Bulletin of the Psychonomic Society
1987, 25 (2), 139-142

Perception of synthetic speech sounds by the budgerigar (*Melopsittacus undulatus*)

ROBERT J. DOOLING, SIGFRID D. SOLI, ROBERT M. KLINE,
THOMAS J. PARK, CAROLINE HUE, and TIMOTHY BUNNELL
University of Maryland, College Park, Maryland

Two budgerigars were trained to respond differently to two synthetic speech tokens, /da/ and /ta/. Voice onset times (VOTs) of these synthetic stimuli were 0 and +70 msec, respectively. After reaching criterion level performance on these stimuli, the budgerigars were tested with synthetic stimuli having VOTs between 0 and +70 msec. The categorization functions derived from each bird's responses to these intermediate stimuli were similar in shape but differed in their boundary locations from the functions typically obtained from humans and other mammals. In addition, the birds found this speech discrimination task extraordinarily difficult. These results define some of the limitations of the budgerigar's auditory system for the processing of complex acoustic stimuli. Since the budgerigar has a distinctly nonmammalian auditory system, these results also have relevance for current theories of speech perception.

The budgerigar, or parakeet, is a popular cagebird known for its ability to mimic complex acoustic signals, including human speech. The budgerigar appears especially adept at discriminating among complex acoustic signals that fall in the spectral region of 2-4 kHz, where most of the species-specific vocalizations fall (Dooling, 1986). The present experiment tested the budgerigar's ability to discriminate among a set of complex acoustic signals that fall outside this narrow spectral region.

The experimental stimuli were synthetic speech sounds from the familiar alveolar VOT (voice onset time) series used in previous tests of speech perception by humans and other mammals (Kuhl & Miller, 1975, 1978; Soli, 1983). The wealth of data available on the perception of this synthetic speech series by humans and other mammals provides an excellent basis for comparison. These previous data suggest that the voicing distinction among stop consonants is perceived in a similar way by humans, monkeys, and chinchillas (Kuhl, 1981, 1986; Sinnott, Beecher, Moody, & Stebbins, 1976).

Since basic differences in auditory sensitivity exist between mammals and birds (Dooling, 1980), the performance of a bird such as the budgerigar may provide new insights about the degree to which the sound system of speech is influenced by the peripheral auditory system. If basic sensitivities of the auditory system do play a major role, then we would expect that an organism with an avian auditory system may place the labeling boundary in a different location.

METHOD

Subjects

The subjects were two experimentally naive, adult male budgerigars of unknown age obtained from commercial sources. They were housed together in a standard wire breeding cage (46 × 22 × 25 cm) within a large aviary. The birds were kept on a day/night cycle correlated with the season.

Stimuli and Calibration

The speech stimuli were tokens from an alveolar VOT series synthesized with the Haskins Laboratories parallel resonance synthesizer using control parameters developed by Lisker and Abramson (1970). The 400-msec stimuli had VOTs ranging from 0 to +70 msec in 10-msec steps. The programmed formant frequencies as well as the sonograms of these stimuli are the same as those shown schematically in Kuhl and Miller (1978, Figure 2, p. 908). All stimuli were presented at a peak level of 86.5 dB SPL. Sonograms of the eight stimuli are shown in Figure 1.

Test Apparatus

A standard operant test chamber and a modified pigeon grain hopper were used. Light-emitting diodes (LEDs) were attached to the arms of microswitches to serve as manipulanda. A bird could easily trip a microswitch by striking the LED with its beak. Acoustic stimuli were presented to the birds through a 13-cm speaker mounted 24.5 cm in front of the food hopper (Park & Dooling, 1985). All experimental events were controlled by a microcomputer. Stimuli were digitized at 10 kHz and stored on a hard disk for computer-controlled presentation.

Training Procedure

The procedure for training and testing the birds was a standard *go/no-go* operant procedure (Park & Dooling, 1985). The birds were first trained to discriminate between two pure tones (2.0 kHz vs. 3.0 kHz). Once criterion was reached, the birds were trained to discriminate between synthetic speech stimuli with VOTs of 0 msec and +70 msec. A trial began with a response on the observing key. Four repetitions of either /da/ or /ta/ were presented with an interstimulus interval of 250 msec. For one bird, /da/ was the *go* stimulus and /ta/ was the *no-go* stimulus. For the other bird, the stimulus classes were reversed to offset possible response bias characteristic of *go/no-go* tasks.

