ORIGINAL PAPER

Asif A. Ghazanfar · Jonathan I. Flombaum
Cory T. Miller · Marc D. Hauser

The units of perception in the antiphonal calling behavior of cotton-top tamarins (*Saguinus oedipus*): playback experiments with long calls

Abstract We investigated how the acoustic structure of the cotton-top tamarin monkey's (*Saguinus oedipus*) combination long call relates to the antiphonal calling behavior of conspecifics. Combination long calls can function as contact calls and are produced by socially isolated individuals. Often conspecifics respond to these calls with their own long calls. Structurally, these calls are always composed of one or more 'chirps' followed by two or more 'whistles'. We compared the antiphonal calling responses to playbacks of complete, naturally produced long calls versus single whistles or single chirps. Subjects responded significantly more to whole calls than to either syllable-type alone. Thus, our data suggest that, in terms of the antiphonal calling behavior of socially isolated conspecifics, the whole long call is the unit of perception.

Key words Primate communication · Contact calls · Vocal communication · Primate auditory cortex · Temporal processing

Introduction

One of the central questions in the study of animal vocal communication is how the structure of acoustic signals relates to the behavior of receivers. To definitively confirm the nature of this relationship, it is necessary to examine the physical structure of the signal, the context in which it is produced, and the behavioral responses of receivers (Bradbury and Vehrencamp 1998; Hauser 1996). Like learning any foreign communication system,

A. A. Ghazanfar (✉) · J. I. Flombaum · C. T. Miller
M. D. Hauser
Primate Cognitive Neuroscience Laboratory,
Department of Psychology, Harvard University,
Cambridge, MA 02138, USA
E-mail: aghazanf@wjh.harvard.edu
Tel.: +1-617-495-9834
Fax: +1-617-495-3728

one of the first problems ethologists confront is the difficulty of parsing acoustic signals into the relevant units of analysis – that is, units which are meaningful in terms of eliciting specific behaviors from the intended receivers (Chomsky and Halle 1968; Green and Marler 1979).

Ethologists studying avian (Emlen 1972; Nelson 1988; Nelson and Marler 1989; Searcy and Marler 1981; Searcy et al. 1999; Shiovitz 1975) and anuran (Gerhardt 1981, 1991; Gerhardt and Schul 1999; Narins and Capranica 1978; Ryan 1980; Ryan and Rand 1990) vocal communication have tackled this problem directly by manipulating features of conspecific signals which are hypothesized to be relevant and then testing subjects with playbacks of these manipulated signals. Emlen (1972) found, for instance, that for indigo buntings (*Passerina cyanea*) the order of syllables within a song were not important for species recognition during agonistic encounters with other males, suggesting that individual syllables are meaningful units for this species and this particular behavioral context. In contrast, Searcy and Marler (1981) tested female song sparrow (*Melospiza melodia*) responsiveness to conspecific male songs and found that females preferred the conspecific temporal patterns of syllables given by males when compared to the temporal patterns of heterospecific male songs. In this case, therefore, the order of elements (thus, the entire song) appears to be a meaningful acoustic unit for eliciting female responses. Male song sparrows also appear to categorize the entire song as a meaningful unit during territorial encounters with other males (Searcy et al. 1999).

Narins and Capranica (1978) analyzed the two-note advertisement call of the tree frog, *Eleutherodactylus coqui*. Although males usually produce both notes together, behavioral observations revealed that when a calling male is approached by an intruding male, he will change his calling pattern from the two-note call to repeating only the first note. Thus, these authors hypothesized that each note within the call conveys a different signal to receivers: the first, shorter note serving as a

territorial signal to nearby males and the second, longer note as a signal to attract females. By conducting playbacks of each note in isolation and measuring the antiphonal calls of nearby males, they indeed found that playbacks of the first note alone elicited antiphonal calls from males, while the second note alone did not elicit any significant amount antiphonal calling responses. Thus, the first note appears to be the meaningful acoustic unit for male receivers. In another anuran species, the Tungara frog (*Physalaemus pustulosus*), Ryan (1980) and Ryan and Rand (1990) have shown that females prefer male calls which have a combination of two acoustically distinct units, the 'whine' and the 'chuck', as opposed to the 'whine' alone (Ryan and Rand 1990). These studies suggest that discrete acoustic units within a call can communicate different things.

Our understanding of the acoustic units within primate vocal repertoires is quite poor, especially when contrasted with other taxonomic groups. Many primate species produce loud and long bouts of vocalizations which contain repetitions of similar acoustic units and/or sequences of different sounding acoustic units. Primate long calls serve as location signals for conspecifics (as their acoustic structure suggests) and are often produced in the context of territorial encounters, mate attraction, and isolation/group cohesion (Marler 1968; Waser 1982). Examples of these multi-syllabic calls include the 'whoop-gobble' of mangabeys (*Cercocebus albigena*; Waser 1977), the 'chirrup-pant-bellow-pump-honk' call sequences of titi monkeys (*Callicebus moloch;* Robinson 1979), the 'twitters' of squirrel monkeys (*Saimiri sciureus*; Newman et al. 1978), the 'songs' of gibbons (*Hylobates agilis*; Mitani 1985; Mitani and Marler 1989), and the 'pant-hoot' of chimpanzees (*Pan troglodytes*; Clark Arcadi 1996; Mitani and Nishida 1993). These calls often have the effect of eliciting similar calls from conspecifics (Waser 1977; Wilson

et al. 2001). Nevertheless, beyond the acoustic classification of these calls and behavioral context in which they are produced, few experimental studies have examined the meaningful units of communication within these bouts of primate vocalizations (Mitani and Marler 1989; Robinson 1979).

