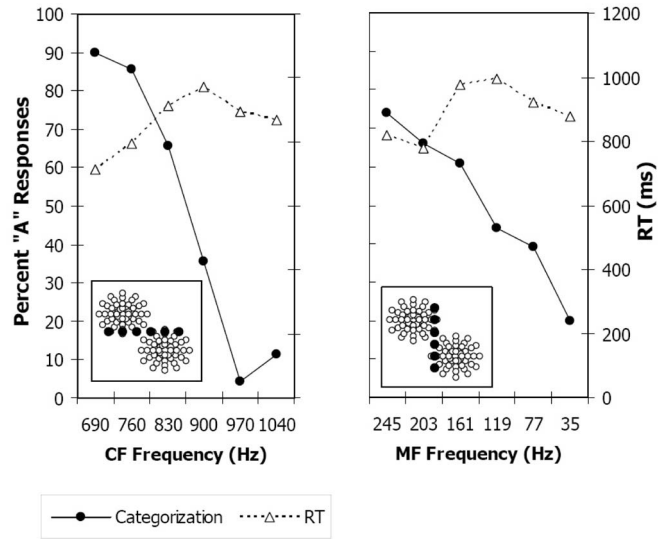

FIG. 2. Experiment 1 categorization responses and reaction time for generalization stimuli. Filled circles in the inset indicate the stimuli for which responses are plotted.

Participants completed both training and testing in a single experimental session lasting approximately 2 h.

B. Results and conclusions

The percent of "A" categorization responses was calculated whereby category "A" was arbitrarily assigned to be the category in the upper left corner of the two-dimensional acoustic space, indicated by open circles in Fig. 1(a). Overall, listeners performed fairly well in learning to categorize these unfamiliar sounds. Across the ten categorization training blocks, listeners' mean accuracy (i.e., responses consistent with feedback assignments) was $M=87.25\%$, $SE=0.016$. The orthogonal arrangement of the novel stimuli presented in the test blocks made it possible to assess categorization across one cue while the other was held constant. Figure 2 illustrates listeners' categorization responses and reaction times to the 12 generalization stimuli. The categorization functions averaged across listeners' responses to these novel stimuli hint that CF may have been more effective in cueing category identity because listeners were more adept at categorizing along the CF dimension than the MF dimension. An analysis of variance (ANOVA) of the percent "A" categorization responses as a function of the two orthogonal acoustic cues reveals that there was a statistically reliable Dimension × Stimulus Step interaction for both categorization response, $F(5,65)=8.45$, $p<0.0001$, $\eta_p^2=0.394$, and the reaction times, $F(5,65)=2.77$, $p=0.025$, $\eta_p^2=0.175$.

Cue weights were computed for each subject as the correlation between dimension values and percentage "A" responses across generalization stimuli. The absolute values of the correlation coefficients were normalized to sum to one.[2] These relative weights confirm the impression from the cat-

egorization functions of the dominance of the CF cue. The average weights were 0.658 for the CF cue and 0.342 for MF. The CF weight was significantly different from 0.5, meaning that listeners did not weight the two cues equivalently as may be expected given the match on informativeness.[3]

Two major conclusions may be made from the present study. First, listeners were able to learn to categorize novel acoustic stimuli effectively and did integrate information across the two dimensions (weights were above 0 for both dimensions). As such, this paradigm provides a useful means for examining cue weighting for categorization in a two-dimensional acoustic space. Second, despite the fact that the discriminability of the exemplars was psychophysically matched across the acoustic dimensions and despite the equivalent informativeness of the acoustic dimensions to the categorization task, listeners weighted CF more heavily than MF.

Of the 14 subjects, only one weighted MF higher than CF. Two participants used CF almost exclusively whereas the majority relied on both cues to some extent but weighted CF higher than MF. The reason for the dominance of CF is not immediately clear; we suggest some possibilities in the Sec. VI. However, it provides us with an opportunity to examine how changes in the training distribution characteristics may encourage listeners to shift their weighting functions.

III. EXPERIMENT 2

The aim of Experiment 2 was to determine whether listeners would make greater use of the MF cue in categorization under circumstances in which CF was rendered less informative. To accomplish this, the input distributions employed for Experiment 1 were shifted toward one another along the CF dimension, which resulted in greater overlap of the distributions on this dimension.

A. Methods

1. Participants

Fifteen volunteers recruited from Carnegie Mellon University participated for course credit or a small payment. All listeners reported normal hearing.

2. Stimuli

The two-dimensional stimulus space constructed for Experiment 1 was used again in Experiment 2, but the input distributions defining the two categories were moved closer to one another along the CF dimension. The centroids of the training distributions in Experiment 1 had CF values of 760 and 970 Hz (differentiated by seven stimulus steps or 210 Hz), respectively. In Experiment 2, these values were 820 and 910 Hz (three stimulus steps, 90 Hz). As a result of this stimulus manipulation, the category input distributions overlapped more along the CF dimension than the MF dimension. In other words, the CF dimension became less informative. By the informativeness metric introduced in Experiment 1, $I_{CF}=0.5$ (75% correct with optimal boundary). All other distribution characteristics were identical to those of Experiment 1. Therefore, the informativeness of the MF dimension was unchanged from Experiment 1 ($I_{MF}=0.916$).

