

Rhesus macaques spontaneously perceive formants in conspecific vocalizations

W. Tecumseh Fitch^{a)}

School of Psychology, University of St. Andrews, St. Andrews, Fife, Scotland, KY169JP, United Kingdom

Jonathan B. Fritz^{b)}

Laboratory of Neuropsychology, NIMH, NIH, Bethesda, Maryland 20742

(Received 7 March 2006; revised 2 July 2006; accepted 5 July 2006)

We provide a direct demonstration that nonhuman primates spontaneously perceive changes in formant frequencies in their own species-typical vocalizations, without training or reinforcement. Formants are vocal tract resonances leading to distinctive spectral prominences in the vocal signal, and provide the acoustic determinant of many key phonetic distinctions in human languages. We developed algorithms for manipulating formants in rhesus macaque calls. Using the resulting computer-manipulated calls in a habituation/dishabituation paradigm, with blind video scoring, we show that rhesus macaques spontaneously respond to a change in formant frequencies within the normal macaque vocal range. Lack of dishabituation to a “synthetic replica” signal demonstrates that dishabituation was not due to an artificial quality of synthetic calls, but to the formant shift itself. These results indicate that formant perception, a significant component of human voice and speech perception, is a perceptual ability shared with other primates. © 2006 Acoustical Society of America. [DOI: 10.1121/1.2258499]

PACS number(s): 43.66.Gf, 43.80.Lb, 43.80.Ka, 43.71.An, 43.70.Bk [JAS] Pages: 2132–2141

I. INTRODUCTION

Formants (the spectral prominences generated by vocal tract resonances) provide a critical acoustic cue to many phonemic differences in human speech (Fant, 1960; Lieberman and Blumstein, 1988; Titze, 1994; Ladefoged, 2001). In addition to their central role in speech perception, recent studies indicate that human formants also act as a perceptual cue to body size (Fitch, 1994; Ives *et al.*, 2005; Smith *et al.*, 2005) and play a role in attractiveness judgments (Collins, 2000; Feinberg *et al.*, 2005). After many years of assuming that formants are a peculiarity of the human speech signal, it is becoming clear to bioacousticians that formants are present in animal calls (Fitch, 1997; Owren *et al.*, 1997; Rendall *et al.*, 1998; Riede and Fitch, 1999; Riede *et al.*, 2005). Thus there is an increasing interest in the roles that formants might play in animal communication, and in the degree to which formant perception in nonhuman animals represents a homologue to the mechanisms involved in human speech perception. Although it is clear that many animals (including dogs, horses, baboons, macaques, and various bird species) can be trained to discriminate between different speech signals, the critical questions for understanding animals' own communication systems are whether they perceive formants in their own, species-specific vocalizations. This requires, first, determining if the spectral prominences examined are, in fact, formants; second, generating an appropriate stimulus set where formants are shifted; and

finally, determining if animal subjects perceive formant changes (preferably spontaneously, without any training) (Fitch and Kelley, 2000; Reby *et al.*, 2005). The specific types of information conveyed by formants (e.g., call type, size, identity, attractiveness) can then be further investigated.

Several approaches have been used to determine whether some particular spectral prominences are formants (derive from vocal tract filtering), rather than representing “pseudo-formants” generated by some other process, or even harmonics of the voice source. Nonlinear phenomena involving the voice source are the clearest alternative mechanism capable of generating pseudo-formants [e.g., in scream vocalizations (Fitch *et al.*, 2002)]. The most direct way to exclude the possibility of source-generated spectral prominences is to place the animal in a light-gas (e.g., heliox) environment and induce it to vocalize (Roberts, 1975; Nowicki, 1987; Amundin, 1991; Rand and Dudley, 1993). If a spectral prominence is caused by formant filtering, the increased speed of sound leads to an increase in vocal tract resonance frequency, and thus to the center frequency of the spectral prominence. Such experiments were critical in revealing that spectral prominences in vocalizations of several frog species are *not* influenced by filtering in the frog's vocal tract and therefore do not represent formants (Rand and Dudley, 1993). These pseudo-formants apparently result instead from frequency-modulation within the laryngeal source (e.g., Martin, 1971). In contrast, heliox testing in many bird species has revealed the predicted shift of frequencies, allowing researchers to conclude that formant filtering plays an important role in avian vocal production (Nowicki, 1987; Suthers and Hector, 1988; Nowicki *et al.*, 1989; Fletcher and Tarnopolsky, 1999). When feasible, heliox testing thus provides the

^{a)}Author to whom correspondence should be addressed; electronic mail: wtsf@st-andrews.ac.uk

^{b)}Currently at: Center for Acoustic and Auditory Research, Institute for Systems Research, ECE, University of Maryland, College Park, MD 20742.

“gold standard” for demonstrating formant filtering in animal vocalizations.

Unfortunately, for many large or free-living animal species, immersion in a heliox atmosphere is impractical or impossible, and many other animals will refuse to vocalize in confined conditions. In such cases other techniques are necessary to demonstrate formant filtering. One alternative approach relies on the acoustic link between vocal tract length and formant frequencies (Fant, 1960; Titze, 1994). If vocalizations from multiple individuals of different body sizes are available, and the different vocal tract lengths for each individual can be measured, a strong correlation between vocal tract length and the frequencies of spectral prominences provides strong evidence for formant filtering (Fitch, 1997; Riede and Fitch, 1999; Reby and McComb, 2003). Although direct anatomical or x-ray measurements of vocal tract length provide the strongest evidence, a correlation of spectral prominences with head length or overall body size provides weaker evidence for vocal tract filtering (Fitch, 2000a). If an animal makes prominent, visible changes in its vocal tract length while vocalizing, and spectral prominences move in synchrony, this is also evidence for formants (particularly if other acoustic variables such as fundamental frequency or higher harmonics do not change in synchrony) [e.g., Hauser *et al.*, 1993; Fitch and Reby, 2001; Harris *et al.* (2006)]. Even simple inspection of spectrograms, combined with basic acoustic considerations and anatomical measurements from museum skulls or preserved specimens, can provide an indication of whether a set of spectral prominences could represent formants. For example, two recent papers have claimed that spectral prominences in mouse vocalizations represent formants (Ehret and Riecke, 2002; Geissler and Ehret, 2002), but the relatively low frequencies of these spectral prominences would entail a vocal tract longer than a mouse’s entire body if they were formants. These authors appear to have confused harmonics of the glottal source with formant frequencies.¹ In summary, there are several possible sources of confirmation that spectral peaks in a given species’ vocalizations represent formants. Converging, consistent data from several sources will of course provide the most convincing demonstration.

