Processing vocal signals for recognition during antiphonal calling in tamarins

CORY T. MILLER, CARMEN G. IGUINA & MARC D. HAUSER

Primate Cognitive Neuroscience Laboratory, Department of Psychology, Harvard University

The ability to recognize a conspecific signal is essential to communication. In addition to recognizing the type of call, receivers extract a range of information from the signal about the producer, including identity, sex and dialect. Despite the apparent ease with which this is accomplished, few available data address the computational processes underlying recognition. While it is possible that recognition of the signal and its information content occur in a single stage of processing, different components of the signal may be processed separately. Here we present a series of experiments designed to examine this issue in the cottontop tamarin, *Saguinus oedipus*. Using the tamarins' natural vocal response to hearing their species-specific combination long call (CLC), antiphonal calling, we presented tamarins with manipulated and unmanipulated CLCs and measured both the number of antiphonal responses and the latency to produce an antiphonal call. Results indicated that recognition of the call type and recognition of the caller occur in separate computational stages of signal processing. These data provide insights into how sensory information is organized by a call recognition system.

Vocal signal recognition systems must contend with an array of problems when processing auditory information. They must parse species-specific vocalizations from the plethora of other sounds in the environment with similar acoustic properties. Because most species produce a number of acoustically distinct vocalizations, the system must also be able to recognize the specific type of vocalization. Vocal signals also provide signal receivers with cues about the caller, such as individual identity, sex, size, motivation and dialect (Gerhardt 1992), and an efficient recognition system must extract this information. Although numerous studies have explored vocal signal recognition and the significance of particular acoustic features (Narins & Capranica 1978; Searcy & Marler 1981; Nelson 1988; Nelson & Marler 1989; Gerhardt 1991; Hauser 1998; Bee & Gerhardt 2002; Miller et al. 2004), few have examined how the details of signal morphology are processed by recognition systems. The primary aim of this paper is to explore the dynamics of a vocal signal recognition system and test how different components of vocalizations are organized by the system for recognition.

Studies of natural call recognition typically rely on a discrete response measure such as whether a respondent turns towards a speaker, approaches, or calls (e.g. Ryan 1980; Nelson & Marler 1989; Gerhardt 1991; Miller & Hauser 2004). Using a graded measure such as latency to respond may, however, yield important insights into the processes underlying recognition. For example, studies of lexical access employing Stroop or Picture–Word Interference tasks use subjects' response latency as a way to determine whether interference occurs when processing different types of information (MacLeod 1991; Costa et al. 2003). In a Stroop task, subjects are slower to state the colour of a printed word if the meaning and the printed colour are mismatched (e.g. 'green' printed in red ink as opposed to green ink). By presenting such stimuli that differ in semantic relatedness, interference in lexical access occurs and can decrease or increase subjects' latency to respond. The pattern of responses can then be used to determine the stage of processing at which the interference occurred. Response latency has also been used to examine specialized processing systems for species recognition in birds. Dooling et al. (1992) showed that budgerigars, *Melospittacus undulatus*, canaries, *Serinus canaria*, and zebra finches, *Taeniopygia guttata castanotis*, show shorter response latencies when recognizing conspecific vocalizations compared to the calls of other species. Because all of the bird species in the experiment were

Correspondence and present address: C. T. Miller, Laboratory of Auditory Neurophysiology, Department of Biomedical Engineering, Johns Hopkins University School of Medicine, 720 Rutland Ave., Ross 424, Baltimore, MD 21205, U.S.A. (email: ctmiller@bme.jhu.edu).

housed in a single aviary, the frequency with which individuals heard their own species' calls versus the other species' calls could not account for this result. The authors argued that these results provided evidence of a specialized system for processing conspecific vocalizations compared to other auditory events, including the calls of other bird species. This suite of studies, as well as others (Perrett et al. 1988; Williams et al. 2004), suggests that response latency can be used to examine the dynamics of information processing systems.

Here we describe a series of experiments in which we used the antiphonal calling behaviour of cottontop tamarins, Saguinus oedipus, to examine how vocal signals are processed for recognition. Antiphonal calling occurs when one individual produces a long-distance vocalization, such as a contact or long call, and another individual responds by producing the same vocalization type. The vocal response is classified as the antiphonal call and typically occurs within a few seconds of receiving the eliciting vocalization. This specialized vocal behaviour represents a natural (i.e. untrained) recognition system because a sound must be recognized as a specific type of conspecific vocal signal to elicit an antiphonal call. Antiphonal calling is known to occur in several non-human primate species (hereafter primate), including squirrel monkeys, Saimiri sciurius (Biben 1993; Soltis et al. 2002), cottontop tamarins (Ghazanfar et al. 2001; Miller et al. 2001a), saddle-back tamarins, Saguinus fuscicollis (Windfelder 2001), emperor tamarins, Saguinus imperator (Windfelder 2001), and common marmosets, Callithrix jacchus (C. T. Miller & X. Wang, unpublished data).

Antiphonal calling exchanges in cottontop tamarins involve their species-specific combination long call (CLC; Fig. 1). The CLC is a multipulsed vocal signal consisting of two acoustically distinct syllable types that occur in a stereotyped order: 1–2 chirps followed by 3–4 whistles (Cleveland & Snowdon 1982; Miller et al. 2003b). Like other primate long calls (Miller & Ghazanfar 2002), CLCs serve to maintain group spacing and may be involved in mate attraction (Miller et al. 2001b, 2004). The CLC has been the focus of numerous studies

of tamarin vocal behaviour. In addition to studies of antiphonal calling in tamarins (Ghazanfar et al. 2001, 2002; Miller et al. 2001a; Miller & Hauser 2004), other investigations have examined aspects of CLC production (Miller et al. 2003a) and perception (Weiss & Hauser 2002). Acoustic analyses indicate that this signal carries information about the caller, such as species, individual identity, sex, and group membership (Weiss et al. 2001). Experimental evidence shows that tamarins are sensitive to these acoustic signatures and use the information to make behavioural decisions (Weiss et al. 2001; Miller et al. 2003; Miller & Hauser 2004).