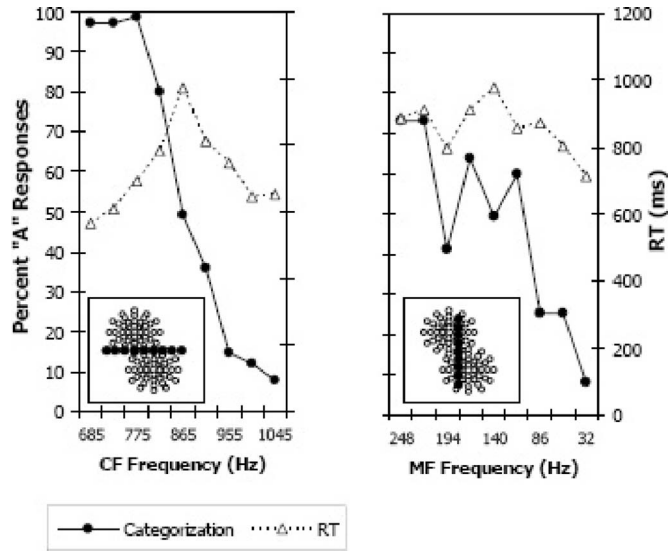

FIG. 3. Experiment 2 categorization responses and reaction time for generalization stimuli. Filled circles in the inset indicate the stimuli for which responses are plotted.

The new input distributions are illustrated as open symbols in Fig. 1(b). The generalization stimuli, which were identical to the stimuli from Experiment 1, are illustrated as filled circles in Fig. 1(b). To provide a finer sampling of categorization across the two cues, several additional stimuli were added to the novel stimulus set of Experiment 2; whereas there were 12 novel stimuli in Experiment 1, there were 17 in Experiment 2.

3. Procedure

The apparatus and procedures were identical to those used in Experiment 1.

B. Results and conclusions

Overall, listeners learned the categories very well, exhibiting a mean of 90.42% correct, $SE=1.46$, across the ten training blocks. Figure 3 illustrates listeners' categorization responses to the 17 generalization stimuli. Again, listeners' responses to the novel stimuli were more categorical along the CF dimension than the MF dimension. Supporting this conclusion, there was a significant Dimension × Stimulus Step interaction for both categorization responses, $F(8,112)=31.83$, $p<0.0001$, $\eta_p^2=0.695$ and reaction times, $F(1,112)=2.818$, $p=0.007$, $\eta_p^2=0.168$.

The average cue weights for the two dimensions were 0.664 and 0.336 for CF and MF, respectively. The relative weighting for CF was statistically equivalent to that obtained for Experiment 1, $t(27)=0.10$, $p=0.92$. Despite the reduced informativeness of CF for the task, listeners continued to rely on it in categorization judgments. Only two of the 15 subjects-gave greater weighting to the more informative cue (MF) and even in these cases the weights for CF and MF were approximately equal. Shifting the informativeness of the CF cue by moving the distribution means closer therefore was not effective in altering listeners' cue weighting.

IV. EXPERIMENT 3

The continued reliance on CF in Experiment 2 is surprising given that the observed pattern of cue weighting was a relatively inefficient weighting function for the training distributions. An obvious question is whether listeners' weighting functions are malleable at all. The goal of Experiment 3 was to present a categorization task that penalized a CF-dominant weighting function more than did the tasks of the previous two experiments. To accomplish this, we changed the variance of the training distributions in addition to changing the position of the centroids, as in Experiment 2.

A. Methods

1. Participants

Fourteen volunteers recruited from Carnegie Mellon University participated for course credit or a small payment. All listeners reported normal hearing.

2. Stimuli

The training distributions for Experiment 3 are shown in Fig. 1(c). The distribution centroids were identical to those of Experiment 2. In Experiments 1 and 2 the distributions spanned nine stimulus steps along each dimension. In Experiment 3, the variability along the CF dimension was increased (distributions spanned 15 stimulus steps) whereas the variability along the MF dimension was reduced (distributions spanned approximately five stimulus steps). The combined changes rendered MF a maximally informative cue for the categorization ($I_{\mathrm{MF}}=1.0$) and CF a marginally informative cue ($I_{\mathrm{CF}}=0.442$). The training distributions were composed of 86 unique stimuli. The generalization stimuli are illustrated as filled circles in Fig. 1(c) and are identical to those of Experiment 2.

3. Procedure

The procedure and apparatus were identical to previous experiments.

B. Results and conclusions

Across the ten training blocks, listeners exhibited a mean categorization accuracy of $M=78.1\%$, $SE=3.38$. Figure 4 illustrates listeners' categorization of the generalization stimuli by acoustic dimension. Unlike the previous experiments, there was no statistically significant dimension (CF versus MF) by stimulus step interaction in listeners' categorization responses, $F(8,104)=1.36$, $p=0.222$, $\eta_p^2=0.095$, or reaction times, $F(8,104)=1.73$, $p=0.10$, $\eta_p^2=0.117$. There was, however, a significant difference in overall reaction time, $F(1,13)=4.69$, $p=0.05$, $\eta_p^2=0.265$, with stimuli varying along the MF dimension categorized somewhat faster ($M=667.63$, $SE=43.01$) than stimuli varying along the CF dimension ($M=738.21$, $SE=68.5$).

The changes in distribution variances were effective in changing the preferred weighting of the dimensions. The average weight for CF was 0.395 with a corresponding MF weight of 0.65. The change in CF weight from Experiment 2 was significant, $t(27)=4.41$, $p<0.0005$. Eleven of the 14 subjects weighted MF greater than CF.

