

Discrimination of synthetic full-formant and sinewave /ra-la/ continua by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*)

Robert J. Dooling

Psychology Department, University of Maryland, College Park, Maryland 20742

Catherine T. Best

Wesleyan University, Middletown, Connecticut 06459 and Haskins Laboratories, New Haven, Connecticut 06511

Susan D. Brown

Psychology Department, University of Maryland, College Park, Maryland 20742

(Received 10 July 1993; revised 25 October 1994; accepted 28 October 1994)

Discrimination of three synthetic versions of a /ra-la/ speech continuum was studied in two species of birds. The stimuli used in these experiments were identical to those used in a previous study of speech perception by humans [Best *et al.*, *Percept. Psychophys.* **45**, 237–250 (1989)]. Budgerigars and zebra finches were trained using operant conditioning and tested on three different series of acoustic stimuli: three-formant synthetic speech, sinewave versions of those tokens, and isolated *F*₃ tones from the sinewave speech. Both species showed enhanced discrimination performance near the /l/-r/ boundary in the full-formant speech continuum, whereas for the *F*₃ continuum, neither species showed a peak near this boundary. These results are similar to human discrimination of the same continua. Budgerigars also showed a peak in discrimination of the sinewave analog continuum paralleling that for full-formant syllables, similar to humans who are induced to perceive sinewave speech as speech. Zebra finches, by contrast, showed a relatively flat function mirroring their performance for *F*₃ sinewaves, similar to humans who are induced to perceive sinewave speech as nonspeech. These data provide new evidence of species similarities and differences in the discrimination of speech and speechlike sounds. These data also strengthen and refine previous findings on the sensitivities of the vertebrate auditory system to the acoustic distinctions between speech sound categories.

PACS numbers: 43.66.Gf, 43.80.Lb, 43.71.Es, 43.71.Hw

INTRODUCTION

It is well known that several nonhuman mammals are quite sensitive to the acoustic features of human speech that define some of the phonetic categories that are used contrastively in spoken language (Burdick and Miller, 1975; Dewson, 1964; Dewson *et al.*, 1969; Kuhl and Miller, 1978; Kuhl, 1981; Sinnott, 1989). This similarity in responses to speech sound contrasts between humans and other mammals has been attributed to mammalian similarities in auditory processing mechanisms (Kuhl, 1986; Delgutte, 1982). It is intriguing, therefore, that several species of birds (redwing blackbirds, cowbirds, pigeons, starlings, budgerigars, and zebra finches) have also been found to be sensitive to the acoustic features of speech that define vowel categories (Dooling and Brown, 1990; Hienz *et al.*, 1981; Dooling, 1992a) and two species of birds (budgerigars and Japanese quail) have been shown to be sensitive to acoustic features which define consonant categories (Kluender *et al.*, 1987; Dooling *et al.*, 1989).

Considering the variety of species used in the above studies, the results suggest that the acoustic features defining speech sound categories may have a broader, more universal salience for the vertebrate auditory system than previously thought. The strength of this conclusion, of course, rests both

on the variety of speech sound contrasts tested and on the range and characteristics of species tested. It is important to collect additional evidence about species comparisons, particularly between humans and nonmammals, on perception of consonant contrasts that differ among languages. This is so because human adults perceive consonants more categorically than vowels, and their perception of consonant contrasts is strongly influenced by experience with their native language. To this end, the present experiments extend the study of consonant perception in birds to a new species (zebra finches) and to a new consonant contrast—/r/ vs /l/. These consonants vary widely across languages in whether they occur in phonemic contrasts, and in their phonetic-articulatory details. Published findings are available on human adult perception of the stimuli used in the present study (Best *et al.*, 1989). For comparative purposes, budgerigars were also tested, because prior perceptual findings with speech contrasts are available for that species, yet none of these studies involved /r/-/l/.

There are several reasons why a comparison of budgerigars and zebra finches on this particular speech contrast might prove enlightening. First, the critical acoustic cues that signal a change in phonetic category from /ra/ to /la/ occur exclusively in the region of 2–3 kHz. This is the frequency region of best absolute auditory sensitivity for most birds

including budgerigars and zebra finches (Dooling, 1982, 1992a). Second, budgerigars appear to have an unusual degree of spectral resolving power in the frequency region around 3 kHz. This is not true for zebra finches nor, as far as we know, for other birds (Okanoya and Dooling, 1987). For budgerigars, frequency difference limens for steady-state pure tones in the region of 3 kHz approach the levels typically reported for humans. In American English, the F_3 in /ra/ and /la/ falls at around 3 kHz for a male speaker. The major acoustic difference between /ra/ and /la/ is a steeply rising transition in F_3 frequency at syllable onset for /ra/, but a flat or slightly falling F_3 onset for /la/ (e.g., O'Connor *et al.*, 1957; Best *et al.*, 1989). Thus on strictly psychoacoustic grounds, we might expect budgerigars to be more sensitive than zebra finches in the discrimination of speech sounds involving changes exclusively in an F_3 with a steady-state frequency of approximately 3 kHz.

Third, one might predict species differences in the perception of /ra-/la/ by virtue of differences in the characteristics of species-specific vocalizations of these two species. Many of the calls and song elements of species-typical budgerigar vocalizations are whistled, tonal patterns often containing considerable frequency modulation (Dooling, 1986). This sort of frequency-modulated tonal patterning is reminiscent of sinewave analog speech which is one of the stimulus types used by Best *et al.* (1989). Zebra finch calls and song elements, on the other hand, contain no frequency-modulated tonal elements but rather consist exclusively of broadband, strongly harmonic syllables, some with and some without frequency modulation (Sossinka and Bohner, 1980; Simpson and Vicario, 1990). The harmonic character of zebra finch vocalizations looks somewhat similar to the harmonic structure of a formant in speech. Spectrograms of sample calls from each species are presented in Fig. 1 for illustration and to permit comparison.

