Pattern recognition mediates flexible timing of vocalizations in nonhuman primates: experiments with cottontop tamarins

ELISABETTA VERSACE*†, **ANSGAR D. ENDRESS**† & **MARC D. HAUSER**†‡§

*Department of Psychology, University of Trieste
†Department of Psychology, Harvard University
‡Department of Biological Anthropology, Harvard University
§Department of Organismic and Evolutionary Biology, Harvard University

To maximize transmission in noisy environments, vocalizing animals have evolved capacities to avoid the masking effects of biotic and abiotic sound sources, such as changing the structure and timing of acoustic signals. Here we explore this problem from a new angle, asking whether animals can extract predictive acoustic cues from an intermittently noisy environment and use this information to guide the timing of their vocalizations. In a series of experiments, we presented cottontop tamarins, *Saguinus oedipus*, with loud noise interrupted by speech syllables and silence. The silences were either long enough to accommodate the full length of their species-specific contact calls or too short. The task, in brief, was for tamarins to pick out the acoustic cues predicting long versus short gaps of silence, and to time their vocalizations to follow cues associated with long gaps. Consistently, and across all subjects tested, tamarins were able to extract the acoustic cues predicting the long gaps, using them to call in the long gaps and consequently, avoiding both the energetic and masking costs of calling in noise. We discuss these findings and the potential applications of this technique to broader questions of pattern recognition and the flexible use of vocal signals in a noisy environment.

© 2008 The Association for the Study of Animal Behaviour. Published by Elsevier Ltd. All rights reserved.

Keywords: auditory learning; cottontop tamarin; masking; pattern recognition; primate; *Saguinus oedipus*; vocal timing

In noisy environments, vocalizations may be easily masked. Selection should therefore favour mechanisms that enable individuals to avoid masking, such as modifying the timing, amplitude or spectral structure of the call. Crucially, there should be both perceptual mechanisms that enable extraction of patterned information from the environment as well as production mechanisms that enable control over the timing and structure of the call. Here we present experiments that explore the capacity of animals to first detect and then use predictable

acoustic patterns to mediate the precise timing of call production, thereby avoiding the costs of masking.

It is well known that many species adjust their vocalizations in different ways to ambient noise. For example, a common response, first noted by Lombard (1911), is to increase call amplitude, a ubiquitous capacity shown in several birds and mammals (Potash 1972; Sinnott et al. 1975; Nonaka et al. 1997; Cynx et al. 1998; Brumm & Todt 2002; Brumm et al. 2004; Scheifele et al. 2005; Egnor & Hauser 2006). In addition to increasing call amplitude, animals often increase the number of vocalizations (e.g. king penguins: Lengagne et al. 1999) as well as their duration (e.g. whales: Foote et al. 2004); a more limited number of species can control the spectral content of their vocalizations to fit an open frequency channel (e.g. bats: Smotherman et al. 2003).

While these mechanisms allow animals to cope with continuous ambient noise, different, and less energetically

Correspondence: E. Versace, Department of Psychology, University of Trieste, 34134 Trieste, Italy (email: eversace@units.it). A. Endress is at the Departments of Psychology and Linguistics, Harvard University, 33 Kirkland Street, Cambridge, MA 02138, U.S.A. M. D. Hauser is at the Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138, U.S.A.

demanding strategies seem more suitable in the context of discontinuous noise. For instance, a signaller can simply wait for a noise offset (i.e. the start of silence), without expending energy to increase the amplitude or number of vocalizations. Indeed, several species are capable of modifying the timing of their calls in order to fit within the periods of silence intervening between periods of noise. For example, several anuran and avian species shift the timing of their calls to avoid mutual acoustic interference (Popp 1989; Garcia-Rutledge & Narins 2001), often using noise offsets as the critical cue for timing call onset. Similar adjustments in call timing have been shown in captive cottontop tamarins, *Saguinus oedipus*, exposed to playbacks of cyclic and predictable background noise (Egnor et al. 2007).

Although virtually all species tested to date appear to use noise offset as a cue to avoid masking, there are conditions in which this strategy will fail. In particular, if the silent intervals are shorter than the length of the desired vocalization, initiating a call at the onset of the silence will result in vocalizations that are at least partially masked. For example, in environments where birds and insects engage in morning chrouses, there will often be narrow windows of opportunity for calling, windows defined by both temporal and spectral dimensions (Green & Marler 1979). In this situation, it is necessary to identify not only silent intervals, but also, to predict whether these will be sufficiently long for a vocalization, either with or without modification. If the length of the silent interval can be predicted by reliable cues, an efficient strategy is to extract this information, using it to adjust the timing of call production. Here, we ask whether cottontop tamarins have this kind of capacity, one that would bring significant advantages in a noisy environment. Specifically, we designed an experiment to explore whether cottontop tamarins can detect acoustic cues that predict long, as opposed to short, intervals of silence and, if so, whether they use this information to preferentially call more often in the longer intervals that maximize signal transmission.