Cotton-top tamarins (*Saguinus oedipus*) are a small, arboreal New World primate species. In response to social isolation, both wild and captive cotton-top tamarins (hereafter simply referred to as tamarins) will emit a long call (Andrew 1963; Cleveland and Snowdon 1982; Neyman 1978). Upon hearing long calls, conspecifics often respond with their own antiphonal long calls. Tamarin long calls are composed of two or more 'whistle'-like syllables each separated by a brief period of silence, and these whistles may be preceded by one or more 'chirp'-like syllables (Cleveland and Snowdon 1982; Fig. 1). In the present study, we examined the unit of perception for antiphonal calling to long calls in a captive colony of tamarins. Using a paradigm similar to Narins and Capranica (1978), we measured the antiphonal long calling responses of isolated tamarin subjects to playbacks of (1) whole long calls; (2) isolated whistles; (3) isolated chirps; (4) white noise bursts; and (5) artificial sequences of chirps. Our goal was to elucidate the perceptual mechanisms underlying call classification in this species and provide the requisite background for subsequent neurophysiological investigations.

Fig. 1 A A captive cotton-top tamarin (*Saguinus oedipus oedipus*) in the midst of a long call during social isolation. **B** Time-amplitude waveform (*y*-axis represents volts) and spectrogram (*y*-axis represents frequency in kilohertz) of a representative combination long call exemplar. Chirps and whistles are labeled. Spectrograms were generated from a 256-pt Fast Fourier Transform (Hanning window applied). *x*-axis represents time in seconds
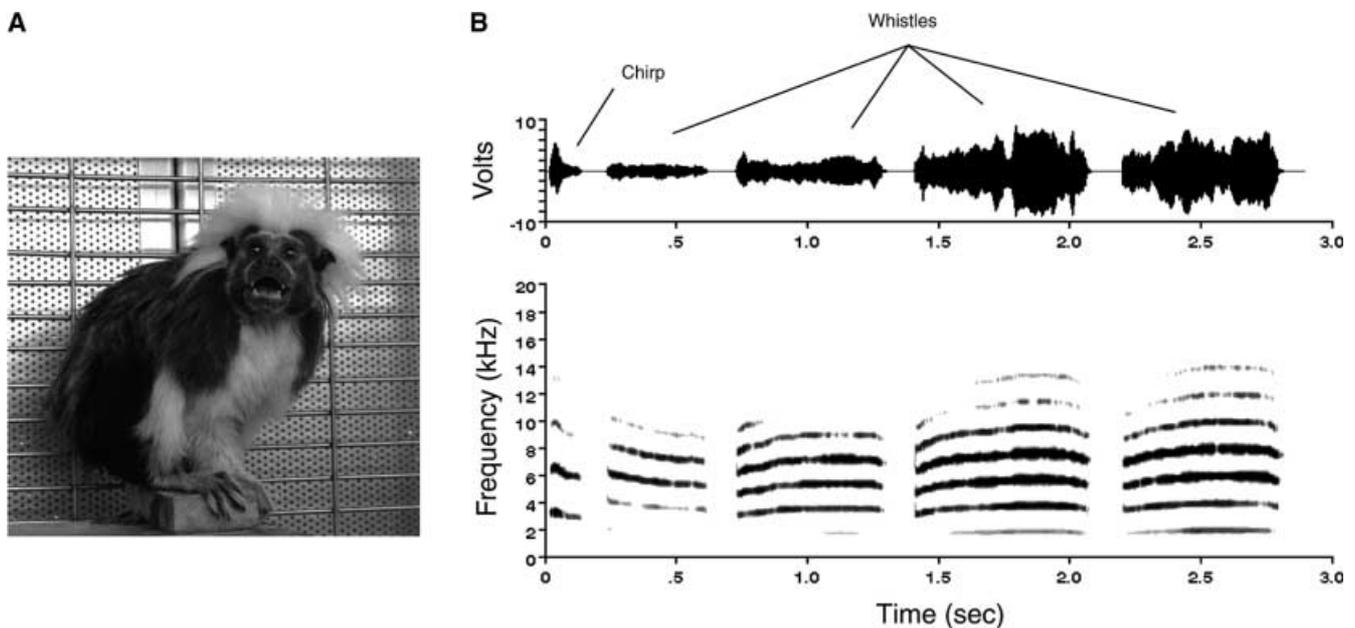
Materials and methods

The subjects were 11 tamarins. They were members of a colony consisting of 13 adults, 2 subadults, and 4 infants. All adults were born in captivity at the New England Regional Primate Center, Southborough, Mass., and then housed at the Primate Cognitive Neuroscience Laboratory at Harvard University. The subadults and infants were all born at the Primate Cognitive Neuroscience Laboratory. Subjects were housed in groups of 2 or 3 adult animals and their offspring in a 6 ft × 5 ft × 2.5 ft home cage made of steel wire and Plexiglas. In all, the colony consists of five mated pairs and three family groups. They were kept on a 12L:12D light cycle and were fed once a day in the early evening. Their diet consisted of Purina tamarin and marmoset chow, crickets, mealworms, supplemental vitamins and sunflower seeds. This diet was supplemented by food received during experiments (typically Noyes banana and Nutrasweet pellets, fruit, occasionally Froot Loops and marshmallows). Subjects had ad libitum access to water.

Apparatus

The test cage used for the playback experiments was housed in an acoustic chamber (Industrial Acoustics Company, Model 400-A). During experiments, the tamarins were placed in a wire and Plexiglas test cage (45 cm × 45 cm × 20 cm) with a cage wire floor. A thin black cotton sheet hung behind the cage. Behind this sheet, an Alesis Monitor One speaker (frequency range 45–18000 Hz ± 3dB) was mounted on a shelf above the box, behind, and to the left.

A video camera was used to record and monitor the sessions. The experimenters watched the session on a monitor outside of the acoustic chamber. An Alesis RA-100 amplifier drove the speaker. The experiment was run using a Hypercard program on a Power Macintosh 7100/80 AV. Calls (sampled at 48 kHz) were played using an Audiomedia II sound card outputting to the Alesis speaker. Experimenters entered data and kept track of time elapsed between trials using a Hypercard-based log. Vocal responses during trials were recorded using a Sennheiser MKH60P48 directional microphone (frequency response, 50–20,000 Hz).

