

Department of Organismic & Evolutionary Biology, Harvard University

Perception of Vocal Tract Resonances by Whooping Cranes *Grus americana*

W. Tecumseh Fitch & J. Patrick Kelley

Fitch, W. T. & Kelley, J. P. 2000: Perception of vocal tract resonances by whooping cranes *Grus americana*. *Ethology* **106**, 559–574.

Abstract

Although formants (vocal tract resonances) can often be observed in avian vocalizations, and several bird species have been shown to perceive formants in human speech sounds, no studies have examined formant perception in birds' own species-specific calls. We used playbacks of computer-synthesized crane calls in a modified habituation–dishabituation paradigm to test for formant perception in whooping cranes (*Grus americana*). After habituating birds to recordings of natural contact calls, we played a synthesized replica of one of the habituating stimuli as a control to ensure that the synthesizer worked adequately; birds dishabituated in only one of 13 cases. Then, we played the same call with its formant frequencies shifted. The birds dishabituated to the formant-shifted calls in 10 out of 12 playbacks. These data suggest that cranes perceive and attend to changes in formant frequencies in their own species-specific vocalizations, and are consistent with the hypothesis that formants can provide acoustic cues to individuality and body size.

Corresponding author: W. Tecumseh Fitch, Department of Organismic & Evolutionary Biology, Harvard University, 33 Kirkland St. Rm, 982, Cambridge, MA 02138 USA. E-mail: tec@wjh.harvard.edu

Introduction

A fundamental goal in the study of animal communication is to identify the physical parameters of signals that convey relevant information to perceivers (Hauser 1996; Bradbury & Vehrencamp 1998). In studies of acoustic communication, one approach to this problem involves measuring various acoustic parameters of signals and then searching for correlations with the behavioral responses to those signals. Unfortunately, the possible number of relevant acoustic parameters is immense, and most natural signals have variations in many parameters simultaneously, making it difficult to know whether the parameters we measure are the same as those to which animals respond. This problem can be avoided by

the use of synthetic signals, where any given parameter can be precisely varied, and all others held constant. Although technological issues once limited synthesis to rather simple sounds (Doherty & Gerhardt 1984; Klump & Gerhardt 1987; Wagner 1992), the combination of powerful desktop computers and flexible signal processing algorithms today permits the creation of arbitrarily complex and realistic synthetic animal sounds (Owren 1990; Owren & Bernacki 1998). This allows us to isolate and vary a particular acoustic parameter and examine the effect of these changes on behavior in a naturalistic setting.

In the current study we use digitally synthesized calls to examine the perception of timbre in whooping cranes (*Grus americana*). Avian vocalizations have typically been analyzed in terms of changes in a single frequency through time, which is appropriate for songs of many passerine species (Greenewalt 1968; Nowicki & Marler 1988; Gaunt & Nowicki 1998). However, the call notes of many passerine species, along with many vocalizations of non-passerine birds, have a more complex frequency structure at any moment, which lead to differences in sound quality termed 'timbre' (Nowicki & Marler 1988). Timbre is a multidimensional concept: any acoustic changes that do not lead to changes in perceived pitch or loudness are considered timbral (Moore 1988). Thus the term subsumes a wide variety of possible acoustic variables under one name, and refers to a large and mostly unexplored acoustic space. Williams et al. (1989) suggested that variation in timbre offers an alternative to variation in pitch for creating variety in a species' vocal repertoire.

One type of timbral variation in bird song has been termed 'harmonic suppression' (Williams et al. 1989; Cynx et al. 1990). Harmonics are frequencies in a sound that are integer multiples of the lowest frequency, which in turn is called the fundamental frequency, and is the physical correlate of the perceptual attribute 'pitch'. Calls characterized by a fundamental plus harmonics are termed 'tonal'. Zebra finches (*Taeniopygia guttata*) produce tonal calls averaging 10 or more harmonics, and the relative amplitudes of each of these varies significantly from one song syllable to another (Williams et al. 1989). Zebra finches perceive these differences, based both on operant testing (Cynx et al. 1990) and on the fact that song learners imitate the patterns of harmonic suppression of their tutors. On the basis of a detailed analysis of the patterns of suppression, Williams et al. (1989) concluded that these harmonic amplitude variations originate at the syrinx, which is the source of acoustic energy in birds. The physical details involved in the production of this phenomenon are still unknown, but the perceptual results demonstrate that zebra finches can resolve individual harmonics and compare their relative amplitudes, a type of timbral discrimination. Similar results have been obtained from playbacks of bird song produced in a helium-oxygen atmosphere, which leads to extra harmonics not normally present in the song (Nowicki 1987). Perceivers show a stronger response to the natural version, which lacks harmonics, indicating that they can discriminate between these signals (Strote & Nowicki 1996). Thus at least one type of timbral perception, involving relative harmonic amplitude, occurs in birds.

Another means by which birds can modify the timbral structure of their

vocalizations is by vocal tract filtering. Any call generated at the syrinx must pass through and be filtered by the resonances (normal modes) of the air column in the trachea and vocal tract (Greenewalt 1968; Nowicki & Marler 1988; Fee et al. 1998; Fletcher & Tarnopolsky 1999). These vocal tract resonances are more concisely termed 'formants'. There is a systematic ambiguity in the use of the term 'formant', which refers both to vocal tract resonances (which exist physically regardless of whether they are excited acoustically) and to their manifestation in acoustic signals (spectral peaks which will reflect these resonances more or less accurately depending on the source excitation). Here, we restrict our use of the term to the former, following general usage in the speech community (Lieberman & Blumstein 1988; Titze 1994). Formants are the primary acoustic parameter conveying meaning in human speech (Fant 1960; Lieberman & Blumstein 1988; Titze 1994), and recent work has also implicated formants as important acoustic cues in non-human primate vocal communication (Owren & Bernacki 1988, Owren 1990; Hauser et al. 1993; Hauser & Schön Ybarra 1994; Fitch & Hauser 1995; Rendall 1996; Fitch 1997; Owren et al. 1997). Formants are completely independent of fundamental frequency, the variable corresponding to pitch changes in periodic calls. Despite a long history of debate over the role of vocal tract resonances in bird sound production (Greenewalt 1968; Gaunt & Gaunt 1985; Nowicki 1987; Nowicki & Marler 1988; Hausberger et al. 1991; Westneat et al. 1993; Gaunt & Nowicki 1998), little work has focused on the perception of formants in bird communication.