To explore this question, we took advantage of the fact that tamarins, like other primates (Marler 1968; Waser 1977, 1982; Newman et al. 1978; Robinson 1979; Mitani 1985; Mitani & Marler 1989; Mitani & Nishida 1993; Clark Arcadi 1996), produce a loud contact call when they are visually isolated from other group members. Their primary vocalization in this context is the combination long call (Cleveland & Snowdon 1982; Weiss et al. 2001; Miller et al. 2003). In the experiments reported below, we exposed tamarins to sequences of loud masking noise, speech syllables and silent intervals. The silent intervals were either long enough or not long enough to fit a combination long call, and the duration of the silences was reliably predicted by the particular syllables preceding them. Hence, if tamarins can use the acoustic cues associated with different syllables to predict the probability of a long or short gap, we would expect them to call more in long than in short silences. Furthermore, given the modal duration of a combination long call, subjects must generally start soon after the relevant predictive cue in order to maximize the amount of signal produced in silence.

GENERAL METHODS

Subjects

We tested eight adult cottontop tamarins (4 males: DD, RK, JM and DW; 4 females: RB, JG, LG and JK) from the Harvard University Cognitive Evolution Laboratory, Cambridge, Massachusetts, U.S.A., ranging in age from 2 to 16 years. All subjects were born in captivity and socially housed, with separate home cages for each breeding pair and their offspring. Subjects were maintained on a diet of monkey chow, fruit, seeds and mealworms, together with free access to water. Subjects voluntarily left their home cages, lured out by a piece of raisin.

Apparatus

During the experimental sessions, subjects were placed in a Plexiglas playback cage (25 × 28 × 51 cm), with a wire mesh front. The playback cage was located inside a double-walled sound-attenuating chamber (Industrial Acoustics, Bronx, NY, U.S.A.; inner dimensions: 1.75 × 1.85 × 1.95 m) with a directional microphone (ME-66, Sennheiser) positioned 10 cm from the playback cage. The microphone signal was amplified (Rane MS-1), and then digitized at a sampling rate of 22 050 Hz and a precision of 32 bit. Data acquisition and sound presentation were controlled by Audacity software (audacity.sourceforge.net); stimuli were presented using an Advent AV570 speaker. Subjects were monitored and tape-recorded with a video camera during the experimental sessions.

Stimuli: General Description

In both experiments, we exposed subjects to playbacks of cyclic and predictable alternations of white noise and silence. The basic organization of the sound streams is shown in Figs 1 and 3. White noise ranged between 800 and 10 000 Hz and had variable length (4 s, 5.5 s, 6 or 7.5 s). The noise was played at 80−85 dB sound pressure level (SPL), a level that should mask the combination long call, which is typically produced at 60 dB SPL in silence (Egnor et al. 2007). Although it is possible that combination long calls produced within such noise are detectable to some extent, based on acoustic analyses and previous studies of calling in noise (Egnor et al. 2006), we expected subjects to generally avoid such loud noise because of the rather severe effects of masking. Noise intervals were separated by silent intervals of 3 s or 1.5 s. Since the duration of the tamarin's combination long call ranges from approximately 2.5 s to 3 s, we used silent intervals of 3 s and 1.5 s, respectively. The duration of the silent intervals was always predicted by either speech syllables or the absence of syllables between the noise and the silent interval.

In each session, we played a different sound stream. The streams were semirandom sequences of trials created according to the following form:

(white noise) + (100 ms silence) + (cue)

+ (silence of 1.5 s or 3 s).

Intervals of white noise were created using Audacity, and saved in the wav file format (sampling rate: 44 100 Hz;
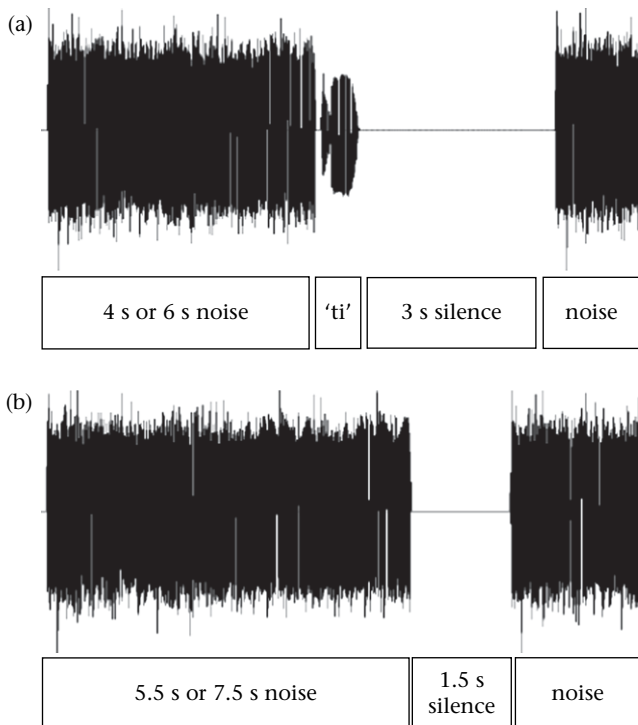
Figure 1. Organization of the sound streams in experiment 1. Streams were organized into trials of the following form: (white noise) + (100 ms silence) + (cue) + (silence of 1.5 s or 3 s). (a) A 4 or 6 s long interval of white noise was followed by the syllable 'ti', and then by a silence of 3 s. This gap is long enough to contain a combination long call. (b) A 5.5 or 7.5 s long interval of white noise was followed by a silence of 1.5 s. This silence is too short to contain a combination long call. The duration of the silent intervals thus was always predicted by either speech syllables or the absence of syllables between the noise and the silent interval.