Antiphonal calling has been used to examine call recognition in tamarins in a number of previous investigations (Ghazanfar et al. 2001, 2002; Miller et al. 2001a; Miller & Hauser 2004). All of these studies, however, used only a discrete measure of recognition: specifically, the number of antiphonal calls produced in response to a variety of natural and manipulated stimuli. Throughout the experiments presented here, we measured both the number of calls produced to various manipulated and unmanipulated CLCs, as well as the latency to call antiphonally to the same stimuli. The rationale for using these two measures is as follows. Antiphonal calls are produced in response to hearing a CLC. When presented with CLCs for which the acoustic structures of the calls have been manipulated outside the naturally produced range, tamarins cease to call antiphonally (Ghazanfar et al. 2002; Miller & Hauser 2004). Therefore, if the number of antiphonal calls produced in response to manipulated and unmanipulated signals is similar, then we can infer that tamarins recognize both stimulus types as representative CLC exemplars. In contrast, the latency to call antiphonally is an indication of the processing time required to determine that the sound is a CLC. If the number of calls produced to each stimulus remains constant, but latencies increase for responses to manipulated calls relative to unmanipulated ones, then the target feature is perceptually salient, and imposes an added signal-processing cost for call recognition to occur. The additional processing time can then be used to infer how
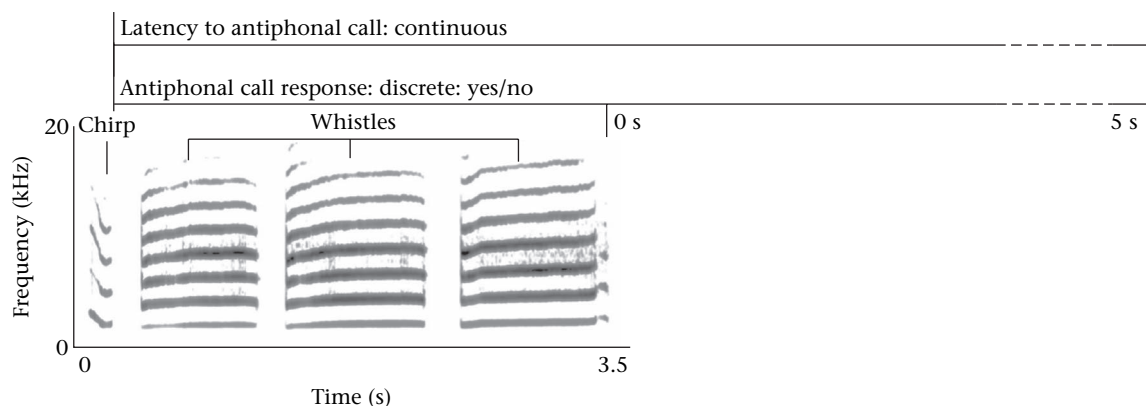

Figure 1. Spectrogram of a combination long call denoting the two syllable types: chirps and whistles. The timing used for both the continuous and discrete response measures are also depicted.

aspects of the vocalization (e.g. call type, caller identity, caller sex, etc.) are parsed by the signal-processing system to enable CLC recognition for antiphonal calling.

GENERAL METHODS

Subjects

Six adult cottontop tamarins (three males, three females) served as subjects for this study. All subjects participated in at least one previous study of antiphonal calling (Ghazanfar et al. 2001, 2002; Miller et al. 2001a; Miller & Hauser 2004) and therefore were familiar with the testing apparatus. There was, however, no evidence that subjects habituated or changed responses to stimuli between experiments. All subjects were born in captivity and housed at the Harvard University Primate Cognitive Neuroscience Laboratory. The animals in this colony were fed a diet of monkey chow, fruit, sunflower seeds, peanuts and yoghurt, and had access to water ad libitum.

Apparatus

The test apparatus was a box consisting of five Plexiglas sides and a wire-mesh front positioned inside an acoustic chamber (Industrial Acoustics Company, Inc., Bronx, New York, U.S.A.; Model 400A). Stimuli were broadcast from an Apple G4 computer using a Digidesign M-Box with 16-bit sound, an Alesis RA-100 Amplifier and an Alesis Monitor One speaker (frequency range 45–18 000 Hz). More details on the experimental set-up are provided elsewhere (Ghazanfar et al. 2001, 2002; Miller et al. 2001a; Miller & Hauser 2004). During test sessions, we used a Sennheiser directional microphone (ME-80) to record all vocalizations to a Tascam DAT recorder. We performed all signal manipulations for all experiments using SoundEdit 16 v. 2.0 (San Francisco, California U.S.A.).

Procedure

We transported subjects individually from the home-room to the testing room via a transport box. Once inside the test chamber, we placed subjects inside the testing apparatus, closed the chamber and started the experiment. We randomized stimuli and then broadcast them at intervals of approximately 30 s using a custom Hypercard program (written by W. T. Fitch). This procedure followed previous antiphonal calling experiments (Ghazanfar et al. 2002; Miller & Hauser 2004). We ran no more than one stimulus set on each subject in a single day.

Stimulus Sets

We generated four stimulus sets for each test condition. As each subject was run on four test sessions for each condition, a different stimulus set was used in each session. All stimulus sets consisted of 10 stimuli, five baseline and five test stimuli. The term 'baseline' is used because stimuli in this class were the point of comparison for the test stimuli. With the exception of the baseline stimuli used in experiment 3, condition B, all baseline stimuli were unmanipulated, naturally produced CLCs. All baseline CLCs were selected from a large corpus of prerecorded CLCs produced by individuals in the colony. These calls were spontaneously produced CLCs and recorded while subjects were isolated in a sound- attenuated chamber. The calls used as 'foreign' CLC stimuli were produced by deceased individuals of the colony. None of the living individuals in the colony had experience with these individuals.

Each stimulus set in experiments 1 and 2 was based on a single naturally produced CLC. Each call was produced by a different individual living in the colony. As such, subjects were familiar with the individuals that produced the calls. For each stimulus set, the natural CLC was manipulated and four copies were generated for each of the original baseline and manipulated CLCs. Together with the original calls, this yielded 10 stimuli for each set: five normal baseline CLCs and five manipulated 'test' CLCs (Fig. 1). In experiment 3, we selected for each stimulus set a CLC produced by each subject's mate (baseline call) and a CLC from a foreign animal (test stimulus) of the opposite sex as the subject's mate. We generated four copies of these calls, yielding a total of 10 CLC stimuli per set: five mate, five foreign.

Analysis

Following previous studies (Ghazanfar et al. 2001, 2002; Miller et al. 2001a; Miller & Hauser 2004), we considered all instances in which a CLC was produced at least 300 ms following stimulus onset and within 5 s of stimulus offset as an antiphonal response. We included for analysis only those sessions in which subjects responded to at least one of each of the stimulus types. If subjects failed to accomplish this level of volubility, the test session was repeated. We implemented these criteria because measures of latency require at least one response to each stimulus type. However, given that most subjects produced multiple antiphonal calls to each stimulus type, few sessions were omitted as a result of failure to satisfy a criterion. In experiment 1, condition D, the test stimulus was manipulated by deleting the second and third whistles in the CLC. Because this stimulus was significantly shorter than baseline stimulus, the response period allowed for scoring the antiphonal call included both the typical 5 s, as well as the duration the stimulus was shortened. As such, the total amount of time subjects could produce an antiphonal call was identical for both the baseline and test stimuli in this condition.

We measured two aspects of subjects' antiphonal responses to test stimuli. First, we compared the total number of antiphonal responses subjects produced for both test stimuli in each session (Fig. 1). Second, we recorded all instances in which a stimulus elicited an antiphonal call using Real Time Spectrogram Display (RTSD; Beeman 2001) and measured the latency in milliseconds from stimulus onset to the onset of the antiphonal response (Fig. 1). Because the number of antiphonal responses to each stimulus type (manipulated/unmanipulated) varied within a test session, we averaged each

subject's latency to respond to each stimulus type within each test session.