It is important to note that a recent comparative perception study also showed that each species processes its own vocalizations more efficiently than those of the other species (Dooling *et al.*, 1992). To the extent that these birds may show enhanced sensitivity to acoustic patterns that have properties similar to those found in their species-specific vocalizations, we might expect zebra finches to discriminate stimuli containing broadbands of harmonic energy, such as the formant structure of /ra-/la/ syllables, better than budgerigars. Analogously, budgerigars should perform better than finches with frequency-modulated tonal stimuli, particularly in the region of their best sensitivity at 3 kHz, as would be represented in sinewave analogs of the F_3 patterns from a /ra-/la/ continuum. They might also show some benefit over finches for perception of sinewave analogs of /ra-/la/ syllables containing frequency-modulated pure-tone replicas of F_1 and F_2 as well as F_3 .

Finally, there are differences between these two species in the acquisition of learned vocal signals, which may also relate in some way to their perception of other auditory stimuli. Both species rely on an external auditory model and social experience to develop a normal vocal repertoire (Eales, 1985; Dooling *et al.*, 1987; Dooling *et al.*, 1990). For both species, the evidence to date also suggests that early

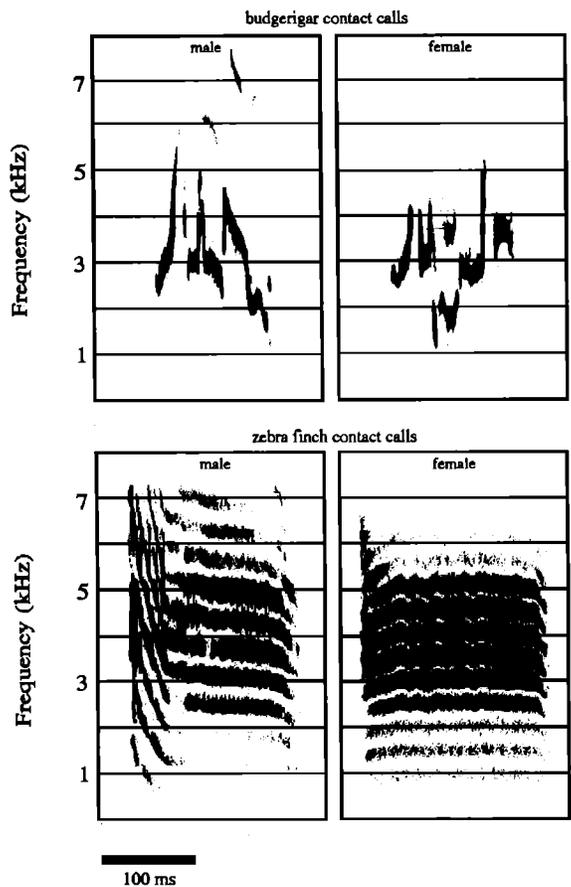


FIG. 1. Spectrograms of selected species-specific contact calls produced by budgerigars and zebra finches.

auditory experience affects the perception of species-specific vocal signals (Cynx and Nottebohm, 1992; Clayton, 1988; Dooling *et al.*, 1990). However, budgerigars learn sounds throughout life from other budgerigars (Brown and Dooling, 1988; Farabaugh *et al.*, 1994) as well as from other species (Gramza, 1970). Moreover, budgerigars can mimic the sounds of human speech; the evidence here, although anecdotal, is widespread.¹ Zebra finches, on the other hand, are far more constrained both in what they will accept as a model and when during development they show vocal learning (Eales, 1985). They show a very strong preference for species-specific song, though under some conditions can be induced to learn the song of a closely related species—the Bengalese finch. There is no evidence that they will learn the songs or calls of more distant species or that they can mimic the sounds of human speech. To the extent that vocal-motor versatility and the demonstrated ability to mimic speech may be related to the perception of speechlike sounds, one might expect that budgerigars would perceive /ra-/la/ syllables, and perhaps sinewave analogs of those syllables as well, more efficiently (i.e., show faster discrimination latencies or a more marked peak in sensitivity at the category boundary) than do zebra finches.

The speech sound contrast between /r/ and /l/ is especially appropriate for a comparative investigation because it is well studied in humans, and its occurrence and phonetic properties differ widely across languages. There is strong

evidence of robust, language-specific, cross-cultural differences in the perception of the American English /r/-/l/ contrast (Miyawaki *et al.*, 1975; MacKain *et al.*, 1981; Best and Strange, 1992). Specifically, monolingual speakers of languages that lack this liquid consonant distinction generally show marked difficulties in discriminating and categorizing /r/ and /l/, although this difficulty may be lessened by extensive conversational experience with English. These observations suggest an important role of language experience in perception of /r/-/l/ by humans. Of greatest interest for the present investigations, a recent study of this speech sound contrast provided compelling evidence consistent with the conclusion that humans perceive the phonetic coherence of the /r/ vs /l/ sounds because they recognize the formant patterns as the acoustic consequences of the articulatory gestures that produce those sounds (Best *et al.*, 1989). In that study, listeners categorized and discriminated a synthetic /la/-/ra/ continuum, a sinewave speech continuum in which the formants of the syllables were replaced with time-varying pure tones that followed the center frequencies of the speech formants, and a continuum of the frequency-modulated F_3 tonal patterns played in isolation. The sinewave speech series was of primary theoretical interest because naive listeners can perceive such stimuli either as distorted speech or as distorted nonspeech sounds such as musical chords. To test whether speechlike categorization and discrimination performance resulted from perception of phonetic coherence in the sinewave syllables, or rather from purely psychoacoustic principles, half of the subjects were induced to hear these stimuli as distorted speech, while the other half were induced to hear them as distorted musical chords with critical variations in the onset glissandi of the highest notes (i.e., the F_3 tones). Consistent with the phonetic coherence hypothesis, only the listeners who had been induced to perceive the sinewave syllable stimuli as speech showed category boundary effects that closely approximated their performance on the full-formant synthetic /la/-/ra/ series. The performance of both groups on the F_3 tone series differed markedly from their performance with the full-formant syllables and coincided with the performance pattern on the sinewave syllables for the listeners who were induced to hear them as distorted chords.