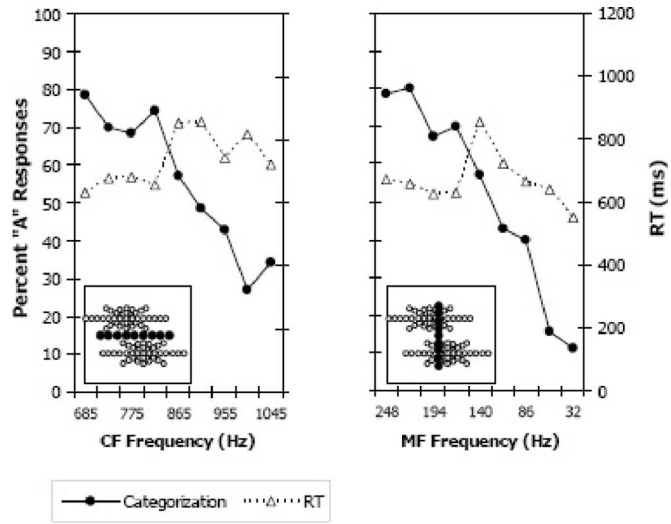

FIG. 4. Experiment 3 categorization responses and reaction time of categorization responses to novel generalization stimuli. Filled circles in the inset indicate the stimuli for which responses are plotted.

The results of Experiment 3 make it clear that weighting functions *can* be changed by the characteristics of the training distributions during a short learning session. Of theoretical and practical interest is the determination of which distributional characteristics can effect change in weighting functions. Distribution variance appears to be a better predictor of relative weighting than are measures of central tendency or our measure of informativeness. Whereas the shift in distribution centroids in Experiment 2 resulted in no change in weighting functions despite a large change in informativeness of the CF cue, the manipulations of variability in Experiment 3 led to a reversal of relative weighting with a smaller change in informativeness of the cues.

From the classic perspective that variability is the essence of information in perception, the lower weighting of the cue with greater within-category variability appears to be counterintuitive. For example, in experiments on sample discrimination in which listeners are asked to listen and respond to tones in a target frequency region, irrelevant tones in distant non-target frequency regions have little effect if they are fixed in frequency. However, when these distracter tones are allowed to vary in frequency, listeners' performance on the target task suffers dramatically (Lutfi, 1992; Neff and Odgaard, 2004). This effect can be thought of as a capture of selective attention by the variability of the irrelevant tones or it can be conceptualized as a greater weighting of highly variable components in the task, as in Lutfi's (1993) CoRE model. In the current experiment, the listeners seemed to be sensitive to increased variability across the CF dimension, as it led to a significant change in weighting functions. However, it was the lower-variance MF dimension that received the greater perceptual weight.

The details of sampling discrimination experiments and the current categorization experiments are too different to make direct comparisons that may illuminate the reason for

these different outcomes. However, the idea that the auditory system would be particularly sensitive to highly variable dimensions, *ceteris paribus*, is intuitively appealing. One may predict that exposure to variance along a dimension, with no attendant feedback, will result in this dimension becoming more heavily weighted. In Experiment 4 we investigated whether an acoustic dimension becomes more heavily weighted perceptually following exposure to its range of variability.

V. EXPERIMENT 4

Prior to categorization training in Experiment 4, listeners were exposed (without feedback) to stimuli varying along the MF dimension while CF was held constant at an intermediate value. It is predicted that this exposure independent of feedback for category membership will lead to a greater weighting of the MF dimension.

A. Methods

1. Participants

Fourteen volunteers recruited from Carnegie Mellon University participated for course credit or a small payment. All listeners reported normal hearing.

2. Stimuli

The distributions defining the categories were identical to those of Experiment 1. In addition, a set of generalization stimuli was created to more finely assess cue weighting across the acoustic space. The grid of filled symbols illustrated in Fig. 1(d) shows the position of these stimuli in acoustic space. This grid of novel stimuli, along with the cross of novel stimuli shown in Fig. 1(a), was presented to listeners in both Experiments 1 and 4.

In addition, another set of stimuli was created especially for Experiment 4. These stimuli were created as described for the previous experiments. They possessed a constant, intermediate CF frequency of 865 Hz and varied in MF from 5 to 275 Hz in 18-Hz steps. They are illustrated in Fig. 1(e) by the large open circle symbols.

3. Procedure

The apparatus and procedure were identical to those of Experiment 1. However, Experiment 4 participants completed a brief passive-listening segment before entering into categorization training. This segment lasted approximately 15 min. In that time, listeners heard the 16 stimuli illustrated in Fig. 1(e) as large open symbols 20 times each in random order. Listeners were instructed to simply listen to the sounds. Immediately after this passive exposure, listeners completed categorization training and generalization tests identical to those of Experiment 1. Thus, Experiments 1 and 4 differed only in the precategorization-training exposure.

B. Results and conclusions

A comparison of categorization accuracy during training across Experiments 1 and 4 reveals that the pretraining exposure to variability along the MF dimension had a significant influence on accuracy, $t(13)=2.19$, $p<0.05$. Across the ten training blocks, Experiment 4 listeners were significantly more accurate ($M=92.74\%$, $SE=0.02$) than Experiment 1 listeners ($M=87.25\%$, $SE=0.02$). Figure 5 illustrates listeners' categorization of the generalization stimuli cross stimuli [see Fig. 1(a)] by acoustic dimension.