These experiments were supported by Public Health Service Grant NS-19006 and Research Career Development Award HD-00512 to R. Dooling. We thank Kevin Reed and Brian Dooling for care of the birds and Cathy Biro for the conduct of important pilot experiments. Sigfrid Soli is now at the Biosciences Laboratory, 270-4S-11, 3-M Center, St. Paul, MN 55144. Timothy Bunnell is now at the Sensory Communication Research Laboratory, Gallaudet College, Washington, DC 20002. Requests for reprints should be sent to R. Dooling, Department of Psychology, University of Maryland, College Park, MD 20742.

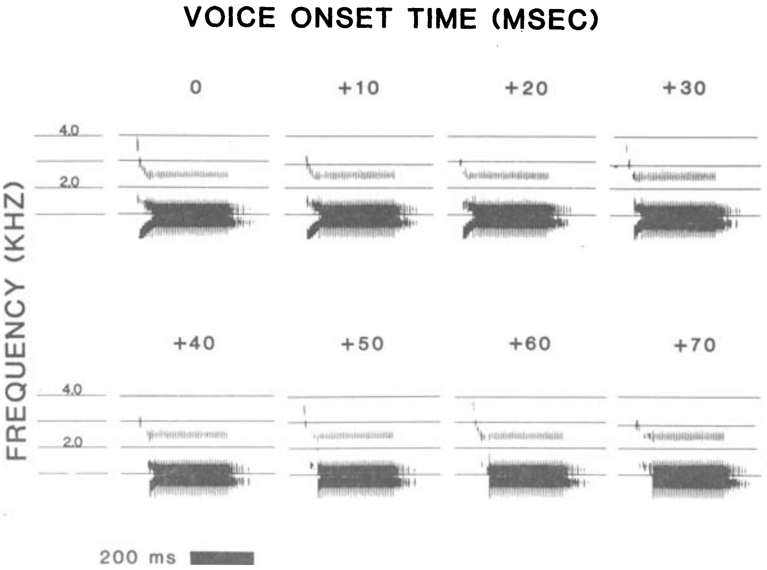

Figure 1. Sonograms of the eight synthetic speech tokens used in this experiment. Frequency is on the ordinate marked in 1-kHz steps. The time marker is 200 msec.

Following the presentation of a *go* stimulus, a response on the left LED within 2 sec resulted in a 2-sec access to grain. A failure to respond within this interval resulted in a 10-sec time-out period during which the lights in the chamber were extinguished. Following the presentation of a *no-go* stimulus, no response on the left LED also resulted in a 2-sec access to grain. An incorrect response to the *no-go* stimulus also resulted in a 10-sec time-out period. Trial type was selected on a quasi-random basis with the restriction that the same trial type could not occur three times in succession. Once performance reached a criterion of 85% correct over several successive sessions, probe trials consisting of stimuli with intermediate VOTs were introduced.

Testing Procedure

The birds were tested in daily sessions consisting of 50 trials: 44 trials were the original training trials and 6 trials were the probe trials consisting of one presentation each of the intermediate VOTs (+10, +20, +30, +40, +50, and +60 msec). To avoid response bias on the probe trials, the birds were reinforced with food regardless of how they responded to the probe stimuli.

Each bird was trained and tested in over 300 sessions, with each session consisting of 50–100 trials. One bird was tested for 75 sessions with probe stimuli, and the other bird was tested for 35 sessions. Probe sessions from the last month of testing in which the birds maintained criterion level performance of 90% correct (or better) were selected for further analysis.

RESULTS

The budgerigars in our study, in contrast with humans and other mammals, had considerable difficulty in learning to discriminate /da/ from /ta/. Both birds required well over 10,000 trials to achieve the modest 85% criterion performance. Even then it was difficult for the birds to consistently maintain a high level of performance.

A generalization or identification function was constructed from the sessions involving presentation of probe stimuli (Kuhl & Miller, 1978). Bird P749 met a criterion of 90% correct in 13 sessions and Bird P14 met the criterion in 17 sessions. The generalization function for each bird was obtained by a visual fitting of a straight line to the data from these sessions plotted on probability paper. The 50% points, or category boundaries, of these curves were at 15 msec for Bird P14 and 25 msec for Bird P749. Thus, the two birds differed in their generalization or identification of the /da/-/ta/ stimuli. These differences were not in the direction predicted by a simple *go* bias in the procedure. The generalization curve for each VOT stimulus and each bird is given in Table 1.

The birds' mean generalization or labeling function is plotted as the percentage of /da/ responses for each stimulus in Figure 2. The labeling functions for humans and chinchillas are also shown. The boundary for the budgerigar is at a smaller VOT value than for English-speaking subjects (Lisker & Abramson, 1970) and chinchillas (Kuhl & Miller, 1978).