For these reasons, it was of interest to determine whether the acoustic changes associated with phonetic category changes along the three synthetic continua used by Best *et al.* (1989) are salient to other vertebrates. We were especially interested in comparing one species which is, versus another which apparently is not, capable of mimicking human speech. From psychoacoustic considerations, we might expect to find that budgerigars process the isolated F_3 tonal stimuli more efficiently than the other continua. On the other hand, because they are such versatile speech mimics, we might expect budgerigars to show some enhancement in discrimination for full-formant syllables and perhaps also for sinewave syllables. Alternatively, on the basis of the harmonic structure of zebra finch songs and calls, we might expect zebra finches to process full-formant speech more efficiently than the other stimuli.

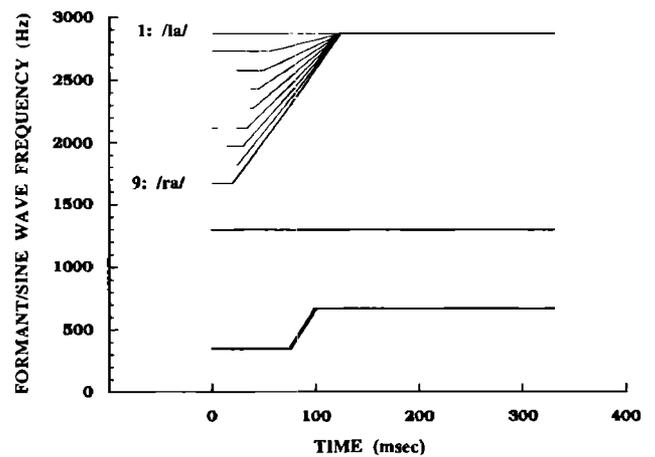


FIG. 2. Schematic diagram of the center frequencies of the three formants in each of the full-formant and sinewave syllable stimuli in the nine-item stimulus continua. Modified from Best *et al.* (1989).

I. METHOD

A. Subjects

The subjects in this experiment were two female and three male adult budgerigars (*Melopsittacus undulatus*) and two female and two male adult zebra finches housed in aviaries at the University of Maryland. All birds were well trained on an auditory detection task and had participated in other psychoacoustic studies involving simple sounds, bird calls, and human speech.

B. Stimuli

The series of stimuli used in this experiment were the following: (1) a full-formant /la/-/ra/ series synthesized with equal-step variations in the F_3 transition only, using the OVE-IIIc serial resonance synthesizer, (2) a matching sinewave syllable continuum generated with a multiple sinewave synthesizer program, and (3) an isolated F_3 tone continuum made up of the F_3 tones from the sinewave analog syllable continuum presented without the tones corresponding to F_1 and F_2 . These were the same stimuli used in the earlier study on perception of phonetic coherence by humans, with the exception that the first stimulus in each set (with a slightly falling F_3 transition) was not used here (Best *et al.*, 1989—see the original article for detailed stimulus description). We omitted the first stimulus in order to reduce the total number of comparisons the birds had to be tested on and because this stimulus was not relevant to the issue of whether birds showed a peak in discrimination near the human boundary in these continua. Because it was the only falling F_3 glide, it had also been eliminated from statistical analyses of boundary slopes in the human study (Best *et al.*, 1989). Schematic representations of the center frequencies of the three formants in sinewave and full-formant stimuli used in the present study (replotted from Best *et al.*, 1989) are shown in Fig. 2; spectrograms of the endpoint stimuli are shown in Fig. 3. F_3 onset frequencies ranged from 2870 for the flat F_3 for the stimulus at the /la/ end of the series to 1670 Hz at the /ra/ end, in steps of approximately 150 Hz (± 9 Hz). The slight variations in step sizes were due to

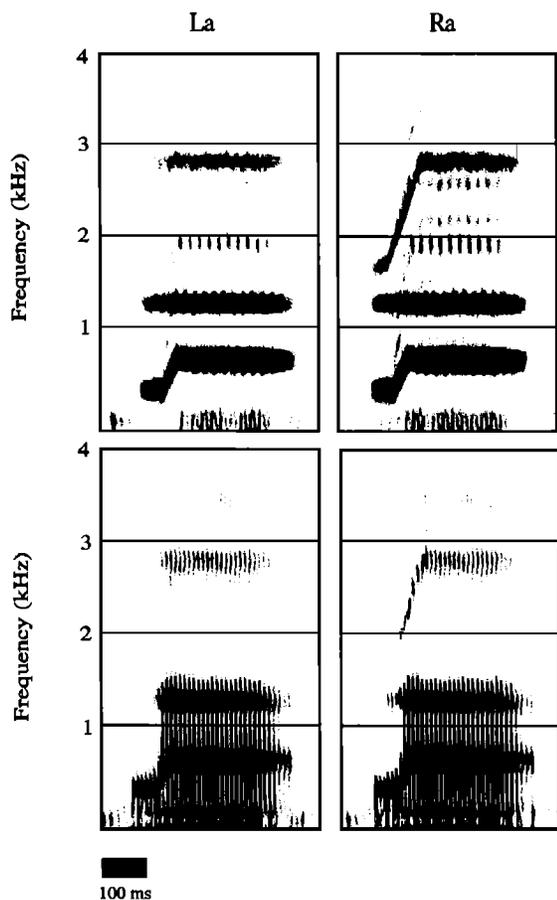


FIG. 3. Spectrograms of the endpoint stimuli from the sinewave analog (top) and the full-formant (bottom) syllable continua used in these experiments.