It is important to emphasize the differences between the two response types. Although subjects' propensity to call antiphonally was based on the 5-s period following the offset of the stimulus, we measured the latency to antiphonally call from the onset of the stimulus (Fig. 1). We justify this measure as follows. The duration of a stimulus varies. When measuring whether subjects responded to the stimulus, we attempted to provide each subject with a comparable time period to respond following stimulus presentation. Because some subjects antiphonally called during playback stimuli, we measured latency from the onset of our broadcast.

All data were analysed using two-tailed repeated measures ANOVAs, comparing subjects' responses (number of antiphonal calls and latency to the antiphonal call) across test sessions; significance was set at $P < 0.05$.

EXPERIMENT 1

We manipulated the CLCs by locally reversing all or some of the individual syllables in the call (Fig. 2). This manipulation served to test whether the recognition system was sensitive to changes in the acoustic structure of individual syllables when the overall temporal pattern is preserved. Furthermore, given the stereotyped temporal structure of CLCs, the conditions in this experiment tested whether individual syllables in the sequence of syllables serve a specific role in call recognition.

Results

Condition A

Test stimuli in this condition consisted of CLCs in which all syllables were reversed in their local position. Results indicated that subjects produced similar numbers of antiphonal calls in response to baseline and test CLCs (ANOVA: $F_{1,5} = 2.73$, $P = 0.16$; Fig. 2) and that there was no interaction between call rate across test sessions ($F_{3,15} = 1.94$, $P = 0.17$). The mean $\pm$ SE latency to produce an antiphonal response was 3364 ± 298.6 ms for baseline calls and 4190 ± 343.1 ms for local-reversal 'test' calls. This difference was statistically significant ($F_{1,5} = 13.32$, $P = 0.01$; Fig. 2). In addition, there was no interaction between latency to respond and session number ($F_{3,15} = 0.88$, $P = 0.47$), suggesting that the difference in latency to the two stimuli was consistent across test sessions.

Condition B

All syllables in test stimuli were locally reversed syllables with the exception of the first whistle. Analyses showed no difference in the number of antiphonal calls between the two stimuli ($F_{1,5} = 2.14$, $P = 0.2$; Fig. 2) and no interaction between response to the stimuli and test session ($F_{3,15} = 1.18$, $P = 0.35$). The mean $\pm$ SE latency to call antiphonally in response to baseline CLCs was 4047 ± 398.5 ms and 4049 ± 357.1 ms in response to test calls, a statistically nonsignificant difference ($F_{1,5} = 0.001$,

$P = 0.99$; Fig. 2). Similarly, no interaction between response latency and test session emerged ($F_{3,15} = 0.49$, $P = 0.69$).

Condition C

Test stimuli consisted of locally reversed chirps and whistles with the exception of the last whistle. No significant difference was observed in the number of antiphonal calls produced between baseline and manipulated CLCs ($F_{1,5} = 0.05$, $P = 0.83$; Fig. 2). Similarly, no interaction occurred between rates of antiphonal calling and test session ($F_{3,15} = 0.58$, $P = 0.64$). The mean $\pm$ SE latency to the antiphonal call was 3505 ± 284.1 ms for natural calls and 4715 ± 261.7 ms for manipulated CLCs. This difference was statistically significant ($F_{1,5} = 38.47$, $P = 0.002$; Fig. 2) and consistent across experimental sessions ($F_{3,15} = 2.36$, $P = 0.16$).

Condition D

We presented subjects with baseline CLCs and calls in which the chirps were reversed, the first whistle was normal and the subsequent whistles were deleted. Subjects showed no difference in the number of antiphonal calls between baseline and test stimuli ($F_{1,5} = 1.25$, $P = 0.31$; Fig. 2) and no interaction between stimulus and test session ($F_{3,15} = 0.94$, $P = 0.45$), suggesting a consistent pattern of response. The mean $\pm$ SE latency to respond antiphonally to baseline calls and test calls was 3698 ± 454.7 ms and 4364 ± 310.5 ms, respectively. This difference, however, was not statistically significant across sessions ($F_{1,5} = 1.05$, $P = 0.35$; Fig. 2). Furthermore, no interaction emerged between stimulus and test session ($F_{3,15} = 0.13$, $P = 0.94$), suggesting a consistent pattern of response over the condition.

Discussion

This experiment tested whether CLC recognition was mediated by a single whistle syllable, or whether a combination of whistles was necessary for recognition without additional processing. For each condition, subjects produced antiphonal calls at the same level for normal and manipulated stimuli, suggesting that all stimuli were recognized as CLCs. The only measure that varied between the stimuli was latency to respond. Results from condition A showed that CLCs in which all syllables were locally reversed elicited a significantly slower antiphonal call response. However, in condition B, when we presented an unmanipulated first whistle followed by reversals of the next two whistles, subjects showed no difference in latency to the call compared with normal CLCs. This result suggests that either the tamarin recognition system places special emphasis on the first whistle in the sequence for recognition or that as long as any single whistle in the sequence is in the normal, forward position, no additional processing is necessary for correct recognition. In condition C, we distinguished between these possibilities by presenting subjects with a CLC in which all whistles were locally reversed except for the final
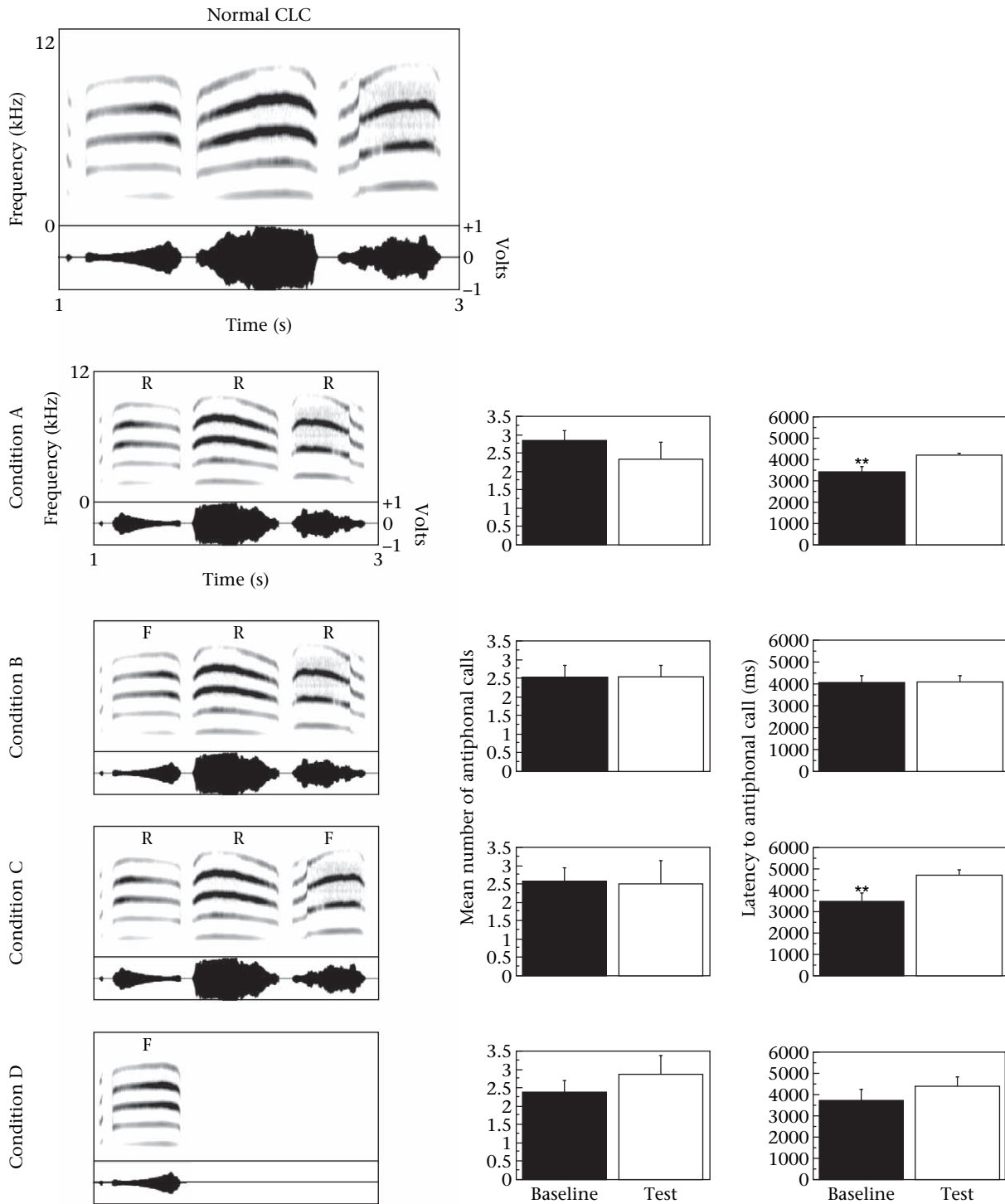
Figure 2. Spectrograms and waveforms of a representative normal 'baseline' combination long call (CLC) and of the test stimuli used for each condition. The direction of each whistle syllable is noted as either being forward (F) or reversed (R) for each of the test stimuli. Bar graphs indicate the mean number of antiphonal calls and the mean latency to antiphonal calling for each condition. Standard error bars are shown for each of these measures. An asterisk denotes a statistically significant difference.