There are good reasons to expect birds to make perceptual use of vocal tract resonances in their own species-specific vocalizations. Formants provide an accurate indication of body size in macaques *Macaca mulatta* (Fitch 1997) and dogs *Canis familiaris* (Riede & Fitch 1999). Hinds & Calder (1971) showed that trachea length (the primary determinant of the frequencies of vocal tract resonances) is correlated with body mass in a variety of bird species, suggesting that formants could also provide a cue to body size in birds. Furthermore, Suthers & Hector (1988) and Suthers (1994) showed that variable asymmetry in oilbird (*Steatornis caripensis*) vocal tracts leads to individual differences in resonance frequencies. They suggested that these acoustic patterns could serve as individual signatures, and pointed out that 'further experiments are needed to determine if these potential acoustic cues are actually used by the birds' (Suthers & Hector 1988; p. 90). Thus, formants could provide information about the body size and/or identity of vocalizing conspecifics.

Previous research on avian formant perception is inconclusive. Birds can react to formant differences in human speech in operant paradigms (e.g. Heinz et al. 1981; Dooling & Brown 1990; Dooling 1992), and 'talking' birds such as parrots and mynahs can imitate human formants (Klatt & Stefanski 1974; Nottebohm 1976; Warren et al. 1996). However, the harmonic perception data cited above suggest that birds could perform these tasks by virtue of differences in harmonic amplitudes. In contrast, human speech perception relies on spectral peaks to extract formant frequencies irrespective of source type; hence, we can understand whispered speech which lacks any harmonics. A perceptual strategy based only on harmonic amplitudes would apply only to tonal signals and, because differences in

harmonic amplitudes could result either from dynamics of the syringeal source or vocal tract filtering, would fail to distinguish source-related spectral information from vocal tract filtering. A harmonic-based strategy would also fail on calls with a noisy or impulsive source (e.g. a great variety of sounds common in birds, variously described as rasps, clicks, rattles, squawks and chatters), which lack harmonics. In fact, such broadband signals are better suited to convey formant information than high-pitched tonal songs because they have energy spread across many different frequencies, making it more likely that any particular resonance will be excited and thus have an observable acoustic effect (Ryalls & Lieberman 1982; Nowicki & Marler 1988; Fitch & Hauser 1995). Thus, unambiguous demonstration of formant perception, as distinct from harmonic amplitude perception, requires the use of signals that lack harmonics.

The goal of the current research was to determine whether whooping cranes are sensitive to formant information in conspecific vocalizations. Specifically, we ask if cranes react to changes to the formant frequencies of conspecific contact calls in a habituation–dishabituation paradigm. Whooping cranes are large territorial birds which spend much of their time foraging terrestrially, making them easy to locate and observe. More importantly, cranes have a broad vocal repertoire, including non-tonal calls possessing clear formants. In particular, crane pairs maintain vocal contact with low-amplitude, purr-like vocalizations termed ‘contact calls’ (Archibald 1976). The pulsatile, non-tonal nature of contact calls makes them ideal for highlighting vocal tract resonances, and means that they lack harmonics or a percept of ‘pitch’. In this study, we used digital signal processing techniques to create synthetic whooping crane contact calls. We used these synthetic calls in a modified habituation–dishabituation paradigm (Cheney & Seyfarth 1988; Nelson & Marler 1989) to test the hypothesis that the cranes perceive and react to changes in vocal tract resonance frequencies in their own, species-specific calls. First, we habituated birds to natural recorded contact calls from one individual. Then, as a control, we played a synthesized replica of one of these habituation calls. Unless the synthesizer introduced perceptible artefacts into the call, the birds should remain habituated. Finally, we played a synthetic test call in which the formant frequencies had been shifted. The only difference between this call and the previous one was in its resonant frequencies; all other acoustic parameters (pulse rate, call length, source amplitude, timing, etc.) remained identical. Thus, consistent dishabituation to this last call would provide strong support for the hypothesis that cranes perceive and attend to formants.

Methods

Study Animals and Site

Seven pairs of whooping cranes living at the International Crane Foundation in Baraboo, Wisconsin were available for this study (details on crane care can be found in Swengel & Carpenter 1996). One pair provided the test calls and the other six were used as subjects in the perceptual experiment. The pair providing calls

lived in a large exhibit located 500 m from the other birds and were thus completely visually and acoustically isolated from them (the quiet contact calls are inaudible beyond about 20 m). The call donors were habituated to human observers and allowed close (within 1–2 m) approaches, necessary to record high-quality contact calls. This pair was also used in pilot experiments to develop and test the procedure and equipment. The six test pairs were housed in ‘Crane City’, a large breeding complex. Here, cranes of various species are housed in large 15 × 18 m outdoor pens which include small heated huts as shelters from bad weather (though all playbacks were performed while cranes were outdoors). Pens are 2.4 m high and are roofed with a 5-cm mesh flexible flight netting which allows clear upward vision and low flight but prevents airborne escape. The mean distance between whooping crane pens was 47 m (minimum 34 m). Cranes in this area are not well-habituated to human intruders; their pens are enclosed in dark green blind material (‘Courtmaster’ tennis netting; J. A. Cissel Mfg., Farmingdale, NJ) that greatly reduces visibility beyond the pen periphery. For all experiments, the black playback speaker and the observer (dressed in black) were hidden behind this blind material to minimize visual disturbance. The large white cranes were still clearly visible through the netting.