sample width: 16 bit, mono, signed linear encoding). Silent intervals were created in the same way.

The speech syllables that were used as predictive cues were pronounced by native speakers of English, recorded using a Sennheiser ME-66 microphone connected to a Sigmatel soundcard, and saved in the wav file format (sampling rate: 44 100 Hz, sample width: 16 bit, mono, signed linear encoding) using Audacity. Trials of the aforementioned form were concatenated into sound streams.

Within each stream, trials were randomized with the constraint of not having more than two noises or silences of equal duration in a row. We then converted all sound files to the raw file format using sox (sox.sourceforge.net), concatenated them according to the randomized lists with the utility cat, and converted the concatenated file into the wav file format (sampling rate: 22 050 Hz; sample width: 16 bit, mono, signed linear encoding), again using sox. In each stream, there were equal numbers of long and short silent intervals for a total of 80 silent gaps.

Experimental Schedule and Procedure

There were two experimental conditions, each run for 8 consecutive days, with no gap between conditions. Each subject was tested once a day.

As pilot data showed that different individuals produced vocalizations at different intensities, we adjusted the amplitude of the white noise according to the amplitude of each subject's combination long call. We thus ran four monkeys (RK, LG, RB, JK) at 80 dB SPL of white noise and four monkeys (DW, JG, DD and JM) at 85 dB SPL.

We placed each subject into the playback cage inside a sound-attenuating chamber. Stimuli were played using Audacity. We also used the same software to record the monkeys' vocalizations and to convert the recordings of each session to the wav file format.

Measures

A combination long call consists of a concatenation of at least two syllables: one or more chirps (i.e. short and highly frequency-modulated syllables) followed by one to four whistles (i.e. longer and less frequency-modulated syllables); as noted above, the average length of a combination long call is 2.5−3 s (Cleveland & Snowdon 1982; Weiss et al. 2001). For each session, we located all combination long calls and measured the interval between the end of the syllable or noise and the beginning of the call.

To distinguish combination long calls from the other vocalization types that subjects produced during the experiments (e.g. alarms, vigilant calls), we counted as a combination long call any vocalization that contained at least two syllables, one of which was at least 400 ms in duration. We labelled the first syllable of at least 400 ms as the 'first whistle'. Syllables preceding the first whistle were considered part of the call when separated by no more than 350 ms either from the onset of the first whistle, or from another syllable included in the call. We also considered as part of the call all the syllables following the first whistle in which the inter-syllable interval was less than 600 ms. We analysed the recordings of the tamarins' vocalizations using Praat (version 4.2.07 for Windows, freely available from http://www.praat.org).

Calls produced during a session can in theory start during silence, during noise, or during one of the syllables. For analyses, we considered calls beginning during the syllable as beginning during noise (because both would be masked). For each session, we counted the number of calls starting in noise and the number of calls starting in long and short silences. We also distinguished between calls starting after each predictive cue, that is, after a syllable or after a noise offset.

To evaluate whether tamarins used the predictive cues to time their vocalizations, the most meaningful comparison is that between calls starting within the first half of the long gap and those starting within the short gap. This is because these two periods have precisely the same duration (1.5 s) and immediately follow sound offsets. Thus, for each subject, we contrasted the number of calls produced in the first half of the long gap with the number of calls produced in the short gap.

EXPERIMENT 1

In experiment 1, each long silent interval was preceded by a speech syllable, whereas each short silent interval directly followed the noise. We asked whether tamarins can adjust

the timing of their vocalizations according to the information conveyed by these simple cues, calling more in the first half of the long silences than in the short silence.

Methods

In experiment 1, the syllable 'ti' predicted long silences, whereas the offset of white noise predicted short silences . We recorded the 'ti' syllable from a female native speaker of English. The white noise preceding 'ti' was either 4 s or 6 s long; either duration occurred equally often. The short silent interval was preceded, equally often, by 5.5 s or 7.5 s of white noise.

