Volume 82 Part Six December 2011

ANIMAL
BEHAVIOUR

ELSEVIER

Published for the Association for the Study of Animal Behaviour
and the Animal Behavior Society

ISSN 0003-3472

Dip listening and the cocktail party problem in grey treefrogs: signal recognition in temporally fluctuating noise

Alejandro Vélez*, Mark A. Bee

Department of Ecology, Evolution, and Behavior, University of Minnesota, St Paul

ARTICLE INFO

Dip listening refers to our ability to catch brief 'acoustic glimpses' of speech and other sounds when fluctuating background noise levels momentarily decrease. Exploiting dips in natural fluctuations of noise contributes to our ability to overcome the 'cocktail party problem' of understanding speech in multitalker social environments. We presently know little about how nonhuman animals solve analogous communication problems. Here, we asked whether female grey treefrogs, *Hyla chrysoscelis*, might benefit from dip listening in selecting a mate in the noisy social setting of a breeding chorus. Consistent with a dip-listening hypothesis, subjects recognized conspecific calls at lower thresholds when the dips in a chorus-like noise masker were long enough to allow glimpses of nine or more consecutive pulses. No benefits of dip listening were observed when dips were shorter and included five or fewer pulses. Recognition thresholds were higher when the noise fluctuated at a rate similar to the pulse rate of the call. In a second experiment, advertisement calls comprising six to nine pulses were necessary to elicit responses under quiet conditions. Together, these results suggest that in frogs, the benefits of dip listening are constrained by neural mechanisms underlying temporal pattern recognition. These constraints have important implications for the evolution of male signalling strategies in noisy social environments.

Humans and nonhuman animals often communicate acoustically in large social groups, such as cocktail parties, choruses or colonies (Schwartz & Freeberg 2008). In such aggregations, high levels of background noise generated by the mixture of signals can lead to errors in communication by impairing signal detection, recognition and discrimination (Klump 1996; Brumm & Slabbekoorn 2005). In humans, the challenge of communicating in multitalker environments is known as the 'cocktail party problem' (reviewed in: Bronkhorst 2000; McDermott 2009). Because the ability to solve cocktail-party-like communication problems has important fitness consequences for many species and in different social contexts, we should expect natural selection to favour both signalling behaviours and perceptual mechanisms that facilitate signal recognition in noisy social aggregations (Klump 1996; Brumm & Slabbekoorn 2005; Langemann & Klump 2005; Bee & Micheyl 2008). Compared to our thorough understanding of strategies signallers use to avoid or ameliorate noise problems (reviewed in Brumm & Slabbekoorn 2005), we know considerably less about how

nonhuman receivers may be adapted to cope with the detrimental effects of noise (Klump 1996; Hulse 2002; Bee & Micheyl 2008).

One way humans and other animals may solve cocktail-party-like problems is by exploiting temporal features of the ambient acoustic environment. A prominent feature of natural soundscapes, including those generated by social aggregations, is that noise levels fluctuate in time (Richards & Wiley 1980; Nelken et al. 1999; Vélez & Bee 2010). Humans can recognize speech at significantly lower thresholds when 'speech-shaped noise' (i.e. noise with the long-term spectrum of speech) fluctuates in amplitude compared with thresholds determined in a control treatment with nonfluctuating noise (Gustafsson & Arlinger 1994; Bacon et al. 1998). This release from auditory masking in fluctuating noise is termed 'dip listening' and is attributed to the listener's ability to catch brief 'acoustic glimpses' of speech cues at those moments when the background noise dips to a low level (Cooke 2006; Vestergaard et al. 2011). Masking release by means of dip listening is generally greater when noise fluctuates at slower, compared with faster, rates (Gustafsson & Arlinger 1994; Bacon et al. 1998). At slower rates, dips in noise levels are usually longer, thereby increasing the probability of catching meaningful glimpses of target speech (Gustafsson & Arlinger 1994; Bacon et al. 1998). Our objective in this study was to examine the extent to which dip listening might benefit receivers in a well-described nonhuman communication system.

* Correspondence: A. Vélez, Department of Ecology, Evolution, and Behavior, University of Minnesota, 100 Ecology, 1987 Upper Buford Circle, St Paul, MN 55108, U.S.A.

E-mail address: velez011@umn.edu (A. Vélez).

Anuran amphibians (frogs and toads) represent ideal models for studying sensory mechanisms that facilitate signal recognition in noisy social environments (reviewed in Vélez et al., in press). In most species, males produce loud sexual advertisement calls that are necessary and sufficient for species recognition and female mate choice (Gerhardt & Huber 2002). In many species, males aggregate in suitable breeding habitats and form dense choruses, which in some cases can be heard from up to 2 km away (Arak 1983). Not surprisingly, the intense noise levels in breeding choruses are a particularly potent source of auditory masking that impairs signal detection, recognition and discrimination by receivers (Gerhardt & Klump 1988a; Wollerman 1999; Wollerman & Wiley 2002; Bee 2007a, 2008a, b; Bee & Schwartz 2009; Vélez & Bee 2010; Vélez et al., in press). Like the background noise at a cocktail party, the level of noise in frog choruses is not constant, but rather exhibits short-term, moment-to-moment fluctuations through time (Vélez & Bee 2010). Could frogs, like humans, exploit amplitude fluctuations in background noise to facilitate communication in noisy social settings?

Here, we tested the hypothesis that females of Cope's grey treefrog, Hyla chrysoscelis, experience a release from auditory masking by listening in the dips of fluctuating chorus noise for the calls of potential mates. We experimentally measured 'signal recognition thresholds' (Bee & Schwartz 2009) in the presence of a chorus-like noise without level fluctuations and in spectrally matched noises in which the level fluctuated over time. In a second experiment, we assessed whether any observable effects of dip listening might be explained by the duration of acoustic glimpses of calls occurring in the dips. Together, results from this study indicate that frogs can exploit dips in chorus noise in ways that parallel our own ability to exploit dips in the background noise of multitalker social environments.

GENERAL METHODS

Subjects

Protocols for collecting, handling and testing animals were approved by the University of Minnesota's Institutional Animal Care and Use Committee (No. 0809A46721, November 2008). We collected a total of 287 females in amplexus between 2100 and 0100 hours in May and June of 2009 and 2010 from wetlands located in east-central Minnesota, U.S.A. (Carver County, 44°52′49.29″N, 93°43′3.10″W; Hennepin County, 45°11′18.71″N, 93°39′9.05″W). Females were returned to the laboratory and kept at 2 °C to delay oviposition until tested (usually within 24 h). We placed females in a 20 °C incubator for at least 30 min prior to testing to allow their body temperature to reach 20 ± 1 °C. We released females at their capture location after testing (usually within 2 days).

Phonotaxis Experiments

We conducted 'no-choice' phonotaxis experiments (Gerhardt 1995) under infrared (IR) illumination inside a walk-in, temperature-controlled (20 ± 1 °C), hemianechoic sound chamber (additional details in Bee & Schwartz 2009). Females were tested in a circular arena (2 m diameter) made of acoustically transparent but visually opaque walls (60 cm height). Behavioural responses were recorded using an IR-sensitive video camera mounted directly above the test arena and viewed on a monitor and scored by two observers in real-time outside the sound chamber. The perimeter of the arena was divided into 24 15° arcs.

All acoustic stimuli were digitally generated using Matlab v7.6 (Mathworks, Natick, MA, U.S.A.; 11.025 kHz, 16-bit). Target signals were broadcast from an A/D/S L210 speaker placed on the floor just outside the wall of the arena, centred in one of the 15° arcs, 1 m away from a release point at the centre of the arena. We varied the position of the speaker around the arena's perimeter between tests of two to four females to eliminate any possibility of directional response bias. No such biases have been observed in our experimental set-up. In experiment 1, masking noises were broadcast from an overhead speaker (Kenwood KFC-1680ie) suspended from the ceiling of the chamber 190 cm above the central release point. Noise levels were uniform (± 2 dB) across the floor of the arena, and the frequency response of the playback system was flat (± 3 dB) between 500 Hz and 4000 Hz. Sound levels were measured and calibrated by placing the microphone of a Brüel & Kjær Type 2250 sound level meter at the approximate position of a subject's head at the release point.

At the beginning of each test, a subject was placed in an acoustically transparent holding cage located at the arena's central release point. Subjects were initially placed with random orientation relative to speaker positions and could freely reorient inside the cage. Broadcasts of the target signal initiated after a 1.5 min silent acclimation period. After 15 s of signal presentation, subjects were released using a rope and pulley system operated from outside the chamber. In tests with masking noise in experiment 1, we started broadcasts of the masker 30 s before the onset of the target signal; the masker was broadcast continuously throughout the remainder of the test. Subjects were tested in series of 6–11 phonotaxis trials and were given a time-out period of 5–15 min inside the incubator between tests. Female grey treefrogs do not show directional biases or carryover effects in repeated tests (Gerhardt et al. 2000).