Vocal Tract Measurements

We collected anatomical data to ensure that our playbacks were within the normal range of variation for whooping cranes. Anatomical data consisted of radiographs of three dead crane specimens’ vocal tracts, including a 10-cm scale. These were scanned into a computer, and vocal tract lengths (from the broncho-syringeal junction to the glottis) were measured along the center of the trachea using the ‘freehand line’ tool in NIH Image 1.61 (National Institute of Mental Health, MA, USA). This measure of vocal tract length is consistent with recent theoretical (Fletcher & Tarnopolsky 1999) and empirical (Fee et al. 1998) data on avian vocal production. As a second source of data on naturally occurring variability in vocal tract length, we recorded calls ad libitum from birds of different sizes and ages, and used the LPC techniques described below to estimate vocal tract lengths and formant dispersion from 12 calls of six different individuals.

Synthesis of Experimental Calls

Contact calls (Fig. 1) are quiet, purr-like vocalizations made frequently by all crane species during normal foraging activities (Archibald 1976). Contact calls from the 4-yr-old female called ‘Oobleck’ were digitally recorded at 48 kHz (16 bits) using a Sennheiser MKH 60 P48 (Old Lyme, CT, USA) microphone and Tascam DA-P1 DAT (TEAC America, Montebello, CA, USA) recorder (filtering and automatic gain control off). These digital recordings were then transferred directly to computer hard disk (via digital input port, Digidesign Audiomeia card in an Apple Macintosh 7100). We used only recordings where background noise was inaudible. Call spectra were examined for high-frequency energy; contact calls contained no appreciable energy above 1 kHz and were thus low-pass filtered

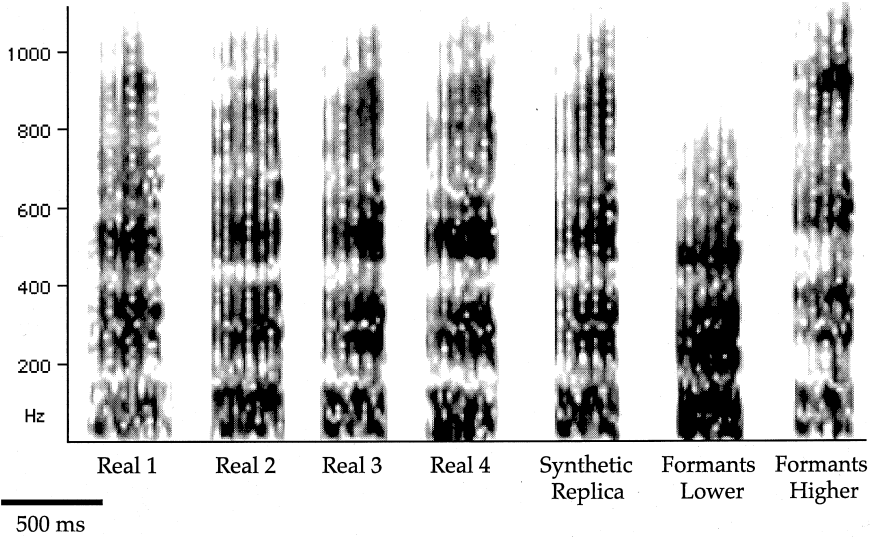


Fig. 1. Spectrograms and time series of whooping crane contact calls (sampling rate 2756 Hz, 16-bit quantization, Hanning window size 256 points); the bar at bottom left indicates 500 ms. Real 1–4 are natural exemplars; the synthetic replica is a synthesized version of Real 3; 'Formants Lower' is the experimentally 'lengthened' vocal tract (formants shifted downward in frequency) and 'Formants Higher' is the 'shortened' tract (formants shifted upward)

(1200 Hz) and downsampled to 2756 Hz sampling rate for further digital processing. Final versions were upsampled and played back at 44.1 kHz.

This description assumes some understanding of LPC analysis; see Owren & Bernacki (1998) for a tutorial introduction. The contact calls were submitted to a 30-pole linear prediction analysis (512 sample window, no pre-emphasis, rectangular window), yielding a filter closely approximating the smoothed magnitude spectrum of the contact calls. The calls were then inverse-filtered using this filter to create an error signal which was an approximation of the syringeal source signal (this 'source signal' was basically a series of broadband pulses). Once this source signal and the filter are separated various modifications of either are independently possible. For this experiment, we 'stretched' the filter function by finding its roots (individual resonances) and multiplying each resonance frequency by a fixed factor of 1.1 or 0.9, which had the effect of increasing or decreasing each resonance by 10%. Increasing the resonance frequencies is analogous to shortening the trachea, thus creating a new filter equivalent to that of a smaller or younger bird. This new 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. This degree of up- or down-shifting was well within the normal anatomical range for whooping cranes

(see 'Results'), and the resulting signals were well within the normal bandwidth range for whooping crane calls in general.

As a control, to check that the signal processing did not result in unintended audible changes to the original signal, we created a 'synthetic replica' signal. All of the same software manipulations (linear prediction analysis, inverse filtering and re-synthesis) were applied to the same call, except that formants were not shifted. To the human ear, these synthetic replicas sound subtly cleaner (less noisy) than the original recordings, but are otherwise identical. The only difference between this synthetic replica call and the purposefully manipulated test stimuli described previously was the frequency shift applied to the digital filter; all other acoustic aspects (e.g. length, amplitude, pulse rate and other timbral cues) were identical (Fig. 1).

Experimental Procedures

Birds were habituated to playbacks of unmodified contact call recordings. The four calls with the highest signal-to-noise ratio were used for this purpose; they were presented in a random order and recycled as many times as necessary for the birds to cease responding for three consecutive playbacks. Then, the synthetic replica was played (as a control); if unintended differences created by the process of analysis and synthesis were audible to the birds, they should dishabituate to this synthetic replica. In the absence of dishabituation (that is, if the synthetic replica was treated as just another habituating stimulus), we continued with a second test trial: a playback of one of the two experimentally manipulated ('test') calls. Since the only difference between this call and the synthetic replica played immediately prior to it was the experimentally induced shift in formant frequencies, dishabituation to this call would represent evidence that the birds perceived the shift in formant frequencies.