whistle. Here subjects showed an increase in latency to call, suggesting that the first whistle is crucial for CLC recognition to occur without additional processing. To further test the significance of the first whistle for CLC recognition, we presented subjects with CLCs in which the last two whistles were omitted in condition D. In this

condition, subjects showed no additional latency to call antiphonally, suggesting that the sound was correctly recognized without additional processing.

In white-crowned sparrows, *Zonotrichia leucophrys oriantha*, the first whistle of the song is the primary cue used early in development for recognizing conspecific

song (Soha & Marler 2000). The results presented here appear to be consistent with this finding. However, in contrast to those observed in the first whistle in white-crowned sparrow song, few differences are apparent in the overall acoustic structure of individual CLC whistles (Weiss et al. 2001). Therefore, the significance of the first whistle for recognition is probably due to its location in the CLC syllable sequence rather than anything unique to the acoustic structure of the first whistle itself.

The antiphonal response is a stereotyped vocal behaviour. And thus it is possible that, to prompt an antiphonal response, the recognition system extracts all of the necessary information from the chirps and first whistle, thus ignoring all information encoded in the subsequent whistles. Because tamarins only produce antiphonal calls to CLCs, the minimal amount of information necessary to produce the antiphonal response is recognition of the call type. However, CLCs also contain a rich array of information about the caller, such as individual identity, sex and group membership (Weiss et al. 2001). Whether this information is involved in antiphonal calling, however, is unclear. This issue is examined in experiment 2.

EXPERIMENT 2

Here we examined the significance of the acoustic information encoded in the whistles following whistle 1 in the following manner. We presented subjects with CLCs in which the chirps and first whistle were from one CLC, and the remaining whistles were from a different CLC. In both conditions, the chirps and first whistle of the manipulated CLC were from calls produced by the subject's mate. The second and third whistles used as stimuli in condition A were produced by a foreign animal of the opposite sex as the subject's mate, whereas the whistles used in condition B were from a different call produced by the same individual (Fig. 3). Condition B served as a control because the stimuli in this condition underwent the same manipulation as in condition A except that the first whistle and the newly attached subsequent whistles were produced by the same individual. Key to this experiment is that, for the test stimuli used in both conditions, call type did not change. In condition A, however, the caller changed after the second whistle. If the tamarin recognition system does not extract information about the caller before the second whistle or if information about the caller is not used in antiphonal calling, then subjects should show comparable latencies to respond to normal and manipulated calls in both conditions. If, however, tamarins extract information about the caller before the second whistle, then we should observe a change in the antiphonal response in condition A, but not in condition B.

Results

Condition A

We presented normal CLCs from each subject's mate and test calls in which the chirps and first whistle were from the mate and the subsequent whistles were from a foreign animal of the opposite sex as the mate. Subjects showed no difference in the number of antiphonal calls produced between baseline and test CLCs ($F_{1,5} = 0.04$, $P = 0.85$; Fig. 3) and there was no interaction between stimulus type and test session ($F_{3,15} = 1.21$, $P = 0.34$). The mean $\pm$ SE latency to antiphonally call was 3719 ± 341.1 ms for baseline CLCs and 4868 ± 328.3 ms for test calls. This difference was statistically significant ($F_{1,5} = 13.6$, $P = 0.01$; Fig. 3). In addition, there was a significant interaction between stimulus type and test session ($F_{3,15} = 6.39$, $P = 0.005$). This interaction effect was due to the first test trial where subjects showed a slight bias to respond more quickly to manipulated calls, but showed a longer latency to respond to manipulated calls on all subsequent trials. Specifically, subjects' mean response latency for baseline stimuli was 4057 ms, 2995 ms, 3942 ms and 3849 ms for each the four test sessions, and 3280 ms, 5016 ms, 5887 ms and 5288 ms for the test stimuli. As such, the trend in the first session was the opposite to that for the other sessions.

Condition B

Here we presented subjects with test CLCs in which the chirps and the first whistle were from a CLC produced by the subject's mate and the subsequent whistles were from a different call exemplar also produced by the mate. No difference in the number of antiphonal calls produced emerged between the stimulus types ($F_{1,5} = 2.65$, $P = 0.17$; Fig. 3), with no interaction between stimulus type and test session ($F_{3,15} = 1.23$, $P = 0.33$). The mean $\pm$ SE latency to respond to baseline calls and test calls was 4010 ± 318.1 ms and 3972 ± 269.2 ms, respectively. Analyses showed no difference in latency to call to either of these stimuli ($F_{1,5} = 0.01$, $P = 0.91$; Fig. 3) and no interaction between stimulus type and test session ($F_{3,15} = 0.34$, $P = 0.79$).

