

**Acoustic communication in noisy environments:
Signal recognition in fluctuating backgrounds.**

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Alejandro Vélez Meléndez

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dr. Mark A. Bee

April 2012

© Alejandro Vélez Meléndez 2012

Acknowledgments

A graduate degree is not an easy task to complete and requires diverse forms of support and assistance from colleagues, family, and friends. I want to thank the people that helped me accomplish this goal.

It is not easy to express how grateful I am to my advisor, Mark Bee, for receiving me in his lab and believing in me. Mark mentored me through my research and his contributions to my professional development are invaluable. Throughout my years at the University of Minnesota, I was lucky to have his support not only on the academic front, but also in my personal life. I am happy to consider him a colleague and friend.

I cannot think of a better place to do a Ph.D. than the Animal Communication lab. I especially want to thank Sandra Tekmen and Beth Pettitt for being by my side during these six years. Words cannot express the appreciation I have for them. Thanks to Elliot Love, Kat Schrode, Noah Gordon, Vivek Nityananda, Michael Caldwell, Jessica Ward, and Norman Lee, my life in the lab was both fun and academically enriching. I am also grateful to all the undergraduate students that helped collecting and testing frogs.

I want to thank committee members Dave Stephens, Marla Spivak, Neal Viemeister, Emilie Snell-Rood, and Keith Barker for their intellectual contributions to my dissertation research. Also, I could not have conducted much of my work without the valuable help offered by Magda Wojtczak, Chris Micheyl, and Peggy Nelson. The EEB Behavior Group and the JSem were both very educational and helpful.

In addition to intellectual contributions and help getting the job done, financial support is always necessary for the successful completion of research projects. I must acknowledge the Florence Rothman Fellowship, the Dayton-Wilkie Fellowship, the Graduate School Thesis Research Grant, and various mini-grants for research and travel expenses from the Graduate and Professional Student Assembly (GAPSA), and the

Department of Ecology, Evolution, and Behavior. My research was also funded by NSF and NIH grants to Mark Bee.

Life as a grad student is fun, but not always easy. Support from friends and family is among the most important factors that help to bear with the stress and difficulties associated with grad school. Luckily, I had the opportunity to continually release tension playing soccer and squash. For that, I am grateful to Caminantes and my squash buddies Juan Jo, María, Fernando, Juancho, and all the members of the University of Minnesota Squash Club. I also want to acknowledge all my friends for their support during difficult periods in my career. I especially want to thank my wonderful family for their love and unconditional support. I would not have been able to get to this point without them and this accomplishment is theirs as it is mine. Last, but definitely not least, I am grateful to Juana, to whom this work I dedicate.

Acknowledgments	i
List of Tables	vii
List of Figures	viii
Introduction	1
Background.....	1
Frogs as a Study System.....	2
Dissertation Research Summary	3
Significance	7
Chapter 1 – Signal recognition by frogs in the presence of temporally fluctuating chorus-shaped noise.....	9
Introduction	10
Methods	15
<i>Subjects</i>	<i>15</i>
<i>General testing procedures</i>	<i>15</i>
<i>Acoustic stimuli</i>	<i>18</i>
<i>Experiment 1: Chorus-shaped noise as a potential signal</i>	<i>20</i>
<i>Experiment 2: Signal recognition thresholds in modulated chorus- shaped noise</i>	<i>21</i>
Results	24
<i>Experiment 1: Chorus-shaped noise as a potential signal</i>	<i>24</i>
<i>Experiment 2: Signal recognition thresholds in modulated chorus- shaped noise</i>	<i>26</i>
Discussion	31
Chapter 2 – Dip listening and the cocktail party problem in grey treefrogs: signal recognition in temporally fluctuating noise.....	38

Introduction	38
General methods	40
<i>Subjects</i>	40
<i>Phonotaxis Experiments</i>	41
Experiment 1: Signal recognition in fluctuating noises	42
<i>Methods</i>	42
<i>Results and Discussion</i>	48
Experiment 2: Minimum call duration for acoustic glimpses	50
<i>Methods</i>	51
<i>Results and Discussion</i>	52
Experiment 3: Chorus-shaped maskers as potential signals	52
<i>Methods</i>	52
<i>Results and Discussion</i>	53
General discussion	56
<i>Dip listening or modulation masking?</i>	56
<i>Solutions to cocktail-party-like problems</i>	58
<i>Proximate level explanations for dip listening</i>	60
<i>Conclusions</i>	61
Chapter 3 – Dip listening or modulation masking? Call recognition by green treefrogs (<i>Hyla cinerea</i>) in temporally fluctuating noise.....	62
Introduction	62
Materials and methods	66
<i>Study sites and subjects</i>	66
<i>Acoustical analysis: Level fluctuations of natural chorus noise</i>	66

<i>General testing procedures</i>	67
<i>Acoustic stimuli</i>	68
<i>Experiment 1: Call recognition in fluctuating chorus-shaped maskers</i>	71
<i>Experiment 2: Chorus-shaped maskers as potential signals</i>	76
Results	77
<i>Acoustical analysis: Level fluctuations of natural chorus noise</i>	77
<i>Experiment 1: Call recognition in fluctuating chorus-shaped maskers</i>	77
<i>Experiment 2: Chorus-shaped maskers as potential signals</i>	81
Discussion	82
<i>Statistical power and experimental design</i>	83
<i>Comparisons with previous studies</i>	85
<i>Signal recognition in natural settings</i>	88
<i>Conclusions</i>	89
Chapter 4 – Signal Recognition by Green Treefrogs (<i>Hyla cinerea</i>) and Cope’s Gray Treefrogs (<i>Hyla chrysoscelis</i>) in Naturally Fluctuating Noise.	91
Introduction	92
General Methods	95
<i>Subjects and study sites</i>	95
<i>Phonotaxis experiments</i>	95
Experiment 1: Signal recognition in fluctuating backgrounds	97
<i>Methods</i>	97
<i>Results and discussion</i>	103
Experiment 2: Behavior in the absence of target signals	105

<i>Methods</i>	105
<i>Results and discussion</i>	108
Experiment 3: Maskers as potential signals	109
<i>Methods</i>	109
<i>Results and discussion</i>	110
General Discussion	112
<i>Signal recognition in natural settings</i>	114
<i>Between-species differences in dip-listening capabilities</i>	116
<i>Conclusions</i>	117
References	118
Appendix 1: Assessing the Potential for Dip Listening in Natural Settings	132
Appendix 2: Analyses of original movement-pattern variables in sham trials	136
Appendix 3: Signal recognition in fluctating noise: Comparison between Cope’s gray treefrogs and green treefrogs	138
Appendix 4: Analyses of signal recognition thresholds excluding attractive maskers	140

List of Tables

Table 1-1. 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.	28
Table 1-2. 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 $p > 0.20$	30
Table 2-1. Results of circular statistical analyses for orientation angles at the wall for the reference and test trials.	55
Table 3-1. Results of circular analyses for orientation angles at the arena wall in response to reference trials and chorus-shaped noises.	82
Table 4-1. Loadings and results of planned comparisons testing each specific hypothesis for Cope's gray treefrogs.	105
Table A2-1. Correlations between the four original movement-pattern variables.	137
Table A3-1. Individual planned comparisons between the non-fluctuating control treatment and all fluctuating treatments for green treefrogs (<i>H. cinerea</i>) and Cope's gray treefrogs (<i>H. chrysoscelis</i>).	139

List of Figures

Figure 1-1. 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. chrysozelis*), (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, where as 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 = 11025 samples/s, Hamming window size = 65,536 points, overlap = 25%) and normalize 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 m 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). 12-13

Figure 1-2. Modulation spectrum of gray treefrog choruses. Shown here on a logarithmic x axis is the mean (bold, solid line) \pm 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 = 11025 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 ($20\log_{10}$ (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. 14

Figure 1-3. Standard call and chorus-shaped noises. (a) Waveform of the synthetic *H. chrysosecelis* standard call with an insert 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. 18

Figure 1-4. 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. 25

Figure 1-5. 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 1-1 for statistical results.)27

Figure 1-6. 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: $F_{3,48} = 1.0$, $P = 0.3982$).33

Figure 2-1. (a) Frequency spectrum of the target signal (black line) and the chorus-shaped maskers (grey line) calculated with 256-point Blackmann–Harris windows. 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.45

Figure 2-2. (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 ($\pm 95\%$ 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 represent less masking than in the nonfluctuating control, and values greater than 0 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 ($\alpha=0.05$).49-50

Figure 3-1. (a) Frequency spectra of the standard call used as a target signal (black line) and the chorus-shaped maskers (gray line) calculated with 256-point Hamming windows. Waveforms of 0.8 s segments of the standard call alone (b), and in the presence of a non-fluctuating masker (c), a 2.5 Hz SAM masker (d), and a 5 Hz SAM masker.69

Figure 3-2. (a, b, c) Waveforms of three representative green treefrog chorus recordings. (d) Mean (solid line) \pm 1 standard deviation (dashed lines) modulation spectrum calculated from 25 90-s segments of green treefrog chorus recordings shown on a logarithmic x axis.78

Figure 3-3. Mean (\pm 95% confidence intervals) signal recognition thresholds as a function of masking treatment. The gray box serves to illustrate the \pm 95% confidence intervals in the non-fluctuating control treatment. Nm = No-masker treatment; Nf = Non-fluctuating treatment. 0.625 – 80 represent SAM rates of each fluctuating noise treatment.79

Figure 3-4. (a) Number of females leaving the holding cage during sham trials as a function of masking treatment. The total number of females tested in each treatment was 24. Median (point), inter-quartile range (box), and non-outlier range (whiskers), values of PCA factor 1 (b) and PCA factor 2 (c) as a function of masking treatment. PCA factor 1 describes angular movements and loads on the variables ‘average turn angle’ and ‘meander’. PCA factor 2 describes non-angular movements and loads on the variables ‘total distance moved’ and ‘average velocity’. Sample sizes for each treatment in (b) and (c) correspond to the number of females leaving the holding cage depicted in (a). Nm = No-masker treatment; Nf = Non-fluctuating treatment. 0.625 – 80 represent SAM rates of each fluctuating noise treatment.80

Figure 4-1. Waveform of 0.8-s segments (top panel) and frequency spectrum (bottom panel) of the synthetic call used as a target signal for tests with green treefrogs (A) and Cope’s gray treefrogs (B). Frequency spectra were calculated with 256-point Hamming windows. For reference, we also show here the frequency spectrum of the maskers (gray line).97

Figure 4-2. Waveform of 10-s segments (top panel) and mean (solid line) \pm 1 standard deviation (dashed lines) modulation spectrum (bottom panel) of 90-s maskers with no level fluctuations (A), synthetically generated random fluctuations (B), or natural level fluctuations characteristic of green treefrog choruses (C) or Cope’s gray treefrog choruses (D). Mean

modulation spectra were calculated from the 25 90-s exemplars of each masker and are shown on a logarithmic x axis. Modulation spectra were generated 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. The modulation spectrum of each masker was calculated as the fast-Fourier transform (FFT) of the envelope (sampling rate = 11025 Hz, Hamming window size = 65536 points, overlap = 25%). We normalized the spectrum of each 90-s segment to the maximum value of the magnitude of the FFT and calculated the mean and standard deviation of the 25 modulation spectra of each type of masker. Mean and ± 1 standard deviation modulation spectra were transformed to a dB scale ($20 \cdot \log_{10}(\text{FFT magnitude})$) and smoothed for plotting purposes using a running average of 11 points.99

Figure 4-3. Schematic representation of the method used to generate the different maskers. This process was repeated 25 times, each one with a different white noise, a different chorus recording of each species, and a different low-pass (10 Hz cut-off) random noise. See text for additional details.100

Figure 4-4. Mean ($\pm 95\%$ confidence intervals) signal recognition thresholds as a function of masking treatment (A), species (B), and masking treatment for each species (C). Nf = Non-fluctuating, Consp. = Conspecific, Heterosp. = Heterospecific, and Rand. = Random. In (C), open circles *H. cinerea* and filled circles represent *H. chrysoscelis*.104

Figure 4-5. (A) Proportions ($\pm 95\%$ confidence intervals) of females leaving the holding cage during sham trials as a function of masking treatment. The total number of females tested in each treatment was 25. Median (point), inter-quartile range (box), and non-outlier range (whiskers), values for total distance moved (B), average velocity (C), average turn angle (D), and average meander (E). Nf = Non-fluctuating, Consp. = Conspecific, Heterosp. = Heterospecific, and Rand. = Random. Open bars and boxes represent *H. cinerea* and filled bars and boxes represent *H. chrysoscelis*.106-107

Figure 4-6. Angles at which subjects reached the arena wall in attraction trials for green treefrogs (top row) and Cope’s gray treefrogs (bottom row) in the nonfluctuating (far left), random (middle left), conspecific (middle right) and heterospecific (far right) treatments. Points depict number of subjects reaching at each 15° bin (solid, short lines) relative to the position of the target speaker (top of each circle). Dotted lines represent 45° bins. See text for additional details111

Figure A1-1. Analyses of the temporal level fluctuations in 60-s recordings of 25 Cope’s grey treefrog choruses. (a) Mean (bold, solid line) \pm 1 standard deviation (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 figure A1-1a 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 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-s 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. (d-f) Examples of individual recordings (N=300 segments per recording) in which the RMS amplitude value and the modulation indices were significantly correlated with a high correlation coefficient ($P < 0.0001$, $r = 0.74$) (d), significantly correlated with a low correlation coefficient ($P < 0.0001$, $r = 0.31$) (e), and not significantly correlated ($P = 0.49$, $r = 0.04$) (f).135

Figure A2-1. Movement patterns in the sham trials. Shown here are the median (point), inter-quartile range (box), and non-outlier range (whiskers) for average turn angle (a), meander (b), total distance moved (c), and average velocity (d) in each treatment. Sample sizes for each treatment are 5 in the Non-fluctuating noise (Nf), 7 in the No-noise (Nn), 2 in the 0.625 Hz SAM noise (0.625), 8 in the 1.25 Hz SAM noise (1.25), 10 in the 2.5 Hz SAM noise (2.5), 5 in the 5 Hz SAM noise (5), 5 in the 10 Hz SAM noise (10), 4 in the 20 Hz SAM noise (20), 5 in the 40 Hz SAM noise (40), and 7 in the 80 Hz SAM noise (80) treatments.137

Figure A3-1. Mean (\pm SE) thresholds as a function of masking treatment for green treefrogs (open circles, dotted lines) and gray treefrogs (filled circles, solid lines). Asterisks represent statistically significant differences using planned comparisons against the non-fluctuating control. Nf = Non-fluctuating treatment; 0.625 – 80 represent SAM rates of each fluctuating noise treatment.139

Introduction

Background

The broad aim of my dissertation research is to address a central question in sensory ecology: how do animals recognize biologically relevant signals in a noisy world? We humans encounter and solve this difficult perceptual task when we follow one conversation in multi-talker social environments. How the human auditory system solves this so-called ‘cocktail-party problem’ (Cherry 1953) has been the focus of research for several decades (reviewed in McDermott 2009). While many nonhuman animals also communicate in noisy social aggregations, we know very little about how they solve analogous problems (Bee and Micheyl 2008). Because the ability to perceptually separate behaviorally relevant signals from the background noise has important evolutionary consequences, natural selection should favor perceptual mechanisms that diminish the detrimental effects of background noise. Nonhuman vertebrates may well solve cocktail-party-like problems with a suit of mechanisms that likely include derived and shared auditory processing strategies for two reasons. First, tympanic hearing evolved independently in three groups of tetrapods (Lombard and Bolt 1979). Second, natural selection has the potential to find different solutions to similar problems. I used frogs as a model system to investigate perceptual mechanisms by which nonhuman animals solve their own cocktail-party-like problems.

I focused on a potential mechanism by which receivers exploit temporal properties of the natural soundscape (i.e., the mixture of sounds in the environment) to overcome problems associated with communication in noise. A well-known characteristic of natural sounds, including the noise generated in animal social aggregations, is that their level fluctuates in time (Richards and Wiley 1980; Nelken et al. 1999). An ability to exploit level fluctuations in background noise constitutes one way by which humans solve the cocktail-party problem. Speech perception is usually better when the level of the background noise fluctuates, compared to when the same noise lacks level

fluctuations (reviewed in Bronkhorst 2000). This phenomenon, often referred to as ‘dip listening’, is attributed to our ability to catch short ‘acoustic glimpses’ of target speech when the background noise dips to low levels (Bacon et al. 1998). Dip listening may be a way by which nonhuman animals also cope with the noise problem. It has even been suggested that animals may be adapted to exploit level fluctuations of the natural soundscape to recognize communication signals (Langemann and Klump 2005). This hypothesis, however, is not yet supported by empirical evidence because (i) there has been no systematic attempt to test signal recognition in the presence of noises with temporal characteristics of natural soundscapes, (ii) very few studies have investigated the ability of nonhuman animals to recognize *communication* signals in the presence of fluctuating backgrounds, and (iii) we still know little about the characteristics of level fluctuations in natural soundscapes. Here, I addressed these gaps in knowledge in two species with well-described communication systems: Cope’s gray treefrogs (*Hyla chrysoscelis*) and green treefrogs (*Hyla cinerea*).

Frogs as a Study System

Frogs represent an ideal model for studying auditory mechanisms that facilitate signal recognition in noisy environments. Male frogs produce loud advertisement calls that are necessary and sufficient for species recognition and mate choice by females. Extensive research has focused on elucidating the type of information transmitted in male advertisement calls, how this information is encoded in different physical properties of the calls, and the mechanisms by which receivers make important adaptive behaviors such as species recognition and mate choice (reviewed in Ryan 2001, Gerhardt & Huber 2002). In addition, we have learned a great deal about the anatomy and physiology of the anuran auditory system (reviewed in Gerhardt & Huber 2002). Importantly, in many species of frogs, males aggregate around suitable habitats and form loud choruses where they produce mating calls to attract females. These choruses can include hundreds of calling males (Murphy 2003) and can be heard from distances of up to 2 km (Arak 1983).

As in other frogs, male Cope's gray treefrogs and green treefrogs form dense choruses where they produce distinctive advertisement calls to attract females (Gerhardt and Huber 2002; reviewed in Schwartz and Bee in press). Importantly, the structure of the advertisement call is very different between these two species. Male Cope's gray treefrogs produce long (~800 ms), trilling advertisement calls comprised of approximately 30 pulses delivered at a rate of 40 – 50 pulses per second. In contrast, green treefrog males produce short (~150 ms) non-pulsed calls. In both species, the noise generated in the chorus hinders the ability of females to recognize and discriminate among individual calls (reviewed in Vélez et al. in press). Therefore, selection pressures should favor mechanisms that ameliorate the detrimental effects of noise in chorus environments, making these species great systems to study evolved mechanisms for communication in noisy social aggregations.

Dissertation Research Summary

The first chapter of my dissertation research was published in *Behavioral Ecology and Sociobiology* (Vélez and Bee 2010). This study sets the ground for the following chapters. I adopted traditional methods from studies on human hearing research to test the hypothesis that female Cope's gray treefrogs (*H. chryoscelis*) listen in the dips of fluctuating background noise. I measured the ability of female frogs to recognize male mating calls in the presence of noises that either lacked level fluctuations, or fluctuated sinusoidally at rates of 5 Hz or 40 Hz. Sinusoidally amplitude modulated (SAM) noises are commonly used in psychophysical and neurophysiological studies on temporal auditory processing due to the high levels of experimental control they offer (Joris 2004). I did not find evidence for dip listening. When the background noise fluctuated at a rate of 5 Hz, the ability of female frogs to recognize calls was not different from that in the presence of noises that lacked level fluctuations. Based on this result, I hypothesized that the 'dips' in the noise fluctuating at 5 Hz were not long enough to included meaningful 'acoustic glimpses' of the mating call. In the presence of noises fluctuating at rates similar to those in the mating call (i.e., 40 Hz), I found that level fluctuations in the

background noise actually impose additional constraints on signal recognition. Because the pulse rate of the call is an important species-recognition cue in Cope's gray treefrogs (Schul and Bush 2002), I interpreted this result as an effect of 'modulation masking.' Used to describe an increased difficulty to recognize fluctuating signals in the presence of fluctuating backgrounds, modulation masking is especially strong when the fluctuation rates of the signals and the backgrounds are similar (Bacon and Grantham 1989). In addition to the main experiment, I recorded and analyzed the sounds of breeding choruses of different frog species. This study was the first to show that the background noise of animal social aggregations fluctuates with species-specific patterns that result from species-specific characteristics of the communication signals and signaling behaviors. For instance, I showed that the noise generated in Cope's gray treefrog choruses consistently fluctuates at slow rates (below ~5 Hz) corresponding to the timing of individual calls, and at faster rates corresponding to the pulse rate of the call (~40 – 60 Hz). These results are particularly important because they highlight the significance of understanding the properties of the soundscape when studying perceptual mechanisms for communication in noisy settings. Results from this study generated a series of hypotheses about the effects on signal recognition of (i) the structure of the signal (i.e., short vs. long, pulsed vs. non-pulsed), (ii) the rates of level fluctuation of the background noise on signal recognition, and (iii) species-specific characteristics of the natural soundscape.