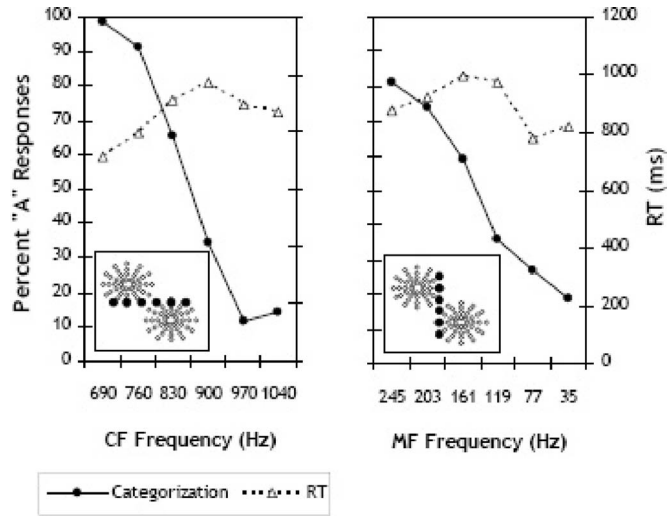

FIG. 5. Experiment 4 categorization responses and reaction time of categorization responses to novel generalization stimuli. Filled circles in the inset indicate the stimuli for which responses are plotted.

Calculations of cue weights also demonstrated that the passive exposure had an effect on categorization. Recall that Experiments 1 and 4 listeners also heard the grid of stimuli illustrated in Fig. 1(d). A contour plot illustrating listeners' categorization responses to the grid of novel generalization stimuli is shown in Fig. 6. When relative cue weights are calculated from responses to this grid of stimuli, Experiment 1 listeners exhibited a greater reliance on the CF dimension (0.744) than they did on MF (0.256). As predicted, the rela-

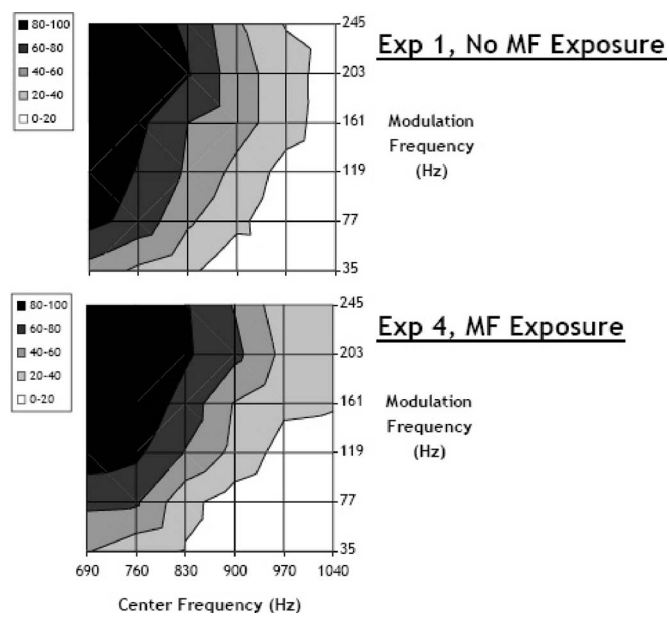

FIG. 6. Experiment 1 (no exposure to MF dimension) and Experiment 4 (exposure to MF dimension) categorization responses to generalization grid stimuli, as shown in Fig. 1(d), plotted as percent "A" category responses.

tive cue weight for CF was significantly lower for Experiment 4 (0.600) than for Experiment 1 (0.744), [$t(26)=1.75$, $p<0.05$, one tailed]. That is, preexposure to variability along the MF dimension with no feedback led listeners to rely upon MF in categorization responses moderately more than listeners who did not experience the preexposure (0.256 versus 0.400). These results suggest that the listeners were sensitive to the range of the MF dimension present in preexposure and treated it as a potential source of information. Note that this was a very conservative estimate of the effect of exposure to cue variability on cue weighting because the dependent variable was the categorization response to generalization stimuli presented at the end of categorization training. Still, a very short period of preexposure had an influence on cue weighting. It is quite possible that the categorization responses at the beginning of categorization training were more heavily weighted toward MF but that the preference for CF rebounded during training. The design of the current experiment makes it difficult to test this explicitly. However, the current evidence is suggestive of a role for unsupervised exposure with cue variability in determining weighting functions.

VI. GENERAL DISCUSSION

Although multiple cues define most auditory categories, including speech categories, how the auditory system integrates multiple acoustic dimensions in perceptual categorization is not yet well understood. The present work investigated cue weighting in auditory categorization, approaching the issue from the perspective that the perceptual system does not merely perceive acoustic dimensions and apply them as evidence for a particular category, but rather weights these dimensions as a function of characteristics of distributions of experienced sounds.