limitations of the synthesizer. All stimuli were 330 ms in length with 100-ms rise and fall times in the amplitude envelope. The stimuli were presented to the birds at a peak level of 72 dB SPL. Sound-pressure level was measured by placing the microphone of a sound level meter just in front of the response panel in the location normally occupied by the bird's head during testing. The order in which each bird was tested on these three continua was randomly selected.

C. Apparatus

The birds were tested in a wire cage mounted in sound-attenuated chambers. One wall of the wire test cage was modified by the addition of a custom-built response panel constructed of two sensitive microswitches with light-emitting diodes (LEDs) attached (Park *et al.*, 1985). A bird could close the microswitch by pecking at the LED. The left microswitch served as an observation key and the right microswitch served as a report key. An IBM AT microcomputer controlled all experimental events. Stimuli were stored on hard disk with 12-bit resolution, output at a sampling rate of 20 kHz, and low-pass filtered at 10 kHz to prevent aliasing.

D. Training and testing procedures

The procedures for training and testing birds are similar to those described in detail previously (Dooling *et al.*, 1989;

Okanoya and Dooling, 1988). We trained birds to peck one key (observation key) repeatedly during the repetitive presentation of one sound (background) and to peck the other key (report key) when a new sound (target) was presented alternately with the background sound. The occurrence of an alternating pattern of target and background occurred randomly within a period of 1–7 s following a peck on the observation key. A peck on the report key during this alternating stimulus pattern was defined as a correct response and was rewarded with a 2-s access to food. If the bird failed to peck either the observation key or the report key within 2 s, the trial was ended and the response latency was recorded at 2 s. About 15% of the trials were sham trials in which the target stimulus was the same as the background stimulus. A response on the report key during a sham trial was punished with a 16-s timeout period. During this period, lights in the test chamber were extinguished but the repeating background sound continued. Each stimulus continuum used in these experiments consisted of nine synthetic tokens. For each continuum, each stimulus in the set served as both a background and a target until all possible combinations of background and target were tested. The order in which stimuli were selected as background or targets was randomized and a different random order was used for each bird. Once one stimulus was selected as a background, all other stimuli in the set served as targets on subsequent trials. Then, a different background stimulus was selected and the procedure repeated. At the end of testing, this resulted in a matrix of response latencies by which each stimulus in the set could be related to every other stimulus in the set. Testing the birds in this way ensured that they were not trained to any particular stimulus contrast—each bird was exposed to each stimulus contrast exactly the same number of times.

Generally, the birds were tested in two daily sessions and about 10–15 of these sessions were required for each bird to have produced at least nine complete matrices of response latencies (i.e., each bird provided at least 18 responses to each stimulus pair). In practice, each session lasted 30 min or until the bird stopped responding, whichever came first. If the bird failed to complete an entire matrix of stimulus combinations in a given session, the remaining trials were completed in the next session. An additional criterion used for accepting a data matrix for subsequent analysis was that the overall percent correct detection for an entire matrix was above 75% with a false alarm rate below 20%. Testing each bird on each continuum generally required between 1 and 2 weeks of daily testing.²

E. Data analysis procedures

The data from these experiments were analyzed as follows. For each bird the average response latencies for pairs of stimuli in a one-step (i.e., adjacent stimuli) discrimination comparison (i.e., 1 vs 2, 2 vs 3, and so on) were computed. These data were then analyzed by a three-way analysis of variance (ANOVA) to compare the relative discriminability of stimulus pairs along each continuum.

II. RESULTS

Results from human experiments using these and other synthetic speech stimuli are typically presented in the form of labeling and discrimination functions (see, for example, Best *et al.*, 1989; Miyawaki *et al.*, 1975). The latency data from our budgerigars and zebra finches are discrimination data as opposed to labeling data. With the paradigm used for our birds, shorter response latencies reflect greater discriminability—a relation that has been empirically validated using simple pure tones of known frequency difference (Dooling *et al.*, 1987). Also, whereas human discrimination results are often based on judgment accuracy on two-step or three-step comparisons (i.e., respectively, 1 vs 3, 2 vs 4, and so on, or 1 vs 4, 2 vs 5, and so on), here we report reaction time data from our birds on one-step comparisons. The main purpose of these experiments is to compare the performance of two avian species rather than to compare birds with humans and one-step comparisons provide the best species comparison. It is important to note, however, that while we report one-step latency data, both percent correct data for the one-step continua as well as latency data and percent correct data for the two-step comparisons closely parallel the data for the one-step continua.

