Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement?

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

Andrew J. Lotto
Department of Psychology, Washington State University, P.O. Box 644820, Pullman, Washington 99164-4820

Keith R. Kluender
Department of Psychology, University of Wisconsin—Madison, 1200 West Johnson Street, Madison, Wisconsin 53706

(Received 11 August 2000; accepted for publication 16 November 2000)

For stimuli modeling stop consonants varying in the acoustic correlates of voice onset time (VOT), human listeners are more likely to perceive stimuli with lower $f0$'s as voiced consonants—a pattern of perception that follows regularities in English speech production. The present study examines the basis of this observation. One hypothesis is that lower $f0$'s enhance perception of voiced stops by virtue of perceptual interactions that arise from the operating characteristics of the auditory system. A second hypothesis is that this perceptual pattern develops as a result of experience with $f0$-voicing covariation. In a test of these hypotheses, Japanese quail learned to respond to stimuli drawn from a series varying in VOT through training with one of three patterns of $f0$-voicing covariation. Voicing and $f0$ varied in the natural pattern (shorter VOT, lower $f0$), in an inverse pattern (shorter VOT, higher $f0$), or in a random pattern (no $f0$-voicing covariation). Birds trained with stimuli that had no $f0$-voicing covariation exhibited no effect of $f0$ on response to novel stimuli varying in VOT. For the other groups, birds' responses followed the experienced pattern of covariation. These results suggest $f0$ does not exert an obligatory influence on categorization of consonants as [VOICE] and emphasize the learnability of covariation among acoustic characteristics of speech. © *2001 Acoustical Society of America.* [DOI: 10.1121/1.1339825]

PACS numbers: 43.71.An, 43.71.Es, 43.71.Pc [CWT]

I. INTRODUCTION

Among the world's languages, fundamental frequency ($f0$) and voicing tend to covary. Cross-linguistically, this observation is extremely reliable; so reliable, in fact, that this relationship has been said to arise as a result of physiological constraints on speech production. However, cross-linguistic analysis demonstrates that $f0$ and the acoustic correlates of voice onset time (VOT) covary only among consonants that are used distinctively by languages (Kohler, 1982, 1984, 1985; Kingston, 1986; Kingston and Diehl, 1994), thus suggesting that the influence is not a mandatory consequence of the speech-production system. Vowels immediately following voiced consonants (e.g., [b], [d], [g]) tend to have lower $f0$'s than those following voiceless consonants (e.g., [p], [t], [k]; House and Fairbanks, 1953; Lehiste and Peterson, 1961; Mohr, 1971; Hombert, 1978; Caisse, 1982; Peterson, 1983; Ohde, 1984).[1] For example, the fundamental frequency of the vowel [ʌ] (as in *bud*) tends to be lower in the utterance [dʌ] than in the syllable [tʌ] (Kingston and Diehl, 1994).

The covariation between $f0$ and voicing[2] in language production has a corresponding regularity in speech perception. When listeners categorize synthetic or digitally manipulated natural speech tokens of a phonetic series varying perceptually from voiced to voiceless (e.g., from [ba] to [pa]) listeners more often identify tokens as voiced (i.e., as [ba]) when $f0$ is low. For higher $f0$'s, listeners more often report hearing voiceless consonants (i.e., [pa]). This finding is extremely robust, and has been reported across multiple phonetic contexts, using a variety of measures (e.g., Chistovich, 1969; Haggard *et al.*, 1970; Fujimura, 1971; Massaro and Cohen, 1976, 1977; Derr and Massaro, 1980; Gruenenfelder and Pisoni, 1980; Haggard *et al.*, 1981; Kohler, 1985; Kohler and van Dommelen, 1986; Whalen *et al.*, 1993; Castleman and Diehl, 1996).

Perception of voiced versus voiceless consonants thus follows the regularities of speech production. Much has been made of this correspondence and a good deal of speculation has surrounded the question of why $f0$ and VOT covary in speech production (e.g., Kingston and Diehl, 1994). However, the mechanisms that govern the perceptual side of this correspondence remain largely unknown.

Diehl and Kluender (1989) offer an hypothesis that accounts for the regularities in speech perception and production. By their *auditory enhancement* account, constellations of articulations (such as those that lead to low $f0$ and other

a)Electronic mail: lholt@andrew.cmu.edu

characteristics of [+voice] consonants) tend to covary because these combinations confer a perceptual advantage to the listener. The covariance of $f0$ and voicing in speech production, they argue, is a result of language communities' tendency to coordinate components of speech that have mutually enhancing auditory effects. Thus, the perceptual interaction of $f0$ and VOT evident in listening studies is due to the stable operating characteristics of the auditory system. For this account to provide an explanation for the perceptual influence of $f0$ on voicing categorization, it must offer the additional hypothesis that the coupling of low $f0$'s with the acoustic characteristics of voiced consonants interacts, in some manner, to create an acoustic signal that is more robust than alternative combinations. In a proposal that builds upon the earlier work of Stevens and Blumstein (1981), Diehl *et al.* (1995) have argued that low $f0$ contributes to the presence of low-frequency energy during and near the consonant, thus enhancing the perception of voicing. This auditory enhancement account therefore implies that the natural covariation of $f0$ and voicing observed across languages confers an advantage in auditory processing by virtue of making low-frequency energy more salient for [+voice] consonants.

In a similar vein, Kingston and Diehl (1994) have argued that mutually enhancing characteristics of speech production are explicitly represented by a level of representation intermediate individual acoustic/phonetic correlates of voicing, like a low $f0$, and higher-level representations of distinctive features (such as [+voice]). The advantage of these *integrated perceptual properties* (IPPs) is in limiting energy expenditure in speech production and producing mutually enhancing acoustic effects, thus aiding communication for both the speaker and the listener. Kingston and Diehl propose that the auditory system literally treats a low $f0$ at vowel onset and a short VOT as perceptually equivalent because both act to increase the percept of low-frequency energy near the stop consonant. Like earlier auditory enhancement accounts, this hypothesis implies that the influence of $f0$ on categorization of consonants as voiced or voiceless is a result of demands upon language to provide a robust, readily intelligible, speech signal. The distinction of this account from earlier treatments of auditory enhancement is that it posits an additional level of representation (Diehl *et al.*, 1995).

These auditory enhancement accounts are in agreement that the reason $f0$ acts to shift listeners' perception of voicing is that the acoustic cues provided by $f0$ and voicing interact to enhance some perceptual property (e.g., the perceived presence of low-frequency energy). By these accounts, stable characteristics of the auditory system are responsible for producing perceptual interactions among the acoustic characteristics of $f0$ and voicing. However, preexisting auditory characteristics may not be the only feasible explanation for the interaction of $f0$ and voicing in speech perception. After all, $f0$ and voicing covary in speech production and, as a result, the language environment is rich with structured covariance. Should listeners be sensitive to this covariance, experience with it could influence categorization of voiced and voiceless consonants that vary in $f0$. That is, perceptual interactions between $f0$ and voicing might arise from learning. Recent results have demonstrated the considerable influence covariation within the language environment has upon speech perception. For example, Pitt and McQueen (1998) have shown that listeners are sensitive to the natural covariation of phonetic segments by demonstrating that experimental manipulation of transitional probabilities between speech sounds can elicit predictable context effects. Saffran *et al.* (1996) likewise have argued that adult as well as infant listeners use natural covariation among syllables in word segmentation.