In the Introduction, we listed four variables that may affect the weighting functions for a categorization task. We summarize here the results of the experiments as they relate to these factors. The first of these potential factors was the informativeness of the cue for categorization. If the listener is weighting cues optimally then relative weightings should be predictable from relative informativeness (at least in terms of the rank ordering of the cues). The data from our experiments do not reveal such a direct relationship. In Experiment 1, CF received a much greater weighting despite the fact that the two cues were equal on our informativeness metric. CF continued to receive a higher weighting even when it was decreased in its informativeness to the categorization task in Experiment 2 by shifting the centroids of the distributions closer together on the CF dimension. We refer to a cue receiving a higher weight than dictated by its informativeness for the current task as having greater *salience* (cue *preference* or *bias* are other possible terms). Christensen and Humes (1996) also examined the categorization of nonspeech sounds differing on several dimensions and found that one of the cues tended to be weighted more despite equal informativeness for the task. However, their dimensions were not equated psychophysically for auditory step size. As a result, the weighting functions could have been reliably reflecting the informativeness of the psychoacoustic cues as opposed to the physical acoustic cues. In our experiments, the cue-weights were computed across equated psychoacoustic scales and thus reflect perceptual cue weighting.

What accounts for this salience? One possibility is that in the natural acoustic environment, CF is a more informative cue to auditory category identity than is MF. That is, in identifying a sound, the carrier frequency may provide more information or may be more reliably related to the distal source than is the rate at which the sound is modulated. Whereas we have no direct evidence that this is the case, it does have intuitive appeal. If this is the case, then listeners may come into the experiment with experience that is relevant to the task. A default higher weighting for CF could be a result of an innate predisposition (because the informativeness of CF has led to a fixed adaptation) or learned through experience. Similar patterns of differences in cue salience have been witnessed in studies of cross-modal integration. Battaglia, Jacobs, and Aslin (2003) report that visual cues are weighted more than auditory cues for spatial location even when the visual cues are not as reliably related to spatial location (see also Ernst and Banks, 2002, for an example of visual cue salience over haptic cues). Battaglia *et al.*. were able to model this influence by proposing that perceivers had *a priori* expectations of the reliability of visual information for spatial location (defined as a prior probability distribution for visual cue variance). Thus, perceivers appear to be weighting suboptimally within the context of the information in an experiment, but they may perceive optimally within the larger context of overall perceptual experience. The interaction of short-term (or local) informativeness and long-term (or global) informativeness will be an important area for future investigation.

A second factor that may affect cue weighting is distributional variance. Because our definition of informativeness is based on a criterion-based decision, variance and informativeness are not equivalent. In Experiment 3, we took the distributions from Experiment 2 and increased the within-category distribution variance along CF and decreased it along MF. The result was a small change in relative informativeness but a large change in the relative within-distribution variance for the two dimensions. Whereas the change in informativeness from Experiments 1 to 2 resulted in no change in perceptual cue weighting, the manipulation of within-category distribution variances led to a substantial change in weighting, with MF now the dominant perceptual cue. Because of the design of the experiments, it is not possible to establish exactly what caused the change in perceptual weighting from Experiment 2 to 3. The change in within-distribution variances resulted in changes in informativeness, the ratios of within- to between-distribution variance and correlations with the feedback for each cue. Further studies will be necessary to manipulate these factors independently. What is clear from these experiments is that changes in the central tendencies of the distributions (with concomitant changes in informativeness) do not appear to be effective in changing perceptual weighting functions, whereas changes in within-category variance can lead to significant perceptual change. Another open question is the influence of the pres-

ence of feedback on perceptual weighting. In Experiment 3, the CF dimension with greater within-category variance during training with feedback received the lower perceptual weighting. In Experiment 4, we presented the range of variability in MF with no feedback prior to the categorization training. This exposure resulted in a moderate increase in the relative weighting of the varied cue.[4] The present experiments do not allow us to fully dissociate the possible roles of variability along an acoustic cue with versus without explicit feedback for categorization; further studies will be necessary to investigate these manipulations independently. However, the present data are interesting in that they indicate that manipulation of distribution characteristics of acoustic categories plays a role in listeners' perceptual cue weighting.

A third factor that can affect cue weighting is how each dimension is represented by the auditory system. We attempted to account for some of the effects of auditory encoding by scaling our dimensions to equivalent auditory step sizes based on pilot testing of the stimuli. Whereas it is true that CF and MF are not independent in frequency modulation detection across a large range of values, we saw no evidence of interaction in our constrained ranges; we avoided very low MF values and created stimuli within a MF range where these dimensions are relatively independent (Demany and Semal, 1989). We did not model the effects of internal noise on the representation of our cue values. We correlated responses with the step size values of our stimuli as though these values are represented perfectly by the auditory system. The General Recognition Theory (GRT, Ashby and Townsend, 1986), a model of categorization developed mainly from visual categorization experiments, represents stimuli as multivariate probability distributions in perceptual space as opposed to points. Likewise, the COSS (conditioned-on-a-single-stimulus) analysis of perceptual weights developed for multicomponent psychoacoustic tasks (Berg, 1989) internal noise is an explicit parameter. It is likely that models of cue weighting in phonetic and other auditory categorization tasks will require a more detailed estimate of the auditory representation of the stimuli than we have used here.

The fourth potential factor affecting cue weighting is task. All of the experiments presented here were categorization tasks in which subjects were asked to maximize accuracy, so the present data do not speak to manipulating cue weighting through task adjustments. Nevertheless, task manipulations could be easily implemented in the present paradigm and would be particularly interesting to investigate in future research. In speech perception research, it is common to examine speech categories with categorization or discrimination paradigms. If one presumes that both paradigms access the same category representation, then they can provide converging evidence about auditory categories. However, if listeners modify their weighting functions to maximize performance given particular task constraints then discrimination and categorization data may not always coincide (e.g., Mirman *et al.*, 2004).

