

Signal recognition by frogs in the presence of temporally fluctuating chorus-shaped noise

Alejandro Vélez · Mark A. Bee

Received: 6 November 2009 / Revised: 4 May 2010 / Accepted: 11 May 2010 / Published online: 30 May 2010
© Springer-Verlag 2010

Abstract The background noise generated in large social aggregations of calling individuals is a potent source of auditory masking for animals that communicate acoustically. Despite similarities with the so-called cocktail party problem in humans, few studies have explicitly investigated how nonhuman animals solve the perceptual task of separating biologically relevant acoustic signals from ambient background noise. Under certain conditions, humans experience a release from auditory masking when speech is presented in speech-like masking noise that fluctuates in amplitude. We tested the hypothesis that females of Cope's gray treefrog (*Hyla chrysoscelis*) experience masking release in artificial chorus noise that fluctuates in level at modulation rates characteristic of those present in ambient chorus noise. We estimated thresholds for recognizing conspecific advertisement calls (pulse rate=40–50 pulses/s) in the presence of unmodulated and sinusoidally amplitude-modulated (SAM) chorus-shaped masking noise. We tested two rates of modulation (5 and 45 Hz) because the sounds of frog choruses are modulated at low rates (e.g., less than 5–10 Hz) and because those of species with pulsatile signals are additionally modulated at higher rates typical of the pulse rate of calls (e.g., between 15 and 50 Hz). Recognition thresholds were similar in the unmodulated and 5-Hz SAM conditions and 12 dB higher in the 45-Hz SAM condition. These results did not support the hypothesis that female gray treefrogs experience masking release in temporally fluctuating

chorus-shaped noise. We discuss our results in terms of modulation masking and hypothesize that natural amplitude fluctuations in ambient chorus noise may impair mating call perception.

Keywords Auditory masking · Cocktail party problem · Gray treefrog · *Hyla chrysoscelis* · Masking release · Modulation masking

Introduction

Acoustic communication in both human and nonhuman animals often takes place in large social groups, such as cocktail parties, choruses, colonies, or crèches (Schwartz and Freeberg 2008). In these social environments, the background “noise” generated by the mixture of acoustic signals from different individuals can be a potent source of auditory masking (reviewed in Brumm and Slabbekoorn 2005). Despite similarity with the human “cocktail party problem,” a phenomenon that describes the difficulty we have following a single conversation in multi-talker social environments (Cherry 1953; Bronkhorst 2000; McDermott 2009), few studies have investigated mechanisms that allow nonhuman animals to solve parallel problems (Hulse 2002; Langemann and Klump 2005; Bee and Micheyl 2008). Likely among these mechanisms is an ability to exploit the spectral, temporal, and spatial relationships between sources of signals and noise (Bee and Micheyl 2008). Human listeners, for instance, experience a release from auditory masking in psychophysical speech recognition tasks when “speech-shaped noise” (i.e., masking noise with the long-term spectrum of speech) fluctuates in amplitude (e.g., Gustafsson and Arlinger 1994; Bacon et al. 1998; Nelson et al. 2003) and originates from a location different from that

Communicated by: J. Christensen-Dalsgaard

A. Vélez (✉) · M. A. Bee
Department of Ecology, Evolution, and Behavior,
University of Minnesota,
100 Ecology, 1987 Upper Buford Circle,
St. Paul, MN 55108, USA
e-mail: velez011@umn.edu

of the target speech (e.g., Shinn-Cunningham et al. 2001; Noble and Perrett 2002). Few studies have explicitly tested the general hypothesis that similar mechanisms operate in the acoustic communication systems of nonhuman animals.

Anuran amphibians (frogs and toads) represent one taxonomic group for which acoustic signal perception in multi-source environments directly impacts evolutionary fitness. In many species, males aggregate in suitable breeding habitats and form choruses in which they produce loud advertisement calls to attract mates (reviews in Gerhardt and Huber 2002; Wells 2007). Advertisement calls are often necessary and sufficient for species recognition and mate choice by females. In addition, females can discriminate among potential conspecific mates based on individual differences in advertisement calls, and discrimination can influence female fitness (Welch et al. 1998). The auditory systems of frogs typically exhibit species-specific tuning to audio frequencies near those present in each species' vocal repertoire (Capranica and Moffat 1983; Gerhardt and Schwartz 2001; Gerhardt and Huber 2002). Thus, the sounds generated in a dense conspecific chorus represent a prominent source of auditory masking that can constrain signal detection, recognition, and discrimination (Gerhardt and Klump 1988a; Narins and Zelick 1988; Wollerman 1999; Schwartz et al. 2001; Wollerman and Wiley 2002; Bee 2008a, b). A fundamental question, then, concerns the extent to which anuran auditory systems may be adapted to cope with such constraints by exploiting the spectro-temporal and spatial features of the acoustic environment.

An important feature of the ambient background noise in a frog chorus is that it fluctuates in amplitude over time (Fig. 1). There are at least three physical causes that contribute to the presence of these amplitude fluctuations. First, the periodicity inherent in the production of repeated and temporally discontinuous acoustic signals can create low-frequency modulations (e.g., <5–10 Hz) that ultimately arise from the call timing behavior of the individuals comprising the chorus (e.g., Nelken et al. 1999). A second and well-known source of low-frequency modulations (e.g., <20 Hz) in ambient noise involves the impacts of the transmission medium (e.g., turbulent air) on sound propagation (Wiley and Richards 1978; Richards and Wiley 1980). Finally, many anuran advertisement calls have periodic amplitude modulations comprising series of discrete pulses that are commonly repeated at rates between 10 and 60 pulses/s (Gerhardt and Huber 2002). Together, these sources of amplitude modulation result in modulation spectra for chorus sounds that can be multi-modal and that differ among species (Fig. 1).

In this study of Cope's gray treefrog (*Hyla chrysoscelis*), we investigated the effects of amplitude modulations in ambient chorus-like noise on the recognition of conspecific advertisement calls. Males of this species produce a pulsed

Fig. 1 Modulation spectra illustrating the temporal fluctuations characteristic of the choruses of different frog species. Each row of the figure has four separate plots for each species showing the following: (*far left*) a 1.2-s waveform of a portion of a single call or an entire single call from one individual, (*middle left*) a 20-s waveform of one call or a series of calls from one individual, (*middle right*) a 20-s segment of a dense chorus, and (*far right*) the modulation spectrum of the chorus segment depicted in the adjacent plot. Amplitude is plotted as a dimensionless normalized value for the waveforms and as a relative value in dB for the modulation spectra. From *top to bottom* are shown examples for **a** Cope's gray treefrog (*H. chrysoscelis*), **b** boreal chorus frogs (*Pseudacris maculata*), **c** American toads (*Bufo americanus*), **d** spring peepers (*Pseudacris crucifer*), **(e)** green treefrogs *Hyla cinerea*, and **f** North American bullfrogs (*Rana catesbeiana*). In all recordings, the nominal species was the dominant species calling at the time of year and at the field sites at which recordings were made. Note how all species depicted here exhibit peaks in their modulation spectra below 5–10 Hz, whereas only those depicted in **a–c** exhibit secondary peaks corresponding to the rates of pulses in the advertisement call (depicted in the *far left plot*). The modulation spectra were generated in Matlab v7.6 by first extracting the Hilbert envelope of the waveform. To correct for the DC offset, we subtracted the mean value of the envelope from each sample of the envelope. We then calculated the fast-Fourier transform of the corrected Hilbert envelope of the waveform (sampling rate=11,025 samples/s, Hamming window size=65,536 points, overlap=25%) and normalized to the maximum value of the magnitude of the FFT. Finally, we converted the magnitude of the FFT to dB ($20\log_{10}(\text{magnitude})$) and smoothed the modulation spectra by using a running average of 11 points. All recordings were made with high-quality audio recorders (e.g., HHb PortaDAT PDR 1000, Marantz PMD 670) and microphones (Sennheiser ME62, ME66, ME67). Recordings of individuals were made at distances near 1 m from the male. Recordings of choruses were made near the peak of calling activity for the night at distances between 4 and 10 m from the nearest calling individual. We chose this distance for recording chorus sounds because female frogs may commonly assess multiple males simultaneously while listening at distances of several meters from the nearest males (e.g., Murphy and Gerhardt 2002)

advertisement call with a pulse rate of about 40–50 pulses/s (Gerhardt 2001), and pulse rate is an important species recognition cue for females (Schul and Bush 2002). As illustrated in Fig. 2, the background noise in gray treefrog choruses is characterized by low-frequency modulations (e.g., <5–10 Hz) as well as higher rates of amplitude modulation (\approx 40–50 Hz) that correspond to the pulse rate of the advertisement call.