Discussion

Results from this experiment indicate that the tamarin vocal signal recognition system does not ignore the second and third whistles of the CLC during antiphonal calling exchanges, but rather appears to extract information about the caller from these syllables. Although this was not evident from data on the number of antiphonal calls produced, which were statistically similar for all normal and manipulated CLCs within each condition, differences in response latency provided evidence that this information was being processed. When presented with a CLC chimera in which the caller changed between the first and second whistles, subjects showed an increase in response latency (condition A). In condition B, however, no latency differences were observed when the chimera involved using the second and third whistles from a different call exemplar produced by the same caller that produced the chirps and first whistle. Because the call type (i.e. CLC) remained constant in the test stimuli in condition A, and only the information about caller changed, the increased latency must have been due to the system detecting an inconsistency in the information
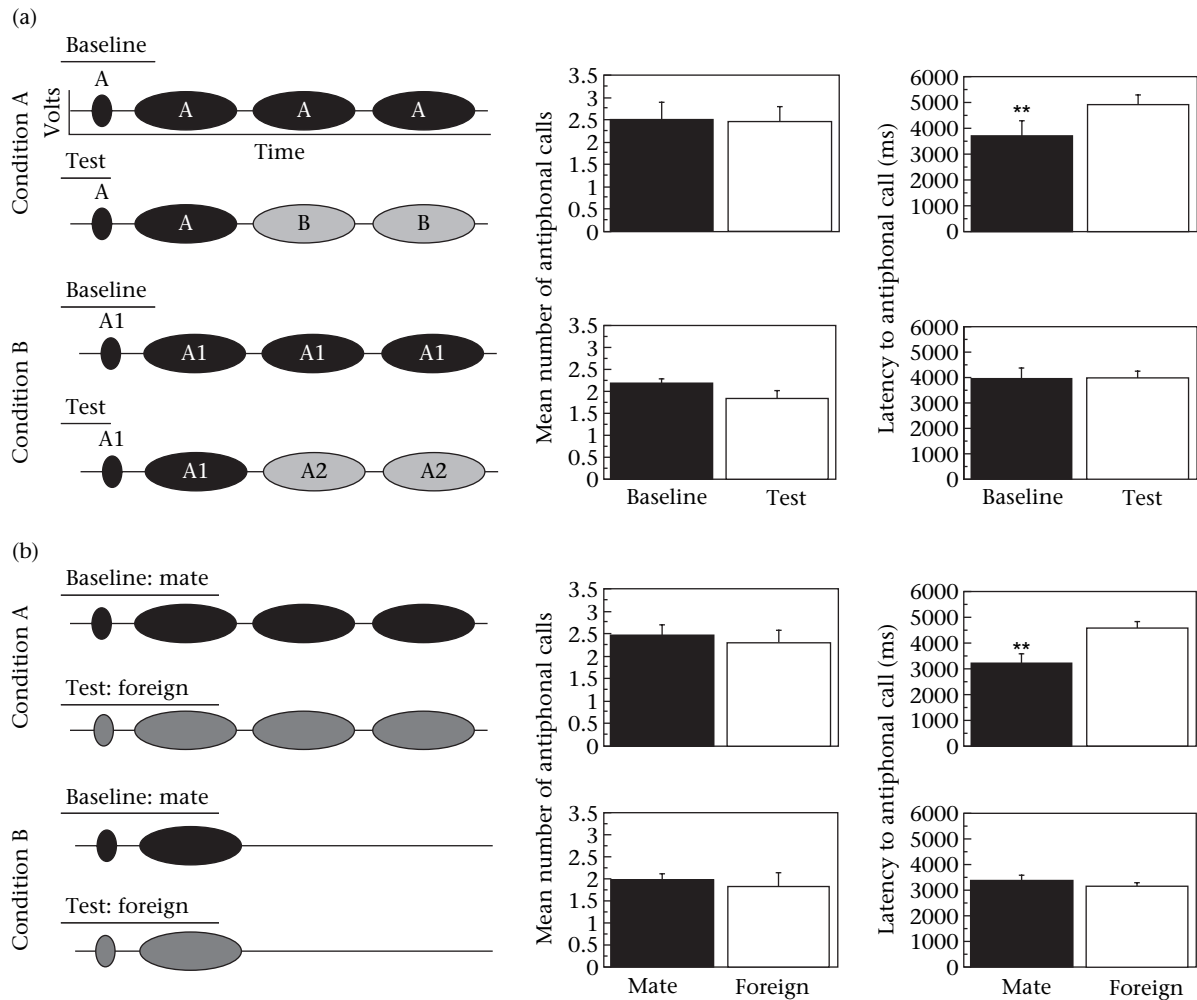
Figure 3. Results from experiments 2 (a) and 3 (b). Schematic drawings of combination long call (CLC) waveforms on the left show the baseline (above) and test (below) stimuli used in each test condition. For experiment 2, condition A, the syllables in each stimulus produced by individuals A and B are shown. For experiment 2, condition B, the syllables in each stimulus from CLC A1 and A2 are noted. Bar graphs indicate the mean number of antiphonal calls and the mean latency to antiphonal call for each condition. Standard error bars are shown for each of these measures. An asterisk denotes a statistically significant difference.

being extracted about the caller. This result could occur because the first whistle provides sufficient information about both the call type and caller. The shift in caller between the first and second whistles was apparently difficult to reconcile for the recognition system, thus causing a longer latency to respond. Alternatively, the recognition system may not have determined caller identity by the end of the first whistle, but had begun extracting this information and determined that an inconsistency had occurred between the first and second whistles. Experiment 3 was designed to test between these possibilities.

EXPERIMENT 3

Results from the first two experiments suggest that the chirps and first whistle are sufficient for eliciting an antiphonal response without additional processing time. Additionally, evidence suggests that antiphonal calling is

also modulated by acoustic information about the caller encoded in the vocalization. What remains unclear is whether both types of information are necessary for the antiphonal response to commence. It could be that recognition of the call type, but not of the caller, is necessary to elicit antiphonal calling. Alternatively, both types of information may be necessary for antiphonal calling to occur. The distinction here may depend on whether recognizing both the caller and call type rely on the same amount of acoustic information. In other words, given that a signal consisting of only chirps and a single whistle did not affect either call rate or latency to call (experiment 1, condition D), the acoustic information encoded in this portion of the call may be sufficient to induce normal patterns of antiphonal calling. If subjects can assess the caller based on this amount of acoustic information, then both call type and information about the caller can be extracted from the first whistle and could be involved in every antiphonal call response. If, however, the chirps and first whistle are insufficient for

recognizing the caller, then the information content of the signal is most likely unnecessary for antiphonal calling. To distinguish between these two possibilities, we conducted the following experiment. In condition A, we presented subjects with unmanipulated CLCs produced by their mates and foreign animals of the opposite sex. The same stimuli were used in condition B, but these stimuli consisted of only the chirps and first whistle. If both call type and caller recognition occur based only on the chirps and first whistle, subjects should show no difference in antiphonal calling between these two conditions. If a difference in antiphonal calling emerges between the two conditions, however, then the information contained in the remaining whistles is necessary for recognition.