It is possible that listeners' experience with the natural covariation between $f0$ and voicing shapes speech perception. By this proposal, $f0$-voicing covariation in speech *production* arises from speakers' tendency to produce voiced consonants with lower $f0$'s than their voiceless counterparts. Listeners may learn and use this covariation such that $f0$ and voicing interact in speech *perception*. This learning account makes fewer predictions than the auditory enhancement account about why speech production should be so patterned. One possibility is that articulations of $f0$ and VOT are not fully independent. However, as Kingston and Diehl (1994) have argued, no one has yet explained how the articulations of $f0$ and VOT depend on one another. Whatever the nature of the linguistic habits of speakers that promote $f0$/VOT covariation, they remain to be fully uncovered.

The following experiment was designed to tease apart the relative roles that stable preexisting auditory characteristics and effects of experience with $f0$-voicing covariation have in explaining the influence of $f0$ on categorization of consonants as voiced or voiceless. Among human listeners, especially native English speakers who have had extensive experience with $f0$-voicing covariation, this is an extraordinarily difficult task. It would be most desirable to have a population of listeners who are inexperienced with covariation between $f0$ and voicing. Among these individuals, it would be possible to exercise complete experimental control over experience and thus assess relative contributions of audition versus learning.

Nonhuman animals are just such a population. An extensive literature now exists to demonstrate the feasibility of using nonhuman animals in experiments aimed at understanding human speech perception. For the most part, nonhuman animals have provided two distinct services in developing our knowledge of speech perception. In one way, they have served as ''pristine'' auditory systems, unblemished by the experience with speech that human listeners bring to the laboratory. In experiments designed to exploit this characteristic, it is possible to examine the contributions of audition to speech perception while factoring out potential effects of experience. From these experiments, we have learned that nonhuman animals respond to speech categorically (Morse and Snowdon, 1975; Kuhl and Miller, 1975, 1978; Waters and Wilson, 1976), exhibit phonetic context effects (Dent *et al.*, 1997; Lotto *et al.*, 1997), and are sensitive to acoustic trading relations (Kluender, 1991; Kluender and Lotto, 1994). Nonhuman animals have also provided a means of directly manipulating experience with speech to test its effect (Kluender *et al.*, 1987, 1998; Lotto *et al.*, 1999). In experiments of this sort, animals have served as a population in which

there is the possibility of exquisite experimental control over speech experience. These methods have allowed rather precise characterization of effects of experience that can be difficult to garner with human adult or infant listeners (see Holt *et al.*, 1998 for a discussion of these issues) and have led to demonstrations that nonhuman animals exhibit learning-dependent hallmarks of speech perception such as phonetic categorization (Kluender *et al.*, 1987) and internal phonetic category structure (Kluender *et al.*, 1998; Lotto *et al.*, 1999).

Under both experimental paradigms, nonhuman animals have served well, demonstrating that they often respond to speech in much the same manner as human listeners. The present experiment is a fusion of these two experimental approaches. Here, the aim is to delineate the relative contributions of perceptual interactions arising from stable operating characteristics of the auditory system and those arising from experience with covariation in the environment. The present design investigates the influence of $f0$ upon nonhuman animals' responses to stimuli that vary in voicing via manipulation of VOT across three conditions; two conditions provide experience with $f0$-voicing covariation and a third strictly eliminates such experience. Nonhuman animal listeners are essential for this endeavor because they allow rigorous experimental control over the characteristics of experience with $f0$-voicing covariation.

II. EXPERIMENT

Japanese quail (*Coturnix coturnix japonica*), an avian species that has been used extensively in auditory and speech perception research, served as listeners. [See Dooling and Okanoya (1995) for behaviorally derived quail audiograms.] Quail were chosen because they have proven to be very capable subjects in auditory learning tasks (Kluender *et al.*, 1987; Lotto *et al.*, 1999) and avian subjects, in general, are known to exhibit phonetic categorization that mimics essential aspects of human phonetic categorization (Kluender *et al.*, 1987, 1998; Lotto *et al.*, 1999). In addition, quail have demonstrated the ability to respond behaviorally to voiced versus voiceless stimuli (Kluender, 1991; Kluender and Lotto, 1994).

The experiment was designed to assess potential influences of auditory constraints and experience with $f0$-voicing covariation on quails' responses to voicing as $f0$ varies. To achieve this aim, quail were assigned to one of three conditions. Birds in each condition first were trained to respond to either voiced or voiceless syllable-initial stop consonants by pecking a key. Conditions were distinguished by the manner that $f0$ and voicing covaried in the set of stimuli used to train the quail. For birds in the first condition, $f0$ and voicing covaried in the natural manner; voiced consonants had lower $f0$'s than voiceless consonants. A second group of quail was trained with stimuli that varied in the reverse manner; voiced consonants had *higher* $f0$'s and voiceless consonants had *lower* $f0$'s.

A final subset of the quail experienced training stimuli that had no orderly covariation between $f0$ and voicing; the two dimensions were uncorrelated among these training stimuli. As a result, the quail in this last group received no orderly experience with $f0$-voicing covariation. Thus, this condition provides a control group with which to compare the other two conditions. If auditory interactions between $f0$ and voicing influence birds' response to novel test stimuli in the control condition, then birds should respond more robustly to voiced consonants with low $f0$ and voiceless consonants with high $f0$ despite that they have no experience with $f0$-voicing covariation. If experience with covariation between $f0$ and voicing is responsible for the perceptual trading relation, then this group of quail should exhibit no effect of $f0$ upon voicing response.

The remaining conditions are critical to the issue of whether experience with $f0$-voicing covariation influences patterns of perception. If experience with covariation is responsible for the effect of $f0$ on voicing identification, quail that experience natural covariation during training should exhibit an influence of $f0$ on response to novel stimuli in the direction of covariation. However, because the pattern of experience for this condition follows that observed naturally in languages, effects observed are difficult to disentangle from putative auditory influences. However, if quail that experience *reversed* covariation exhibit a ''reversed'' influence of $f0$ on response to voicing such that consonants with *higher* $f0$'s are more often responded to as voiced consonants, then the pattern of behavior should mirror that of the input covariance and diverge from the pattern of results predicted by auditory interactions.

A. Method

1. Subjects

Twenty-one adult Japanese quail (*Coturnix coturnix japonica*) served as listeners in the experiment. Some of the quail with smaller body weights failed to reach criterion performance (pecking ten times more often to positive stimuli than to negative stimuli, see Sec. II B) leaving 16 quail to enter into testing. Free-feed weights ranged from 104 to 160 g.

2. Stimuli

a. Stimulus synthesis. A bilabial stop-consonant series of 19 stimuli varying perceptually from [ba]-[pa] was synthesized using the parallel branch of the Klatt (1980) speech synthesizer. Endpoint stimuli were based upon the productions of a single male talker who uttered ''ba'' and ''pa'' in isolation. For all stimuli in the series, nominal formant frequencies were equivalent. First through third formants ($F1-F3$) were 150, 800, and 2100 Hz, respectively, at stimulus onset. Frequencies for all three formants changed linearly over the next 40 ms to 750, 1220, and 2600 Hz for $F1-F3$. Formant frequencies remained at these values for the remainder of the 250-ms total duration.