## EXPERIMENT 1: SIGNAL RECOGNITION IN FLUCTUATING NOISES

In experiment 1, we tested the hypothesis that females listen in the dips of temporally fluctuating background noise. We compared signal recognition thresholds in response to a target signal presented in the presence of a nonfluctuating (control) masker and maskers differing in rates of temporal fluctuation. Based on the dip-listening hypothesis, we predicted (1) that subjects would experience masking release in fluctuating noise, as evidenced by relatively lower signal recognition thresholds in the presence of fluctuating maskers compared with the nonfluctuating (control) masker, and (2) that the magnitude of masking release would be greater in the presence of maskers that fluctuated at slower compared with faster rates.

Methods

Acoustic stimuli

The target signal was a synthetic advertisement call with values of spectral and temporal properties close to the averages of calls recorded in local populations (Fig. 1a, b). The call had 32 pulses (11 ms pulse duration) delivered at a rate of 45.5 pulses/s (22 ms pulse period) and repeated with a period of 5 s during a phonotaxis trial. Each pulse consisted of two phase-locked sinusoids with frequencies (and relative amplitudes) of 1.25 kHz (-9 dB) and 2.5 kHz (0 dB). The amplitude envelope of each pulse was shaped with a 4 ms inverse exponential rise time and a 7 ms exponential fall time.

The maskers consisted of artificial 'chorus-shaped noises' created by filtering white noise so that it had the long-term frequency spectrum of the sound of natural Cope's grey treefrog breeding choruses (Fig. 1a). Between May and July, 2007–2010, we recorded 25 different grey treefrog choruses using a Marantz PMD
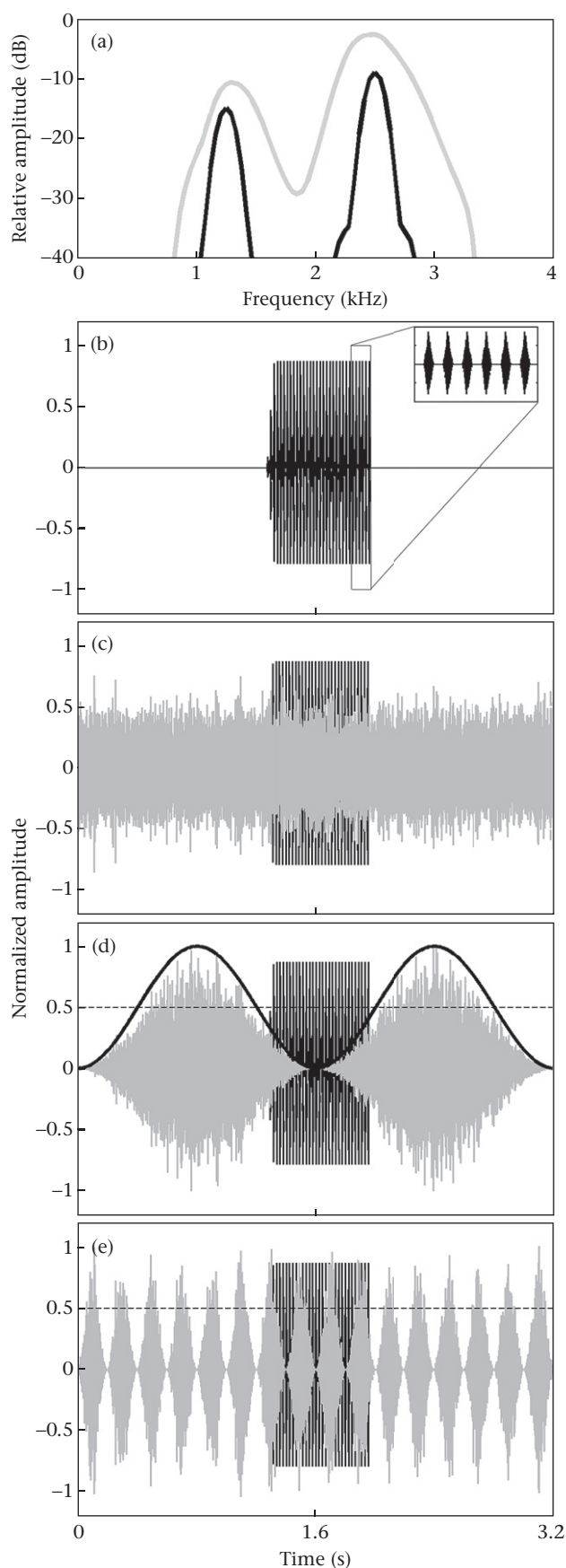
670 digital recorder and a Sennheiser ME62 microphone. We made recordings (44.1 kHz sampling rate, 16-bit resolution) of at least 1.5 min duration near the nightly peak of calling activity, at distances of 4–10 m from the nearest calling male, and at a height of 5 cm above ground or water level. We chose these distances and this microphone height because females at our field sites probably assess males from such positions. We only used recordings of choruses in which other frog species were not heard calling.
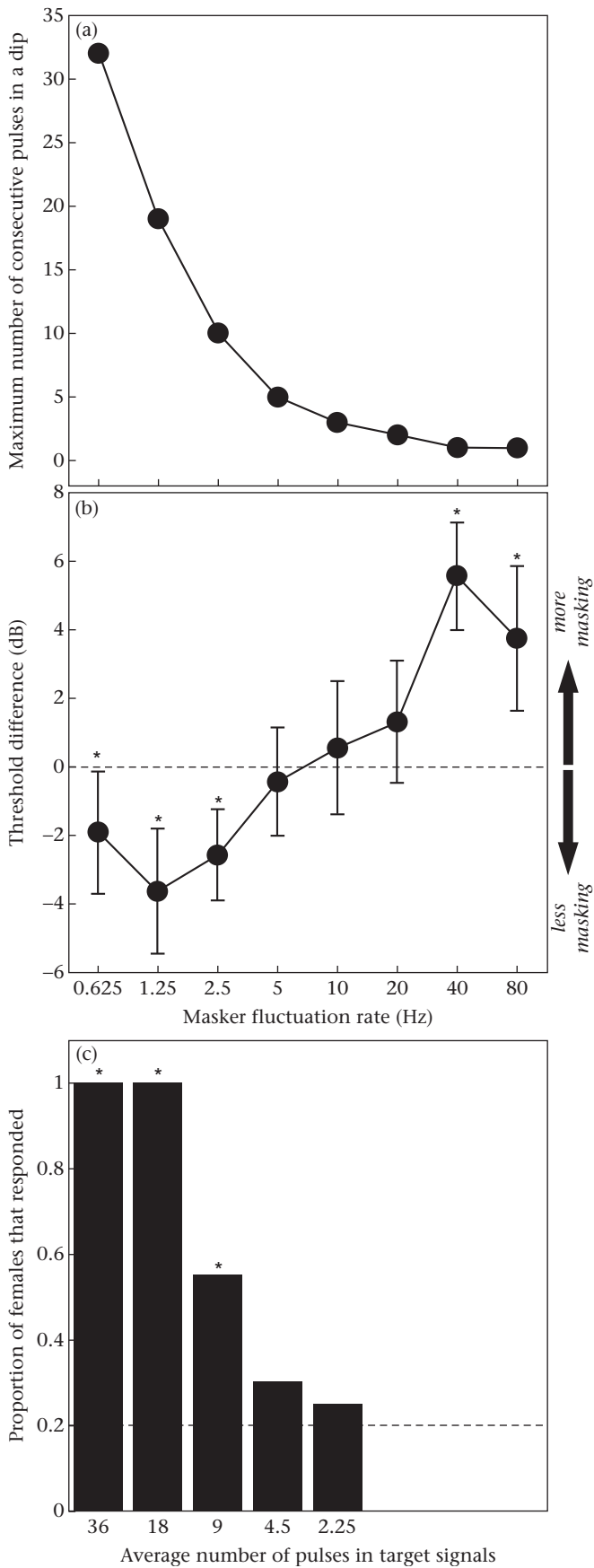
For experiment 1, we selected 14 1.5 min recordings made between 2007 and 2009 as models for creating chorus-shaped maskers for use in phonotaxis experiments conducted in 2009 and 2010. Each recording was first down-sampled to 11.025 kHz and transformed to the spectral domain before determining its power spectrum (FFT size = 4 194 304). We then calculated the average spectrum of the 14 recordings. To create a chorus-shaped noise, we generated a 6 min white noise, transformed it to the spectral domain (FFT size = 4 194 304) and multiplied it by the average frequency spectrum determined from the 14 natural choruses. We then band-pass filtered the resulting frequency spectrum between 850 Hz and 3300 Hz (Fig. 1a) and transformed it to the temporal domain with an inverse FFT (Fig. 1c).