Results

Condition A

In this condition, we presented subjects with CLCs produced by their mates and foreign individuals of the opposite sex as their individual mates. Analyses revealed no difference in the number of antiphonal calls produced in response to the two stimulus types ($F_{1,5} = 0.45$, $P = 0.53$; Fig. 3) and no interaction between stimulus type and test session ($F_{3,15} = 0.32$, $P = 0.81$). Furthermore, subjects showed a significantly longer mean $\pm$ SE latency to call antiphonally to foreign tamarin CLCs compared to subjects' mates (mate: 3213 ± 280.2 ms; foreign: 4579 ± 293.7 ms; $F_{1,5} = 22.97$, $P = 0.005$; Fig. 3) and there was no interaction between stimulus type and test session ($F_{3,15} = 0.10$, $P = 0.96$).

Condition B

Here we presented subjects with CLCs consisting of only the chirps and first whistle from calls produced by their mates and foreign tamarins of the opposite sex as their individual mates. Results showed no difference in rates of antiphonal calling between the stimulus types ($F_{1,5} = 0.20$, $P = 0.67$; Fig. 3) and no interaction between stimulus type and test session ($F_{3,15} = 0.38$, $P = 0.77$). The mean $\pm$ SE latency to call antiphonally to cagemate and foreign calls was 3360 ± 282.9 ms and 3162 ± 250.4 ms, respectively. This difference was not statistically significant ($F_{1,5} = 1.82$, $P = 0.23$; Fig. 3). In addition, no interaction between stimulus class and test sessions emerged ($F_{3,15} = 1.43$, $P = 0.27$).

Discussion

This condition tested whether subjects were able to recognize the caller from either the whole call or from the chirps and first whistle alone. For both test conditions, subjects responded at statistically similar levels to cagemate and foreign calls, suggesting that both stimulus types were recognized as CLCs. In contrast, when presented with whole CLCs produced by their cagemates or foreign animals, subjects showed an increase in latency to antiphonally call in response to foreign CLCs.

However, this difference disappeared when we presented subjects with only the chirps and first whistle. These data suggest that the chirps and first whistle of CLCs are insufficient to recognize the caller. Multiple whistles are required for the recognition system to extract this information from the signal. Overall, these results indicate that recognizing the call type (i.e. CLC) alone is sufficient for tamarins to produce the antiphonal call response. Although recognizing the caller of the signal influences signal-processing speed, it is not necessary for eliciting the antiphonal calling response. We hypothesize, therefore, that whereas call-type recognition is sufficient for antiphonal calling, caller recognition is neither necessary nor sufficient to elicit an antiphonal call.

BASELINE RESPONSE COMPARISON

The experiments presented here compared subjects' latency to respond and the number of antiphonal calls produced in response to various manipulated and unmanipulated stimuli in different test conditions. The antiphonal response, however, is untrained and therefore potentially quite variable. If we are to draw strong conclusions through comparisons between conditions, it is important to test whether subjects' responses to baseline CLCs were similar across the entire experiment. If the response properties changed significantly over the course of the experiment, it would not be appropriate to make direct comparisons between test conditions. If, however, subjects' baseline response behaviours were consistent across the test conditions, then comparisons between the conditions would be more valid.

Results

Overall, subjects produced a mean of 2.4 antiphonal calls in response to baseline CLCs. A repeated measures ANOVA revealed no main effect of condition ($F_{7,35} = 1.33$, $P = 0.26$) or test session ($F_{3,15} = 0.09$, $P = 0.96$), as well as no interaction between these two factors ($F_{21,105} = 1.29$, $P = 0.19$). This suggests that subjects produced a statistically similar number of antiphonal calls in response to baseline stimuli. Subjects' mean $\pm$ SE latency to respond to baseline stimuli was 3638 ± 130.3 ms. A repeated measures ANOVA revealed no significant effect of test condition ($F_{7,35} = 0.76$, $P = 0.61$) or test session ($F_{3,15} = 1.5$, $P = 0.24$), suggesting that subjects' responses did not vary across the different test conditions and sessions. Additionally, no interaction between condition and session emerged ($F_{21,105} = 1.38$, $P = 0.14$), suggesting that subjects' latency to respond to baseline stimuli was similar for the test sessions within each condition.

Discussion

This analysis suggests that subjects' response to baseline CLCs was consistent over the course of this study. Given this result, however, we must ask why subjects responded to CLCs with such consistent speed following stimulus presentation. Our data do not permit us to

ascertain the specific mechanism that motivated the speed of this response. Previous analyses show that most vocal responses to playbacks of CLCs occur quickly following the stimulus (Ghazanfar et al. 2002). This suggests that an innate impetus to promptly produce antiphonal calls in response to conspecifics may be the motivational force that drives subjects' response latency. Because the antiphonal call is essentially a long-distance contact behaviour, a prompt antiphonal response may communicate to the initial caller a willingness to interact, at least vocally.

Given the range of response latency (515–8480 ms), it is clear that the mean latency is not the threshold at which subjects can respond. However, requiring subjects to respond as quickly as they are capable may not be necessary because the comparison between subjects' responses to baseline and test calls is a relative measure, not an absolute one. Subjects' latency to respond to baseline CLCs was consistent over the course of the experiments. As such, motivation was effectively controlled. Our comparison between the baseline and test stimuli is not indicative of the absolute capacity of the tamarin recognition system, but rather permits us to determine whether a manipulated call requires additional processing time for recognition relative to a baseline unmanipulated call.

GENERAL DISCUSSION

The ability to quickly recognize the communicative signals of conspecifics is imperative to the survival of many animals. These experiments sought to examine the principles underlying vocal signal recognition in an effort to better understand how recognition systems organize sensory information for rapid recognition. One possible mechanism for facilitating recognition is to parse components of the signal and process them separately. For example, the structure of a vocalization (call type) and information content (caller identity) represent a logical separation (e.g. Endler 1993). Whether recognition systems actually process this information in different computational stages or whether all acoustic information involved in vocal signal recognition occurs in a single processing stage is unknown. Results presented here indicate that the first whistle provides tamarins with sufficient information to recognize the call type (i.e. combination long call), but multiple whistles are necessary to recognize the caller. With regard to antiphonal calling, we consider call-type recognition and caller recognition to be separable computational processes for the following two reasons. First, only call-type recognition is necessary to initiate the antiphonal response. Caller recognition can modulate the antiphonal response, but the antiphonal response may occur without this information. Second, only call-type recognition requires the chirps and first whistle, whereas recognizing caller identity involves the redundancy of multiple whistles. This suggests that these two components of the signal are encoded with distinct suites of acoustic features and require different computational algorithms to extract these cues from the signal.

One possible explanation for the pattern of results observed in this study is that stimulus novelty induces a longer latency to call antiphonally. Although possible, the data presented here are inconsistent with this interpretation. In experiment 1, the test stimuli used in condition B and condition C each consisted of only one forward whistle, with all other syllables being reversed. As such, both test stimuli consisted of an equal amount of 'novel' acoustic information, yet subjects showed a latency to respond in only one of these conditions (condition C). Furthermore, in experiment 3, we presented subjects with CLCs produced by a familiar cagemate and an unfamiliar foreign animal. Again, subjects showed a latency to call in only one of these conditions (condition A). Together, these results suggest that a 'novelty' explanation is insufficient to explain the results presented here. Rather, the results are consistent with the interpretation that a longer latency to respond to a signal is indicative that the algorithms used for recognition require additional acoustic information to determine the call type and caller identity.