A. Implications for L2 phonetic acquisition

In the Introduction, we discussed the proposal that one of the obstacles to acquiring L2 phonetic categories could be a mismatch between weighting functions appropriate for L1 and L2. For example, Lotto *et al.* (2004) estimated the phonetic distributions for syllable-initial /r/ and /l/ from native-English productions and determined that the optimal weighting strategy for categorization was a very heavy perceptual weight on F3-onset frequency and a much lower weight on F2-onset frequency. In a follow up, Lotto *et al.* demonstrated that the optimal weighting pattern for the Japanese distinction between /w/ and /r/ (a distinction often considered to interfere with acquisition of the English liquid contrast) was just the opposite, with a higher weighting of F2-onset frequency. Thus, one may predict that Japanese listeners would have difficulty with the English contrast (and vice versa) because of the mismatch in learned weighting functions. In fact, perception and production studies both indicate that native Japanese listeners weight F2-onset higher than F3-onset for the English contrast (Yamada and Tohkura, 1990; Iverson *et al.*, 2003, 2005; Lotto *et al.*, 2004). This inappropriate weighting of F2 resembles the salience of CF in the current studies. Given the similarity, our results detailing effective and ineffective means of shifting weighting functions may have relevance for training L2 learners. In particular, the lack of weighting shift between Experiments 1 and 2 suggests that changing the informativeness of cues by shifting the average cue values (e.g., enhancing the difference in F3-onset values or decreasing the differences in F2-onset for /l/-/r/ stimuli presented to Japanese listeners) may not result in a significant change in categorization. On the other hand, the reversal of the perceptually dominant cue from Experiments 2 to 3 demonstrates that adding variance to an over-utilized cue (e.g., allowing F2 onset to vary independent of whether the sound is /l/ or /r/) may be an effective strategy to change listeners' weighting functions. In addition, the effect of pre-exposure to cue variance in Experiment 4, emphasizes the notion that the proper manipulation of variance may be essential to appropriate category acquisition.

In some earlier attempts to train adults on non-native contrasts the training sets did not contain much variance (e.g., Strange and Jenkins, 1978). As a result, the trainees were able to learn to discriminate the training stimuli but were not able to transfer this learning to new stimuli and contexts. In more recent attempts at L2 training, it has become clear that added variance in the training set (e.g., more speakers, more phonetic contexts) aids learning and generalization (Bradlow *et al.*, 1997; Jamieson and Morosan, 1989; Lively *et al.*, 1993). It is quite possible that a major benefit of high-variability training is that less informative cues will vary more within a category across multiple exemplars while more informative cues will be relatively more stable. As in Experiment 3, these changes in relative variance may be most effective at changing weighting functions. The importance of variance for category learning may even be seen in L1 acquisition. The speech that adults direct to their children ("motherese" or infant-directed speech) is actually *more* acoustically variable than speech directed to other adults

(Kuhl *et al.*, 1997). This seems counter intuitive unless one appreciates that infants must use the variance in speech input to determine which features are phonetically-relevant and which are not.

VII. SUMMARY

Of course, it is important to keep in mind that the categories of the present experiments were formed on the basis of an hour or so of experience whereas phonetic categories have the benefit of much greater experience, even among the youngest listeners. Nevertheless, we believe there is much to be gained for an understanding of phonetic cue weighting from laboratory studies that investigate cue weighting in general auditory (nonspeech) categorization. The purpose of the experiments presented here was to examine cue weighting in auditory categorization with well-defined training distributions for which experience could be entirely controlled. This allowed us to define the informativeness of each cue for the task and examine how changes in training distributions affect weighting functions. The long-term goals of this project are to establish a framework for testing models of cue weighting in phonetic categorization. It should be noted that there have been many models of perception, in general, and speech perception, in particular that have included cue weighting explicitly or implicitly. For example, Massaro's (1987; 1998) fuzzy logic model of perception (FLMP), Nosofsky's (1992) generalized context model (GCM), and connectionist models (e.g., TRACE, McClelland and Elman, 1986; Damper and Harnad, 2000) inherently weight cues in coming to a category decision. In order to provide a strong test of these models as a basis of speech perception, one needs good estimates of the distributions of phonetic category exemplars across the various dimensions. In addition to auditory cues, any full specification of phonetic categories would include visual cues, which can also be weighted in a categorization decision (Massaro, 1998). These kinds of estimates are rare.

We believe that nonspeech category tasks such as the ones presented here will play an important role in the development of models of auditory cue weighting. Whereas they do not have the ecological advantages of speech perception tasks, nonspeech tasks allow a level of stimulus and distribution control that allows one to investigate the constraints of the categorization process in detail.

ACKNOWLEDGMENTS

This work was supported by a National Science Foundation grant (NSF BCS-0078768) to L.L.H. and A.J.L. and by a James S. McDonnell Foundation award for Bridging Mind, Brain, and Behavior, to L.L.H. The authors thank Christi Adams for her essential role in conducting the experiments.

[1]An additional grid of 36 novel stimuli [pictured as filled circles in Fig. 1(d)] was also presented to Experiment 1 listeners for comparison with Experiment 4. Responses to these stimuli are discussed with the results of Experiment 4.