The stimuli were constructed as described in Fig. 1. We ran subjects on experiment 1 for 8 consecutive days. We aborted session 8 for subject DW because of signs of distress. To show statistical significance at the individual level, we required a minimum of six calls. We therefore excluded from the analyses one subject (JG) who produced fewer than six calls beginning in silence over the course of each experiment (i.e. calls starting in silence per session: $\overline{X} = 1.87$, range 0–4).

Results

Group results

For each subject, we calculated the total number of calls started in the short silent interval and in the first part of the long silent interval, respectively. We then compared these counts. As shown in Fig. 2, subjects tended to call more in the first part of the long silent intervals than in the short silences (Wilcoxon signed-ranks test: $T = 0$, $N = 7$, $P = 0.018$).

There was no difference in reaction times for initiating calls in the short gap ($\overline{X} \pm SE = 0.87 \pm 0.06$) as opposed to the long gap ($\overline{X} \pm SE = 0.89 \pm 0.07$).

Individual results

We present the results of each subject in Table 1. For each individual we compared the number of calls starting in the short silent interval and in the first part of the long silent interval using a two-tailed binomial test.

Five of seven subjects (DW, RB, DD, JM and LG) showed a statistically significant preference for calling in the long interval compared to the short interval, suggesting that they were able to discriminate between the long and the short intervals. These subjects showed a significant preference for calling in the first half of the long silent intervals. RK and JK showed a nonsignificant tendency towards calling more in the first half of the long silent intervals. JK showed no preference for either interval type.

Discussion

The group and individual results of experiment 1 show that most subjects called more in the first half of the long gap than in the short gap. These results show that tamarins can modify the timing of their calls as a function of an arbitrary predictive cue, with the consequence that they reduce the masking effects of background noise.
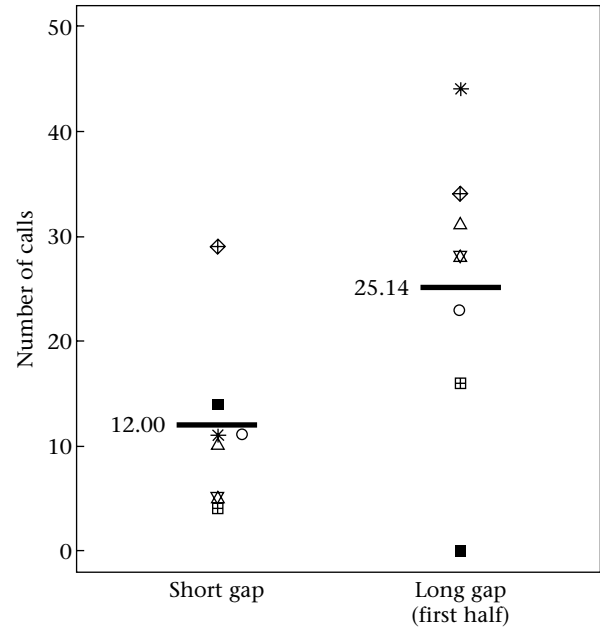

Figure 2. Results of experiment 1: number of calls starting in the short silent gap (1.5 s) and in the first half of the long gap (1.5 s). Each subject is indicated by a different symbol. The thick horizontal lines represent the population averages in the two conditions.

Results further showed that at the group level, subjects tended to call more often in the first half of the long intervals compared to the short intervals, indicating that they were not merely attending to offsets of sound.

Individual success in calling more in one of the two gaps can be accounted for by two different mechanisms: subjects either attended to and used the different predictive cues (i.e. syllable versus noise), or they responded to the different noise durations preceding the silent intervals. Experiment 2 was designed to test between these possibilities by presenting periods of noise that were always followed by speech syllables. Thus, the only predictive cue to gap length was the acoustic morphology of the particular syllables associated with both long and short silent gaps; furthermore, the periods of white noise

Table 1. Summary of the individual results in experiment 1

| Subjects | Number of calls | | P | Total no. of calls | % Calls in silence |
	Short gap	Long gap			
DW	10	31	<0.01	106	70
RB	11	23	0.011	68	87
DD	11	44	<0.01	107	81
RK	29	34	0.532	148	58
JM	6	28	<0.01	110	90
JK	14	17	0.72	92	51
LG	4	18	<0.01	59	80

Number of calls starting in the short gap and in the first half of the long gap, P level associated with a two-tailed binomial test, total number of calls and percentage of calls starting in silence are specified for each individual.

preceding the syllables were equally long in both conditions. As a result, to succeed on this task, tamarins would have to discriminate among the speech syllables and selectively initiate calls immediately following those syllables predicting the long gap.

