Role of experience for language-specific functional mappings of vowel sounds

Keith R. Kluender[a)]
Department of Psychology, University of Wisconsin—Madison, 1202 West Johnson Street, Madison, Wisconsin 53706

Andrew J. Lotto
Department of Psychology and Parmly Hearing Institute, Loyola University—Chicago, 6525 North Sheridan Road, Chicago, Illinois 60626

Lori L. Holt
Department of Psychology, University of Wisconsin—Madison, 1202 West Johnson Street, Madison, Wisconsin 53706

Suzi L. Bloedel
Department of Audiology and Speech Pathology, VA Medical Center, 5000 West National Avenue, Milwaukee, Wisconsin 53295

(Received 28 January 1997; revised 16 July 1998; accepted 20 August 1998)

Studies involving human infants and monkeys suggest that experience plays a critical role in modifying how subjects respond to vowel sounds between and within phonemic classes. Experiments with human listeners were conducted to establish appropriate stimulus materials. Then, eight European starlings (*Sturnus vulgaris*) were trained to respond differentially to vowel tokens drawn from stylized distributions for the English vowels /i/ and /ɪ/, or from two distributions of vowel sounds that were orthogonal in the $F1-F2$ plane. Following training, starlings' responses generalized with facility to novel stimuli drawn from these distributions. Responses could be predicted well on the bases of frequencies of the first two formants and distributional characteristics of experienced vowel sounds with a graded structure about the central ''prototypical'' vowel of the training distributions. Starling responses corresponded closely to adult human judgments of ''goodness'' for English vowel sounds. Finally, a simple linear association network model trained with vowels drawn from the avian training set provided a good account for the data. Findings suggest that little more than sensitivity to statistical regularities of language input (probability–density distributions) together with organizational processes that serve to enhance distinctiveness may accommodate much of what is known about the functional equivalence of vowel sounds.
 [S0001-4966(98)00312-9]

PACS numbers: 43.71.An [WS]

INTRODUCTION

Two significant characteristics of the way listeners perceive speech sounds are that experience in a particular language environment has profound effects, and that some acoustic instantiations of a phoneme are perceptually more compelling or effective than others. Although this latter observation has been a common one ever since speech researchers first manipulated natural and synthetic speech signals, for a long while relatively little was made of this fact. Perhaps this was owing to the historical influence of ''categorical perception'' of speech sounds, by which within-category differences were considered largely irrelevant. A number of studies have revealed the importance of differences between different examples of the same phoneme. For example, some speech stimuli served as more effective adapters in selective adaptation studies (Miller *et al.*, 1983; Samuel, 1982), and some stimuli served as better competitors in dichotic competition experiments (Miller, 1977; Repp, 1977). More recent studies have incorporated explicit judgments of the degree to which particular stimulus is perceived as a good example of a particular phonetic segment (Grieser and Kuhl, 1989; Iverson and Kuhl, 1995; Kuhl, 1991; Miller and Volaitis, 1989; Volaitis and Miller, 1992).

The research effort reported here concerns developmental aspects of responding to vowel-sound distributions in a graded and language-specific manner. By the age of 6 months, infants respond to vowel sounds in a language-appropriate fashion even when stimuli overlap considerably along acoustic dimensions that are less directly relevant to vowel identity (Kuhl, 1983). Using a reinforced head turn paradigm, Kuhl trained infants to turn their heads only when the phonemic quality of a repeating background stimulus changed between the relatively similar synthesized vowels /ɑ/ and /ɔ/ modeled after male utterances. When tested on novel synthesized vowels /ɑ/ and /ɔ/ modeled after utterances by women and children (adding variation in pitch contour in addition to shifting absolute frequency of formants), infants provided the correct response as defined by phonemic (functional) equivalence despite talker and fundamental-frequency ($f0$) changes.

While this earlier study attests to the ability of infants to

a)Electronic mail: kluender@macc.wisc.edu

respond equivalently to vowels in the face of phonemically irrelevant variation, more recent studies by Kuhl and her colleagues (Grieser and Kuhl, 1989; Iverson and Kuhl, 1995; Kuhl, 1991) have investigated how responses to vowel sounds vary across acoustic/auditory dimensions that are directly relevant phonemically. In these cases, instances of the same vowel differing in acoustic/auditory dimensions seem not to be perceptually equivalent for either 6-month olds or adults. Using a reinforced head-turn paradigm, Grieser and Kuhl (1989) examined the extent to which six-month-old infants responded to a change from a repeating background /i/ stimulus to another variant of /i/ drawn from a distribution of /i/ examples. They found that the degree to which infants responded to a change from the background stimulus to another variant of the vowel was less when the background stimulus was a vowel judged by adult listeners to be near ideal or ''prototypical,'' Kuhl (1991) conceptualized this as a ''perceptual magnet effect'' and suggested that infants come to internalize vowel category prototypes similar to those for adults, and that variants of the vowel category are perceptually assimilated to the prototype or ''Native Language Magnet'' (Kuhl, 1993) to a greater degree than could be explained by psychophysical distance alone.

As might be expected, whether one of the comparison stimuli was a ''prototype'' or not, greater acoustic/auditory distance resulted in greater discriminability, and infants were generally more likely to respond when acoustic/auditory differences were greater. This fact makes the results a bit more difficult to interpret with regard to the process by which infants respond differentially. In a sense, the paradigm pitted the infant's ability to discriminate two vowel tokens against the infant's tendency to respond equivalently to discriminably different vowels that share some functional equivalence. By analogy, one would not wish to suggest that infants were incapable of detecting gender and age differences in Kuhl's (1983) experiments with the vowels /ɑ/ and /ɔ/. In any event, these studies (Grieser and Kuhl, 1989; Kuhl, 1991) demonstrate that infants were less likely to respond (indicating a stimulus change) when the background stimulus represented a relatively good example of the vowel /i/. Kuhl (1991) took this as evidence that there is an internal organization of phonetic categories around prototypic members that is an ontogenetically early aspect of the speech code.

This conclusion is consonant with the ubiquitous finding in psychological studies of categorization that instances of categories or concepts, whether dogs or birds or automobiles, are not equally exemplary. If one infers the existence and nature of internal representations for categories from responses on a variety of tasks, such representations would seem to have a graded structure—often described as being centered around an ideal or ''prototypical'' instance of the category (Rosch, 1975, 1978). For now, the present authors are agnostic with regard to the existence of internal representations for categories and are not prepared to require their existence when the data mostly consist of differential responses to functionally near-equivalent instances. Some reservations regarding the utility of posting representations such as phonetic categories will be conveyed in later discussion in this report. Here, the term ''category'' will be used only when it is necessary to portray the intentions of other investigators.[1] Also, the terms ''prototype'' and ''prototypical'' will be used only for consistency with formulation of these issues by others.[2] Instead, descriptions of stimulus materials will hew more closely to physical dimensions, and more neutral terminology such as ''functional equivalence'' or ''functional mapping'' will be used.

Considerations of terminology aside, Kuhl's measurements of infants' differential responses to contrasts between acoustically different instantiations of a given phoneme constitute an important step in understanding how infants come to perceptually organize sounds in a fashion appropriate to their language. More fine-grained analyses of the overall structure of infant functional mappings for vowel sounds (in contrast to establishing only the centroid or prototype) will be especially important, in part because, to a large extent, it is the hallmark of other studies of categorization that such equivalence classes have graded structures with some stimuli (not only prototypes) being better exemplars than others. Goodness judgments by adult listeners (Kuhl, 1991) suggest that, not only do equivalence classes for vowels have the appearance of being structured around a best example or prototype, but also that instances nearer to the best exemplar or prototype are ''better'' members of the category—an archetypal category structure. Analogous data has been collected for adult classification of consonants (e.g., Iverson and Kuhl, 1995; Massaro, 1987; Miller *et al.*, 1983; Miller and Volaitis, 1989; Samuel, 1982).

What has become apparent is that the degree to which infants treat instances of a vowel distribution equivalently is conditioned by their experience with a particular language. Evidence supporting a role for learning can be found in a study (Kuhl *et al.*, 1992) using the same paradigm as Grieser and Kuhl (1989; Kuhl, 1991) with infants from different language environments. Six-month-old infants raised in Swedish- and English-speaking environments exhibit quite different tendencies to respond to changes from a relatively good example to a relatively poor example of a vowel when tokens are drawn from a distribution corresponding to the Swedish high front rounded vowel /y/ versus a distribution corresponding to the English vowel /i/. Again, for both groups of infants, larger acoustic differences were detected more easily for both native and non-native[3] vowel sets. Importantly, however, English infants were much more likely to respond to differences between the relatively good ''prototype'' high front rounded Swedish vowel /y/ and variants of /y/ than they were to respond to differences between the relatively good ''prototype'' English /i/ and its variants. The complementary pattern was found for Swedish infants' responses. The fact that infants are less likely to respond differentially to examples of a vowel common within their language environment is taken as evidence that, by six months of age, infants have begun to treat similarly sounds that correspond to functional groupings in their native-language environment.

By contrast, Kuhl (1991) found that, for rhesus monkey subjects discriminating /i/ and /i/-like sounds in a task methodologically analogous to that used with infants, there was little or no evidence that relatively good /i/ stimuli are per-

ceived as any more similar to other vowel sounds drawn from a distribution of /i/ sounds than are relatively poor instances. Taken together, results with human infants and with monkeys have encouraged a number of researchers to propose that infants possess initial ''language-universal'' categories that are modified through exposure to the native language to become language-specific categories of adult language users (Miller and Eimas, 1996; for reviews, see Best, 1994; Werker, 1994). Most specifically, Eimas (1991) has argued that there exists an innately given, universal set of phonemes together with processes that enable the infant to map acoustic variants onto phonemic category representations. In addition to being consistent with traditional nativist accounts of language competence (e.g., Halle, 1990; Pinker and Bloom, 1990), innate phonetic categories may be congenial to some essentialist accounts of concepts more generally (Atran, 1987; Gelman and Wellman, 1991; Keil, 1987; Medin and Ortony, 1989).

Kuhl (1991) was initially circumspect with regard to such questions, presenting two potential explanations: ''First, infants at birth could be biologically endowed with mechanisms that define vowel prototypes for certain vowels (e.g., the 'quantal' vowels) or for all of the vowels in all the languages of the world . . . A second alternative is that the effects are due to experience in listening to a specific language. (p. 105)'' More recently, Kuhl (1993) suggests that native-language prototypes are most likely the product of early experience in a language environment. Relatively little has been revealed, however, about putative processes by which learning and experience would shape development of functional (phonemic) equivalence among vowel sounds if, in fact, functional equivalence can be learned. It is necessary to elucidate the processes by which functional equivalences for vowel sounds might arise through experience and learning if they are to arise *de novo*. If equivalence classes (phonetic categories) for different acoustic instantiations of vowels can be a function of experience and general processes of learning, what are the salient characteristics of the resulting response structure? In particular, do patterns of responses to learned vowel equivalence classes bear close resemblance to response patterns measured for infant and adult humans?