Once it has been established that formants are present in a particular type of vocalization, perceptual experiments are necessary to test whether the species in question attends to these cues. While it is possible to perform such experiments with natural call exemplars with varying formant frequencies, this leaves open the possibility that changes in other, unmeasured variables were noticed instead. Thus, synthetic signals in which only formants are changed are preferable. To the extent that the source/filter theory of vocal production applies to many animal vocalizations (Fitch and Hauser, 1995; Fitch and Hauser, 2002), various well-understood signal processing techniques, developed by speech scientists, are available to modify formants without changing other aspects of the signal. For example, LPC-based analysis and resynthesis can be used to artificially separate the signal into source and filter components, if certain preconditions are fulfilled (Fitch, 1997; Owren and Bernacki, 1988). Then, the filter component can be modified in specific ways, and

source and filter can be recombined to create a natural-sounding vocalization where only the spectral prominences have been shifted (Moorer, 1979; Moore, 1990; Fitch, 2002; Smith *et al.*, 2005). These and similar techniques can easily be implemented on desktop systems using software such as MATLAB or PRAAT. Such techniques have been successfully applied to whooping cranes (Fitch and Kelley, 2000) and red deer (Reby *et al.*, 2005), and are used in the present study with macaques. It is important to recognize, however, that not all vocalization types are appropriate for such techniques. If present, source-determined spectral peaks (in particular the fundamental frequency and low harmonics) must be lower than the lowest formant (as is the case in adult speech). Even in such ideal cases, however, it is difficult to distinguish formant frequency perception from the perception of differences in harmonic amplitudes [as seen in some birds Cynx *et al.*, 1990]. With very high fundamentals it also remains possible that source/filter interactions could occur, violating the independence assumption of the source/filter theory and of LPC (Fitch and Hauser, 1995). The ideal calls for resynthesis techniques and are therefore those in which no source-related peaks exist, e.g., calls with source components that consist of broadband noisy excitation or impulse trains. Given appropriate precautions, however, linear prediction or similar digital processing techniques allow researchers to generate a set of stimuli in which formants are manipulated, but all other acoustic variables are held constant.

Given an appropriate set of synthesized animal vocalizations, we can finally proceed to experimentally determine whether animals of a particular species perceive formant changes in their own species’ calls. Presumably, given adequate prolonged training, any animal with adequate spectral sensitivity should be able to learn to distinguish between sounds with shifted formants (for example, many different vertebrate species have been trained to distinguish between synthetic human vowels differing only in formant frequencies). But if animals naturally make use of formants in their communication system, they should react to changes in formant frequency spontaneously, without requiring specific training or reinforcement. Thus, tests of spontaneous perception, such as habituation/discrimination techniques, provide the strongest evidence for a species’ use of formants as a meaningful communicative parameter. Such techniques were introduced for perceptual experiments with human infants (Eimas *et al.*, 1971), they have been successfully used with many different animal species, including nonhuman primates (e.g., Seyfarth and Cheney, 1990; Rendall, 1996; Fischer, 1998; Hauser, 1998). However, results from such experiments employing resynthesized calls, where only formants change, are currently available for only two nonhuman species: whooping cranes (Fitch and Kelley, 2000) and red deer (Reby *et al.*, 2005).

In this study, we test the hypothesis that rhesus macaques (*Macaca mulatta*) spontaneously perceive formant frequencies in conspecific vocalizations. We also use a control condition to test the adequacy of our monkey call synthesis techniques, specifically to ensure that these techniques introduce no perceptually salient artificial quality to our synthetic calls. We chose the rhesus macaque, a common labo-

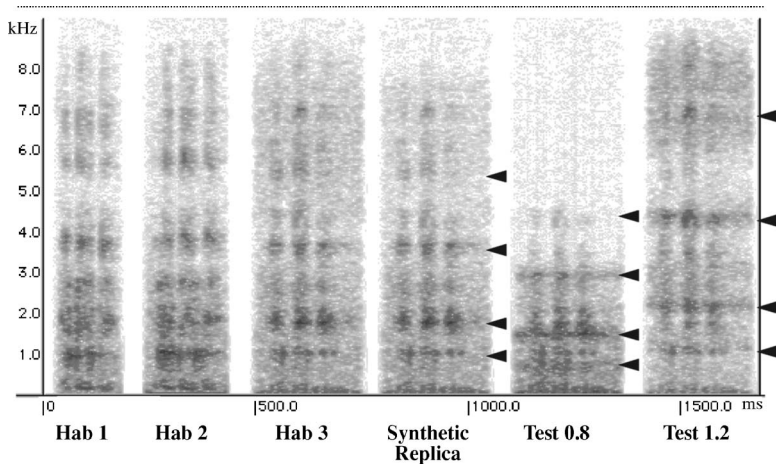


FIG. 1. Spectrogram illustrating the calls used in this experiment. The first three calls (“Hab”) form the habituation set and are natural, recorded calls. The last three are synthetic calls. “Synthetic replica” has the formants unchanged from Hab 3, while in the two test calls, formants have been shifted down (0.8) or up (1.2) in frequency by 20%. Black arrows indicate the lowest four formant frequencies

ratory primate, as our test species because prior studies have demonstrated that formants are present in some calls of this species, and could potentially carry information about body size, call type, and/or individual identity (Hauser *et al.*, 1993; Fitch, 1997; Rendall *et al.*, 1998). Furthermore, with intensive training in the laboratory, macaques of a closely related species (*Macaca fuscata*) are able to recognize changes in formant frequencies in synthesized vowels with an accuracy rivaling humans’ (Sommers *et al.*, 1992). However, no previous study has combined spontaneous perceptual testing with a synthetic stimulus set to conclusively demonstrate formant perception in this (or any other nonhuman primate) species.

Experimental Approach. We used a custom-designed monkey call synthesizer, based on the source/filter theory of vocal production (Fant, 1960; Titze, 1994; Fitch and Hauser, 1995; Fitch, 2002), to create natural-sounding macaque calls. Our central hypothesis was that rhesus monkeys spontaneously perceive changes in formant frequencies in conspecific vocalizations, and find these differences significant enough to warrant dishabituation. This was tested by habituating an animal to a set of natural stimuli from a single individual, and then playing a test call in which the formants had been digitally shifted while holding other acoustic variables constant. If a subject dishabituated to the modified “test” call, we concluded that it had perceived the test call as different from the preceding stimuli, and therefore perceived the formant changes. A nontonal noisy vocalization type, the aggressive “pant threat,” was used to avoid the issues with harmonics cited earlier (Fig. 1).

Nonetheless, the cause of dishabituation in this case would remain ambiguous because a subject might simply perceive the formant-shifted call as “sounding synthetic” in general, rather than noticing the formant changes in particular. To exclude this possibility, we first tested a “synthetic replica” call created by applying the same software manipulations to one of the habituation calls, but without modifying formants (following Fitch and Kelley, 2000). To the human ear, synthetic replicas sound subtly cleaner (less noisy) than the original recordings, but otherwise identical. If the percept is similar for monkeys, we predicted that they would transfer habituation to this stimulus. If the subject did dishabituate to the synthetic replica, responding to resynthesis *per se*, the

session was ended. A series of such results would indicate an inadequacy in our monkey call synthesis techniques. However, if the subject transferred habituation to the synthetic replica, we proceeded to test our core hypothesis by playing the synthetic, formant-shifted call, which differed from the synthetic replica presented previously only in that the formant frequencies were experimentally modified. All other acoustic aspects (e.g., duration, amplitude, pulse rate, other timbral cues, etc.) were identical to the call played just previously (Fig. 1).

II. MATERIALS AND METHODS

A. Animal subjects

The subjects in these experiments were 13 adult rhesus macaques (*Macaca mulatta*, age 4–9 yr, 7 females, 6 males). All experiments were conducted under an approved NIMH study proposal in accord with NIH Guidelines on the Care and Use of Primates. The monkeys were on a 12 h light/dark cycle (7 am–7 pm) and the experiments were conducted between 10 am and 5 pm. The monkeys were housed at NIH in individual cages in colony rooms, and had received no previous exposure to playback of vocal stimuli before these experiments were conducted. Each monkey was run individually.