DISCUSSION

This study addressed two issues: whether budgerigars could learn to discriminate among phonetically distinct synthetic speech sounds, and whether their generalization or labeling function, especially the 50% point or category boundary, resembles that of humans and chinchillas.

The results show that with extensive training, budgerigars can indeed distinguish the phonetically distinct synthetic /da/ and /ta/. The more revealing finding, however, is the difficulty both budgerigars had in learning this discrimination. Even when criterion performance was achieved, the birds had difficulty maintaining consistent, high-level performance throughout the experiment.

Table 1
Percent of Stimuli Labeled /da/ as a Function of Voice Onset Time (in Milliseconds)

	Voice Onset Time							
Bird	0	10	20	30	40	50	60	70
P14	94	82	0	11	1	1	6	8
P749	82	85	69	76	15	8	0	2

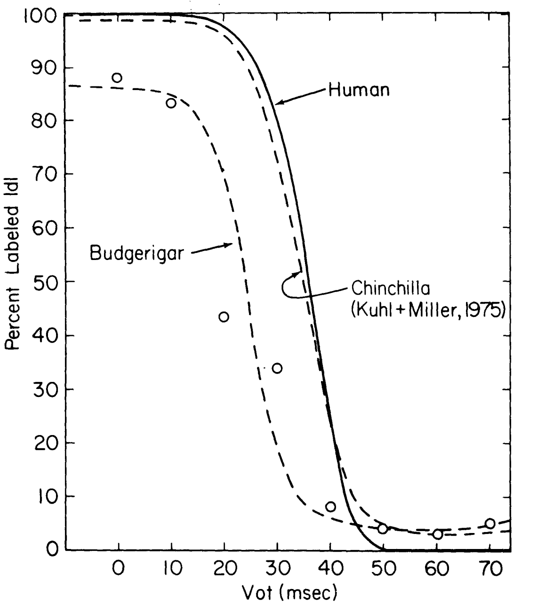

Figure 2. Serial spectral plots of the 0-msec, +30-msec, and +70-msec VOT tokens from Figure 1. Two sets of plots are shown scaled in human critical bandwidths (bark) and budgerigar critical bandwidths. The abscissa is scaled in species-appropriate critical band units. The ordinate is total power per critical bandwidth (dB/z).

The birds' performance on the speech task stands in marked contrast to the performance of budgerigars on similar tasks involving discrimination of bird calls. A budgerigar requires approximately 1,000–1,400 trials to learn a discrimination between two pure tones, two budgerigar contact calls, or two canary calls (Dooling, 1986; Park & Dooling, 1985). The birds in the present experiment, however, required well over 10,000 trials to learn to discriminate between /da/ and /ta/.

In defining the category boundary for budgerigars, we used only those sessions where performance on the endpoints was high enough to ensure that the birds were accurately distinguishing the stimuli. Even with this conservative approach, the category boundaries for the two budgerigars differed from one another and from the boundaries for humans and chinchillas. In spite of the difference between birds, however, it is clear that the location of the boundary is different from that observed in humans and chinchillas. These results suggest that the /da/-/ta/ synthetic speech continuum is perceived differently by budgerigars than by humans and chinchillas.

The mechanism underlying the differences between budgerigar and mammalian categorization functions is not yet clear; however, there are important differences in auditory sensitivity between budgerigars and mammals. Budgerigars hear best between 1.0 and 4.0 kHz, while thresholds below 1 kHz are higher than those of humans and most other mammals (Dooling, 1980). Thus, some of the most important acoustic contrasts among the stimuli (e.g., the spectral and temporal cues associated with F1 cutback [Soli, 1983]) may simply be less audible to the budgerigar.

Budgerigars are not dramatically different from humans and other mammals on most psychoacoustic measures of frequency, intensity, and temporal discrimination, but they are quite different in terms of critical bandwidth (Dooling, 1980). This may be an important factor in the species differences observed in the present experiment. Budgerigars have the narrowest critical bandwidths, between 2.0 and 4.0 kHz, with larger bandwidths (i.e., less frequency selectivity) at lower and higher frequencies (Dooling, 1980, 1986). Below 1.0 kHz, in the spectral region containing important voicing cues, the critical bandwidths of the budgerigar auditory system are much larger than those of humans (Scharf, 1970).