Stimuli

#### Whole calls versus isolated syllables

According to acoustic and behavioral criteria, there are three general long call types: the normal long call, the quiet long call, and the combination long call (CLC; Cleveland and Snowdon 1982; Weiss et al. 2001). CLCs consist of one or more introductory chirps followed by multiple whistles and appear to function as a 'lost' call for individuals who are socially isolated. Normal and quiet long calls

are composed only of two or more whistles, but are not preceded by chirps. They are given in different behavioral contexts than the CLCs. Upon hearing any of three long call types, animals will often respond with antiphonal long calls (Cleveland and Snowdon 1982). CLCs and their constituents were used as stimuli in the present study (Fig. 1).

Calls were recorded during spontaneous vocal behavior during social isolation using a Tascam digital audio tape recorder (DP-1) and a Sennheiser ME-60 microphone. The calls were digitized and hi-pass filtered using Sound Designer II software. In order to determine the cut-off frequency for the filter, a spectrogram (1024 FFT) was used to determine the lowest frequency of the call. After filtering the call below that frequency, all of the calls were normalized to 100%. Only calls judged to be of high recording quality were selected for use in this experiment. We screened each call by examining the spectrograms and listening to each exemplar in the playback chamber to ensure that they were free of any artifacts.

We collected two CLC exemplars from eight different adult individuals (four males and four females) from our colony, yielding a total of 16 calls. Call sequences for playback consisted of CLCs alternated with either single whistles or single chirps. Whistles and chirps were edited from whole CLCs. For whistles, we alternated between using the first and the last whistle of the calls. For chirps, there was usually only one present, but if there were two, we used the one with the highest amplitude. In the call sequence, subjects were always presented with a whole CLC and a whistle or chirp from each of the eight individuals. Two call sequences were generated so that each whole CLC exemplar was used as a stimulus and contributed a single whistle (whistle condition) or chirp (chirp condition). A call sequence contained only whole calls and single whistles or whole calls and single chirps. Since animals may habituate to vocalizing in general over the course of an experiment and because we wanted to space the exemplars from the same animal as far apart as possible, stimuli were presented in a non-random order so that we could sample responses to whole calls versus single syllables evenly (see Fig. 2 for a schematic of the call sequence structure).

Fig. 2 Structure of call sequences used in playback experiments. During an experiment, whole long calls were alternated with single syllables, white noise bursts, or sequences of chirps. There were 16 playback stimuli, 8 from exemplar set no. 1 and 8 from exemplar set no. 2 (2 exemplars each from 8 animals). Eight of the stimuli were whole calls, while the other 8 were the alternate stimuli. Playbacks of each stimuli were separated by a minimum of 30 s. Subjects always heard both whole calls and alternate stimuli (single whistle, single chirp, or a sequence of chirps) from each of the 8 individuals during the course of an experiment. Exemplars from the same individual were spaced temporally as far apart as possible in the call sequence. Call sequence no. 2 was presented in the reverse order as call sequence no. 1 (e.g., the isolated syllable preceded the whole call)
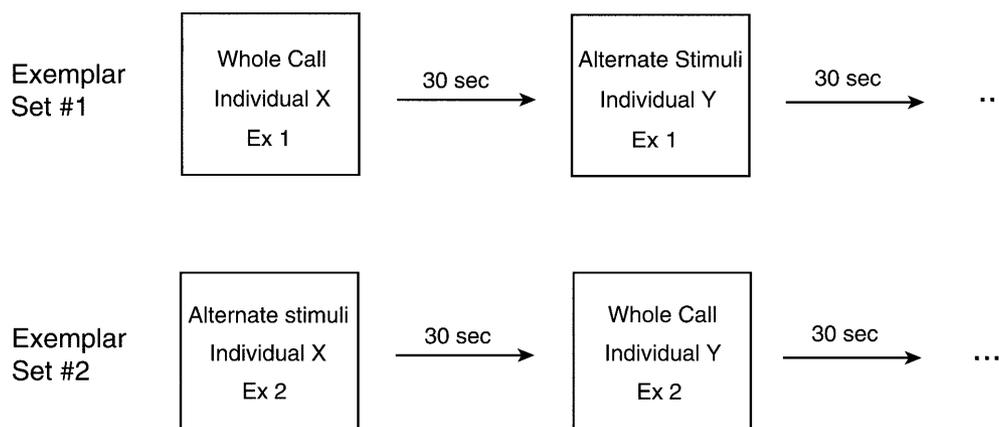
Control conditions

For the control conditions, we used a new set of 16 long call exemplars taken from the same individuals as in the previous conditions. This was done in order to minimize any habituation effects that may occur through the repeated use of the same exemplar set. In order to establish a baseline level of responses, we ran a control condition in which the call sequence consisted of whole CLCs alternating with a burst of white noise. The duration of the noise burst was 421 ms – the average length of a tamarin long call syllable (average value includes both chirps and whistles; Weiss et al. 2001). The noise burst was amplified until the root mean square value was equivalent to that of the average CLC. Responses to this white noise stimulus were then compared to responses to isolated whistles or chirps.

Assessing the strong claim that the whole call is the unit of perception required an additional control stimulus: one composed of an equal number of syllable units and of an equal duration as a whole long call. Therefore, we generated an artificial stimulus composed of sequences of chirps. To make a chirp sequence, a single chirp was edited from a whole long call and then copied using SoundForge 4.5 (Sonic Foundry). Each chirp sequence consisted of five chirps which were spaced evenly apart in time such that the duration of the chirp sequence approximated the duration of the long call from which the chirp was edited. A novel chirp sequence was made for every CLC in our stimulus set in order to control for variation in both the sounds of chirps across individuals and the durations of CLCs. Note that while a chirp sequence may be of similar duration as a CLC, it contains less energy (see Results). As in all other conditions, the call sequence played back to subjects consisted of the chirp sequence stimuli interleaved with normal CLCs.