In the second chapter of my dissertation research, published last year in *Animal Behaviour* (Vélez and Bee 2011), I built upon results and hypotheses generated from chapter 1. First, results from the analysis of chorus recordings suggested that the rates of level fluctuation used in the previous study (5 Hz and 40 Hz) were not entirely representative of the rates of fluctuation present in the natural soundscape. Here, I tested the hypothesis that level fluctuation rates that span those in the sounds of frog choruses affect signal recognition in Cope's gray treefrogs. I measured the ability of female frogs to recognize male mating calls in the presence of SAM backgrounds ranging from very slow fluctuation rates (0.625 Hz) to fast fluctuation rates (80 Hz). Female Cope's gray

treefrogs experienced dip listening in the presence of fluctuating backgrounds in which noise dips were long enough to include nine or more consecutive pulses of the call, but not when the dips included five pulses or less. I also found that subjects experienced modulation masking in backgrounds fluctuating at rates similar to those in the signal. These results are consistent with my previous findings. In addition, my results partially support the hypothesis that female frogs need to hear a minimum number of pulses in noise dips for dip listening to occur. I conducted a second experiment to determine whether the number of pulses falling within dips of fluctuating backgrounds could explain the patterns of dip listening observed. In the absence of background noise, I found that the minimum number of pulses a female frog can hear and still recognize the signal as a conspecific call is between six and nine. Together, my results suggest that the ability of female Cope's gray treefrogs to listen in the dips depends, in part, on the number of pulses falling within dips of fluctuating backgrounds. At a proximate level, this study generated important hypotheses about the underlying neural mechanisms for temporal processing and signal recognition in noise in frogs. At an ultimate level, it generated hypotheses about the implications of dip listening on male calling behavior. Finally, to the best of my knowledge, this is the first study with nonhuman vertebrates to demonstrate that temporally structured backgrounds improve the ability of receivers to recognize communication signals in noisy settings.

Chapter three is a study I conducted with green treefrogs (*H. cinerea*) in parallel with the experiments of chapter two and the resulting manuscript is in review at the *Journal of Comparative Physiology A* (Vélez et al. in review). Here, I tested the hypothesis that female green treefrogs experience dip listening in the presence of fluctuating backgrounds using the same protocol of chapter two. Because the advertisement call of green treefrogs is short and not pulsed, patterns of dip listening and modulation masking different from those observed with Cope's gray treefrogs could be uncovered. In particular, I did not expect to observe modulation masking effects. My results showed that female green treefrogs did not experience dip listening or modulation

masking in the presence of any of the fluctuating backgrounds tested. In addition, I analyzed recordings of the sounds of green treefrog choruses and showed that their levels fluctuate primarily at rates below 10 Hz, with a peak at 5 Hz that corresponds to male calling interactions in a chorus. Together, these results suggest that female listeners in natural settings should receive no benefits, nor experience any additional constraints, as a result of level fluctuations in the soundscape of green treefrog choruses. The differences in dip-listening capabilities between Cope's gray treefrogs and green treefrogs, two closely related species, highlight the importance of comparative approaches to explore and understand mechanisms for communication in noise.

In chapter four, I report results of a comparative study examining signal recognition in the presence of naturally fluctuating backgrounds. There is growing evidence to suggest that the auditory system is adapted to process physical properties of natural sounds (Rieke et al. 1995; Lewicki 2002; Woolley et al. 2005; Smith & Lewicki 2006). In species that communicate in social aggregations, differences in the physical properties across soundscapes could potentially lead to species-specific adaptations that facilitate signal perception in the local environment. In the previous chapters, I used sinusoidally amplitude-modulated noises to test hypotheses about the ability of female frogs to listen in the dips of fluctuating backgrounds. However, level fluctuations in natural soundscapes are not sinusoidal. Furthermore, my results revealed that level fluctuations in the soundscape of breeding choruses differ across species. Here, I tested three hypotheses about signal recognition in fluctuating backgrounds. I tested the 'dip-listening' hypothesis, which predicts facilitated signal recognition in the presence of fluctuating maskers, compared to non-fluctuating maskers. I also tested the 'natural-soundscapes advantage' hypothesis, which predicts facilitated signal recognition when level fluctuations resemble those of natural soundscapes, compared to artificial level fluctuations. The third hypothesis I tested is the 'species-specific advantage' hypothesis, which predicts an improvement in signal recognition when level fluctuations resemble species-specific patterns of conspecific soundscapes, compared to other naturally

fluctuating backgrounds. I tested these three hypotheses by measuring the ability of female Cope's gray treefrogs and green treefrogs to recognize male conspecific calls in the presence of background noises with (i) no level fluctuations, (ii) random fluctuations, or level fluctuations characteristic of (iii) conspecific choruses and (iv) heterospecific choruses. Support for the dip listening hypothesis was weak; only Cope's gray treefrogs experienced dip listening in the presence of randomly fluctuating noises. I found no evidence to support the natural-soundscapes advantage hypothesis or the species-specific advantage hypothesis. These results corroborate my previous findings showing that Cope's gray treefrogs, but not green treefrogs, experience dip listening under some noise conditions. My results also suggest level fluctuations in the soundscape of breeding choruses may offer few dip-listening opportunities in these two species. Together, my findings offer little support for the hypothesis that receivers are adapted to exploit level fluctuations of the natural soundscape to recognize communication signals. Chapter four is currently in manuscript form and will be submitted before my defense to the *Journal of Comparative Psychology*.

Significance

In my dissertation research, I integrated fundamental questions in sensory ecology with hypotheses from psychophysics and human hearing studies. An important outcome of my research concerns the contrasting results between experiments using backgrounds with artificial (random or sinusoidal) and natural level fluctuations. I showed that Cope's gray treefrog females listen in the dips of artificially fluctuating backgrounds. This ability, however, offers little benefit in the presence of naturally fluctuating backgrounds. These results stress the importance of studying the physical properties of natural soundscapes and their effects on signal recognition before generalizing from studies using solely artificial stimuli. Another important result from my research is the differences in dip-listening capabilities between the two species studied. This result highlights the importance of comparative approaches in studies of signal recognition in noise. The between-species differences revealed in my studies urge that generalizations across taxa,

even between closely related species, are made with caution. Together, my results suggest that, in order to understand the evolution of auditory processing mechanisms for communication in noisy environments, it is important to (i) take comparative approaches, (ii) investigate the properties of natural soundscapes and how they affect signal recognition, and (iii) use both natural and artificial stimuli to better understand the capabilities and limitations of the auditory system.

The variety of disciplines of the journals at which I am targeting my research evidences the significance of my research in diverse fields, such as behavioral ecology, sensory biology, auditory neuroscience, and hearing research. Results from my research increase our understanding of how the vertebrate auditory system may evolve to solve problems associated with communication in noisy settings. My approach is particularly important and novel because it establishes a connection between the characteristics of natural acoustic environments and the possible adaptations that facilitate signal recognition in social aggregations. My research complements that on the human cocktail party problem by adding an evolutionary perspective.

Chapter 1 – Signal recognition by frogs in the presence of temporally fluctuating chorus-shaped noise.¹

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 non-human 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 modulations rates characteristic of those present in ambient chorus noise. We estimated thresholds for recognizing conspecific advertisement calls (pulse rate=40-50 pulse/s) in the presence of unmodulated and sinusoidally amplitude modulated (SAM) chorus-shaped masking noise. We tested two rates of modulation (5 Hz 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 15-50 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.

¹ This chapter is published as:
Vélez A, and Bee MA. 2010. Signal recognition by frogs in the presence of temporally fluctuating chorus-shaped noise. *Behavioral Ecology and Sociobiology*, 64: 1695-1709.

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 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; Bee 2008b). 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. 1-1). 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 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 Figure 1-2,

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 (≈ 40 - 50 Hz) that correspond to the pulse rate of the advertisement call.

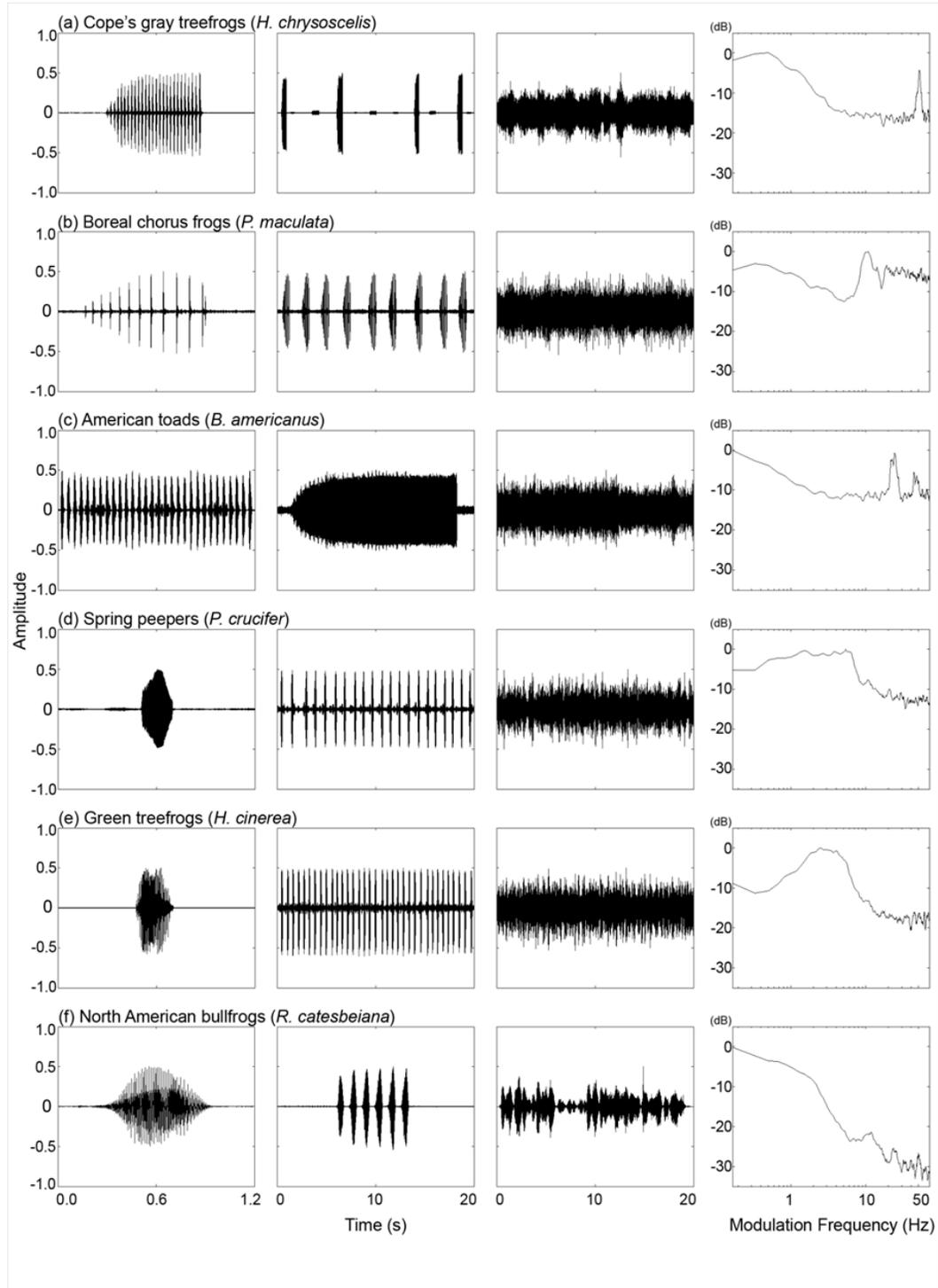


Figure 1-1 Modulation spectra illustrating the temporal fluctuations characteristic of the choruses of different frog species. Each row of the figure has four separate plots for each species showing the following: (far left) a 1.2-s waveform of a portion of a single call or an entire single call from 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, where as 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 = 11025 samples/s, Hamming window size = 65,536 points, overlap = 25%) and normalize 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 m 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).

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 recognition

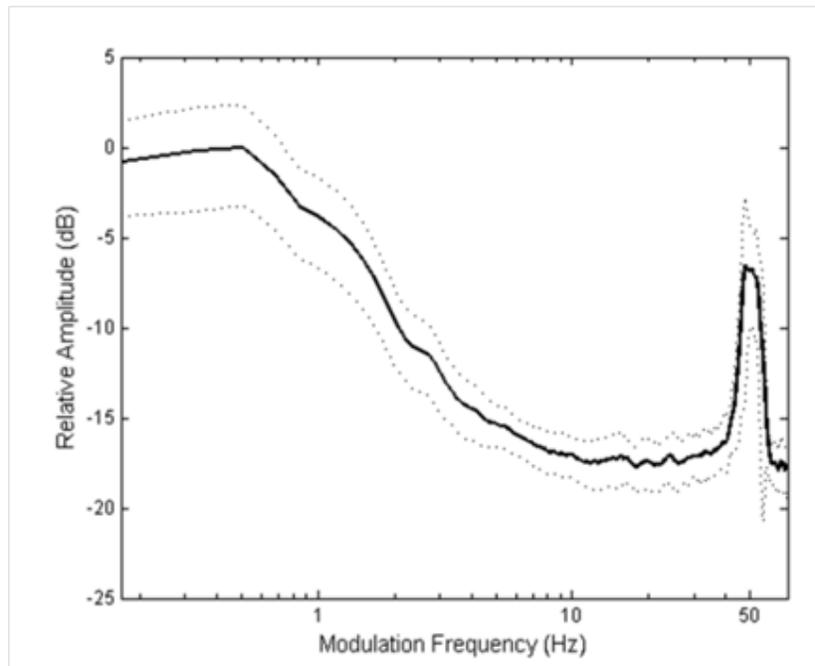


Figure 1-2 Modulation spectrum of gray treefrog choruses. Shown here on a logarithmic x axis is the mean (bold, solid line) ± 1 standard deviation (thin, dotted lines) modulation spectrum determined by averaging the spectra of eight 60-s segments of Cope's gray treefrog choruses. Each segment was taken 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 = 11025 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 omni-directional 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.

task in the presence of speech-shaped noise (e.g., Festen and Plomp 1990; Bronkhorst and Plomp 1992; Shinn-Cunningham et al. 2001). Typical maskers in such studies often comprise noises that fluctuate in amplitude with the envelope of sine waves and are referred to as sinusoidally amplitude-modulated (SAM) noises (e.g., Takahashi and Bacon 1992; Gustafsson and Arlinger 1994; Füllgrabe et al. 2006). In our study, the target

signal 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, Minnesota, U.S.A.) and the Tamarack Nature Center (45° 6'9.81"N, 93° 2'27.56", Ramsey County, Minnesota, U.S.A.). We returned females to the lab and kept them at 2°C to delay egg deposition until they were tested (usually within 24 hrs). 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, 2008b; 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, 2008b; 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 $20^{\circ}\text{C} \pm 2^{\circ}\text{C}$ in two temperature-controlled, hemi-anechoic sound chambers (see Bee and Schwartz (2009) for details). Tests were conducted under infrared (IR) 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-minute 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).

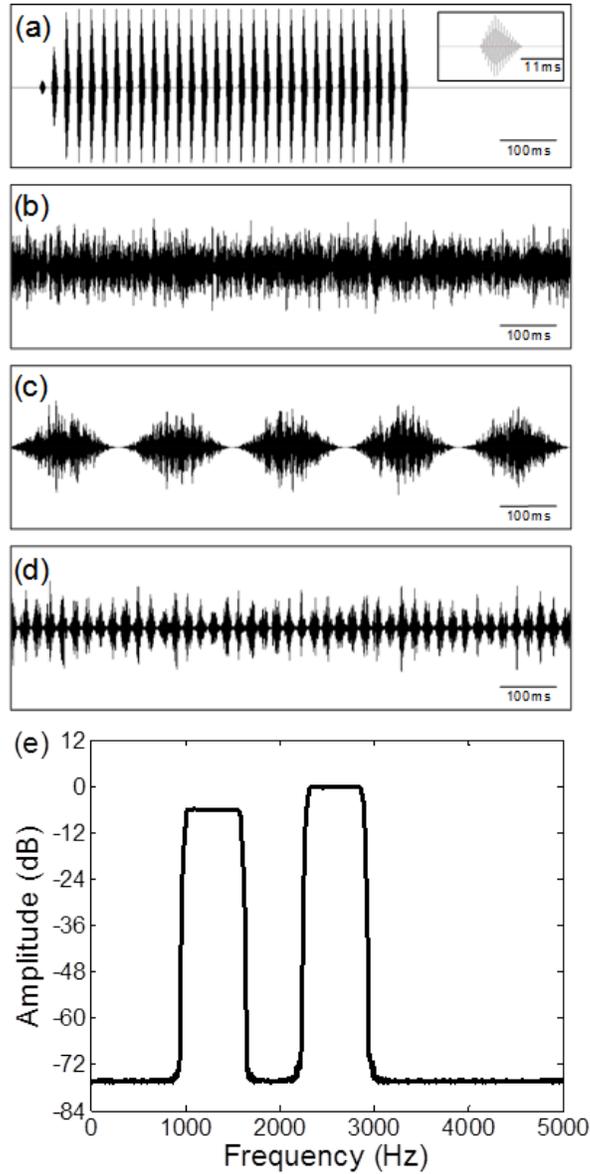


Figure 1-3 Standard call and chorus-shaped noises. (a) Waveform of the synthetic *H. chrysoscelis* standard call with an insert showing the waveform of a single pulse; (b-d) Waveforms of the unmodulated masker (b), the 5-Hz SAM masker (c); and the 45-Hz SAM masker (d); (e) Power spectrum showing the spectral profile of the chorus-shaped noises.