Voice onset time was modeled acoustically by varying amplitude of voicing (Klatt parameter AV, set to zero during VOT duration), noise in the signal (AH=55 during VOT duration) and amplitude of the first formant (A1=0 during VOT duration, to model $F1$ cutback). These parameters were varied in 5-ms steps to mimic acoustic changes from 5 to 95 ms VOT. From this base set of stimuli, fundamental frequency ($f0$) was manipulated to create a full stimulus corpus that varied in VOT (5–95 ms) and $f0$. To accomplish

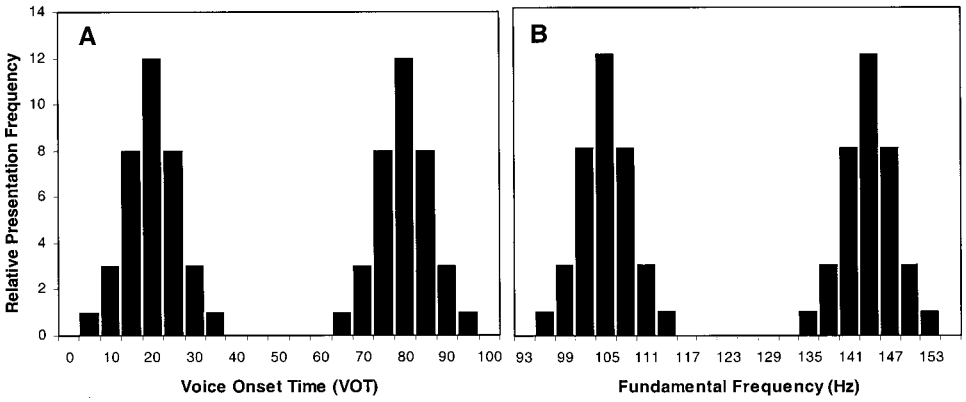

FIG. 1. Sampling distributions for voice onset time [VOT, panel (a)] and fundamental frequency [$f0$, panel (b)] from which stimuli were drawn for presentation to quail during training.

this, the 19-member series varying perceptually from [ba]-[pa] was synthesized at 14 different $f0$'s, 96–114 Hz in 3-Hz steps and 135–153 Hz in 3-Hz steps. This created an $f0 \times$VOT stimulus space consisting of a corpus of 266 stimuli (19 VOT values $\times 14$ $f0$'s). Two distinct ranges for $f0$ were created to provide a ''low'' versus ''high'' distinction, with non-overlapping values roughly corresponding perceptually to male and female voice. For each stimulus, $f0$ was constant across the entire stimulus duration.[3]

b. Stimulus sampling. Typically, speech perception experiments that examine phonetic labeling do so using one or several phonetic series that vary perceptually from one clearly identifiable phonetic endpoint to another via an acoustic manipulation. This is generally true of both human and nonhuman studies of speech perception. Here, one of the primary goals is to examine the role of experience in shaping perception. As a result, the traditional approach is adapted to better model some of the statistical characteristics of speech sound distributions that human listeners encounter.

Although there have been few large-scale efforts to measure the acoustic characteristics of multiple phonetic segments across multiple speakers, those that exist (e.g., Peterson and Barney, 1952; Lisker and Abramson, 1964) suggest that there is a good deal of variability in the acoustic characteristics of speech sounds across speakers. Fortunately for listeners, there is also a good deal of regularity. In an early inventory of cross-language stop-consonant voicing, for example, Lisker and Abramson (1964) observed between- and within-speaker variation in VOT. However, they also reported very regular underlying acoustic patterns for the phonemes of a particular language. Across speakers and productions, estimated VOT values tended to cluster around a particular mean value that occurred most frequently across productions of a particular phoneme (see Newman, 1997). In addition, there was variance such that values adjacent to the mean were also observed, but less frequently. To put this observation in more concrete terms, the measured values roughly approximated a normal (Gaussian) distribution. Therefore, there is reason to believe that normal distributions are a reasonable approximation of the distributions underlying variability in speech production.

In line with these observations, stimuli from the VOT $\times f0$ stimulus space were not presented equally often during the experiment. Rather, there was a statistical structure to the manner in which stimuli were sampled from the space. Independent distributions for sampling $f0$ and VOT were created by modeling Gaussian (normal) distributions with variance of 1.25 stimulus steps (based upon series of 3-Hz steps for $f0$ and 5-ms steps for VOT). For each dimension, two distributions were created. These distributions corresponded to high versus low $f0$ (distribution means were 105 and 144 Hz, respectively) and voiced versus voiceless consonants (with distribution means of 20 and 80 ms).[4] Stimuli were presented with relative frequencies that were discrete approximations of these continuous Gaussian distributions. Figure 1 illustrates relative frequencies for $f0$ and VOT values. Using distributions of $f0$ and VOT values rather than individual stimuli (e.g., one stimulus with a low $f0$ versus one with a high $f0$) allowed for a more sensitive test of interactions between $f0$ and VOT because it encouraged quail to generalize to novel stimuli. Likewise, it provided a more realistic model of $f0$/VOT covariation.

c. Stimulus presentation. Stimuli were synthesized with 12-bit resolution at a 10-kHz sampling rate, matched in rms energy and stored on a computer disk. Stimulus presentation was under the control of an 80386 computer. After D/A conversion (Ariel DSP-16), stimuli were low-pass filtered (4.8-kHz cutoff frequency, Frequency Devices #677), amplified, and presented to quail via a single 13-cm speaker (Peerless 11592) in a tuned enclosure providing flat frequency response from 40 to 5000 Hz. Sound level was calibrated by placing a small sound-level meter (Bruel & Kjaer 2232) in the chamber with the microphone positioned at approximately the same height and distance from the speaker as a bird's head.

B. Procedure

1. Training stimuli

Reinforcement contingencies were structured to train quail to respond differentially to training stimuli drawn from the VOT distributions shown in Fig. 1(a) (i.e., 5–35 ms VOT versus 65–90 ms VOT). Half of the birds in each condition were rewarded for pecking in response to voiced stimuli (5–35 ms VOT, designated +voice birds). Longer VOT (65–95 ms) signaled reinforcement for the remaining quail (−voice birds). The exact stimuli that were presented to quail in a given training session were determined by randomly sampling from the Gaussian distributions described earlier (see Fig. 1). Stimuli were constrained in the number

of positive versus negative stimuli that could occur in a session, but otherwise $f0$ and VOT were randomly selected from the appropriate distribution.

During training, stimuli varied by condition. One group of quail was trained with stimuli exhibiting ''natural'' covariance. That is, voiced stimuli had low $f0$ and voiceless stimuli had high $f0$. Quail in this group heard stimuli created to mimic 5–35-ms VOT that were synthesized with an $f0$ varying between 96 and 114 Hz and 65–95-ms VOT stimuli with an $f0$ of 135–153 Hz. The matching between $f0$ and VOT adhered to these constraints, although the precise $f0$/VOT stimuli varied randomly from within the sampling distributions (see Fig. 1). Another group of quail was trained on the ''reverse'' of this covariation. These birds heard 5–35-ms VOT stimuli synthesized with an $f0$ of 135–153 Hz and 65–95-ms VOT stimuli synthesized with an $f0$ of 96–114 Hz. For a final group of ''control'' quail, $f0$ and VOT did not covary. Stimuli presented to these quail had random assignment of $f0$ to VOT. For this group of quail, $f0$ and VOT values were chosen from the sampling distributions in the same manner as for the other quail. However, no constraints upon $f0$ to VOT mapping were enforced; $f0$ was assigned randomly to VOT values. For any presentation, $f0$ could be chosen from either the high or the low distribution, independent of VOT value.