B. Behavioral protocol

This behavioral protocol followed standard habituation/dishabituation paradigms (e.g., Seyfarth *et al.*, 1980; Hauser, 1998) in most respects. We concealed a loudspeaker in a testing room, and then introduced a monkey subject, sitting in a monkey chair facing directly away from the speaker. Using a video camera to monitor head position, we waited until the subject was looking directly away from the speaker ($180^\circ \pm 20^\circ$) before initiating each playback event. The criterion for response was a head turn in the direction of the speaker initiated within 3 s of sound playback. An experiment started with repeated playback of calls from the habituation set until the subject habituated (defined as a failure to respond to three successive playbacks). The average time between playbacks was 30 s (range 8–200 s; within the range of variation of pant-threat rates observed in free-living

populations (Fitch, unpublished data). After successful habituation, we played the synthetic replica. A dishabituation response to this “control” stimulus terminated the session. However, if subjects transferred habituation to the control, we then played the formant-shifted stimulus. If the animal dishabituated, we could safely conclude that the dishabituation was caused by the formant shifts, and not by other acoustic cues or artifacts of synthesis, and the experimental session was concluded: dishabituation to the “test” stimulus terminated the session. During the habituation phase, the same sound was sometimes played twice in a row, so dishabituation to the test stimulus could not have resulted simply from the monkey perceiving repetition. Thus, consistent dishabituation to the formant-shifted stimulus constitutes strong evidence that monkeys spontaneously perceive formants in conspecific calls. However, a null result (failure to dishabituate to either test stimulus) could be caused simply by general habituation: sensory fatigue, distraction, adaptation to the playback setting, or other confounds. To exclude this possibility, sessions in which the subject failed to dishabituate to either of the previous two stimuli were ended with a “post-test” stimulus, a monkey “shrill bark” alarm call, expected to reliably elicit a response, and thus reject this final control hypothesis (following Hauser, 1998).

C. Materials and testing procedure

The experiment was performed in an empty playback room (approximately 4 × 3 m, 2.5 m h) acoustically treated with Sonex 1 in. foam (9.4 × 2 ft panels, Sonex #10897, Illbruck, Minneapolis, MN). The experimenter, computer equipment, and playback speaker were hidden by a 2.3 × 1.9 m curtain made of heavy opaque cloth that bisected the room diagonally. This curtain was acoustically transparent (reducing audio levels by only 2 dB SPL and introducing no audible distortion). Monkeys were seated in custom-built plexiglas primate chairs (20 × 25 cm, 56 cm high) which allowed free head movement in the horizontal plane. Monkeys were seated in a fixed standard position in the room, facing away from the curtain and loudspeaker, with the loudspeaker 1.5 m directly behind them. Responses were filmed using a Panasonic Digital 5000 VHS video camera mounted on a tripod, 1.2 m away and directly facing the subject, monitored via a Sony Trinitron monitor during playbacks, and simultaneously recorded to VHS tapes (Fuji HQ-120, SP) with a Sony VHS Hi-Fi recording deck. Playbacks were performed using an Apple Powerbook computer and custom playback software, using the built-in sound output (44.1 kHz sampling rate, 16 bit quantization) attached to a Bose Roommate II self-powered speaker. Playback levels were determined with a Radio Shack Sound Level Meter, set for C-weighted, fast response measurement. Sound level measurements were made by mounting the SPL meter at the location occupied by the monkey’s head during experiments. Broadband ambient ventilation noise in the playback room was 62–65 dB SPL, effectively masking computer-generated fan noise and keypresses. Playback levels were adjusted to 75–78 dB SPL and were very clearly audible above this background. Since playback time was initiated by the monkey facing calmly away

from the speaker, a maximum delay of 4 min between playbacks was chosen. If exceeded, the experiment would have been terminated, but this never occurred.

D. Stimuli and signal processing

Monkey calls were synthesized using techniques developed by WTF (Fitch and Hauser, 1995; Fitch and Kelley, 2000; Fitch, 2002). The pant-threat vocalizations used in this study were recorded from a single adult female macaque, unknown to our monkey subjects, from a rhesus population on the island of Cayo Santiago near Puerto Rico (monkey 74B, recorded by Marc D. Hauser, Harvard University, using a Sennheiser microphone and Sony Walkman Professional cassette recorder). Call spectra were examined for high frequency energy; threat calls contained no appreciable energy above 8 kHz. The three highest-quality pant-threat calls were selected as habituation stimuli, low-pass filtered (8500 Hz) and downsampled to 18.5 kHz sampling rate for further digital processing. Final versions were upsampled and played back at 44.1 kHz. One habituation call was submitted to an 18-pole linear prediction analysis (512 sample window, no preemphasis, rectangular window), yielding a filter closely approximating the smoothed magnitude spectrum of the call. The calls was then inverse filtered using this filter, yielding an error signal which approximates the laryngeal source signal (this “source signal” consisted of three impulses with some attendant noise). Once the source signal and the filter were separated, various modifications of either are independently possible. In this experiment, we scaled the entire filter function, increasing or decreasing each resonance by 20% by finding its roots (corresponding to individual formants) and then multiplying each formant frequency by a fixed factor (1.2 or 0.8). Increasing (or decreasing) the formant frequencies is analogous to shortening (or lengthening) the vocal tract, respectively. The original model call had intermediate formant frequency values, so 20% up- or downshifting of formants corresponds to a VTL of approximately 8, or 9.5 cm (about 6 and 12 kg body weight, respectively), remaining well within the normal acoustic range for adult macaques (Fitch, 1997). This relatively large change was chosen to increase the chances that the acoustic difference would be not just perceptible, but also behaviorally meaningful to our subjects, and thus to induce dishabituation (Nelson and Marler, 1989). The modified filter function was then recombined with the original source (polynomialized back into a filter function and used to create a new synthetic signal by filtering the source signal). All signal processing was performed in MATLAB 5.1 (The Mathworks, Inc., Natick, MA) using the Signal Processing Toolbox and custom software.