[2]It is typical in cue-weighting experiments in psychoacoustics and speech perception to normalize the weights (e.g., Berg, 1989; Christensen and Humes, 1996; Doherty and Turner, 1996; Lutfi, 1992). The result provides a comparison of the relative weighting of cues. Because this normalization scheme does not take into account the accuracy of listeners' categorizations, it can lead to problems with interpretation. For example, guessing randomly in Experiment 1 or responding perfectly with a rational weighting strategy would both result in equal normalized weights of 0.5. Given the high level of accuracy attained by listeners in the experiments presented here, we believe that this ambiguity is less of an issue than it may be for more complex tasks.

[3]All statistical tests for cue weights were performed only on the CF weights, as the MF weights are simply 1 minus the CF weight.

[4]These results may relate to an issue that we have avoided for the most part. There has been a tendency in the literature to equate cue weighting with selective attention. Some models of categorization explicitly equate weighting with attention (e.g., Nosofsky, 1986). The sensitivity of the auditory system to variability as demonstrated in psychoacoustic tasks (e.g., Lutfi, 2003; Neff and Odgaard, 2004) can be interpreted as variability capturing selective attention. The results of Experiment 4 also fit this interpretation. The change in weighting with variance manipulations between Experiments 2 and 3 demonstrate a possible dissociation between weighting and attention. Francis and Nusbaum (2002) have also called into question the viability of simple attentional models in phonetic category learning.

Ashby, F. G., and Townsend, J. T. (**1986**). "Varieties of perceptual independence," Psychol. Rev. **93**, 154–179.

Battaglia, P. W., Jacobs, R. A., and Aslin, R. N. (**2003**). "Bayesian integration of visual and auditory signals for spatial localization," J. Opt. Soc. Am. A **20**, 1391–1397.

Benguerel, A.-P. and Bhatia, T. K. (**1980**). "Hindi stop consonants: An acoustic and fiberscopic study," Phonetica **37**, 134–148.

Berg, B. G. (**1989**). "Analysis of weights in multiple observation tasks," J. Acoust. Soc. Am. **86**, 1743–1746.

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., and Tohkura, Y. (**1997**). "Training the Japanese listener to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production," J. Acoust. Soc. Am. **101**, 2299–2310.

Christensen, L. A., and Humes, L. E. (**1996**). "Identification of multidimensional complex sounds having parallel dimension structure," J. Acoust. Soc. Am. **99**, 2307–2315.

Crowther, C. S., and Mann, V. (**1992**). "Native language factors affecting use of vocalic cues to final consonant voicing in English," J. Acoust. Soc. Am. **92**, 711–722.

Crowther, C. S., and Mann, V. (**1994**). "Use of vocalic cues to consonant voicing and native language background: the influence of experimental design," Percept. Psychophys. **55**, 513–525.

Damper, R. I., and Harnad, S. R. (**2000**). "Neural network modeling of categorical perception," Percept. Psychophys. **62**, 843–867.

Demany, L., and Semal, C. (**1989**). "Detection thresholds for sinusoidal frequency modulation," J. Acoust. Soc. Am. **85**, 1295–1301.

Doherty, K. A., and Turner, C. W. (**1996**). "Use of a correlational method to estimate a listener's weighting function for speech," J. Acoust. Soc. Am. **100**, 3769–3773.

Ernst, M. O., and Banks, M. S. (**2002**). "Humans integrate visual and haptic information in a statistically optimal fashion," Nature (London), **415**, 429–433.

Flege, J. E., and Wang, C. (**1989**). "Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final /t/-/d/ contrast," J. Phonetics **17**, 299–315.

Francis, A. L., Baldwin, K., and Nusbaum, H. C. (**2000**). "Effects of training on attention to acoustic cues," Percept. Psychophys. **62**, 1668–l680.

Francis, A. L., and Nusbaum, H. C. (**2002**). "Selective attention and the acquisition of new phonetic categories," J. Exp. Psychol. Hum. Percept. Perform. **28**, 349–366.

Grieser, D., and Kuhl, P. K. (**1989**). "Categorization of speech by infants: Support for speech-sound prototypes," Dev. Psychol. **25**, 577–588.

Hazan, V., and Barrett, S. (**2000**). "The development of phonemic categorization in children aged 6-12," J. Phonetics **28**, 377–396.

Hillenbrand, J. M., Clark, M. J., and Houde, R. A. (**2000**). "Some effects of duration on vowel recognition," J. Acoust. Soc. Am. **108**, 3013–3022.

Holt, L. L., Lotto, A. J., and Diehl, R. L. (**2004**). "Auditory discontinuities interact with categorization: Implications for speech perception," J. Acoust. Soc. Am. **116**, 1763–1773.

Iverson, P., Hazan, V., and Bannister, K. (**2005**). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English

/r/-/l/ to Japanese adults. J. Acoust. Soc. Am. **118**, 3267–3278.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (**2003**). "A perceptual interference account of acquisition difficulties for non-native phonemes," Cognition **87**, B47–B57.

Jamieson, D., and Morosan, D. (**1989**). "Training new, normative speech contrasts: A comparison of the prototype and perceptual fading techniques," Can. J. Psychol. **43**, 88–96.

Jusczyk, P. W. (**1997**). *The Discovery of Spoken Language* (MIT Press, Cambridge, MA).

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," Lang. **60**, 286–319.