2. Training procedure

Following 18 to 22 h of food deprivation (adjusted to each bird individually for optimal performance)[5] birds were weighed and placed in a small sound-attenuated chamber within a larger single-wall sound-attenuated booth (Suttle Acoustics Corp.). In a go/no-go identification task, quail pecked a lighted key (1.2 cm square) located 15 cm above the floor and centered below a speaker from which stimuli were presented. A computer recorded responses and controlled reinforcement.

Following magazine training and autoshaping procedures, reinforcement contingencies were gradually introduced over 8 days in sessions of 60 to 72 trials. During this time, average amplitude of the stimuli was increased from 50 to 70 dB SPL to introduce 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 increased from 5 to 30 s, and the ratio of positive to negative trials decreased from 4:1 to 1:1. After the gradual introduction of reinforcement contingencies over eight days, daily training sessions consisted of 72 stimuli (36 positive and 36 negative).

On each trial, a stimulus was presented repeatedly once per 1550 ms at an average peak level of 70 dB SPL. On a trial-by-trial basis, the intensity of stimuli varied randomly from the mean of 70 dB by ± 0–5 dB via a computer-controlled digital attenuator (Analog Devices 7111). The average duration of each trial was 30 s, varying geometrically from 10 to 65 s. The intertrial interval was 15 s. Responses to positive stimuli were reinforced on a variable-interval schedule by 1.5–2.5 s of access to food from a hopper beneath the peck key. Duration of reinforcement was also adjusted for each bird for consistent performance. Average interval to reinforcement was 30 s (10–65 s) so that positive stimuli were reinforced on an average of once per trial. Note that when a trial was long (e.g., 65 s) and times to reinforcement were short (e.g., 10 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. Any reinforcement interval that did not expire during one positive trial carried over to the next positive trial. Such intermittent reinforcement encouraged consistent peck rates during later non-reinforced testing trials. During negative trials, birds were required to refrain from pecking for 5 s for the trial to be terminated. This procedure has been used successfully to train Japanese quail in other speech perception tasks (Kluender, 1991; Kluender and Lotto, 1994; Lotto *et al.*, 1997, 1999).

3. Testing

All birds learned quickly to respond differentially to VOT. Birds continued to train with the distributions until they responded with 10:1 performance for positive versus negative stimuli. Among the 21 quail that began magazine and autoshape training, 16 quail made it to criterion performance and were entered in the testing procedure. Of these birds, five were in the ''natural'' condition, seven were in the ''reverse'' condition, and four were in the ''control'' condition.

Following training, quail were tested on novel, intermediate members of the [ba]-[pa] series (VOT from 40 to 60 ms in 5-ms steps) synthesized with $f0$ of 105 and 141 Hz, the modes of the $f0$ sampling distributions of training stimuli. In all, ten stimuli were tested (2 $f0$'s$\times$5 novel intermediate series members). In each daily test session, the ten test trials (each with a fixed trial duration of 30 s) were randomly interspersed among normal training trials. Each testing session began with 15 trials of training stimuli. Novel test trials could not occur until after these 15 trials as an assurance that birds had ''settled into'' the task before responding to test stimuli. After these initial trials, 10 test trials were randomly interspersed with 60 training trials for a total of 70 trials per test session. Contingencies remained the same for training stimuli, but during test trials no contingencies were in effect. Birds neither received food reinforcement nor needed to refrain from pecking for presentation to terminate after 30 s. Training and testing stimuli were randomly ordered for each bird. Testing continued for 20 daily sessions, providing a data set consisting of birds' peck responses to 20 repetitions of each of the ten test stimuli.

C. Results

The data set was submitted to an analysis of peck responses across high and low $f0$ for all test stimuli. For each bird, raw pecks were collected for each test trial. There is inherent variance in peck rates across individual birds. Therefore, total pecks to each test stimulus were summated across the 20 repetitions of the novel stimuli. Mean peck rates (i.e., pecks per 30-s trial) were calculated for both high- and low-$f0$ test stimuli. These means were then transformed into normalized peck rates by dividing peck rates to test

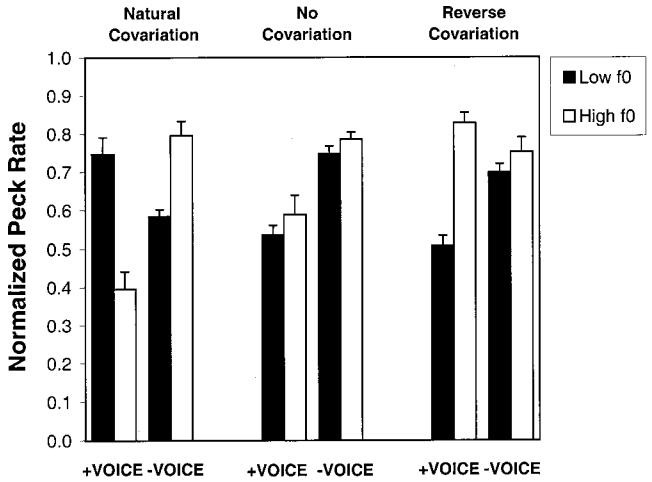

FIG. 2. Average normalized peck rates to low (105 Hz, black bars) versus high (141 Hz, white bars) $f0$. For each VOT$\times f0$ covariation condition (natural, reverse, control), data is presented for birds reinforced to peck to voiced stimuli (+voice) and for those reinforced to peck to voiceless stimuli (−voice).

stimuli by the individual quail's highest peck rate to the ten test stimuli. This transformation adjusted peck rates to a scale between zero and one for each bird, thus minimizing the variance that arises from the fact that some birds are ''heavier'' peckers than others (Bush *et al.*, 1993). This normalization method has been used previously to mitigate natural variance across individual animals (e.g., Lotto *et al.*, 1997).

Normalized mean peck rates and corresponding standard errors are presented in Fig. 2. Data are displayed by $f0$ (black bars correspond to $f0=105$ Hz, white bars show $f0=141$ Hz), sorted by condition (natural, control, reverse), and presented for +voice and −voice birds. Matched-pairs *t*-tests were computed for the difference between normalized peck rates to low- and high-$f0$ test stimuli for +/−voice birds in each condition.

1. Influence of ''natural'' f0/VOT covariation

Birds that were trained to peck in response to *voiced* consonants and heard natural covariation of $f0$ and VOT during training demonstrated a difference in their response to novel stimuli as a function of $f0$. These quail pecked significantly more ($t=4.80$, $p<0.01$) to novel, intermediate VOT series members synthesized with a *low* $f0$ (0.75 average normalized rate) than to the same stimuli synthesized with a *higher* $f0$ (0.39 average normalized rate). Quail trained to peck to *voiceless* consonants in the natural condition also exhibited a significant shift in behavior as a function of $f0$ ($t=3.02$, $p<0.01$), pecking more robustly to novel stimuli with a *higher* $f0$ (0.79 average normalized rate) than to those with a *lower* $f0$ (0.58 average normalized rate).