Acoustic stimuli

The standard call

We used a standard synthetic call (Fig. 1-3a) as the target signal in all reference conditions and in several of our 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. 1-3b-e). An unmodulated noise was created by filtering white noise into two 600-Hz-wide spectral bands centered at 1.3 kHz 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 Hz 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-Hz SAM 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 one minute. This sound level falls within the range of background noise levels

that we and others have recorded in natural *H. chrysoscelis* 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) x 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 modulated 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 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:

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

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 equation 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

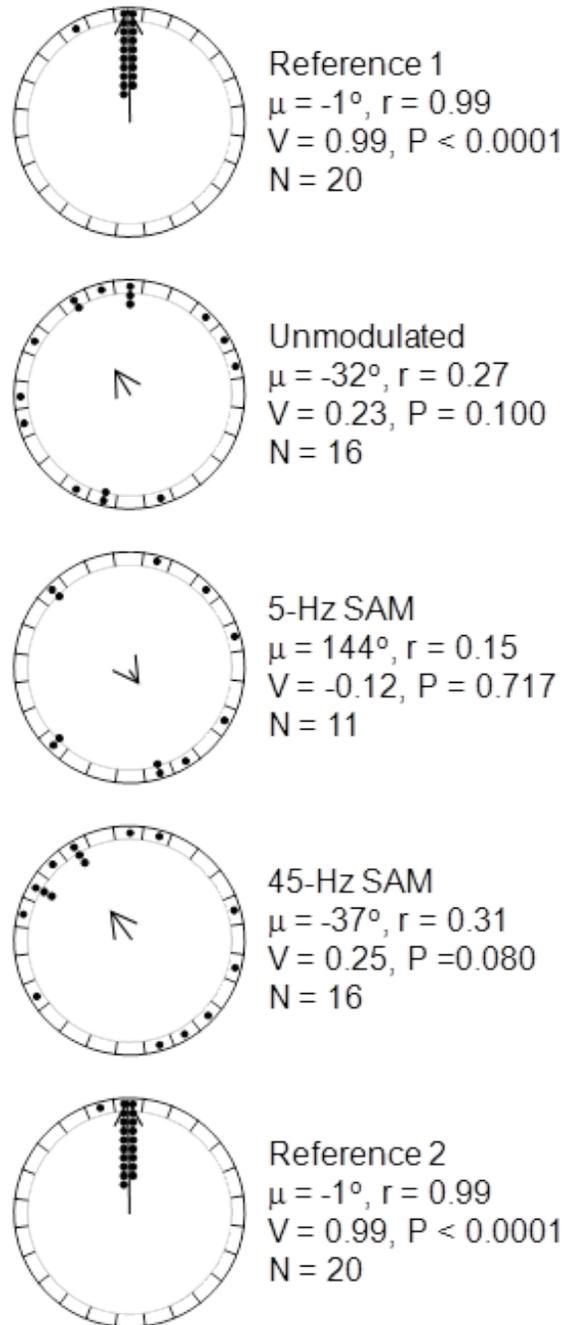


Figure 1-4 Chorus-shaped noise as a potential signal. Points depict the angles at which females first touched the wall of the arena relative to the position of the target speaker (top of each circle) in the two reference conditions, and in response to the unmodulated, 5-Hz SAM, and 45-Hz SAM chorus-shaped 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.

not exhibit phonotaxis toward unmodulated and SAM chorus-shaped noises (Fig. 1-4). 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. 1-4). 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 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 1-1). 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. 1-5). In addition, response probabilities were above 0.80 at all signal levels (all $P_s < 0.05$ 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

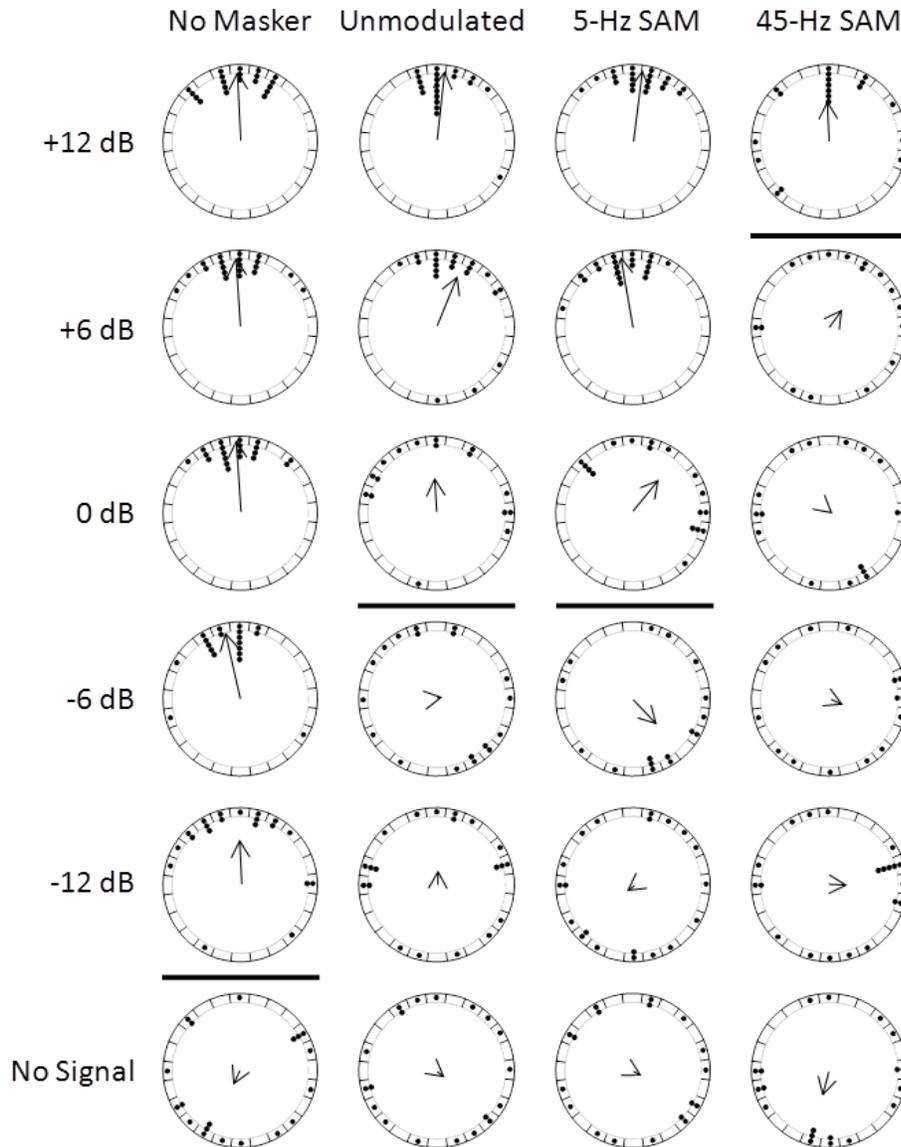


Figure 1-5 Angular orientation in response to synthetic calls presented in the presence or absence of chorus-shaped noise. Points depict the angles at which individual females first left a circle with radius 20 cm centered on the release point at the center of the test arena. Data are shown for the 24 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 1-1 for statistical results.)

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

Table 1-1. 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). Asterisks indicate statistically significant orientation in the masking conditions ($\alpha = 0.05$).

Condition	SNR	Mean Vector (m°)	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*	-3	0.89	27	20	0.89	< 0.0001
	+6 dB*	-4	0.89	27	20	0.89	< 0.0001
	0 dB*	-4	0.92	23	20	0.92	< 0.0001
	-6 dB*	-13	0.87	38	19	0.79	< 0.0001
	-12 dB*	-3	0.57	60	20	0.57	< 0.0001
	No signal	-156	0.20	103	19	-0.18	0.8690
Unmodulated	+12 dB*	6	0.89	28	20	0.86	< 0.0001
	+6 dB*	22	0.68	50	20	0.63	< 0.0001
	0 dB*	-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*	7	0.92	23	20	0.92	< 0.0001
	+6 dB*	-10	0.91	25	20	0.89	< 0.0001
	0 dB*	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*	-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

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. 1-5; Table 1-1). 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. 1-5; Table 1-1). 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 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 orientation also would have occurred at even higher SNRs in the presence of the 45-SAM masker, we used +12 dB 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: 4 of 20; unmodulated: 2 of 20; 5-Hz SAM: 4 of 20; 45-Hz SAM: 3 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 wall in front of the arena 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 1-2). We used SNRs

of -6 dB 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 dB and +12 dB, but not at lower SNRs (Table 1-2). Using +6 dB and 0 dB as the upper and lower bounds, respectively, we estimated a signal recognition threshold of 4 dB for the 45-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. 1-6) based on comparing mean values of total distance moved, velocity of movement, turn angle, and meander in a repeated measures MANOVA (Wilks' $\lambda = 0.40$, $F_{12,5} = 0.6$, $P =$

Table 1-2. 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

0.7652). Subsequent univariate tests also failed to reveal differences in each of these behavioral response measures (Fig. 1-6). 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: (i) increased potential for communication errors (e.g., missed detection or incorrect classification) and (ii) 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. 1-6). 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.

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 *Ch. 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 Hz, 2.5 Hz, and 5 Hz) had little effect on signal recognition. In contrast, higher modulation rates (15 Hz, 50 Hz, 70 Hz, and 150 Hz) significantly impaired recognition relative to that in the unmodulated condition. Notably, there was a steep decline in recognition between the 5 Hz and 15 Hz modulation frequencies, which encompasses the slower rate of amplitude modulation in the signal. Together, these results suggest that masking release in

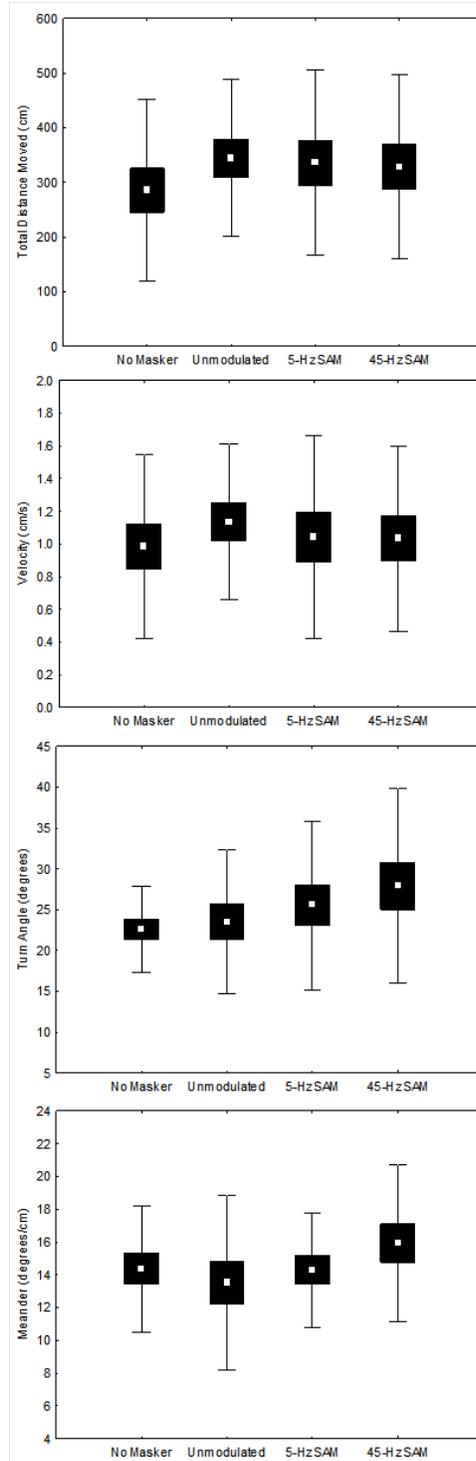


Figure 1-6 Patterns of movement in the no signal conditions. Each plot depicts the mean (point), ± 1 SE (box), and ± 1 SD (whiskers) for one of the movement variables across the four different masking conditions. From top to bottom the results are shown for total distance moved (univariate ANOVA: $F_{3,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

temporally fluctuating noise had little influence on signal recognition in *Ch. 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, non-human 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 five 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. 1-1). 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, *H. 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.

Chapter 2 – Dip listening and the cocktail party problem in grey treefrogs: signal recognition in temporally fluctuating noise.²

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.

Introduction

Humans and nonhuman animals often communicate acoustically in large social groups, such as cocktail parties, choruses or colonies (Schwartz and 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 and Slabbekoorn 2005). In humans, the challenge

² This chapter is published as:
Vélez A, and Bee MA. 2011. Dip listening and the cocktail party problem in grey treefrogs: signal recognition in temporally fluctuating noise. *Animal Behaviour*, 82: 1319-1327.

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 and Slabbekoorn 2005; Langemann and Klump 2005; Bee and Micheyl 2008). Compared to our thorough understanding of strategies signallers use to avoid or ameliorate noise problems (reviewed in Brumm and 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 and 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 and Wiley 1980; Nelken et al. 1999; Vélez and 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 and 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 and 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 and 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.

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 and 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 and Klump 1988a; Wollerman 1999; Wollerman and Wiley 2002; Bee 2007a, 2008a, b; Bee and Schwartz 2009; Vélez and 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 and 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 and 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 and 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 spectral and temporal properties close to the average of calls recorded in local populations (Fig. 2-1a, 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. 2-1a). 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 and 3300 Hz (Fig. 2-1a) and transformed it to the temporal domain with an inverse FFT (Fig. 2-1c).

Across nine different treatments, chorus-shaped maskers differed in their temporal structure but had equivalent long-term root-mean-square (RMS) amplitudes (Fig. 2-1c–

e). In the nonfluctuating control treatment (Fig. 2-1c), 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. 2-1c, d, e). The remaining treatments comprised eight experimental treatments in which we imposed sinusoidal fluctuations in masker level at rates of 0.625 Hz (Fig. 2-1d), 1.25 Hz, 2.5 Hz, 5 Hz (Fig. 2-1e), 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. 2-1d, e). The duration of each ‘dip’ in these maskers, as measured between the 6 dB down points (Fig. 2-1d, 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. 2-2a). The rates of fluctuation in our maskers spanned the range present in the sounds of grey treefrog choruses (Vélez and Bee 2010) and encompassed the pulse repetition rate of the advertisement call ($\approx 40\text{--}50$ 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).

We used artificial SAM maskers in this experiment because they are commonly used in psychophysical and neurophysiological studies of temporal auditory processing (Joris 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

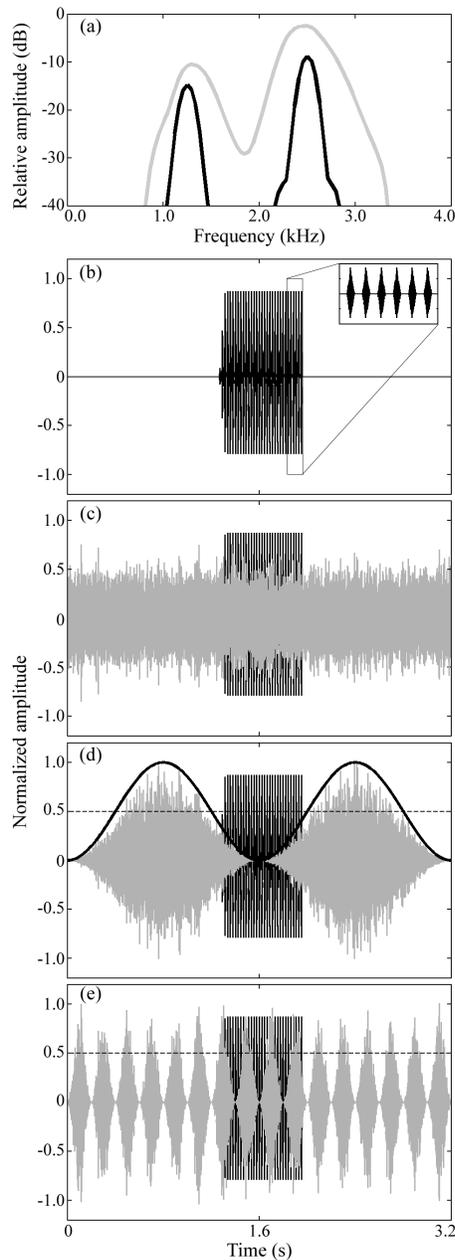


Figure 2-1. (a) Frequency spectrum of the target signal (black line) and the chorus-shaped maskers (grey line) calculated with 256-point Blackmann–Harris windows. Waveforms of 3.2 s segments of the target signal alone (b), and in the presence of maskers representing the nonfluctuating control treatment (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.

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 ($\rho = 0^\circ, 90^\circ, 180^\circ$ 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 and 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 semicircle 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 and 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 (fast RMS, C-weighted), 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 and Schwartz 2009; Vélez and Bee 2010), we replaced any subject that failed to respond on any reference trial ($N = 11$). This procedure ensured that subjects remained highly motivated to respond across trials, as evidenced by similar mean response latencies in the initial (\pm SD = 111.58 ± 39.89 s) and final (\pm SD = 110.54 ± 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 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 and 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 (\pm SE = 67.9 ± 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: $P_s > 0.20$) and homogeneity of variance (Levene’s tests: $P_s > 0.24$) 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. 2-2b), 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. 2-2b), 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

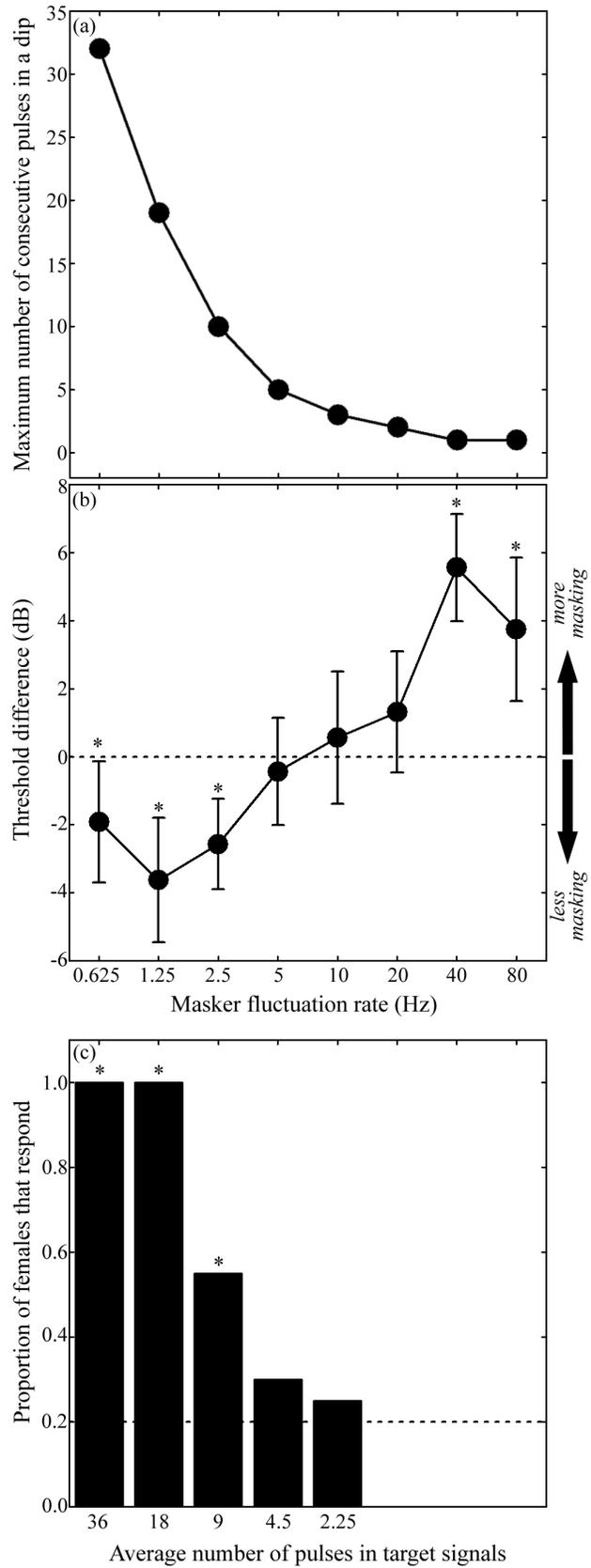


Figure 2-2. (a) Maximum number of consecutive pulses falling in the dips of fluctuating maskers of experiment 1 as a function of masker fluctuation rate. A pulse was considered as falling in a dip when its maximum amplitude fell between the 6 dB down points of the masker. Maximum values were calculated 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 ($\pm 95\%$ 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 represent less masking than in the nonfluctuating control, and values greater than 0 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 ($\alpha=0.05$).

pattern depicted in Fig. 2-2b. 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 and Schwartz 2009; Vélez and 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 semicircle 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 and 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. 2-2c) was significantly greater than the null expectation of 0.20 for calls containing nine or more pulses (one-tailed binomial tests: $P_s \leq 0.01$), but not five or fewer pulses ($P_s > 0.19$). 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 and 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 (fast RMS, C-weighted). 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 and 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 and 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 ($\leq 5^\circ$), and

lengths of mean vectors (r) close to 1.0 (Table 2-1). 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 78° (Table 2-1). 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 uniformly distributed, and instead would be directed, respectively, towards or away from the stimulus. As illustrated in Table 2-1 (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 and 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 2-1, 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

Table 2-1. Results of circular statistical analyses for orientation angles at the wall for the reference and test trials. 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).

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 (μ°)	Length of mean vector (r)	Circular SD ($^\circ$)	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

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 and Bee (2010) were significantly different from those measured in response to an attractive target signal in the first reference trial (see Table 2-1, two-sample Watson's U^2 tests, 'rel. Reference 1'). Responses in the two reference trials were not significantly different (Table 2-1). 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 treefrogs 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. 2-2a). 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. chrysosceslis*) 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; Bee, et al. in press.). 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. 2-2b). Females of *H. chrysozelis* use pulse rate as an acoustic property for species recognition (Schul and 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 and Marshall 2006). We attribute these results to a phenomenon known from the human hearing literature as 'modulation masking' (Bacon and Grantham 1989; Kwon and 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 and 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 and 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 and 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 short signals.

A key question arising from our results and those of Ronacher and 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 ($\approx 40\text{--}60$ Hz) (Vélez and 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 and Wiley 1980) and the production of temporally discrete calls by individual males in the chorus (Vélez and Bee 2010). The high rates of fluctuation reflect the pulse repetition rate of the grey treefrog advertisement call (Vélez and Bee 2010). Additional acoustic analyses of recordings of grey treefrog choruses revealed that high fluctuation rates ($\approx 40\text{--}60$ 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 (Appendix 1). 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 and 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 and 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 and Jouventin 1998). While this hypothesis has not been directly tested in penguins, our results and those of Ronacher and 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 and Micheyl 2008). In treefrogs, for example, spatial release from masking contributes to recognition of conspecific signals (Schwartz and Gerhardt 1989; Bee 2007a; Nityananda and Bee 2012) 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 and 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 and Slabbekoorn 2005; Schwartz and 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 and 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 and 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 and 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. 2-1e). 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.