The issue can be addressed by providing a comparative look at how the filtering properties of the budgerigar and human auditory systems transform the acoustic characteristics of three exemplars from the synthetic speech series. Serial spectral plots of the 0-, +30-, and +70-msec tokens are shown in Figure 3. These plots provide a species-appropriate auditory representation (i.e., scaled in critical band units) of the acoustic changes among these tokens. Two serial plots are shown for each token; abscissas are in linear-critical band units appropriate for each species rather than in log-frequency units (Dooling, Bunnell, & Clark, 1984). Critical bandwidths for tokens on the left are appropriate for the budgerigar (Dooling & Saunders, 1975; Saunders, Rintelmann, & Bock, 1979), whereas critical bandwidths for tokens on the right are appropriate for humans (Zwicker, 1965). In all plots, the ordinate is the sum or total power per critical bandwidth.[1]

The most obvious species difference in these plots is that the energy in speech sounds is spread over many more critical bandwidths in the human plot than in the budgerigar plot. Human critical bandwidths are narrower at low frequencies, suggesting that humans may be able to resolve low-frequency spectral details that budgerigars cannot. For example, Figure 3 shows that it is difficult to differentiate between the first and second formants in the serial spectral plots of budgerigars, whereas three formants are clearly visible in the serial spectral plots for humans.

If budgerigars do not use the low-frequency spectral cues that humans use in partitioning the VOT continuum, this could influence the location of the category boundary. Other factors, however, must also contribute to the boundary difference between budgerigars and mammals since chinchilla critical ratios at low frequencies are also considerably larger than those of humans (Miller, 1964). Nevertheless, chinchillas and humans have similar boundaries for the /da/-/ta/ continuum (Kuhl & Miller, 1975).

The results of several previous studies of VOT discrimination with human subjects (e.g., Lisker, 1975; Soli, 1983) suggest that spectral cues, defined by the onset frequency and transition of the first formant, play an important role in the labeling and discrimination of VOT con-

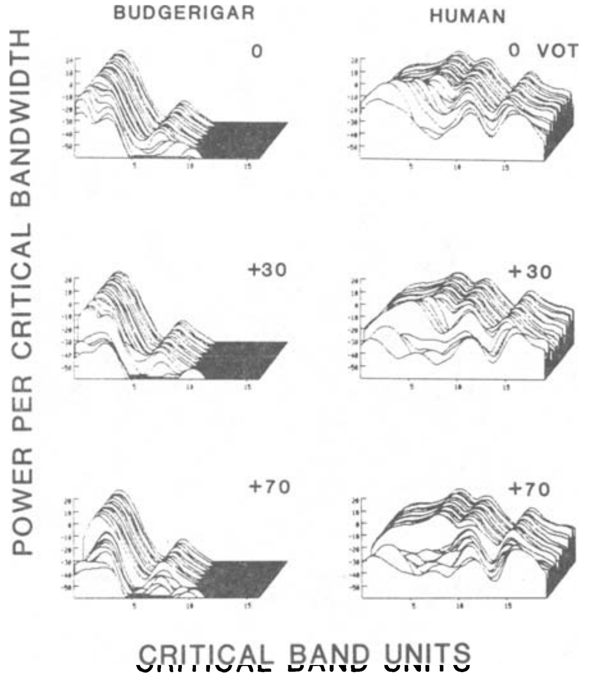

Figure 3. The average identification function is shown for the two budgerigars. For comparison, identification functions are also shown for the human and the chinchilla (replotted from Kuhl & Miller, 1975).

tinua. Figure 3 shows that first (and second) formant frequencies are poorly represented in the budgerigar critical band plot, whereas the third formant transitions are quite well represented. The third formant transitions begin around 4 kHz and drop to about 2.5 kHz during the first 30-40 msec of the stimulus. Budgerigars exhibit their greatest spectral resolution in precisely this spectral region. In stimuli with short VOTs, almost all of the third formant transition is voice-excited, whereas in stimuli with longer VOTs (i.e., greater than 20 msec), this transition is excited by low-amplitude aspiration. The budgerigar critical band spectra, when considered in conjunction with the budgerigars' category boundary at 20 msec VOT, suggest that the birds may have distinguished the stimuli based on the presence or absence of a voiced third formant transition at stimulus onset.

A test of this hypothesis would involve construction of a new stimulus set in which only the third formant transition changed. Another test of this hypothesis might be a simple filtering out of the third formant. Thus, by isolating or eliminating the third formant cue, its influence on the perception of speech by budgerigars could be determined. However, given the difficulty in training budgerigars to discriminate speech, these additional experiments may not be practical.

Although it remains uncertain what the acoustic cue or constellation of cues are that budgerigars use to distinguish the /da/-/ta/ continuum, on average, budgerigars treat the /da/-/ta/ continuum differently than do humans. For this reason, the present results are relevant to current theoretical views about speech perception. A number of animal studies have shown that mammals with auditory systems similar to those in humans also categorize synthetic speech sounds in a way similar to the way humans would (Kuhl & Miller, 1975, 1978; Kuhl & Padden, 1982, 1983). It is significant that a bird—the budgerigar—perceives the /da/-/ta/ continuum differently than do the mammalian species tested so far. These findings support the notion that speech sound contrasts may have evolved to conform to the sensitivities of the mammalian auditory system.