The use of nonhuman animal subjects afforded Kuhl (1991) the opportunity to assess vowel discrimination in a model unfettered by extensive experience with distributional properties of vowel sounds. In the present studies, animal models are used to address explicitly questions relating to experience with vowel sounds when language-specific properties are strictly controlled. The aim is to understand better the nature of explicitly learned equivalence classes for vowel sounds to afford comparison with extant measures from human infant and adult listeners. The nonhuman species used is European starlings (*Sturnus vulgaris*), a bird that has been demonstrated to have hearing comparable to humans within the frequency range of human vowel sounds (Dooling *et al.*, 1986; Kuhn *et al.*, 1980, 1982) and appears to share a common mechanism of spectral analysis with many other vertebrates including humans (Dooling *et al.*, 1986).

Because nonhuman animals have been shown to be reasonably adept at responding differentially when presented

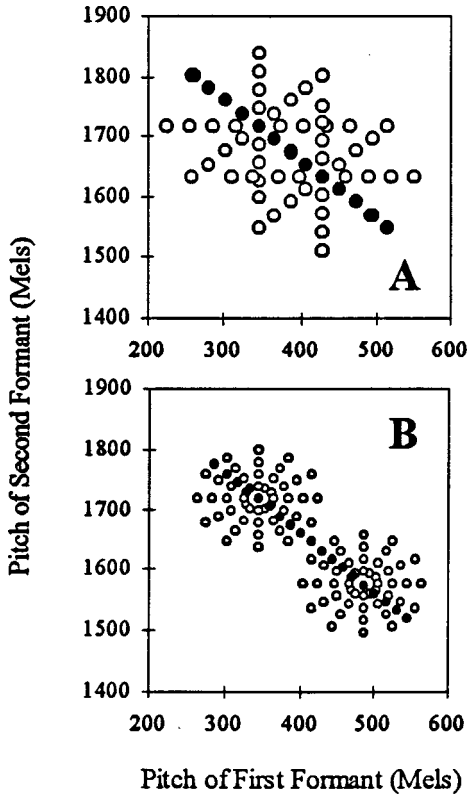

FIG. 1. In the top panel (a), 13 stimuli used in preliminary study (Lotto *et al.*, in press) are represented by filled circles and plotted in mel coordinates corresponding to synthesizer frequency values for $F1$ and $F2$. All circles, filled and unfilled, correspond to stimuli used by Kuhl (1991). In the bottom panel (b), 19 vowel stimuli used in experiment 1 are plotted as filled circles in a mel-scaled $F1$–$F2$ space. Unfilled circles correspond to $F1$ and $F2$ values for /i/ and /ɪ/ stimuli used in experiments 2 and 3.

with contrasts between speech sounds (Kluender, 1991; Kluender *et al.*, 1987; Kluender and Lotto, 1994; Kuhl and Miller, 1975, 1978; Kuhl and Padden, 1982, 1983) including vowel sounds (Burdick and Miller, 1975; Kluender and Diehl, 1987), it is not enough simply to demonstrate that starlings can respond differentially to different vowel sounds. If general processes of learning serve to explain perceptual development of speech perception by human infants, then one needs to demonstrate that responses to vowel equivalence classes learned by nonhuman subjects bear close resemblance to response patterns measured for infant and adult humans.

I. EXPERIMENT 1

The present effort began with synthesis of stimuli in accord with the descriptions given by Grieser and Kuhl (1989) and Kuhl (1991). In a preliminary experiment (Lotto *et al.*, in press) a series of stimuli was drawn from their two overlapping distributions of vowel sounds. [See top panel (a) of Fig. 1.] With the exception of durational differences, stimuli were synthesized in accordance with their descriptions. Tokens for each of Kuhl's (1991) distributions lay on eight spokes radiating from a centroid in a mel-scaled $F1$–$F2$ space. Lotto *et al.* (in press) used only the 13 stimuli along the diagonal (filled symbols). Sixteen listeners were asked to judge the quality of these 13 vowel sounds with

respect to whether each sounded most like the vowel in ''heat,'' ''hat,'' ''hate,'' ''hit,'' ''head,'' ''hood,'' or ''none of the above.'' The most important feature of subjects' reported percepts is that at least two of the stimuli from Grieser and Kuhl's (1989) and Kuhl's (1991) /i/ distribution [top panel (a) of Fig. 1] typically were not perceived as /i/ by this group of listeners. Much more common for these sounds were percepts of /ɛ/, /e/, and /ɪ/. Because stimuli were not included from other spokes from Kuhl's (1991) distribution, these data do not address the degree of which other stimuli drawn from that distribution would be perceived as /i/. This observation that some of the tokens intended to be perceived as /i/ by Grieser and Kuhl (1989) and by Kuhl (1991) are not perceived as /i/ is consistent with earlier reports (Iverson and Kuhl, 1995; Lively, 1993; Sussman and Lauckner-Morano, 1995).

In the interest of employing a set of stimuli that would constitute a distribution of reasonably compelling instances of the vowel /i/, a second series of stimuli were synthesized. In order to better delineate a range of acceptable tokens of the vowel /i/, these stimuli were presented to naive listeners for identification.

A. Method

1. Subjects

Sixteen college-age adults served as subjects. For all experiments reported here involving human objects, individuals learned English as their first language and reported normal hearing. All subjects received Introductory Psychology class credit for their participation.

2. Stimuli

Nineteen five-formant vowel stimuli were synthesized using the cascade branch of the Klatt (1980) software synthesizer implemented in CSRE (CSYNTR16; Jamieson *et al.*, 1992) on a microcomputer with 12-bit resolution at a 10-kHz sampling rate and were stored on computer disk. Stimuli were synthesized with parameters chosen from along a diagonal in a mel-scaled *F*1–*F*2 space [see filled circles in bottom panel (b) of Fig. 1]. In contrast to earlier efforts and in the interest of better circumscribing a region of perceptually acceptable instances of /i/, stimuli were spaced only 20 mel apart along the diagonal. The diagonal was at 45° relative to the mel-scaled *F*1–*F*2 plane, so the *F*1 and *F*2 mel values of each stimulus were of equal increments or decrements relative to adjacent stimuli. The fifth stimulus from the most extreme (low *F*1, high *F*2) end of the diagonal shared the same *F*1 and *F*2 values as the centroid of the /i/ distribution used previously (Grieser and Kuhl, 1989; Kuhl, 1991) and conformed to mean values for male talkers measured by Peterson and Barney (1952). Center frequencies (Hz) and mel values for *F*1 and *F*2 are listed in Table I. Synthesizer values for *F*3, *F*4, and *F*5 were held constant at 2780, 3300, and 3850 Hz, respectively. Formant bandwidths, *B*1, *B*2, *B*3, *B*4, and *B*5, were 50, 70, 110, 250, and 200 Hz, respectively. Duration of each stimulus was 300 ms. Although Grieser and Kuhl (1989) and Kuhl (1991) used 500-ms stimuli, and Iverson and Kuhl (1995) used 435-ms stimuli, 300 ms was chosen as a reasonable compromise between those rather extreme durations and shorter more natural durations. Fundamental frequency was held constant at 120 Hz. A 25-ms linear amplitude ramp was imposed on the beginning and end of each stimulus.

TABLE I. Synthesis parameters for first and second formants of stimuli used in experiment 2 depicted as both Hz and mel.

Hertz		mel	
*F*1	*F*2	*F*1	*F*2
221	2421	288	1775
233	2388	303	1760
246	2355	317	1746
258	2322	331	1732
270	2290	345	1718
283	2258	359	1704
295	2226	373	1690
308	2194	387	1676
321	2163	402	1661
334	2132	416	1647
347	2102	430	1633
360	2071	444	1619
374	2041	458	1605
387	2011	472	1590
401	1982	486	1576
415	1953	500	1562
429	1924	515	1548
443	1896	529	1534
457	1868	543	1520

3. Procedure

Given the intended application for these stimuli (experiment 2) and the present emphasis upon the range of acceptable /i/ stimuli, a forced-choice identification task was used. Subjects were asked to identify stimuli as /i/ or as /ɪ/. The choice of /ɪ/ as an alternative was based upon the authors' perception of many of these shorter (300 vs 435 ms) stimuli being better examples of /ɪ/ than of /e/ or /ɛ/. Stimulus presentation was under control of a microcomputer. Following D/A conversion (Ariel DSP-16), stimuli were low-pass filtered (Frequency Devices 677, cutoff frequency 4.8 kHz) prior to being attenuated (Analog Devices AD7111 digital attenuator), amplified (Stewart HDA4), and played over headphones (Beyer DT-100) at 70 dB SPL. Calibration of presentation level was achieved by first matching the rms level of all stimuli to a 1-kHz tone prior to D/A conversion. Subjects were instructed to press either of two buttons labeled ''heat'' and ''hit'' to indicate which of these alternatives best characterized the vowel sound heard on a trial.

B. Results

Data pooled across 16 listeners are presented in Fig. 2. The first nine stimuli were labeled as /i/ quite reliably (greater than 85% of presentations). As *F*1 and *F*2 values increase and decrease, respectively, more stimuli are identified as /ɪ/, not /i/. Based upon this distribution of responses, it was now possible to construct with confidence distributions of /i/ and /ɪ/ exemplars for presentation in the following equivalence class learning experiment.

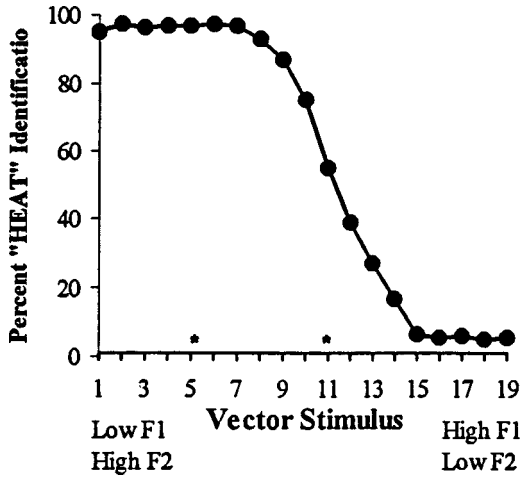

FIG. 2. Identification data from 16 listeners responding ''heat'' or ''hit'' when identifying 19 vowel stimuli in experiment 1 and depicted in the bottom panel (b) of Fig. 1. Stimuli 5 and 11 (indicated by *) share *F*1 and *F*2 values with ''prototype /i/'' and ''nonprototype /i/'' stimuli, respectively, used by Kuhl (1991).