Our objective was to test the hypothesis that female gray treefrogs experience masking release in the presence of amplitude-modulated “chorus-shaped noise” (i.e., masking noise containing the audio frequencies characteristic of conspecific breeding choruses). We used no-choice phonotaxis tests to measure a signal recognition threshold that is conceptually analogous to the “speech reception threshold” (SRT) measured in human psychoacoustic studies of masked speech perception (see discussion in Bee and Schwartz 2009). Briefly, the SRT in human studies is determined as the minimum signal level necessary to elicit a pre-defined level of correct responses on a speech

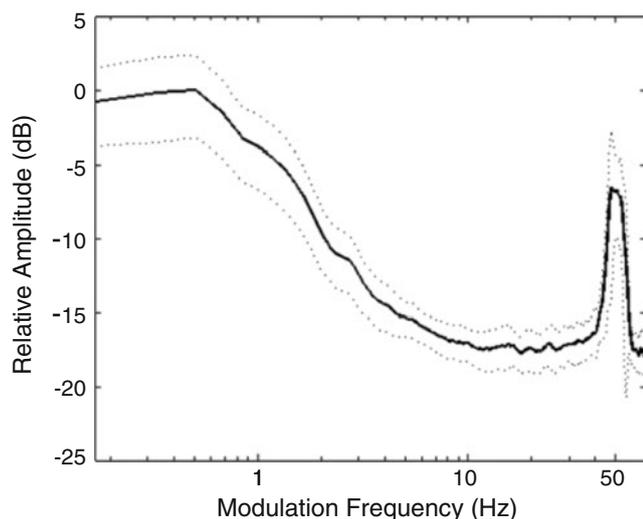


Fig. 2 Modulation spectrum of gray treefrog choruses. Shown here on a logarithmic x axis is the mean (*bold, solid line*) ± 1 standard deviation (*thin, dotted lines*) modulation spectrum determined by averaging the spectra of eight 60-s segments of Cope's gray treefrog choruses. Each segment was taken from a recording of a different chorus recorded in central Minnesota between 1 May and 1 July 2007–2009. For each 60-s segment, we first computed the Hilbert envelope of the waveform and corrected for the DC offset by subtracting the mean value of the envelope from each sample of the envelope. Then, we calculated the fast-Fourier transform of the envelope (sampling rate=11,025 samples/s, Hamming window size=65,536 points, overlap=25%) and normalized the spectrum to the maximum value of the magnitude of the FFT. We then calculated the mean and standard deviation of the modulation spectra of the eight segments, transformed these values to a dB scale ($20\log_{10}(\text{magnitude})$), and smoothed the spectrum with a running average of 11 points. Recordings were made with a Marantz PMD 670 and an omnidirectional Sennheiser ME62 that was positioned 5 cm above the ground at distances ranging between 5 and 10 m from the nearest calling male. A recording position close to the ground was used because females in our populations commonly approach choruses of calling males from such positions

was a synthetic advertisement call with a pulse rate of about 45 pulses/s that was presented at different sound levels in the presence of chorus-shaped noise. The noise was either unmodulated or sinusoidally amplitude-modulated at a low rate (5-Hz SAM) and at a higher rate similar to the pulse repetition rate of the advertisement call (45-Hz SAM). Our prediction was that if females experienced masking release in modulated noise backgrounds, then signal recognition thresholds would be lower in the presence of SAM noise compared with unmodulated noise.

Methods

Subjects

All collections, handling, and testing of animals were approved by the University of Minnesota's Institutional

Animal Care and Use Committee (#0809A46721, November 21, 2008). Nightly collections of gravid females were made between 2100 and 0100 hours in May and June of 2007 and 2008 from wetlands located in the Carver Park Reserve (44°52'49.29"N, 93°43'3.10"W; Carver County, MN, USA) and the Tamarack Nature Center (45°6'9.81"N, 93°2'27.56", Ramsey County, MN, USA). We returned females to the lab and kept them at 2°C to delay egg deposition until they were tested (usually within 24 h). We released females at their original location of capture after testing. In total, 162 females were collected and tested as part of this study. Of these females, 140 met all of our criteria (see below) for inclusion in the datasets used for statistical analyses. Additional descriptions of our field sites and collecting procedures are provided elsewhere (Bee 2007b, 2008a, b; Bee and Swanson 2007; Bee and Schwartz 2009).

General testing procedures

Our testing equipment and general protocols were the same as those described in other recent studies of gray treefrogs, and readers are referred to those studies for additional details not reported here (e.g., Bee 2008a, b; Bee and Schwartz 2009). Briefly, on the day of testing, females were placed in a 20°C incubator where they remained at least 1 h before testing to allow their body temperatures to reach 20°C ($\pm 1^\circ\text{C}$). Phonotaxis tests were conducted at a temperature of $20 \pm 2^\circ\text{C}$ in two temperature-controlled, hemi-anechoic sound chambers (see Bee and Schwartz 2009 for details). Tests were conducted under infrared illumination, and behavioral responses were observed using a video camera mounted from the center of each sound chamber's ceiling. The video feed was simultaneously encoded to MPEG digital files and monitored in real time from outside each chamber. Digital acoustic stimuli (44.1 kHz sampling rate, 16-bit resolution) were broadcast from a computer outside each chamber through a multichannel soundcard, amplified using a multichannel amplifier, and then output to A/D/S L210 speakers (target signals) or Kenwood KFC-1680ie speakers (maskers). The frequency responses of the playback systems were flat (± 3 dB) over the frequency range of interest.

We conducted phonotaxis tests in circular test arenas (2 m diameter) with acoustically transparent but visually opaque walls. The floor of the sound chamber served as the floor of the test arena. The perimeter of the arena was divided into 24 15° arcs. The speaker used to broadcast the target signal was placed on the floor just outside the wall of the arena, centered in one of the 15° arcs, 1 m away from a release point at the center of the arena. We varied the position of the speaker around the arena's perimeter between tests of two to four subjects to eliminate any

possibility of a directional response bias in our sound chambers. The speaker used to broadcast the masking noises was suspended from the ceiling of the chamber 190 cm above the central release point. The overhead speaker created a uniform (± 2 dB) noise level across the floor of the circular arena. Sound levels were measured and calibrated by placing the microphone of a Larson-Davis System 834 or a Brüel and Kjær Type 2250 sound level meter at the approximate position of a subject's head at the central release point. Sound levels were calibrated at the start of each testing day and after each repositioning of the target speaker.

At the beginning of each test, the subject was placed in an acoustically transparent holding cage at the arena's central release point. Stimulus broadcasts began after a 1.5-min silent acclimation period and were continued throughout the duration of a test. After 30 s of signal presentation, we remotely released the subject using a rope and pulley system that could be operated from outside the chamber. In phonotaxis tests in which a masking noise was presented, broadcast of the masker was initiated 30 s before the onset of the target signal and was broadcast continuously over the duration of the test.

Each subject was tested individually in a sequence of tests and was given a 5–15-min timeout period inside the incubator between consecutive tests. A test sequence always began with a “reference condition” and then alternated between two or three consecutive tests of various “treatment conditions” followed by another reference condition, and so on, until all designated treatments had been tested. Each test sequence always ended with a final test of the reference condition. During the reference condition, we broadcast a standard synthetic call (see below) at 85 dB SPL (re. 20 μ Pa, fast RMS, C-weighted) without broadcasting any additional masking noise. This signal level corresponds to a natural call amplitude measured at 1 m (Gerhardt 1975). Unless noted otherwise, we scored responses as follows. We scored a “correct response” if the subject touched the wall of the arena inside the 15° arc in front of the speaker that was broadcasting the target signal within 5 min of being released. We scored a “no response” in a treatment condition if the female failed to meet our response criterion in that condition, but responded during all of the reference conditions. Any subject that failed to respond in a reference condition was excluded from further testing and statistical analyses. We also excluded a subject from statistical analyses if its latency to respond in the final reference condition was more than twice that in the first reference condition. These procedures ensure the validity of no responses in treatment conditions by confirming that females remain responsive over the duration of the test sequence (Bush et al. 2002; Schul and Bush 2002).