Across nine different treatments, chorus-shaped maskers differed in their temporal structure but had equivalent long-term root-mean-square (RMS) amplitudes (Fig. 1c–e). In the nonfluctuating control treatment (Fig. 1c), the masker had no fluctuations beyond the inherent random fluctuations characteristic of band-limited noise, which were present in all of our chorus-shaped maskers (cf. Fig. 1c–e). The remaining treatments comprised eight experimental treatments in which we imposed sinusoidal fluctuations in masker level at rates of 0.625 Hz (Fig. 1d), 1.25 Hz, 2.5 Hz, 5 Hz (Fig. 1e), 10 Hz, 20 Hz, 40 Hz and 80 Hz. These sinusoidally amplitude-modulated (SAM) maskers were created by multiplying the nonfluctuating masker by a modulating sine wave according to the following equation:

$$n(t) = A(1 + m \sin(2\pi f_m t + \rho))c(t),$$

where $n(t)$ is the resulting SAM masker, A is a scaling factor, m is the modulation depth (1.0 in all cases), f_m is the modulation frequency (from 0.625 Hz to 80 Hz in one-octave steps), ρ is the starting phase of the modulator, $c(t)$ is the nonfluctuating chorus-shaped masker, and t is time in seconds (Fig. 1d, e). The duration of each 'dip' in these maskers, as measured between the 6 dB down points (Fig. 1d, e), were 800 ms, 400 ms, 200 ms, 100 ms, 50 ms, 25 ms, 12.5 ms and 6.25 ms, respectively. Consequently, the maximum number of consecutive pulses in the 32-pulse target signal that could fit within these dips decreased from 32 pulses to one pulse as the fluctuation rate increased from 0.625 Hz to 80 Hz (Fig. 2a). The rates of fluctuation in our maskers spanned the range present in the sounds of grey treefrog choruses (Vélez & Bee 2010) and encompassed the pulse repetition rate of the advertisement call ($\approx$40–50 pulses/s). In playback tests, the equivalent long-term RMS amplitude of all maskers was set to a sound pressure level (SPL re. 20 μPa) of 73 dB (LC_{eq}). This level falls within the range of chorus noise levels measured in the field (Schwartz et al. 2001; Swanson et al. 2007).

Waveforms of 3.2 s segments of the target signal alone (b), and in the presence of maskers representing the nonfluctuating control treatment (c), or experimental treatments with maskers that fluctuated with sinusoidal modulation at rates of 0.625 Hz (d) and 5 Hz (e). Inset in (b) shows six pulses of the target signal. The solid black line in (d) depicts the sine wave used to modulate the masker. The dashed lines in (d) and (e) serve to illustrate the values at which the amplitude of the fluctuating maskers reaches 50% of the maximum amplitude; this line marks the 6 dB down points used to measure the maximum number of consecutive pulses falling within dips of fluctuating maskers.

Figure 1. (a) Frequency spectrum of the target signal (black line) and the chorus-shaped maskers (grey line) calculated with 256-point Blackman–Harris windows.

We used artificial SAM maskers in this experiment because they are commonly used in psychophysical and neurophysiological studies of temporal auditory processing (Joris et al. 2004), they offer a high level of experimental control, and because the sounds of natural choruses can act as an attractive signal (e.g. Swanson et al. 2007) and thereby potentially confound the results of call recognition experiments (see experiment 3 below). We used four exemplars of each masker generated with a different white noise; for SAM maskers, each exemplar was created using a sinusoidal modulator with a different starting phase ($\rho = 0°$, $90°$, $180°$ or $270°$) because, depending on the phase of the masker and the timing of the target signal, starting phase could have affected the number of pulses of the signal falling within dips during a phonotaxis trial. In this experiment, however, the number of pulses that fell in dips was the same in all exemplars. We tested equal numbers of subjects with each exemplar and treated starting phase as a between-subjects factor in our statistical analyses. Subjects were assigned randomly to each combination of starting phase and masker type.

Signal recognition thresholds

We used a between-subjects design ($N = 24$ per treatment) to measure signal recognition thresholds in response to a target signal presented in the control treatment and eight experimental treatments (total $N = 216$). Following Bee & Schwartz (2009), we operationally defined the signal recognition threshold as the lowest signal-to-noise ratio eliciting a positive phonotaxis response. We scored a response when (1) the subject's first contact with the arena wall was in the hemicircle of the arena containing the target speaker, (2) the subject touched the wall inside the 15° bin in front of the target speaker within 5 min of being released and (3) it remained in a 30° bin centred on the speaker for 30 s. We would note that signal recognition thresholds are not the same as signal detection thresholds because positive phonotaxis indicates that the subject (1) has detected the signal, (2) can localize it and (3) recognizes it as the call of an appropriate mate.

We determined signal recognition thresholds using an adaptive tracking procedure in which we tested subjects in a series of 'reference trials' and 'test trials' (Bee & Schwartz 2009). Each series began and ended with a reference trial in which we broadcast the target signal in the absence of noise at a level of 85 dB SPL (LCF), which corresponds to natural call amplitudes measured at 1 m (Gerhardt 1975). Reference trials elicit robust phonotaxis from motivated females. We also tested subjects on a reference trial after any two successive test trials failed to elicit a response to ensure that the subject was indeed still responsive. As in previous studies (Bush et al. 2002; Bee & Schwartz 2009; Vélez & Bee 2010), we replaced any subject that failed to respond on any reference trial ($N = 11$). This procedure ensured that only highly motivated subjects were included in the analyses as evidenced by similar mean response

Figure 2. (a) Maximum number of consecutive pulses falling in the dips of fluctuating maskers of experiment 1 as a function of masker fluctuation rate. A pulse was considered as falling in a dip when its maximum amplitude fell between the 6 dB down points of the masker. Maximum values were calculated from 256 calls for each masking condition (64 calls per exemplar, which corresponds to the total number of calls presented in 5 min during the broadcast of a masker in a phonotaxis test). (b) Mean (±95% confidence intervals) threshold differences as a function of masker fluctuation rate in experiment 1; these differences are relative to the threshold measured in the control condition with a nonfluctuating masker, and the dashed line in (b) represents no difference (i.e. 0 dB) from the control. As depicted by the arrows, values lower than 0 dB represent less masking than in the nonfluctuating control, and values greater than 0 dB represent more masking. (c) Proportion of subjects responding to the target signal as a function of the average number of pulses in the signal in experiment 2. In experiment 2, signals were broadcast in the absence of background noise, and the numbers of pulses were chosen to match the number of consecutive pulses falling in dips of fluctuating maskers of experiment 1. The dashed line in (c) represents the null expectation of 0.2. Asterisks in (b) and (c) represent statistically significant differences from null expectations ($\alpha = 0.05$).

latencies in the initial ($\overline{X} \pm SD = 111.58 \pm 39.89$ s) and final ($\overline{X} \pm SD = 110.54 \pm 38.80$ s) reference trials (paired-sample t test: $t_{0.05\ (2),\ 215} = 0.2179$, $P > 0.6$).

In the first test trial in a series, we broadcast the target signal and the masking noise at a signal-to-noise ratio of 0 dB, that is, the signal was also calibrated to 73 dB SPL (LCF). In subsequent trials, we kept the level of the masker constant at 73 dB SPL (LC$_{eq}$) and systematically varied the level of the target signal contingent upon the subject's behaviour in the previous test. If the subject responded in the previous test trial, the level of the signal was decreased by 3 dB in the next trial. If the subject failed to respond, the level of the target signal was instead increased by 3 dB in the next trial. This pattern continued until the subject changed its behaviour between two consecutive test trials (i.e. from response to no response, or vice versa). On the subsequent test trial following a change in behaviour, we reversed the direction of signal level change and reduced the step size to 1.5 dB. A subject's recognition threshold was calculated as the average between the minimum signal level that elicited a response and the maximum signal level that failed to elicit a response (Bee & Schwartz 2009). Recognition thresholds were calculated in a linear scale and converted back to decibels.

Release from auditory masking is typically calculated as the relative difference (in dB) between thresholds measured in two or more conditions. To examine the magnitude of any masking release due to dip listening, we subtracted the average threshold measured for subjects in the nonfluctuating control treatment ($\overline{X} \pm SE = 67.9 \pm 0.8$ dB; $N = 24$) from those of each subject tested in the eight experimental treatments with fluctuating maskers. We predicted that these 'threshold differences' would be significantly lower than 0 dB if subjects experienced masking release in the presence of a fluctuating masker (i.e. negative threshold differences indicate masking release). We tested this prediction using two-tailed t tests assessing differences from zero. We analysed threshold differences using a two-way ANOVA to investigate the influence of differences in masker fluctuation rates (8 levels) and masker starting phase (4 levels). Threshold differences met the assumptions of normality (Kolmogorov−Smirnov tests: $Ps > 0.20$) and homogeneity of variance (Levene's tests: $Ps > 0.24$) for parametric testing. We used an experiment-wide significance criterion of $\alpha = 0.05$.