EXPERIMENT 2

Experiment 2 further explored the limits of the relationship between pattern recognition and vocal control in tamarins. As in experiment 1, tamarins could avoid the masking of their contact calls by starting them in windows of silence of sufficient duration. In contrast to experiment 1, however, experiment 2 presented tamarins with a more difficult discrimination task, requiring a distinction between two speech syllables associated with the long gap and a third, different syllable associated with the short gap. In this context, if tamarins call more frequently in the first half of the long gaps than in the short gaps, they must be attending to the differences between syllables associated with these gaps. Moreover, if performance improves from that observed in experiment 1, then either the additional cues provide better predictors of gap duration, or subjects have learned more about the structure of these experiments (i.e. that there were predictable gaps that either were or were not sufficiently long for a combination long call), or some combination of these factors.

Methods

As illustrated in Fig. 3, the design of experiment 2 was similar to that of experiment 1 with two exceptions. First, the duration of the silent gaps was predicted by three distinct syllables. The same syllable as in experiment 1 'ti' as well as 'lu' (a new syllable) predicted the long silence, and the syllable 'ja' predicted the short silence. The syllables 'ti' and 'lu' were pronounced by a female native speaker of English, whereas the syllable 'ja' was pronounced by a male native speaker of English. Second, all noise segments were either 4 s or 6 s long; these durations occurred equally often with either silence duration. Each session lasted roughly 10 min.

Experiment 2 was run immediately after experiment 1 for 8 consecutive days. There was a 1 day delay in starting subject DD because he could not be lured out of his homecage on day 1. Because of technical problems, we could not run the last session for subjects DW, JG, DD and JM. One subject, JG, was excluded from analysis because of an insufficient number of calls starting in silence (i.e. calls starting in silence per session: $\overline{X} = 1.28$, range 1−3).

Results

Group results

Figure 4 shows the group results for experiment 2. For each subject, we calculated the total number of calls starting in the short silent interval and in the first part of the long silent interval, respectively. We then compared these counts. As a group, the tamarins called more in the first
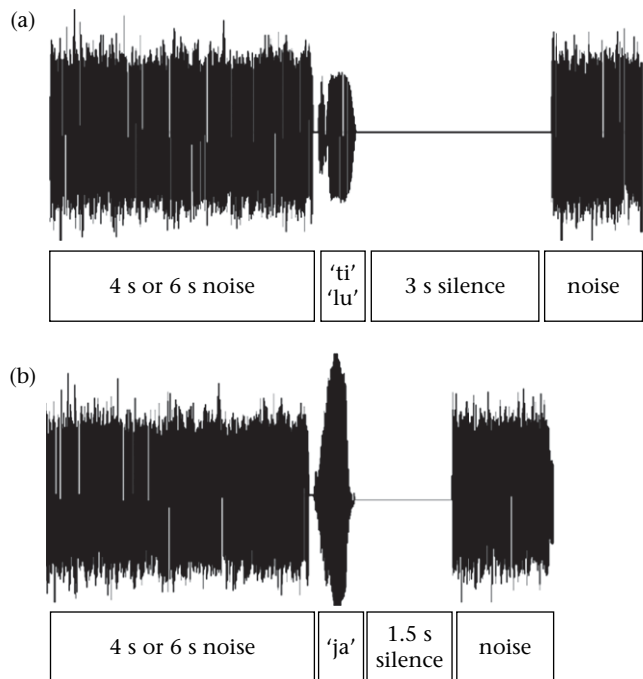

Figure 3. Organization of the sound streams in experiment 2. Streams were organized into trials of the following form: (white noise) + (100 ms silence) + (cue) + (silence of 1.5 s or 3 s). (a) A 4 or 6 s long interval of white noise was followed by either the syllable 'ti' or the syllable 'lu', and then by a silence of 3 s. This gap is long enough to contain a combination long call. (b) A 4 or 6 s long interval of white noise was followed by the syllable 'ja', and then by a silence of 1.5 s. This silence is too short to contain a combination long call. The duration of the silent intervals thus was always predicted by different speech syllables.

half of the long silence than during the short silence (Wilcoxon signed-ranks test: $T = 0$, $N = 7$, $P = 0.018$).

In experiment 2, two syllables predicted the long gap: one was familiar ('ti'), because it was presented in experiment 1, and one ('lu') was novel. It is possible that the effect of calling more in the long gap was driven by the familiar syllable, and that tamarins restricted their calls to gaps following this syllable. There was no statistically significant difference between the responses to the familiar and the novel or new syllable (Wilcoxon signed-ranks test: $T = 8$, $N = 7$, $P = 0.31$).

Reaction times associated with calls starting after each of these cues were virtually identical (familiar syllable before long gap: $\overline{X} \pm \text{SE} = 0.99 \pm 0.05$; new syllable before long gap: $\overline{X} \pm \text{SE} = 0.96 \pm 0.05$; new syllable before short gap: $\overline{X} \pm \text{SE} = 0.92 \pm 0.098$).

Interestingly, in experiment 2 all subjects but one increased the proportion of calls produced in the long gap in comparison with experiment 1 (see Fig. 4). Although this difference was not statistically significant (Wilcoxon signed-ranks test: $T = 3$, $N = 7$, $P = 0.12$; see Table 3), it was in the direction expected if subjects learned the functional problem set up by these experiments.