Chapter 3 – Dip listening or modulation masking? Call recognition by green treefrogs (*Hyla cinerea*) in temporally fluctuating noise.³

Despite the importance of perceptually separating signals from background noise, we still know little about how nonhuman animals solve this problem. Dip listening, an ability to catch meaningful ‘acoustic glimpses’ of a target signal when fluctuating background noise levels momentarily drop, constitutes one possible solution. Amplitude-modulated noises, however, can sometimes impair signal recognition through a process known as modulation masking. We asked whether fluctuating noise simulating a breeding chorus affects the ability of female green treefrogs (*Hyla cinerea*) to recognize male advertisement calls. Our analysis of recordings of the sounds of green treefrog choruses reveal that their levels fluctuate primarily at rates below 10 Hz. In laboratory phonotaxis tests, we found no evidence for dip listening or modulation masking. Mean signal recognition thresholds in the presence of fluctuating chorus-like noises were never statistically different from those in the presence of a non-fluctuating control. An analysis of statistical effects sizes indicates that masker fluctuation rates, and the presence versus absence of fluctuations, had negligible effects on subject behavior. Together, our results suggest that female listeners in natural settings should receive no benefits, nor experience any additional constraints, as a result of level fluctuations in the soundscape of green treefrog choruses.

Introduction

In natural environments, acoustic communication is often compromised because abiotic noise (e.g. wind, rushing water, rustling leaves) and biotic noise (e.g. other signaling animals) impair signal detection, recognition, and discrimination (Klump 1996; Brumm

³ This chapter is currently under review as:
Vélez A, Höbel G, Gordon NM, and Bee MA. *In review*. Dip listening or modulation masking? Call recognition by green treefrogs (*Hyla cinerea*) in temporally fluctuating noise. *Journal of Comparative Physiology A*.

and Slabbekoorn 2005, Langemann and Klump 2005; Arch and Narins 2008; Schwartz and Freeberg 2008). Noise generated by other conspecific signals is probably an important source of auditory masking in animal social aggregations. Since the frequency spectrum of this type of biotic noise overlaps with that of the target signal, frequency filtering renders little gain in signal-to-noise ratio (Patterson and Moore 1986). Therefore, perceptual mechanisms that exploit spatial and temporal features of signals and noise could be particularly important in species that communicate in dense social aggregations of conspecifics. Accordingly, several studies of ‘spatial unmasking’ show that physical separation between target signals and background noise improves a receiver’s ability to detect or recognize acoustic signals, even when both have similar spectral content (reviewed in Bee and Micheyl 2008). In contrast, we still know little about perceptual mechanisms by which animals potentially exploit temporal features of signals and noise during signal recognition (Klump 1996; Hulse 2002; Bee and Micheyl 2008).

A well-known feature of natural sounds, including the noise generated by aggregations of signaling animals, is that their amplitudes fluctuate in time (Richards and Wiley 1980; Nelken et al. 1999; Vélez and Bee 2010). An ability to exploit level fluctuations in background noise contributes to human speech perception in noisy social settings. Speech recognition thresholds are usually lower when the level of speech-shaped noise (i.e. noise with the frequency spectrum of speech) fluctuates, compared to those when the same noise lacks level fluctuations (Gustafsson and Arlinger 1994; Bacon et al. 1998; Füllgrabe et al. 2006). Release from masking in temporally fluctuating noise is often referred to as ‘dip listening’ and is attributed to our ability to catch short ‘acoustic glimpses’ of target speech when the level of the background noise dips to low levels (Cooke 2006; Vestegaard et al. 2011). Masking release by means of dip listening is generally greater when noises fluctuate at slower, compared with faster, rates because the dips in noise levels are usually longer, increasing the probability of catching meaningful glimpses of target speech (Gustafsson and Arlinger 1994; Bacon et al. 1998). Importantly, however, level fluctuations in background noise do not always ameliorate speech

recognition in noise. Psychophysical studies of ‘modulation masking’ in humans reveal that fluctuating maskers can impair recognition of temporally modulated target signals, particularly when the fluctuation rates of signal and noise are similar (Bacon and Grantham 1989). Because speech recognition heavily depends on the signal’s temporal structure (Shannon et al. 1995), fluctuating background noise may sometimes impair speech recognition by means of modulation masking (Kwon and Turner 2001).

Psychophysical and neurophysiological studies have shown that level fluctuations in background noise can improve the ability of nonhuman animals to detect simple tonal signals and narrowband noises (Klump and Langemann 1995; Nelken et al. 1999; Langemann and Klump 2001; Nieder and Klump 2001; Hofer and Klump 2003; Bee et al. 2007; Jensen 2007; Langemann and Klump 2007; Branstetter and Finneran 2008; Fay 2011). However, only a few studies have investigated similar abilities in animals’ recognition of their acoustic communication signals. Studies of Bow-winged grasshoppers (*Chorthippus biguttulus*; Ronacher and Hoffmann 2003) and Cope’s gray treefrogs (*Hyla chrysoscelis*; Vélez and Bee 2011) indicate that receivers recognize temporally structured signals at lower thresholds in slowly fluctuating noise backgrounds. That is, Bow-winged grasshoppers and Cope’s gray treefrogs can listen in the dips under some noise conditions. In both species, signalers produce calls comprised of a series of distinct pulses and pulse structure mediates species recognition (Helvesen and Helvesen 1997; Schul and Bush 2002). Importantly, both Bow-winged grasshoppers and Cope’s gray treefrogs also experience modulation masking (i.e. higher thresholds relative to non-fluctuating controls) when masker fluctuation rates are similar to the pulse rates that mediate recognition of conspecific calls. These results raise important questions about the effects of fluctuating noise backgrounds in species with acoustic communication signals that do not comprise series of pulses.

Our primary objective in this study of green treefrogs (*Hyla cinerea*) was to investigate the extent to which dip listening and modulation masking affect signal recognition in a frog species with an advertisement call composed of a single note instead

of a pulse train. Male green treefrogs form dense breeding choruses in which they produce a short (120-160 ms) advertisement call repeated about once or twice per second (Oldham and Gerhardt 1975). The advertisement call produced by males is composed of a harmonic series with spectral peaks around 900 Hz and 3000 Hz (Oldham and Gerhardt 1975). While the interaction of spectral components generates some amplitude modulations in the call, the call is delivered as a single note and not as a series of pulses like the calls of Cope's gray treefrogs and Bow-winged grasshoppers. As in other anurans (Gerhardt and Huber 2002; reviewed in Vélez et al. in press), background chorus noise constrains signal recognition in green treefrogs (Gerhardt and Klump 1988a) and spatial unmasking leads to improvements in signal recognition by females (Schwartz and Gerhardt 1989).

We first recorded the sounds of green treefrog choruses and calculated their average modulation spectrum (Vélez and Bee 2010) to determine the rates at which chorus noise levels typically fluctuate. These analyses revealed that chorus noise levels fluctuate predominantly at rates below 10 Hz. We next conducted phonotaxis experiments to test the overarching hypothesis that level fluctuations in background noise affect the ability of female frogs to recognize male advertisement calls. We measured 'signal recognition thresholds' (Bee and Schwartz 2009) in the presence of 'chorus-shaped' noises that either lacked amplitude fluctuations or were sinusoidally amplitude modulated (SAM) at rates between 0.625 Hz and 80 Hz. We found little evidence to indicate that subjects experienced either dip listening (i.e., lower thresholds) or modulation masking (i.e., higher thresholds) in the fluctuating noise conditions. Together, our results suggest that fluctuation rates typical of the noise generated by green treefrog choruses should have little effect in terms of either facilitating (dip listening) or constraining (modulation masking) signal recognition by females in a chorus environment.

Materials and methods

Study sites and subjects

We made acoustic recordings and collected female green treefrogs (*H. cinerea*) in amplexus from artificial ponds at the Jasper State Fish Hatchery in Jasper, Texas (30° 57' 4.01" N, 94° 7' 39.53" W), between 2200 and 0200 hours in April and May 2010 and 2011, and July 2010. Females were kept at approximately 2 – 4°C in ice-filled coolers to delay oviposition and were transported overnight (within 36 – 48 hours of collection) to the St. Paul campus of the University of Minnesota, where they were tested using the same general protocols and under the same general conditions as in our previous study of dip listening and modulation masking in Cope's gray treefrogs (Vélez and Bee 2011). At least 1 hr prior to testing, females were placed in a 24°C incubator to allow their body temperature to reach 24°C ± 1°C. At the completion of testing, subjects were returned to the pond from which they were collected. All procedures for collecting, handling, and testing animals were approved by the University of Minnesota's Institutional Animal Care and Use Committee (IACUC No. 0809A46721, November 2008).

Acoustical analysis: Level fluctuations of natural chorus noise

Recordings of green treefrog breeding choruses were made during peaks of calling activity. Digital recordings (44100 Hz sampling rate, 16-bit resolution) were made using a Marantz PMD 670 recorder and an omnidirectional Sennheiser M62 microphone. We positioned the microphone 5 cm above ground level at distances ranging between 5 and 15 m from the nearest calling male. We chose a position close to the ground because females at our field site usually approach choruses from such positions. The range of distances to the nearest calling male was chosen to minimize the predominance of calls with high signal-to-noise ratios produced by a single nearby male. We chose for analysis 25 90-s segments of green treefrog choruses, each one taken from chorus recordings made on different days, at different ponds, or at different times and locations within a pond. We used recordings in which other species were absent or calling at very low

densities and away from the microphone (overall amplitude of heterospecific signals relative to green treefrog chorus noise less than -6 dB).

Chorus recordings were analyzed using custom-written scripts in Matlab v7.6 (Mathworks, MA, USA). We estimated level fluctuation rates by calculating a modulation spectrum for each chorus recording, which represents the power spectrum of the envelope of the chorus noise. To this end, we first down-sampled the 90-s recorded segments to a 11025 Hz sampling rate and normalized them to the same root-mean-square (RMS) amplitude to compensate for differences in microphone positions and recorder gain settings at the times recordings were made. We then extracted the Hilbert envelope of the waveform and, to correct for the DC offset, subtracted the mean value of the envelope from each sample of the envelope. To determine the rates of level fluctuation, we calculated the long-term modulation spectrum of each 90-s segment as the fast-Fourier transform (FFT) of its envelope (sampling rate = 11025 samples/s, Hamming window size = 65536 points, overlap = 25%). We normalized each spectrum to the maximum value of the magnitude of its FFT and then calculated the mean and standard deviation of the 25 modulation spectra. Mean and ± 1 standard deviation modulation spectra were transformed to a dB scale ($20 \cdot \log_{10}(\text{FFT magnitude})$) and smoothed for plotting purposes using a running average of 11 points.

General testing procedures

Our general testing procedures have been described previously and readers are referred to those studies for additional details (Bee and Schwartz 2009; Vélez and Bee 2011).

Briefly, we conducted no-choice phonotaxis experiments (Gerhardt 1995) under infrared (IR) illumination in a circular test arena (2 m diameter) made of acoustically transparent but visually opaque walls (60 cm height). The test arena was placed inside a walk-in, temperature-controlled ($24^{\circ}\text{C} \pm 1^{\circ}\text{C}$), hemi-anechoic sound chamber (details in Bee and Schwartz 2009). Using an IR-sensitive video camera mounted above the center of the test arena, behavioral responses were simultaneously encoded to digital video files and scored in real time by two observers on a monitor outside the chamber.

We divided the perimeter of the test arena into 24 15° arcs. The speaker used to broadcast target signals (a/d/s/ L210 or Orb Mod1) was placed on the floor just outside the wall of the arena, 1 m away from a release point at the center of the arena, and centered in one of the 15° arcs. To eliminate any possibility of directional response bias, the position of the target speaker was randomly varied around the arena's perimeter between tests of two to four females. Maskers were broadcast from a Kenwood KFC-1680ie speaker suspended from the ceiling of the chamber 190 cm above the central release point. This overhead speaker created a fairly uniform (± 2 dB) sound field across the floor of the circular test arena. Our playback system had a flat (± 3 dB) frequency response between 500 Hz and 5000 Hz. We used a Brüel and Kjær Type 2250 sound level meter to measure and calibrate sound levels at the approximate position of a subject's head at the central release point.

At the beginning of each test, we placed a subject in an acoustically transparent holding cage located at the arena's central release point. After a 1.5-min silent acclimation period, broadcasts of the target signal were initiated. Subjects were released after 15 s of signal presentation using a rope and pulley system operated from outside the chamber. In tests in which a masking noise was used, broadcasts of the masker began 30 s before the onset of the target signal and continued throughout the entire test. Each subject participated in a series of 6-17 phonotaxis tests and was given a timeout period of 5 to 15 minutes inside the incubator between tests. Unless noted otherwise, we scored a correct response in a phonotaxis test if (i) the subject first touched the wall of the test arena in the hemi-circle containing the speaker broadcasting the target signal, (ii) the subject made contact with the arena's wall inside the 15° bin in front of the speaker broadcasting the target signal within 5 min of being released, and (iii) the subject remained for 30 s in a bin of approximately 30° width centered in front of the speaker.

Acoustic stimuli

All acoustic stimuli used in this study were generated in Matlab v7.6 (Mathworks, MA, USA) with a sampling rate of 11025 Hz and 16-bit resolution.

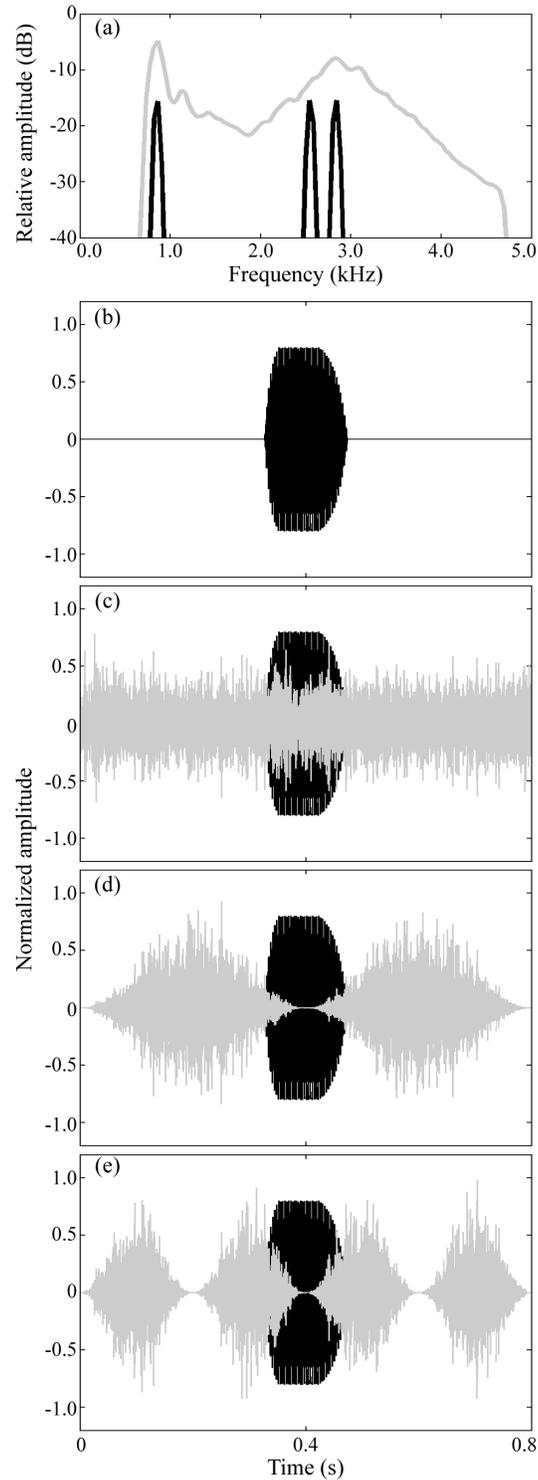


Figure 3-1. (a) Frequency spectra of the standard call used as a target signal (black line) and the chorus-shaped maskers (gray line) calculated with 256-point Hamming windows. Waveforms of 0.8 s segments of the standard call alone (b), and in the presence of a non-fluctuating masker (c), a 2.5 Hz SAM masker (d), and a 5 Hz SAM masker (e).

The standard call

The target signal was a standard synthetic call with values of temporal and spectral properties close to the averages of green treefrog calls recorded at our study site (Höbel unpublished data; Fig. 3-1a, b). The standard call comprised three phase-locked sinusoids with equal relative amplitudes and frequencies of 850 Hz, 2550 Hz, and 2833 Hz. The call was 147 ms in duration and repeated at a rate of 1.78 calls/s (562 ms call period). The amplitude-envelope of the call was shaped with a 25-ms rise time and a 50-ms fall time (50% amplitudes were reached at 25% of the rise time and at 75% of the fall time).

Chorus-shaped maskers

We created chorus-shaped maskers by filtering white noise so that the long-term frequency spectrum resembled that of natural green treefrog choruses (Fig. 3-1a). As models of chorus noise, we selected 10 100-s recordings of green treefrog choruses obtained at our study site near the peak of nightly vocal activity using the recording protocols described above. Recordings were down-sampled to 11025 Hz, normalized to a common RMS amplitude, and transformed to the spectral domain (FFT size = 4194304). We created an exemplar of chorus-shaped noise by generating a 6-min white noise (sampling rate = 11025 Hz), transforming it to the spectral domain (FFT size = 4194304), and multiplying its frequency spectrum by the average frequency spectrum determined from our 10 recordings of natural choruses. The resulting frequency spectrum was band-pass filtered between 700 Hz and 4,700 Hz and transformed to the time domain with an inverse FFT. The resulting noise had a frequency spectrum typical of green treefrog choruses and a flat temporal envelope with no amplitude fluctuations beyond the random inherent fluctuations typical of band-limited noise (Fig. 3-1c). We refer to this type of noise as ‘non-fluctuating’ chorus-shaped noise because we did not impose any modulations on the resulting waveforms. For our experiments, we created four exemplars of non-fluctuating chorus-shaped noises using a different white noise for each.

Fluctuating maskers were generated by imposing sinusoidal amplitude modulation (SAM) at rates of 0.625 Hz, 1.25 Hz, 2.5 Hz (Fig. 3-1d), 5 Hz (Fig. 3-1e), 10 Hz, 20 Hz,

40 Hz, and 80 Hz on all four exemplars of non-fluctuating chorus-shaped noise. The use of SAM noise is common in psychophysical and neurophysiological studies of temporal auditory processing due to the high level of experimental control they offer (Joris 2004). We created SAM noises by multiplying the non-fluctuating chorus-shaped noises by a modulating sine wave following the equation:

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

where $n(t)$ is the resulting SAM noise, 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 non-fluctuating chorus-shaped noise, and t is time in s. The starting phase of the modulator differed for each of the four exemplars ($\rho = 0^\circ, 90^\circ, 180^\circ, \text{ or } 270^\circ$). We used a different starting phase for each exemplar because, during phonotaxis tests, the number of calls falling within dips of fluctuating maskers could depend on the starting phase of the masker and the timing of the target signal, potentially affecting signal recognition. Equal numbers of subjects were tested with each exemplar and starting phase was treated as a between-subjects factor in preliminary statistical analyses. Neither ‘starting phase’ nor the interaction between ‘starting phase’ and ‘masking treatment’ had a significant effect on signal recognition thresholds; therefore, we dropped these factors from the analysis of variance (ANOVA) models reported below (starting phase: $F_{3,180} = 1.36, P = 0.26, \text{ partial } \eta^2 = 0.02$; starting phase \times masking treatment: $F_{24,180} = 0.44, P = 0.99, \text{ partial } \eta^2 = 0.06$).

Experiment 1: Call recognition in fluctuating chorus-shaped maskers

In two previous studies of green treefrogs, call recognition thresholds were estimated in the presence (Gerhardt and Klump 1988a) and absence (Gerhardt 1981) of chorus noise using two different procedures. Therefore, our first objective in this experiment was to measure signal recognition thresholds in the presence and absence of non-fluctuating chorus-shaped maskers using the same standardized protocol (Bee and Schwartz 2009).

The second, and main, objective of this experiment was to determine the extent to which level fluctuations in background noise affect the ability of female green treefrogs

to recognize advertisement calls. We tested the null hypothesis that level fluctuations do not affect signal recognition thresholds against two alternative hypotheses: the dip-listening hypothesis and the modulation-masking hypothesis. According to the dip-listening hypothesis, we predicted lower signal recognition thresholds in the presence of fluctuating maskers compared with those measured in the presence of non-fluctuating maskers. In contrast, if females experienced modulation masking, we predicted higher signal recognition thresholds in the presence of fluctuating maskers compared with those measured using non-fluctuating maskers.