Results and Discussion

Average recognition thresholds measured in the presence of fluctuating noise ranged from about 4 dB lower to 6 dB higher than the nonfluctuating control (Fig. 2b), and the magnitude of these threshold differences depended on the masker's fluctuation rate (ANOVA: $F_{7,160} = 15.03$, $P < 0.0001$, partial $\eta^2 = 0.396$). Consistent with the dip-listening hypothesis, subjects experienced a release from masking of about 2−4 dB, on average (Fig. 2b), relative to the control treatment when the maskers fluctuated at rates of 0.625 Hz ($t_{0.05\ (2),\ 23} = -2.22$, $P = 0.036$), 1.25 Hz ($t_{0.05\ (2),\ 23} = -4.08$, $P < 0.001$) and 2.5 Hz ($t_{0.05\ (2),\ 23} = -3.98$, $P < 0.001$). Mean threshold differences were not significantly different from 0 dB when masker fluctuation rates were 5 Hz ($t_{0.05\ (2),\ 23} = -0.58$, $P = 0.571$), 10 Hz ($t_{0.05\ (2),\ 23} = 0.59$, $P = 0.556$) and 20 Hz ($t_{0.05\ (2),\ 23} = 1.52$, $P = 0.142$). In contrast to the dip-listening hypothesis, threshold differences were significantly greater than 0 dB and averaged about 4−6 dB higher than in the control treatment when the masker fluctuated at 40 Hz ($t_{0.05\ (2),\ 23} = 7.31$, $P < 0.001$) and 80 Hz ($t_{0.05\ (2),\ 23} = 3.67$, $P < 0.002$).

We found no significant main effect of the fluctuating masker's starting phase on threshold differences (ANOVA: $F_{3,160} = 2.06$, $P = 0.107$, partial $\eta^2 = 0.04$); there was a weak but significant interaction between a masker's fluctuation rate and its starting phase (ANOVA: $F_{21,160} = 1.63$, $P = 0.047$, partial $\eta^2 = 0.18$). Responses elicited using one exemplar (270° starting phase) were relatively constant across the different masker fluctuation rates, whereas responses with the other three exemplars showed the general pattern depicted in Fig. 2b. This interaction cannot be explained by differences in the maximum number of consecutive pulses falling within dips, which was invariant across all four starting phases. We currently lack a suitable explanation for this interaction.

EXPERIMENT 2: MINIMUM CALL DURATION FOR ACOUSTIC GLIMPSES

The basic idea behind the dip-listening hypothesis is that listeners are able to catch brief glimpses of target signals during dips in ambient noise levels. The results of experiment 1 suggested dip listening occurred at the slowest fluctuation rates tested (≤ 2.5 Hz), which were associated with the longest dips (≥ 200 ms at the 6 dB down points). In experiment 2 we asked, what is the shortest glimpse of an advertisement call that a female grey treefrog could hear in the absence of noise and still recognize it as the call of an appropriate mate? Answering this question was necessary to interpret results from experiment 1. To address this question, we used no-choice phonotaxis experiments to estimate the 'minimum effective call duration', which we operationally defined as the minimum number of pulses in a target signal required to elicit positive phonotaxis.

Methods

The target signals used in this experiment were identical to those of experiment 1 in all respects except pulse number. Here, we varied pulse number across five treatments so that the target signal had on average either 2.25, 4.5, 9, 18 or 36 pulses. These pulse numbers were chosen because they are close to the maximum number of consecutive pulses that fit between the 6 dB down points of the five maskers with the longest dips used in experiment 1 (i.e. 0.625−10 Hz). For the 2.25 pulses treatment, one three-pulse signal was broadcast after every third two-pulse signal; for the 4.5 pulses treatment, every other signal had either four or five pulses. Signals were broadcast at 70 dB SPL (LCF), which is close to the minimum amplitude necessary to elicit phonotaxis in the presence of a chorus-shaped masker presented at 73 dB SPL (Bee & Schwartz 2009; Vélez & Bee 2010; experiment 1 of this study). During a phonotaxis trial, the signal repeated with a period of 5 s for all treatments.

Using a within-subjects design, we tested females ($N = 20$) in all five treatments (in a different randomized order for each subject) and calculated the proportion of subjects that responded. We scored a response if (1) the subject's first contact with the arena wall was in the hemicircle containing the target speaker and (2) the subject touched the wall inside the 15° bin in front of the target speaker within 5 min of being released. We reasoned that if females recognized the signal as a conspecific mating call, they should show positive phonotaxis. We tested this prediction using one-tailed binomial tests and an expected (null) false alarm rate of 0.2. This false alarm rate was determined from data obtained in a previous study (Vélez & Bee 2010). In that study, which used only the second of the two response criteria used here, we determined that the proportion of females that touched the wall of the arena in front of a silent target speaker was 0.20 in the absence of any acoustic stimulation. For purposes of the present study, we reanalysed those data using both of our response criteria and found that the false alarm rate remained 0.20 using both criteria.

Results and Discussion

The proportion of females that met the response criteria (Fig. 2c) was significantly greater than the null expectation of 0.20 for calls

containing nine or more pulses (one-tailed binomial tests: $Ps \leq 0.01$), but not five or fewer pulses ($Ps > 0.19$). These results indicate that in the absence of noise, the minimum effective call duration was between six and nine pulses. Given the gross-temporal properties of our target stimuli, this corresponds to a minimum acoustic glimpse of about 120–190 ms in duration, which is less than the duration of the dips in the fluctuating noises for which masking release was observed in experiment 1 (≤ 2.5 Hz). In other words, the dips in the slowest fluctuating maskers used in experiment 1 were sufficiently long to allow females to catch glimpses of calls that exceeded the minimum effective call duration for eliciting positive phonotaxis in experiment 2.

EXPERIMENT 3: CHORUS-SHAPED MASKERS AS POTENTIAL SIGNALS

Experiment 1 revealed significant differences in signal recognition thresholds depending on the presence/absence and rates of fluctuation in chorus-shaped noise maskers. There was a potential confound in this experiment, however, related to the observation that the 'noise' of natural breeding choruses can act as acoustic 'signals' that allow males and females of some frog species, including our study species, to locate breeding aggregations (Gerhardt & Klump 1988b; Bee 2007b; Swanson et al. 2007; Christie et al. 2010). Therefore, estimates of signal recognition thresholds in the presence of simulated chorus noise could potentially be confounded if the maskers by themselves were not behaviourally neutral stimuli. One way in which such confounds could be introduced would be if subjects were attracted to (or repelled by) some chorus-shaped maskers but not others. In experiment 3, we tested the null hypothesis that nonfluctuating and SAM chorus-shaped maskers, when used as potential target signals, fail to elicit positive or negative phonotaxis.

Methods

In this experiment, there were nine noise stimuli, each of which corresponded to one of the chorus-shaped maskers used in experiment 1 (i.e. the nonfluctuating control and the eight SAM maskers). Females were tested in a series of no-choice phonotaxis trials that began and ended with a reference trial in which we broadcast our attractive 32-pulse target signal in the absence of noise at a level of 85 dB SPL (LCF). In each test trial, a noise stimulus

(i.e. a chorus-shaped masker from experiment 1) was broadcast continuously from a speaker on the floor of the sound chamber, just outside the wall of the test arena, similar to broadcasts of target signals in experiments 1 and 2. We tested eight of the nine noise stimuli (nonfluctuating noise and 1.25 Hz through 80 Hz SAM noise) using a within-subjects design ($N = 20$) in which females were tested with all eight stimuli in randomized orders. A second group of females ($N = 20$) was tested using the 0.625 Hz SAM noise as a stimulus. The long-term RMS amplitude of all noise stimuli was calibrated to 73 dB SPL (LC_{eq}) at the central release point of the test arena. As an additional comparison group ($N = 20$), we reanalysed data from a 'no-noise' treatment conducted in a previous study of ours (Vélez & Bee 2010) that used the same experimental set-up to measure female behaviour in our test arena in the absence of any acoustic stimulation.

As a response variable, we measured orientation angles corresponding to the angle (measured in 5° bins) at which subjects first touched the wall of the arena relative to the location of the playback speaker (designated as 0°). Following previous studies (Swanson et al. 2007; Vélez & Bee 2010), test trials ended as soon as subjects touched the arena wall or after 5 min of stimulus presentation. We employed circular statistics to investigate the effect of the different noise stimuli on orientation angles. We did not use Bonferroni corrections for multiple comparisons in these analyses because doing so would make it more difficult to detect whether the noise stimuli affected female phonotaxis behaviour in this control experiment.