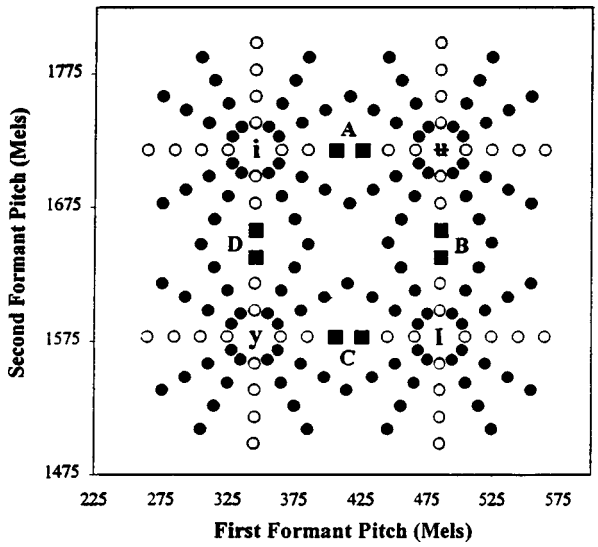

FIG. 3. Mel-scaled plot of 196 vowel stimuli synthesized for experiment 2 representing equal 49-token distributions of the English vowels /i/ and /ɪ/ and vowels approximate to /y/ and /ʉ/. Filled circles represent stimuli used in training. Unfilled circles, filled squares, and the symbols /i/, /ɪ/, /y/, and /ʉ/ (centroids) correspond to stimuli withheld until the testing phase of experiment 2. Squares labeled A, B, C, D correspond to pairs of stimuli used in comparison of between- and within-distribution response strengths.

II. EXPERIMENT 2

Experiment 2 was designed to answer the primary question addressed in this report. If equivalence classes for different acoustic instantiations of a given vowel can be a function of experience and learning, what are the salient characteristics of the resulting response structure? In particular, are response gradients acquired through learning comparable to response gradients measured for infant and adult human listeners? Animal studies of speech perception have been used to assess auditory processes without confounds of effects of experience (e.g., Dooling *et al.*, 1995; Kluender and Lotto, 1994; Kuhl, 1981, 1986, 1991). In contrast, the present study is designed explicitly to engage processes of learning in an animal for which experience with speech sounds can be precisely controlled. European starlings (*Sturnus vulgaris*) were trained to respond differentially to stimuli drawn from distributions of vowel sounds representative of English vowels, /i/ and /ɪ/, or from distributions constructed to be orthogonal to the /i/ and /ɪ/ distributions in a mel-scaled *F*1–*F*2 plane. These orthogonal distributions roughly correspond to high front rounded vowel /y/ and high mid rounded vowel /ʉ/ like those occurring in Swedish. Half of the birds were assigned as /i–ɪ/ birds, and half were assigned as /y–ʉ/ birds.

A. Method

1. Subjects

Eight European starlings (*Sturnus vulgaris*) served as subjects in the learning experiment. Free-feed weights ranged from 66 to 102 g.

2. Stimuli

A total of 196 vowel stimuli were synthesized representing equal 49 token distributions of the English vowels /i/ and /ɪ/ and of the two orthogonal distributions /y/ and /ʉ/. Distributions for /i/ and /ɪ/ vowels alone are represented in the bottom panel (b) of Fig. 1, and distributions for all four vowels are shown in Fig. 3. From experiment 1, it is inferred that all 49 examples of /i/ were reasonably good versions of the English vowel /i/. The centroid of /ɪ/ was synthesized with values very close to the mean values for /ɪ/ measured by Peterson and Barney (1952), and the authors perceived all members of the /ɪ/ distribution to be acceptable versions of the English vowel /ɪ/. However, owing partially to the duration of the stimuli, a few instances were not particularly compelling versions of a lax vowel. Centroids for /y/ and /ʉ/ were determined on the basis of considerations other than appropriateness as exemplars of vowels from Swedish or any other language. Instead, centroids for /y/ and /ʉ/ were chosen so that the cluster of four distributions fulfilled a number of experimental desiderata including denser sampling, orthogonality, and evaluation of discrimination versus functional equivalence (categorization).

Of course, none of the vowels closely mimic realistic productions representative of infant experience. Steady-state vowels, with variations of only *F*1 and *F*2 and excluding diphthongal patterns, consonantal contexts, and durational differences, may be pale imitations of the real thing; however, static monophthongal vowels are consistent with previous studies addressing the same and related questions. Although formant values for /y/ closely approximate stimuli used for Kuhl *et al.* (1992), it is unlikely that any of the stimuli in this distribution would constitute particularly good examples of Swedish /y/ for two reasons. First, for Swedish, /y/ is heavily diphthongized, and these sounds are monophthongal. Second, high front rounded vowels have relatively low-frequency *F*3, and the range of *F*2 frequencies used across the distribution preclude the use of *F*3 values appropriate for a high front rounded vowel. Both of these reservations hold for the synthetic versions of Swedish /y/ used by Kuhl *et al.* (1992).

The four distributions of stimuli differed in several ways from the original Grieser and Kuhl (1989) and Kuhl (1991) stimuli, First, 12 spokes of stimuli, instead of 8, emanated from the centroids. Second, stimuli were synthesized along each spoke with *F*1 and *F*2 frequency values corresponding to four 20-mel increments. As seen from experiment 1, the smaller step size afforded a more realistic approximation of the perceptually acceptable area in the *F*1–*F*2 plane for a given vowel sound. Half again as many spokes and the smaller step size together contributed to more compact distributions that provided a denser sampling of the perceptual space.

There are two other important aspects of these stimulus distributions that bear note. First, each pair of vowel distributions (/i/–/ɪ/, /y/–/ʉ/) is orthogonal to the other in a mel-scaled space. One virtue of this arrangement is that any confounds related to predispositions of the auditory system or to effects from experience with other sounds can be detected or eliminated.

Second, vowel pairs overlap sufficiently to assess separately the contributions of discrimination versus functional equivalence. This is because some subsets of stimuli that require differential responding by half the subjects do not require differential responding to the other half of the subjects. For example, one can see from stimuli marked by filled squares in Fig. 3 that stimuli to which /i–ɪ/ birds should respond differentially (A vs B and C vs D) do not require differential responding by /y–ʉ/ birds (A,B both /ʉ/ and C,D both /y/). These comparisons afford direct measurement of whether subjects respond similarly due to functional equivalence or due to lack of discriminability.

All stimuli for experiment 2 were synthesized with the same values for duration, amplitude contour, *f*0 contour, formant bandwidth, *F*3, *F*4, and *F*5 as stimuli from experiment 1. Formant-frequency values for *F*1 and *F*2 at the centroids for the /i/ were 270 and 2290 Hz (344.8 and 1718.1 mels as in experiment 1), and 389 and 1986 Hz (484.8 and 1578.1 mels) for /ɪ/. First and second formant values for /ɪ/ differ minimally from Peterson and Barney (1952) average values of 390 and 1990 Hz for men. Values of *F*1 and *F*2 at the centroids were 270 and 1986 Hz (344.8 and 1578.1 mels) for /y/, and 389 and 2290 Hz (484.8 and 1718.1 mels) for /ʉ/. Formant frequencies for the other 48 stimuli for each distribution were placed at 20-mel intervals measured from the centroid, 4 on each of the 12 spokes for each distribution.

3. Procedure

Birds were first trained by means of operant procedures to peck differentially to vowels either drawn from distributions for /i/ or /ɪ/, or drawn from distributions for /y/ and /ʉ/. Following 5 to 20 h of food deprivation (adjusted to each bird individually for optimal performance[4]), birds were placed in a sound-proof operant chamber (Industrial Acoustics Corp. AC1) inside a larger single-wall sound-proof booth (Suttle Acoustics Corp). In a go/no-go task, birds pecked a single lighted 1.2-cm-square key located 15 cm above the floor and centered below the speaker. For two of the /i–ɪ/ birds, pecks to /i/ were positively reinforced, while, for the other two, pecks to /ɪ/ were positively reinforced. For two of the /y–ʉ/ birds, pecks to /y/ were positively reinforced, while, for the other two, pecks to /ʉ/ were positively reinforced. Stimuli were presented, responses were recorded, and reinforcement was controlled by a 80286 microcomputer with an Ariel DSP16 A/D–D/A board and custom parallel I/O.

On each trial, a single vowel sound was presented repeatedly once per 1.3 s at an average A-weighted peak level of 70 dB SPL measured at the approximate location of the bird's head (Bruel & Kjaer type 2232). Stimuli were equated for rms energy level prior to attenuation. On a trial-by-trial basis, the intensity of the sound was varied randomly from 70 dB by ±0–5 dB [mean=70 dB SPL] through a computer-controlled digital attenuator (Analog Devices 7111). This roving intensity level mitigated the opportunity for responding correctly on the basis of relative loudness. Average duration of each trial was 30 s, varying geometrically from 10 to 65 s. Intertrial interval was 15 s. No sound or light (other than normal chamber illumination) was presented during the intertrial interval. Responses to positive stimuli were reinforced on a variable interval schedule by 1.5–2.0 s access to food from a hopper beneath the peck key. Duration of hopper access was adjusted for each bird for consistent performance across a session. Average interval to reinforcement was 30 s (10 to 65 s), so that positive stimuli were reinforced on an average of once per trial. Note that when a trial was long (e.g., 57 or 65 s) and times to reinforcement were short (e.g., 10 or 12 s), reinforcement was available more than once. Likewise, on shorter positive trials, reinforcement did not become available if time to reinforcement was longer than the trial. Such intermittent reinforcement encouraged consistent peck rates during subsequent non-reinforced test trials. During negative trials, birds were required to refrain from pecking for 5 s in order for presentation of the stimulus to be terminated.

Following magazine training and autoshaping procedures, reinforcement contingencies were gradually introduced over a one-week period in sessions of 60 to 72 trials each. During that first week: mean amplitude of the stimuli was increased from 35 to 70 dB SPL in order to introduce the sound without startling the birds; average trial duration increased from 5 to 30 s; intertrial interval decreased from 40 to 15 s; average time to reinforcement was increased from 5 to 30 s; access to the food hopper was decreased from 4.0 to 2.0 s; and the ratio of positive to negative trials decreased from 4:1 to 1:1.