Individual results

All seven subjects showed a significant preference for calling in the first half of the long silent interval (Table 2),
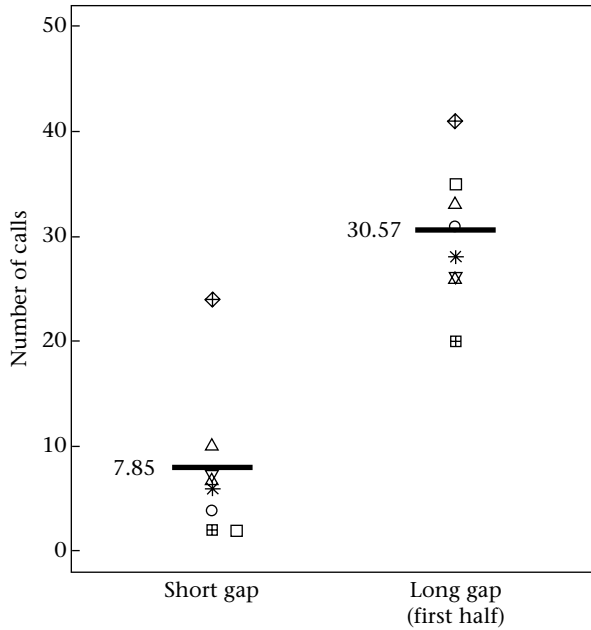
Figure 4. Results of experiment 2: number of calls starting in the short silent gap (1.5 s) and in the first half of the long gap (1.5 s). Each subject is indicated by a different symbol. The thick horizontal lines represent the population averages in the two conditions.

indicating that they were able to discriminate between the cues and their predictive values.

Although all subjects called more in the first half of the long than short gap (Table 2), there were individual differences in overall accuracy. That is, while over 50% of each subject's calls during the eight experimental sessions were produced in silence, some subjects largely restricted their calls to silence (e.g. DW, RB, JM and LG).

Discussion

The results of experiment 2 show that all subjects called significantly more after the cues predicting the long silent interval than after the cue predicting the short silent interval. Crucially, because in experiment 2 both the short and the long gap were predicted by speech syllables, the tamarins had to rely on these arbitrary cues to adjust the timing of their vocalizations and avoid masking. The

Table 3. Percentage of calls in the first half of the long gap in experiments 1 and 2

	% Long gap	
Subjects	Exp. 1	Exp. 2
DW	75.61	76.74
RB	67.65	88.57
DD	80.00	82.35
RK	53.97	63.08
JM	82.35	79.41
JK	54.84	94.59
LG	81.81	81.81

results of experiment 2 suggest that tamarins were able to discriminate between acoustical cues and to use them to locate appropriate opportunities for calling. Although not statistically significant, there is suggestive evidence that the tamarins either improved their performance as a function of learning between experiments 1 and 2, or that the cues provided in experiment 2 facilitated their performance.

GENERAL DISCUSSION

The natural world is noisy. To avoid the effects of masking, animals communicating in the auditory domain must find opportunities to call that maximize signal transmission to intended receivers. Two strategies to avoid masking are used by a wide variety of animals, namely (1) to increase the amplitude of the vocalizations in the presence of noise and (2) to wait for the noise to stop before starting to vocalize (e.g. Potash 1972; Nonaka et al. 1997; Lengagne et al. 1999; Brumm & Todt 2002; Foote et al. 2004; Egnor & Hauser 2006). A more complex strategy has been observed in certain birds that start vocalizations only after the end of specific kinds of noise (Green & Marler 1979); in this way, they can make sure that the subsequent silent period will be long enough to complete their vocalizations. Here, we ask whether cottontop tamarins have a similar ability. While nonhumans primates are generally thought to be poor vocal learners, with at best limited evidence for vocal plasticity (reviewed in Janik & Slater 1997; Seyfarth & Cheney 1999; Egnor & Hauser

Table 2. Summary of the individual results in experiment 2

	Number of calls			Number of calls			
Subjects	Short gap	Long gap	P	Familiar cue	Novel cue	Total no. of calls	% Calls in silence
DW	10	33	<0.01	5	3	98	75
RB	4	31	<0.01	17	14	83	90
DD	6	28	<0.01	11	17	78	79
RK	24	41	0.046	21	10	133	63
JM	7	27	<0.01	16	11	102	83
JK	2	35	<0.01	24	11	118	58
LG	4	18	<0.01	8	12	49	86

Number of calls starting in the short gap and in the first half of the long gap, P level associated with a two-tailed binomial test, number of calls starting in the first half of the long gap after the familiar cue and the novel cue, total number of calls and percentage of calls starting in silence are specified for each individual.