Testing procedure

Subjects were tested between 7:00 a.m. and 10:00 a.m., prior to feeding. A subject was placed into the playback chamber and allowed to acclimate for one minute before the experiment began. Stimuli were then played back every 30 s, and long calls given by the subject following playback were scored as responses. As long calls are often given when the subjects are isolated, a period of 5–7 s of silence was required before presenting any stimuli at the start of the experiment and during the experiment. In general, subjects responded with CLCs, but occasionally produced normal long calls. We did not quantify the possible differences in long call types across the different conditions.

During a session, behaviors (orienting responses, scanning, etc.) as well as calls other than long calls (e.g., chirps) were produced by subjects. However, according to Cleveland and Snowdon (1982), the only unambiguous response to CLCs by an individual is antiphonal calling. Thus, we used antiphonal calling as our sole response measure. If a long call was produced concurrently with the onset of a stimulus presentation, it was scored as a spontaneous call, not an antiphonal response. This occurred only three times in the entire study.

There were a total of four experiments with each subject in the 'whistle' and 'chirp' conditions, as each subject was tested with both call sequences in two sessions (1 and 2). Each experimental session lasted approximately 9 min. Subjects were never tested twice on the same day. Sessions 1 and 2 were separated by at least a 1-week interval. The 'whistle' condition and the 'chirp' condition were separated by a month. Four months later, the two control conditions were only conducted 1 month apart from each other. Subjects were tested once with both call sets of each control condition. A subject was excluded from the analysis if it did not respond with an antiphonal call to at least 2 of the 16 stimuli in a given call sequence. This occurred only in one condition, in which two tamarins failed to respond.

Analysis

A long call produced by a subject within 5 s of the presentation of a stimulus was categorized as an antiphonal call, while long calls produced beyond 5 s post-stimulus and until the next stimulus presentation were categorized as 'spontaneous'. This criterion allowed us to measure both immediate vocal responses to playbacks and overall long call calling rate following a particular stimulus.

The number of antiphonal calls (within 5 s post-stimulus) or the number of calls produced in 30 s post-stimulus were entered into a repeated-measures ANOVA with session number and call sequence as between subject factors and call type (whole call versus whistle, or chirp) as a within subject factor. Statistical significance was set at $P < 0.05$. Data are reported in terms of means and standard errors.

Results

We tested 11 captive tamarins over a period of 4 months, for their responsiveness to playbacks of sequences of whole long calls alternated with single long call syllables (Fig. 2). In the whistle condition ($n = 9$), two males did not meet the criterion for responding, while in the chirp condition ($n = 11$) all subjects met the criterion. The control conditions ($n = 10$) were conducted 4 months later.

Antiphonal calling responses

In the whistle condition, tamarins had an opportunity to respond to eight whole long calls and eight single whistles. Subjects were more likely to produce antiphonal calls in response to hearing whole CLCs ($39.24 \pm 2.60\%$) than to hearing single whistles ($25.69 \pm 2.90\%$; Fig. 3A). This difference was statistically significant ($F_{(1, 8)} = 16.49$, $P = 0.004$).

In the chirp condition, subjects again responded more readily to whole long calls than to single long call chirps. Antiphonal calls were given $45.74 \pm 3.04\%$ of the time to whole calls and $25.00 \pm 2.79\%$ of the time to single chirps (Fig. 3B). This difference was statistically significant ($F_{(1,10)} = 20.8$, $P = 0.001$).

For both the whistle and chirp conditions, there were no significant main effects for antiphonal calling in either session number or the call sequence used. Thus, there was no evidence of habituation across sessions.

Frequency of long calling

In addition to measuring the immediate vocal responses of the tamarins following playback of long calls or their single syllables, we measured the overall number of long calls given from the time of the playback of one stimulus to the onset of the next stimulus in the sequence. This measurement includes those calls that occurred within 5 s of playback. This provided a measure of the overall rate of calling following whole calls or single syllables.

In terms of overall long calling rate, tamarins long called significantly more following playbacks of whole
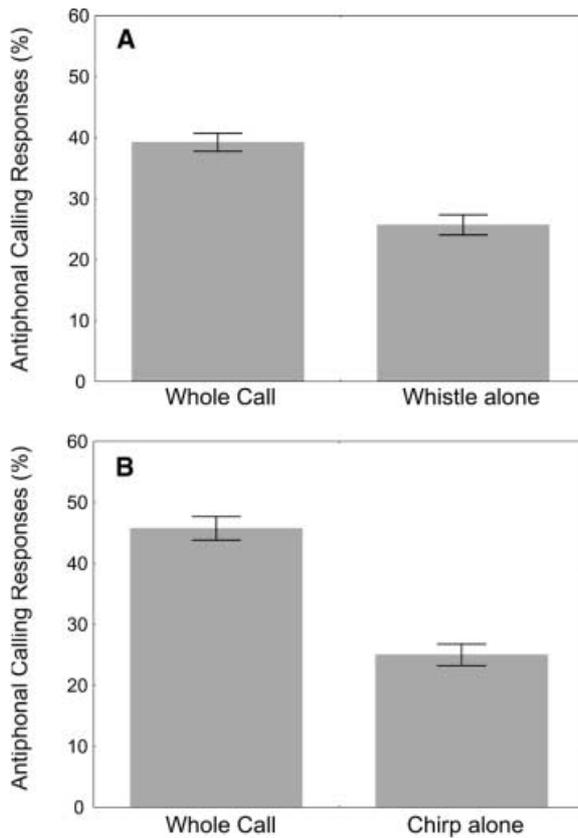
Fig. 3A, B Antiphonal calling responses to playbacks of whole long calls versus single syllables. The average number of antiphonal calls to either whole calls or single syllables was divided by 8 (the number of possible antiphonal calls in either category) to give a percentage. **A** Whole calls versus single whistles. **B** Whole calls versus single chirps. Error bars show 1 SEM

Fig. 4A, B Long calling rate following playbacks of whole long calls versus single syllables. Calling rates were calculated by dividing the average number of all long calls produced during the inter-stimulus interval following either whole calls or single syllables. This number was then divided by 30 s to give 'calls per second', represented by the *y*-axis. **A** Whole calls versus single whistles. **B** Whole calls versus single chirps. Error bars show 1 SEM

calls than following either single whistles or chirps. In the whistle condition, the average rate of long calling in response to whole calls was 0.26 ± 0.02 calls per second, while the average calling rate following single whistles was 0.21 ± 0.02 (Fig. 4A). This difference was statistically significant ($F_{(1, 8)} = 9.57$, $P = 0.015$).