Acoustic stimuli

The standard call

We used a standard synthetic call (Fig. 3a) as the target signal in all reference conditions and in several of our

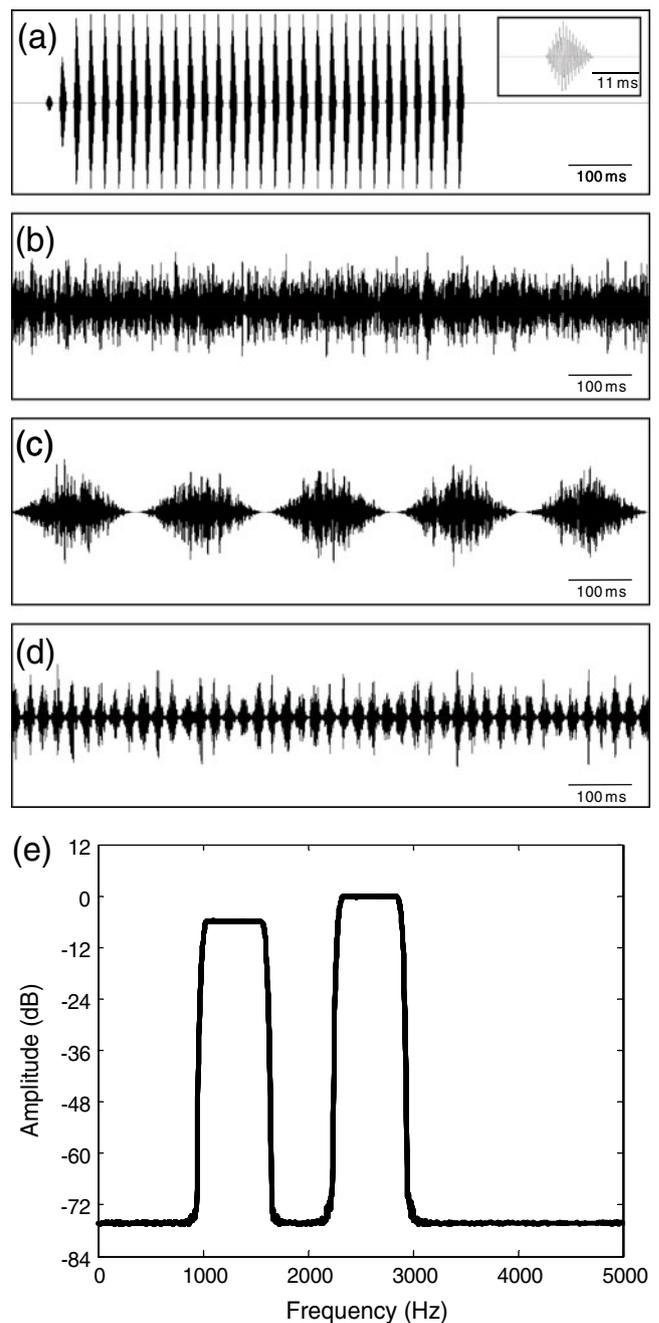


Fig. 3 Standard call and chorus-shaped noises. **a** Waveform of the synthetic *H. chrysocelis* standard call with an inset showing the waveform of a single pulse; **b–d** waveforms of the unmodulated masker (**b**), the 5-Hz SAM masker (**c**), and the 45-Hz SAM masker (**d**); **e** power spectrum showing the spectral profile of the chorus-shaped noises

treatment conditions. The standard call was synthesized using custom-made software (courtesy of J. J. Schwartz) and had values of spectral and temporal properties close to the averages of calls recorded in local Minnesota populations (corrected to 20°C; M. A. Bee unpublished data). The call consisted of 30 pulses (11-ms pulse duration) delivered at a rate of 45.5 pulses/s (22-ms pulse period). Each pulse consisted of two harmonically related, phase-locked sinusoids with frequencies (and relative amplitudes) of 1.3 kHz (−9 dB) and 2.6 kHz (0 dB). The amplitude envelope of each pulse was shaped with a 4-ms rise time and 7-ms fall time with shapes characteristic of calls from local populations. The first 50 ms of the call was shaped with a linear onset. Within a particular test, the standard call repeated with a period of 5 s, which is within the range of call periods measured in local populations (corrected to 20°C).

Chorus-shaped noises

We used Adobe Audition v1.5 to create three chorus-shaped noises. Each noise had the same long-term spectrum and had acoustic energy at the audio frequencies characteristic of gray treefrog choruses (Fig. 3b–e). An unmodulated noise was created by filtering white noise into two 600-Hz-wide spectral bands centered at 1.3 and 2.6 kHz, with the latter having a relative amplitude that was 6 dB greater. The stop-band attenuation was −80 dB. Two modulated noises were created by multiplying white noise by either a 5 or 45-Hz sinusoid with a DC offset that resulted in a modulation depth of 100%. These modulated white noises were then filtered to create 5- and 45-Hz SAM chorus-shaped noises having the same long-term frequency spectrum as the unmodulated noise. We used four different frozen-noise exemplars of the unmodulated and SAM maskers; for the latter, each exemplar had a different starting phase (0°, 90°, 180°, or 270°). Equal numbers of subjects were tested with each exemplar. We explored the use of starting phase as a between-subjects factor in our statistical analyses, but it was never significant, and so was dropped from the final models reported below. In all playback tests, the RMS amplitude of all three maskers was set to a sound pressure level of 73 dB at the subject release site by calibrating the long-term equivalent noise level (LC_{eq}) over at least 1 min. This sound level falls within the range of background noise levels that we and others have recorded in natural *H. chrysocelis* choruses (Schwartz et al. 2001; Swanson et al. 2007; Vélez and Bee unpublished data).

Experiment 1: chorus-shaped noise as a potential signal

Some frogs, including our study species, show positive phonotaxis toward the natural sounds of a chorus, suggesting that chorus “noise” can actually function as a biologically

relevant “signal” for localizing breeding aggregations (Gerhardt and Klump 1988b; Bee 2007a; Swanson et al. 2007). The efficacy of chorus sounds—and by extension, our chorus-shaped noises—as an attractive signal could potentially confound results from studies of masked signal recognition in the presence of chorus-shaped noise. One way that such a confound could be introduced into the data would be if different noises varied in their relative attractiveness to females. To evaluate this possibility, we performed a control experiment in which the three chorus-shaped noises were presented to subjects as potential target signals.

Each subject ($N=20$) was tested in a sequence comprising an initial reference condition followed by three treatment conditions and a final reference condition. Recall that the standard call (see above) was the target signal during the reference conditions. During each treatment condition, one of the three chorus-shaped noises (unmodulated, 5-Hz SAM, or 45-Hz SAM) was presented as the target signal. The order of the three noises across treatment conditions was randomized for each subject. Each masker was broadcast continuously during the treatment condition from a speaker located on the floor just outside the wall of the test arena. In a previous study conducted in our laboratory (Swanson et al. 2007), similar procedures successfully elicited phonotaxis from female gray treefrogs in response to broadcasts of the sounds of real choruses.

We required subjects to touch the arena wall in the 15° arc in front of the target speaker during the reference conditions. Following Swanson et al. (2007), we ended each treatment condition as soon as a subject touched the wall anywhere in the arena. We used circular statistics (V tests; Zar 1999) to test the null hypothesis that the angles at which subjects first touched the arena wall were uniformly distributed around the arena. The alternative hypothesis was that responses were oriented in the direction of the target speaker broadcasting the chorus-shaped noise. For these analyses, we designated the position of the target speaker as 0° and used a significance criterion of $\alpha=0.05$.

Experiment 2: signal recognition thresholds in modulated chorus-shaped noise

We estimated “signal recognition thresholds” (Bee and Schwartz 2009) by presenting the standard call at various signal-to-noise ratios (SNRs) in the presence of chorus-shaped noise. Following Bee and Schwartz (2009), we operationally defined signal recognition as occurring when females exhibited phonotaxis with respect to the standard call. We operationally defined the signal recognition threshold as the minimum signal level required to elicit phonotaxis behavior exceeding a pre-determined criterion level of response. We describe these threshold criteria in

more detail in subsequent sections. Our estimates of signal recognition thresholds for a particular masking condition are based on pooling data from the entire group of subjects tested in that condition. Hence, we regard these estimates as “population-level thresholds” (Bee and Schwartz 2009). This method of threshold estimation differs from those used in traditional psychoacoustic experiments (e.g., adaptive tracking) for estimating thresholds for individual subjects (Klump et al. 1995). We recently showed, however, that population-level thresholds estimated using the methods described below are similar to those estimated using an adaptive tracking procedure (Bee and Schwartz 2009).

Experimental design

We tested 120 females using a 4 masking condition (within subjects) × 6 SNR (between-subjects) factorial design. The target signal was the standard call. In three of the four masking conditions, we broadcast either the unmodulated, 5-Hz SAM, or 45-Hz SAM chorus-shaped noises from the overhead speaker; the fourth condition was a “no-masker” condition in which no masking noise was broadcast. This no-masker condition served as a control to assess the effects of our unmodulated and SAM maskers on subjects’ responses to the target signal. The level of the masking noises was fixed at 73 dB SPL (at the central release site). We tested five signal levels (61, 67, 73, 79, and 85 dB SPL) that corresponded to SNRs of −12, −6, 0 +6, and +12 dB. In the no-masker condition, the target signal was broadcast at the same sound pressure level required to realize the nominal SNR. As one additional level of the SNR factor, we included a “no-signal” condition, in which we muted the audio channel for the target signal so that no signal was broadcast. Different groups of 20 subjects were tested at each SNR. Individual subjects were tested in a sequence comprising three reference conditions and four treatment conditions (one for each masking condition). Subjects were randomly assigned to a SNR, and the order of the treatment conditions was randomized separately for each subject.