2004), several primates show at least the capacity for rapid forms of vocal modification, especially with respect to modifying the duration, timing and amplitude of their calls. For instance, Egnor et al. (2007) showed that tamarins can time their calls to fit windows of silences when exposed to regular alternations of loud noises and periods of silences, and like many other vertebrates, can increase the amplitude of their vocalizations in the face of predictable increases in background noise (Miller et al. 2003; Egnor et al. 2006, 2007). The results presented here add to this literature by showing that cottontop tamarins can use arbitrary acoustic cues to adaptively guide their vocal behaviour and avoid masking from ambient noise.

Tamarins were exposed to alternations between periods of loud noise and periods of silence that either were or were not long enough to contain a combination long call. The durations of the silences could be predicted based on auditory cues that were inserted between the noise and the silences. Specifically, in experiment 1, the speech syllable 'ti' predicted a long silence while short silences were predicted by the mere offset of white noise. In experiment 2, in contrast, the syllables 'ti' and 'lu' predicted a long silence, while the syllable 'ja' predicted a short silence. Given the costs of increasing call amplitude above the white noise, the most efficient strategy to avoid masking was the use of acoustic cues preceding the silences (i.e. the syllables), calling immediately within the first half of the long silent intervals. In experiment 1, most subjects succeeded in using the syllable 'ti' to predict the long intervals, and to time their calls accordingly. An alternative explanation for these results is, however, that tamarins were using differences in noise duration to predict the differences in the gaps. That is, the noise duration for the long gap was shorter than the noise preceding the short gap, and subjects may have used this information to guide the timing of their calls. Experiment 2 ruled out this account. In experiment 2, both the short and the long gaps were preceded by equally long periods of noise. The crucial difference was the spectral structure of the particular syllable preceding long and short gaps. All subjects were able to discriminate between these different acoustical cues and to use this information to regulate the timing of their calls, as evidenced by individual and group analyses. Hence, tamarins can use rather arbitrary cues to dynamically modulate the timing of their vocalizations.

The results presented here have potentially interesting implications for future studies of both vocal control and pattern recognition, and raise the possibility that predictive calling in intermittent noise may provide a general bioassay for exploring such problems in a variety of species. Methodologically, and as briefly reviewed in the Introduction, several studies have presented animals with intermittent noise, designed to explore the capacity to time the onset of species-specific vocalizations within periods of silence. These studies reveal that a wide variety of species are endowed with the capacity to use noise offsets to initiate vocal production (Popp 1989; Garcia-Rutledge & Narins 2001; Egnor et al. 2007). An issue that emerges from such work is the extent to which there are interspecific differences in not only the timing of signal production, but in the capacity to modify other aspects of signal structure as well as differences in the capacity to extract nonoffset cues or patterns to efficiently guide signal output.

The methodological approach pursued here has potentially interesting implications for exploring more complex forms of pattern recognition, providing a natural, nontrained assay for a variety of calling animals. In particular, once an animal has learned that there are different windows of opportunity for calling, and that particular acoustic patterns predict such opportunities, we can exploit this ability to ask about the limits on pattern recognition. For example, in the case presented here, we found that tamarins discriminated three different speech syllables and they recognized that two of the three syllables were associated with the preferable long gap, and that hearing these two syllables provided a perfectly predictable cue to call. Given this result, it is now possible to ask about other perceivable cues or patterns. For example, we can ask about the number of syllables (e.g. 3 syllables predict long silences, 2 predict short silences) to probe issues of number representation, the duration of syllables (2–3 s long versus 4–6 s) to explore issues of timing, repetitions of syllables (e.g. ti-ti versus ti-la) to assess the capacity to extract identity, and even syllables arranged in rule-like patterns (e.g. ti-la-yu-ti versus la-ti-ti-yu, translating into identical sounds at the edges of a string to identical sounds internal to the string). This approach may thus provide answers not only to questions about vocal production, but also about more general computational capacities related to the abstractness of patterns that can be recognized by a variety of species.

Acknowledgments

References

Brumm, H. & Todt, D. 2002. Noise-dependent song amplitude regulation in a territorial songbird. *Animal Behaviour*, **63**, 891–897.

Brumm, H., Voss, K., Koeller, I. & Todt, D. 2004. Acoustic communication in noise: regulation of call characteristics in a New World monkey. *Journal of Experimental Biology*, **207**, 443–448.

Clark Arcadi, A. 1996. Phrase structure of wild chimpanzee pant hoots: patterns of production and interpopulation variability. *American Journal of Primatology*, **39**, 159–178.

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

Cynx, J., Lewis, R., Tavel, B. & Tse, H. 1998. Amplitude regulation of vocalizations in noise by a songbird, *Taeniopygia guttata*. *Animal Behaviour*, **56**, 107–113.

Egnor, R. S. E. & Hauser, M. D. 2004. A paradox in the evolution of primate vocal learning. *Trends in Neurosciences*, **27**, 649–654.

Egnor, R. S. E. & Hauser, M. D. 2006. Noise-induced vocal modulation in cotton-top tamarins (*Saguinus oedipus*). *American Journal of Primatology*, **68**, 1183–1190.