Birds were trained first to respond differentially to a subset of 64 of the sounds included in their respective pairs of vowel distributions. Training stimuli are represented as filled symbols in Fig. 3. Some stimuli (unfilled symbols and filled squares), including the centroids of the distributions, were withheld from presentation during the training phase of the study. These stimuli were reserved for the test phase in order to be used as novel exemplars to assess the degree of generalization to novel tokens and to assess the response structure in a way that is unconfounded with history of reinforcement. All birds learned quickly to respond correctly to training tokens of /i/ versus /ɪ/, or /y/ versus /ʉ/, pecking at least twice as often to positive stimuli versus negative stimuli

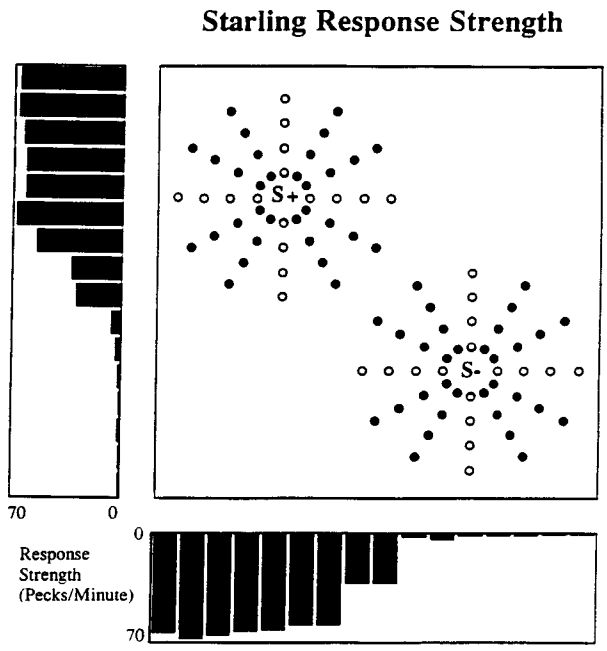

FIG. 4. From experiment 2, adjusted peck rates as a function of $F1$ and $F2$ values are plotted as histograms on x ($F1$) and y ($F2$) axes following rotation/reflections to align responses in the /i/–/ɪ/ diagonal. Bar heights correspond to mean peck rates for stimuli with a given $F1$ or $F2$ value following scaling to each bird's maximum peck rate for any stimulus.

at the end of 80 days of training (5120 trials). Birds continued to be trained with the subset of representatives of their distributions for a total of 101 training sessions.

The eight birds were then tested on novel examples [the centroids and other stimuli that had not yet been presented from the birds' respective distributions (/i/–/ɪ/, /y/–/ʉ/)]. A subset of previously reinforced training stimuli (eight from each vowel-sound distribution) also were tested as test stimuli in this second stage of the experiment to make possible comparisons between experienced exemplars and novel instances of the distributions. Across 50 daily sessions, all 50 test stimuli (34 novel+16 non-novel) were presented 20 times each. During a single test session, 20 novel stimuli were presented individually in 30-s trials. During presentation of novel stimuli, no contingencies were in effect. Birds neither received food reinforcement nor needed to refrain from pecking in order for presentation to terminate after 30 s. Trials with novel stimuli were interspersed among the 64 reinforced trials using non-novel training stimuli. Test trials could not occur until after 15 non-novel stimulus trials had been presented. This assured that each bird ''settled in'' to the task before responding to test stimuli.

B. Results

Data for all birds across four conditions are displayed in Fig. 4. For each subject, the two highest and two lowest response rates to a given stimulus were not entered into the analyses.[5] Whether birds were reinforced for pecking to /i/, /ɪ/, /y/, or /ʉ/, the same basic patterns of data were seen. There were no systematic differences between /i–ɪ/ and /y–ʉ/ birds, nor were there any systematic differences as a consequence of which vowel in a pair was designated positive. Consequently, in order to evaluate performance across the eight avian subjects, data as a function of $F1$ and $F2$ values were reflected and/or rotated to align positive and negative vowel clusters in the $F1$–$F2$ plane. Data for cases when /ɪ/ was positive were rotated 180 degrees to conform with data for cases when /i/ was positive. For /y/ and /ʉ/, a reflection is required to meet the same end. Values from cases for which /y/ was positive were reflected over an $F2$ axis separating /y/ from /i/. Values from cases for which /ʉ/ was positive were reflected over an $F1$ axis separating /ʉ/ from /i/. Analogous reflections were performed prior to analysis for negative categories. Finally, in order to normalize for individual differences in peck rates, mean peck rates in pecks per minute were converted to percentages of the maximum mean peck rate measured for each bird in response to any test trial.

Multiple linear regression analyses were conducted separately for peck rates to novel positive stimuli and to novel negative stimuli. Three independent variables were entered into the multiple regression analyses: $F1$ value (mels); $F2$ value (mels); and distance from centroid of the distribution (mels). These dimensions are orthogonal, thus avoiding many of the usual concerns regarding multivariate measures. For stimuli to which birds were reinforced for pecking (positive), all three variables contributed significantly to prediction of peck rate. The three-variable regression was statistically significant ($F\text{ratio}_{3,132}=26.37$, $p<0.0001$, multiple $R=0.61$). The value of $F2$ had the greatest contribution ($r=0.52$, $p<0.001$) followed by $F1$ value ($r=-0.28$, $p<0.001$) followed by distance from the centroid ($r=-0.14$, $p<0.05$). Using as an example the two birds for which /i/ was the positive vowel, regression analysis indicates that birds pecked most vigorously in response to stimuli with higher $F2$ and lower $F1$, and overlaid upon this pattern is a tendency to peck more rapidly to stimuli closer to the centroid of the distribution of /i/ tokens. The same pattern was seen for each vowel distribution: highest rates for high $F1$ and low $F2$ for /ɪ/, low $F1$ and low $F2$ for /y/, and for high $F1$ and high $F2$ for /ʉ/, with enhanced responding near the centroid for all cases.

For stimuli to which birds were trained to refrain from pecking (negative), the same basic pattern was found with all three variables again contributing significantly to prediction of peck rate. The overall regression was significant ($F\text{ratio}_{3,132}=33.22$, $p<0.0001$, multiple $R=0.66$). The ordinal relation of the three variables predicting peck rates was the same as for the positive cases. The value of $F2$ had the greatest contribution ($r=0.46$, $p<0.001$) followed by $F1$ value ($r=-0.37$, $p<0.001$) followed by distance from the centroid ($r=0.28$, $p<0.001$). Using the same example of the two birds for which /i/ was positive and /ɪ/ was the negative vowel, the regression analysis indicates that birds pecked least in response to stimuli with lower $F2$ and higher $F1$, and overlaid upon this pattern is a tendency to peck relatively less to stimuli closer to the centroid of the distribution of /ɪ/ tokens. The same pattern was seen for each vowel distribution: lowest rates for low $F1$ and high $F2$ for /ɪ/, high $F1$ and high $F2$ for /y/, and for low $F1$ and low $F2$ for /ʉ/, with diminished responding near the centroid for all cases. For both positive and negative stimuli, response rate

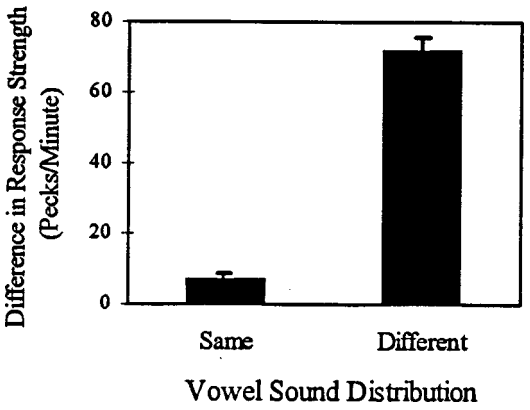

FIG. 5. Average differences in peck rates in response to pairs of stimuli when drawn from the same vowel sound distribution or from different distributions.

for any given stimulus can be reasonably well predicted on the bases of *F*1 and *F*2 values and on the distance from the centroids of the distributions.

The reader may recall that one of the difficulties in interpreting infant responses is that one cannot know whether infants fail to respond because they cannot discriminate two stimuli or because they are treating discriminably different stimuli equivalently. To address this question in the present experiment, distributions had been constructed to overlap in a manner such that some pairs of stimuli were included in a single distribution for one set of birds, but were divided between the two distributions for the other set of birds. Labels **A**, **B**, **C**, and **D** denoted these four pairs in Fig. 3. Analyses of peck rates for these pairs of stimuli indicate that pairs of vowels from the same distribution for one set of birds (e.g., /i–ɪ/) were, indeed, discriminably different for the other set of birds (e.g., /y–ʉ/). Average absolute-value differences in normalized peck rates are plotted in Fig. 5.

When stimuli were assigned to different distributions (**B**–**C** and **A**–**D** for /y–ʉ/; **A**–**B** and **C**–**D** for /i–ɪ/) the average difference was 71.94 pecks per minute, a significantly greater response difference ($t_{14}=15.58$, $p<0.0001$) as compared to an average difference of 7.14 pecks for minute when stimuli were drawn from the same distribution (**A**–**B** and **C**–**D** for /y–ʉ/; **B**–**C** and **A**–**D** for /i–ɪ/). The fact that differences in peck rates were so much greater for stimuli assigned to different distributions (/i/ vs /ɪ/, /y/ vs /ʉ/) compared to stimuli drawn from the same distribution (/i/, /ɪ/, /y/, or /ʉ/) can be taken as strong evidence that the degree to which stimuli elicit the same response cannot be explained simply as a lack of discriminability. It appears that birds learned to treat discriminably different stimuli as functionally equivalent.

C. Discussion

From the data for the eight birds, several observations can be made. First, relative frequencies of the two primary spectral prominences (*F*1 and *F*2) were good predictors of how these two-vowel spaces became organized for starling subjects. Within the context of general principles of learning, analogous effects are well established and may remind the reader of classical theories of discrimination learning (e.g., Spence, 1936, 1937, 1952, 1960). One of the essential facts that these early learning theorists wished to explain was that a positive response to one stimulus (S+) was affected by the nature of a second stimulus (S−) which discouraged responding. A classic experiment in this regard (Hanson, 1959) demonstrated that the peak of the discrimination function for responses by pigeons that were trained to respond to a visual stimulus at one wavelength (S+) would shift to a longer wavelength when S− was a shorter wavelength. Basically, this ''peak shift effect'' consisted of the response pattern to S+ (excitatory) being skewed away from S− (inhibitory).