The no-signal condition deserves additional comment, as it was included to address two specific issues. First, at the factorial combination of the no-signal and no-masker conditions, we tested subjects in our arena without broadcasting any sounds. This allowed us to estimate a false alarm rate for our response criterion by assessing how frequently subjects touched the wall in the 15° bin centered on the silent target speaker within 5 min. Second, having a no-signal condition crossed with each of the other masking conditions allowed us to assess the extent to which subjects might have behaved differently in the test arena depending on the type of masking noise (unmodulated, 5-Hz SAM, or 45-Hz SAM) that was broadcast from the overhead speaker. For instance, subjects could have behaved differently in the presence of one of the

maskers and in ways that affected their responses to the standard call, such as exhibiting less overall movement (e.g., waiting and listening) or more directionally varied movements (e.g., increased searching behavior). Thus, the no-signal conditions served as additional controls that allowed us to assess subject behavior while in the presence of chorus-shaped noise and in the absence of the standard call.

Behavioral response measures

Thresholds based on angular orientation From video analyses of phonotaxis tests, we assessed the directedness of phonotaxis toward the target signal by measuring the angle (relative to the target speaker at 0°) at which subjects first exited a circle of 20-cm radius centered on the release cage. Following Bee and Schwartz (2009), we chose a distance of 20 cm as a compromise between analyzing the angles at which subjects exited the release cage and the angles at which they first touched the arena wall 1 m away. Our rationale was as follows. Subjects in our testing apparatus sometimes exit the release cage in one direction and then quickly reorient and initiate movement in a different direction while still physically located immediately adjacent to the release cage. Subjects typically do not make multiple reorientation movements while positioned within 20 cm of the release cage. Thus, we believe that measuring angular orientation upon exiting our release cage is not an entirely reliable measure of the subject's directed movements. However, allowing subjects to freely move about over the entire arena floor potentially introduces spatial cues that could influence estimates of signal recognition thresholds. Restricting the measurement distance to 20-cm minimizes any cues related to the variation in SNRs experienced by moving about in the sound field. According to both our own empirical measurements in the sound chambers and the inverse square law, moving 20 cm closer to a source originally located 1 m away results in a gain in signal level that is less than 2 dB, which is less than the 6-dB step-size we used between adjacent signal levels.

We used circular statistics (V tests; Zar 1999) to test the null hypothesis that angles at 20 cm from the release point were uniformly distributed against the alternative hypothesis that subjects oriented toward the target signal (0°). We estimated an upper threshold bound as the lowest SNR at which subjects exhibited significant orientation toward the target signal at that SNR and also at all higher SNRs. We estimated a lower threshold bound as the next lowest SNR. We then computed the signal recognition threshold as the average of the upper bound (UB) and lower bound (LB) using the following equation:

$$\text{signal recognition threshold} = 10 \log_{10} \left(\frac{10^{\text{UB}/10} + 10^{\text{LB}/10}}{2} \right) \quad (1)$$

Thresholds based on response probabilities Following Bee and Schwartz (2009), we also estimated signal recognition thresholds based on the proportion of subjects that met our response criterion of touching the arena wall in the 15° arc in front of the speaker within 5 min. We estimated an upper threshold bound as the lowest SNR at which the proportion of subjects that met our response criterion was significantly greater than 0.20 (one-tailed binomial tests) at that SNR and also at all higher SNRs. We used a null expectation of 0.20 because we empirically determined that 10–20% of subjects met our response criterion even when no target signal was presented (see below). The next lowest SNR below the upper bound was taken as the lower bound, and signal recognition thresholds were estimated using Eq. 1.

Movement patterns To assess the possibility that phonotaxis behavior was directly influenced by differences between the three types of chorus-shaped noise, we used the animal tracking software EthoVision[®] v3.1 (Noldus 2005) to analyze patterns of subject movement in the no-signal conditions. We measured the total distance (in cm) that subjects moved during a test and the average velocity (in cm/s) of their movements. We measured two additional behaviors potentially related to sound localization (Rheinlaender and Klump 1988). These included the average absolute turn angles (in degrees) associated with movements greater than 1.0 cm during a phonotaxis test and a second measure called “meander.” The latter quantifies (in degrees/cm) the magnitude of changes in the direction of movements relative to the distance moved (Noldus 2005). We compared these response measures using repeated measures multivariate analysis of variance (MANOVA).

Results

Experiment 1: chorus-shaped noise as a potential signal

We found no indication that subjects treated the artificial chorus-shaped noises as behaviorally relevant signals. In contrast to how female gray treefrogs respond to recordings of natural choruses (see Swanson et al. 2007), subjects in this experiment did not exhibit phonotaxis toward unmodulated and SAM chorus-shaped noises (Fig. 4). Nevertheless, subjects were clearly motivated to respond during this experiment, as evidenced by their uniformly strong orientation toward the standard call in the reference conditions that preceded and followed the three treatment conditions (Fig. 4). The mean (\pm SD) response latency in the reference conditions was 77.8 ± 23.5 s, and latencies did not differ between the two reference conditions (paired-sample *t* test: $t = -0.58$, $P = 0.5688$). The results of experiment 1 thus

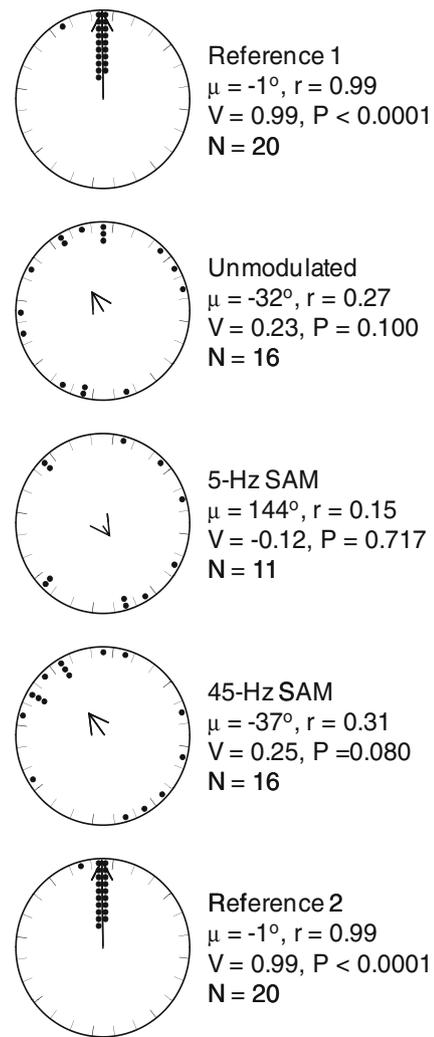


Fig. 4 Chorus-shaped noise as a potential signal. Points depict the angles at which females first touched the wall of the arena relative to the position of the target speaker (top of each circle) in the two reference conditions, and in response to the unmodulated, 5-Hz SAM, and 45-Hz SAM chorus-shaped noises. Also shown are descriptive circular statistics for the mean vector (μ) and the length of the mean vector (r), and the results of V tests of the null hypothesis that angles were uniformly distributed. The direction and length of each arrow depict the mean vector angle (μ) and the length of the mean vector (r), respectively

confirmed that the three chorus-shaped noises were not attractive to females. Therefore, any differences in signal recognition thresholds in the presence of these three noises in experiment 2 could not be attributed to a confound related to competition between an attractive standard call and attractive chorus-shaped noises.

Experiment 2: signal recognition thresholds in modulated chorus-shaped noise

Subjects tested in experiment 2 also remained motivated to respond over the entire duration of the test sequence, as

evidenced by their consistently strong orientation toward the target signal in the three reference conditions (Table 1). The mean response latency in the reference conditions, averaged across all conditions and subjects, was 83.4 ± 31.2 s. There were no significant differences in latency across the three reference conditions (ANOVA: $F_{2,228}=0.2$, $P=0.8310$). There were also no significant differences in latency between the six groups of subjects tested at different SNRs (ANOVA: $F_{5,114}=0.5$, $P=0.7625$), nor were there any significant interactions between SNR and the repeated measure of reference condition (ANOVA: $F_{10,228}=1.7$, $P=0.0992$).

In the no-masker condition, subjects exhibited significant orientation toward the speaker at all SNRs that included broadcasts of the target signal (Fig. 5). In addition, response probabilities were above 0.80 at all signal levels (all P values <0.05 in one-tailed binomial tests of the hypothesis that $P>0.20$). Hence, signal recognition thresholds could not be calculated for this condition. Elsewhere, we and others have shown that signal recognition thresholds are in the range of about 35–45 dB in response to synthetic

advertisement calls presented without masking noise (Beckers and Schul 2004; Bee and Swanson 2007; Bee and Schwartz 2009). In contrast to these results in the no-masker condition, we were able to estimate signal recognition thresholds for all three of the conditions with masking noise based on angular orientation and response probabilities.