In the chirps condition, the same pattern was evident. Subjects called more overall to whole calls (0.30 ± 0.02 calls per second) than to single chirps (0.26 ± 0.02 calls per second; Fig. 4B). This difference was statistically significant ($F_{(1, 10)} = 7.22$, $P = 0.023$).

For both the whistle and chirps conditions, there were no significant main effects for overall long calling rate in either session number or the call sequence used.

Single whistles versus single chirps

A comparison of the frequency of antiphonal calling in the whistles conditions with the chirps conditions revealed that there were no significant differences between responses to whole calls across the two conditions nor between responses to single whistles versus single chirps. However, when we compared call frequency to single
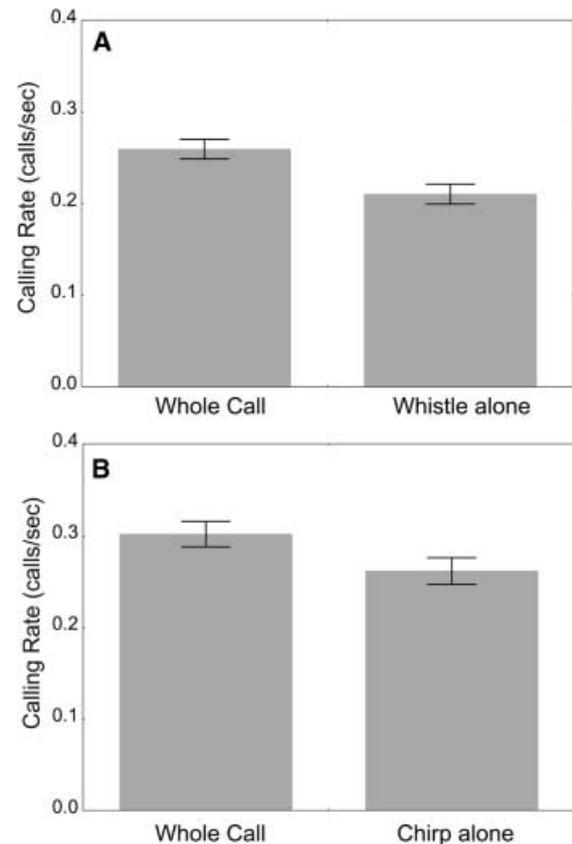
whistles versus single chirps, we found that tamarins called at a significantly higher rate following chirps than to whistles ($F_{(1, 8)} = 7.62$, $P = 0.025$).

Antiphonal calling responses to white noise

Given that our subjects did respond to single whistles and chirps (though at a reduced level relative to whole calls), it was necessary to establish a baseline response measure. We used a white noise stimulus with a duration and amplitude equal to the average duration of a long call syllable (Weiss et al. 2001). Figure 5 shows the antiphonal calling rates to noise bursts relative to whole calls. Subjects responded significantly more to whole calls ($44.38 \pm 4.58\%$) than to noise bursts ($23.75 \pm 4.43\%$; $F_{(1, 9)} = 9.60$, $P = 0.013$). A comparison of the antiphonal call responses to single whistles, single chirps and noise bursts revealed no differences between whistles versus noise or between chirps versus noise. Thus, the antiphonal calling responses to single whistles and single chirps do indeed represent the baseline level of responding by our subjects.
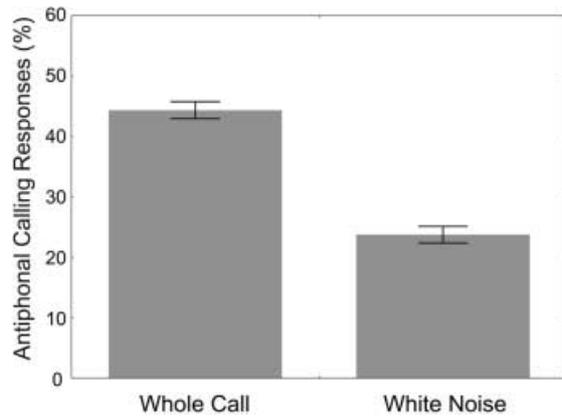
Fig. 5 Antiphonal calling responses to playbacks of whole long calls versus white noise bursts. The average number of antiphonal calls to either whole calls or white noise bursts was divided by 8 (the number of possible antiphonal calls in either category) to give a percentage. Error bars show 1 SEM

Antiphonal calling responses to sequences of chirps

As a first step in determining the features important for eliciting long calling behavior, we constructed artificial sequences of five chirps which equaled the duration of a whole long call. As evident in Fig. 6A, such stimuli, while of equal duration, have far less energy than a long call composed of both whistles and chirps. Subjects responded equally to whole long calls ($42.50 \pm 6.04\%$) and chirp sequences ($45.63 \pm 6.03\%$; $F_{(1, 9)} = 0.506$, $P = 0.495$) (Fig. 6B). These results suggest that the overall amount of energy and spectral contour are not important features in eliciting long calls in our experimental context.

Discussion

Although cotton-top tamarins can produce chirps and whistles independently, we hypothesized that the combination of both syllable types in the form of a CLC would be a more effective stimulus for eliciting antiphonal long calls by our subjects than either unit alone. This was indeed the case. Whole calls were much more effective in eliciting antiphonal long calls than either single whistles or single chirps. In terms of overall long call production rate, whole call playbacks were also able to elicit more calls than single whistles or chirps. In addition, chirps elicited higher rates of long calling than whistles. Response levels to whole calls versus sequences of chirps were equivalent, providing clues to the acoustic cues necessary in eliciting antiphonal calls by isolated tamarins. Overall, these data suggest that the most meaningful unit from the perspective of socially isolated receivers is the whole call. This study represents an important first step in deciphering the perceptually salient features of the tamarin's long call.