In the present experiment, strength of responses to stimuli from positive distributions became greater as the frequencies of spectral prominences for *F*1 and *F*2 were more distant from those for the negative distributions. For the example of the vowels /i/ (S+) and /ɪ/ (S−), response strength increased with decreasing *F*1 and increasing *F*2 frequencies. This pattern is consistent with what one would expect on the basis of precedents in the learning literature. Lest one consider this point to be of significance only as it pertains to a trivial consistency between pigeon and starling performance, it bears note that such behavior is consistent with classic perspectives in phonetics. As Jakobson and Halle wrote in *The Fundamentals of Language* (1971, p. 22) ''All phonemes denote nothing but mere otherness.'' In this case, the degree to which a stimulus is treated as /i/, /ɪ/, /y/, or /ʉ/ depends considerably upon the degree to which the stimulus is not /ɪ/, /i/, /ʉ/, or /y/, respectively. This tradition was extended, for example, in the simulation studies by Liljencrantz and Lindblom (1972) and later by Lindblom (1986) in which many of the consistencies in vowel systems across languages could be explained by the principle of languages using vowel sounds that are as mutually distinctive as possible in acoustic and/or auditory space. When one considers the present experiment as one for which the task for subjects is to organize a very small vowel space, such ''mere otherness'' plays an influential role.

However consistent the data may be with regard to precedents in the learning and phonetics literature, there exists a potentially disquieting difference between starling response patterns and previous reports of adult human goodness ratings for distributions of /i/ tokens. Following the necessary reflections, all eight starlings exhibited graded response structures with increasing response strength as the frequencies of *F*1 and *F*2 became more distant from formant frequencies for the opposing vowel distribution. However, Kuhl (1991) found no strong evidence for this sort of anisotrophism for goodness judgments by adult humans for stimuli distributed around the same centroid but with 30-mel step sizes. While it is true that the present experiment employed distributions of more densely packed stimuli relative to earlier efforts (e.g., Grieser and Kuhl, 1989; Kuhl, 1991), this difference between human and starling data bears note. Experiment 3 of this report provides adult human judgments of the stimuli used in Experiment 2, and discussion of these discrepancies will receive fuller attention.

Turning now to the third predictor of response strength, consider the fact that response rates were greater for positive distributions and lesser for negative distributions when

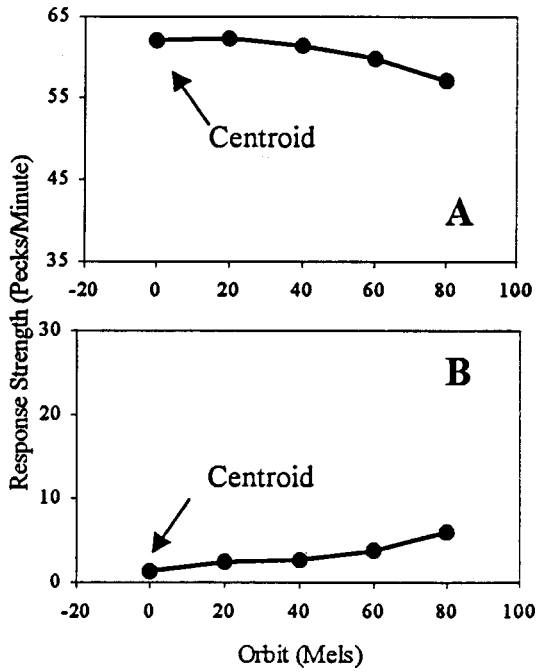

FIG. 6. From experiment 3, adjusted peck rate data averaged across conditions for stimuli at different distances from the centroids of positive (top A) and negative (bottom B) distributions.

stimuli were nearer to centroids of the distributions. Figure 6 displays adjusted peck rate data averaged across conditions for stimuli at different distances from the centroids of positive and negative distributions. Such response patterns—whether derived from ratings of ''goodness,'' response times in category judgment tasks, or response rates/probabilities—are frequently considered among the hallmarks of ''category'' structure.

Gradients present for starling data stand in contrast to Kuhl's (1991) finding that rhesus monkeys showed no evidence of response differences beyond those predicted simply by acoustic/auditory distance. Monkeys were equally proficient discriminating pairs of vowel stimuli when one stimulus was the prototype /i/ as when one stimulus was the poorer rendition (nonprototype) of /i/. Despite the fact that distributions were more densely sampled in the present case, the centroids for /i/ distributions in both studies were near identical to those for Kuhl (1991). There are two reasons not to consider the present results to be at odds. First, Kuhl (1991) used a within-distribution discrimination task; monkeys were reinforced for responding to within-distribution stimulus differences. The present case is more akin to actual use of phonetic distinctions with starlings reinforced for responding to between-distribution differences with no encouragement to respond differentially to within-distribution differences. Second, monkeys did not have the benefit of extensive experience with the distributions of vowel sounds. Of course, this is not analogous to the case for starlings in the present study nor for the comparison case of six-month-old human infants who have been bathed in a half-year exposure to distributions of vowel sounds.

Overall, there is little to recommend a sensory account. The monkey data suggest that differential effects, *vis a vis* the centroid, are not a consequence of any general auditory predeposition. The fact that starling data did not differ systematically as a function of vowel (/i/, /ɪ/, /y/, /ʉ/) suggests that none of these sounds is privileged in acoustic/auditory terms.

It is beyond the scope of the present report to review theories of categorization; however, it bears note that all theories of categorization strive, at least in part, to explain the ubiquitous finding of graded structure. This is true for the class of probabilistic models which include spreading activation (e.g., Collins and Loftus, 1975) or feature comparison (Smith *et al.*, 1974) and which often include hypothesis of some internal prototype with which particular instances are compared (see, e.g., Posner and Keele, 1968; Strange *et al.*, 1970). Others have proposed that graded structure can be accommodated in exemplar-based models by which stimuli are categorized with reference to stored exemplars of individual experienced instances (e.g., Hintzman and Ludlam, 1980; Medin and Schaffer, 1978). Finally, more recent connectionist models of distributed memory (e.g., Knapp and Anderson, 1984) also result in graded category structure.

One explanation offered for the results of the earlier studies by Kuhl and her colleagues (Grieser and Kuhl, 1989; Iverson and Kuhl, 1995; Kuhl, 1991, 1993) is that vowel ''categories'' could be conceptualized as being organized around an ideal or prototypical version of the vowel. Kuhl (1993) argues for experience-based versus innate prototypes, and the present data are consistent with this in as much as starlings would be unlikely genetic recipients of prototypes for the human vowel sounds /i/, /ɪ/, /y/, /ʉ/. With respect to humans, Kluender (1994) has made the argument that, in general, principles of natural selection would not encourage innate predispositions for speech sounds that are relatively infrequent among the worlds languages. In this respect, none of the vowels used in this study, with the exception of /i/, occurs with great frequency among languages. Even for very common /i/, acoustic properties can vary considerably across languages.

For the most part, theories of human categorization behavior do not rely upon endowment with innate prototypes or concepts; although, some essentialist accounts of concepts have been influential (Atran, 1987; Gelman and Wellman, 1991; Keil, 1987; Medin and Ortony, 1989). Instead, most attempts to explain categorization behavior make do with the assumption that the environment provides ample structure for experience to define and shape internalized category structure. In the present case with starlings, one would infer that experience with distributional properties of these vowel sounds served as the basis for development of the graded response structures. More specifically, behavior comes to reflect experienced probability-density functions in as much as vowel-sound distributions were more dense nearer to the centroid. Following experiment 3, a simple linear learning model will be presented that tests how, for starlings (and humans), experience with distributional properties of vowel sounds may give rise to graded response structures.

III. EXPERIMENT 3

Starling response gradients, both for /i–ɪ/ birds and for /y–ʉ/ birds, differed from the /i/ category gradient inferred

from adult goodness judgments in Kuhl (1991). In particular, the majority of the variance in human judgments measured in that earlier study by Kuhl could be attributed to distance from the centroid (prototype) with little observable influence of *F*1 and *F*2 frequency *per se*. In the present experiment, a goodness judgment task much like that used by Kuhl (1991) with adult human subjects was used to assess the pattern of relative ''goodness'' judgments of /i/ and /ɪ/ stimuli used in experiment 2.

A. Method

1. Subjects

Thirteen college-age adults served as subjects. All subjects learned English as their first language, reported normal hearing, and received Introductory Psychology class credit for their participation.

2. Stimuli

All 98 stimuli from the distributions for /i/ and /ɪ/ employed in experiment 2 were used in experiment 3.

3. Procedure

The subjects' task was to judge all vowel tokens with regard to the extent to which each token constituted a ''good'' example of the vowel /i/ or the vowel /ɪ/. One to three subjects were tested concurrently in three single-subject sound-proof chambers (Suttle Equipment Corp.) during a single half-hour experimental session. Each of the 98 stimuli was presented six times in random order at an intensity level of 70 dB SPL at a rate of about one stimulus every 3 s. To avoid any bias being introduced by the particular pronunciation of the experimenter, all instructions were written. Subjects were instructed to press one of seven buttons labeled ''1 good hit,'' ''2,'' ''3,'' ''4,'' ''5,'' ''6,'' and ''7 good heat'' to indicate the degree to which each token sounded like a good example of /ɪ/ or /i/. After selecting one of the seven alternatives, subjects pressed an eighth button to indicate that they were satisfied with their selection. To make certain that subjects were familiar with the range and distribution of the stimulus tokens, the first two blocks of 98 responses were considered practice and were not subjected to further analysis.

B. Results

All subjects had no problems conforming with instructions and completing the task. Patterns of average ratings across the 13 subjects are displayed in Fig. 7. Analogous to the analysis for experiment 2, *F*1 value (mels), *F*2 value (mels), and distance from centroid of the distribution (mels) were entered into the multiple regression analyses. Regression analyses were run separately for responses to stimuli from the /i/ and /ɪ/ distributions. For responses to /i/ stimuli, regression on only two variables was statistically significant (F ratio$_{3,634}=30.20, p<0.0001$, multiple $R=0.35$). The value of *F*2 had the greatest contribution ($r=0.33$, $p<0.001$) followed by distance from the centroid ($r=-0.13$, $p<0.001$). The value of *F*1 did not contribute significantly to predicting the relative goodness of /i/ tokens ($r=-0.04$, $p=0.34$).