III. DATA ANALYSIS

All trials were videotaped for additional offline analysis. All critical trials (last habituation, synthetic replica, test, and post-test) and a randomly selected set of habituation trials from each animal were digitized (Apple iMovie software) and scored by two observers (one blind to condition). Inter-observer agreement was very high (97% agreement). To further quantify the strength of response we measured both la-

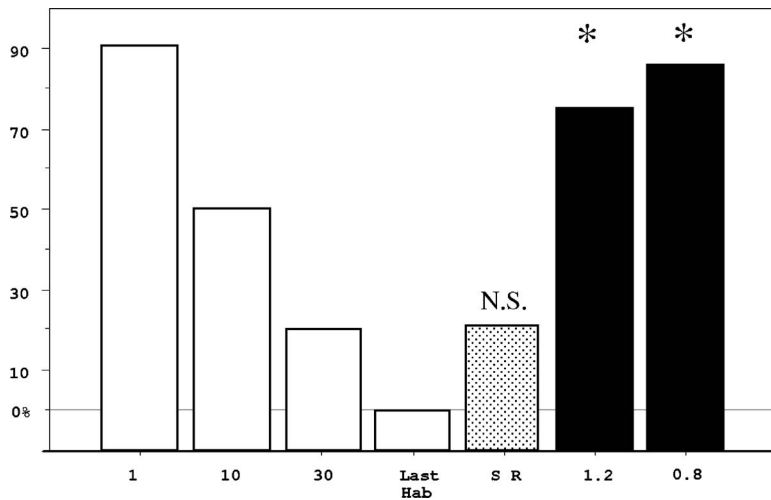


FIG. 2. Results of playbacks: Chance of response drops across habituation trials (1, 10, and 30 represent the first, tenth, and thirtieth trials), reaching 0 (by definition in our protocol) on the last habituation trial. Dishabitation to the synthetic replica (SR), where formants were not shifted, rarely occurred. Dishabitation nearly always occurred to the formant shifted stimuli (1.2 and 0.8), thus indicating perception of the formant shift imposed on these stimuli.

tency to look (in video frames, 30 frames/s) and magnitude of looks (in degrees from the initial head direction at playback to the maximum position during the look, max 180°). In two cases, monkeys' responses were ill-suited to an analysis of look latency and magnitude, once because a monkey began a head movement just before sound playback (negative latency), then followed it with a 135° turn toward the speaker in the opposite direction, and in another case because the monkey made a small 15° turn very quickly, followed by a large turn (90°) 6 s later (outside our arbitrary maximal cutoff window for scoring "looks"). In both cases, we then repeated the trial and received clear, short-latency looks that were used in the analysis.

Our original plan was to run each monkey twice, using both the up- or downshifted stimulus (randomly assigned). This would have enabled a statistical test to determine if up- or downshifting was more salient. Unfortunately, due to circumstances beyond our control, only 6 of the 13 monkeys were run a second time, yielding 19 experiments total. To ensure that repeated playbacks to a subset of monkeys did not influence our results we present analyses for both the first trials alone ($N=13$, which we term the "first" trials) and this total number ($N=19$). For statistical analysis we used one-tailed binomial tests to determine whether the proportion of looks (indicating dishabitation) relative to nonlooks (continued habituation) differed from chance. One-tailed tests are appropriate in this paradigm because the direction of the response is predicted in advance, and differs between conditions ("no" for synthetic replicas and "yes" for test stimuli). The significant p values we obtained with one-tailed statistics in general remain significant with a two-tailed test. The binomial test requires the specification of the base frequency of looking after habituation to three trials, which we took conservatively to be 50%. Statistics were performed in MATLAB 5.1 and STATVIEW 5.0 (SAS Institute, Cary, NC).

IV. RESULTS

We completed a total of 19 playback experiments with 13 different monkey subjects (six were tested twice as discussed above). The mean number of trials to habituation was 22.5 (min 4, max 45). The pattern of dishabitation is illustrated in Fig. 2.

Only four subjects dishabituated to the synthetic replica stimulus (50% binomial test 19(15), N.S.), indicating that the monkeys did not significantly dishabituate to this control stimulus. Only three monkeys dishabituated in the 13 "first" trials (50% binomial test 13(10), N.S.). Failure to consistently dishabituate to synthetic replica calls shows that the speech synthesis techniques used here are capable of generating realistic-sounding macaque calls. In the 15 remaining trials, 12 subjects who heard the formant-shifted test stimulus dishabituated (50% binomial test 15(3), $p=0.018$). For "first" trials, 8 of the remaining 10 subjects dishabituated to the test stimulus (50% binomial test 10(2), $p=0.055$). These results indicate that the monkeys heard and responded to the change in formant frequencies in these stimuli. Of the three subjects who did not dishabituate to the test stimulus, all responded to a post-test alarm call, indicating that no monkeys had habituated to the playback situation in general.

To further quantify the strength of response we measured both latency to look (in video frames, 30 frames/s) and magnitude of looks (in degrees from the initial head direction at playback to the maximum position during the look, max 180°). Monkeys looked faster to formant shifted stimuli than to synthetic replicas. Most looks (11/12) to test stimuli had latencies of less than 20 frames (667 ms) (mean 12 frames or 400 ms). In contrast, half of the looks to synthetic replicas (2/4) had latencies of more than 1 s (mean 26 frames or 867 ms). To further explore the reaction of subjects to synthetic calls, we compared looks to natural calls (last habituation, and post-test stimuli) with those to synthetic calls (the synthetic replica and test stimuli). There were no significant differences in the latency to look (Mann-Whitney $U=112$, $p=0.92$) or magnitude of looks (Mann-Whitney $U=109$, $p=0.84$) between synthetic and natural calls. We found no difference in monkeys' dishabitation to the upshifted versus downshifted test stimuli: of 12 dishabituations 6 were to upshifted and 6 to downshifted, and of the 3 failures to dishabituate 1 was to the upshifted and 2 were to the downshifted stimulus. Although monkeys looked more rapidly to the downshifted stimuli (mean 4.8 frames or 160 ms) than to the upshifted stimuli (mean 18.4 frames or 620 ms), this difference was not significant (unpaired t -test,

$t=1.45$, $p=0.174$). These findings thus reinforce the basic finding that monkeys ignored the synthetic replica and strongly responded to formant shifts in either direction.

V. DISCUSSION

The results of these experiments show that rhesus macaque perceive changes in formant frequencies in their own species-specific pant threat vocalizations. A significant proportion (12 out of 15) of our monkey subjects found formant shifts of 20% up or down to be salient enough to warrant dishabituation. Our use of a call type that lacks harmonics shows that this result cannot be attributed to perception of relative harmonic amplitude. The use of digital synthesis techniques allowed us to vary only formant frequencies, without varying other aspects of the signal. A failure to dishabituate to “synthetic replicas” which had been digitally processed *without* shifting formants shows that this result is not due to some unintended artificial quality imposed by the analysis/resynthesis technique. We conclude that rhesus macaques are sensitive to formants in vocalizations without training or reinforcement. It thus appears highly likely that formants play some role in the communication system of rhesus macaques.

These results are compatible with a number of previous results from rhesus macaques. Fitch (1997) showed that the spectral prominences in this species are formants, and showed that formant frequencies correlate with body size. Rendall *et al.* (1996) showed that free-ranging rhesus macaques distinguish identity of individuals by acoustic cues in their vocalizations, and acoustic analysis indicated that formant frequencies are potentially important cues for identity in this species (Rendall *et al.*, 1998). Injections of xylocaine into the perioral region in this species, which block the ability to produce the lip-rounding associated with “coo” calls, presumably affecting formants, led to differential reactions of conspecifics in this species (Hauser and Schön Ybarra, 1994). Combined, all of these studies converge on the conclusion that rhesus macaques perceive formant frequencies in their own vocalizations, though the information they extract from these cues remains uncertain (see the following).