The timing of call playbacks was determined by the birds' behavior. We waited until both birds had their heads down in either a foraging position (bill near the ground; Fig. 2a) or a locomoting position (curved neck, with the head near body level). We recorded a 'looking up' response if, within 2 s of playback, the bird raised its head from this lowered position to the highest possible position above the body, with the neck straightened completely (Fig. 2b). This behavior is clear and unambiguous, and is believed to represent a vigilant 'alert' response (Swengel et al. 1996; p. 106). Each pair was treated as a single subject: a 'looking up' response from either or both members of the pair was counted as a single positive response. To score individual responses separately would violate statistical assumptions of independence for these mated pairs, which appear highly sensitive to each other's behavior. Birds which responded did so virtually immediately (< 1 s) after playback. There was a slight delay from keypress to sound playback as the sound file was loaded into RAM; occasionally (2% of trials; see 'Results'), birds looked up just before the sound played. These premature responses were termed 'uninterpretable' and were ignored, except that habituation required three unambiguous failures to respond in a row.

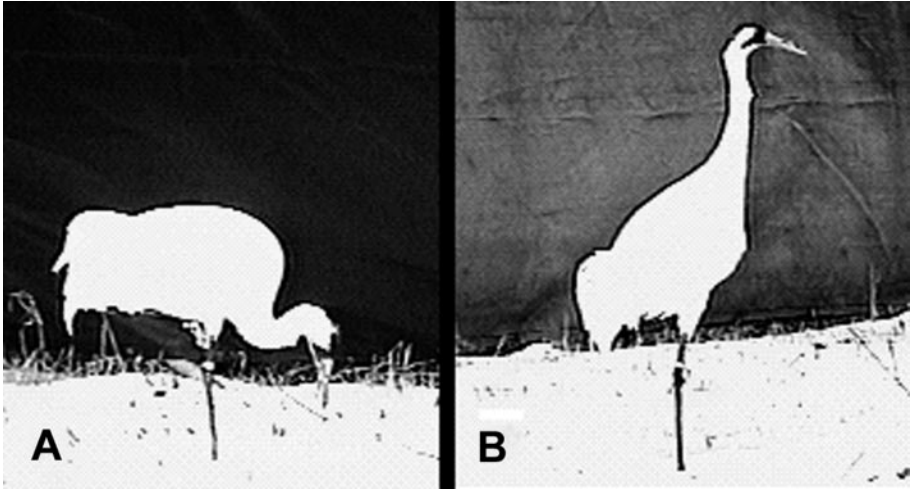


Fig. 2: Two successive video stills of vigilant response in a whooping crane. In frame A, the bird is in the 'foraging' position, with bill near the ground. In frame B, the bird has assumed the 'alert' position. If either bird in a pair assumed this final position in response to playbacks, a positive response was scored. See 'Methods' for details

Playbacks of each call were separated by an average of 26.5 s (min = 9 s, max = 130 s, SD = 17.5 s), well within the spontaneous contact-calling rate of whooping crane pairs (in 10 1-min periods, the spontaneous contact call rate of two different pairs averaged 6.4 calls min⁻¹). Although spontaneous contact calls between the subject pair were common during these experiments, other vocalizations (guard or unison calls; Archibald 1976) indicated a high level of arousal, and we ceased playbacks in the event of guard or unison calling to prevent further agitation of these valuable breeding animals. Agitation also occurred when we removed the tennis netting surrounding the pens in order to get a clear enough view for videotaping; therefore the experiments could not be videotaped and scored blindly later. However, the 'looking up' behavior that we scored was clear and unambiguous, and we believe that it is unlikely that our scoring could be affected by unconscious bias.

Playbacks were performed through a Bose Roommate II self-powered loudspeaker (Bose Corp., Framingham, MA) attached to a Macintosh Powerbook 5300 running custom software written by the first author using Hypercard 2.3 (Apple Computer). This software plays the sounds on keypress, automatically cycling through the habituation stimuli and allowing the user to enter the subjects' behavior as single-key codes, recording the time of all keypresses to the nearest 17 ms.

Results

The synthetically manipulated calls used in this study fell well within the normal anatomical and acoustical range for whooping cranes. The unmodified

contact calls we used had a formant dispersion of 240 Hz; the up- and down-shifted synthetic calls had measured dispersions of 258 and 208 Hz, respectively. Vocal tract lengths measured from radiographs of dead specimens varied from 10.7 cm (a 4-d-old chick) to 79.3 (1-yr-old female) to 102.2 cm (28-yr-old female), with corresponding predicted formant dispersions of 1640 Hz down to 171 Hz. Johnsgard (1983) reported a *G. americana* with a 147-cm trachea, supporting an even greater range of vocal tract lengths than documented here.

Measured formant dispersion in contact calls from six birds aged 4 mo to 7 yr varied from 160 to 360 Hz (mean 266 Hz), although their weights only varied from 5 to 7 kg. These data indicate that both up- and down-shifted synthetic calls were well within the normal range. Our acoustic measurements also suggest that there is substantial age- or identity-related variability in whooping cranes, as there was substantial variability in formant dispersion among birds of the same weight. This additional variance may be related to age (Niemeier 1979). However, our current sample size is inadequate to analyze these different possibilities statistically.

Thirteen playback experiments were successfully completed; eight for the shorter and five for the longer vocal tract test stimulus. Five other experiments were aborted due to calling and aggressive responses directed at the observer, neighbors or overflying wild sandhill cranes *Grus canadensis*. The average number of trials to habituation (defined as three failures to respond in a row) was 23.7 (range = 7–62; SD = 14.1). A total of 324 playback trials were made. Of these, eight were uninterpretable responses (subjects looked up before the sound played; see 'Methods') and were discarded. The average duration of an experiment was 639 s (range = 189–1887 s). Any consecutive playback experiments to the same pair of birds occurred on different days; three pairs were tested successfully three times, one pair twice and two pairs once.