Egnor, R. S. E., Iguina, C. & Hauser, M. D. 2006. Perturbation of auditory feedback causes systematic perturbation in vocal structure in adult cotton-top tamarins. *Journal of Experimental Biology*, **209**, 3652–3663.

Egnor, R. S. E., Wickelgren, J. G. & Hauser, M. D. 2007. Tracking silence: adjusting vocal production to avoid acoustic interference. *Journal of Comparative Physiology A*, **193**, 477–483.

Foote, A. D., Osborne, R. W. & Hoelzel, A. R. 2004. Whale-call response to masking boat noise. *Nature*, **428**, 910.

Garcia-Rutledge, E. J. & Narins, P. M. 2001. Shared acoustic resources in an old world frog community. *Herpetologica*, **57**, 104–116.

Green, S. & Marler, P. 1979. The analysis of animal communication. In: *Social Behavior and Communication, Handbook of Behavioral Neurobiology. Vol. 3* (Ed. by P. Marler & J. Vandenbergh), pp. 73–158. New York: Plenum.

Janik, V. M. & Slater, P. J. B. 1997. Vocal learning in mammals. *Advances in the Study of Behavior*, **26**, 59–99.

Lengagne, T., Aubin, T., Lauga, J. & Jouventin, P. 1999. How do king penguins (*Aptenodytes patagonicus*) apply the Mathematical Theory of Information to communicate in windy conditions? *Proceedings of the Royal Society London, Series B*, **266**, 1623–1628.

Lombard, E. 1911. Le signe de l'elevation de la voix. *Annales des maladies de l'oreille, du larynx, du nez et du pharynx*, **37**, 101–119.

Marler, P. 1968. Aggregation and dispersal: two functions in primate communication. In: *Primates: Studies in Adaptation and Variability* (Ed. by P. C. Jay), pp. 420–438. New York: Holt, Rinehart & Winston.

Miller, C. T., Flusberg, S. & Hauser, M. D. 2003. Interruptibility of long call production in tamarins: implications for vocal control. *Journal of Experimental Biology*, **206**, 2629–2639.

Mitani, J. C. 1985. Responses of gibbons (*Hylobates muelleri*) to self, neighbor, and stranger duets. *International Journal of Primatology*, **6**, 193–200.

Mitani, J. C. & Marler, P. 1989. A phonological analysis of male gibbon singing behavior. *Behaviour*, **109**, 20–45.

Mitani, J. & Nishida, T. 1993. Contexts and social correlates of long distance calling by male chimpanzees. *Animal Behaviour*, **45**, 735–746.

Newman, J. D., Lieblich, A., Talmage-Riggs, G. & Symmes, D. 1978. Syllable classification and sequencing in twitter calls of squirrel monkeys (*Saimiri sciureus*). *Zeitschrift für Tierpsychologie*, **47**, 77–88.

Nonaka, S., Takahashi, R., Enomoto, K., Katada, A. & Unno, T. 1997. Lombard reflex during PAG-induced vocalizations in decerebrate cats. *Neuroscience Research*, **29**, 283–289.

Popp, J. W. 1989. Temporal aspects of singing interactions among territorial ovenbirds (*Seiurus aurocapillus*). *Ethology*, **82**, 127–133.

Potash, L. M. 1972. Noise-induced changes in calls of the Japanese quail. *Psychonomic Science*, **26**, 252–254.

Robinson, J. G. 1979. An analysis of the organization of vocal communication in the titi monkey *Callicebus moloch*. *Zeitschrift für Tierpsychologie*, **49**, 381–405.

Scheifele, P. M., Andrew, S., Cooper, R. A., Darre, M., Musiek, F. E. & Max, L. 2005. Indication of a Lombard vocal response in the St. Lawrence river beluga. *Journal of the Acoustical Society of America*, **117**, 1486–1492.

Seyfarth, R. M. & Cheney, D. L. 1999. Production, usage, and response in nonhuman primate vocal development. In: *The Design of Animal Communication* (Ed. by M. D. Hauser & M. Konishi), pp. 391–418. Cambridge, Massachusetts: MIT Press.

Sinnott, J. M., Stebbins, W. C. & Moody, D. B. 1975. Regulation of voice amplitude by the monkey. *Journal of the Acoustical Society of America*, **58**, 412–414.

Smotherman, M., Zhang, S. Y. & Metzner, W. 2003. A neural basis for auditory feedback control of vocal pitch. *Journal of Neuroscience*, **23**, 1464–1477.

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

Waser, P. M. 1982. The evolution of male loud calls among mangabeys and baboons. In: *Primate Communication* (Ed. by C. T. Snowdon, C. H. Brown & M. R. Peterson), pp. 117–143. New York: Cambridge University Press.

Weiss, D., Garibaldi, B. & 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*, **115**, 258–271.