Thus, these birds' responses to novel stimuli mirrored natural covariation of $f0$ and VOT. Quail trained to peck in response to voiced stimuli pecked most vigorously to novel stimuli with a *low* $f0$, whereas quail trained to peck to voiceless stimuli responded most to stimuli with a *high* $f0$. This avian pattern of results mirrors data that have been observed in human perception (Chistovich, 1969; Haggard *et al.*, 1970; Fujimura, 1971; Massaro and Cohen, 1976, 1977; Haggard *et al.*, 1981; Whalen *et al.*, 1993; Castleman and Diehl, 1996). Stimuli with high $f0$ tend to be labeled as voiceless whereas otherwise similar stimuli with a low $f0$ are more often labeled as voiced.

These data demonstrate that the influence of $f0$ upon voicing need not arise from species-specific mechanisms. However, this single condition does not allow determination of whether the general mechanisms that may govern the interaction of $f0$ and voicing arise from auditory perceptual interactions or from experience with $f0$/VOT covariation. Quail in this condition experienced $f0$ and VOT covariation that modeled the covariation found among many of the world's languages. Although it is possible that experience with this pattern of covariation may have influenced their perception of novel stimuli, it is also possible that auditory interactions led them to respond to stimuli with a higher $f0$ as better exemplars of voiceless consonants than those with lower $f0$. Thus, it is necessary to turn to the behavior of quail in the remaining conditions to evaluate the relative influence of auditory constraints and learning.

2. Outcome of the control condition: No f0/VOT covariation

First, consider the control condition in which $f0$ and VOT did not covary during training. For stimuli presented to quail in this condition, $f0$ and voicing characteristics varied independently. If experience with covariation rather than auditory enhancement explains the effect of $f0$ upon VOT labeling for ''natural'' quail, there should be no effect of $f0$ upon responses of control quail in this condition. However, if the auditory system conspires to bias low-$f0$ stimuli to be responded to as short VOT stimuli, control quail should peck in a manner that mirrors effects typically observed for English listeners.

In fact, quail in the control condition did *not* demonstrate an influence of $f0$ upon their pecking behavior. With no covariation between $f0$ and VOT in the training stimuli, quail exhibited no effect of $f0$ on response to novel stimuli. Neither the birds trained to peck to *voiced* consonants ($t=0.56$, $p=0.30$, average normalized peck rates of 0.53 and 0.58 for low and high $f0$, respectively) nor those trained to respond to *voiceless* consonants ($t=1.05$, $p=0.18$, average normalized peck rates of 0.75 and 0.79 for low and high $f0$, respectively) were influenced by $f0$.

3. Influence of ''reverse'' f0/VOT covariation

Consider, now, the final condition. If experience with $f0$/VOT covariation is responsible for effects observed in the ''natural'' quail, then covariation in the opposite direction (i.e., low $f0$ paired with voiceless consonants) should influence birds' behavior in the opposite manner. In fact, birds trained to peck to *voiced* consonants for which covariation with $f0$ was in the direction opposite natural covariation exhibited a significant difference in their pecking behavior contingent on $f0$ ($t=4.26$, $p<0.01$), pecking most vigorously to stimuli with high $f0$ (0.83 average normalized rate) and less to stimuli with low $f0$ (0.51 average normalized rate). Birds' behavior mirrored the covariation inherent in

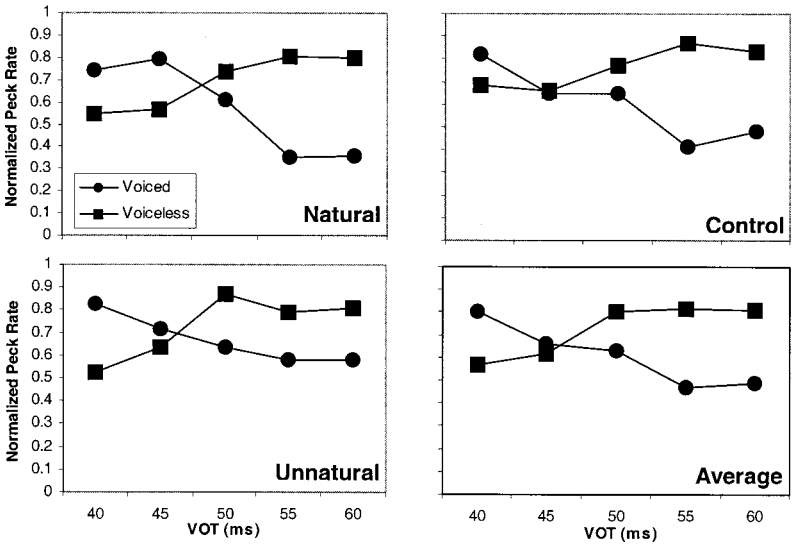

FIG. 3. Average normalized peck rates collapsed across $f0$ for each condition and averaged across conditions. Square symbols indicate −voice birds responses to novel test stimuli that vary across the VOT series. Circles correspond to +voice birds responses. These data indicate that quail did not perform the task simply by responding to changes in $f0$.

their training stimuli and was opposite the direction predicted by auditory enhancement hypothesis. Thus, these data suggest that the influence of $f0$ on VOT labeling is not bound to correspondence with the pattern typically observed in the world's languages, but rather is influenced by the pattern of $f0$/VOT covariation present in the speech input.

The results for birds trained to peck to *voiceless* consonants are less clear in that they did not exhibit a significant effect of $f0$ upon their response to novel stimuli ($t=.85$, $p=0.22$). Low $f0$ (0.70 average normalized rate) and high $f0$ (0.75 average normalized rate) did not differentially influence quails' pecking behavior to novel stimuli. It appears likely that the failure to find an influence of $f0$ upon birds' response to novel stimuli in this condition is related to rather high variability. Two of the four birds in this condition did exhibit a modest influence of $f0$ upon response to novel stimuli in the direction predicted by their experience. However, the remaining quail did not differentially respond as a function of $f0$.

To summarize these results across conditions and counterbalancing, birds in five of the six conditions tested (3 types of experience×2 mappings to voicing) exhibited behavior that supports the hypothesis that experience with covariation is fundamental to effects of $f0$ upon voicing perception.

4. Influence of VOT on birds' responses

The results are therefore in line with the hypothesis that the influence of $f0$ is related to experience with $f0$/VOT covariance. However, the data presented thus far have reported only the influence of $f0$. It is possible that these results reflect a tendency by birds to respond solely on the basis of $f0$, to the exclusion of VOT. If this is the case, then the results do not bear upon the hypotheses under investigation, but rather merely reflect the ability of quail to map $f0$ variation to a pecking response. To examine this possibility, it is necessary to inspect birds' responses across novel VOT stimuli, independent of $f0$. Figure 3, which illustrates these data, suggests that the birds' responses were sensitive to VOT. Data points in Fig. 3 correspond to average normalized peck rates to novel stimuli collapsed across the two test $f0$'s for +voice and −voice quail. Quail differentially responded to test stimuli varying in VOT, demonstrating that their behavior was not solely the result of sensitivity to $f0$. Note that the data points illustrated are drawn from the middle of the series and can thus be expected to be ''boundary'' stimuli. Broader identification functions are common for animal subjects. Typically, this characteristic is taken as indicative of attentional differences between animals and humans (e.g., Kuhl and Miller, 1978), but it also may be due to the animals' more limited experience with speech sounds. Quail trained to respond to +voice exhibited a very different pattern of response to novel test stimuli than did their −voice counterparts. This indicates that although $f0$ influenced quails' responses, they used both $f0$ and VOT to perform the task.