There was no relationship between the probability of responding and the interstimulus interval ($n = 316$, unpaired t -test, $t = 1.34$, ns; 'yes' mean = 27.3 s, SD = 18 s, 'no' mean = 24.6 s, SD = 17 s). As expected in a habituation paradigm, responses were progressively less likely as trials continued (Fig. 3); examining all experiments together, the probability of a positive response decreased from 81% in the first five trials ($n = 64$) to 68% in trials 6–10 ($n = 60$), and had dropped to 47% by trials 16–20 ($n = 43$).

Subjects dishabituated to the synthetic replica in only one of 13 playbacks (binomial test for one or fewer, $p = 0.002$). Thus, any unintended artefacts introduced by the process of analysis and synthesis were ignored by the birds. For 10 of the 12 remaining experiments, at least one bird of the pair dishabituated to the formant-shifted call (binomial test for 10 or more, $p = 0.019$). In four of these trials, both birds looked up to the test stimulus after both birds had failed to respond in the four previous trials.

In five of the eight test trials in which the sound was modified to simulate a shorter vocal tract, birds showed a distinctive approach response: they reduced their distance to the loudspeaker by at least one half within 10 s of call playback, typically approaching directly to the pen periphery. This never occurred during

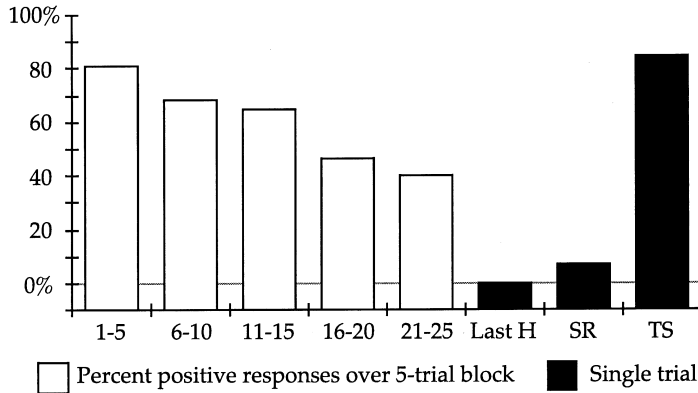


Fig. 3: Chance of response by trial, combining data from all experiments. 'Chance of response' is simply the number of times the birds looked up (a positive response was scored) divided by the total number of positive and negative responses (uninterpretable responses excluded). For the first five (open) bars, the values shown are averages over successive five-trial blocks. For the last three (solid) bars, only the data for the particular trial are shown. 'Last H' is the final habituation trial, which is by definition zero since this was the criterion for habituation. SR is the synthetic replica and TS is the test stimulus (the experimentally modified vocalization)

playback of natural adult sounds or the synthetic replica, or for the five test trials simulating a longer vocal tract.

Discussion

Six pairs of adult whooping cranes were played real and synthetic contact calls in a habituation–dishabituation paradigm. After habituating to the real contact calls of a single unfamiliar female, birds dishabituated to a synthetic replica of one of these calls only once out of 13 trials, indicating that any artefacts inadvertently introduced by the synthesizer were ignored by the birds. In 10 of the remaining 12 sessions, the birds dishabituated to a synthetic version of this same call which had the frequencies of its vocal tract resonances experimentally modified, clearly indicating that they perceived and reacted to this change. Because the calls had no harmonics, discrimination of the stimuli in this experiment could not have relied upon differences in harmonic amplitudes, as has been suggested for other species (Williams et al. 1989; Cynx et al. 1990). Furthermore, because the other aspects of the synthetic replica and test calls were acoustically identical, the birds' reactions could not have been based upon differences in pulse rate or timing, amplitude contour, or other source-related variability. These data thus provide the first experimental evidence that an avian species can perceive and respond to changes in vocal tract resonances in conspecific calls.

Before turning to the implications of the findings reported here, we will address some potential criticisms of our methodology. The first is that, by recycling the same playback stimuli repeatedly, using only two test stimuli, and using the

same subjects for more than one trial, we committed the error that Kroodsmá (1986, 1989, 1990) and others have labeled 'pseudoreplication'. Pseudoreplication typically occurs in a playback experiment when researchers interested in responses to broad classes of stimuli (e.g. the response of territorial males to known vs. unknown individuals) use a single call to represent each class. Regardless of the number of subjects tested, an experiment playing just two call exemplars in fact provides but a single data point relevant to the hypothesis under test. While 'known' calls may differ from 'unknown' calls in myriad different parameters, only one or a few vary between the two calls actually tested. Thus the real hypothesis being tested in such an experiment is whether there is a differential response to the *particular* calls employed in the experiment. Whether or not these calls indicate anything about the broader class from which they are chosen remains unknown. Such data provide very limited external validity (ability to generalize beyond the particulars of the experiment) and thus a weak test of the hypothesis of interest.

Our experiment differs from such experiments in two critical respects. First, we are interested in responses to a single acoustic cue, which we control, rather than to a broad class of stimuli varying in multiple unknown and uncontrolled cues. The analysis and synthesis techniques used here allow us to isolate a single acoustic variable in a natural vocalization and examine subjects' responses to changes in that variable alone. Secondly, our main concern here, as in most psychoacoustics experiments, is with the internal validity of our experiment: how safely can we conclude that our subjects did indeed respond to the change we made? The hypothesis under test in the current experiment, that cranes can detect changes in formant frequencies, requires repeated playback of a single test stimulus for high internal validity. If we had used a different test stimulus for each playback, other uncontrolled cues (besides formants) would have varied, reducing our certainty that dishabituations resulted exclusively from formant changes. Finally, repeated playbacks to a single subject are standard in psychoacoustics research, because this increases our confidence that a particular subject's response is due to our manipulation and not to some chance event. Indeed, many psychoacoustics experiments, both with humans and with animals, use only two or three subjects and involve hundreds of repetitions of the same stimuli. Our repeated playbacks to seven pairs of birds in this experiment thus are quite reasonable, and make it unlikely that the statistically significant response we observed is limited to one or two 'odd birds'.