Results and Discussion

In the reference trials, subjects were strongly oriented towards the target signal, as evidenced by mean angles (μ) close to 0°, small circular standard deviations ($\leq 5°$), and lengths of mean vectors (r) close to 1.0 (Table 1). In contrast, there was little evidence that subjects oriented towards the noise stimuli during the test trials, with lengths of mean vectors below 0.4 and circular standard deviations above 77° (Table 1). We statistically assessed these patterns in three separate analyses. First, we tested the null hypothesis that orientation angles measured in the reference trials and in response to each noise stimulus in the test trials were uniformly distributed using one-sample Watson's U^2 tests. Our reasoning was that, if subjects were attracted to or repelled by a noise stimulus in the test trials, orientation angles would not be

Table 1
Results of circular statistical analyses for orientation angles at the wall for the reference and test trials

Condition	Angle at wall				One-sample Watson's U^2 test		Two-sample Watson's U^2 test (rel. No noise)		Two-sample Watson's U^2 test (rel. Reference 1)	
	Mean vector ($\mu°$)	Length of mean vector (r)	Circular SD (°)	N	U^2	P	U^2	P	U^2	P
No noise*	136	0.08	129	17	0.03	>0.5			0.78	<0.001
Reference 1	359	0.99	5	20	1.56	<0.01	0.78	<0.001		
Reference 2	0	0.99	3	20	1.67	<0.01	0.97	<0.001	0.01	>0.5
Nonfluctuating	65	0.20	102	16	0.05	>0.5	0.03	>0.5	0.83	<0.001
0.625 Hz SAM	49	0.35	83	14	0.12	$0.15<P<0.25$	0.09	$0.2<P<0.5$	0.48	<0.001
1.25 Hz SAM	261	0.16	110	16	0.05	>0.5	0.04	>0.5	0.85	<0.001
2.5 Hz SAM	326	0.39	78	16	0.16	$0.05<P<0.1$	0.11	$0.2<P<0.5$	0.70	<0.001
5 Hz SAM	28	0.25	96	18	0.08	$0.25<P<0.5$	0.06	>0.5	0.71	<0.001
10 Hz SAM	35	0.25	96	15	0.08	$0.25<P<0.5$	0.07	$0.2<P<0.5$	0.74	<0.001
20 Hz SAM	31	0.24	97	16	0.07	$0.5<P<0.9$	0.05	>0.5	0.64	<0.001
40 Hz SAM	48	0.15	111	16	0.09	$0.1<P<0.5$	0.06	>0.5	0.84	<0.001
80 Hz SAM	61	0.29	90	17	0.09	$0.5<P<0.9$	0.05	>0.5	0.68	<0.001

N represents the number of subjects that touched the wall of the arena in each type of condition; the total number of subjects tested per condition was 20. SAM: sinusoidally amplitude-modulated maskers.
 * Data reanalysed from Vélez & Bee (2010).

uniformly distributed, and instead would be directed, respectively, towards or away from the stimulus. As illustrated in Table 1 (using one-sample Watson's U tests), significant directionality was found only in responses to the target signals in the reference trials and not in responses to the noise stimuli on the test trials. Thus, females were not oriented towards or away from the speaker when they first touched the arena wall.

In a second analysis, we used two-sample Watson's U^2 tests to determine whether the distributions of orientation angles measured in response to the target signals in reference trials and noise stimuli in test trials differed from an actual null distribution obtained from a separate 'no-noise' treatment in which no sound was broadcast (Vélez & Bee 2010). We reasoned that if the noise stimuli affected female behaviour in the arena, then this might be reflected in a difference between the distributions of orientation angles measured in the presence and absence of the noises. Responses during the reference trials differed significantly from the no-noise treatment, but those measured in response to the noise stimuli did not (see Table 1, two-sample Watson's U^2 tests, 'rel. No noise'). Hence, female orientation in the presence of the noise stimuli was not different from a quiet condition lacking the noise stimuli.

Finally, we assessed whether females treated the noise stimuli as attractive target signals. We tested the null hypothesis that the distribution of orientation angles towards the noise stimuli was not different from that observed in the reference trials using two-sample Watson's U^2 tests. Importantly, the two-sample Watson's U^2 test assumes independent samples. Our data set allowed us to meet this assumption in the following way. We used responses on the first reference trial from the 20 subjects tested only with the 0.625 Hz noise stimulus for making comparisons with responses to the eight noise stimuli made by the other group of 20 subjects. Likewise, responses of this second group of 20 subjects on their first reference trial served for making comparisons with responses to the 0.625 Hz noise stimulus by the first group of 20 subjects. The distributions of orientation angles in response to the noise stimuli and the no-noise condition of Vélez & Bee (2010) were significantly different from those measured in response to an attractive target signal in the first reference trial (see Table 1, two-sample Watson's U^2 tests, 'rel. Reference 1'). Responses in the two reference trials were not significantly different (Table 1). Together, results from these three analyses confirm that the maskers used in experiment 1 were, by themselves, behaviourally neutral stimuli.

GENERAL DISCUSSION

Dip Listening or Modulation Masking?

Consistent with the dip-listening hypothesis, female grey tree-frogs experienced about a 2–4 dB release from masking when a chorus-like noise fluctuated at slow rates (≤2.5 Hz), at which the dips were relatively long (≥200 ms). We believe this masking release can be explained, in part, by the number of consecutive pulses that fell in the dips of the fluctuating maskers. At fluctuation rates of 2.5 Hz and slower, the maximum number of consecutive pulses that fell in the dips was nine or more pulses; at the faster rates we tested (≥5 Hz), five or fewer consecutive pulses were able to fit in a dip (Fig. 2a). These results are in line with those of our second experiment, in which we found that calls with five or fewer pulses failed to elicit responses, whereas calls with nine or more pulses reliably elicited positive phonotaxis. Together, these results are consistent with the hypothesis that females experienced masking release by catching short but meaningful acoustic glimpses of advertisement calls as a result of listening in the dips of fluctuating chorus-like noise. While the benefits of dip listening reported here are small (2–4 dB), they are large enough to be

biologically important. In previous two-choice discrimination experiments, females of Cope's grey treefrog, *H. chrysoscelis*, and the closely related eastern grey treefrog, *Hyla versicolor*, strongly preferred mating calls broadcast at higher sound levels in both the presence or absence of chorus-like noise, even when the difference between two alternatives was as small as 2–4 dB (Fellers 1979; see also Gerhardt et al. 2000; M. A. Bee, A. Vélez & J. D. Forester, unpublished data). In addition, differences in signal amplitude on the order of 2–4 dB can eliminate or reverse female preferences for certain call properties (Gerhardt et al. 2000). Hence, a 2–4 dB masking release due to dip listening could be quite important for a female grey treefrog listening to potential mates in the din of a breeding chorus.

Not all of our results, however, were consistent with the dip-listening hypothesis. We found that females actually experienced additional masking when the background noise fluctuated at some of the faster rates we tested (>20 Hz). This additional masking was most pronounced (≈6 dB) when the masker's fluctuation rate (40 Hz) was most similar to the pulse rate of the target signal (45.5 pulses/s) (Fig. 2b). Females of *H. chrysoscelis* use pulse rate as an acoustic property for species recognition (Schul & Bush 2002). Previous studies of call interference have shown that overlapping calls can disrupt a female's perception of pulse rate (Marshall et al. 2006; Schwartz & Marshall 2006). We attribute these results to a phenomenon known from the human hearing literature as 'modulation masking' (Bacon & Grantham 1989; Kwon & Turner 2001). Used to describe the increased difficulty we have perceiving amplitude-modulated signals in the presence of amplitude-modulated noise, modulation masking is most pronounced when signals and maskers fluctuate at similar rates (Bacon & Grantham 1989). Importantly, studies of human speech communication have found that, under some circumstances, both modulation masking and masking release due to dip listening can occur simultaneously while listening to speech in fluctuating noise (Kwon & Turner 2001).

The effects of fluctuating noise on the ability to recognize communication signals shown here are not unique to vertebrates. Following similar methods, Ronacher & Hoffmann (2003) found that male grasshoppers (*Chorthippus biguttulus*) experience masking release by means of dip listening when the background noise fluctuates at slow rates. Interestingly, the dip-listening effect was observed when the target signal was long (12 subunits) but not when it was short (4 subunits). In addition, male grasshoppers also experienced modulation masking when the background noise and the signal fluctuated at similar rates. The modulation masking effect was observed with both long and shorts signals.

A key question arising from our results and those of Ronacher & Hoffmann (2003) concerns the extent to which dip listening, modulation masking, or both, might influence signal recognition in the natural setting of a breeding aggregation. Answering this question requires a detailed understanding of the animal's acoustic environment. We recently showed that the sounds of grey treefrog choruses fluctuate both at slow rates (<2 Hz) and at a faster rate (≈40–60 Hz) (Vélez & Bee 2010). The slow rates of fluctuation present in natural choruses are due to a combination of low-frequency environmental perturbations (e.g. air turbulence; Richards & Wiley 1980) and the production of temporally discrete calls by individual males in the chorus (Vélez & Bee 2010). The high rates of fluctuation reflect the pulse repetition rate of the grey treefrog advertisement call (Vélez & Bee 2010). Additional acoustic analyses of recordings of grey treefrog choruses revealed that high fluctuation rates (≈40–60 Hz) are more prominent when sound levels are high (i.e. at the peaks in amplitude) and far less prominent during dips in sound level (Supplementary Material). Thus, in a chorus setting, modulation masking should be most pronounced at those moments when the ambient noise level is already high

(e.g. when a nearby male calls) and less pronounced during momentary dips in noise level (e.g. between the calls of nearby males). These results suggest that female grey treefrogs, like humans (Kwon & Turner 2001; Vestergaard et al. 2011), may at times experience a combination of dip listening and modulation masking in noisy social environments.