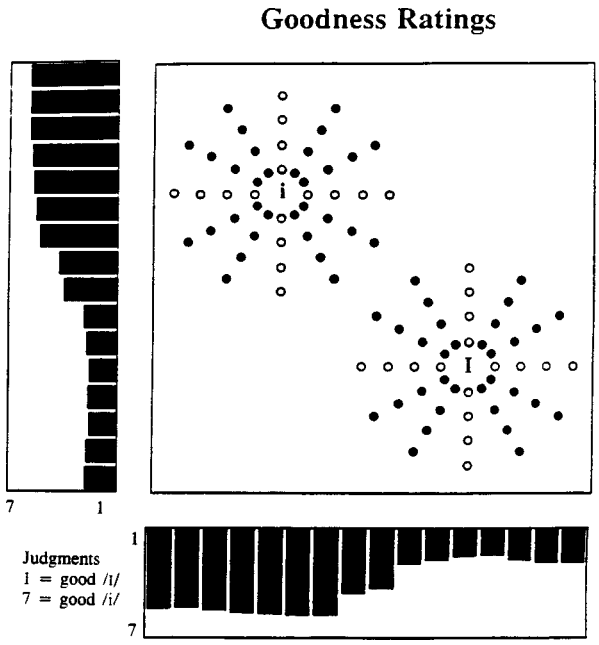

FIG. 7. From experiment 3, average ratings by 13 human listeners of good /i/ (7) to good /ɪ/ (1) for 34 stimuli presented to starlings as novel test stimuli. Histograms are as a function of *F*1 and *F*2 values.

For judgments of /ɪ/, all three variables contributed significantly to prediction of ratings. The three-variable regression was statistically significant (F ratio$_{3,634}=45.88$, $p<0.0001$, multiple $R=0.42$). Distance from the centroid had the greatest contribution ($r=0.30$, $p<0.001$) followed by *F*2 value ($r=0.24$, $p<0.001$) followed by value of *F*1 ($r=-0.18$, $p<0.001$).

Regression analyses were conducted to quantify the correspondence between starling responses to distributions for /i/, /ɪ/, /y/, and /ʉ/ and adult human goodness ratings for the /i/ and /ɪ/ distributions. For starlings, the data consisted of the responses to novel test tokens drawn from respective positive and negative distributions following reflections as before for /i/, /ɪ/, /y/, and /ʉ/. Response rates for these 34 tokens, 17 novel positive tokens and 17 negative, were compared with goodness ratings for corresponding tokens of /i/ and /ɪ/, respectively. The correlation between responses to tokens drawn from positive and negative distributions (for starlings) and goodness judgments of corresponding /i/ and /ɪ/ tokens (for humans) was extremely high ($r=0.999$, $p<0.0001$) indicating that, across the two distributions, starling responses and adult human judgments were in generally close correspondence.

Given the source of much of the variance in the data for both human and avian subjects, such substantial correlation may not be surprising. Much of the variance across response rates and across ratings for the two distributions is related to differential responses to two distributions of sounds. Starlings were trained to respond differentially to contrasts between /i/ vs /ɪ/ or /y/ vs /ʉ/, and humans were asked to rate instances of phonemically distinct classes of sounds /i/ and

/ɪ/. Consequently, much of the total variance entered into the correlation analysis can be interpreted with respect to responses being of two distinct types owing to the use of two distinct distributions of sounds. As such, the extremely high degree of shared variance may have more to do with variance between vowel distributions than with variance within vowel distributions, a central focus of this effort.

In order to address correspondences between response patterns within individual vowel distributions, separate regression analyses were conducted for starling responses to novel positive tokens and human judgments of corresponding tokens of /i/, and for starling responses to negative tokens and human judgments of corresponding tokens of /ɪ/. The correlation between starling peck rates to the 17 novel stimuli drawn from positive distributions and human goodness judgments of the corresponding stimuli drawn from the /i/ distribution was substantial ($r=0.671$, $p<0.01$). The correlation between starling peck rates to the 17 novel stimuli drawn from negative distributions and human goodness judgments of stimuli drawn from the /ɪ/ distribution was still greater ($r=0.784$, $p<0.001$).

The choice of /i/ as the benchmark positive distribution was an arbitrary one; /ɪ/ could have been used. Although the lack of systematic variation as a consequence of which vowel was designated positive suggests that this choice ought not matter, two additional regression analyses were conducted. One analysis compared responses to tokens drawn from positive distributions for starlings with goodness judgments for /ɪ/, and one compared responses to tokens drawn from negative distributions for starlings with goodness judgments for /i/. Both of these correlations were comparable to those computed for the previous complementary relationships. The correlation between peck rates in response to novel positive tokens and goodness judgments for corresponding /ɪ/ tokens was significant yielding $r=0.697$ ($p<0.002$). The correlation between responses to novel negative tokens and goodness judgments for corresponding /i/ tokens also was significant yielding $r=0.703$ ($p<0.002$).

Overall, there was a remarkable correspondence between human goodness judgments and starling peck rates. With the exception of the negligible contribution of $F1$ frequency on goodness judgments for /i/, overall pattern of human responses is quite consistent with the starling measures.

Figure 8 displays mean goodness ratings as a function of distance from the centroids of the distributions for /i/ and /ɪ/. This tendency to attribute a greater degree of ''goodness'' to tokens with formant-frequency values nearer the centroids of these distributions is consistent with Kuhl's (1991) measurements for a broader distribution of tokens which shared the same /i/ centroid as used in these experiments. Although broad gradients corresponding to values of $F2$ for /i/ judgments and to $F1$ and $F2$ for /ɪ/ judgments do not correspond well to Kuhl's data, the effects, particularly for /i/ were not unanticipated. Lively (1993) synthesized /i/ stimuli comparable (30-mel rings) to those used by Kuhl (1991), and while his adult human subjects demonstrated a significant effect of distance from the centroid for both ''prototype'' and ''nonprototype'' conditions, vowels with higher $F2$ values were given the highest goodness ratings. From Lively's figures,

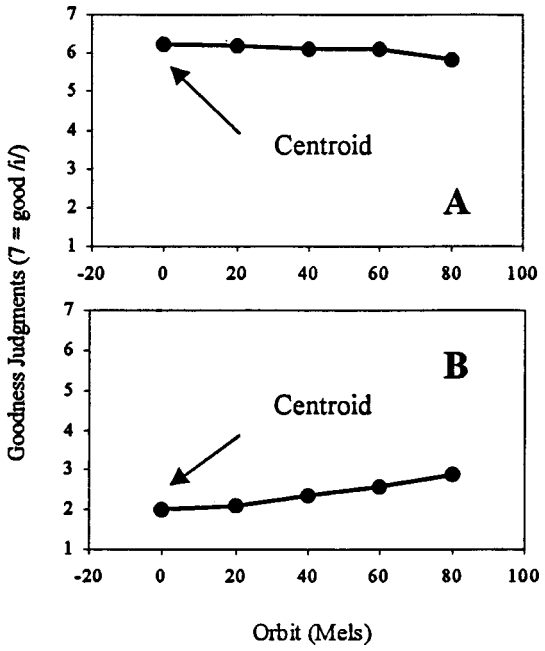

FIG. 8. From experiment 3, average goodness ratings from human listeners for 34 stimuli presented to starlings as novel test stimuli as a function of distance from the centroids of /i/ (top A) and /ɪ/ (bottom B) distributions.

one also can observe that the value of $F1$ played a negligible role in goodness ratings for /i/. Human goodness judgments elicited for tokens from tighter, denser distributions in experiment 3 correspond well with Lively's (1993) measures of goodness for broader distributions (as used by Kuhl, 1991). Further, starling responses are in close accord with these findings.

One possible explanation for the sizable influence of $F2$ could be that, when $F2$ is relatively high and nearer $F3$, there is some auditory interaction creating a functional $F2$ ($F2'$) that serves to warp perceptual distance in a fashion not captured solely by mel distance (e.g., Johnson, 1989). Typically, the assumption is that $F2'$ can be described as the weighted average of $F2$ and $F3$, thus equal mel steps when $F2$ is near $F3$ result in disproportionately large perceptual distances.

One way in which starling and human performance differed is informative. For human listeners, $F2$ and, to a lesser extent, distance from the centroid accounted for much of the variance in goodness judgments of /i/. For the same listeners, variance in goodness judgments of /ɪ/ was best described in terms of distance from the centroid followed by $F2$ and finally $F1$, all being statistically significant predictors. This contrasts with starling data for which distance from the centroid always is less predictive than $F2$ or $F1$. The more compelling effect of distance from the /ɪ/ centroid for native-English listeners may be due to the fact that for /ɪ/, but not /i/, close neighbors (/ɛ/, /e/, /ɜ˞/, and /i/) surround all sides in the $F1$–$F2$ plane. There is a smaller effect of distance from the centroid for goodness judgments of /i/, which lies at an extreme corner of the vowel space with no neighboring vowels with lower $F1$ or higher $F2$. While extreme versions of /i/ (low $F1$, high $F2$) are most distinctive relative to other English vowel sounds, acoustic instances of /ɪ/ that are most central to the distribution would be maximally distinctive from surrounding vowel sound distributions. What is the

same whether one considers either the minimalist two-vowel spaces presented to starlings or the more-populated English vowel space, a perceptual principle akin to ''mere otherness'' is observed much as it was suggested to exist in the postulation of ''adaptive dispersion'' (Lindblom, 1986) as a predictor of the structure of vowel inventories.[6]

C. Minimalist computational model

In order to better understand how a relatively simple organism such as a starling can come to learn a functional mapping of vowel sounds that is so similar to that for humans, a simple linear association network model was simulated using elementary matrix and vector operations. Because the data for starlings and for humans were fairly well accommodated in the linear operations of regression analysis, there was reason to believe that a learning model based solely upon linear operations might provide an adequate and potentially informative account for the data. The model used here can be considered an instantiation of the Hebbian synapse rule (Hebb, 1949), and is a tightly constrained model in which all operations are local and there is no need for the ''back-propagation'' of errors common to many current network models.

A linear network can be conceptualized as a system of linear algebra equations of the form:

$$\mathbf{Aw}=\mathbf{b}, \tag{1}$$

where **A** is a matrix of input exemplars, **w** is a vector of connection weights, and **b** is a vector of output values. The weights of the network can then be solved for by

$$\mathbf{w}=\mathbf{A}^{+}\mathbf{b}, \tag{2}$$

where $\mathbf{A}^{+}$ is the pseudo-inverse of **A**. (For review, see Jordan, 1986.)