Threshold estimates based on angular orientation

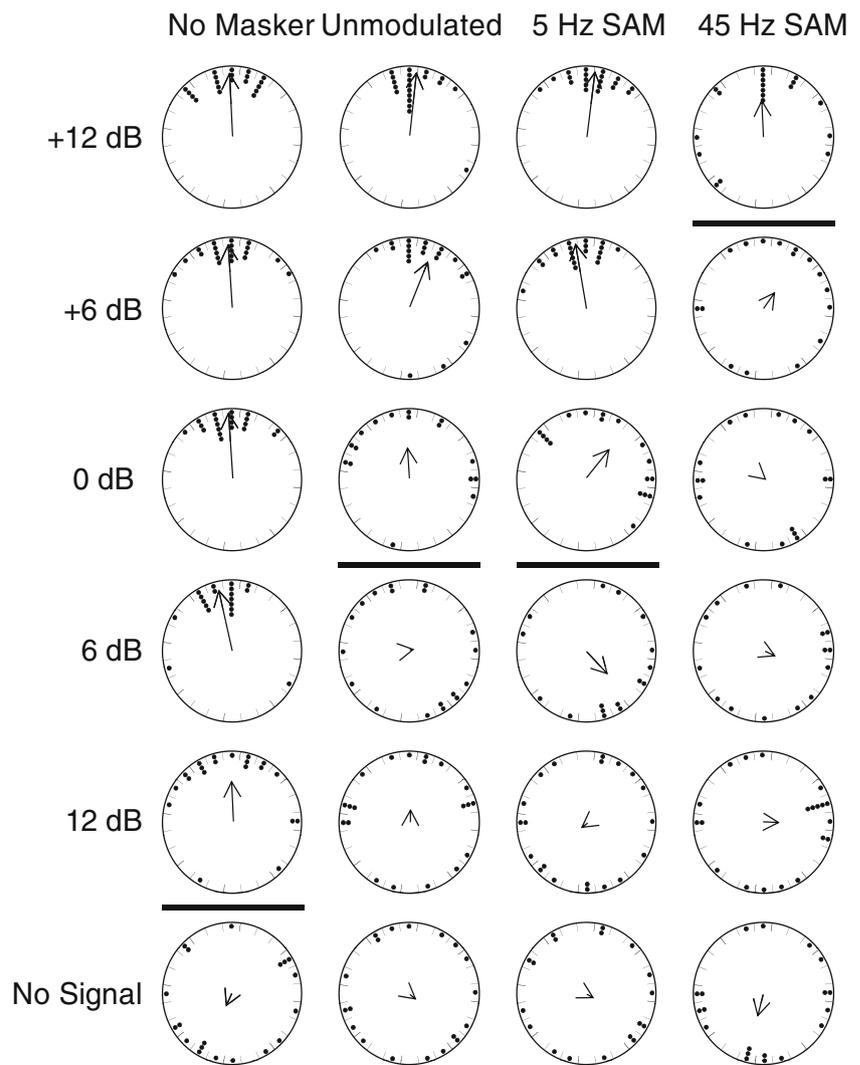
In the no-signal conditions, there was no evidence that subject movements were oriented toward the silent speaker when they reached a point located 20 cm away from the central release point (Fig. 5; Table 1). Subjects oriented toward the signal at relatively lower SNRs in the presence of the unmodulated and 5-Hz SAM maskers than in tests conducted with the 45-Hz SAM masker (Fig. 5; Table 1). In both the unmodulated and 5-Hz SAM conditions, we found significant orientation at SNRs of 0 dB and higher, but not at SNRs of -6 dB and lower. Using 0 dB as the upper bound and -6 dB as the lower bound, we calculated a

Table 1 Results of circular statistical analyses for response angles at 20 cm in the three reference conditions and in the four masking conditions as a function of the signal-to-noise ratio (SNR; see text for additional details)

Condition	SNR	Mean vector (μ°)	Length of mean vector (r)	Circular SD ($^\circ$)	N	V	P
Reference 1	–	1	0.94	21	120	0.94	<0.0001
Reference 2	–	2	0.92	24	120	0.92	<0.0001
Reference 3	–	–2	0.92	24	120	0.92	<0.0001
No masker	+12 dB ^a	–3	0.89	27	20	0.89	<0.0001
	+6 dB ^a	–4	0.89	27	20	0.89	<0.0001
	0 dB ^a	–4	0.92	23	20	0.92	<0.0001
	–6 dB ^a	–13	0.87	38	19	0.79	<0.0001
	–12 dB ^a	–3	0.57	60	20	0.57	<0.0001
	No signal	–156	0.20	103	19	–0.18	0.8690
Unmodulated	+12 dB ^a	6	0.89	28	20	0.86	<0.0001
	+6 dB ^a	22	0.68	50	20	0.63	<0.0001
	0 dB ^a	–4	0.43	75	16	0.43	0.007
	–6 dB	78	0.06	136	18	0.01	0.471
	–12 dB	0	0.18	106	20	0.18	0.1340
	No signal	127	0.12	118	19	–0.07	0.6700
5-Hz SAM	+12 dB ^a	7	0.92	23	20	0.92	<0.0001
	+6 dB ^a	–10	0.91	25	20	0.89	<0.0001
	0 dB ^a	38	0.51	66	17	0.40	0.009
	–6 dB	136	0.39	79	16	–0.28	0.943
	–12 dB	–127	0.11	121	20	–0.06	0.655
	No signal	117	0.12	118	19	–0.06	0.6320
45-Hz SAM	+12 dB ^a	–3	0.51	67	19	0.50	<0.001
	+6 dB	35	0.29	90	16	0.24	0.0910
	0 dB	120	0.02	160	16	–0.01	0.5220
	–6 dB	113	0.17	108	15	–0.07	0.6420
	–12 dB	92	0.23	97	20	–0.01	0.5170
	No signal	–166	0.31	88	19	–0.3	0.968

^a Statistically significant orientation in the masking conditions ($\alpha=0.05$)

Fig. 5 Angular orientation in response to synthetic calls presented in the presence or absence of chorus-shaped noise. Points depict the angles at which individual females first left a circle with radius 20 cm centered on the release point at the center of the test arena. Data are shown for the 24 factorial combinations of six signal-to-noise ratios and four masking conditions. The position of the target speaker was designated as 0° and corresponds to the top of each circular graph. The direction and length of each arrow depict the mean vector angle (μ) and the length of the mean vector (r), respectively. In each noise condition, significant orientation was observed at all SNRs above the horizontal line in each column; for the three conditions with masking noise, the horizontal line separates the upper and lower bounds used to estimate signal recognition thresholds. (See Table 1 for statistical results.)



signal recognition threshold of -2 dB for these two conditions. In the 45-Hz SAM condition, significant orientation was found only at the highest SNR of $+12$ dB. Assuming that significant orientation also would have occurred at even higher SNRs in the presence of the 45-SAM masker, we used $+12$ and $+6$ dB as the upper and lower bound SNRs, respectively, and calculated a signal recognition threshold of 10 dB for this condition. Hence, based on measures of angular orientation, our estimates of signal recognition thresholds were 12 dB higher in the 45-Hz SAM condition than in both the unmodulated and 5-Hz SAM conditions.

Threshold estimates based on response probabilities

In the no-signal conditions, 10–20% of subjects touched the wall of the arena in front of the silent speaker within 5 min (no masker, four of 20; unmodulated, two of 20; 5-Hz SAM, four of 20; 45-Hz SAM, three of 20). There was no significant

difference in the proportion of subjects exhibiting these “false alarms” across the four masking conditions (Cochran’s Q test, $Q=1.00$, $df=3$, $P=0.8013$). We used the proportion of subjects that touched the arena wall in front of the silent speaker in the factorial combination of the no-signal and no-masker conditions ($P=0.20$) as an estimate of a false-alarm rate for detecting a correct response from subjects using our testing methods. The proportion of subjects meeting the response criterion was significantly greater than this false alarm rate at SNRs of -6 dB and higher in both the unmodulated and 5-Hz SAM masking conditions (Table 2). We used SNRs of -6 and -12 dB for the upper and lower bounds, respectively, and computed a signal recognition threshold of -8 dB for the unmodulated and 5-Hz SAM conditions. In the 45-Hz SAM condition, a proportion of subjects significantly greater than 0.20 responded at SNRs of $+6$ and $+12$ dB, but not at lower SNRs (Table 2). Using $+6$ and 0 dB as the upper and lower bounds, respectively, we estimated a signal recognition threshold of 4 dB for the 45-

Table 2 Proportions (p) of subjects exhibiting correct responses as a function of SNR in the four masking conditions and results from one-tailed binomial tests of the hypothesis that $P > 0.20$

SNR	No masker		Unmodulated		5-Hz SAM		45-Hz SAM	
	p	Binomial P	p	Binomial P	p	Binomial P	p	Binomial P
-12 dB	0.85	<0.0001	0.05	0.9884	0.25	0.3703	0.30	0.1957
-6 dB	0.95	<0.0001	0.55	0.0005	0.45	0.0099	0.30	0.1957
0 dB	1.00	<0.0001	0.55	0.0005	0.70	<0.0001	0.35	0.0867
+6 dB	1.00	<0.0001	1.00	<0.0001	1.00	<0.0001	0.60	0.0001
+12 dB	1.00	<0.0001	1.00	<0.0001	1.00	<0.0001	0.90	<0.0001

Hz SAM masking condition. Hence, based on measures of response probabilities, estimates of signal recognition thresholds were 12 dB higher in the 45-Hz SAM condition than in both the unmodulated and 5-Hz SAM conditions, for which thresholds were again similar.