There are two potential ways that a signal-processing system could be structured to process both caller identity and call type using distinct computational algorithms for recognition. The first type of recognition system would involve a serial processing mechanism. In such a system, a second stage of processing to identify the caller could begin only after the sound has been determined to be a species-specific vocalization. The prediction here is that a temporal separation should occur between processing call type and caller identity, with the former being processed first followed by the latter. A second type of recognition system, in contrast, would process the call type and caller identity in parallel. For this system, all the acoustic features from the sound would be processed simultaneously by the call-type recognition component and caller-recognition component of the system. Here, whether a temporal difference occurs between call-type and caller recognition would be less clear because the different components of the signal may require more or less acoustic information for recognition.

Results from the tamarin experiments presented here suggest that vocal signal recognition involves parallel processing mechanisms for call-type and caller recognition. Data suggest that during the first whistle, the recognition system extracts the acoustic information pertinent to both call-type and caller information simultaneously. However, failure to determine the caller does not affect subjects' propensity to respond to the stimulus as a CLC. Namely, in all conditions, subjects produced antiphonal calls to all baseline and test stimuli at statistically similar rates, suggesting that they recognized all stimuli as CLCs. Therefore, interrupting one level of information processing (i.e. caller recognition) does not appear to interfere with call-type recognition. This does not mean that call-type recognition is impervious to perturbations of the signal. Other results show that manipulating the acoustic structure of CLCs outside the natural range of the call will lead to a decrease in antiphonal call rate and thus, CLC recognition (Miller &

Hauser 2004). Furthermore, results of experiment 2, condition A suggest that, although the caller cannot be recognized using the first whistle alone, information about the caller's identity is being accrued during the first whistle. Hence, when caller identity was changed following the first whistle, subjects' latency to call increased because an inconsistency was detected in that processing stream. If serial mechanisms were employed for signal processing during antiphonal calling, we would not expect a longer response latency when caller was switched at that point in the call because information about caller identity would have begun following the first whistle. Overall, this pattern of results suggests that call-type and caller recognition are parallel processes during CLC recognition (Fig. 4). Although the temporal difference between processing these two aspects of the signal does occur, with call type being recognized before caller, this is probably a by-product of the signal's design rather than evidence of a serial processing system. In other words, the acoustic structure of CLCs provides sufficient information for recognizing the call type early in the call, but to recognize a caller, more of the signal must be processed.

The results presented here are analogous to data from experiments on face recognition in primates. In a neuro-imaging experiment examining face recognition in humans, Liu et al. (2002) showed that a first neural response correlated with recognition of the object as a face, whereas a second response occurred when subjects were able to recognize the individual identity of the face. Similar results were obtained in an electrophysiology study of rhesus monkeys, *Macaca mulatta* (Sugase et al. 1999). The similarities between these face-recognition studies and the data presented here for tamarins could occur for several reasons. Because both vocalizations and faces convey communicative information, it may be that sensory systems evolved similar mechanisms to extract and categorize information encoded in communicative signals.
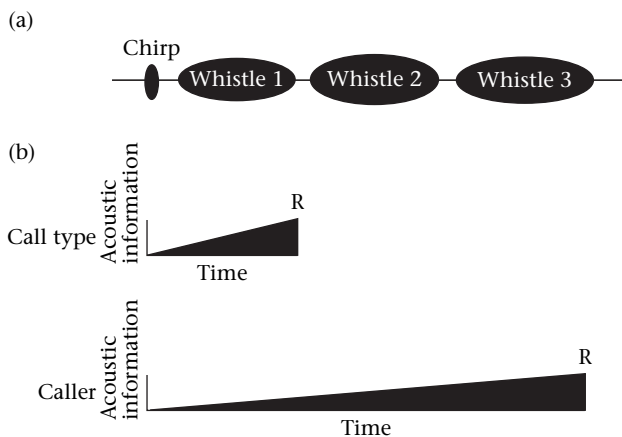
This categorization process, however, may be more general. Studies of visual object categorization by Rosch et al. (1976) suggest that all objects consist of a basic-level category that refers to the general category in which it belongs (e.g. animal) and a subordinate-level category (e.g. bird or penguin) for more specific information, including colour, texture, size and shape. Consistent with this level of analysis, our results suggest that the basic-level categorical information (call-type recognition) is processed separately from, but in parallel with, the subordinate-level category (caller recognition). The similarities between species, sensory modalities and level of processing in the categorization process suggest that a basic function of recognition and categorization systems may be to parse the basic-level and subordinate-level information because it is an adaptive strategy for processing the relevant information quickly.

Because of the importance of communicating with conspecifics, selection is likely to favour neural systems with mechanisms that enable rapid recognition of species-specific signals (e.g. Doupe & Kuhl 1999). The similarities between visual and auditory processing systems suggest that dividing a signal into basic and subordinate levels may be an efficient system for object recognition. At present, however, we lack the necessary data to determine whether these results represent real or superficial parallels between modalities. To address this possibility, it is necessary to examine the neurophysiological mechanisms that underlie object recognition and categorization. Whereas data on this topic exist for visual objects (Perrett et al. 1985; Gross 1992; O'Scalaidhe et al.1997; Logothetis 2000; Freedman et al. 2001), there is a paucity of comparable data on audition. The antiphonal calling assay should provide the centerpiece for neuroethological investigations of vocal signal recognition. Because the behaviour is both natural, repetitive, and elicited under captive conditions, it is likely that neural pathways underlying call recognition will be well developed and show specificity towards the key features of the signal that are important for recognition (e.g. Wang & Kadia 2001; Sigala & Logothetis 2002). Recording neural activity while the animal engages in the antiphonal calling is likely to yield significant data on the mechanisms that underlie vocal signal recognition. Such a study would not only provide insights into cross-modality similarities in signal processing, but also address whether the recognition process involves a serial or parallel process.

Figure 4. Diagram of parallel processing mechanisms for vocal signal recognition system. (a) Schematic drawing of the amplitude waveform of a CLC with syllable types labelled. (b) The relationships between acoustic information and time necessary for recognizing 'caller' and 'call type'. 'R' denotes the point at which recognition of that component of the signal occurs.

Acknowledgments

References

Bee, M. A. & Gerhardt, H. C. 2002. Individual voice recognition in a territorial frog (*Rana catesbeiana*). *Proceedings of the Royal Society of London, Series B*, **269**, 1443–1448.

Beeman, K. 2001. *SIGNAL User's Guide*. Belmont, Massachusetts: Engineering Design.

Biben, M. 1993. Recognition of order effects in squirrel monkey antiphonal call sequences. *American Journal of Primatology*, **29**, 109–124.

Cleveland, J. & Snowdon, C. T. 1982. The complex vocal repertoire of the adult cotton-top tamarin, *Saguinus oedipus oedipus*. *Zeitschrift für Tierpsychologie*, **58**, 231–270.

Costa, A., Mahon, B. Z., Savova, V. & Caramazza, A. 2003. Level of categorization effect: a novel effect in the picture–word interference paradigm. *Language & Cognitive Processes*, **18**, 204–233.