A second potential flaw in our experiment concerns the sequential playback of the two test stimuli, the synthetic replica and the formant-shifted call. If cranes do *not* perceive changes in formant frequencies, these two calls would be perceived as a repeat of the exact same stimulus. Because our habituation sequence did not contain exact repeats (our randomization procedure sampled the habituation stimuli without replacement), it could have been this perceived repetition that caused dishabituation, rather than the formant change. To evaluate this hypothesis we examined another set of repetitions, cases in which the synthetic replica followed directly after the natural stimulus upon which it was based. This occurred in four cases, and dishabituation occurred in none of these. While this number of trials is

too small for a valid statistical test, we believe that this observation renders the 'repetition' hypothesis extremely unlikely.

Thus, we argue that the experimental methodology employed in this study provides a strong test of the hypothesis 'cranes responded to changes in the formant frequencies of conspecific contact calls'. The use of habituation to a pre-specified criterion ensures that all birds were brought to the same level of habituation before hearing the test stimuli, reducing the chance that individual differences in excitability or vigilance could account for our results. The failure of our subjects to respond to the synthetic replica makes it unlikely that their renewed response was due to artefacts we unwittingly induced with our synthesis procedure. The reliable dishabituation observed when the formants, but nothing else, changed provides strong support for the hypothesis. Thus, we can safely conclude from our data that our *particular* subjects responded to changes in these *particular* calls, and that the changes they responded to were changes in formant frequencies.

Further work will be required to increase the generality (external validity) of this result. However, our results suggest that the perceptual abilities demonstrated by our subjects would be applicable to natural crane vocalizations from different individuals. The acoustical and anatomical data we collected indicate that our synthetic formant changes fell well within the range of variability observed among normal cranes, so the magnitude of formant change they were able to perceive in this experiment is comparable to that between different normal individuals. Of course, smaller changes in formants (or, theoretically, even larger ones) might elicit no response. Formant changes may elicit response in contact calls, but not in other crane vocalizations. It could be that the contact calls of our 'donor' crane were particularly well suited to reveal formant changes, and those of other cranes would be less so (although if this were the case it would not refute the hypothesis that cranes can hear formant changes). Such limits to generalization exist in any single experiment, and can only be overcome by repeated testing using different subjects, call types, calls, and techniques. However, both the descriptive anatomical and acoustic data and the experimental data are consistent with the broader hypothesis that whooping cranes can and do perceive formants in the calls of conspecifics.

In what ways could formant cues be used by listeners? One possibility is that vocal tract resonances are a cue to individuality. Suthers & Hector (1988) and Suthers (1994) described vocal tract resonances in oilbirds that were due to individually variable differences in the length of the two bronchi. They speculated that the resultant differences in formant frequencies could provide a means of recognizing other birds, as well as a way of distinguishing one's own echolocation calls. Similarly, in primates, Rendall (1996) suggested that formants in macaque vocalizations might provide cues to individual identity due to individual differences in vocal tract length or shape. Our results are compatible with this idea: after habituating to one bird's calls, subjects may have interpreted the transformed contact call as coming from a different bird, and thus shown renewed interest.

A second possibility is that vocal tract resonances provide information about body size. Tracheal length is positively correlated with body size in many birds (Hinds & Calder 1971), and vocal tract length should be the main determinant of

the spacing between adjacent tracheal resonances (Lieberman & Blumstein 1988; Titze 1994). Thus, resonance spacing or 'formant dispersion' (Fitch 1997) could provide a cue to the body size of a calling bird. A suggestive finding of the current study was that an approach reaction was observed in five of eight playbacks of the shortened-trachea calls, but was never observed to the normal- or long-trachea calls. Adult cranes appear extremely interested in young birds (Voss 1974; Swengel, pers. comm.), and playbacks of calls from young cranes often elicit an approach from adults (Fitch, unpubl. data). The frequent approach to the synthetic short-trachea calls observed in the current study is consistent with the hypothesis that the calls with the shorter trachea were perceived as a smaller bird by the subjects. While clearly inconclusive, these data provide some support for the notion that formant information is used by cranes to gauge body size.

This hypothesis is particularly interesting in light of the wide variety of bird species that exhibit tracheal elongation (55 species in 10 families; Berndt 1938; Clench 1978; Niemeier 1979; Frith 1994; Fitch 1999). In these species, the adult trachea is greatly elongated, typically forming loops or coils within the bird's body. If birds use information on vocal tract resonances to gauge body size, tracheal elongation could provide a way to 'fake' larger size: a small bird could duplicate the acoustic structure of a much larger bird's calls (Fitch 1999). Although experimental work with cranes by Gaunt et al. (1987) showed that tracheal length has little effect on voice pitch, and demonstrated a decrease in loudness with either lengthening or shortening of the trachea, their evidence with regards to formant frequencies was ambiguous. Despite the fact that these workers concluded that the crane trachea has little role in determining the spectral structure of crane vocalizations, they did not address the potential adaptive functions of passive vocal tract filtering (Gaunt et al. 1987; p. 56), as discussed here.

Formant perception might also provide information on caller age (consequent to size information), reproductive maturity (Marion 1977 showed that tracheal length provides an indication of male sexual maturity in the plain chachalaca *Ortalis vetula*) or sex (in many species with tracheal elongation, only one sex exhibits the trait; Niemeier 1979). In any case, these hypotheses are not mutually exclusive, since formants could provide information relevant to multiple caller characteristics (such as sex and body size), even in a single species.