Experimental design

Using a between-subjects experimental design, we tested 24 subjects in each of ten randomly assigned treatments (total $N = 240$). A ‘no-masker’ treatment, in which no masking noise was broadcast, served as a control to measure signal recognition thresholds in the absence of background noise. This control treatment allowed us to evaluate the effects of our chorus-shaped maskers on subjects’ responses to the target signal. (We describe generation of signal recognition thresholds in the next section.) In a second treatment, we determined thresholds in the presence of non-fluctuating chorus-shaped noise broadcast from the overhead speaker. This ‘non-fluctuating noise’ treatment served two purposes. First, we used this treatment to estimate the difference in signal recognition thresholds in the presence and absence of chorus-shaped noise. Second, this treatment served as a control to assess the effects of level fluctuations in masking noise on subjects’ ability to recognize the target signal. In the remaining eight treatments, we measured thresholds in the presence of one of the eight SAM chorus-shaped maskers (i.e. 0.625 Hz – 80 Hz SAM in octave steps) broadcast from the overhead speaker. We refer to these eight treatments as ‘fluctuating noise’ treatments. In all nine treatments that involved broadcasts of a chorus-shaped masker, the equivalent long-term RMS amplitude of the masker was calibrated at the central release point of the arena to a sound pressure level (SPL re. 20 μ Pa) of 73 dB (LC_{eq}). This level falls within the range of chorus noise levels measured in the field (Vélez and Bee unpublished data). Individual subjects were

tested in a series of six to 17 tests that included ‘reference’, ‘test’, and ‘sham’ trials (detailed below).

Signal recognition thresholds

As in previous studies (Bee and Schwartz 2009; Vélez and Bee 2011; Nityananda and Bee 2012), we operationally defined the signal recognition threshold as the lowest signal level necessary to elicit positive phonotaxis to the target signal. As defined here, signal recognition thresholds differ from traditional signal detection thresholds because positive phonotaxis requires that subjects (i) detect the signal, (ii) recognize it as a conspecific advertisement call, and (iii) localize its source. Following Bee and Schwartz (2009), we estimated signal recognition thresholds using an adaptive tracking procedure in which subjects were tested in a series of reference trials and test trials. The total number of trials in a series depended on the subjects’ behavioral responses. Each series of trials began and ended with a reference trial. In reference trials, the target signal was the standard call broadcast at 85 dB SPL (LCF) in the absence of masking noise. This signal level corresponds to natural call amplitudes measured at 1 m (Gerhardt 1975). Subjects were also tested in a reference trial if they failed to meet our response criteria in any two consecutive test trials during a series of trials. Subjects failing to meet the response criteria in any reference trial were replaced and their data were discarded. This procedure ensures that all subjects included in the analyses were motivated to respond throughout all trials and is necessary to validate test trials in which subjects fail to respond to the target signal (Bush et al. 2002).

The trial following the first reference trial was a ‘sham trial,’ in which we tested the subject without presenting a target signal. (Sham trials are described in more detail in a subsequent section.) Following the sham trial, subjects began a series of ‘test trials.’ In the first test trial of a series, the target signal was broadcast at 48 dB SPL (LCF) in the no-masker treatment and 73 dB SPL (LCF) in all treatments with a masker. These levels were selected based on previous estimates of signal recognition thresholds in the presence and absence of noise for our study species (Gerhardt 1981; Gerhardt and Klump 1988a).

In all subsequent test trials, the level of the target signal was systematically varied contingent upon the subject's behavior in the previous test trial. If the subject responded in a test trial, the level of the signal was decreased by 3 dB on the next test trial. If, on the other hand, the subject failed to respond in a test trial, the level of the target signal was increased by 3 dB on the next test trial. We continued this pattern until there was a change in the subject's behavior between two consecutive test trials (i.e. from response to no response, or vice versa). After a subject changed its behavior between two test trials, we conducted a final test trial in which the direction of signal level change was reversed and the step size was reduced to 1.5 dB. We calculated a subject's recognition threshold as the average between the minimum signal level that elicited positive phonotaxis and the maximum signal level that failed to do so (Bee and Schwartz 2009). Signal recognition thresholds were calculated in a linear scale and converted back to decibels.

Statistical analyses

We compared signal recognition thresholds in the presence and absence of non-fluctuating chorus-shaped maskers with a Mann-Whitney U test. Signal recognition thresholds in these two treatments were normally distributed (Kolmogorov-Smirnov tests $P_s > 0.2$) but did not meet the homogeneity of variance assumption for parametric testing (Levene's test: $F_{1, 46} = 18.827, P < 0.001$).

To examine the extent to which level fluctuations affected masked signal recognition thresholds, we first analyzed thresholds in the non-fluctuating control treatment and all the fluctuating noise treatments using a one-way ANOVA. Signal recognition thresholds in these treatments were normally distributed (Kolmogorov-Smirnov tests $P_s > 0.05$) and homoscedastic (Levene's test: $F_{8, 207} = 0.78, P = 0.62$). We used planned contrasts to test the dip-listening and modulation-masking hypotheses, which make opposite predictions about the differences between thresholds measured in fluctuating and non-fluctuating maskers. The dip listening hypothesis predicts relatively lower thresholds in the presence of fluctuating maskers, whereas the modulation-masking

hypothesis predicts relatively higher thresholds in fluctuating maskers. A significance criterion of $\alpha = 0.05$ was used in all tests.

Sham trials: Subject behavior in the absence of target signals

We tested females in sham trials to observe their behavior in our experimental apparatus in the presence of noise but in the absence of a specific target signal. This control treatment was important because the ‘noise’ generated in a chorus can actually serve as a ‘signal’ that allows individuals of some species of frogs to locate breeding sites (Gerhardt and Klump 1988b; Bee 2007; Swanson et al. 2007; Christie et al. 2010). Consequently, chorus sounds and chorus-like noises potentially affect subject behavior during signal recognition experiments. In the present experiment, such behavior could have biased our results if subjects behaved in ways that affected estimates of recognition thresholds (e.g., by exhibiting more active searching behavior) in the presence of some, but not all, of the maskers. Sham trials allowed us to evaluate this possibility.

Note that sham trials in the no-masker treatment were designed to provide a baseline of subject behavior in the sound chamber without any acoustic stimulation. During sham trials of all other treatments, each masker was broadcast continuously from the overhead speaker with long-term RMS amplitude of 73 dB SPL (LC_{eq}). Since no stimuli were broadcast from speakers on the floor, there was no ‘correct’ response in any sham trials. Therefore, we did not apply any response criteria in sham trials and recorded female behavior for five minutes after her release.

We used two approaches to test the null hypothesis that movement patterns were not affected by the presence/absence or type of masker broadcast from the overhead speaker. First, we compared the proportion of subjects that left the holding cage and showed any type of movement in the arena. Our reasoning was that, if the sounds of some chorus-shaped maskers caused the subjects to show more or fewer overall movements, we would find a difference in the proportion of females leaving the holding cage across treatments. We used Fisher’s exact tests to test the null hypothesis that the proportion of

females leaving the holding cage in each masking treatment was not different from the proportion of females doing so in the no-masker treatment (i.e., when there was no sound present at all). Second, we used the animal tracking software EthoVision v3.1 (Noldus 2005) to compare movement patterns of the subjects that left the holding cage. We measured the total distance moved by the subjects (in cm), the average velocity of their movements (in cm/s), the average turn angles (in degrees) and the meander (in degrees/cm), which quantifies the magnitude of changes in direction of movements relative to the distance moved (see Bee and Riemersma 2008). Because these four variables were intercorrelated, we used principal component analysis (PCA) to examine the extent of covariation and to obtain independent factors describing movement patterns. We investigated movement patterns across treatments using Kruskal-Wallis tests.

Experiment 2: Chorus-shaped maskers as potential signals

As previously mentioned, chorus sounds may act as signals used by individuals to locate breeding sites (Gerhardt and Klump 1988b; Bee 2007; Swanson et al. 2007; Christie et al. 2010). Ambient sounds can also be used by some frogs to evaluate potential threats. For instance, the temporal structure of the cracking sounds of burning vegetation appears to be an important cue used by juvenile reed frogs (*Hyperolius nitidulus*) to flee from approaching fire (Grafe et al. 2002). Consequently, temporally structured sounds other than individual communication signals can have behavioral salience to frogs and potentially ‘compete’ with the target signal in signal recognition experiments. We evaluated this possibility in a control experiment in which we tested the null hypothesis that the noises we used as maskers in experiment one fail to elicit positive or negative phonotaxis when broadcast as potential target signals.

Using a within-subject experimental design, we tested 20 subjects in nine treatments, each corresponding to one of the chorus-shaped noises used in experiment one. Individual females were tested in a series of 11 trials: two reference trials (initial and final) and nine intervening test trials. In each test trial, a chorus-shaped noise was broadcast continuously from a speaker on the floor just outside the arena wall. We

calibrated the long-term RMS amplitude of the chorus-shaped noises to 73 dB SPL (LCeq) at the central release point of the test arena. The 15° bin where the speaker was placed was designated 0°, and we measured the angle at which females first touched the wall of the arena in 15° steps. The position of the target speaker was randomly varied between groups of two or three subjects. As in previous similar studies (Swanson et al. 2007; Vélez and Bee 2010, 2011), we ended the test trials as soon as subjects made contact with the wall anywhere around the test arena or after five minutes of stimulus presentation. For each subject, we randomized the order in which the chorus-shaped noises were presented. We tested the null hypothesis that orientation angles were uniformly distributed using Rayleigh tests. We reasoned that, if subjects were attracted to or repelled by the chorus-shaped noises, orientation angles would not be uniformly distributed, but would be directed toward or away from the stimulus, respectively.

Results

Acoustical analysis: Level fluctuations of natural chorus noise

The noise generated by green treefrog choruses fluctuates in level (Fig. 3-2a, b, c). The average modulation spectrum of green treefrog choruses is characterized by low-frequency modulation rates (below approximately 10 Hz) with a prominent peak around 5 Hz (Fig. 3-2d). Analyses of modulation spectra also revealed a second peak of level fluctuations around 300 Hz, with an amplitude of approximately -18 dB relative to the dominant peak at 5 Hz (Fig. 3-2d).

Experiment 1: Call recognition in fluctuating chorus-shaped maskers

Signal recognition thresholds

Signal recognition thresholds were significantly higher (by approximately 20 dB) in the presence of the non-fluctuating masker than in the absence of noise (Mann-Whitney U Test: $U = 19.0$, $P < 0.0001$). The median signal recognition threshold was 53.3 dB (mean = 54.6 dB; Fig. 3-3) in the no-masker control treatment and 73.8 dB (mean = 73.4; Fig. 3-3) dB in the non-fluctuating control treatment.

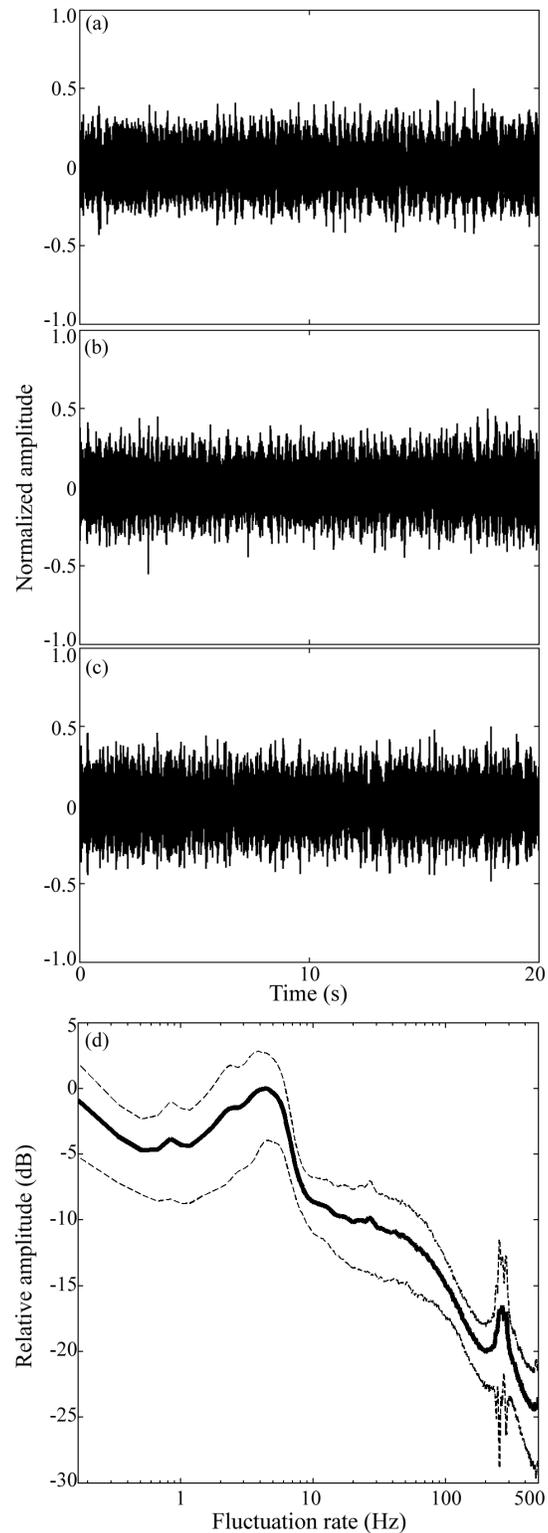


Figure 3-2. (a, b, c) Waveforms of three representative green treefrog chorus recordings. (d) Mean (solid line) ± 1 standard deviation (dashed lines) modulation spectrum calculated from 25 90-s segments of green treefrog chorus recordings shown on a logarithmic x axis.

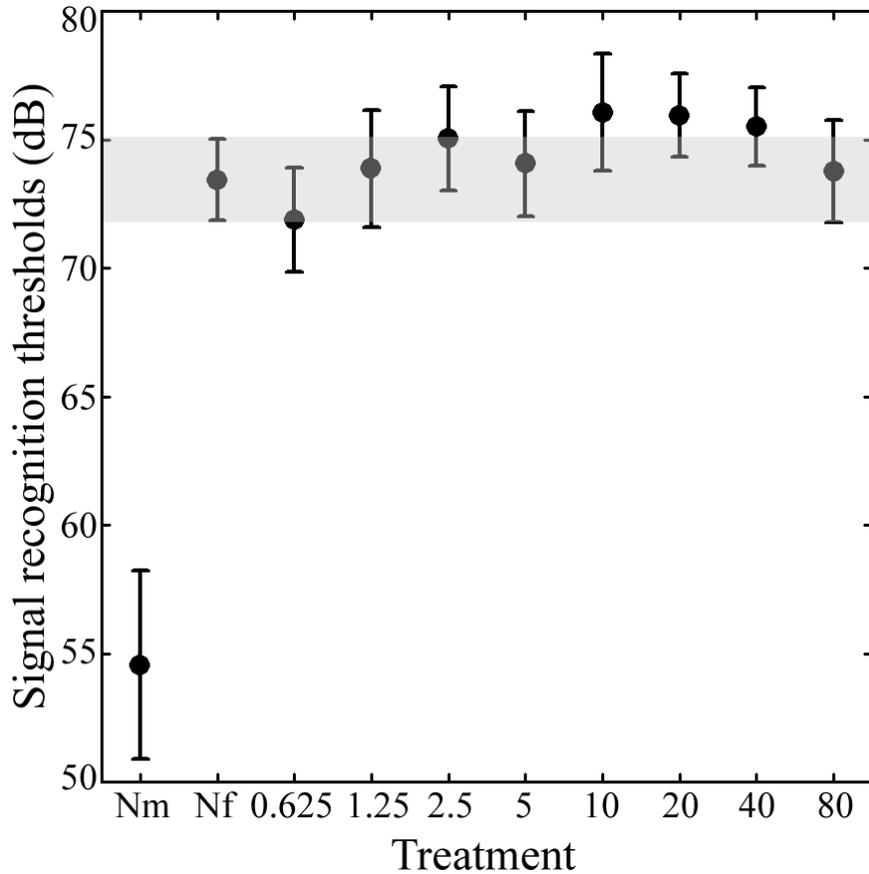


Figure 3-3. Mean ($\pm 95\%$ confidence intervals) signal recognition thresholds as a function of masking treatment. The gray box serves to illustrate the $\pm 95\%$ confidence intervals in the non-fluctuating control treatment. Nm = No-masker treatment; Nf = Non-fluctuating treatment. 0.625 – 80 represent SAM rates of each fluctuating noise treatment.

On average, signal recognition thresholds in the presence of fluctuating maskers ranged from about 1.5 dB lower to 2.5 dB higher than the average recognition threshold in the non-fluctuating control treatment (Fig. 3-3). We found a significant effect of masker type on signal recognition thresholds (one-way ANOVA: $F_{8, 207} = 2.09$, $P = 0.038$, $\eta^2 = 0.08$). Individual planned comparisons failed to reveal any significant differences between signal recognition thresholds in the non-fluctuating control treatment and each fluctuating noise treatment; these differences approached statistical significance at masker fluctuation rates of 10 Hz ($F_{1, 207} = 3.87$, $P = 0.0504$, partial $\eta^2 = 0.018$) and 20 Hz ($F_{1, 207} = 3.51$, $P = 0.0623$, partial $\eta^2 = 0.017$), but not any of the other fluctuation rates (all $F_{S1, 207} < 2.4$, $P_s > 0.1$, partial $\eta^2_s < 0.012$).

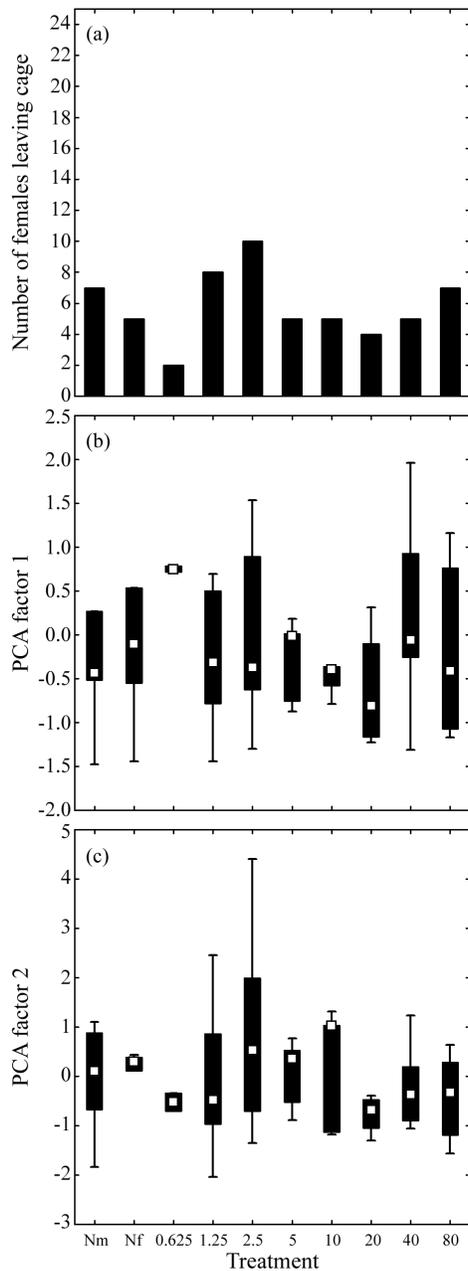


Figure 3-4. (a) Number of females leaving the holding cage during sham trials as a function of masking treatment. The total number of females tested in each treatment was 24. Median (point), inter-quartile range (box), and non-outlier range (whiskers), values of PCA factor 1 (b) and PCA factor 2 (c) as a function of masking treatment. PCA factor 1 describes angular movements and loads on the variables ‘average turn angle’ and ‘meander’. PCA factor 2 describes non-angular movements and loads on the variables ‘total distance moved’ and ‘average velocity’. Sample sizes for each treatment in (b) and (c) correspond to the number of females leaving the holding cage depicted in (a). Nm = No-masker treatment; Nf = Non-fluctuating treatment. 0.625 – 80 represent SAM rates of each fluctuating noise treatment.

Sham trials: Female behavior in the absence of target signals

We found no significant differences in movement patterns between masking treatments in the sham trials (Fig. 3-4). Seven of 24 subjects (29.2%) left the holding cage during the sham trial in the no-masker treatment. In treatments with masking noise, between two (8.3%) and 10 (41.7%) of 24 subjects left the holding cage during the sham trials (Fig. 3-4). The proportions of females that left the holding cage in treatments that included a masker were not significantly different from that in the no-masker treatment (two-tailed Fisher's exact tests: all $P_s > 0.13$).