Results from various other primate species provides additional converging evidence for formant perception in non-human primates. Several authors have termed spectral prominences in the calls of nonhuman primates “formants” with little further discussion or justification (Lieberman, 1968; Andrew, 1976; Richman, 1976). In baboons, species-specific “grunt” vocalizations have an acoustic structure quite similar to human vowels (Owren *et al.*, 1997) with spectral prominences hypothesized to represent formants, and significant correlations between body size and formants support this hypothesis (Rendall, 2005). A recent study shows that spectral prominences in guereza monkeys both correlate with body size, and change in close synchrony with lip movements, strongly suggesting that they represent formants (Harris *et al.*, 2006). Regarding perception, laboratory studies with the closely related Japanese macaque *Macaca fuscata*, although using synthetic vowel stimuli rather than

conspecific calls, with training demonstrated an exquisite sensitivity to formants in this species, rivaling or exceeding that of humans (Sommers *et al.*, 1992). In a training paradigm baboons were shown to be quite sensitive to formant changes in grunts synthesized with a human Klatt synthesizer (and were similarly sensitive to formant changes in human vowels) (Hienz *et al.*, 2004). Finally, vervet monkeys possess spectral prominences that may represent vocal tract resonances (Owren and Bernacki, 1998), and perceptual tests show that vervets respond to these prominences in a classification task which involved training but did not demand that the subjects attend to that particular cue (Owren, 1990b, a). All of these data converge to suggest that formant perception may be a widespread capability in Old World monkeys.

Techniques similar to those used here have recently been utilized to demonstrate spontaneous formant perception in cranes and deer (Fitch and Kelley, 2000; Reby *et al.*, 2005). In both cases the species was tested because they possess unusual vocal adaptations hypothesized to modify formants. In whooping cranes *Grus americana*, the trachea is greatly elongated. Due to the anatomy of the avian vocal production system, tracheal elongation lowers formant frequencies, and has been hypothesized as a means of exaggerating size (Fitch, 1999). This hypothesis was tested by modifying formants in a nonharmonic call (the “contact call”) using computer resynthesis, and demonstrating that listening cranes noticed this change (Fitch and Kelley, 2000). In red deer *Cervus elaphus*, adult stags have a permanently-descended larynx, and during territorial roars they lower the larynx even further to its anatomical limit. Again, this lowers formants, and was hypothesized to exaggerate projected body size (Fitch and Reby, 2001). This hypothesis was tested via playback of resynthesized calls, which demonstrated formant perception and use of formants as cues to size (Reby *et al.*, 2005). Thus, in at least two nonprimate species with formant-modifying vocal anatomy, formant perception is present.

These previous studies of animal formant perception leave open two evolutionary possibilities. First, formant perception and modification in primates, cranes and deer may represent convergent evolution. There are many examples of such convergence among vertebrate vocal communication, the most prominent being complex vocal imitation, which has evolved convergently several times (e.g., in humans, songbirds, and sea mammals), but is lacking in other primates (Janik and Slater, 1997; Fitch, 2000b; Marler and Slabbekoorn, 2004). Alternatively, formant perception in all of these species may be homologous, present in these widely separated species by virtue of inheritance from a common ancestor [the ancestral amniote, who lived some 300 million years ago (Smithson, 1989)]. If this latter hypothesis is correct, formant perception is predicted to be widespread among birds and mammals, even those which lack special formant-modifying anatomy. In particular, the crucial groups for testing the hypothesis that formant perception mechanisms are homologous in all these species are other nonhuman mammals, further bird species, and vocal reptiles such as the American alligator *Alligator mississippiensis*, which has clear formant-like bands in its vocalizations that correlate nicely

with body size (Fitch, unpublished data). It is also possible that amphibians may also perceive formants, though as discussed in Sec. I there is no evidence at present that the spectral prominences present in many anuran species represent formants (Rand and Dudley, 1993).

The ability to generate and control realistic animal acoustic signals (Fitch, 2002) opens new and exciting vistas in understanding the acoustic cues utilized in animal communication. The invention of speech synthesizers was a necessary prerequisite for the advances in speech perception starting in the 1970s, but off-the-shelf vocal synthesizers for animal calls still do not exist [although synthesizers designed for humans may be adequate in some cases (Hienz *et al.*, 2004)]. Fortunately, recent advances in our understanding of vertebrate vocal production have provided a major step forward in resolving this problem (Fitch and Hauser, 1995; Owren and Bernacki, 1988; Fitch, 2002). In particular, the realization that the source-filter theory of speech production also applies to animal vocalizations means that many of the algorithms developed by the speech community can now be applied, with the appropriate modifications, to nonhuman vocalizations. The ability to generate highly realistic animal vocalizations by computer allows us to choose and manipulate specific acoustic variables, leaving all other cues unchanged. With a careful choice of appropriate calls from a species' vocal repertoire, and new techniques for perceptual testing that do not involve training, we can now explore animal's perception of their own species-specific vocalizations at a level of detail previously impossible. Resynthesized calls can also be used in more traditional operant settings to allow accurate determination of difference limens and perceptual sensitivities (e.g., Owren, 1990b; Sinnott and Kreiter, 1991; Sommers *et al.*, 1992; Hienz *et al.*, 2004), or even in choice settings to explore perceptual preferences (McComb, 1991). Thus, the combination of digital synthesis of animal calls with playback experiments can provide a rich source of insight into the acoustic cues that play important roles in animal communication, in a wide variety of nonhuman vertebrate species.

What information might formants be providing? Differences in formant frequencies provide the primary cue to vowel identity in all human languages, and formant transitions also cue many important consonantal distinctions (Lieberman and Blumstein, 1988; Titze, 1994). Pitch information, though *present* in speech signals, is not necessary for their perception: even in so-called "tonal languages" like Chinese or Thai, formants are the key acoustic cue for most phonetic distinctions. Stimuli containing formant information alone are adequate to decode the phonetic content of speech (Remez *et al.*, 1981; Tartter, 1991), and the human auditory system automatically normalizes for vocal tract length and size information in speech from different speakers (Ives *et al.*, 2005; Smith *et al.*, 2005). Thus, of all the various acoustic cues that make up the complex speech signal, formants (or their synthetic analogues) are both necessary and sufficient for speech perception.

Previous studies suggest that formants may provide a similarly rich source(s) of information for animals (e.g., Sommers *et al.*, 1992; Fitch, 1997; Owren *et al.*, 1997; Ren-

dall *et al.*, 1998; Riede and Fitch, 1999). Formants might provide several types of information to macaques. Formants are correlated with body size in macaques, baboons, and many mammals (Fitch, 1997; Riede and Fitch, 1999; Fitch, 2000a; Rendall, 2005), so formant perception could provide information about body size, as it does in humans (e.g., Fitch, 1994; Fitch and Giedd, 1999; Ives *et al.*, 2005; Smith *et al.*, 2005). In sexually dimorphic species such as rhesus macaques, size may also provide an indirect indication of sex or other secondary factors such as age, or degree of potential threat. Formants also may provide a reliable cue to individual identity: static differences in vocal tract anatomy (particularly in the nasal region) may remain invariant across calls and thus indicate identity (Rendall *et al.*, 1998). Formants may also provide one of the basic cues that differentiate different call types, similar to the way they distinguish different vowels in human speech (Lieberman, 1968). Hauser and colleagues (Hauser *et al.*, 1993; Hauser and Schön Ybarra, 1994) found that the vocal tract movements (specifically lip-protrusion associated with coo calls, and the retracted lips associated with screams), had well-defined acoustic effects on rhesus calls acoustics. Thus, macaques could potentially extract information about size, sex, identity, and call type from formants.