Whistles, chirps and long calls in the cotton-top tamarin vocal repertoire

In their extensive study of the vocal behavior of the tamarin, Cleveland and Snowdon (1982) demonstrated that the entire tamarin repertoire in captivity can be reduced to variations and combinations of two elements: the chirp and the whistle. In contrast to previous descriptions of the tamarin vocal repertoire (Andrew 1963), these authors suggest that the variants of chirps and whistles do not necessarily represent points along a single motivational continuum, rather each variant may represent a different message. For example, while single whistle-like vocalizations are often produced during close range interactions (including antiphonal calling exchanges), spectrographic analyses revealed that these whistles are not the same as the whistles that make up long calls. We provide some experimental support for the classification scheme of Cleveland and Snowdon (1982) by demonstrating that response rates to single whistles or single chirps of long calls are equal to response rates to white noise bursts. Further experiments will be necessary to determine directly if naturally produced single whistles are more effective stimuli than whistles edited from long calls. In general, our findings suggest that syllables isolated from long calls are not, by themselves, effective signals for eliciting antiphonal long calls.

The units of perception in primate vocalizations

While several studies have focused on the role of spectral (e.g., Gouzoules and Gouzoules 1989, 2000; Hauser 1993; Owren 1990a, 1990b; Rendall et al. 1996, 1998; Sugiura 1998) and temporal (e.g., A.A. Ghazanfar et al. unpublished observations; Green 1975; Hauser et al. 1998; May et al. 1988; Zoloth et al. 1979) parameters in primate vocal recognition, only a few experimental studies have explored the units of perception in multi-syllabic vocalizations (Hauser et al. 1998; Mitani and Marler 1989; Robinson 1979).

The studies of Mitani and Marler (1989) and Robinson (1979) examined the territorial calls of gibbons and titi monkeys, respectively. Both studies found that males of these species recombine different note types to create unique phrases or songs, and that particular orders of notes have a higher probability of occurring than others. Field playbacks in which the notes were arranged in a low probability sequence revealed that both gibbons and titi monkeys recognize the order of notes in conspecific vocalizations. Indeed, gibbons produced significantly more 'squeak' calls (given during inter-group encounters) upon hearing a song with experimentally rearranged notes (Mitani and Marler 1989), and titi monkeys produced significantly more 'moaning' responses (given in response to interspecies encounters) following playbacks of conspecific vocalizations with abnormal arrangements of syllables (Robinson 1979).
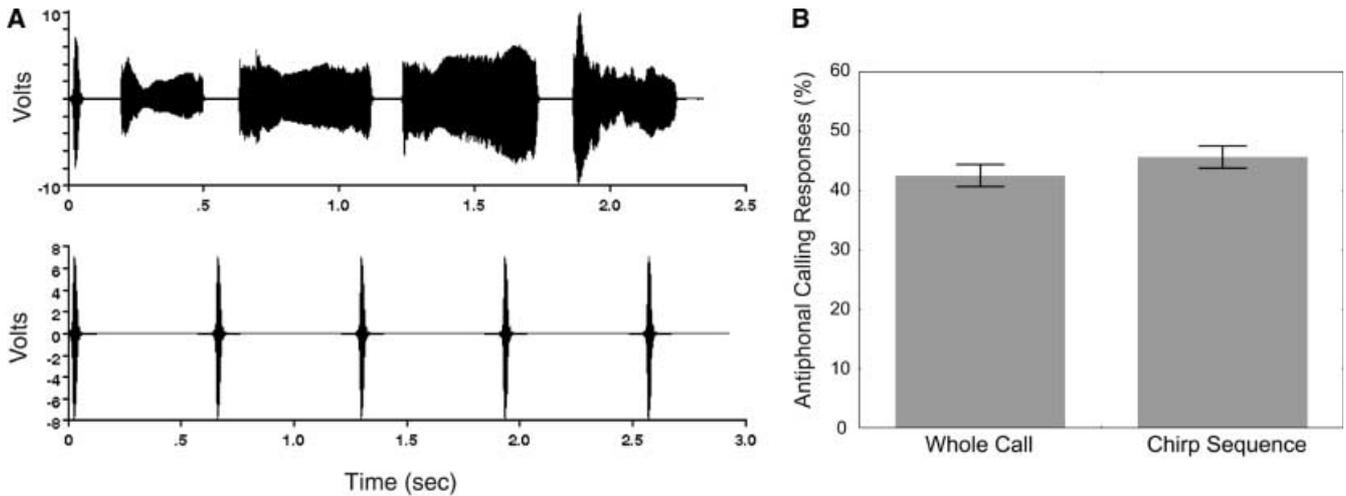
Fig. 6A, B Antiphonal calling responses to playbacks of whole long calls versus sequences of five chirps. **A** Time-amplitude waveform (*y*-axis represents volts) of a combination long call and, below it, the artificial sequence of chirps constructed from its chirp syllable. Such sequences of chirps were constructed for every exemplar in our call set. **B** The average number of antiphonal calls to either whole calls or sequences of chirps. Error bars show 1 SEM

Why are whole long calls more effective stimuli than its components?