Solutions to Cocktail-party-like Problems

The perceptual mechanisms by which nonhuman animals solve cocktail-party-like problems potentially include abilities to exploit various spectral, temporal and spatial cues (Hulse 2002; Bee & Micheyl 2008). In aggregations of conspecific signallers, in which signals and noise overlap in acoustic frequency, temporal and spatial cues should be particularly important. For instance, the outstanding ability of king penguin, *Aptenodytes patagonicus*, chicks to recognize parental calls in cacophonous crèches was hypothesized to result from adaptations of the auditory system that exploit amplitude fluctuations in background noise (Aubin & Jouventin 1998). While this hypothesis has not been directly tested in penguins, our results and those of Ronacher & Hoffmann (2003) provide robust support for the idea that nonhuman receivers can also listen in the dips of fluctuating noise in large social aggregations.

Another series of studies has shown that receivers from diverse taxa experience a significant release from masking when signals and sources of noise are spatially separated (reviewed in Bee & Micheyl 2008). In treefrogs, for example, spatial release from masking contributes to recognition of conspecific signals (Schwartz & Gerhardt 1989; Bee 2007a; Nityananda & Bee, in press) and discrimination between conspecific and heterospecific signals (Bee 2008a). Furthermore, in chorus-like situations, spatial separation between male treefrogs improves the ability of females to discriminate among conspecific calls varying in attractiveness (Richardson & Lengagne 2010). In natural settings, we would predict that receivers benefit from synergistic interactions between spatial release from masking and dip listening, as is the case in humans (Hawley et al. 2004).

Sexual selection (either inter- or intrasexual) has also favoured the evolution of various signalling strategies to overcome the detrimental effects of noise (reviewed in: Brumm & Slabbekoorn 2005; Schwartz & Bee, in press). For example, although they do not avoid call overlap with their neighbours in a chorus (Schwartz et al. 2002), male grey treefrogs increase both the duration and redundancy of their signals by adding pulses to lengthen their calls as noise levels and local caller densities increase (Schwartz et al. 2002; Love & Bee 2010). The functional significance of increasing call duration from the male perspective may stem from dip listening on the part of females; males that produce longer calls should have a higher probability of placing more consecutive pulses in the dips of temporally fluctuating chorus noise. This hypothesis is in line with the 'interference risk hypothesis' (Schwartz et al. 2001, 2002, 2008), which states that males lengthen their calls because doing so increases the likelihood that, on average, they will broadcast more pulses per call free of overlap by the calls of other males. This appears to be the case in grasshoppers; males experience a release from masking in the presence of fluctuating noise with long signals, but not with short signals (Ronacher & Hoffmann 2003), suggesting an advantage for females that produce longer signals in the presence of fluctuating noise. Available studies of the eastern grey treefrog, *H. versicolor*, however, offer little support for this functional explanation for such change in male signalling behaviour (Schwartz et al. 2002, 2008). Future studies of signal recognition with fluctuating maskers and signals of different duration might shed light on why grey treefrog males increase the duration of their calls in noisy environments.

Proximate Level Explanations for Dip Listening

Our results are consistent with previous behavioural and neurophysiological studies of sound pattern recognition in frogs. Subjects in experiment 2 required a minimum effective call duration of six to nine consecutive pulses before responding. In the eastern grey treefrog (*H. versicolor*), in which the pulses and interpulse intervals of the call are about twice as long as those of Cope's grey treefrogs (*H. chrysoscelis*), the threshold number of pulses necessary to elicit phonotaxis is between three and six pulses (Bush et al. 2002). A neurophysiological mechanism that may explain these behavioural data involves 'pulse-integrator' neurons in the frog midbrain that only fire in response to a threshold number of consecutive pulses produced with specific interpulse intervals (Alder & Rose 1998; Edwards et al. 2002; Schwartz et al. 2010). Importantly, different neurons can have different threshold numbers of interpulse intervals. Hence, a possible mechanistic explanation for our data is that recognition thresholds were lower in the presence of slowly fluctuating maskers, compared with faster fluctuating maskers, because the number of consecutive pulses and interpulse intervals fitting in the dips exceeded the thresholds of relatively larger populations of pulse-integrator neurons in the midbrain. We hypothesize that the response properties of pulse-integrator neurons might also explain the lack of masking release at faster modulation rates, where five or fewer consecutive pulses fell in the dips (see Fig. 1e). Similar to the 'resetting' of pulse-integrator neurons after an interpulse interval with anomalous duration (Edwards et al. 2002), momentary peaks in amplitude that mask pulses and interpulse intervals might also reset these neurons. Future neurophysiological studies of the responses of pulse-integrator neurons in the context of dip listening will be an important next step in understanding how the anuran auditory system exploits temporal fluctuations in background noise.

Conclusions

To our knowledge, this is the first study to demonstrate that a nonhuman vertebrate benefits from listening in the dips of ambient noise to perceive communication signals. We suspect dip listening is a widespread hearing mechanism shared by different vertebrate and invertebrate taxa that facilitates acoustic communication in noisy environments. The ability of female frogs to exploit temporal fluctuations in background noise may be constrained at a proximate level by specific sensory mechanisms involving pulse-integrator neurons responsible for encoding temporal properties critical for species recognition. At an ultimate level, dip listening by females may act as a source of selection favouring males that are able to produce longer calls. Additional experiments with maskers having natural amplitude fluctuations and calls varying in duration will shed additional light on the relative benefits and constraints experienced by signallers and receivers in natural acoustic environments.

Supplementary Material

Supplementary material for this article is available, in the online version, at doi:10.1016/j.anbehav.2011.09.015.

References

Alder, T. B. & Rose, G. J. 1998. Long-term temporal integration in the anuran auditory system. *Nature Neuroscience*, **1**, 519–523.

Arak, A. 1983. Sexual selection by male–male competition in natterjack toad choruses. *Nature*, **306**, 261–262.

Aubin, T. & Jouventin, A. 1998. Cocktail-party effect in king penguins. *Proceedings of the Royal Society B*, **265**, 1665–1673.

Bacon, S. P. & Grantham, D. W. 1989. Modulation masking: effects of modulation frequency, depth, and phase. *Journal of the Acoustical Society of America*, **85**, 2575–2580.

Bacon, S. P., Opie, J. M. & Montoya, D. Y. 1998. The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds. *Journal of Speech Language and Hearing Research*, **41**, 549–563.

Bee, M. A. 2007a. Sound source segregation in grey treefrogs: spatial release from masking by the sound of a chorus. *Animal Behaviour*, **74**, 549–558.

Bee, M. A. 2007b. Selective phonotaxis by male wood frogs (*Rana sylvatica*) to the sound of a chorus. *Behavioral Ecology and Sociobiology*, **61**, 955–966.

Bee, M. A. 2008a. Finding a mate at a cocktail party: spatial release from masking improves acoustic mate recognition in grey treefrogs. *Animal Behaviour*, **75**, 1781–1791.

Bee, M. A. 2008b. Parallel female preferences for call duration in a diploid ancestor of an allotetraploid treefrog. *Animal Behaviour*, **76**, 845–853.

Bee, M. A. & Micheyl, C. 2008. The 'cocktail party problem': what is it? How can it be solved? And why should animal behaviorists study it? *Journal of Comparative Psychology*, **122**, 235–251.

Bee, M. A. & Schwartz, J. J. 2009. Behavioral measures of signal recognition thresholds in frogs in the presence and absence of chorus-shaped noise. *Journal of the Acoustical Society of America*, **126**, 2788–2801.

Bronkhorst, A. W. 2000. The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions. *Acustica*, **86**, 117–128.

Brumm, H. & Slabbekoorn, H. 2005. Acoustic communication in noise. *Advances in the Study of Behavior*, **35**, 151–209.

Bush, S. L., Gerhardt, H. C. & Schul, J. 2002. Pattern recognition and call preferences in treefrogs (Anura: Hylidae): a quantitative analysis using a no-choice paradigm. *Animal Behaviour*, **63**, 7–14.

Christie, K., Schul, J. & Feng, A. S. 2010. Phonotaxis to male's calls embedded within a chorus by female gray treefrogs, *Hyla versicolor*. *Journal of Comparative Physiology A*, **196**, 569–579.

Cooke, M. 2006. A glimpsing model of speech perception in noise. *Journal of the Acoustical Society of America*, **119**, 1562–1573.