In this study we used a relatively gross manipulation—shifting all formant frequencies—that corresponds to a lengthening or shortening of overall vocal tract length. The positive response to these changes by our macaque listeners clearly opens the door to a more detailed exploration of formant perception in this and related species. In particular, while overall formant dispersion may be a cue to body size (and secondarily sex or age), more detailed aspects of the formant pattern, or changes in specific formants, may provide information about individual identity (Rendall *et al.*, 1998), call type (Hauser and Schön Ybarra, 1994), or other factors. The techniques developed in the current study should be seen as first steps, but clearly open the door to much more detailed study of these and other questions. It may be particularly interesting to learn whether the lowest three formants, which carry virtually all of the phonetic information in human speech, are preferentially attended to by macaques or other primates.

Neural basis of formant perception. If formants do indeed provide a multifaceted source of relevant information to nonhuman primate listeners, the corresponding neural mechanisms involved in decoding them might be quite complex, and thus may provide a richer neural substrate relevant to the evolution of speech perception than previously suspected. At least two broad regions of auditory cortex are likely to be involved in macaque formant perception, perhaps forming part of a more complex network for vocal perception. First, neurophysiological studies reveal neurons that show robust responses to bandpass-filtered noise stimuli (which are acoustically similar to formant frequencies) in the macaque lateral belt of auditory cortex, that may also play a role in analysis of conspecific vocalizations (Rauschecker *et al.*, 1995; Rauschecker and Tian, 2004), and recent fMRI analyses indicate potentially homologous activations in humans (von Kriegstein *et al.*, 2006). Second, multisensory

neurons in the superior temporal sulcus of the temporal lobe (STS) give robust responses to vocalizations (Ghazanfar, 2005), and neuroimaging data have provided additional evidence for activation of the STS in voice recognition and perception in both humans (Belin and Zatorre, 2003; Belin *et al.*, 2004; Uppenkamp *et al.*, 2006) and macaques (Poremba *et al.*, 2004). If formants play a crucial role in cueing vocal identity (Rendall *et al.*, 1998), such STS neurons should respond strongly and specifically to formant changes. In addition to lateral belt areas and STS, other cortical areas in the macaque and human may also play a role in vocal processing, including the rostral superior temporal gyrus (Poremba *et al.*, 2004) and lateral prefrontal cortex (Averbeck and Romanski, 2004; Gifford *et al.*, 2005; Romanski *et al.*, 2005). Whatever the neural substrates, the discovery that macaques perceive formants in their own vocalizations, and thus share a crucial component of human speech perception, opens the door to neuroscientific studies of formant perception that would be difficult or impossible in humans.

In conclusion, our experiments demonstrate that rhesus macaques are both capable of perceiving changes in formant frequencies in their own species-specific vocalizations, and that they spontaneously do so, without any training. Combined with previous data, this finding supports the hypothesis that formant perception is present in nonhuman primates, thus evolving prior to human speech, and indeed may be widespread among vertebrate species. Due to the importance of formants in speech, the mechanisms underlying macaque formant perception may represent an evolutionary precursor to the more sophisticated mechanisms underlying human speech perception, and this finding thus has important potential implications for our understanding of both primate communication systems and the evolution of spoken language. Wang has proposed that there may be an auditory cortical pathway specialized for processing vocal communication sounds in primates (Wang, 2000). Many scholars have suggested that human speech perception built upon pre-existing sensory mechanisms in our primate ancestors (Snowdon, 1982; Hauser and Fitch, 2003), while others suggest that at least some aspects evolved in humans *de novo* (Sinnott and Williamson, 1999; Pinker and Jackendoff, 2005). The ability to synthesize and manipulate specific aspects of primate calls using a variety of transformations [such as the source/filter model used here (Fitch, 2002), the Mellin transform (Smith *et al.*, 2005), or a parametric “virtual vocalization” model approach (DiMattina and Wang, 2005)] provides a new tool to help resolve these debates empirically, and opens the door to detailed exploration of the information-processing mechanisms underlying vocal perception in nonhuman primates.

ACKNOWLEDGMENTS

We thank Marc D. Hauser for advice on experimental design and for generously sharing his macaque recordings, and Ricardo Gil da Costa, Asif Ghazanfar, Mortimer Mishkin, Drew Rendall, and Richard C. Saunders and several anonymous reviewers for comments on the manuscript. We thank Ludise Malkova, Richard Saunders, and Mortimer Mishkin for their support in completing these experiments,

conducted in the Laboratory of Neuropsychology, NIMH, NIH, and John Newman, Steven Suomi, Peggy O’Neill-Wagner, and Jennifer Stowe of the NIH Animal Care Center in Poolesville, MD for their help in piloting this paradigm. This work was funded by the NIH (NIDCD T32 DC00038 to W.T.F. and an NIH IRTA Postdoctoral Fellowship to J.B.F.) and by the NIMH IRP.

¹A simple analysis demonstrates that the spectral peaks in the synthetic signals used in Ehret and Riecke’s study could not represent the formants of a mouse, since a vocal tract capable of generating the frequency components that Ehret and Riecke term “formants” would have to be longer than a mouse pup’s entire body. The lowest predicted formant frequencies for a mouse pup (assuming a 1 cm vocal tract length) would be about 9 kHz, but the synthetic mouse calls had frequency components at 3.8, 7.6, and 11.4 kHz, a frequency spacing of 3.8 kHz, which (if the components were formants) would correspond to a vocal tract length of 4.6 cm. This is longer than the entire body length of a neonatal mouse (about 3 cm), whose calls the synthetic stimuli are supposed to mimic, and are indeed considerably beyond the values predicted for an adult mouse with a vocal tract length of (liberally) 2 cm (4.3, 13.1, 21.9 kHz). It would thus be physically impossible for a neonatal mouse, or even an adult mouse, to produce formants with the stated frequencies. The individual frequency components manipulated in Ehret and Riecke’s synthetic stimuli, therefore represent harmonics: aspects of the laryngeal source and not the vocal tract filter (Roberts, 1975; Weisz *et al.*, 2001; Liu *et al.*, 2003). The distinction between formants and harmonics is central in speech science, and redefining such terms in a different species “to stress {their} perceptual significance” (Geissler and Ehret, 2002) both deviates from long-accepted usage and is highly misleading if such data are to be related to speech in humans, or to vocalizations in other species.