Dooling, R. J., Brown, S. D., Klump, G. M. & Okanoya, K. 1992. Auditory perception of conspecific and heterospecific vocalizations in birds: evidence or special processes. *Journal of Comparative Psychology*, **106**, 20–28.

Doupe, A. & Kuhl, P. 1999. Birdsong and human speech: common themes and mechanisms. *Annual Review of Neuroscience*, **22**, 567–631.

Endler, J. 1993. Some general comments on the evolution and design of animal communication systems. *Proceedings of the Royal Society of London, Series B*, **340**, 215–225.

Freedman, D. J., Anderson, K. C. & Miller, E. K. 2001. Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, **291**, 312–316.

Gerhardt, H. C. 1991. Female mate choice in treefrogs: static and dynamic acoustic criteria. *Animal Behaviour*, **42**, 615–636.

Gerhardt, H. C. 1992. Multiple messages in acoustic signals. *Seminars in the Neurosciences*, **4**, 391–400.

Ghazanfar, A. A., Flombaum, J. I., Miller, C. T. & Hauser, M. D. 2001. The units of perception in cotton-top tamarin (*Saguinus oedipus*) long calls. *Journal of Comparative Physiology A*, **187**, 27–35.

Ghazanfar, A. A., Smith-Rohrberg, D., Pollen, A. & Hauser, M. D. 2002. Temporal cues in the antiphonal calling behaviour of cotton-top tamarins. *Animal Behaviour*, **64**, 427–438.

Gross, C. G. 1992. Representation of visual stimuli in inferior temporal cortex. *Philosophical Transactions of the Royal Society of London*, **335**, 3–10.

Hauser, M. D. 1998. Orienting asymmetries in rhesus monkeys: the effect of time-domain changes on acoustic perception. *Animal Behaviour*, **56**, 41–47.

Liu, J., Harrison, A. & Kanwisher, N. 2002. Stages of processing in face perception: an MEG study. *Nature Neuroscience*, **5**, 910–916.

Logothetis, N. K. 2000. Object recognition: holistic representations in the monkey brain. *Spatial Vision*, **13**, 165–178.

MacLeod, C. M. 1991. Half a century of research on the Stroop effect: an integrative review. *Psychological Bulletin*, **109**, 163–203.

Miller, C. T. & Ghazanfar, A. A. 2002. Meaningful acoustic units in nonhuman primate vocal behavior. In: *The Cognitive Animal*

(Ed. by M. Bekoff, C. Allen & G. Burghardt), pp. 265–274. Cambridge, Massachusetts: MIT Press.

Miller, C. T. & Hauser, M. D. 2004. Multiple acoustic features underlie vocal signal recognition in tamarins: antiphonal calling experiments. *Journal of Comparative Physiology A*, **190**, 7–19.

Miller, C. T., Dibble, E. & Hauser, M. D. 2001a. Amodal completion of acoustic signals by a nonhuman primate. *Nature Neuroscience*, **4**, 783–784.

Miller, C. T., Miller, J., Costa, R. G. D. & Hauser, M. D. 2001b. Selective phonotaxis by cotton-top tamarins (*Saguinus oedipus*). *Behaviour*, **138**, 811–826.

Miller, C. T., Flusberg, S. & Hauser, M. D. 2003a. Interruptibility of cotton-top tamarin long calls: implications for vocal control. *Journal of Experimental Biology*, **206**, 2629–2639.

Miller, C. T., Weiss, D. J. & Hauser, M. D. 2003b. Mechanisms of acoustic perception in cotton-top tamarins. In: *Primate Audition: Behavior and Neurobiology* (Ed. by A. A. Ghazanfar), pp. 43–60. Boca Raton, Florida: CRC Press.

Miller, C. T., Scarl, J. S. & Hauser, M. D. 2004. Sensory biases underlie sex differences in tamarin long call structure. *Animal Behaviour*, **68**, 713–720.

Narins, P. M. & Capranica, R. R. 1978. Communicative significance of the two-note call of the treefrog *Eleutherodactylus coqui*. *Journal of Comparative Physiology A*, **127**, 1–9.

Nelson, D. A. 1988. Feature weighting in species song recognition by the field sparrow (*Spizella pusilla*). *Behaviour*, **106**, 158–182.

Nelson, D. A. & Marler, P. 1989. Categorical perception of a natural stimulus continuum: birdsong. *Science*, **244**, 976–978.

O'Scalaidhe, S. P., Wilson, F. A. W. & Goldman-Rakic, P. S. 1997. Areal segregation of face-processing neurons in prefrontal cortex. *Science*, **278**, 1135–1138.

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D. & Jeeves, M. A. 1985. Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London, Series B*, **223**, 293–317.

Perrett, D. I., Mistlin, A. J., Chitty, A. J., Smith, P. A., Potter, D. D., Broennimann, R. & Haries, M. 1988. Specialized face processing and hemispheric asymmetry in man and monkey: evidence from single unit and reaction time studies. *Behavioural Brain Research*, **29**, 245–258.

Rosch, E., Mervis, C. B., Gray, W. B., Johnson, D. M. & Boyes-Braem, P. 1976. Basic objects in natural categories. *Cognitive Psychology*, **8**, 349–382.

Ryan, M. J. 1980. Female mate choice in a neotropical frog. *Science*, **209**, 523–525.

Searcy, W. A. & Marler, P. 1981. A test for responsiveness to song structure and programming in female sparrows. *Science*, **213**, 926–928.

Sigala, N. & Logothetis, N. K. 2002. Visual categorization shapes feature selectivity in primate temporal cortex. *Nature*, **415**, 318–320.

Soha, J. A. & Marler, P. 2000. A species-specific acoustic cue for selective song learning in the white-crowned sparrow. *Animal Behaviour*, **60**, 297–306.

Soltis, J., Berhards, D., Donkin, H. & Newman, J. D. 2002. The squirrel monkey chuck call: vocal response to playback chucks based on acoustic structure and affiliative relationship with the caller. *American Journal of Primatology*, **57**, 119–130.

Sugase, Y., Shigeru, Y., Ueno, S. & Kawano, K. 1999. Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, **400**, 869–873.

Wang, X. & Kadia, S. C. 2001. Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. *Journal of Neurophysiology*, **86**, 2616–2620.

Weiss, D. J. & Hauser, M. D. 2002. Perception of harmonics in the combination long call of cotton-top tamarins (*Saguinus oedipus*). *Animal Behaviour*, **64**, 415–426.

Weiss, D. J., Garibaldi, B. T. & Hauser, M. D. 2001. The production and perception of long calls by cotton-top tamarins (*Saguinus oedipus*): acoustic analyses and playback experiments. *Journal of Comparative Psychology*, **11**, 258–271.

Williams, M. A., Moss, S. A. & Bradshaw, J. L. 2004. A unique look at face processing: the impact of masked faces on the processing of facial features. *Cognition*, **91**, 155–172.

Windfelder, T. L. 2001. Interspecific communication in mixed-species groups of tamarins: evidence from playback experiments. *Animal Behaviour*, **61**, 1193–1201.