Inputs to the network were synthesizer values for *F*1 and *F*2 for each English vowel (/i/ and /ɪ/). Thus, each vowel sound was described as a two-value vector. The 64 vectors for the training stimuli used for /i–ɪ/ birds in experiment 3 were entered into an array **A** yielding a 64×2 element input matrix. A 64×1 output matrix **b** was created by entering a ''1'' for the positive stimuli (/i/) and a ''0'' for the negative stimuli (/ɪ/). Then, Eq. (2) was solved for the 2×1 weight vector **w**. This completed the ''training'' phase. This matrix-algebraic solution is formally equivalent to using a single-layer network for which weights are determined through multiple iterations of exposure to training tokens (Jordan, 1986). In this case, advantage was taken of the fact that the same weights can be derived by solving equations in closed form.

The model then was tested using the 34 novel test stimuli and 16 training stimuli that were presented to birds in trials without contingencies. This constituted a new input matrix **A** of dimensions 50×2. This matrix was multiplied by the 2×1 weight vector derived in the ''training'' phase to yield a 1×50 vector of values corresponding to the output values (between 0 and 1) for each of the test stimuli.

Comparisons between the model output for test stimuli and avian peck rates to the same stimuli reveal a number of similarities. As was the case for bird responses, model output exhibited a gradient across values of *F*1 and *F*2 such that greater and lower outputs occur for those stimuli with extreme formant values. Also, there is a similar ''prototype'' effect in as much as output values respect the probability–density distribution of the input with relatively higher or lower values nearer the centroids of the positive and negative distributions, respectively. Correlation coefficients were computed comparing model output and peck rates for /i–ɪ/ birds reinforced for pecking either to /i/ or to /ɪ/. For the /i/ positive case, $r=0.926$, $p<0.0001$. For the /ɪ/ positive case, $r=0.919$, $p<0.0001$. As was the case for comparisons between avian responses and human goodness judgments, comparisons also were made between model outputs and avian responses to stimuli within individual vowel distributions. The correlation between starling peck rates to the 17 novel stimuli drawn from positive distributions and model predictions for the corresponding stimuli drawn from the /i/ distribution was $r=0.678$, $p<0.01$. The correlation between starling peck rates to the 17 novel stimuli drawn from negative distributions and model predictions for stimuli drawn from the /ɪ/ distribution was greater $r=0.730$, $p<0.01$. It appears that this minimalist perceptron model may provide respectable predictive power.

It should be emphasized that the purpose of this simulation exercise was not to propose that the learning process for either human infants or starlings must reduce to a simple model of this type. Although this model can be cast as a ''neural network'' model and could enjoy the allusion to neural processing, no such claims are being made here. It does bear note, however, that such a model engenders biologically plausible operations in the sense that connections are local and weight adjustments follow simple Hebbian rules. Nevertheless, the model is likely too simplistic and contexturally isolated at present to be suggested as a model for neural activity in avian, let alone human brains. What is important is that there is reason to hope that the processes by which human infants (and starlings) come to organize vowel sounds in a language-specific fashion may be explainable by rather elegant and possibly linear processes.

IV. GENERAL DISCUSSION

The present effort began with the fundamental question of how perceptual behavior of human infants could come to respect language-specific equivalence classes for speech sounds through experience and learning if such classes were to arise *de novo*. Earlier findings (Kuhl, 1991) using nonhuman subjects suggested that, for vowel sounds at least, properties of mammalian auditory sensory systems do not, by themselves, give rise to functional equivalence classes appropriate to linguistic sound systems. This evidence, together with studies (Kuhl, 1983; Kuhl *et al.*, 1992) demonstrating that, by six months of age, infants respond to acoustically different vowel sounds in a fashion that respects their functional equivalence within a language environment, suggests an essential role of early experience. While Kuhl (1991) took advantage of animal subjects to minimize effects of experience with speech sounds in order to evaluate raw sensory

abilities, the present studies exploited the opportunity to embrace and control experience with approximations to natural distributions of speech sounds.

Following an experiment which established the appropriateness of stimulus materials for this effort, it was found that starlings could learn functional equivalence classes of vowel sounds that were representative of the English vowels /i/ and /ɪ/ as well as control stimuli /y/ and /ʉ/. Starlings generalized to novel instances of these distributions, and there was evidence that equivalent responses to different tokens drawn from the same distributions of vowel sounds were not indicative of a lack of discriminative capacity. In fact, avian subjects that learned to treat orthogonal (in $F1-F2$ space) distributions equivalently were facile in responding differentially to the same pairs of stimuli treated equivalently by other subjects.

Both across and within vowel distributions, there was remarkable agreement between measures of starling response strength (peck rate) and human goodness judgments of the same English vowel sounds. To the extent that divergence between starling and human performance was found (greater effect of distance from centroid for /ɪ/), it is likely explainable on the basis of experience with vowel sounds encountered by native-English human listeners but not by starling subjects. Taken together, human and avian results suggest that the process of mapping a space of vowel sounds may be in accord with long-held principles of ''mere otherness'' and ''adaptive dispersion.''

A very simple linear associative model was used to assess starling performance. When the model was permitted the same range of ''experience'' with distributions of vowel sounds as starlings were, response strengths to individual vowel sounds from the model and birds were in close agreement. Although no claims should be made about the verisimilitude of the computational simulation as compared with biological instantiations of these processes by humans or by birds, the model does present an existence proof that a simple linear system can result in functional mappings of vowel sounds in similarly graded and language-specific fashion. In particular, simulation results do suggest that relatively elegant solutions may exist to explain how subjects with brains of little volume come to exhibit response patterns that are strikingly like those measured for human subjects for the same vowel sounds. Avian and computational performances taken together, it may be appropriate to exhibit some caution before one either posits the requirement of innate specific predispositions for phonetic categories, or hypothesizes the existence of internal prototypes for phonetic categories through whatever process. Neither starlings nor perceptrons have the privilege of inheriting human phonetic categories, and peaks in response gradients allude to, but do not require, putative prototypes. In a similar spirit (Lacerda, 1998) has introduced an exemplar-based model inspired by neuronal group selection theory (Edelman, 1987) that further demonstrates that constructs such as prototypes are unnecessary to account for extent data for human adults and infants responding to vowel sounds.

It well may be the case that rather general processes of learning can accommodate much of what is known about the functional equivalence of vowel sounds within the vowel space of a language environment. This demonstration of the efficacy of simple learning via distributional properties at the phonetic level is consonant with recent demonstrations that statistical relationships between neighboring speech sounds can be used by 8-month-old infants at the morphemic level for word segmentation (Saffran *et al.*, 1996). When one considers the task assigned to the infant language learner, it may be possible for young listeners to establish their nascent lexicons through little more than sensitivity to statistical regularities of language input together with organizational processes that serve to enhance distinctiveness of regions in that input.

ACKNOWLEDGMENTS

The authors thank Carol Fowler, Francisco Lacerda, Arthur Samuel, and Winifred Strange for thoughtful comments on an earlier draft of this manuscript. This work was supported by NIDCD Grant No. DC-00719 and NSF Young Investigator Award DBS-9258482 to the first author.

[1]Among concerns one may have regarding the use of the term ''category'' is the fact that the term is used in at least three different ways. First, the term most commonly is used to refer to a group of objects or events in the world that more or less share some set of attributes. Second, the term ''category'' often is used to refer to a set of objects or events that give rise to similar behavior, i.e., functional equivalence. Third, the term is sometimes used with reference to some internal cognitive representation which may or may not be defined by a prototype. Putatively, this internal representation serves to mediate the relation between sensory information and behavior.

[2]Kuhl (1991) was similarly circumspect with regard to her use of ''prototype'' with respect to internal representations of phonetic categories. In the cognitive psychology categorization literature, behavior that has been attributed to the existence of prototypes also has been attributed to exemplar models that do not require categories to be defined by reference to a single representation of the category (see, e.g., Brooks, 1978; Knapp and Anderson, 1984; Medin and Schaffer, 1978; Nelson, 1974; Reed, 1972). While there are significant differences both within and between different prototype and exemplar models of categorization, for now, it will be adequate to understand that, for all of these theoretical approaches, categories are taken to exhibit structure such that not all instances of a category constitute equally good category members. Kuhl (1991, 1993) accepts both prototype and exemplar models as plausible with respect to phonetic categories.

[3]The Swedish vowel system does include a variant of the vowel [i]; however, Swedish /i/ is substantially different acoustically from English /i/ and the /i/ ''prototype'' used in Kuhl *et al.*'s (1992) study was not typical of Swedish /i/.

[4]Optimal performance was defined as the highest ratio of pecks to positive versus negative stimuli. Birds were idiosyncratic with regard to the amount of deprivation that resulted in the most stable performance, and weights ranged from 80% to 90% of free-feed weights at the time of training/testing.

[5]Extreme values were deleted to account for the fact that behavior of the birds can be affected by motivational factors irrelevant to the questions of interest. For example, a very hungry subject will peck more vigorously and indiscriminantly, often early in a test session, and relatively satiated birds will decrease peck rates overall toward the end of some sessions. Truncation of both extremes is an unbiased method of avoiding such aberrant data.

[6]One would expect that, if nonhuman subjects must respond differentially to more than two vowels (e.g., the four front vowels /i/, /ɪ/, /ɛ/, and /æ/), a greater effect of distance from the centroid should be found for /ɪ/ and /ɛ/ owing to the requirement of distinctiveness from flanking neighbors. The authors presently are conducting such an experiment.

Atran, S. (**1987**). ''Folkbiological universals as common sense,'' *Noam Chomsky: Consensus and Controversy*, edited by S. Modgil and C. Modgil (Falmet, Philadelphia).

Best, C. T. (**1994**). ''The emergence of native-language phonological influences in infants: A perceptual assimilation model,'' in *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, edited by J. V. Goodman and H. C. Nusbaum (MIT, Cambridge, MA), pp. 167–224.

Brooks, L. (**1978**). ''Non-analytical concept formation and memory for instances,'' in *Cognition and Categorization*, edited by E. Rosch and B. Lloyd (Erlbaum, Hillsdale, NJ), pp. 169–211.

Burdick, C. K., and Miller, J. D. (**1975**). ''Speech perception by the chinchilla: Discrimination of sustained /a/ and /i/,'' J. Acoust. Soc. Am. **58**, 415–427.

Collins, A. M., and Loftus, E. F. (**1975**). ''A spreading-activation theory of semantic processing,'' Psychol. Rev. **82**, 407–428.