There are several possible acoustic features that may make whole long calls more effective than its syllable components. For example, it may be that a certain amount of acoustic energy is necessary to elicit antiphonal long calling. Specifically, the greater the duration of the stimuli, the more likely an antiphonal call is produced by a receiver. The frequency bandwidth may also play a similar role: chirps tend to be higher in frequency than whistles and thus together they span a greater spectral range than either syllable alone. Another factor may be the specific temporal sequence of syllables within the tamarin long call. Cleveland and Snowdon (1982) report that combination long calls always begin with chirps followed by whistles, and reordering these syllables may be one way to further test whether whole long calls are the units of perception for this species. Our control experiment demonstrated that tamarins will call back as frequently to sequences of chirps as they do to whole long calls. Based on this result, we can eliminate syllable duration, overall energy and ordering of elements as important cues. We recognize, however, that such cues may be important for antiphonal calling in other behavioral contexts. Thus, further playback experiments using acoustically manipulated long calls and their constituents will be necessary to precisely determine the perceptual mechanisms used by tamarins for long call recognition (e.g., number of units, bandwidth, inter-syllable intervals, and overall duration). Such experiments may lead to the development of testable predictions regarding the neural basis for vocal perception

in this primate species as it has for other taxonomic groups (Ghazanfar and Hauser 1999).

Neurophysiological predictions

There are many behavioral experiments in both humans and other taxa that demonstrate that a common strategy for signal identification and localization is sensitivity to two or more spectral or temporal components (Capranica 1965; Knudsen and Konishi 1979). Detection of such sound combinations often lead to the release of species-specific behaviors. For example, the female Tungara frog is most sensitive to male advertisement calls that have both the 'whine' and 'chuck' notes; when given a choice, they will preferentially move towards individuals producing both notes than to individuals producing only one note (Ryan and Rand 1990). Similarly, tamarins call antiphonally to long calls containing both the chirp and multiple whistles more readily than to either syllable alone.

One candidate neuronal mechanism that may underlie this behavioral selectivity is the non-linear response of neurons to the combined elements of a signal as compared to responses to the individual elements alone. Combination-sensitive neurons have been described extensively in the auditory systems of bats (O'Neill and Suga 1979; Esser et al. 1997), frogs (Fuzessery and Feng 1983), and songbirds (Margoliash and Fortune 1992) using species-specific vocalizations as stimuli. Recently, neurons that exhibit combination-sensitivity have been described in species of both New World (deCharms et al. 1998) and Old World (Brosch et al. 1999) monkeys (for review see Ghazanfar and Nicolelis 2001). However, these experiments only used artificial tone sequences as stimuli.

We predict that, like other organisms, primates with long, multisyllabic vocalizations will have neurons which are combination-sensitive, particularly when the unit of perception is the whole sequence, not the individual syllables. In the present case, we predict that the

central auditory system of tamarins should contain many neurons which are combination-sensitive to both temporal and spectral components of the chirp-whistle sequence. Indeed, there is some evidence for this at the neural level in the closely related species, the common marmoset (*Callithrix jacchus*). In this species, time-reversing 'twitter' calls used as stimuli dramatically changed neural activity patterns in the primary auditory cortex (Wang et al. 1995). Recently, Bieser (1998) has shown that neurons in squirrel monkey auditory cortex respond more faithfully to natural conspecific 'twitter' calls than to synthetic versions which matched the spectral contour. The primary difference between the two stimuli was the presence of strong amplitude modulation in the natural calls which formed syllable-like elements. Thus, we predict that changing call durations, increasing or decreasing the intersyllable interval, or altering the harmonic structure of the long call may not only result in a decrease in the number of antiphonal calls by subjects, but also may result in a decrease in the responsiveness of a subset of auditory neurons to such stimuli.

References

Andrew RJ (1963) Origins and evolution of the calls and facial expressions of the primates. Behaviour 20: 1–108

Bieser A (1998) Processing of twitter-call fundamental frequencies in insula and auditory cortex of squirrel monkeys. Exp Brain Res 122: 139–148

Bradbury JW, Vehrencamp SL (1998) Principles of animal communication. Sinauer, Sunderland, MA

Brosch M, Schulz A, Scheich H (1999) Processing of sound sequences in macaque auditory cortex: response enhancement. J Neurophysiol 82: 1542–1559

Capranica R (1965) The evoked vocal response of the bullfrog. Res Monogr 33. MIT Press, Cambridge, MA

Chomsky N, Halle M (1968) The sound patterns of English. MIT Press, Cambridge, MA

Clark Arcadi A (1996) Phrase structure of wild chimpanzee pant hoots: patterns of production and interpopulation variability. Am J Primatol 29: 159–178

Cleveland J, Snowdon CT (1982) The complex vocal repertoire of the adult cotton-top tamarin, *Saguinus oedipus oedipus*. Z Tierpsychol 58: 231–270

deCharms RC, Blake DT, Merzenich MM (1998) Optimizing sound features for cortical neurons. Science 280: 1439–1443

Emlen ST (1972) An experimental analysis of the parameters of bird song eliciting species recognition. Behaviour 41: 130–171

Esser KH, Condon CJ, Kanwal J, Suga N (1997) Syntax processing by auditory cortical neurons in the FM-FM area of the mustached bat *Pteronotus parnellii*. Proc Natl Acad Sci USA 94: 14019–140124

Fuzessery ZM, Feng AS (1983) Mating call selectivity in the thalamus and midbrain of the leopard frog (*Rana p. pipiens*): single and multiunit analyses. J Comp Physiol A 150: 333–344

Gerhardt HC (1981) Mating call recognition in the barking treefrog (*Hyla gratiosa*): responses to synthetic mating calls and comparisons with the green treefrogs (*Hyla cinerea*). J Comp Physiol A 144: 17–25

Gerhardt HC (1991) Female mate choice in treefrogs: static and dynamic acoustic criteria. Anim Behav 42: 615–635

Gerhardt HC, Schul J (1999) A quantitative analysis of behavioral selectivity for pulse rise-time in the gray treefrog, *Hyla versicolor*. J Comp Physiol A 185: 33–40

Ghazanfar AA, Hauser MD (1999) The neuroethology of primate vocal communication: substrates for the evolution of speech. Trends Cognit Sci 3: 377–384

Ghazanfar AA, Nicolelis MAL (2001) The structure and function of dynamic cortical and thalamic receptive fields. Cerebral Cortex (in press)

Gouzoules H, Gouzoules S (1989) Design features and developmental modifications of pigtail macaque, *Macaca nemestrina*, agonistic screams. Anim Behav 37: 383–401