Movement patterns

To assess the possibility that subjects behaved differently depending both on whether or not a masker was presented from the overhead speaker, and on which of the three maskers was presented, we analyzed videos of movements for 17 subjects during tests of the no-signal conditions. (Three subjects were excluded either because they did not leave the release cage during tests of a masking condition, or videos for one or more masking conditions were unavailable due to software encoding errors that occurred during the tests.) We found no significant difference across the four masking conditions (Fig. 6) based on comparing mean values of total distance moved, velocity of movement, turn angle, and meander in a repeated measures MANOVA (Wilks' $\lambda=0.40$, $F_{12,5}=0.6$, $P=0.7652$). Subsequent univariate tests also failed to reveal differences in each of these behavioral response measures (see Fig. 6). Thus, there was little evidence to suggest that the threshold differences reported above were somehow an artifact of differences in how subjects behaved in the presence of the different masking noises.

Discussion

Two important and related consequences of auditory masking in noisy social environments can impact evolutionary fitness: (1) increased potential for communication errors (e.g., missed detection or incorrect classification) and (2) reduced signal active space (Wiley 1994, 2006; Brumm and Slabbekoorn 2005; Langemann and Klump 2005). We should generally expect natural selection to favor mechanisms that function to ameliorate these consequences. Two such mechanisms that improve human speech perception in noise involves exploiting spatial separation between signals

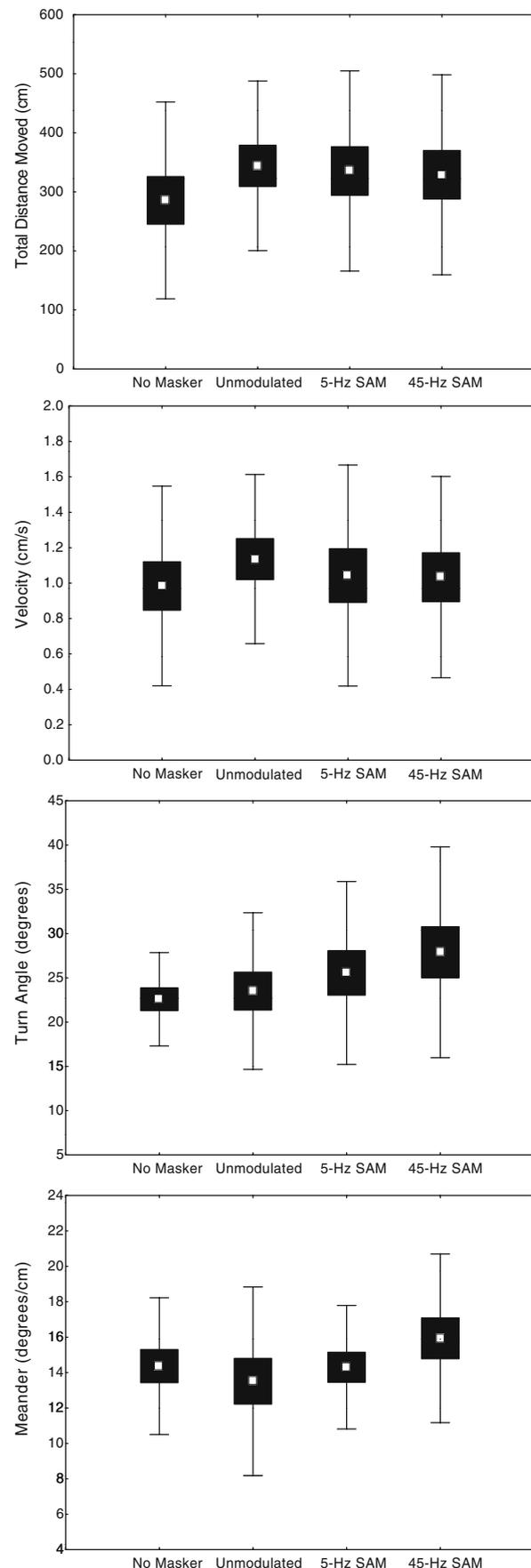
and noise and the fluctuating amplitude of speech-like masking sounds (e.g., Festen and Plomp 1990; Bronkhorst and Plomp 1992; Gustafsson and Arlinger 1994). We have previously reported that female gray treefrogs experience spatial unmasking when there is physical separation between a source of advertisement calls and sources of unmodulated chorus-shaped noise (Bee 2007b, 2008a). Our aim here was to extend these earlier findings by testing the hypothesis that females also experience masking release in temporally fluctuating noise.

According to the masking release hypothesis, we predicted that signal recognition thresholds would be lower in the presence of chorus-shaped maskers that were modulated at rates of 5 Hz, 45 Hz, or both, when compared with those measured in unmodulated noise. Our results are inconsistent with this prediction. Under the conditions tested in this study, we found little evidence that female gray treefrogs experienced masking release in fluctuating chorus-shaped maskers when compared with an unmodulated noise background. We found instead that signal recognition thresholds in the unmodulated and 5-Hz SAM conditions were the same, and those in the 45-Hz SAM condition were 12 dB higher. These patterns were consistent when thresholds were estimated using data for both angular orientation or response probabilities. The relatively higher signal recognition thresholds in the 45-Hz SAM condition did not result from among-treatment differences in the relative attractiveness of the standard call and the chorus-shaped noise. As demonstrated in experiment 1, none of the chorus-shaped noises were attractive to females. Nor did differences in threshold result from any response biases introduced because subjects behaved differently in the 45-Hz SAM condition compared with the other masking conditions, as evidenced by similar patterns of movement across the no-signal conditions of experiment 2 (Fig. 6). From these data, we can conclude that masking release in temporally fluctuating noise had little influence on signal recognition thresholds in gray treefrogs under the conditions tested in the present study. Instead, a rate of modulation in the masker that was similar to the pulse rate of the call (40–50 Hz) impaired signal recognition beyond that caused by the unmodulated masker.

Fig. 6 Patterns of movement in the no-signal conditions. Each plot depicts the mean (point), ± 1 SE (box), and ± 1 SD (whiskers) for one of the movement variables across the four different masking conditions. From top to bottom, the results are shown for total distance moved (univariate ANOVA: $F_{3, 48}=1.01$, $P=0.3855$), velocity (univariate ANOVA: $F_{3, 48}=0.5$, $P=0.6762$), turn angles (univariate ANOVA: $F_{3, 48}=1.4$, $P=0.2624$), and meander (univariate ANOVA: $F_{3, 48}=1.0$, $P=0.3982$)

Our results closely parallel those of Ronacher and Hoffmann (2003), who investigated the extent to which temporally fluctuating noise affected the ability of male grasshoppers (*Chorthippus biguttulus*) to recognize the stridulatory signals of females. These signals comprise a series of pulsed syllables. Each syllable is repeated at a rate of about 10 times per second and contains several pulses that are produced at a rate of about 70 pulses/s (Ronacher and Krahe 1998; Ronacher and Hoffmann 2003). Like many frogs, the temporal structure of the signal is critically important for sound pattern recognition in *C. biguttulus*. Ronacher and Hoffmann (2003) found that signal recognition was impaired when the modulation frequencies in the masker were most similar to those present in the signal. Compared to an unmodulated noise condition, SAM maskers that fluctuated at rates slower than the modulations present in the signal (1.5, 2.5, and 5 Hz) had little effect on signal recognition. In contrast, higher modulation rates (15, 50, 70, and 150 Hz) significantly impaired recognition relative to that in the unmodulated condition. Notably, there was a steep decline in recognition between the 5- and 15-Hz modulation frequencies, which encompasses the slower rate of amplitude modulation in the signal. Together, these results suggest that masking release in temporally fluctuating noise had little influence on signal recognition in *C. biguttulus*. Instead, these findings are consistent with the hypothesis that similar modulation rates in signals and noise can result in greater masking than that elicited by an unmodulated masker.