III. DISCUSSION

The objective of the present experiment was to examine the relative contributions of stable auditory perceptual interactions and experience with covariation in understanding the influence of $f0$ upon [VOICE] labeling. Many studies have demonstrated that listeners' categorization of synthetic or digitally manipulated natural speech varying perceptually from voiced to voiceless across a phonetic series can be shifted by changes in $f0$. Stimuli with lower $f0$'s are more often categorized as voiced consonants whereas stimuli with otherwise identical acoustic characteristics, but with higher $f0$'s, tend to be labeled as voiceless. This pattern of perception mimics patterns of speech production commonly observed across languages. Voiceless consonants tend to be produced with higher $f0$'s than their voiced counterparts.

The mechanisms behind this phenomenon are largely unknown, but there are at least two prominent hypotheses. The first hypothesis, in line with the tenets of auditory enhancement (Diehl and Kluender, 1989; Kingston and Diehl, 1994; Diehl *et al.*, 1995), suggests that general auditory interactions among mutually enhancing acoustic characteristics of $f0$ and VOT couple to improve the intelligibility of voiced consonants with low $f0$ and voiceless consonants

with high $f0$. Another possibility is that experience with covariation of $f0$ and voicing within the speech signal is responsible for effects of $f0$ upon [VOICE] labeling.

Using a nonhuman animal model, these two hypotheses were teased apart in the present experiment. Results from quail in the ''natural'' and ''reverse'' covariation conditions demonstrate that experience with $f0$/VOT covariance influences quails' response to novel stimuli. Quail that responded to +voice and −voice in the ''natural'' and those who pecked to +voice in the ''reverse'' condition pecked more often to $f0$/VOT combinations that matched their pattern of experience. One subset of these quail (those who were trained to peck to −voice stimuli in the ''reverse'' condition) did not adhere to this pattern of results. However, $f0$ had a null effect on responses to novel stimuli for these quail, so it is difficult to interpret these data. Overall, three of the four groups of quail in conditions where $f0$ covaried with VOT demonstrated an effect of $f0$ upon response to novel stimuli that mirrored the covariance experienced during training. The data of the +voice quail in the reverse condition, especially, are difficult to explain from an account that relies upon auditory interactions because they suggest $f0$ does not exert an obligatory influence on perception of [VOICE] consonants in the absence of covariation with VOT.

The results of the ''control'' condition complement these findings. Quail that did not experience regularity in $f0$/VOT covariation during training showed no shift in response to test stimuli contingent on $f0$. These results suggest that the influence of $f0$ is not strictly related to stable mutually enhancing interactions that are auditory in nature.

A. Ties to previous experiments

The present results may relate well to findings of some previous experiments that examined human listeners' perception of voicing as a function of $f0$. For example, Bernstein (1983) found that adult listeners make use of $f0$ in identifying words that vary in word-initial voicing (*gate* versus *Kate*), but 4- and 6-year-old children do not. These results are consistent with the hypothesis that experience may play a role in determining the influence of $f0$ upon perception of consonantal voicing.

Likewise, Haggard *et al.* (1981) reported an intriguing cross-linguistic difference that arose serendipitously from a study of the influence of $f0$ upon phoneme boundaries between voiced and voiceless consonants. One of their 35 listeners identified members of a series of stimuli varying in VOT quite differently from the rest of the listeners tested. This listener exhibited uncharacteristically flat identification functions across VOT, suggesting that she relied almost entirely upon the $f0$ variation and almost completely disregarded variation in VOT. Upon recalling the listener to determine whether she suffered from a hearing deficit, Haggard *et al.* found that the listener had been born in Italy and had emigrated to the United States at age five, though she had no command of Italian as an adult. Haggard *et al.* suggested that these odd data might underscore the importance of early learning in determining phoneme boundaries and acoustic cue weighting in speech perception. Massaro and Cohen (1976, 1977) have also reported marked individual differences in the influence of $f0$ on listeners' categorization of fricatives as /s/ versus /z/.

Although neither of these results (one a null finding for children, the other based on a single listener) is exceptionally strong evidence that experience plays a role in determining the influence of $f0$ upon perception of voicing, coupled with the findings of the present experiment, they hint at a role for experience and offer intriguing possibilities for further research.

Sinnott and Saporita (2000) recently have presented data from American English and Spanish adults as well as monkeys (*Macaca fuscata*) on the perceptual influence of first formant ($F1$) transition onset frequency covariation with gap duration in a speech series that varied from *say* to *stay*. Unlike English, Spanish does not have native consonantal cluster contrasts like *say* versus *stay*. Incremental increases in gap duration in Sinnott and Saporita's stimuli caused perception to change from *say* to *stay* for all subjects. However, subjects varied in their use of the $F1$ onset cue, which covaried with gap duration. American English listeners exhibited a strong influence of $F1$-onset frequency on stimulus identification. Spanish listeners differed in the degree to which they used $F1$ onset as a cue. Monkeys did not appear to use $F1$ onset at all. Sinnott and Saporita (2000) suggest that the important factor delineating these subject populations is degree of exposure to English and thus to $F1$ onset and gap duration covariation. These findings thus appear to suggest a perceptual learning component for another example of cue covariation in speech perception.

B. The generality of auditory mechanisms

There is at least one important criticism that could be leveled against the current experiment. The present results are much less clear if the avian auditory system of the Japanese quail is sufficiently different from the human auditory system so as not to capture putatively important characteristics that may contribute to auditory interactions between $f0$ and voicing. Though there are significant anatomical and physiological distinctions between avian and human auditory systems (see, e.g., Popper and Fay, 1980), there are several reasons to believe that the data presented here reliably represent audition quite generally. First, avian species have been shown to exhibit a number of effects in speech perception that rely upon audition, with no influence of learning (e.g., Kluender and Lotto, 1994; Dooling *et al.*, 1995; Dent *et al.*, 1997; Lotto *et al.*, 1997). These effects quite closely mirror human perceptual results, so despite distinctions between avian and human auditory systems, it appears that there is a great deal of functional correspondence. Furthermore, the present stimuli differed critically in VOT and $f0$—two acoustic cues that rely on rather low-frequency hearing. Previously, Kluender and Lotto (1994; Kluender, 1991) have shown that quail are quite capable of this sort of task, exhibiting effects of F1 on their response to voiced and voiceless stimuli. Thus, the conclusions drawn from the observation that quail in the ''control'' condition failed to exhibit an effect of $f0$ upon response to novel stimuli, although depen-

dent upon the functional correspondence of human and avian audition, are supported by these previous positive results.

Holt *et al.* (1999) have presented data from a mammalian model (the chinchilla, *Chinchilla villidera*) performing a similar task that provide further support for the present conclusions. Chinchillas' audiograms closely model those of humans (Henderson *et al.*, 1969). Furthermore, the properties of their auditory system are well mapped and, psychoacoustically, their behavior corresponds quite well with that of humans. For these reasons, chinchillas are one of the most common species used in systems auditory neurophysiology research. As a part of a larger project examining auditory cues to voicing, Holt *et al.* (1999) tested chinchilla perception of VOT as a function of changes in $f0$. Unlike quail in the ''natural'' and ''reverse'' conditions of the current experiment, chinchillas were not trained with covariance between $f0$ and VOT. Thus, the chinchillas' experience closely modeled that of quail in the ''control'' condition. Their pattern of response was also similar. Like control quail, chinchillas did not exhibit an effect of $f0$ upon response to voicing differences. Thus, the generalizability of the current results is supported by the fact that a mammalian species (in possession of an auditory system that more closely models that of humans) also fails to exhibit an effect of $f0$ upon VOT response when there is no history of experience with covariation between $f0$ and VOT.