The average relative response latencies for budgerigars and zebra finches for adjacent stimuli (one-step continuum) for the three sets of stimuli are shown in Fig. 4. Both budgerigars and zebra finches showed a response latency minimum (i.e., peak in sensitivity to stimulus differences, or discrimination) in the synthetic speech continuum, while only budgerigars showed a response latency minimum in the sinewave syllable continuum. For the F_3 continuum, both species found the difference between the first and second token (flat versus nearly flat F_3) in the continuum essentially non-discriminable (i.e., three budgerigars and two finches consistently timed out at 2000 ms without responding; the other birds' mean latencies fell between 1860 and 1940). Beyond this there was a tendency for zebra finches to find the differences among the remaining adjacent tokens in the F_3 tone continuum equally discriminable—roughly the same effect as that observed when these birds were tested with sinewave syllables. Budgerigars, on the other hand, showed marked response latency minima (i.e., peaks of greater discriminability) between tokens 2 and 3 (and, to a lesser extent, pair 7–8) in this continuum. The enhanced discriminability in the region of tokens 5–7 for full-formant and sinewave speech stimuli was not observed for F_3 presented alone. Conversely, the budgerigars' minima at 2–3 and 7–8 on the F_3 tone series were not evident in their full-formant or sinewave syllable performance.

These data were analyzed by a three-way analysis of variance (continuum \times pairs \times species). The overall ANOVA had several significant main effects and several significant interactions. The continuum effect was only marginal [$F(2,14)=2.901$, $p<0.08$] but showed that the birds, overall, were slowest in responding to the speech continuum ($M=1497$ ms) and fastest on the F_3 continuum ($M=1345$ ms). The significant continuum \times species interaction [$F(2,14)=4.14$, $p<0.03$] showed that budgerigars were fastest on the F_3 continuum ($M=1229$ ms), were much

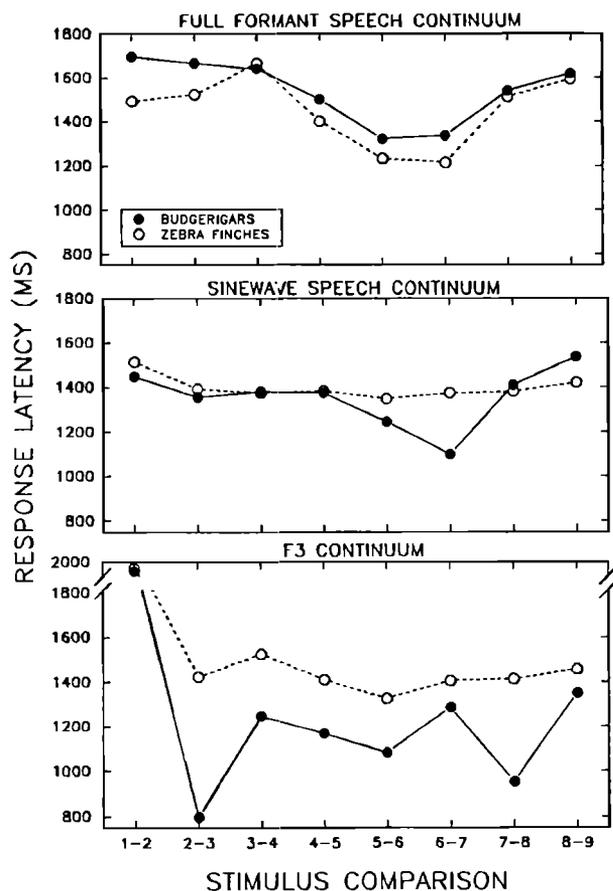


FIG. 4. Relative response latencies for budgerigars and zebra finches are shown for adjacent tokens along the full-formant speech continuum (top), sinewave syllable continuum (middle), and F_3 alone continuum (bottom). Response latency minima in continua reflect peaks in discrimination sensitivity.

slower on the full-formant speech continuum ($M=1532$ ms), and were intermediate on the sinewave syllable continuum ($M=1356$ ms). Zebra finches performed at more similar speeds on all three continua ($M=1398$, 1454, and 1491 ms). According to simple effects tests of that interaction, budgerigars were marginally faster than finches in their mean latencies for the F_3 tone continuum [$F(1,12)=3.92$, $p=0.07$]; no other simple effects were significant for that interaction.

The pairs main effect shows that when the three continua are averaged across species, there was a highly significant effect [$F(7,49)=19.378$, $p<0.0001$] of a response latency minimum (i.e., sensitivity peak) which coincides with pair 5 and 6. The continuum \times pairs interaction [$F(14,98)=8.397$, $p<0.0001$] revealed significant differences in the shape of the latency functions across continua. Tukey tests of pairwise differences among stimulus contrasts showed that on the F_3 tone continuum, comparison 1–2 had a significantly longer latency than all other comparison pairs ($p<0.01$), and that latency on comparison pair 2–3 was faster than on pairs 3–4 and 4–5 ($p<0.05$) and on pair 8–9 ($p<0.01$). F_3 tone performance was relatively flat but variable along the continuum between comparison pairs 3–4 and 8–9. Tukey tests indicated that the speech continuum

showed a clear peak in sensitivity at pairs 5–6 and 6–7 relative to other pairs ($p < 0.01$).

The sinewave speech continuum also showed an apparent peak in sensitivity for pair 6–7. However, the continuum \times pairs \times species interaction [$F(2, 14) = 2.164$, $p < 0.02$] further revealed a species difference in the shape of the discrimination function, but only on the sinewave speech continuum. Simple effects tests on that interaction indicate that the two species performed similarly on the $F3$ tones, although with somewhat poorer finch performance overall (noted above), which was significantly poorer than budgerigars on pairs 2–3 [$F(1, 16) = 15.175$, $p < 0.001$] and 7–8 [$F(1, 16) = 8.25$, $p < 0.01$]. As noted above, both species showed a single peak on speech pairs about midway through the continuum. Most notably, however, on the sinewave syllables, budgerigars showed a significant sensitivity peak at around the same place as the full-formant speech peak, i.e., pair 6–7 [$F(7, 49) = 5.55$, $p < 0.0001$], but finches showed a relatively flat discrimination function without significant latency differences among any stimulus pairings.