The main finding of our study, that at least one bird species perceives and responds to formant changes in conspecific vocalizations, has clear implications for the study of bird vocal communication in general. Although 'whistled' song is often the most obvious part of the vocal repertoire of many species, this should not lead researchers to ignore the large and potentially highly informative world of broad bandwidth calls (Nowicki & Marler 1988). The acoustic structure of such calls will typically be strongly affected by vocal tract filtering, and thus more acoustically similar to the sounds of human speech (or the growls, purrs and barks of many mammals). Fortunately, digital signal processing algorithms designed for formant analysis (such as LPC and cepstral analysis) are now widely available, and can be readily modified to work with bird vocalizations. The use of computer-synthesized vocalizations permits the controlled manipulation of individual acous-

tic variables, allowing animal researchers the kind of experimental precision which was invaluable in understanding the perception of human speech (see e.g. Liberman 1996; Liberman & Blumstein 1988). Because the production acoustics of broadband avian vocalizations are so similar to those of human speech, future studies on the production and perception of vocal tract resonances in birds may be expected to reveal further interesting parallels between bird vocal communication and human language.

Acknowledgements

Sincere thanks are due to George Archibald, Director and founder of the International Crane Foundation, for permission to work at this unique center, to Julie Langenberg for access to the anatomical data, and to Scott Swengel, Curator of Birds, for his logistic support at the ICF and invaluable insights into crane behavior. Thanks also to Leigh Taylor for impromptu logistic support. This report benefited from comments by Asif Ghazanfar, Jean-Guy Godin, Marc Hauser, Roderick Suthers, Scott Swengel, Daniel Weiss, and three anonymous reviewers. This work was supported by an NIH/NIDCD Postdoctoral Training Fellowship (T32 DC00038) to the first author.

Literature Cited

- Archibald, G. W. 1976: The unison call of cranes as a useful taxonomic tool. PhD Thesis, Cornell Univ., Ithaca.
- Berndt, R. 1938: Intersternale Trachealschlingen bei Vögeln. *Gegenb. Morph. Jahrbuch*, **82**, 27—117.
- Bradbury, J. W. & Vehrencamp, S. L. 1998: *Principles of Animal Communication*. Sinauer Associates, Sunderland, MA.
- Cheney, D. L. & Seyfarth, R. M. 1988: Assessment of meaning and detection of unreliable signals by vervet monkeys. *Anim. Behav.* **36**, 477—486.
- Clench, M. H. 1978: Tracheal elongation in birds-of-paradise. *Condor* **80**, 423—430.
- Cynx, J., Williams, H. & Nottebohm, F. 1990: Timbre discrimination in Zebra finch (*Taeniopygia guttata*) song syllables. *J. Comp. Psychol.* **104**, 303—308.
- Doherty, J. A. & Gerhardt, H. C. 1984: Evolutionary and neurobiological implications of selective phonotaxis in the spring peeper (*Hyla crucifer*). *Anim. Behav.* **32**, 875—881.
- Dooling, R. J. 1992: Perception of speech sounds by birds. *Adv. Biosci.* **83**, 407—413.
- Dooling, R. J. & Brown, S. D. 1990: Speech perception by budgerigars (*Melopsittacus undulatus*): Spoken vowels. *Perception Psychophysics* **47**, 568—574.
- Fant, G. 1960: *Acoustic Theory of Speech Production*. Mouton, The Hague.
- Fee, M. S., Shraiman, B., Pesaran, B. & Mitra, P. P. 1998: The role of nonlinear dynamics of the syrinx in the vocalizations of a songbird. *Nature* **395**, 67—71.
- 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. *J. Zool. (Lond.)* **248**, 31—49.
- Fitch, W. T. & Hauser, M. D. 1995: Vocal production in nonhuman primates: acoustics, physiology, and functional constraints on 'honest' advertisement. *Am. J. Primatol.* **37**, 191—219.
- Fletcher, N. H. & Tarnopolsky, A. 1999: Acoustics of the avian vocal tract. *J. Acoust. Soc. Am.* **105**, 35—49.
- Frith, C. B. 1994: Adaptive significance of tracheal elongation in manucodes (Paradisaeidae). *Condor* **96**, 552—555.
- Gaunt, A. S. & Gaunt, S. L. L. 1985: Syringeal structure and avian phonation. In: *Current Ornithology*, Vol. 3 (Johnson, R. F., ed.). Plenum Press, New York, pp. 213—245.
- Gaunt, A. & Nowicki, S. 1998: Sound production in birds: Acoustics and physiology revisited. In: *Animal Acoustic Communication: Sound Analysis and Research Methods* (Hopp, S. L., Owren, M. J. & Evans, C. S., eds), Springer, New York, pp. 291—321.