C. Role of learning in speech perception

The present results should not be taken as evidence against an auditory enhancement account of speech perception. Though its emphasis is very clearly upon general auditory perceptual mechanisms, the auditory enhancement account does not fail to posit a role for learning. Diehl *et al.* (1990), for example, explicitly point out that a successful theory of speech perception must provide both an account of ''the transfer function of the auditory system'' as well as ''the listener's tacit knowledge of speech and language-specific facts that are relevant to phonetic categorization'' (p. 244). They go on to argue that it is presently possible to be much more unequivocal about influences of the auditory system than it is to be explicit about speech-relevant knowledge of listeners. As a consequence, they begin by seeking explanation ''in terms of general auditory mechanisms before appealing to speech-specific tacit knowledge'' (p. 245). We agree that the determination of the role of experience and learning ought to be a fundamental pursuit in understanding speech perception. Data from nonhuman species and those from nonspeech studies (e.g., Kluender *et al.*, 1998; Saffran *et al.*, 1999) suggest that very general learning processes may play an important role. Furthermore, phonetic categorization is unlikely to be the only domain where experience is important to speech perception.

An important lesson from the present work is that structured experience influences perception. The speech signal, in general, is richly structured with regularities imposed both by physical constraints on articulatory processes and by linguistic constraints that shape the habits of talkers. Experience with this structure shapes perception, and nonhuman animal models can contribute to our understanding of these processes.

ACKNOWLEDGMENTS

This work was supported in part by a National Science Foundation Predoctoral Fellowship to the first author. Additional support was provided by NSF Young investigator Award DBS-9258482 to the third author. Some of the data were presented at the 138th Meeting of the Acoustical Society of America in Columbus, OH. The authors thank Eric P. Lotto for his assistance in conducting the experiment and serving as the male voice upon which the experiment's stimuli were based. The authors also gratefully acknowledge the helpful comments of Randy L. Diehl, John Kingston, and Joan Sinnott. Correspondence and requests for reprints should be addressed to Lori L. Holt, Department of Psychology, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213 (e-mail: lholt@andrew.cmu.edu).

[1]Although it is accurate to refer to the syllable-initial English voiceless stops that have mainly been examined in these studies as ''voiceless aspirated'' stops, they will be referred to here simply as ''voiceless'' stops for ease of reading.

[2]There are many acoustic cues that correlate with voicing. For example, presence of voicing during consonant constriction, low first formant ($F1$) near consonant constriction, absence of significant aspiration after consonant release, relatively short closure interval, and relatively long preceding vowel are all effective perceptual cues to the [VOICE] contrast. Lisker and Abramson (1964) demonstrated that the primary acoustic correlate of voicing is variation in voice onset time (VOT) in utterance-initial position. In its most precise usage, VOT refers to an articulatory characteristic—namely, the interval of time between the release of a stop consonant and the onset of voicing of the following vowel. Hereafter, we abandon the cumbersome phrase ''acoustic correlates of voice onset time'' and extend the usage of VOT to include the acoustic effects of the VOT as well, sacrificing precision but preserving readability. The *Stimulus* section (11.A.2) describes the precise acoustic cues synthesized to model voicing in the present experiment.

[3]There has been a good deal of debate about whether overall frequency of $f0$ or direction of $f0$ contour contributes more reliable cues to voicing (e.g., Lea, 1973; Umeda, 1981; Ohde, 1982; Silverman, 1986; Castleman and Diehl, 1996). Rather than modeling these more complex aspects of $f0$, the present stimuli had a flat $f0$. As a result, these stimuli might be thought to model initial or peak $f0$, each of which has been found to systematically vary as a function of voicing (Umeda, 1981). There is substantial evidence that stimuli synthesized with flat $f0$ contours influence listeners' perception of voicing.

[4]These distributions differ in some ways from what is typical of English voicing categories. For example, the modal VOT values for the voiced and voiceless distributions were 20 and 80 ms, respectively. The mid-point ''boundary'' for these stimuli was thus approximately 50 ms VOT whereas, in English, the labial VOT boundary is approximately 25 ms (Lisker and Abramson, 1964). This difference was tolerated in an effort to avoid inclusion of stimuli with VOT values less than or equal to 0 ms for fear that the 0-ms boundary might introduce unwanted discontinuities in the stimulus set. Another difference is that the standard deviations of our stimuli were equivalent across voiced and voiceless modes. Lisker and Abramson (1964) have shown that the standard deviations of VOT distributions depend on modal VOT; shorter VOT categories tend to have smaller standard deviations than longer VOT categories. This detail of distributions was not modeled here. Manipulations of both distribution modes and standard deviations may prove to be interesting variables for further research of how complex phonetic categories are learned. However, as a starting point, we chose to test the hypothesis in the most straightforward manner.

[5]Optimal performance was operationally defined as the highest ratio of pecks to positive versus negative stimuli. Birds are idiosyncratic in the amount of food deprivation necessary to achieve stable optimal perfor-

mance. Weights ranged from 85% to 100% of free-feed weight during training and testing.

Bernstein, L. (**1983**). ''Perceptual development for labeling words varying in voice onset time and fundamental frequency,'' J. Phonetics **11**, 383–393.

Bush, L. L., Hess, U., and Wolford, G. (**1993**). ''Transformations for within-subjects designs: A Monte Carlo investigation,'' Psychol. Bull. **113**, 566–579.

Caisse, M. (**1982**). ''Cross-linguistic differences in fundamental frequency perturbation induced by voiceless unaspirated stops,'' M. A. thesis, Univ. of California—Berkeley.

Castleman, W. A., and Diehl, R. L. (**1996**). ''Effects of fundamental frequency on medial and final [voice] judgments,'' J. Phonetics **24**, 383–398.

Chistovich, L. A. (**1969**). ''Variations of the fundamental voice pitch as a discriminatory cue for consonants,'' Sov. Phys. Acoust. **14**, 372–378.

Dent, M. L., Brittan-Powell, E. F., Dooling, R. J., and Pierce, A. (**1997**). ''Perception of synthetic /ba/-/wa/ speech continuum by budgerigars (*Melopsittacus undulates*),'' J. Acoust. Soc. Am. **102**, 1891–1897.

Derr, M. A., and Massaro, D. W. (**1980**). ''The contribution of vowel duration, f_0 contour, and frication duration as cues to the /juz/-/jus/ distinction,'' Percept. Psychophys. **27**, 51–59.

Diehl, R. L., and Kluender, K. R. (**1989**). ''On the objects of speech perception,'' Ecological Psychol. **1**, 121–144.

Diehl, R. L., Castleman, W. A., and Kingston, J. (**1995**). ''On the internal perceptual structure of phonological features: The [voice] distinction,'' J. Acoust. Soc. Am. **97**, 3333–3334.

Diehl, R. L., Kluender, K. R., and Walsh, M. A. (**1990**). ''Some auditory bases of speech perception and production,'' in *Advances in Speech, Hearing, and Language Processing, Volume 1*, edited by W. A. Ainsworth (JAI, London), pp. 243–267.

Dooling, R. J., and Okanoya, K. (**1995**). ''The method of constant stimuli in testing auditory sensitivity in small birds,'' in *Methods in Comparative Psychoacoustics*, edited by G. M. Klump, R. J. Dooling, R. R. Fay, and W. C. Stebbins (Birkhäuser Verlag, Basel, Switzerland), pp. 161–169.