Edwards, C. J., Alder, T. B. & Rose, G. J. 2002. Auditory midbrain neurons that count. *Nature Neuroscience*, **5**, 934–936.

Fellers, G. 1979. Aggression, territoriality, and mating behaviour in North American treefrogs. *Animal Behaviour*, **27**, 107–119.

Gerhardt, H. C. 1975. Sound pressure levels and radiation patterns of vocalizations of some North American frogs and toads. *Journal of Comparative Physiology A*, **102**, 1–12.

Gerhardt, H. C. 1995. Phonotaxis in female frogs and toads: execution and design of experiments. In: *Methods in Comparative Psychoacoustics* (Ed. by G. M. Klump, R. J. Dooling, R. R. Fay & W. C. Stebbins), pp. 209–220. Basel: Birkhäuser Verlag.

Gerhardt, H. C. & Huber, F. 2002. *Acoustic Communication in Insects and Anurans: Common Problems and Diverse Solutions*. Chicago: Chicago University Press.

Gerhardt, H. C. & Klump, G. M. 1988a. Masking of acoustic signals by the chorus background noise in the green treefrog: a limitation on mate choice. *Animal Behaviour*, **36**, 1247–1249.

Gerhardt, H. C. & Klump, G. M. 1988b. Phonotactic responses and selectivity of barking treefrogs (*Hyla gratiosa*) to chorus sounds. *Journal of Comparative Physiology A*, **163**, 795–802.

Gerhardt, H. C., Tanner, S. D., Corrigan, C. M. & Walton, H. C. 2000. Female preference functions based on call duration in the gray tree frog (*Hyla versicolor*). *Behavioral Ecology*, **11**, 663–669.

Gustafsson, H. A. & Arlinger, S. D. 1994. Masking of speech by amplitude modulated-noise. *Journal of the Acoustical Society of America*, **95**, 518–529.

Hawley, M. L., Litovski, R. Y. & Culling, J. F. 2004. The benefit of binaural hearing in a cocktail party: effect of location and type of interferer. *Journal of the Acoustical Society of America*, **115**, 833–843.

Hulse, S. H. 2002. Auditory scene analysis in animal communication. *Advances in the Study of Behavior*, **31**, 163–200.

Joris, P. X., Schreiner, C. E. & Rees, A. 2004. Neural processing of amplitude-modulated sounds. *Physiological Reviews*, **84**, 541–577.

Klump, G. M. 1996. Bird communication in the noisy world. In: *Ecology and Evolution of Acoustic Communication in Birds* (Ed. by D. E. Kroodsma & E. H. Miller), pp. 321–338. Ithaca, New York: Cornell University Press.

Kwon, B. J. & Turner, C. W. 2001. Consonant identification under maskers with sinusoidal modulation: masking release or modulation interference? *Journal of the Acoustical Society of America*, **110**, 1130–1140.

Langemann, U. & Klump, G. M. 2005. Perception and acoustic communication networks. In: *Animal Communication Networks* (Ed. by P. K. McGregor), pp. 451–480. Cambridge: Cambridge University Press.

Love, E. K. & Bee, M. A. 2010. An experimental test of noise-dependent voice amplitude regulation in Cope's grey treefrog (*Hyla chrysoscelis*). *Animal Behaviour*, **80**, 509–515.

McDermott, J. H. 2009. The cocktail party problem. *Current Biology*, **19**, R1024–R1027.

Marshall, V. T., Schwartz, J. J. & Gerhardt, H. C. 2006. Effects of heterospecific call overlap on the phonotactic behaviour of grey treefrogs. *Animal Behaviour*, **72**, 449–459.

Nelken, I., Rotman, Y. & Yosef, O. B. 1999. Responses of auditory-cortex neurons to structural features of natural sounds. *Nature*, **397**, 154–157.

Nityananda, V. & Bee, M. A. In press. Spatial release from masking in a free-field source identification task by gray treefrogs. *Hearing Research*.

Richards, D. G. & Wiley, R. H. 1980. Reverberations and amplitude fluctuations in the propagation of sound in a forest: implications for animal communication. *American Naturalist*, **115**, 381–399.

Richardson, C. & Lengagne, T. 2010. Multiple signals and male spacing affect female preference at cocktail parties in treefrogs. *Proceedings of the Royal Society B*, **277**, 1247–1252.

Ronacher, B. & Hoffmann, C. 2003. Influence of amplitude modulated noise on the recognition of communication signals in the grasshopper *Chorthippus biguttulus*. *Journal of Comparative Physiology A*, **189**, 419–425.

Schul, J. & Bush, S. L. 2002. Non-parallel coevolution of sender and receiver in the acoustic communication system of treefrogs. *Proceedings of the Royal Society B*, **1502**, 1847–1852.

Schwartz, J. J. & Bee, M. A. In press. Anuran signal production in noisy environments. In: *Animal Communication and Noise* (Ed. by H. Brumm). New York: Springer.

Schwartz, J. J. & Freeberg, T. M. 2008. Acoustic interaction in animal groups: signaling in noisy and social contexts: introduction. *Journal of Comparative Psychology*, **122**, 231–234.

Schwartz, J. J. & Gerhardt, H. C. 1989. Spatially mediated release from auditory masking in an anuran amphibian. *Journal of Comparative Physiology A*, **166**, 37–41.

Schwartz, J. J. & Marshall, V. T. 2006. Forms of call overlap and their impact on advertisement call attractiveness to females of the gray treefrog, *Hyla versicolor*. *Bioacoustics*, **16**, 39–56.

Schwartz, J. J., Buchanan, B. W. & Gerhardt, H. C. 2001. Female mate choice in the gray treefrog (*Hyla versicolor*) in three experimental environments. *Behavioral Ecology and Sociobiology*, **49**, 443–455.

Schwartz, J. J., Buchanan, B. W. & Gerhardt, H. C. 2002. Acoustic interaction among male gray treefrogs, *Hyla versicolor*, in a chorus setting. *Behavioral Ecology and Sociobiology*, **53**, 9–19.

Schwartz, J. J., Brown, R., Turner, S., Dushaj, K. & Castano, M. 2008. Interference risk and the function of dynamic shifts in calling in the gray treefrog (*Hyla versicolor*). *Journal of Comparative Psychology*, **122**, 283–288.

Schwartz, J. J., Huth, K., Hunce, R. & Lentine, B. 2010. Effect of anomalous pulse timing on call discrimination by females of the gray treefrog (*Hyla versicolor*): behavioral correlates of neurobiology. *Journal of Experimental Biology*, **213**, 2066–2072.

Swanson, E. M., Tekmen, S. M. & Bee, M. A. 2007. Do female anurans exploit inadvertent social information to locate breeding aggregations? *Canadian Journal of Zoology*, **85**, 921–932.

Vélez, A. & Bee, M. A. 2010. Signal recognition by frogs in the presence of temporally fluctuating chorus-shaped noise. *Behavioral Ecology and Sociobiology*, **64**, 1695–1709.

Vélez, A., Schwartz, J. J. & Bee, M. A. In press. Anuran signal perception in noisy environments. In: *Animal Communication and Noise* (Ed. by H. Brumm). New York: Springer.

Vestergaard, M. D., Fyson, N. R. C. & Patterson, R. D. 2011. The mutual roles of temporal glimpsing and vocal characteristics in cocktail-party listening. *Journal of the Acoustical Society of America*, **130**, 429–439.

Wollerman, L. 1999. Acoustic interference limits call detection in a Neotropical frog *Hyla ebraccata*. *Animal Behaviour*, **57**, 529–536.

Wollerman, L. & Wiley, R. H. 2002. Background noise from a natural chorus alters female discrimination of male calls in a Neotropical frog. *Animal Behaviour*, **63**, 15–22.

ASSESSING THE POTENTIAL FOR DIP LISTENING IN NATURAL SETTINGS

Results from experiment 1 in the main article (Fig. 2) suggested that, compared with a nonfluctuating control condition, slow rates of amplitude fluctuation (< 5 Hz) lead to masking release, whereas faster rates (> 20 Hz) result in additional masking. The question therefore arises as to which of these processes (masking release versus additional masking) is more likely to prevail in the natural environment. A combination of low-frequency environmental perturbations (Richards & Wiley 1980) and the production of temporally discrete calls by individual males generate slow fluctuations (< 2 Hz) in chorus noise (Fig. S1a; Vélez & Bee 2010). The sounds of grey treefrog choruses also fluctuate at a faster rate (≈ 40 Hz–60 Hz) characteristic of the pulsed structure of the mating call (Fig. S1a; Vélez & Bee 2010). Therefore, we might expect that faster rates of modulation should be most pronounced during momentary 'peaks' (not 'dips') in chorus noise levels coincident with the calls of males in a listener's vicinity. If so, then both dip listening and modulation masking might occur at different times. To evaluate this possibility, we analysed chorus recordings to determine the predominance of fluctuation rates corresponding to the pulsatile structure of the call as a function of the overall chorus noise level. We predicted a positive relationship between amplitude of short waveform segments of the chorus noise (as a proxy for identifying peaks and dips) and the modulation depth of the 40–60 Hz component of the modulation spectrum (as a measurement of the prominence of the pulsatile component of the noise).