Our results and those of Ronacher and Hoffmann (2003) contrast with previous studies of human speech perception that have investigated the influences of temporal fluctuations in masking noise. Most of these studies have demonstrated improvements in masked speech understanding when the maskers are modulated compared to an unmodulated masker presented at the same level (e.g., Festen and Plomp 1990; Bronkhorst and Plomp 1992; Takahashi and Bacon 1992; Gustafsson and Arlinger 1994; Bacon et al. 1998; Nelson et al. 2003; Füllgrabe et al. 2006). It is generally believed that human listeners can exploit periods of low amplitude in fluctuating maskers to improve the detection and recognition of speech signals (e.g., so-called dip listening; Buus 1985). Generally, the longer the dip, the greater the improvement in performance compared to unmodulated maskers. Masking release for the



detection of simpler signals (e.g., pure tones) in temporally fluctuating noise has also been demonstrated in humans, nonhuman mammals, and songbirds (reviewed in Langemann and Klump 2005).

In the present study, the period of the slower fluctuating masker (i.e., the 5-Hz SAM noise) was 200 ms, and the duration of the dip (measured at the 6-dB down points) was 100 ms. Given the pulse durations and inter-pulse intervals of 11 ms in the standard call, the maximum number of complete pulses that would occur during a 100-ms dip is about 5 pulses. We do not presently know the lowest number of pulses that elicit phonotaxis from females of Cope's gray treefrog. In the eastern gray treefrog, in which calls have, on average, about 18–20 pulses, a call with just three pulses elicits almost no phonotaxis, whereas a call with 6 pulses elicits a phonotaxis response that is only about 45% of the strength of that elicited by an 18-pulse call (Bush et al. 2002). In addition, Alder and Rose (1998) showed that stimuli comprising at least 8 pulses were necessary to elicit a response from “pulse-integrator” neurons in the midbrains of the Pacific treefrog (*Pseudacris regilla*) and the leopard frog (*Rana pipiens*) and suggested that temporal integration of several pulses was necessary for signal recognition. We hypothesize that hearing only five pulses in a “dip” may be an insufficient number for females of Cope's gray treefrog to realize any benefit of dip listening on signal recognition in a masker modulated at a rate of 5 Hz.

Instead of masking release, our data suggest that our subjects experienced “modulation masking” (Bacon and Grantham 1989) when the target signal and masking noise were modulated at the same rates. Psychophysical studies of modulation masking in humans have shown that amplitude-modulated maskers can, under some conditions, impair perception of modulated sounds (Bacon and Grantham 1989, 1992; Millman et al. 2002), including speech (Kwon and Turner 2001). Importantly, the negative effects of modulation masking are most pronounced when signals and maskers have similar modulation rates (Bacon and Grantham 1989).

We believe findings from human studies of modulation masking are relevant to the interpretation of our results. In many species of frogs and insects (Gerhardt and Huber 2002), gross temporal properties of the amplitude envelopes of signals are critical for sound pattern recognition. In Cope's gray treefrog, for example, the rate of pulses in the male advertisement call is an important temporal property that mediates species recognition (Schul and Bush 2002). Temporal overlap between the calls of two nearby males interferes with call recognition by female gray treefrogs and is largely due to the disruption of a female's perception of the pulsed structure of the call (Schwartz 1987; Schwartz and Gerhardt 1995; Marshall et al. 2006; Schwartz and Marshall 2006). Our findings extend these earlier studies by

showing that the high-rate (i.e., 40–50 Hz) amplitude modulations present in the ambient noise of a chorus could also interfere with perception of the pulsed structure of advertisement calls and thus impair call recognition by females. Within the acoustic scene of a breeding chorus, modulation masking by the ambient background noise could be a persistent problem for communication.

Previous studies of modulation masking in humans have revealed two additional findings that are relevant to our study. First, the degree of modulation masking is directly related to the depth of modulation in the masker (Bacon and Grantham 1989). As in several studies of masking release and modulation masking in humans (e.g., Takahashi and Bacon 1992; Kwon and Turner 2001; Füllgrabe et al. 2006) and in the grasshopper study by Ronacher and Hoffmann (2003), we used SAM maskers with 100% modulation depth. Our use of 100% modulation depth was designed to facilitate the most direct comparisons possible with these previous studies. We caution, however, that during periods of active calling in dense gray treefrog choruses, when masking is expected to be most severe, the depth of modulations in the ambient chorus noise would typically not approach 100% (see Fig. 1). It would, therefore, be premature to use findings from this study to attempt to estimate accurately the magnitude of masking (e.g., in dB) or anticipated reductions in signal active space (e.g., in meters) that might occur under more natural listening conditions. Before such estimates would be meaningful, additional research should be conducted using a wider range of modulation rates and depths, as well as both periodic and randomly modulated noises, including natural noises.

Second, the magnitude of modulation masking is inversely related to the duration of the target signal (Millman et al. 2002). In the closely related eastern gray treefrog, *Hyla versicolor*, females have a non-linear, directional preference for average and longer-than-average calls over shorter-than-average calls in two-choice laboratory experiments (Gerhardt et al. 2000) and in natural choruses (Schwartz et al. 2001). In addition, females receive indirect genetic benefits in the form of increased offspring fitness by mating with males that produce longer calls (Welch et al. 1998). We recently showed that females of Cope's gray treefrog have preferences for call duration that parallel those of the eastern gray treefrog (Bee 2008b). Additionally, we showed that this preference was relaxed when calls were broadcast in the presence of an unmodulated chorus-shaped noise. We speculated that the noise of the chorus could constrain the expression of adaptive female preferences. However, if the severity of modulation masking in real choruses decreases with increasing signal duration, then modulation masking might be one mechanism that could actually restore the advantage of longer

over shorter calls and enable females to choose the best males. Future studies should concentrate on elucidating the effect of natural amplitude modulations of the background noise of the chorus on the ability of female frogs to perceive and discriminate between behaviorally relevant signals that differ in pulse number.

In conclusion, we found little evidence to support the masking release hypothesis. Instead, we found that similar rates of amplitude fluctuations in signals and maskers resulted in modulation masking. These findings contrast with most studies of human speech perception in temporally fluctuating maskers, but closely parallel results from a similar study of a grasshopper. We hypothesize that modulation masking could operate as a constraint on acoustic signal perception in the noisy social environment of a chorus that simultaneously provides a relative advantage to signalers producing longer signals.

Acknowledgments We are grateful to J. Berrío, H. Chun, L. Corcoran, D. Heil, J. Henderson, J. Henly, M. Kuczynski, J. Lane, A. Leightner, E. Love, R. Olsen, K. Riemersma, D. Rittenhouse, N. Rogers, K. Speirs, E. Swanson, S. Tekmen, A. Thompson, and J. Walker-Jansen for their assistance in collecting and testing frogs; to S. Humfeld and G. Höbel for recordings of green treefrogs included in Fig. 1; to K. Riemersma and K. Speirs for their extensive help analyzing videos; and to two anonymous reviewers for their helpful feedback on an earlier version of the manuscript. This work was conducted under Special Use Permit #14902 from the Minnesota Department of Natural Resources and with Special Use Permits issued by M. Linck at the Three Rivers Park District and John Moriarty at the Ramsey County Department of Parks and Recreation. This research was approved by the University of Minnesota Institutional Animal Care and Use Committee (#0809A46721) and funded by a Grant-in-Aid of Research from the University of Minnesota Graduate School and NIH R03DC008396 to M. Bee and an EEB Block Grant, a UMN Graduate School Thesis Research Grant, and a Florence Rothman Fellowship to A. Vélez. Preparation of the manuscript was further supported by a fellowship from the McKnight Foundation to M. Bee.