III. DISCUSSION

This study sought to provide comparative data from birds on the perception of synthetic speech sounds. Budgerigars and zebra finches were tested on three differing synthetic versions of the critical acoustic category differences along a /ra-la/ speech continuum. The results were analyzed by examining the relative discriminability of adjacent stimuli along each continuum, with comparisons across species and across continua. These results provide interesting comparative data that are relevant to a number of issues in speech perception research.

A. Species differences in speech perception

In general, budgerigars responded more efficiently to $F3$ tones than to sinewave speech or to full-formant speech. Relative to $F3$ tones, budgerigars performed worse (longer response latencies) on the endpoint stimuli for both sinewave speech and full-formant speech. Said another way, budgerigars showed a peak in discrimination performance near the center of both the sinewave speech and the full-formant speech continua at about the same point as do human listeners (see Fig. 6 in Best *et al.*, 1989—but recall that stimulus item 1 from each continuum in that report was dropped for the present study). In general, budgerigars were also slightly better at discriminating sinewave syllables than full-formant speech (i.e., response latencies for sinewave syllables were shorter than for full-formant speech syllables).

Zebra finches, on the other hand, showed a different pattern of results in that they performed with about the same degree of efficiency (i.e., around 1400 ms) on all three continua. The peak in discrimination performance observed for zebra finches discriminating among full-formant speech syllables compared to sinewave syllables or $F3$ tones comes about because zebra finches actually show enhanced performance near the center of the full-formant speech continuum relative to the endpoints of the continuum. This enhanced performance occurs near the same place in the continuum as it does for humans. Perhaps even more intriguing, zebra

finches did not show such a peak in discrimination when tested with sinewave syllables, thus mirroring the pattern of results seen with human listeners who had been induced to hear the sinewave syllables as distorted musical chords rather than as speech (Best *et al.*, 1989).

How can we account for these species differences? Several findings from this study are consistent with what we know of the psychoacoustics of hearing in these two species, or with the characteristics of their species-specific vocal signals. That budgerigars discriminate among $F3$ tones, and to some extent sinewave syllables, much better than do zebra finches might be expected from the budgerigar's enhanced sensitivity to spectral changes in the region of 2–4 kHz. Zebra finches, on the other hand, generally outperform budgerigars on full-formant speech. Zebra finch vocalizations are distinguished by harmonic bands of energy bearing some resemblance to a formant of human speech. Given the “formantlike” structure of zebra finch vocalizations, one might expect zebra finches to generally show facility in discriminating among speech sounds differing in formant patterns.

It is difficult to account for the peak in discrimination that both species show for full-formant speech but that only budgerigars show for sinewave syllables. The fact that both species of birds, like English-speaking humans, show a peak in discrimination in full-formant speech at approximately the human /r-l/ boundary could be taken as evidence that the particular constellation of acoustic cues occurring near the center of this continuum have robust perceptual consequences for the vertebrate auditory system.

That only budgerigars show a similar peak in discrimination in sinewave speech may have a psychoacoustic explanation related to their well-documented enhanced spectral resolving ability in the region around 2–4 kHz. Critical bandwidths in budgerigars are considerably smaller than those in zebra finches in the frequency region of the $F3$ formant transition (although budgerigars did have extraordinary difficulty in discriminating the 1–2 contrast in the $F3$ continuum relative to other contrasts, suggesting a differential sensitivity to frequency glide rate).

Another more intriguing possibility is that the coexistence of excellent auditory discrimination, an extraordinarily versatile vocal-motor system, and a highly social lifestyle lead to an unusual ability to attend to, remember, and mimic acoustic patterns produced in social circumstances by other conspecifics and by other vertebrates including humans. Perhaps exposure to talking humans sensitizes these birds to some of the critical acoustic features of human speech. Experiments with isolate-reared budgerigars or budgerigars raised in a monolingual or Japanese- or Korean-speaking environment may answer this question, as well as tests on other bird species that are broad mimics, including speech mimicry (e.g., starlings) versus those that are more constrained in what they will accept as a vocal learning model.

B. Categorical perception and natural categories

The present discrimination data are not alone sufficient to establish whether birds perceive the /ra-la/ speech continuum in a categorical fashion. A strict demonstration would require evidence that the birds also categorized tokens from

this continuum appropriately. But the present data are certainly consistent with the notion that birds would perceive these sounds categorically. It is perhaps worth noting that in studies of VOT perception by animals, labeling functions follow closely what would be expected from the discriminability of adjacent tokens along the continuum (see, for example, Kuhl and Miller, 1978; Kuhl, 1981).

Much previous work on perception of speech sounds by animals is susceptible to the criticism that animals were "trained" or "overtrained" to respond to speech sound categories with subsequent tests for category boundaries perhaps influenced by this training. In other words, what may emerge from such tests is what an animal "can be trained" to do rather than what it "does" do naturally. The procedures used in the present experiments were developed specifically to avoid this problem. The birds in these experiments were exposed to every possible stimulus comparison, in random order, exactly the same number of times. This ensures that the bird is not trained to any particular discrimination. Moreover, the region of maximum discriminability along the full-formant speech continua is evident upon first exposure. We conclude that both zebra finches and budgerigars have a "natural" tendency to find the acoustic differences between certain tokens in this continuum more discriminable than others. Particularly relevant to this "training" concern is the fact of differences in performance among continua and between species.