We analysed the temporal level fluctuations in 60 s segments of 25 different Cope's grey treefrog choruses recorded between 2007 and 2010. All recordings were first down-sampled to 11.025 kHz sampling rate and then normalized to the same root-mean-square (RMS) amplitude to compensate for differences in microphone and recorder gain settings at the times of recording. Each 60 s segment was then subdivided into 300 smaller, nonoverlapping segments (200 ms each), and the waveform RMS amplitude for each of these shorter segments was calculated. This duration was selected because results from experiments 1 and 2 in the main article suggested that female frogs need to hear calls of about 200 ms (e.g. nine pulses of 22 ms pulse period), either in silence or during dips in the background noise, for signal recognition to occur. For each recording, the RMS amplitude of each 200 ms segment was then normalized to the maximum RMS amplitude of the 300 segments for that recording. Therefore, for each 60 s recording, the

segment with the highest RMS amplitude was designated a 0 dB RMS amplitude, with lower-amplitude segments assigned negative values in decibels. As a measurement of modulation depth, we calculated the modulation index (m). The modulation index describes the amount of variation in sound level around the level of a nonfluctuating carrier with the same RMS amplitude. Values close to 0 represent little fluctuation in sound level, while those with values close to 1 represent high variation in sound level. To calculate the modulation index of the 40–60 Hz component of the modulation spectrum of each 200 ms segment, we first extracted the Hilbert envelope of the segment's waveform. We then generated a modulation spectrum for each segment by calculating the fast Fourier transform of the envelope (Hamming window, FFT size = 4096). For each frequency bin of the modulation spectrum (bin size = 2.69 Hz), we calculated the modulation index as twice the magnitude of the Fourier coefficient of the modulation spectrum normalized by the DC component, using the following equation:

$$m(\mathrm{Mr}) = 2 \times \left| \frac{\mathrm{Fc\ envelope\ (Mr)}}{\mathrm{Fc\ envelope\ (0)}} \right|$$

where $m(\mathrm{Mr})$ is the modulation index as a function of modulation rate, Fc envelope (Mr) is the Fourier coefficient for the modulation rate corresponding to each frequency bin of the modulation spectrum, and Fc envelope (0) is the Fourier coefficient of the modulation spectrum at DC. We then recorded the maximum modulation index between 40 and 60 Hz. We analysed the correlation between the 200 ms segments' RMS amplitude and the maximum modulation index of the 40–60 Hz modulation frequency component for the 25 recordings separately (N = 300 segments per recording) and pooled together (N = 7500 segments). We also ranked the 300 segments of each recording by RMS amplitude (0 dB = rank 1) and calculated the average, minimum and maximum modulation index for each rank (1 to 300) across the 25 recordings.

There was a significant positive correlation between the RMS amplitude of the 200 ms segment's waveform and the highest modulation index of the 40–60 Hz component of the segment's modulation spectrum when data from the 25 recordings were pooled (r = 0.33, N =7500 segments, $P < 0.0001$; Fig. S1b). When the 200 ms segments were ranked by RMS amplitude, average modulation indices increased with increasing RMS amplitude (Spearman rank correlation: r_S = -0.928, N = 300, $P < 0.05$; Fig. S1c). When each recording was analysed separately, we found a significant positive correlation for 22 chorus recordings ($0.15 \leq r \leq 0.74$,

$N = 300$ segments per recording, Ps < 0.01; Fig. S1d, e) and no significant correlation for three chorus recordings (-0.06 $\leq r \leq 0.05$, $N = 300$ segments per recording, Ps > 0.28; Fig. 3f). The interpretation of these results is that the fast fluctuations present in chorus noise are most prevalent during peaks, and not the dips, in overall sound level. By extension, these analyses indicate that female grey treefrogs should be more likely to experience conditions favouring dip listening in choruses at times when modulation masking is generally much less pronounced.

References

Richards, D. G. & Wiley, R. H. 1980. Reverberations and amplitude fluctuations in the propagation of sound in a forest: implications for animal communication. *American Naturalist*, **115**, 381–399.

Vélez, A. & Bee, M. A. 2010. Signal recognition by frogs in the presence of temporally fluctuating chorus-shaped noise. *Behavioral Ecology and Sociobiology*, **64**, 1695–1709.
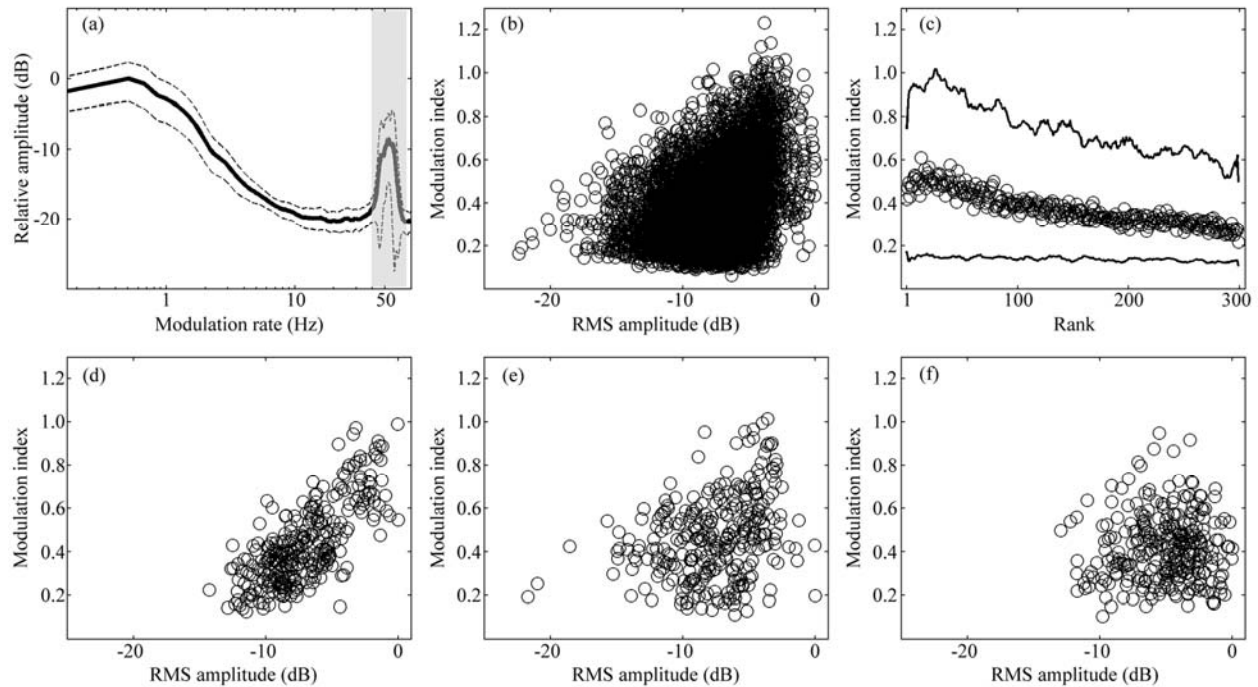
Figure S1. Analyses of the temporal level fluctuations in 60 s recordings of 25 Cope's grey treefrog choruses. (a) Mean (bold, solid line) ± 1 SD (thin, dotted lines) modulation spectrum of grey treefrog choruses shown on a logarithmic X axis. This modulation spectrum depicts the rates of temporal fluctuation in chorus noise in the frequency domain. The grey box depicts the 40–60 Hz component of the modulation spectrum due to the pulsatile structure of the advertisement call. Additional details about the methods to generate Fig. 3a can be found in Vélez & Bee (2010). (b) Modulation depth, expressed as the modulation index, of the 40–60 Hz component from the modulation spectrum as a function of the waveform RMS amplitude for data pooled across all 25 recordings. The modulation index of the 40–60 Hz component was used as a measurement of the prominence of the pulsatile component of the noise, and the waveform RMS amplitude was used as a proxy for identifying peaks and dips in the noise. (c) Average (open circles) and minimum and maximum (solid lines) modulation indices for the 25 60-second recordings ranked by the RMS amplitude of the 200 ms segments; on the X axis, a value of 1 represent segments ranked as having the highest RMS values (i.e. peaks in amplitude) and a value of 300 represent segments ranked with lowest RMS values (i.e. dips in amplitude).

Minimum and maximum values are smoothed with a moving average of 11 points. Examples of individual recordings ($N = 300$ segments per recording) in which the RMS amplitude value and the modulation indices were (d) significantly correlated with a high correlation coefficient ($r = 0.74$, $P < 0.0001$), (e) significantly correlated with a low correlation coefficient ($r = 0.31$, $P < 0.0001$) and (f) not significantly correlated ($r = 0.04$, $P = 0.49$).