Gouzoules H, Gouzoules S (2000) Agonistic screams differ among four species of macaques: the significance of motivation-structure rules. Anim Behav 59: 501–512

Green S (1975) Variation of vocal pattern with social situation in the Japanese monkey (*Macaca fuscata*): a field study. In: Rosenblum LA (ed) Primate behavior, vol 4. Academic Press, New York, pp 1–102

Green S, Marler P (1979) The analysis of animal communication. In: Marler P, Vandenbergh J (eds) Social behavior and communication. Handbook of behavioral neurobiology, vol 3. Plenum Press, New York, pp 73–158

Hauser MD (1993) The evolution of nonhuman primate vocalizations: effects of phylogeny, body weight, and social context. Am Nat 142: 528–542

Hauser MD (1996) The evolution of communication. MIT Press, Cambridge, pp 92–101

Hauser MD, Agnetta B, Perez C (1998) Orienting asymmetries in rhesus monkeys: effect of time-domain changes on acoustic perception. Anim Behav 56: 41–47

Knudsen EI, Konishi M (1979) Mechanisms fo sound localization in the barn owl (*Tyto alba*). J Comp Physiol 133: 13–21

Margoliash D, Fortune ES (1992) Temporal and harmonic combination-sensitive neurons in the zebra finch's HVc. J Neurosci 12: 4309–4326

Marler P (1968) Aggregation and dispersal: two functions in primate communication. In: Jay PC (ed) Primates: studies in adaptation and variability. Holt, Rinehart, and Winston, New York, pp 420–438

May B, Moody DB, Stebbins WC (1988) The significant features of Japanese macaque coo sounds: a psychophysical study. Anim Behav 36: 1432–1444

Mitani J, Nishida T (1993) Contexts and social correlates of long distance calling by male chimpanzees. Anim Behav 45: 735–746

Mitani JC (1985) Responses of gibbons (*Hylobates muelleri*) to self, neighbor, and stranger duets. Int J Primatol 6: 193–200

Mitani JC, Marler P (1989) A phonological analysis of male gibbon singing behavior. Behaviour 109: 20–45

Narins PM, Capranica RR (1978) Communicative significance of the two-note call of the treefrog *Eleutherodactylus coqui*. J Comp Physiol A 127: 1–9

Nelson DA (1988) Feature weighting in species song recognition by the field sparrow (*Spizella pusilla*). Behaviour 88: 145–189

Nelson DA, Marler P (1989) Categorical perception of a natural stimulus continuum: birdsong. Science 244: 976–978

Newman JD, Lieblich A, Talmage-Riggs G, Symmes D (1978) Syllable classification and sequencing in twitter calls of squirrel monkeys (*Saimiri sciureus*). Z. Tierpsychol 47: 77–88

Neyman PF (1978) Aspects of the ecology and social organization of free-ranging cotton-top tamarins (*Saguinus oedipus*) and the conservation status of the species. In: Kleiman DG (ed) The biology and conservation of the Callitrichidae. Smithsonian Institute Press, Washington

O'Neill WE, Suga N (1979) Target range-sensitive neurons in the auditory cortex of the mustache bat. Science 251: 565–568

Owren MJ (1990a) Acoustic classification of alarm calls by vervet monkeys (*Cercopithecus aethiops*) and humans (*Homo sapiens*). I. Natural calls. J Comp Psychol 104: 20–28

Owren MJ (1990b) Acoustic classification of alarm calls by vervet monkeys (*Cercopithecus aethiops*) and humans (*Homo sapiens*). II. Synthetic calls. J Comp Psychol 104: 29–40

Rendall D, Rodman PS, Emond RE (1996) Vocal recognition of individuals and kin in free-ranging rhesus monkeys. Anim Behav 51: 1007–1015

Rendall D, Owren MJ, Rodman PS (1998) The role of vocal tract filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations. J Acoust Soc Am 103: 602–614

Robinson JG (1979) An analysis of the organization of vocal communication in the titi monkey *Callicebus moloch*. Z Tierpsychol 49: 381–405

Ryan MJ (1980) Female mate choice in a neotropical frog. Science 209: 523–525

Ryan MJ, Rand AS (1990) The sensory basis of sexual selection for complex calls in the Tungara frog, *Physalaemus pustulosus* (sexual selection for sensory exploitation). Evolution 44: 305–314

Searcy WA, Marler P (1981) A test for responsiveness to song structure and programming in female sparrows. Science 213: 926–928

Searcy WA, Nowicki S, Peters S (1999) Song types as fundamental units in vocal repertoires. Anim Behav 58: 37–44

Shiovitz KA (1975) The process of species-specific song recognition by the indigo bunting, *Passerina cyanea*, and its relationship to the organization of avian acoustical behavior. Behaviour 55: 128–179

Sugiura H (1998) Matching of acoustic features during the vocal exchange of coo calls by Japanese macaques. Anim Behav 55: 673–687

Wang X, Merzenich MM, Beitel R, Schreiner CE (1995) Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. J Neurophysiol 74: 2685–2706

Waser PM (1977) Individual recognition, intragroup cohesion, and intergroup spacing: evidence from sound playback to forest monkeys. Behaviour 60: 28–74

Waser PM (1982) The evolution of male loud calls among mangabeys and baboons. In: Snowdon CT, Brown CH, Peterson MR (eds) Primate communication. Cambridge University Press, New York

Weiss DJ, Garibaldi BT, Hauser MD (2001) The production and perception of long calls in cotton-top tamarins (*Saguinus oedipus*). J Comp Psychol (in press)

Wilson ML, Hauser MD, Wrangham RW (2001) Does participation in cooperative intergroup conflict depend on numerical assessment, range location or rank for wild chimpanzees? Anim Behav (in press)

Zoloth SR, Petersen MR, Beecher MD, Green S, Marler P, Moody DB et al (1979) Species-specific perceptual processing of vocal sounds by monkeys. Science 204: 870–873