The four variables of movement patterns measured (i.e., total distance moved, average velocity, average turn angle, meander) could be summarized by two principal components with eigenvalues greater than 1.0, which together explained 83.2% of the total variance. The first PCA factor accounted for 49.04% of the total variance, had an eigenvalue of 1.96, and loaded most heavily on variables related to angular movements: meander (factor-variable correlation = 0.96), mean turn angle (0.95), mean velocity (-0.31), and total distance moved (0.18). The second PCA factor had an eigenvalue of 1.37, described an additional 34.16% of the total variance, and loaded most heavily on variables related to non-angular movements: total distance moved (0.84), mean velocity (0.79), mean turn angle (0.16), and meander (-0.07). We found no differences across treatments in PCA factor 1 (angular movements; Kruskal-Wallis test: $H_{9, 58} = 5.28$, $P = 0.81$) or PCA factor 2 (non-angular movements; $H_{9, 58} = 5.82$, $P = 0.76$) during sham trials (Figure 3-4). Parallel analyses on the original movement-pattern variables confirmed these results (Appendix 2).

Experiment 2: Chorus-shaped maskers as potential signals

Results from this control experiment are summarized in Table 3-1. Subjects were strongly oriented to the target speaker in reference trials, as evidenced by mean vector angles close to 0° and lengths of mean vectors close to 1.0 (Rayleigh test: $Z_s > 18$, $P_s < 0.001$; Table 3-1). Subjects were neither attracted to, nor repelled by, the non-fluctuating noise or by noises that fluctuated at rates of 0.625 Hz, 10 Hz, 20 Hz, 40 Hz, and 80 Hz, as evidenced

Table 3-1. Results of circular analyses for orientation angles at the arena wall in response to reference trials and chorus-shaped noises. *N* represents the number of subjects that touched the wall of the arena in each treatment; the total number of subjects tested per treatment was 20.

Treatment	Circular analysis					
	N	Mean vector (°)	Length of mean vector (r)	Circular standard deviation (°)	Rayleigh test <i>Z</i>	Rayleigh test <i>P</i>
Reference 1	20	0	0.99	5	19.9	<0.0001
Non-fluctuating	1	135	1.00	--	1.0	0.51
0.625 Hz SAM	3	249	0.70	48	1.5	0.25
1.25 Hz SAM	13	3	0.62	56	4.9	<0.01
2.5 Hz SAM	14	3	0.96	16	12.9	<0.01
5 Hz SAM	11	7	0.81	37	7.3	<0.01
10 Hz SAM	3	341	0.88	29	2.3	0.09
20 Hz SAM	5	226	0.21	101	0.2	0.81
40 Hz SAM	5	220	0.64	54	2.0	0.13
80 Hz SAM	3	115	0.45	73	0.6	0.59
Reference 2	20	4	0.96	16	18.5	<0.0001

by uniformly distributed orientation angles ($P_s \geq 0.09$; Table 3-1). Orientation angles, however, were not uniformly distributed when chorus-shaped noises fluctuated at rates of 1.25 Hz, 2.5 Hz, and 5 Hz ($P_s < 0.01$; Table 3-1). While the proportions of subjects that touched the wall of the arena in these three treatments (between 0.55 and 0.70) were lower than that in the two reference trials (1.0 in both cases), mean angles close to 0° and lengths of mean vectors greater than 0.6 suggest that the females that touched the wall of the arena were attracted to these chorus-shaped noises.

Discussion

In this study of green treefrogs, we found an increase of approximately 20 dB in signal recognition thresholds in the presence of chorus-shaped maskers broadcast at natural levels. However, we also found that level fluctuations in background noise had little effect on the ability of female green treefrogs to recognize male advertisement calls. Mean signal recognition thresholds in fluctuating noise treatments were never statistically different from those in the non-fluctuating control treatment. These results are thus inconsistent with both the dip-listening and the modulation-masking hypotheses. Studies

that report null results, like this one, are often open to criticisms regarding inadequate sample sizes (and thus, low statistical power), inappropriate experimental designs, or both. Therefore, we consider these issues next.

Statistical power and experimental design

Statistical power refers to the probability of rejecting a false null hypothesis and varies as a function of effect size and sample size at a given alpha level (Cohen 1988; Rosenthal and Rosnow 1991). A common effect size for the F statistic from ANOVA is partial η^2 , which describes the proportion of total variance in the dependent variable accounted for by a given factor. Our study had adequate statistical power to detect a relatively small overall effect of masker type on signal recognition thresholds (partial $\eta^2 = 0.08$). Even with our relatively large sample size (total $N = 240$), however, the statistical power of our focused contrasts was not high enough to detect the very small effects corresponding to differences between the non-fluctuating control treatment and each of the fluctuating noise treatments. The differences between mean signal recognition thresholds in the non-fluctuating control treatment and all fluctuating noise treatments ranged between 0.3 dB and 2.6 dB, and the corresponding statistical effect sizes (partial η^2) ranged from less than 0.001 to 0.018 (Appendix 3). These effect sizes are considerably smaller than some of those found in a previous study of Cope's gray treefrogs using the same experimental design, sample size and statistical analyses (Vélez and Bee 2011; Appendix 3). In Cope's gray treefrogs, mean threshold differences between the non-fluctuating control treatment and all fluctuating noise treatments ranged from 0.4 dB to 5.6 dB. The corresponding statistical effect sizes ranged between 0.001 and 0.096. Therefore, the sample size of the present study provided sufficient statistical power to detect the small to moderate effect sizes observed in Cope's gray treefrogs, but not the smaller effects reported here for green treefrogs. Together, our results suggest the effects of temporally fluctuating maskers on signal recognition thresholds are smaller in green treefrogs compared with Cope's gray treefrogs. Previous studies suggest that differences of 2 to 4 dB in signal levels are biologically important for females listening to potential mates in natural settings

(Gerhardt et al. 2000; Vélez and Bee 2011; Bee et al. in review). Smaller effects, as the ones we report for green treefrogs, are less likely to be biologically meaningful in the din of a breeding chorus.

Regarding the experimental design, we followed the same protocol as in Vélez and Bee (2011), which proved adequate to demonstrate both dip listening and modulation masking in Cope's gray treefrogs. There is, however, one caveat potentially affecting interpretations of results from the present study of green treefrogs: three of the eight fluctuating maskers (1.25 Hz, 2.5 Hz and 5 Hz) were behaviorally salient when presented as potential target signals in experiment two. In that experiment, females exhibited significant orientation toward these three maskers (Table 3-1), indicating that these sounds were attractive and had some potential to 'compete' with the target signals used to measure signal recognition thresholds in experiment one. Interpreting results from the 1.25 Hz, 2.5 Hz and 5 Hz treatments of experiment 1, therefore, should be done with some caution. For two reasons, we believe the degree of 'competition' for subjects' responses between these three maskers and the standard call used to measure signal recognition thresholds in experiment one was likely small. First, across the 1.25 Hz, 2.5 Hz and 5 Hz treatments of experiment two, only between 55% and 70% of subjects reached the arena wall within five minutes (Table 3-1). By comparison, approximately 96% of females did so in response to the standard call presented at an equivalent (i.e., 73 dB) or *lower* level in the no-masker treatment of experiment one (data not shown). Hence, the 1.25 Hz, 2.5 Hz and 5 Hz maskers should be regarded as marginally attractive at best. Second, and perhaps more importantly, results from the sham trials of experiment one showed that subject movement patterns were not differentially affected by the type of chorus-shaped noise broadcast from the overhead speaker (Figure 3-4, Appendix 2). Together, these results suggest that the potential confounding effects of the 1.25 Hz, 2.5 Hz and 5 Hz maskers used in measuring signal recognition thresholds in experiment one were probably small. Nevertheless, we cannot completely rule out the possibility that these three maskers competed with the target signal, affecting our estimates of signal

recognition thresholds in these three (of eight) treatments. With this caveat in mind, it is worthwhile discussing how our results relate to those of previous studies and the implications of our results for communication in natural environments in green treefrogs.

Comparisons with previous studies

Signal recognition in green treefrogs

In the absence of noise, we estimated an average signal recognition threshold of 54.6 dB (median = 53.3 dB). This estimate is similar, but somewhat higher, than the 48 dB behavioral recognition threshold reported by Gerhardt (1981). The difference between the two threshold estimates is likely due to methodological differences. Gerhardt (1981) did not follow an adaptive tracking method as we did, but instead tested females at different fixed signal levels. He determined 48 dB as a reasonable behavioral threshold because approximately 36% of the females approached the target speaker within 30 cm at this signal level. In our experiment, 33% of the females tested approached the speaker within 30 cm at a signal level of 48 dB. Therefore, even though the criteria used in our study and that of Gerhardt (1981) yield different threshold estimates, female behavior in both studies was nevertheless quite similar. Our estimates of signal recognition thresholds in the presence of background noise, when expressed as signal-to-noise ratios, are similar to those reported previously. In the presence of non-fluctuating chorus-shaped maskers, we estimated an average signal recognition threshold of 73.4 dB. Since our maskers were broadcast at 73 dB SPL (LC_{eq}), this masked signal recognition threshold corresponds to a signal-to-noise ratio of 0.4 dB. Using a short recording of chorus noise as a masker, Gerhardt and Klump (1988a) determined that female green treefrogs recognize calls embedded in background noise when the signal-to-noise ratio was at least 0 dB. While there is a difference in the temporal structure between our non-fluctuating chorus-shaped masker and their chorus recording, it may not have a strong effect on signal recognition, as will be discussed shortly. Our results corroborate Gerhardt and Klump's (1988a) conclusion that males that produce calls at sound levels lower than the level of the ambient chorus noise have little chance of attracting females.

Dip listening and modulation masking in nonhuman communication systems

Dip listening and modulation masking in the context of nonhuman animal communication have only been investigated in Bow-winged grasshoppers (Ronacher and Hoffmann 2003) and Cope's gray tree frogs (Vélez and Bee 2010, 2011). Female Bow-winged grasshoppers produce stridulatory mating signals composed of a series of pulsed syllables. Each syllable is repeated at a rate of approximately 10 times per second and the pulses in each syllable are delivered at a rate of about 70 pulses per second (Ronacher and Krahe 1998; Ronnacher and Hoffmann 2003). In the presence of slowly fluctuating SAM maskers, male Bow-winged grasshoppers experienced masking release when the target signal was long (12 syllables), but not when it was short (4 syllables). These results offer partial support for the dip-listening hypothesis and suggest that signal duration affects signal recognition in fluctuating noise in grasshoppers. Interestingly, males experienced 5 dB to 10 dB of modulation masking with both short and long signals when background noise levels fluctuated at rates above 10 Hz. Consistent with studies on modulation masking in humans (Bacon and Grantham 1989, 1992; Millman et al. 2002), these results suggest that modulation masking in male grasshoppers is especially pronounced when there are similar rates of fluctuation in the background noise and the target signals (e.g. pulse and syllable repetition rates).

Likewise, females of Cope's gray treefrogs (*H. chrysoxcelis*) exhibit patterns of dip listening and modulation masking similar to those reported in humans. Male gray treefrogs produce pulsed advertisement calls repeated about once every five seconds. Each call comprises a series of about 30 pulses delivered at a rate of approximately 45 pulses per second. In a first study with Cope's gray treefrog females, Vélez and Bee (2010) found an effect of modulation masking of approximately 12 dB in the presence of 40 Hz SAM maskers. In a subsequent study of Cope's gray treefrogs, Vélez and Bee (2011) used the same protocol as that used in the present study of green treefrogs. Female Cope's gray treefrogs experienced a release from masking of 2 to 4 dB in the presence of

SAM chorus-shaped maskers that fluctuated at rates below 5 Hz, and about 4 dB to 6 dB of modulation masking when maskers fluctuated at rates above 20 Hz. Masking release in gray treefrogs in slowly fluctuating backgrounds was attributed to the ability of female frogs to catch meaningful acoustic glimpses of the advertisement call. Nine or more consecutive pulses of the gray treefrog call fell within dips (measured at the 6-dB downpoints) of maskers fluctuating at rates of 2.5 Hz or slower. Vélez and Bee (2011) showed that, in the absence of background noise, calls with nine or more consecutive pulses were necessary to elicit phonotaxis. The strong effect of modulation masking observed in gray treefrogs in the 40 Hz and 80 Hz SAM treatments was attributed to the interference caused by background noise level fluctuations at rates similar to those in the pulsed call. This result is consistent with other studies of call interference in gray treefrogs showing that overlapping calls disrupt perception of pulse rate by females (Marshall et al. 2006; Schwartz and Marshall 2006).

In contrast, our results with green treefrogs offer little support for the operation of dip listening or modulation masking in this species. In the present study, the maximum number of consecutive calls (call duration = 147 ms, call period = 562 ms) falling between the 6-dB down points of dips in our maskers was two calls in the 0.625 Hz SAM treatment and fell to one call in the 1.25 Hz and 2.5 Hz SAM treatments; complete calls did not acoustically ‘fit’ in the shorter dips of maskers modulated at rates faster than 2.5 Hz. Therefore, we might have expected dip listening to occur in the presence of the most slowly fluctuating maskers, as reported for Cope’s gray treefrogs and Bow-winged grasshoppers. However, this was not the case (Fig. 3-3). One possible explanation for the lack of masking release in the 0.625 Hz, 1.25 Hz, and 2.5 Hz SAM treatments is that one or two consecutive calls without intervening noise are not enough to elicit phonotaxis in green treefrogs. In other words, even in the most slowly fluctuating maskers used in this study, the ‘acoustic glimpses’ afforded females in these conditions might have been too short to gain any benefit of listening in the dips. At present, we do not know how many consecutive calls are necessary to elicit phonotaxis in green treefrogs. As noted earlier, a

second, alternative explanation for the lack of masking release in the presence of slowly fluctuating maskers (at least in the 1.25 Hz and 2.5 Hz treatments) is that the maskers competed with the target signal, thereby abolishing or confounding any benefit of listening in the dips. The lack of attraction to the 0.625 Hz masker in experiment two (Table 3-1), however, would be inconsistent with this general interpretation for a lack of dip listening in the presence of slowly fluctuating maskers. Future studies should concentrate on elucidating which of these or other explanations account for the results obtained.

Signal recognition in natural settings

The noise generated in green treefrog choruses is characterized by slow fluctuation rates between about 2 Hz and 7 Hz, with a peak centered at 5 Hz; a second and relatively smaller (< -15 dB) peak was located around 300 Hz (Fig. 3-2d). The 300-Hz component is likely due to the amplitude modulations in the advertisement call. The predominance of low-frequency fluctuations in green treefrog chorus noise can be attributed to two factors. First, environmental perturbations on the transmission medium, such as turbulent air, can impose low-frequency fluctuations in ambient noise (Wiley and Richards 1978; Richards and Wiley 1980). Second, the production of repeated and temporally discontinuous calls creates low-frequency modulations that arise from call timing behavior of individuals in a chorus (Nelken et al. 1999). Note, however, that the peak around 5 Hz in the modulation spectrum does not match the rate of approximately 1.8 Hz at which individual males produce advertisement calls. Interestingly, the 5-Hz peak of the modulation spectrum more accurately represents vocal interactions among males in a chorus. Green treefrog males actively avoid call overlap with up to two of their nearest neighbors (Höbel and Gerhardt 2007; Jones et al. 2009; Höbel 2011). Given that individual green treefrog males produce advertisement calls about 150 ms long and repeated with a silent interval of approximately 410 ms (560 ms call period), the rate at which bouts of sound are produced by three interacting males that avoid call overlap is approximately 5.3 Hz. Therefore, the

peak around 5 Hz in the modulation spectra of green treefrog choruses may be due, in part, to behavioral vocal interactions among males.

Together with our analysis of chorus sounds, our results from experiment one suggest that natural level fluctuations in green treefrog chorus noise would have little effect on signal recognition by females for two main reasons. First, we tested fluctuation rates that span the range of predominant level fluctuations in green treefrog choruses and found no evidence of dip listening or modulation masking. Second, studies with human listeners indicate that (i) sinusoidal amplitude modulations provide greater dip-listening benefits than irregular amplitude modulations, and (ii) the amount of masking release and modulation masking increases as a function of increasing modulation depth (Bacon and Grantham 1989; Gustaffson and Arlinger 1994; Bacon et al 1998). Level fluctuations in chorus settings are not expected to be sinusoidal or to have modulation depths of 100% (see Fig. 3-2a-c). Therefore, our SAM maskers potentially offered a ‘best case scenario’ for observing dip listening and modulation masking compared with natural fluctuations in chorus noise. The fact that neither dip listening nor modulation masking were observed under potentially ideal conditions in the laboratory suggests these processes likely have small or even negligible effects on signal recognition in natural settings.

Conclusions

We found that female green treefrogs did not experience masking release or modulation masking in the presence of temporally fluctuating backgrounds, even under potentially ideal conditions for detecting these phenomena. The differences between relatively closely related species (i.e., *H. chrysoscelis* and *H. cinerea*) highlight the importance of comparative approaches in behavioral and neurophysiological studies to understand the underlying mechanisms for communication in noisy environments. Several previous studies have shown the existence of temporally selective neurons and temporal filters in the frog auditory system (Rose and Capranica 1984; Rose et al. 1985; Gooler and Feng 1992; Diekamp and Gerhardt 1995; Alder and Rose 1998; Rose and Gooler 2006).

Unfortunately, however, most of these studies have focused solely on how the signal’s

temporal structure is processed in species with pulsed advertisement calls (but see Klump et al. 2004). Identifying the neurophysiological mechanisms by which a greater diversity of species process amplitude-modulated sounds would help us understand the pattern of results observed in studies of signal recognition in fluctuating noise.

Chapter 4 – Signal Recognition by Green Treefrogs (*Hyla cinerea*) and Cope’s Gray Treefrogs (*Hyla chrysoscelis*) in Naturally Fluctuating Noise.

An ability to exploit level fluctuations of the background noise represents one way by which acoustically communicating animals may recognize signals in noisy social environments. We tested three hypotheses on signal recognition in fluctuating backgrounds: The ‘dip-listening’ hypothesis predicts facilitated signal recognition in the presence of fluctuating maskers, compared to non-fluctuating maskers. The ‘natural-soundscapes advantage’ hypothesis predicts facilitated signal recognition when level fluctuations resemble those of natural soundscapes, compared to artificial level fluctuations. In naturally fluctuating backgrounds, the ‘species-specific advantage’ hypothesis predicts facilitated signal recognition when level fluctuations resemble species-specific patterns of conspecific soundscapes. Using phonotaxis tests, we measured the ability of female green treefrogs (*Hyla cinerea*) and Cope’s gray treefrogs (*Hyla chrysoscelis*) to recognize male conspecific calls in the presence of noise maskers with (i) no level fluctuations, (ii) random fluctuations, or level fluctuations characteristic of (iii) conspecific choruses and (iv) heterospecific choruses. Support for the dip listening hypothesis was weak; only Cope’s gray treefrogs experienced dip listening in the presence of randomly fluctuating maskers. We found no evidence to support the natural-properties advantage hypothesis or the species-specific advantage hypothesis. These results corroborate previous findings showing that Cope’s gray treefrogs, but not green treefrogs, experience dip listening under some noise conditions. Our results also suggest level fluctuations in the soundscape of breeding choruses may offer few dip-listening opportunities in these two species. Together, our findings offer little support for the hypothesis that receivers are adapted to exploit level fluctuations of natural soundscapes to recognize communication signals.

Introduction

Animals often communicate acoustically in social aggregations that include several signalers and receivers (Schwartz and Freeberg 2008). In such environments, the background noise generated by the mixture of unattended signals impairs signal detection, recognition, and discrimination (Klump 1996; Brumm and Slabbekoorn 2005). How nonhuman animals overcome the problems associated with communication in noisy environments has received little attention (Klump 1996; Hulse 2002; Brumm and Slabbekoorn 2005; Bee and Micheyl 2008; Miller and Bee 2012). Among the potential mechanisms by which nonhuman animals solve such problems includes an ability to exploit level fluctuations of the background noise to improve signal perception in noisy environments (Klump 1996; Langemann and Klump 2005; Bee and Micheyl 2008).