- Gaunt, A. S., Gaunt, S. L. L., Prange, H. D. & Wasser, J. S. 1987: The effects of tracheal coiling on the vocalizations of cranes (Aves: Gruidae). *J. Comp. Physiol. A* **161**, 43—58.
- Greenewalt, C. H. 1968: *Bird Song: Acoustics and Physiology*. Smithsonian Institution Press, Washington.
- Hausberger, M., Black, J. M. & Richard, J.-P. 1991: Bill opening and sound spectrum in barnacle goose loud calls: individuals with 'wide mouths' have higher pitched voices. *Anim. Behav.* **42**, 319—322.
- Hauser, M. D. 1996: *The Evolution of Communication*. MIT Press, Cambridge, MA.
- Hauser, M. D. & Schön Ybarra, M. 1994: The role of lip configuration in monkey vocalizations: Experiments using xylocaine as a nerve block. *Brain Language* **46**, 232—244.
- Hauser, M. D., Evans, C. S. & Marler, P. 1993: The role of articulation in the production of rhesus monkey (*Macaca mulatta*) vocalizations. *Anim. Behav.* **45**, 423—433.
- Heinz, R. D., Sachs, M. B. & Sinnott, J. M. 1981: Discrimination of steady-state vowels by blackbirds and pigeons. *J. Acoust. Soc. Am.* **70**, 699—706.
- Hinds, D. S. & Calder, W. A. 1971: Tracheal dead space in the respiration of birds. *Evolution* **25**, 429—440.
- Johnsgard, P. A. 1983: *Cranes of the World*. Indiana Univ. Press, Bloomington.
- Klatt, D. H. & Stefanski, R. A. 1974: How does a mynah bird imitate human speech? *J. Acoust. Soc. Am.* **55**, 822—832.
- Klump, G. M. & Gerhardt, H. C. 1987: Use of non-arbitrary acoustic criteria in mate choice by female gray tree frogs. *Nature* **326**, 286—288.
- Kroodsma, D. E. 1986: Design of song playback experiments. *Auk* **103**, 640—642.
- Kroodsma, D. E. 1989: Suggested experimental designs for song playbacks. *Anim. Behav.* **37**, 600—609.
- Kroodsma, D. E. 1990: Using appropriate experimental designs for intended hypotheses in 'song' playbacks, with examples for testing effects of song repertoire sizes. *Anim. Behav.* **40**, 1138—1150.
- Liberman, A. M. 1996: *Speech: A Special Code*. MIT Press, Cambridge, MA.
- Lieberman, P. & Blumstein, S. E. 1988: *Speech Physiology, Speech Perception, and Acoustic Phonetics*. Cambridge Univ. Press, Cambridge.
- Marion, W. R. 1977: Growth and development of the plain chachalaca in south Texas. *Wilson Bull.* **89**, 47—56.
- Moore, B. C. J. 1988: *An Introduction to the Psychology of Hearing*. Academic Press, New York.
- Nelson, D. A. & Marler, P. 1989: Categorical perception of a natural stimulus continuum. *Birdsong Science* **244**, 976—978.
- Niemeier, M. M. 1979: Structural and functional aspects of vocal ontogeny in *Grus canadensis* (Gruidae: Aves). PhD Thesis, Univ. of Nebraska, Lincoln.
- Nottebohm, F. 1976: Phonation in the orange-winged Amazon parrot, *Amazona azonica*. *J. Comp. Physiol. A* **108**, 157—170.
- Nowicki, S. 1987: Vocal tract resonances in oscine bird sound production: evidence from birdsongs in a helium atmosphere. *Nature* **325**, 53—55.
- Nowicki, S. & Marler, P. 1988: How do birds sing? *Music Perception* **5**, 391—426.
- Owren, M. J. 1990: Acoustic classification of alarm calls by vervet monkeys (*Cercopithecus aethiops*) and humans: I. Natural calls. *J. Comp. Psychol.* **104**, 20—28.
- Owren, M. J. & Bernacki, R. 1988: The acoustic features of vervet monkey (*Cercopithecus aethiops*) alarm calls. *J. Acoust. Soc. Am.* **83**, 1927—1935.
- Owren, M. J. & 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* (Hopp, S. L., Owren, M. J. & Evans, C. S., eds). Springer, New York, pp. 130—162.
- Owren, M. J., Seyfarth, R. M. & 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.
- Rendall, D. 1996: Social communication and vocal recognition in free-ranging rhesus monkeys (*Macaca mulatta*). PhD Thesis, Univ. of California, Davis.
- Riede, T. & Fitch, W. T., 1999: Vocal tract length and acoustics of vocalization in the domestic dog *Canis familiaris*. *J. Exp. Biol.* **202**, 2859—2867.
- Ryalls, J. H. & Lieberman, P. 1982: Fundamental frequency and vowel perception. *J. Acoust. Soc. Amer.* **72**, 1631—1634.

- Strote, J. & Nowicki, S. 1996: Responses to songs with altered tonal quality by adult song sparrow (*Melospiza melodia*). *Behaviour* **133**, 161—172.
- Suthers, R. A. 1994: Variable asymmetry and resonance in the avian vocal tract: a structural basis for individually distinct vocalizations. *J. Comp. Physiol.* **175**, 457—466.
- Suthers, R. A. & 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* (Nachtigall, P. E. & Moore, P. W. B., eds). Plenum Press, New York, pp. 87—91.
- Swengel, S. R. & Carpenter, J. W. 1996: General husbandry. In: *Cranes: Their Biology, Husbandry, and Conservation* (Ellis, D. H., Gee G. F. & Mirande, C. M., eds), National Biological Service, Washington, pp. 31—43.
- Swengel, S. R., Archibald, G., Ellis, D. & Smith, D. 1996: Behavior management. In: *Cranes: Their Biology, Husbandry, and Conservation* (Ellis, D. H., Gee, G. F. & Mirande, C. M., eds). National Biological Service, Washington, pp. 105—122.
- Titze, I. R. 1994: *Principles of Voice Production*. Prentice Hall, Englewood Cliffs.
- Voss, K. S. 1974: Observations on the adoption of a sandhill crane chick by a pair of Manchurian cranes. *Avicultural. Mag.* **80**, 14—16.
- Wagner, W. E., Jr, 1992: Deceptive or honest signalling of fighting ability? A test of alternative hypotheses for the function of changes in call dominant frequency by male cricket frogs. *Anim. Behav.* **44**, 449—462.
- Warren, D. K., Patterson, D. K. & Pepperberg, I. M. 1996: Mechanisms of American English vowel production in a gray parrot (*Psittacus erithacus*). *Auk* **113**, 41—58.
- Westneat, M. W., Long, J. H. J., Hoese, W. & Nowicki, S. 1993: Kinematics of birdsong: Functional correlation of cranial movements and acoustic features in sparrows. *J. Exp. Biol.* **182**, 147—171.
- Williams, H., Cynx, J. & Nottebohm, F. 1989: Timbre control in Zebra finch (*Taeniopygia guttata*) song syllables. *J. Comp. Psychol.* **103**, 366—380.

Received: May 26, 1999

Initial acceptance: September 6, 1999

Final acceptance: December 18, 1999 (J.-G. Godin)