References

- Alder TB, Rose GJ (1998) Long-term temporal integration in the anuran auditory system. *Nat Neurosci* 1:519–523
- Bacon SP, Grantham DW (1989) Modulation masking: effects of modulation frequency, depth, and phase. *J Acoust Soc Am* 85:2575–2580
- Bacon SP, Grantham DW (1992) Fringe effects in modulation masking. *J Acoust Soc Am* 91:3451–3455
- Bacon SP, Opie JM, Montoya DY (1998) The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds. *J Speech Lang Hear Res* 41:549–563
- Beckers OM, Schul J (2004) Phonotaxis in *Hyla versicolor* (Anura, Hylidae): the effect of absolute call amplitude. *J Comp Physiol A* 190:869–876
- Bee MA (2007a) Selective phonotaxis by male wood frogs (*Rana sylvatica*) to the sound of a chorus. *Behav Ecol Sociobiol* 61:955–966
- Bee MA (2007b) Sound source segregation in grey treefrogs: spatial release from masking by the sound of a chorus. *Anim Behav* 74:549–558
- Bee MA (2008a) Finding a mate at a cocktail party: spatial release from masking improves acoustic mate recognition in grey treefrogs. *Anim Behav* 75:1781–1791
- Bee MA (2008b) Parallel female preferences for call duration in a diploid ancestor of an allotetraploid treefrog. *Anim Behav* 76:845–853
- Bee MA, Micheyl C (2008) The cocktail party problem: what is it? How can it be solved? And why should animal behaviorists study it? *J Comp Psychol* 122:235–251
- Bee MA, Schwartz JJ (2009) Behavioral measures of signal recognition thresholds in frogs in the presence and absence of chorus-shaped noise. *J Acoust Soc Am* 126:2788–2801
- Bee MA, Swanson EM (2007) Auditory masking of anuran advertisement calls by road traffic noise. *Anim Behav* 74:1765–1776
- Bronkhorst AW (2000) The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions. *Acustica* 86:117–128
- Bronkhorst AW, Plomp R (1992) Effect of multiple speech-like maskers on binaural speech recognition in normal and impaired hearing. *J Acoust Soc Am* 92:3132–3139
- Brumm H, Slabbekoom H (2005) Acoustic communication in noise. *Adv Study Behav* 35:151–209
- Bush SL, Gerhardt HC, Schul J (2002) Pattern recognition and call preferences in treefrogs (Anura: Hylidae): a quantitative analysis using a no-choice paradigm. *Anim Behav* 63:7–14
- Buus S (1985) Release from masking caused by envelope fluctuations. *J Acoust Soc Am* 78:1958–1965
- Capranica RR, Moffat JM (1983) Neurobehavioral correlates of sound communication in anurans. In: Ewert JP, Capranica RR, Ingle DJ (eds) *Advances in vertebrate neuroethology*. Plenum, New York, pp 701–730
- Cherry EC (1953) Some experiments on the recognition of speech, with one and with two ears. *J Acoust Soc Am* 25:975–979
- Festen JM, Plomp R (1990) Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J Acoust Soc Am* 88:1725–1736
- Füllgrabe C, Berthommier F, Lorenzi C (2006) Masking release for consonant features in temporally fluctuating background noise. *Hear Res* 211:74–84
- Gerhardt HC (1975) Sound pressure levels and radiation patterns of vocalizations of some North American frogs and toads. *J Comp Physiol* 102:1–12
- Gerhardt HC (2001) Acoustic communication in two groups of closely related treefrogs. *Adv Study Behav* 30:99–167
- Gerhardt HC, Huber F (2002) Acoustic communication in insects and anurans: common problems and diverse solutions. Chicago University Press, Chicago
- Gerhardt HC, Klump GM (1988a) Masking of acoustic signals by the chorus background noise in the green treefrog: a limitation on mate choice. *Anim Behav* 36:1247–1249
- Gerhardt HC, Klump GM (1988b) Phonotactic responses and selectivity of barking treefrogs (*Hyla gratiosa*) to chorus sounds. *J Comp Physiol A* 163:795–802
- Gerhardt HC, Schwartz JJ (2001) Auditory tuning, frequency preferences and mate choice in anurans. In: Ryan MJ (ed) *Anuran communication*. Smithsonian Institution Press, Washington DC, pp 73–85
- Gerhardt HC, Tanner SD, Corrigan CM, Walton HC (2000) Female preference functions based on call duration in the gray tree frog (*Hyla versicolor*). *Behav Ecol* 11:663–669
- Gustafsson HA, Arlinger SD (1994) Masking of speech by amplitude-modulated noise. *J Acoust Soc Am* 95:518–529

- Hulse SH (2002) Auditory scene analysis in animal communication. *Adv Study Behav* 31:163–200
- Klump GM, Dooling RJ, Fay RR, Stebbins WC (1995) Methods in comparative psychoacoustics. Birkhäuser, Basel
- Kwon BJ, Turner CW (2001) Consonant identification under maskers with sinusoidal modulation: masking release or modulation interference? *J Acoust Soc Am* 110:1130–1140
- Langemann U, Klump GM (2005) Perception and acoustic communication networks. In: McGregor PK (ed) *Animal communication networks*. Cambridge University Press, Cambridge, pp 451–480
- Marshall VT, Schwartz JJ, Gerhardt HC (2006) Effects of heterospecific call overlap on the phonotactic behaviour of grey treefrogs. *Anim Behav* 72:449–459
- McDermott JH (2009) The cocktail party problem. *Curr Biol* 19:R1024–R1027
- Millman RE, Lorenzi C, Apoux F, Fullgrave C, Green GGR, Bacon SP (2002) Effect of duration on amplitude-modulation masking. *J Acoust Soc Am* 111:2551–2554
- Murphy CG, Gerhardt HC (2002) Mate sampling by female barking treefrogs (*Hyla gratiosa*). *Behav Ecol* 13:472–480
- Narins PM, Zelik R (1988) The effects of noise on auditory processing and behavior in amphibians. In: Frittsch B, Ryan MJ, Wilczynski W, Hetherington TE, Walkowiak W (eds) *The evolution of the amphibian auditory system*. Wiley, New York, pp 511–536
- Nelken I, Rotman Y, Bar Yosef O (1999) Responses of auditory-cortex neurons to structural features of natural sounds. *Nature* 397:154–157
- Nelson PB, Jin SH, Carney AE, Nelson DA (2003) Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners. *J Acoust Soc Am* 113:961–968
- Noble W, Perrett S (2002) Hearing speech against spatially separate competing speech versus competing noise. *Percept Psychophys* 64:1325–1336
- Noldus (2005) EthoVision[®] Video Tracking System for automation of behavioral experiments: reference manual version 3.1. Noldus Information Technology, Wageningen
- Rheinlaender J, Klump GM (1988) Behavioral aspects of sound localization. In: Frittsch B, Ryan MJ, Wilczynski W, Hetherington T (eds) *The evolution of the amphibian auditory system*. Wiley, New York, pp 297–305
- Richards DG, Wiley RH (1980) Reverberations and amplitude fluctuations in the propagation of sound in a forest: implications for animal communication. *Am Nat* 115:381–399
- Ronacher B, Hoffmann C (2003) Influence of amplitude modulated noise on the recognition of communication signals in the grasshopper *Chorthippus biguttulus*. *J Comp Physiol A* 189:419–425
- Ronacher B, Krahe R (1998) Song recognition in the grasshopper *Chorthippus biguttulus* is not impaired by shortening song signals: implications for neuronal encoding. *J Comp Physiol A Sens Neural Behav Physiol* 183:729–735
- Schul J, Bush SL (2002) Non-parallel coevolution of sender and receiver in the acoustic communication system of treefrogs. *Proc R Soc Lond B Biol Sci* 269:1847–1852
- Schwartz JJ (1987) The function of call alternation in anuran amphibians: a test of three hypotheses. *Evolution* 41:461–471
- Schwartz JJ, Buchanan BW, Gerhardt HC (2001) Female mate choice in the gray treefrog (*Hyla versicolor*) in three experimental environments. *Behav Ecol Sociobiol* 49:443–455
- Schwartz JJ, Freeberg TM (2008) Acoustic interaction in animal groups: signaling in noisy and social contexts—introduction. *J Comp Psychol* 122:231–234
- Schwartz JJ, Gerhardt HC (1995) Directionality of the auditory system and call pattern recognition during acoustic interference in the gray treefrog, *Hyla versicolor*. *Audiot Neurosci* 1:195–206
- Schwartz JJ, Marshall VT (2006) Forms of call overlap and their impact on advertisement call attractiveness to females of the gray treefrog, *Hyla versicolor*. *Bioacoustics* 16:39–56
- Shinn-Cunningham BG, Schickler J, Kopco N, Litovsky R (2001) Spatial unmasking of nearby speech sources in a simulated anechoic environment. *J Acoust Soc Am* 110:1118–1129
- Swanson EM, Tekmen SM, Bee MA (2007) Do female anurans exploit inadvertent social information to locate breeding aggregations? *Can J Zool* 85:921–932
- Takahashi GA, Bacon SP (1992) Modulation detection, modulation masking, and speech understanding in noise in the elderly. *J Speech Hear Res* 35:1410–1421
- Welch AM, Semlitsch RD, Gerhardt HC (1998) Call duration as an indicator of genetic quality in male gray treefrogs. *Science* 280:1928–1930
- Wells KD (2007) *The ecology and behavior of amphibians*. University of Chicago Press, Chicago
- Wiley RH (1994) Errors, exaggeration, and deception in animal communication. In: Real LA (ed) *Behavioural mechanisms in evolutionary ecology*. Chicago University Press, Chicago, pp 157–189
- Wiley RH (2006) Signal detection and animal communication. *Adv Study Behav* 36:217–247
- Wiley RH, Richards DG (1978) Physical constraints on acoustic communication in the atmosphere: implications for the evolution of animal vocalizations. *Behav Ecol Sociobiol* 3:69–94
- Wollerma L (1999) Acoustic interference limits call detection in a Neotropical frog *Hyla ebraccata*. *Anim Behav* 57:529–536
- Wollerma L, Wiley RH (2002) Background noise from a natural chorus alters female discrimination of male calls in a Neotropical frog. *Anim Behav* 63:15–22
- Zar JH (1999) *Biostatistical analysis*, 4th edn. Prentice-Hall, Upper Saddle River