In this comparative study, we examined the extent to which female green treefrogs (*Hyla cinerea*) and Cope's gray treefrogs (*Hyla chrysoscelis*) exploit level fluctuations of the chorus noise in recognizing male sexual signals. In both species, reproduction takes place in dense choruses in which males produce distinctive advertisement calls that mediate species recognition and female mate choice (Gerhardt and Huber 2002; reviewed in Schwartz and Bee in press). The noise generated in such choruses impairs the ability of females to recognize and discriminate among individual calls (reviewed in Vélez et al. in press). Importantly, level fluctuations of green treefrog and Cope's gray treefrog choruses differ and reveal species-specific patterns (Vélez and Bee 2010, 2011; Vélez et al. in review). Our objective was to test three hypotheses about signal recognition in fluctuating backgrounds: the 'dip listening' hypothesis, the 'natural-soundscapes advantage' hypothesis, and the 'species-specific advantage' hypothesis.

The dip-listening hypothesis states that the ability to recognize signals is facilitated in the presence of fluctuating maskers, compared to non-fluctuating maskers. Dip listening refers to our ability to catch brief 'acoustic glimpses' of speech and other target signals when the level of the background noise momentarily decreases (Gustafsson

& Arlinger 1994; Bacon et al. 1998; Cooke 2006; Füllgrabe et al. 2006; Vestegaard et al. 2011). Dip listening may represent a way by which nonhuman animals recognize communication signals in noisy social aggregations (Langemann and Klump 2005). Behavioral and neurophysiological studies with animals of diverse taxa show that simple tonal and narrowband noise signals can be detected at lower thresholds in the presence of maskers with *artificial* (e.g., sinusoidal or random) level fluctuations, compared with those in non-fluctuating maskers (Klump and Langemann 1995; Nelken et al. 1999; Langemann and Klump 2001, 2007; Nieder and Klump 2001; Hofer and Klump 2003; Bee et al. 2007; Jensen 2007; Branstetter and Finneran 2008; Fay 2011; Goense and Feng 2012). The ability to recognize communication signals in fluctuating backgrounds has received much less attention (but see Ronacher and Hoffmann 2003; Vélez and Bee 2010, 2011; Vélez et al. in review). Furthermore, the extent to which receivers exploit *natural* level fluctuations to improve signal recognition is still unclear. Here, we asked whether female frogs experience dip listening when recognizing communication signals in the presence of maskers with artificial and natural level fluctuations.

The natural-soundscapes advantage hypothesis states that signal recognition is facilitated when level fluctuations in the background noise resemble those of natural soundscapes, compared to artificial level fluctuations. Growing evidence suggests auditory systems are adapted to process physical properties of natural sounds (Rieke et al. 1995; Lewicki 2002; Woolley et al. 2005; Smith & Lewicki 2006). For example, auditory neurons transmit information more efficiently when the stimuli have properties of natural sounds, compared to stimuli with artificial properties (Rieke et al. 1995; Woolley et al. 2005). In species that communicate in social aggregations, the auditory system may be adapted to exploit physical properties of the natural soundscapes to improve signal recognition. Here, we asked whether the ability to recognize signals is better when level fluctuations in the background noise resemble those of natural soundscapes, compared to artificial level fluctuations.

The ‘species-specific advantage’ hypothesis states that signal recognition is facilitated when natural level fluctuations resemble species-specific patterns of conspecific soundscapes, compared to other naturally fluctuating maskers. Langemann and Klump (2005) suggested that animals may be adapted to exploit level fluctuations in the background noise that are typical of the natural soundscape during signal recognition. However, there is little support for this hypothesis for three main reasons. First, no study has examined the effects of natural level fluctuations on signal recognition. Second, the few available studies on recognition of communication signals in fluctuating backgrounds used artificial (sinusoidal) level fluctuations (Ronacher and Hoffmann 2003; Vélez and Bee 2010, 2011; Vélez et al. in review). Third, we still know little about level fluctuations in natural soundscapes. We recently showed that level fluctuations in the background noise of animal social aggregations differ across species and reveal species-specific patterns that result from the characteristics of the signals and signaling behaviors (Vélez and Bee 2010, 2011; Vélez et al. in review). Therefore, differences in the physical properties across soundscapes could potentially lead to species-specific adaptations that facilitate signal perception in the local soundscape. The extent to which an ability to recognize communication signals in fluctuating backgrounds is ‘tuned’ to species-specific fluctuation patterns of the soundscape is still an open question.

In this study, we used phonotaxis experiments to measure ‘signal recognition thresholds’ (Bee and Schwartz 2009) in the presence maskers that had (i) no level fluctuations, (ii) random fluctuations, or level fluctuations characteristic of (iii) conspecific choruses and (iv) heterospecific choruses. Our results corroborate previous studies showing that female Cope’s gray treefrogs, but not green treefrogs, listen in the dips of some fluctuating backgrounds (Vélez and Bee 2011; Vélez et al. in review), and reveal that level fluctuations of natural soundscapes offer little benefits for signal recognition.

General Methods

Subjects and study sites

Procedures used to collect, handle, and test animals were approved by the University of Minnesota's Institutional Animal Care and Use Committee (IACUC No. 0809A46721). We collected female green treefrogs (*H. cinerea*) in amplexus between 2200 and 0200 hours in April and May 2011 from artificial ponds at the Jasper State Fish Hatchery in Jasper, Texas (30° 57' 4.01" N, 94° 7' 39.53" W; Jasper County, Texas, U.S.A.). Female Cope's gray treefrogs (*H. chrysoscelis*) were collected in amplexus between 2100 and 0100 hours in May and June 2011 from wetlands at the Carver Park Reserve (44° 52' 49.29" N, 93° 43' 3.10" W; Carver County, Minnesota, U.S.A.) and the Crow Hassan Park Reserve (45° 11' 18.71" N, 93° 39' 9.05" W; Hennepin County, Minnesota, U.S.A.). Females were kept at approximately 2-4°C to delay oviposition until tested (usually within 24-48 hours) in the laboratory on the St. Paul campus of the University of Minnesota. Prior to testing, we placed females in an incubator for at least 45 minutes to allow their body temperature to reach 24 ± 1°C (green treefrogs) or 20 ± 1°C (Cope's gray treefrogs). After completing all tests, females were released at the pond from which they were collected (usually within three days).

Phonotaxis experiments

We used the same general protocol and testing equipment as described in recent studies of green treefrogs and Cope's gray treefrogs (Bee and Schwartz 2009; Vélez and Bee 2011; Vélez et al. in review). Readers are referred to those studies for additional details not reported here. Briefly, single-speaker phonotaxis tests (Gerhardt 1995) were conducted under infrared illumination (IR) inside a walk-in, temperature-controlled, hemi-anechoic sound chamber (details in Bee and Schwartz 2009). The temperature inside the chamber was set to 20 ± 1°C for tests with Cope's gray treefrogs and to 24 ± 1°C for tests with green treefrogs; these temperatures are close to the average temperatures at which each species breeds in local populations (Vélez and Bee unpublished data). We tested females in a circular test arena (2 m diameter) made of a 60-

cm high, acoustically transparent but visually opaque wall. The perimeter of the arena was divided into 24 15° arcs. An IR-sensitive video camera mounted from the ceiling of the chamber above the center of the test arena was used to score behavioral responses in real time by two observers viewing a monitor outside the chamber. The video feed was simultaneously encoded to digital video files.

Digital acoustic stimuli (11025-Hz sampling rate, 16-bit resolution) were generated using Matlab v7.6 (Mathworks, MA, USA) and broadcast through Orb Mod 1 speakers. The speaker used to broadcast target signals was placed on the floor, just outside the arena wall, centered in one of the 15° arcs, 1 m away from a subject release point at the center of the arena. The position of the target speaker was randomly varied around the arena's perimeter between tests of two to four subjects to eliminate any possibility of directional response bias. We have not observed such biases in our experimental set-up. Masking noises were broadcast from an overhead speaker mounted from the ceiling of the chamber, 190 cm above the central release point. The frequency response of our playback system was flat (± 2.5 dB) between 500 Hz and 5000 Hz, and the overhead speaker created uniform noise levels (± 2 dB) across the floor of the test arena. We calibrated sound levels by placing the microphone of a Brüel & Kjær Type 2250 sound level meter at the approximate position of a subject's head at the central release point.

We initiated a phonotaxis test by placing a subject in an acoustically transparent holding cage located at the central release point of the arena. Subjects could freely reorient inside the holding cage. Broadcast of a masking noise initiated after a 1min silent acclimation period and continued throughout the entire test. In tests with target signals in experiment 1, broadcast of the target signal initiated 30 s after the onset of the masker. Using a rope and pulley system operated from outside of the chamber, subjects were released after 15 s of signal presentation.

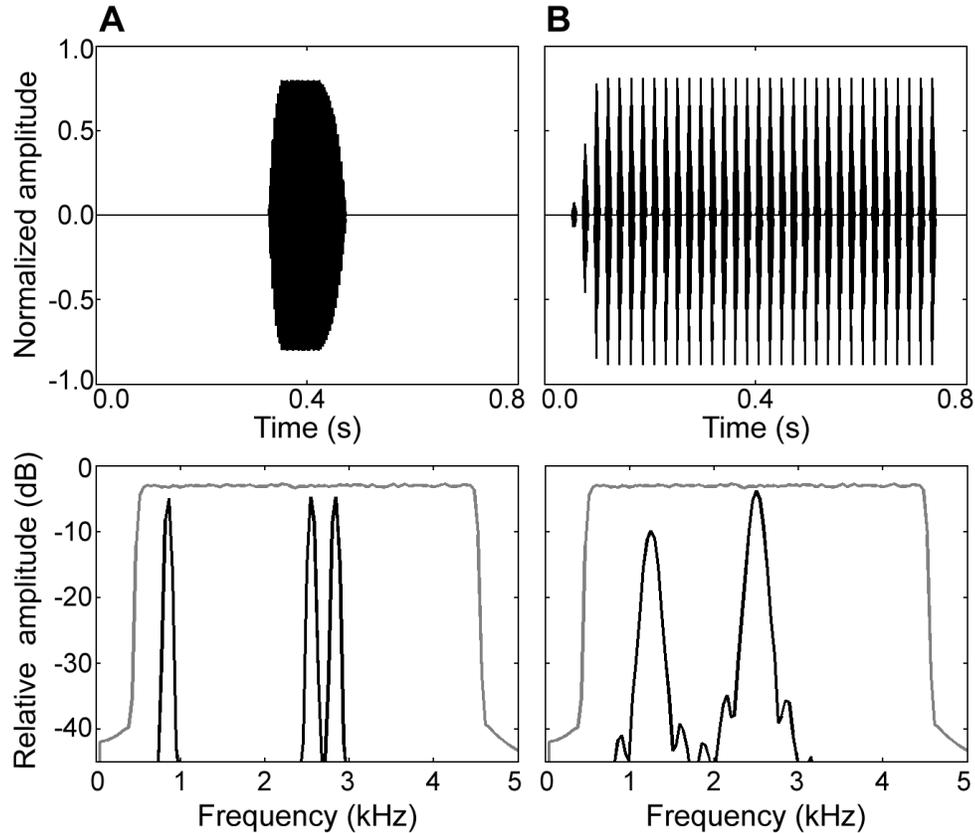


Figure 4-1. Waveform of 0.8-s segments (top panel) and frequency spectrum (bottom panel) of the synthetic call used as a target signal for tests with green treefrogs (A) and Cope's gray treefrogs (B). Frequency spectra were calculated with 256-point Hamming windows. For reference, we also show here the frequency spectrum of the maskers (gray line).

Experiment 1: Signal recognition in fluctuating backgrounds

Methods

Acoustic stimuli

Target signals. – For each species, the target signal was a synthetic advertisement call with values of temporal and spectral properties close to the averages of calls recorded in local populations of that species (Fig. 4-1; Bee, Höbel, Vélez unpublished data). For green treefrogs, the target signal was a 147-ms long synthetic call that repeated at a rate of approximately 107 calls/min (562 ms call period). The call comprised three phase-locked sinusoids with equal relative amplitudes and frequencies of 850 Hz, 2550 Hz, and

2833 Hz. The amplitude-envelope of the call was shaped with a 25-ms inverse exponential rise time and a 50-ms inverse exponential fall time. For Cope's gray treefrogs, the target signal comprised 32 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 1250 Hz (-9 dB) and 2500 Hz (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 first 50 ms of the call was shaped with a linear onset. The Cope's gray treefrog target signal repeated at a rate of 12 calls/min (5 s call period).

Masking noises. – We measured signal recognition thresholds (see below) in four different masking treatments. Across treatments, the masking noises had the same long-term frequency spectrum (500 Hz – 4500 Hz; Fig. 4-1) but different temporal envelopes (Fig. 4-2). The masking noise in the control treatment lacked level fluctuations; we refer to this masker as the 'non-fluctuating' treatment. The second treatment was a 'randomly fluctuating' noise treatment in which the masker had a random temporal envelope. The third and fourth treatments were 'naturally fluctuating' treatments in which the maskers had envelopes extracted from recordings of green treefrog choruses or Cope's gray treefrog choruses. We refer to these maskers as the conspecific treatment and the heterospecific treatment, depending on the species identity of the subject. In the conspecific treatment, green treefrogs experienced maskers with envelopes from green treefrog choruses, and Cope's gray treefrogs experienced maskers with envelopes from Cope's gray treefrog choruses. In the heterospecific treatment, each species experienced maskers with envelopes from choruses of the other species.

To create naturally fluctuating maskers, we used the Hilbert transform to extract the envelopes from 50 different 90-s recordings of green treefrog choruses and Cope's gray treefrog choruses (25 recordings per species). Full details about recording equipment and protocols are described in Vélez et al. (in review) for green treefrogs and in Vélez

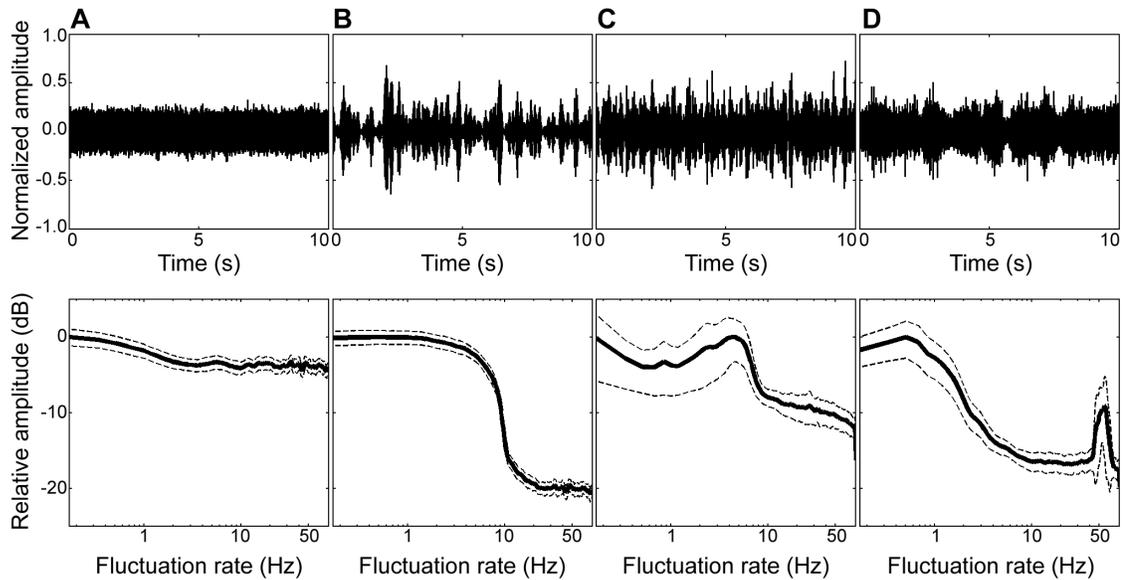


Figure 4-2. Waveform of 10-s segments (top panel) and mean (solid line) ± 1 standard deviation (dashed lines) modulation spectrum (bottom panel) of 90-s maskers with no level fluctuations (A), synthetically generated random fluctuations (B), or natural level fluctuations characteristic of green treefrog choruses (C) or Cope's gray treefrog choruses (D). Mean modulation spectra were calculated from the 25 90-s exemplars of each masker and are shown on a logarithmic x axis. Modulation spectra were generated 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. The modulation spectrum of each masker was calculated as the fast-Fourier transform (FFT) of the envelope (sampling rate = 11025 Hz, Hamming window size = 65536 points, overlap = 25%). We normalized the spectrum of each 90-s segment to the maximum value of the magnitude of the FFT and calculated the mean and standard deviation of the 25 modulation spectra of each type of masker. Mean and ± 1 standard deviation modulation spectra were transformed to a dB scale ($20 \cdot \log_{10}(\text{FFT magnitude})$) and smoothed for plotting purposes using a running average of 11 points.

and Bee (2010, 2011) for Cope's gray treefrogs. Briefly, recordings were obtained at nights and times of high calling activity using a Marantz PMD 670 recorder and an omnidirectional Senheizer M62 microphone. The microphone tip was placed 5 cm above ground or water level, between four and 15 m from the nearest calling male. We used only recordings in which other species were absent or calling at very low densities and away from the microphone. Chorus recordings of each species were made on different nights, at different ponds, or at different times and locations within a pond.

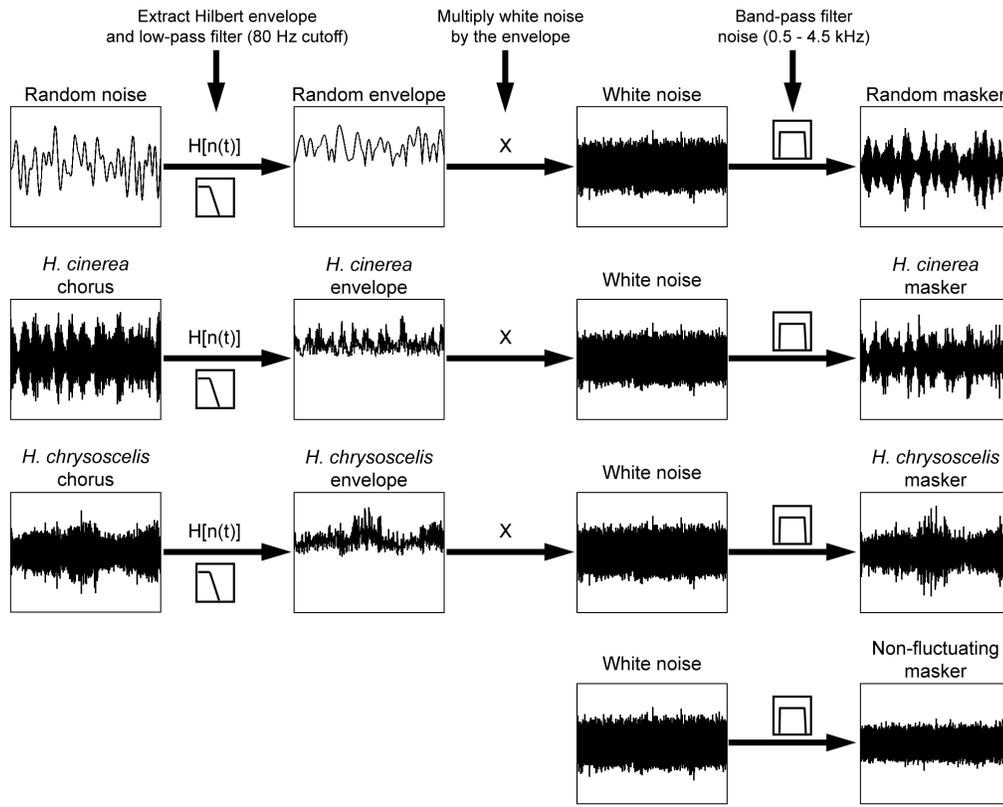


Figure 4-3. Schematic representation of the method used to generate the different maskers. This process was repeated 25 times, each one with a different white noise, a different chorus recording of each species, and a different low-pass (10 Hz cut-off) random noise. See text for additional details.