- Amundin, M. (1991). “Helium effects on the click frequency spectrum of the Harbor porpoise, *Phocoena phocoena*,” *J. Acoust. Soc. Am.* **90**, 53–59.
- Andrew, R. J. (1976). “Use of formants in the grunts of baboons and other nonhuman primates,” *Ann. N.Y. Acad. Sci.* **280**, 673–693.
- Averbeck, B. B., and Romanski, L. M. (2004). “Principal and independent components of macaque vocalizations: Constructing stimuli to probe high-level sensory processing,” *J. Neurophysiol.* **91**, 2897–2909.
- Belin, P., Fecteau, S., and Bedard, C. (2004). “Thinking the voice: neural correlates of voice perception,” *Trends in Cognitive Science* **8**, 129–135.
- Belin, P., and Zatorre, R. J. (2003). “Adaptation to speaker’s voice in right anterior temporal lobe,” *NeuroReport* **14**, 2105–2109.
- Collins, S. A. (2000). “Men’s voices and women’s choices,” *Anim. Behav.* **60**, 773–780.
- Cynx, J., Williams, H., and Nottebohm, F. (1990). “Timbre discrimination in Zebra finch (*Taeniopygia guttata*) song syllables,” *J. Comp. Psychol.* **104**, 303–308.
- DiMattina, C., and Wang, X. (2005). “Virtual vocalization stimuli for investigating neural representations of species-specific vocalizations,” *J. Neurophysiol.* **95**, 1244–1262.
- Ehret, G., and Riecke, S. (2002). “Mice and humans perceive multiharmonic communications sound in the same way,” *Proc. Natl. Acad. Sci. U.S.A.* **99**, 479–482.
- Eimas, P. D., Siqueland, P., Jusczyk, P., and Vigorito, J. (1971). “Speech perception in infants,” *Science* **171**, 303–306.
- Fant, G. (1960). *Acoustic Theory of Speech Production* (Mouton, The Hague).
- Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., and Perrett, D. I. (2005). “Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices,” *Anim. Behav.* **69**, 561–568.
- Fischer, J. (1998). “Barbary macaques categorize shrill barks into two call types,” *Anim. Behav.* **55**, 799–807.
- Fitch, W. T. (1994). *Vocal Tract Length Perception and the Evolution of Language* (UMI Dissertation Services, Ann Arbor, MI).
- Fitch, W. T. (1997). “Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques,” *J. Acoust. Soc. Am.* **102**, 1213–1222.
- Fitch, W. T. (1999). “Acoustic exaggeration of size in birds by tracheal elongation: Comparative and theoretical analyses,” *Journal of Zoology (London)* **248**, 31–49.

- Fitch, W. T. (2000a). "Skull dimensions in relation to body size in nonhuman mammals: The causal bases for acoustic allometry," *Zoology* **103**, 40–58.
- Fitch, W. T. (2000b). "The evolution of speech: A comparative review," *Trends in Cognitive Science* **4**, 258–267.
- Fitch, W. T. (2002). "Primate vocal production and its implications for auditory research," in *Primate Audition: Ethology and Neurobiology*, edited by A. A. Ghazanfar (CRC Press, Boca Raton, FL), pp. 87–108.
- Fitch, W. T., and Giedd, J. (1999). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," *J. Acoust. Soc. Am.* **106**, 1511–1522.
- Fitch, W. T., and Hauser, M. D. (1995). "Vocal production in nonhuman primates: Acoustics, physiology, and functional constraints on 'honest' advertisement," *Am. J. Primatol.* **37**, 191–219.
- Fitch, W. T., and Hauser, M. D. (2002). "Unpacking 'Honesty': Vertebrate vocal production and the evolution of acoustic signals," in *Acoustic Communication*, edited by A. M. Simmons, R. F. Fay, and A. N. Popper (Springer, New York), pp. 65–137.
- Fitch, W. T., and Kelley, J. P. (2000). "Perception of vocal tract resonances by whooping cranes, *Grus americana*," *Ethology* **106**, 559–574.
- Fitch, W. T., Neubauer, J., and Herzel, H. (2002). "Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production," *Anim. Behav.* **63**, 407–418.
- Fitch, W. T., and Reby, D. (2001). "The descended larynx is not uniquely human," *Proc. R. Soc. London, Ser. B* **268**, 1669–1675.
- Fletcher, N. H., and Tarnopolsky, A. (1999). "Acoustics of the avian vocal tract," *J. Acoust. Soc. Am.* **105**, 35–49.
- Geissler, D. B., and Ehret, G. (2002). "Time-critical integration of formants for perception of communication calls in mice," *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9021–9025.
- Ghazanfar, A. A. (2005). "Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex," *J. Neurosci.* **25**, 5004–5012.
- Gifford, G. W., III, MacLean, K. A., Hauser, M. D., and Cohen, Y. E. (2005). "The neurophysiology of functionally meaningful categories: Macaque ventrolateral prefrontal cortex plays a critical role in spontaneous categorization of species-specific vocalizations," *J. Cogn. Neurosci.* **17**, 1471–1482.
- Harris, T. R., Fitch, W. T., Goldstein, L. M., and Fashing, P. J. (2006). "Black and white colobus monkey (*Colobus guereza*) roars as a source of both honest and exaggerated information about body mass," *Ethology* **112**, 911–920.
- Hauser, M. D. (1998). "Functional referents and acoustic similarity: Field playback experiments with rhesus monkeys," *Anim. Behav.* **55**, 1647–1658.
- Hauser, M. D., Evans, C. S., and Marler, P. (1993). "The role of articulation in the production of rhesus monkey (*Macaca mulatta*) vocalizations," *Anim. Behav.* **45**, 423–433.
- Hauser, M. D., and Fitch, W. T. (2003). "What are the uniquely human components of the language faculty?," in *Language Evolution*, edited by M. Christiansen and S. Kirby (Oxford University Press, Oxford), pp. 158–181.
- Hauser, M. D., and Schön Ybarra, M. (1994). "The role of lip configuration in monkey vocalizations: Experiments using xylocaine as a nerve block," *Brain Lang.* **46**, 232–244.
- Hienz, R. D., Jones, A. M., and Weerts, E. M. (2004). "The discrimination of baboon grunt calls and human vowel sounds by baboons," *J. Acoust. Soc. Am.* **116**, 1692–1697.
- Ives, D., Smith, D. R., and Patterson, R. D. (2005). "Discrimination of speaker size from syllable phrases," *J. Acoust. Soc. Am.* **118**, 3816–3822.
- Janik, V. M., and Slater, P. B. (1997). "Vocal learning in mammals," *Advances in the study of behavior* **26**, 59–99.
- Ladefoged, P. (2001). *Vowels and Consonants: An Introduction to the Sounds of Languages* (Blackwell, Oxford).
- Lieberman, P. (1968). "Primate vocalization and human linguistic ability," *J. Acoust. Soc. Am.* **44**, 1574–1584.
- Lieberman, P., and Blumstein, S. E. (1988). *Speech Physiology, Speech Perception, and Acoustic Phonetics* (Cambridge University Press, Cambridge, UK).
- Liu, R. C., Miller, K. D., Merzenich, M. M., and Schreiner, C. E. (2003). "Acoustic variability and distinguishability among mouse ultrasound vocalizations," *J. Acoust. Soc. Am.* **114**, 3412–3422.
- Marler, P., and Slabbekoorn, H. (2004). *Nature's Music: The Science of Birdsong* (Academic, New York).
- Martin, W. F. (1971). "Mechanics of sound production in toads of the genus *Bufo*: Passive elements," *J. Exp. Zool.* **176**, 273–294.
- McComb, K. E. (1991). "Female choice for high roaring rates in red deer, *Cervus elaphus*," *Anim. Behav.* **41**, 79–88.
- Moore, F. R. (1990). *Elements of Computer Music* (Prentice Hall, Englewood Cliffs, NJ).
- Moorer, J. A. (1979). "The use of linear prediction of speech in computer music applications," *J. Audio Eng. Soc.* **27**, 134–140.
- Nelson, D. A., and Marler, P. (1989). "Categorical perception of a natural stimulus continuum: Birdsong," *Science* **244**, 976–978.
- Nowicki, S. (1987). "Vocal tract resonances in oscine bird sound production: Evidence from birdsongs in a helium atmosphere," *Nature (London)* **325**, 53–55.
- Nowicki, S., Mitani, J. C., Nelson, D. A., and Marler, P. (1989). "The communicative significance of tonality in birdsong: Responses to songs produced in helium," *Bioacoustics* **2**, 35–46.
- Owren, M. J. (1990a). "Acoustic classification of alarm calls by vervet monkeys (*Cercopithecus aethiops*) and humans. I. Natural calls," *J. Comp. Psychol.* **104**, 20–28.
- Owren, M. J. (1990b). "Acoustic classification of alarm calls by vervet monkeys (*Cercopithecus aethiops*) and humans. II. Synthetic calls," *J. Comp. Psychol.* **104**, 29–40.
- Owren, M. J., and Bernacki, R. (1988). "The acoustic features of vervet monkey (*Cercopithecus aethiops*) alarm calls," *J. Acoust. Soc. Am.* **83**, 1927–1935.
- Owren, M. J., and Bernacki, R. H. (1998). "Applying linear predictive coding (LPC) to frequency-spectrum analysis of animal acoustic signals," in *Animal Acoustic Communication: Sound Analysis and Research Methods*, edited by S. L. Hopp, M. J. Owren, and C. S. Evans (Springer, New York), pp. 130–162.
- Owren, M. J., Seyfarth, R. M., and Cheney, D. L. (1997). "The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cyncephalus ursinus*): Implications for production processes and functions," *J. Acoust. Soc. Am.* **101**, 2951–2963.
- Pinker, S., and Jackendoff, R. (2005). "The faculty of language: what's special about it?," *Cognition* **95**, 201–236.
- Poremba, A., Malloy, M., Saunders, R. C., Carson, R. E., Herscovitch, P., and Mishkin, M. (2004). "Species-specific calls evoke asymmetric activity in the monkey's temporal poles," *Nature (London)* **427**, 448–451.
- Rand, A. S., and Dudley, R. (1993). "Frogs in helium: The anuran vocal sac is not a cavity resonator," *Physiol. Zool.* **66**, 793–806.
- Rauschecker, J., and Tian, B. (2004). "Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey," *J. Neurophysiol.* **91**, 2578–2589.
- Rauschecker, J. P., Tian, B., and Hauser, M. (1995). "Processing of complex sounds in the macaque nonprimary auditory cortex," *Science* **268**, 111–114.
- Reby, D., and McComb, K. (2003). "Anatomical constraints generate honesty: Acoustic cues to age and weight in the roars of red deer stags," *Anim. Behav.* **65**, 519–530.
- Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., and Clutton-Brock, T. (2005). "Red deer stags use formants as assessment cues during intrasexual agonistic interactions," *Proc. R. Soc. London, Ser. B* **272**, 941–947.
- Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (1981). "Speech perception without traditional speech cues," *Science* **212**, 947–950.
- Rendall, C. A. (1996). "Social communication and vocal recognition in free-ranging rhesus monkeys (*Macaca mulatta*)," University of California, Davis.
- Rendall, D. (2005). "Pitch (Fo) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry," *J. Acoust. Soc. Am.* **117**, 944–955.
- Rendall, D., Owren, M. J., and Rodman, P. S. (1998). "The role of vocal tract filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations," *J. Acoust. Soc. Am.* **103**, 602–614.
- Rendall, D., Rodman, P. S., and Emond, R. E. (1996). "Vocal recognition of individuals and kin in free-ranging rhesus monkeys," *Anim. Behav.* **51**, 1007–1015.
- Richman, B. (1976). "Some vocal distinctive features used by gelada monkeys," *J. Acoust. Soc. Am.* **60**, 718–724.
- Riede, T., Bronson, E., Hatzikirou, H., and Zuberbühler, K. (2005). "Vocal production mechanisms in a non-human primate: Morphological data and a model," *J. Hum. Evol.* **48**, 85–96.
- Riede, T., and Fitch, W. T. (1999). "Vocal tract length and acoustics of vocalization in the domestic dog *Canis familiaris*," *J. Exp. Biol.* **202**,