We also note that both species show the lack of a clear peak on the *F3* tone stimuli. This species congruence is compatible either with an argument that the full-formant /r-l/ "categories" reflect the birds' detection of acoustic categories or, though less likely, the detection of phonetic categories. However, the difference in performance of the two avian species on the third stimulus continuum—the sinewave syllables—suggests at least that budgerigars may hear something more in these stimuli. Budgerigars showed a discrimination peak in the sinewave speech continuum like human listeners who had been induced to hear these same sinewave patterns as speech (Best *et al.*, 1989). Clearly budgerigars have a proclivity to detect categories in those stimuli while zebra finches utterly failed to do so and humans only do so when induced to perceive those stimuli as speech. To return to the speculative point raised earlier, budgerigars may be particularly sensitive to acoustic patterns that fall within range of their broad mimicry capability. Given the frequency-modulated tonal properties of their species-specific vocalizations, sinewave analogs of speech may be particularly salient to budgerigars because they may detect in these sinewave speech stimuli certain tonal modulations that they are able to control with their own vocal tract movements. This possibility could be further assessed by comparing the extent to which budgerigars prefer sinewave speech over full-formant speech as a model for vocal learning.

C. Cultural differences, speech sounds, and the vertebrate auditory system

It is well known that there are language-community differences in the way these particular speech sounds are used and perceived by humans. Previous research has examined

the effect of linguistic experience on the perception of this continuum, and suggested that the effect of linguistic experience is unique to a special "speech mode" of perception (Miyawaki *et al.*, 1975). Other studies have shown that Japanese listeners, with sufficient training, can come to hear the /r-l/ contrast nearly as categorically as do English speakers (Best and Strange, 1992; MacKain *et al.*, 1981). In both cases, it is phonologically appropriate experience that leads to English-like perception of this continuum. The present data from two species of birds and recent data from Sinnott and Barnett (1992) on primates are the first to report the perception of this consonant contrast by nonlinguistic animals. Clearly the present results cannot be accounted for by appealing to phonological mechanisms or experience with speech sounds in their natural communicative context.

Rather, the peak in discrimination of full-formant speech continuum evidenced by our birds must be due to nonlinguistic auditory processes. These data provide new evidence of the sensitivities of the vertebrate auditory system to the acoustic distinctions associated with speech sound categories. This is not to say that human discrimination and perception of these same sounds cannot also involve other specifically human, phonetic mechanisms—the weight of evidence favors the view that at least some aspects of speech perception are unique to phonetic processing and/or humans (e.g., Best *et al.*, 1989; Best *et al.*, 1981; Kuhl, 1991; Tomiak *et al.*, 1987; Whalen and Liberman, 1987; cf. Diehl and Kluender, 1989; Kluender *et al.*, 1987). But, together with other studies of speech sound discrimination of consonants and vowels by mammals and birds, these results suggest that many of the properties of speech are more salient to a wider range of vertebrate auditory systems than previously thought. At the same time, this first report of a nonhuman species difference in perception of a highly modified set of speech-like stimuli (the sinewave syllables) raises a number of intriguing questions that warrant further investigation.

ACKNOWLEDGMENTS

This work was supported by Grants No. NIH DC-00198 and No. MH-00982 to R. Dooling, No. DC-00403 to C. Best, and No. HD-01994 to the Haskins Laboratories. We thank G. Klump for comments on an earlier draft and D. Reidell for care of the birds.

¹Our impression, in fact, from listening to mimicked speech produced by budgerigars, is that it sounds similar to sinewave speech.

²At the conclusion of testing, one zebra finch was found to have much higher false alarm rates for the speech continuum than for the other two continua. This bird was retested on all contrasts of the speech continuum presented in random order. The results of this second run showed false alarm rates more consistent with results from the other two continua and these data were used in all subsequent analyses.

Best, C. T., Morrongiello, B., and Robson, R. (1981). "Perceptual equivalence of acoustic cues in speech and nonspeech perception." *Percept. Psychophys.* **29**, 345–360.

Best, C. T., and Strange, W. (1992). "Effects of language-specific phonological and phonetic factors on cross-language perception of approximants." *J. Phon.* **20**, 305–330.

Best, C. T., Studdert-Kennedy, M., Manuel, S., and Rubin-Spitz, J. (1989). "Discovering phonetic coherence in acoustic patterns." *Percept. Psychophys.* **45**, 237–250.