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.

Fujimura, O. (**1971**). ''Remarks on stop consonants: Synthesis experiments and acoustic cues,'' in *Form and Substance: Phonetic and Linguistic Papers Presented to Eli Fischer-Jørgensen*, edited by L. L. Hammerich, R. Jakobson, and E. Zwirner (Akademisk Forlag, Copenhagen), pp. 221–232.

Gruenenfelder, T. M., and Pisoni, D. B. (**1980**). ''Fundamental frequency as a cue to postvocalic consonantal voicing: Some data from speech perception and production,'' Percept. Psychophys. **28**, 514–520.

Haggard, M., Ambler, S., and Callow, M. (**1970**). ''Pitch as a voicing cue,'' J. Acoust. Soc. Am. **47**, 613–617.

Haggard, M., Summerfield, Q., and Roberts, M. (**1981**). ''Psychoacoustical and cultural determinants of phoneme boundaries: Evidence from trading F_0 cues in the voiced-voiceless distinction,'' J. Phonetics **9**, 49–62.

Henderson, D., Onishi, S., Eldredge, D. H., and Davis, H. (**1969**). ''A comparison of chinchilla auditory evoked response and behavioral response thresholds,'' Percept. Psychophys. **5**, 41–45.

Holt, L. L., Lotto, A. J., and Kluender, K. R. (**1998**). ''Incorporating principles of general learning in theories of language acquisition,'' in *Chicago Linguistic Society, Volume 34: The Panels*, edited by M. Gruber, C. Derrick Higgins, K. S. Olson, and T. Wysocki (Chicago Linguistics Society, Chicago), pp. 253–268.

Holt, L. L., Lotto, A. J., and Kluender, K. R. (**1999**). ''Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement?'' J. Acoust. Soc. Am. **106**, 2247(A).

Hombert, J. M. (**1978**). ''Consonant types, vowel quality, and tone,'' in *Tone: A Linguistic Survey*, edited by V. Fromkin (Academic, New York), pp. 77–111.

House, A. S., and Fairbanks, G. (**1953**). ''The influence of consonant environment on the secondary acoustical characteristics of vowels,'' J. Acoust. Soc. Am. **25**, 105–135.

Kingston, J. (**1986**). ''Are F_0 differences after stops deliberate or accidental?'' J. Acoust. Soc. Am. **79**, S27(A).

Kingston, J., and Diehl, R. L. (**1994**). ''Phonetic knowledge,'' Lang. **70**, 419–454.

Klatt, D. K. (**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., 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.

Kluender, K. R., Lotto, A. J., Holt, L. L., and Bloedel, S. L. (**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.

Kohler, K. J. (**1982**). ''F_0 in the production of lenis and fortis plosives,'' Phonetica **39**, 199–218.

Kohler, K. J. (**1984**). ''Phonetic explanation in phonology. The feature fortis/lenis,'' Phonetica **41**, 150–174.

Kohler, K. J. (**1985**). ''F_0 in the perception of lenis and fortis plosives,'' J. Acoust. Soc. Am. **78**, 21–32.

Kohler, K. J., and van Dommelen, W. A. (**1986**). ''Prosodic effects on lenis/fortis perception: Preplosive F_0 and LPSC synthesis,'' Phonetica **43**, 70–75.

Kuhl, P. K., and Miller, J. D. (**1975**). ''Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants,'' Science **90**, 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.

Lea, W. (**1973**). ''Segment and suprasegmental influences of fundamental frequency contour,'' Conson. Types and Tone, South. Cal. Occasional Papers in Ling. **1**, 17–69.

Lehiste, I., and Peterson, G. E. (**1961**). ''Some basic considerations in the analysis of intonation,'' J. Acoust. Soc. Am. **33**, 419–425.

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

Lotto, A. J., Holt, L. L., and Kluender, K. R. (**1999**). ''Structure of phonetic categories produced by general learning mechanisms,'' J. Acoust. Soc. Am. **106**, 2247.

Lotto, A. J., Kluender, K. R., and Holt, L. L. (**1997**). ''Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*),'' J. Acoust. Soc. Am. **102**, 1134–1140.

Massaro, D. W., and Cohen, M. M. (**1976**). ''The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction,'' J. Acoust. Soc. Am. **60**, 704–717.

Massaro, D. W., and Cohen, M. M. (**1977**). ''Voice onset time and fundamental frequency as cues to the /zi/-/si/ distinction,'' Percept. Psychophys. **22**, 373–383.

Mohr, B. (**1971**). ''Intrinsic variations in the speech signal,'' Phonetica **23**, 65–93.

Morse, P. A., and Snowdon, C. T. (**1975**). ''An investigation of categorical speech discrimination by rhesus monkeys,'' Percept. Psychophys. **17**, 9–16.

Newman, R. S. (**1997**). ''Individual differences and the link between speech perception and speech production,'' unpublished doctoral dissertation, SUNY Buffalo.

Ohde, R. N. (**1982**). ''The effects of linguistic context on temporal and F0 properties of speech,'' J. Acoust. Soc. Am. **72**, LL5(A).

Ohde, R. N. (**1984**). ''Fundamental frequency as an acoustic correlate of stop consonant voicing,'' J. Acoust. Soc. Am. **75**, 224–230.

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

Peterson, N. R. (**1983**). ''The effect of consonant type on fundamental frequency and larynx height in Danish,'' Annu. Report Inst. Phon., Univ. Copenhagen **17**, 55–86.

Pitt, M. A., and McQueen, J. M. (**1998**). ''Is compensation for coarticulation mediated by the lexicon?'' J. Mem. Lang. **39**, 347–370.

Popper, A. N., and Fay, R. R. (**1980**). *Comparative Studies of Hearing in Vertebrates* (Springer-Verlag, New York).

Saffran, J. R., Aslin, R. N., and Newport, E. L. (**1996**). ''Statistical learning by 8-month-olds,'' Science **274**, 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. (**1999**). ''Statistical learning of tone sequences by human infants and adults,'' Cognition **70**, 27–52.
Silverman, K. E. A. (**1986**). ''F_0 segmental cues depend on intonation: The case of the rise after voiced stops,'' Phonetica **43**, 76–91.
Sinnott, J. M., and Saporita, T. A. (**2000**). ''Differences in American English, Spanish, and monkey perception of the *say-stay* trading relation,'' Percept. Psychophys. **62**, 1312–1319.
Stevens, K. N., and Blumstein, S. E. (**1981**). ''The search for invariant acoustic correlates of phonetic features,'' in *Perspectives on the Study of Speech*, edited by P. D. Eimas and J. L. Miller (Erlbaum, Hillsdale, NJ), pp. 1–38.
Umeda, N. (**1981**). ''Influence of segmental factors on fundamental frequency in fluent speech,'' J. Acoust. Soc. Am. **70**, 350–355.
Waters, R. A., and Wilson, Jr., W. A. (**1976**). ''Speech perception by rhesus monkeys: The voicing distinction in synthesized labial and velar stop consonants,'' Percept. Psychophys. **19**, 285–289.
Whalen, D. H., Abramson, A. S., Lisker, L., and Mody, M. (**1993**). ''$F0$ gives voicing information even with unambiguous voice onset times,'' J. Acoust. Soc. Am. **93**, 2152–2159.