Dooling, R. J., Best, C. T., and Brown, S. D. (**1995**). ''Discrimination of synthetic full-formant and sinewave /ra-la/ continua by budgerigars (*Melopsittacus undulatus)* and zebra finches (*Taeniopygia guttata*),'' J. Acoust. Soc. Am. **97**, 1839–1846.

Dooling, R. J., Okanoya, K., Dowling, J., and Hulse, S. (**1986**). ''Hearing in the starling (*Sturnus vulgaris*): Absolute thresholds and critical ratios,'' Bull. Psychonomic Soc. **24**, 462–464.

Edelman, G. (**1987**). *Neural Darwinism: The Theory of Neuronal Group Selection* (Basic Books, New York).

Elimas, P. D. (**1991**). ''Comment: Some effects of language acquisition on speech perception,'' in *Modularity and the Motor Theory of Speech Perception*, edited by I. G. Mattingly and M. Studdert-Kennedy (Erlbaum, Hillsdale, NJ), pp. 111–116.

Gelman, S. A., and Wellman, H. M. (**1991**). ''Insides and essences: Early understandings of the non-obvious,'' Cognition **23**, 183–209.

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

Halle, M. (**1990**). ''Phonology,'' in *An Invitation to Cognitive Science: Language*, edited by D. N. Osherson and H. Lasnik (MIT, Cambridge, MA), pp. 43–68.

Hansen, H. M. (**1959**). ''Effects of discrimination training on stimulus generalization,'' J. Exp. Psychol. **58**, 321–372.

Hebb, D. O. (**1949**). *The Organization of Behavior* (Wiley, New York).

Hintzman, D. L., and Ludlam, G. (**1980**). ''Differential forgetting of prototypes and old instances: Simulation by an exemplar-based classification model,'' Mem. Cog. **8**, 378–382.

Iverson, P., and Kuhl, P. (**1995**). ''Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling,'' J. Acoust. Soc. Am. **97**, 553–562.

Jakobson, R., and Halle, M. (**1971**). *The Fundamentals of Language* (Mouton, The Hague).

Jamieson, D. J., Ramji, K. V., Kheirallah, I., and Nearey, T. M. (**1992**). ''CSRE: A speech research environment,'' in *Proceedings ICSLP 92*, edited by J. Ohala, T. Nearey, B. Derwing, M. Hodge, and G. Wiebe (Univ. Alberta, Edmonton, AB), pp. 1127–1130.

Johnson, K. (**1989**). ''Higher formant normalization results from auditory integration of *F*2 and *F*3, Percept. Psychophys. **46**, 174–180.

Jordan, M. I. (**1986**). ''An introduction to linear algebra in parallel distributed processing,'' in *Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*, edited by D. E. Rumelhart and J. L. McClelland (MIT, Cambridge, MA), pp. 365–422.

Keil, F. C. (**1987**). ''Conceptual development and category structure,'' in *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*, edited by U. Neisser (Cambridge U. P., Cambridge), pp. 175–200.

Klatt, D. H. (**1980**). ''Software for a cascade/parallel formant synthesizer,'' J. Acoust. Soc. Am. **67**, 971–995.

Kluender, K. R. (**1991**). ''Effects of first formant onset properties on voicing judgments result from processes not specific to humans,'' J. Acoust. Soc. Am. **90**, 83–96.

Kluender, K. R. (**1994**). ''Speech perception as a tractable problem in cognitive science,'' in *Handbook of Psycholinguistics*, edited by M. A. Gernsbacher (Academic, San Diego, CA), pp. 173–217.

Kluender, K. R., and Diehl, R. L. (**1987**). ''Use of multiple speech dimensions in concept formation by Japanese quail,'' J. Acoust. Soc. Am. Suppl. 1 **82**, S84.

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.

Kluender, K. R., Diehl, R. L., and Killeen, P. R. (**1987**). ''Japanese Quail can learn phonetic categories,'' Science **237**, 1195–1197.

Knapp, A. G., and Anderson, J. A. (**1984**). ''Theory of categorization based on distributed memory storage,'' J. Exp. Psychol. **10**, 616–637.

Kuhl, P. K. (**1981**). ''Discrimination of speech by nonhuman animals: Basic sensitivities conducive to the perception of speech-sound categories,'' J. Acoust. Soc. Am. **70**, 340–349.

Kuhl, P. K. (**1983**). ''Perception of auditory equivalence classes for speech in early infancy,'' Inf. Beh. Dev. **6**, 263–285.

Kuhl, P. K. (**1986**). ''Theoretical contributions of tests on animals to the special-mechanisms debate in speech,'' Exp. Biol. **45**, 233–265.

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 D. de Boysson-Bardies *et al.* (Kluwer Academic, The Hague), pp. 259–274.

Kuhl, P. K., and Miller, J. D. (**1975**). ''Speech perception by the chinchilla: Voiced-voiceless distinction in the 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., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (**1992**). ''Linguistic experience alters phonetic perception in infants six-months of age,'' Science **255**, 606–608.

Kuhn, A., Leppelsack, H. J., and Schwartzkopff, J. (**1980**). ''Measurement of frequency discrimination in the starling (*Sturnus vulgaris*) by conditioning of heart rate,'' Naturwissenschaften **67**, 102–103.

Kuhn, A., Muller, C. M., Leppelsack, H. J., and Schwartzkopff, J. (**1982**). ''Heart-rate conditioning used for determination of auditory threshold in the starling,'' Naturwissenschaften **69**, 245–246.

Lacerda, F. (**1998**). ''An exemplar-based account of emergent phonetic categories,'' J. Acoust. Soc. Am. **103**, 2980(A).

Liljencrantz, J., and Lindblom, B. (**1972**). ''Numerical stimulation 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.

Lively, S. E. (**1993**). ''An examination of the perceptual magnet effect,'' J. Acoust. Soc. Am. **93**, 2423.

Lotto, A. J., Kluender, K. R., and Holt, L. L. (**in press**). ''Effects of language experience on perceptual organization of vowel sounds,'' in *Papers in Laboratory Phonology V*, edited by M. Broe and J. Pierrehumbert (Cambridge U. P., Cambridge).

Massaro, D. W. (**1987**). ''Categorical partition: A fuzzy logical model of categorization behavior,'' in *Categorical Perception*, edited by S. Harnad (Cambridge U. P., Cambridge), pp. 254–283.

Medin, D. L., and Ortony, A. (**1989**). ''Psychological essentialism,'' in *Similarity and Analogical Reasoning*, edited by S. Vosniadou and A. Ortony (Cambridge U. P., New York), pp. 179–195.

Medin, D. L., and Schaffer, M. M. (**1978**). ''A context theory of classification learning,'' Psychol. Rev. **85**, 207–238.

Miller, J. L. (**1977**). ''Properties of feature detectors for VOT: The voiceless channel of analysis,'' J. Acoust. Soc. Am. **62**, 641–648.

Miller, J. L., and Eimas, P. D. (**1996**). ''Internal structure of voicing categories in early infancy,'' Percept. Psychophys. **58**, 1157–1167.

Miller, J. L., and Volaitis, L. E. (**1989**). ''Effect of speaking rate on the perceptual structure of a phonetic category,'' Percept. Psychophys. **46**, 505–512.

Miller, J. L., Connine, C. M., Schermer, T. M., and Kluender, K. R. (**1983**). ''A possible auditory basis for internal structure of phonetic categories,'' J. Acoust. Soc. Am. **73**, 2124–2133.

Nelson, K. (**1974**). ''Concept, word, and sentence: Interrelations in acquisition and development, Psychol. Rev. **81**, 267–248.

Peterson, G. E., and Barney, H. L. (**1952**). ''Control methods used in a study of the vowels,'' J. Acoust. Soc. Am. **24**, 175–184.

Pinker, S., and Bloom, P. (**1990**). ''Natural language and natural selection,'' Behav. Brain Sci. **13**, 707–784.
Posner, M. I., and Keele, S. W. (**1968**). ''On the genesis of abstract ideas,'' J. Exp. Psychol. **77**, 28–38.
Posner, M. I., and Keele, S. W. (**1970**). ''Retention of abstract ideas,'' J. Exp. Psychol. **83**, 304–308.
Reed, S. K. (**1972**). ''Pattern recognition and categorization,'' Cogn. Psychol. **3**, 383–407.
Repp, B. H. (**1977**). ''Dichotic competition of speech sounds: The role of acoustic stimulus structure,'' J. Exp. Psychol. **3**, 37–50.
Rosch, E. H. (**1975**). ''Cognitive representations of semantic categories,'' J. Exp. Psychol. **3**, 193–233.
Rosch, E. H. (**1978**). ''Principles of categorization,'' in *Cognition and Categorization*, edited by E. Rosch and B. Lloyd (Erlbaum, Hillsdale, NJ).
Saffran, J. R., Aslin, R. N., and Newport, E. L. (**1996**). ''Statistical learning by 8-month-old infants,'' Science **274**, 1926–1928.
Samuel, A. G. (**1982**). ''Phonetic prototypes,'' Percept. Psychophys. **31**, 307–314.
Smith, E. E., Shoben, E. J., and Rips, L. J. (**1974**). ''Structure and processing in semantic memory: A feature model for semantic decision,'' Psychol. Rev. **81**, 214–241.
Spence, K. W. (**1936**). ''The nature of discrimination learning in animals,'' Psychol. Rev. **43**, 427–449.
Spence, K. W. (**1937**). ''The differential response in animals to stimuli varying within a single dimension,'' Psychol. Rev. **44**, 430–444.
Spence, K. W. (**1952**). ''The nature of the response in discrimination learning,'' Psychol. Rev. **59**, 89–93.
Spence, K. W. (**1960**). *Behavior Theory and Learning* (Prentice–Hall, Englewood Cliffs, NJ).
Strange, W., Keeney, T., Kessel, F. S., and Jenkins, J. J. (**1970**). ''Abstraction over time from distortions of random dot patterns—a replication,'' J. Exp. Psychol. **83**, 508–510.
Sussman, J., and Lauckner-Morano, V. (**1995**). ''Further tests of the 'perceptual magnet effect' in the perception of [i]: Identification and change-no-change discrimination,'' J. Acoust. Soc. Am. **97**, 539–552.
Volaitis, L. E., and Miller, J. L. (**1992**). ''Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories,'' J. Acoust. Soc. Am. **92**, 723–735.
Werker, J. F. (**1994**). ''Cross-language speech perception: Developmental change does not involve loss,'' in *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, edited by J. V. Goodman and H. C. Nusbaum (MIT, Cambridge, MA), pp. 93–120.