- Brown, S. D., Dooling, R. J., and O'Grady, K. (1988). "Perceptual organization of acoustic stimuli by Budgerigars (*Melopsittacus undulatus*): III. Contact calls," *J. Comp. Psychol.* **102**, 236–247.
- 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.
- Clayton, N. S. (1988). "Song discrimination learning in zebra finches," *Anim. Behav.* **36**, 1016–1024.
- Cynx, J., and Nottebohm, F. (1992). "Role of gender, season, and familiarity in discrimination of conspecific song by zebra finches (*Taeniopygia guttata*)," *Proc. Natl. Acad. Sci. (USA)* **89**, 1368–1371.
- Delgutte, B. (1982). "Some correlates of phonetic distinctions at the level of the auditory nerve," in *The Representation of Speech in the Peripheral Auditory System*, edited by R. Carlson and B. Granstrom (Elsevier, Amsterdam).
- Dewson, J. H. (1964). "Speech sound discrimination by cats," *Science* **144**, 555–556.
- Dewson, J. H., Pribram, K. H., and Lynch, J. C. (1969). "Effects of ablations and temporal cortex upon speech sound discrimination in the monkey," *Exp. Neurol.* **24**, 579–591.
- Diehl, R., and Kluender, K. (1989). "On the objects of speech perception," *Ecol. Psychol.* **1**, 1–45.
- Dooling, R. J. (1982). "Auditory perception in birds," in *Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller Academic Press (New York), Vol. 1.
- Dooling, R. J. (1986). "Perception of vocal signals by the budgerigar (*Melopsittacus undulatus*)," *Exp. Biol.* **45**, 195–218.
- Dooling, R. J. (1992a). "Hearing in birds," in *The Evolutionary Biology of Hearing*, edited by D. B. Webster, R. R. Fay, and A. N. Popper (Springer-Verlag, New York).
- Dooling, R. J. (1992b). "Perception of speech sounds by birds," in *Advances in Biosciences: Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, London).
- Dooling, R. J., and Brown, S. D. (1990). "Speech perception by budgerigars (*Melopsittacus undulatus*): Spoken vowels," *Percept. Psychophys.* **47**, 568–574.
- Dooling, R. J., Brown, S. D., Klump, G., and Okanoya, K. (1992). "Auditory perception of conspecific and heterospecific vocalizations in birds: Evidence for special processes," *J. Comp. Psychol.* **106**, 20–28.
- Dooling, R. J., Park, T. J., Brown, S. D., Okanoya, K., and Soli, S. D. (1987). "Perceptual organization of acoustic stimuli by budgerigars (*Melopsittacus undulatus*): II. Vocal signals," *J. Comp. Psychol.* **101**, 367–381.
- Dooling, R. J., Okanoya, K., and Brown, S. D. (1989). "Perception of speech sounds by the budgerigar (*Melopsittacus undulatus*): The voiced–voiceless distinction," *Percept. Psychophys.* **46**, 65–71.
- Dooling, R. J., Park, T. J., Brown, S. D., and Okanoya, K. (1990). "Perception of species-specific vocalizations by isolate-reared budgerigars (*Melopsittacus undulatus*)," *Int. J. Comp. Psychol.* **4**, 57–78.
- Eales, L. A. (1985). "Song learning in zebra finches: Some effects of song model availability on what is learnt and when," *Anim. Behav.* **33**, 1293–1300.
- Farabaugh, S. M., Linzenbold, A., and Dooling, R. J. (1994). "Vocal plasticity in budgerigars (*Melopsittacus undulatus*)—Evidence for social factors in the learning of contact calls," *J. Comp. Psychol.* **108**, 1–11.
- Gramza, A. F. (1970). "Vocal mimicry in captive budgerigars (*Melopsittacus undulatus*)," *Z. Tierpsychol.* **27**, 971–983.
- Hienz, R. D., Sachs, M. B., and Sinnott, J. M. (1981). "Discrimination of steady-state vowels by blackbirds and pigeons," *J. Acoust. Soc. Am.* **70**, 699–706.
- Kluender, K. R., Diehl, R. L., and Killeen, P. R. (1987). "Japanese quail can learn phonetic categories," *Science* **237**, 1195–1197.
- 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. (1981). "Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories," *J. Acoust. Soc. Am.* **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," *Exp. Biol. Sens. Percept. Process.* **45**, 233–265.
- 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.
- MacKain, K. S., Best, C. T., and Strange, W. (1981). "Categorical perception of English /t/ and /l/ by Japanese bilinguals," *Appl. Psycholinguist.* **2**, 369–390.
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., and Fujimura, O. (1975). "An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English," *Percept. Psychophys.* **18**, 331–340.
- O'Connor, J. D., Gerstman, L. J., Liberman, A. M., Delattre, P. C., and Cooper, F. S. (1957). Acoustic cues for the perception of initial /w,j,r,l/ in English," *Word* **15**, 24–43.
- Okanoya, K., and Dooling, R. J. (1987). "Hearing in passerine and psittacine birds: A comparative study of absolute and masked auditory thresholds," *J. Comp. Psychol.* **101**, 7–15.
- Okanoya, K., and Dooling, R. J. (1988). "Obtaining acoustic similarity measures from animals: A method for species comparisons," *J. Acoust. Soc. Am.* **83**, 1690–1693.
- Park, T., Okanoya, K., and Dooling, R. (1985). "Operant conditioning of small birds for acoustic discrimination," *J. Ethol. (Japan)* **3**, 5–9.
- Simpson, H. B., and Vicario, D. S. (1990). "Brain pathways for learned and unlearned vocalizations differ in zebra finches," *J. Neurosci.* **10**, 1541–1556.
- Sinnott, J. M. (1989). "Detection and discrimination of synthetic English vowels by Old World monkeys (*Cercopithecus*, *Macaca*) and humans," *J. Acoust. Soc. Am.* **86**, 557–565.
- Sinnott, J. M., and Barnett, D. M. (1992). "Comparison of humans and monkey sensitivity to the English liquid /ra-la/ contrast," *J. Acoust. Soc. Am.* **92**, 2340 (A).
- Sossinka, R., and Bohner, J. (1980). "Song types in the zebra finch (*Poephila guttata castanotis*)," *Z. Tierpsychol.* **53**, 123–132.
- Tomiak, G. R., Mullenix, J. W., and Sawusch, J. R. (1987). "Integral processing of phonemes: Evidence for a phonetic mode of perception," *J. Acoust. Soc. Am.* **81**, 755–764.
- Whalen, D. H., and Liberman, A. M. (1987). "Speech perception takes precedence over nonspeech perception," *Science* **237**, 169–171.