Kluender, K. R., Lotto, A. J., Holt, L. L., and Bloedel, S. B. (**1998**). "Role of experience in language-specific functional mappings for vowel sounds as inferred from human, nonhuman, and computational models," J. Acoust. Soc. Am. **104**, 3568–3582.

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., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., and Lacerda, F. (**1997**). "Cross-language analysis of phonetic units in language addressed to infants," Science **277**, 684–686.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (**1992**). "Linguistic experience alters phonetic perception in infants by six-months of age," Science **255**, 606–608.

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

Liberman, A. (**1996**). *Speech: A Special Code* (MIT Press, Cambridge, MA).

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (**1967**). "Perception of the speech code," Psychol. Rev. **74**, 431–461.

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.

Lively, S. E., Logan, J. S., and Pisoni, D. B. (**1993**). "Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories," J. Acoust. Soc. Am. **94**, 1242–1255.

Lotto, A. J., Sato, M., and Diehl, R. L. (**2004**). "Mapping the task for the second language learner: The case of Japanese acquisition of /r/ and /l/," *From Sound to Sense:* 50+ *Years of Discoveries in Speech Communication*, Research Laboratory of Electronics at MIT, Cambridge, MA.

Lutfi, R. A. (**1992**). "Informational processing of complex sound. III: Interference," J. Acoust. Soc. Am. **91**, 3391–3401.

Lutfi, R. A. (**1993**). "A model of auditory pattern analysis based on component-relative-entropy," J. Acoust. Soc. Am. **94**, 748–758.

Lutfi, R. A. (**2003**). "Perturbation analysis of acoustic cues for sound source identification," J. Acoust. Soc. Am. **113**, 2326.

Lutfi, R. A., and Doherty, K. A. (**1994**). "Effect of component-relative-entropy on the discrimination of simultaneous tone complexes," J. Acoust. Soc. Am. **96**, 3443–3450.

Massaro, D. W. (**1987**). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry* (Lawrence Erlbaum Associates, Hillsdale, NJ).

Massaro, D. W. (**1998**). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle* (MIT Press, Cambridge, MA).

Mayo, C. and Turk, A. (**2004**). "Adult-child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions," J. Acoust. Soc. Am. **115**, 3184–3194.

McClelland, J. L., and Elman, J. L. (**1986**). "The TRACE Model of Speech Perception," Cogn. Psychol. **18**, 1–86.

Mirman, D., Holt, L. L., and McClelland, J. M. (**2004**). "Categorization and discrimination of non-speech sounds: Differences between steady-state and rapidly-changing acoustic cues," J. Acoust. Soc. Am. **116**, 1198–1207.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., and Fujimura, O. (**1975**). "An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English," Percept. Psychophys. **18**, 331–340.

Morrongiello, B. A., Robson, R. C., Best, C. T., and Clifton, R. K. (**1984**). "Trading relations in the perception of speech by five-year-old children," J. Exp. Child Psychol. **37**, 231–250.

Neff, D. L., and Odgaard, E. C. (**2004**). "Sample discrimination of frequency differences with distracters," J. Acoust. Soc. Am. **116**, 3051–3061.

Nittrouer, S. (**2004**). "The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults," J. Acoust. Soc. Am. **115**, 1777–1790.

Nittrouer, S., Crowther, C. S., and Miller, M. E. (**1998**). "The relative weighting of acoustic properties in the perception of [s]+stop clusters by children and adults," Percept. Psychophys. **60**, 51–64.

Nittrouer, S., Miller, M. E., Crowther, C. S., and Manhart, M. J. (**2000**). "The effect of segmental order on fricative labeling by children and adults," Percept. Psychophys. **62**, 266–284.

Nosofsky, R. M. (1992). "Similarity scaling and cognitive process models," Annu. Rev. Psychology 43, 25–53.

Nosofsky, R. M. (**1986**). "Attention, similarity, and the identification-categorization relationship," J. Exp. Psychol. Gen. **115**, 39–57.

Parnell, M. M., and Amerman, J. D. (**1978**). "Maturational influences on the perception of coarticulatory effects," J. Speech Hear. Res. **21**, 682–701.

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.

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.

Simos, P. G., and Molfese, D. L. (**1997**). "Electrophysiological responses from a temporal order continuum in the newborn infant," Neuropsychologia **35**, 89–98.

Sinex, D., and McDonald, L. (**1989**). "Synchronized discharge rate representation of voice-onset time in the chinchilla auditory nerve," J. Acoust. Soc. Am. **85**, 1995–2000.

Sinex, D., McDonald, L., and Mott, J. (**1991**). "Neural correlates of non-monotonic 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.

Stevens, K. N. (**1989**). "On the quantal nature of speech," J. Phonetics **17**, 3–45.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).

Strange, W., and Jenkins, J. J. (**1978**). "Role of linguistic experience in the perception of speech," in *Perception and Experience*, edited by Pick, H. L. and Walk, R. D. (Plenum, New York), pp. 125–169.

Walley, A. C., and Carrell, T. D. (**1983**). "Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants," J. Acoust. Soc. Am. **73**, 1011–1022.

Werker, J. F., and Tees, R. C. (**1984**). "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life," Circuit Tree **7**, 49–63.

Yamada, R. A., and Tohkura, Y. (**1990**). "Perception and production of syllable-initial English /r/ and /l/ by native speakers of Japanese," *Proceedings of International Conference on Spoken Language Processing*, Kobe, Japan, 757–760.