- Roberts, L. H. (1975). "The rodent ultrasound production mechanism," *Ultrasonics* **13**, 83–88.
- Romanski, L. M., Averbeck, B. B., and Diltz, M. (2005). "Neural representation of vocalizations in the primate ventrolateral prefrontal cortex," *J. Neurophysiol.* **93**, 734–747.
- Seyfarth, R. M., and Cheney, D. L. (1990). "The assessment by vervet monkeys of their own and another species' alarm calls," *Anim. Behav.* **40**, 754–764.
- Seyfarth, R. M., Cheney, D. L., and Marler, P. (1980). "Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication," *Science* **210**, 801–803.
- Sinnott, J. M., and Kreiter, N. A. (1991). "Differential sensitivity to vowel continua in Old World monkeys (*Macaca*) and humans," *J. Acoust. Soc. Am.* **89**, 2421–2429.
- Sinnott, J. M., and Williamson, T. L. (1999). "Can macaques perceive place of articulation from formant transition information?," *J. Acoust. Soc. Am.* **106**, 929–937.
- Smith, D. R. R., Patterson, R. D., Turner, R., Kawahara, H., and Irino, T. (2005). "The processing and perception of size information in speech sounds," *J. Acoust. Soc. Am.* **117**, 305–318.
- Smithson, T. R. (1989). "The earliest known reptile," *Nature (London)* **342**, 676–678.
- Snowdon, C. T. (1982). "Linguistic and psycholinguistic approaches to primate communication," in *Primate Communication*, edited by C. T. Snowdon, C. H. Brown, and M. R. Petersen (Cambridge University Press, New York), pp. 171–211.
- Sommers, M. S., Moody, D. B., Prosen, C. A., and Stebbins, W. C. (1992). "Formant frequency discrimination by Japanese macaques (*Macaca fuscata*)," *J. Acoust. Soc. Am.* **91**, 3499–3510.
- Suthers, R. A., and Hector, D. H. (1988). "Individual variation in vocal tract resonance may assist oilbirds in recognizing echoes of their own sonar clicks," in *Animal Sonar: Processes and Performances*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum, New York), pp. 87–91.
- Tartter, V. C. (1991). "Identifiability of vowels and speakers from whispered syllables," *Percept. Psychophys.* **49**, 365–372.
- Titze, I. R. (1994). *Principles of Voice Production* (Prentice Hall, Englewood Cliffs, NJ).
- Uppenkamp, S., Johnsrude, I. S., Norris, D., Marslen-Wilson, W., and Patterson, R. D. (2006). "Locating the initial stages of speech-sound processing in human temporal cortex," *Neuroimage* **31**, 1284–1296.
- von Kriegstein, K., Warren, J., Ives, D., Patterson, R. D., and Griffiths, T. D. (2006). "Processing the acoustic effect of size in speech sounds," *Neuroimage* (to be published).
- Wang, X. (2000). "On cortical coding of vocal communication sounds in primates," *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11843–11849.
- Weisz, D. J., Yang, B. Y., Fung, K., and Amirali, A. (2001). "The mechanism of ultrasonic vocalization in the rat," *Abstr. Soc. Neurosci.* **27**, 88.19.