REFERENCES

Dooling, R. J. (1980). Behavior and psychophysics of hearing in birds. In A. Popper & R. Fay (Eds.), *Comparative studies of hearing in vertebrates* (pp. 261-288). New York: Springer-Verlag.

Dooling, R. J. (1986). Perception of vocal signals by budgerigars (*Melopsittacus undulatus*). *Experimental Biology*, **45**, 195-218.

Dooling, R. J., Bunnell, H. T., & Clark, C. (1984). Critical band analysis of avian vocalizations. *Journal of the Acoustical Society of America*, Suppl. 1: Program of the 107th Meeting, S27.

Dooling, R. J., & Saunders, J. C. (1975). Hearing and vocalizations in the parakeet (*Melopsittacus undulatus*). Absolute thresholds, critical ratios, frequency difference limens, and vocalizations. *Journal of Comparative & Physiological Psychology*, **88**, 1-20.

Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America*, **70**, 340-349.

Kuhl, P. K. (1986). The special-mechanisms debate in speech: Contributions of tests on animals (and the relation of these tests to studies using non-speech signals). *Experimental Biology*, **45**, 233-265.

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

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, **63**, 905-917.

Kuhl, P. K., & Padden, D. M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception & Psychophysics*, **32**, 542-550.

Kuhl, P. K., & Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *Journal of the Acoustical Society of America*, **73**, 1003-1010.

Lisker, L. (1975). Is it VOT or a first formant transition detector? *Journal of the Acoustical Society of America*, **57**, 1547-1551.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the 6th International Congress of Phonetic Sciences* (pp. 563-567). Prague, Czechoslovakia: Academia.

Miller, J. D. (1964). Auditory sensitivity of the chinchilla in quiet and in noise. *Journal of the Acoustical Society of America*, **36**(A), 2010.

Park, T. J., & Dooling, R. J. (1985). Perception of species-specific contact calls by budgerigars (*Melopsittacus undulatus*). *Journal of Comparative Psychology*, **99**, 391-402.

Saunders, J. C., Rintelmann, W. F., & Bock, G. R. (1979). Frequency selectivity in bird and man: A comparison among critical ratios, critical bands, and psychophysical tuning curves. *Hearing Research*, **1**, 303-323.

Scharf, B. (1970). Critical bands. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (Vol. 1, pp. 157-202). New York: Academic Press.

Sinnott, J. M., Beecher, M. D., Moody, D. B., & Stebbins, W. C. (1976). Speech sound discrimination by monkeys and humans. *Journal of the Acoustical Society of America*, **60**, 687-695.

Soli, S. D. (1983). The role of spectral cues in discrimination of voice onset time differences. *Journal of the Acoustical Society of America*, **73**, 2150-2165.

Zwicker, E. (1965). Temporal effects in simultaneous masking and loudness. *Journal of the Acoustical Society of America*, **38**, 132-141.

NOTE

1. The procedures for generating the serial spectral plots shown in Figure 3 were as follows. For each frame in the serial display, the magnitude of the discrete Fourier transform (DFT) of a windowed (25.6-msec Hamming window) and first-differenced segment of the waveform was computed via a standard Fast Fourier Transform routine. One hundred twenty-eight regions, each 1 critical bandwidth wide, were then selected. These regions were chosen to be spaced evenly in critical band units (either budgerigar or human), and the energy in the DFT was summed within each selected region.

This process simulates the summation of energy within a relatively large number (128) of overlapping critical band filters. Critical bandwidths were estimated for the human using a formula derived by Zwicker (1965) and for the budgerigar using data obtained by Dooling and Saunders (1975) and Saunders et al. (1979). Following the summation process, the spectral data were further filtered using a three-term binomial smoothing function to remove roughness due to the rectangular summation window.

Computation of the budgerigar critical band spectra differed in one additional way from that of the human critical band spectra. The human audiogram, compared to that of the budgerigar, is relatively flat up to 5 kHz. The budgerigar shows sharply decreasing sensitivity below 1 kHz and an abrupt decrease in sensitivity for frequencies above 4 kHz. Consequently, in our computing the budgerigar spectra, the initial DFT was shaped by a filter function modeled after the budgerigar audiogram. In effect, this tapered the spectrum by about 14 dB/octave below 1 kHz and about 40 dB/octave above 4 kHz.

(Manuscript received September 19, 1986;
revision accepted for publication December 5, 1986.)