Figure 4-3 summarizes the method used to generate the maskers. First, we generated four copies of a 90-s white noise (11025-Hz sampling rate, 16-bit resolution). To create the randomly fluctuating masker, we generated a 90-s low-pass noise with a cut-off frequency of 10 Hz and extracted its Hilbert envelope. We then multiplied a copy of the white noise by this random amplitude-envelope. The two naturally fluctuating noises were generated by multiplying a copy of the white noise by the Hilbert envelope of a 90-s recording of a green treefrog or a Cope’s gray treefrog chorus after low-pass filtering the envelope with a cut-off frequency of 80 Hz. The three fluctuating (i.e., one

random and two natural) and the non-fluctuating copies of the white noise were then converted to the spectral domain with a fast Fourier transform (FFT; window size = 992250), band-pass filtered between 500 Hz and 4500 Hz by setting the Fourier coefficients outside this range to zero, and transformed back to the time domain with an inverse FFT. We created six-minute-long maskers of each type using four copies of the 90-s noises generated. To avoid pseudoreplication (Kroodsma 1989, 1990), we created 25 exemplars of each masker by repeating the process 25 times, each time with a different white noise, different chorus recording of each species, and different random envelope. Each exemplar was tested only once per treatment for each species. Masking treatment and exemplar were randomly assigned for each subject. In all treatments, maskers were broadcast from the overhead speaker with a long-term root-mean-square (RMS) amplitude of 76 dB sound pressure level (SPL re. 20 μ Pa; L_{Ceq}; 40 dB spectrum level) calibrated at the central release point of the arena.

Signal recognition thresholds

Using a between-subjects experimental design, we measured signal recognition thresholds for 25 subjects of each species in each of the four masker treatments described above: non-fluctuating, random, conspecific, and heterospecific (2 species x 4 treatments x 25 subjects per treatment; total N = 200). As in our previous studies, we operationally defined the signal recognition threshold as the minimum signal level necessary to elicit positive phonotaxis to a target signal (Bee and Schwartz 2009; Vélez and Bee 2011; Vélez et al. in review). Positive phonotaxis to a target signal was scored when the subject (i) first touched the wall of the test arena in the hemi-circle containing the target speaker, (ii) made contact with the arena wall inside the 15° bin in front of the target speaker within five minutes of being released, and (iii) remained in a 30° arc centered in front of the target speaker for 30 s. Therefore, our signal recognition thresholds differ from more traditional signal detection thresholds because, for positive phonotaxis to occur, the target signal has to be (i) detected by the subject, (ii) recognized as a conspecific advertisement call, and (iii) successfully localized.

We tested each subject in a series of 8-14 trials and gave them a timeout period of 5-15 min inside the incubator between trials. Each series always began and ended with a test of a reference trial. In the reference trial, we broadcast the target signal at 85 dB SPL (LCF, measured at 1 m) and in the absence of masking noises. This signal level corresponds to natural call amplitudes measured at 1 m for both species (Gerhardt 1975). We also performed a test of a reference trial after two consecutive test trials (described below) failed to elicit a response. We replaced subjects that failed to meet the response criteria in any reference trial. This procedure ensures that all subjects included in the analyses were highly motivated throughout all trials and validates test trials in which subjects fail to respond to the target signal (Bush et al. 2002).

After the first reference trial, we ran two control trials: a ‘sham’ trial and an ‘attraction’ trial. These two control trials are explained in detail in experiments 2 and 3, respectively. A second reference trial was conducted after the two control trials. After the second reference trial, we started a series of test trials designed to estimate a signal recognition threshold for each subject. In the first test trial, the target signal was broadcast at a level of 76 dB SPL (LCF). In all subsequent test trials, the level of the target signal was varied contingent upon the response of the subject in the previous test trial. If the subject met the response criteria in the first test trial, the level of the signal in the following test trial was lowered by 3 dB. If, on the other hand, the subject failed to respond, the level of the target signal was increased by 3 dB. We continued increasing or decreasing the level of the target signal contingent on the subject’s response in the previous test trial until there was a change in the subject’s behavior between two trials (i.e., from response to no-response, or vice versa). Following the trial in which the subject’s behavior changed, we ran a final test trial in which the direction of level change was reversed, and the step size of the change was reduced to 1.5 dB. For each subject, we calculated the signal recognition threshold as the average between the minimum signal level that elicited positive phonotaxis and the maximum signal level that failed to do so.

We calculated the signal recognition threshold in a linear scale and then converted it back to decibels.

Statistical analyses

We compared signal recognition thresholds using a 2 species \times 4 masking treatment factorial analysis of variance (ANOVA). Signal recognition thresholds met the assumptions of normality (Kolmogorov-Smirnov tests: all $P_s > 0.2$) and homogeneity of variance (Levene's test: $F_{7, 192} = 1.86$, $P = 0.08$) for parametric testing. We used planned contrasts to test three specific hypotheses. Based on the dip-listening hypothesis, we predicted lower signal recognition thresholds in all fluctuating treatments, compared to the non-fluctuating treatment. According to the natural-properties advantage hypothesis, we predicted lower signal recognition thresholds in naturally fluctuating treatments, compared to those in the random treatment. Based on the species-specific advantage hypothesis, we predicted lower signal recognition thresholds in the conspecific treatment, compared to those in the heterospecific treatment.

Results and discussion

Mean signal recognition thresholds varied within one dB across masking treatments. The main effect of masker treatment was not significant (ANOVA: $F_{3, 192} = 1.27$, $P = 0.287$, partial $\eta^2 = 0.02$; Fig. 4-4A). Both the main effect of species ($F_{1, 192} = 31.49$, $P < 0.001$, partial $\eta^2 = 0.14$; Fig. 4-4B) and the species \times masking treatment interaction ($F_{3, 192} = 5.65$, $P < 0.001$, partial $\eta^2 = 0.08$; Fig. 4-4C) were significant. Overall, signal recognition thresholds were 3.5 dB higher in green treefrogs than in Cope's gray treefrogs (Fig. 4-4B). Because we found significant effects of species and species \times masking treatment, but no effect of masking treatment, we compared signal recognition thresholds across masking treatments separately for each species. In green treefrogs, mean signal recognition thresholds varied within 2 dB (73.1 dB to 75.0 dB) across treatments and these differences were not significant ($F_{3, 96} = 0.91$, $P = 0.439$, partial $\eta^2 = 0.028$; Fig. 4-4C). Therefore, we did not perform planned comparisons testing specific hypotheses

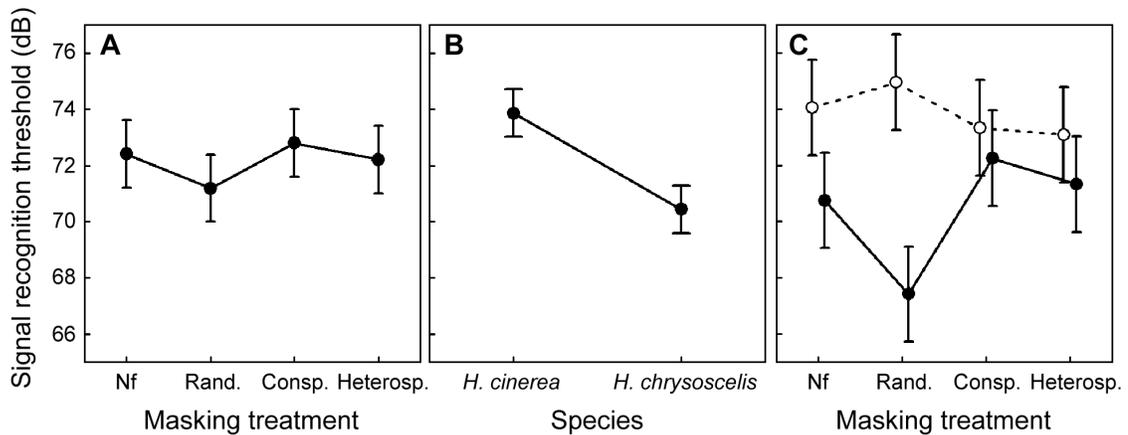


Figure 4-4. Mean ($\pm 95\%$ confidence intervals) signal recognition thresholds as a function of masking treatment (A), species (B), and masking treatment for each species (C). Nf = Non-fluctuating, Consp. = Conspecific, Heterosp. = Heterospecific, and Rand. = Random. In (C), open circles *H. cinerea* and filled circles represent *H. chrysoscelis*.

for this species. In contrast, the effect of masking treatment was significant in Cope's gray treefrogs ($F_{3, 96} = 6.19$, $P < 0.001$, partial $\eta^2 = 0.162$; Fig. 4-4C), with mean signal recognition thresholds varying over a 5 dB range between 67.4 dB and 72.4 dB across treatments.

Table 4-1 summarizes results of individual planned comparisons testing each hypothesis. Consistent with the dip listening hypothesis, signal recognition thresholds were lower in the random treatment than in the non-fluctuating treatment. Mean signal recognition thresholds in all other fluctuating treatments were not different from those in the non-fluctuating control. Together, these results suggest that female Cope's gray treefrogs can listen in the dips of fluctuating maskers, but this ability offers little benefits in naturally fluctuating backgrounds.

In contrast to the natural-soundscapes advantage hypothesis, mean signal recognition thresholds were actually higher in naturally fluctuating treatments than in the random treatment (Table 4-1). These results suggest that signal recognition in fluctuating backgrounds was better with artificial level fluctuations, compared to natural ones. Therefore, our results lead us to reject the natural-soundscapes advantage hypothesis.

Similarly, we found no evidence to support the species-specific advantage hypothesis; signal recognition thresholds did not differ between the conspecific and heterospecific treatments. Together, our results suggest that the ability of Cope’s gray treefrog females to listen in the dips of fluctuating maskers offer little benefit in naturally fluctuating backgrounds. Furthermore, our results suggest that dip listening in Cope’s gray treefrogs is not tuned to level fluctuations of the local soundscape.

Experiment 2: Behavior in the absence of target signals

Estimates of signal recognition thresholds could be affected if the maskers were not neutral stimuli and affected female movement patterns. For instance, maskers that sound more like a conspecific breeding chorus could elicit higher levels of searching behavior, such as higher over all movements around the arena, affecting the probability of recognizing low-level signals. To measure the extent to which female behavior in our testing apparatus was differently affected by the maskers, we conducted a ‘sham’ trial in which no target signal was broadcast.

Methods

During sham trials, each masker was broadcast continuously from the overhead speaker with a long-term RMS amplitude of 76 dB SPL (LCeq). Because no target signals were broadcast from speakers on the floor, there was no ‘correct response’ in the sham trial.

Table 4-1. Loadings and results of planned comparisons testing each specific hypothesis for Cope's gray treefrogs.

Hypothesis	Treatment loadings				Results		
	Non-Fluctuating	Random	Conspecific	Heterospecific	F _{1, 96}	P	partial η^2
Dip listening	-3	1	1	1	0.18	0.669	0.002
	-1	1	0	0	7.76	0.006	0.075
	-1	0	1	0	1.57	0.213	0.016
	-1	0	0	1	0.23	0.631	0.002
Natural-soundscapes advantage	0	-2	1	1	17.79	<0.001	0.156
	0	-1	1	0	16.31	<0.001	0.145
	0	-1	0	1	10.68	0.002	0.1
species-specific advantage	0	0	-1	1	0.59	0.443	0.006

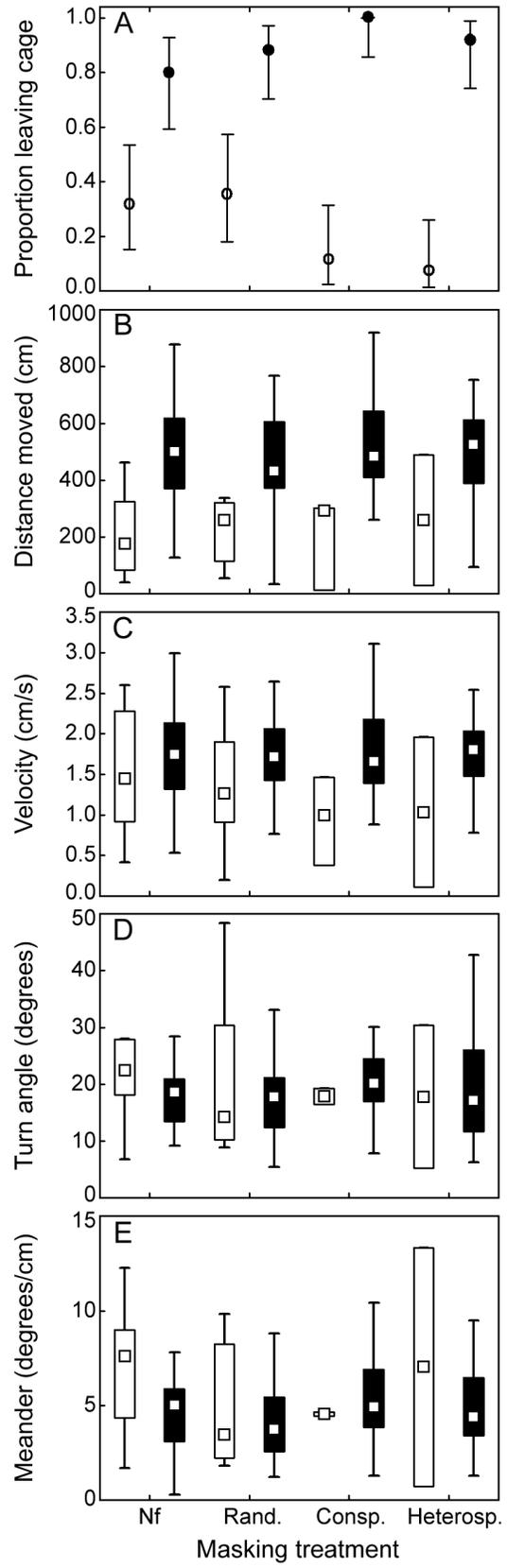


Figure 4-5. (A) Proportions ($\pm 95\%$ confidence intervals) of females leaving the holding cage during sham trials as a function of masking treatment. The total number of females tested in each treatment was 25. Median (point), inter-quartile range (box), and non-outlier range (whiskers), values for total distance moved (B), average velocity (C), average turn angle (D), and average meander (E). Nf = Non-fluctuating, Consp. = Consppecific, Heterosp. = Heterospecific, and Rand. = Random. Open bars and boxes represent *H. cinerea* and filled bars and boxes represent *H. chrysoscelis*.

Therefore, we did not apply any response criteria during sham trials and recorded subject behavior over 5 min of masker presentation after being released.

Statistical analyses

We analyzed movement patterns in two ways. First, we tested the hypothesis that the maskers did not affect subjects' tendency to move in the test arena. We scored whether or not subjects left the holding cage during the five minutes of masker presentation. We calculated the proportions of subjects leaving the holding cage in the four masking treatments and their corresponding 95% binomial confidence intervals. We reasoned that, if some maskers elicited more or less overall movements, then the proportions of subjects leaving the cage would differ across treatments.

Second, we tested the hypothesis that movement patterns in the test arena of the subjects that left the holding cage did not differ across masking treatments. We used the software Ethovision vXT (Noldus 2010) to measure the total distance moved (in cm), the average velocity of their movements (in cm/s), the average turn angles (in degrees), and the meander (in cm/s), which is a measurement of the amount of change in direction relative to the distance moved (see Bee and Riemersma 2008). The last two measurements were included because they are potentially related to sound localization behaviors (Rheinlaender and Klump 1988). We used non-parametric tests to investigate the effect of masker type (Kruskal-Wallis tests) and species (Mann-Whitney U tests) on movement patterns.

Results and discussion

Figure 4-5 summarizes results from this experiment. The number of subjects leaving the holding cage across treatments varied between two and nine in green treefrogs, and between 17 and 25 in Cope's gray treefrogs. For each species, the 95% binomial confidence intervals of the corresponding proportions show considerable overlap. These results suggest that subject tendencies to move in our testing apparatus were not differently affected by the maskers. The 95% confidence intervals do not overlap between species in any treatment, suggesting Cope's gray treefrogs left the holding cage more often than green treefrogs.

Analyses of movement patterns of the subjects that left the cage during sham trials revealed no differences across masking treatments for either species in total distance moved (Cope's gray treefrogs: Kruskal-Wallis $H_{3, 90} = 0.96$, $P = 0.81$; green treefrogs: $H_{3, 22} = 0.35$, $P_S = 0.95$), mean velocity (Cope's gray treefrogs: $H_{3, 90} = 0.41$, $P = 0.99$; green treefrogs: $H_{3, 22} = 1.77$, $P_S = 0.62$), average turn angles (Cope's gray treefrogs: $H_{3, 89} = 1.87$, $P = 0.60$; green treefrogs: $H_{3, 19} = 0.46$, $P_S = 0.93$), and mean meander ($H_{3, 89} = 2.76$, $P = 0.43$; green treefrogs: $H_{3, 19} = 0.74$, $P_S = 0.86$). Because there were no differences across treatments and sample sizes are very small for green treefrogs, we pooled data across treatments to investigate between-species differences in each of the four variables measured. We found no differences between species in average turn angles (Mann-Whitney U Test: $U = 822.0$, $P = 0.85$) or mean meander ($U = 758.0$, $P = 0.48$). However, the total distance moved ($U = 303.0$, $P < 0.001$) and the mean velocity ($U = 647.0$, $P = 0.0$) were significantly higher in Cope's gray treefrogs (medians = 482 cm; 1.7 cm/s) than in green treefrogs (medians = 248 cm; 1.3 cm/s).

Together, results from these two analyses indicate that the maskers by themselves did not differentially affect subject movement patterns (Fig. 4-5). Hence, the effect of the maskers on subject behavior had little potential to confound our estimates of signal recognition thresholds within each species. The between-species differences on the proportion of subjects leaving the holding cage and on the total distance moved can

account for the 3.5 dB difference in overall signal recognition thresholds between green treefrogs and Cope's gray treefrogs (Fig. 4-4B). At low signal levels, females of both species may not recognize the target signal while in the holding cage. However, because Cope's gray treefrog females tend to wander around the test arena, they have a greater probability to reach places where the signal-to-noise ratio is higher than in the holding cage. Therefore, Cope's gray treefrogs also have a higher probability to recognize the signal at lower signal levels in our experimental setup.

Experiment 3: Maskers as potential signals

Another potential confound in estimating signal recognition thresholds stems from the possibility that maskers could act like salient and biologically relevant signals that 'compete' with the target signals. For instance, males and females of some frog species can use the sound of conspecific choruses as acoustic beacons to locate breeding aggregations (Gerhardt and Klump 1988; Bee 2007; Swanson et al. 2007; Christie et al. 2010). The aim of the attraction trials was to evaluate the possibility that our maskers competed with the target signal. Specifically, we tested the null hypothesis that the maskers used in experiment 1 were not attractive to females.

Methods

In attraction trials, we broadcast each noise alone from a speaker on the floor, just outside the wall of the arena, centered in the 15° bin that was 180° away from the speaker used to broadcast the target signal in experiment 1. Noises were broadcast continuously and their long-term RMS amplitude was calibrated to 76 dB SPL (LCeq). After release, we measured the angle at which subjects first touched the wall of the arena, in 15° bins. As in previous similar studies (Swanson et al. 2007; Vélez and Bee 2010, 2011; Vélez et al. in review), we ended the attraction trials as soon as subjects touched the wall anywhere around the test arena or after five minutes of stimulus presentation.

Statistical analyses

We investigated subjects' behavior in attraction trials using a categorical analysis. We first grouped the angles at which subjects reached the arena wall in eight 45° bins. In response to attractive sounds, subjects typically reach the arena wall in the bin in front of the speaker or one bin to either side (Vélez and Bee 2010). We then grouped all responses into two categories. The first category included responses in which subjects first touched the arena wall inside the 45° bin in front of the target speaker. The second category included all other angles and trials in which subjects failed to touch the arena wall. The inclusion of the latter trials is important because these data are biologically relevant and cannot be included in analyses using circular statistics. We used a Fisher's exact test to compare categorical responses in all treatments to those in the absence of noise. Our reasoning was that if maskers were attractive, then we would expect more females to touch the wall inside the 45° bin in front of the speaker than those expected by chance. From two previous studies (Vélez and Bee 2010, Vélez et al. in review), we reanalyzed data and calculated the proportion of subjects first touching the arena wall inside the 45° bin in front of a designated, but silent, target speaker. We found that these proportions are 0.083 (two out of 20 subjects) and 0 (0 out of 24 subjects) for Cope's gray treefrogs and green treefrogs, respectively (Vélez and Bee 2010, Vélez et al. in review).

Results and discussion

Results for this experiment are summarized in Figure 4-6. The proportion of female green treefrogs reaching the arena wall inside the 45° bin in front of the target speaker was not significantly different from the null expectation in the non-fluctuating, conspecific, and heterospecific treatments (two-tailed Fisher's exact test, all P s = 1.0). The proportion of subjects approaching the target speaker was significantly higher than the expected by chance in the random treatment ($P < 0.01$). In Cope's gray treefrogs, the proportion of subjects reaching the target speaker was not significantly different from that in the absence of sound in the non-fluctuating ($P = 0.269$), random ($P = 0.269$), and heterospecific ($P = 0.437$) treatments. In the conspecific treatment, the proportion of

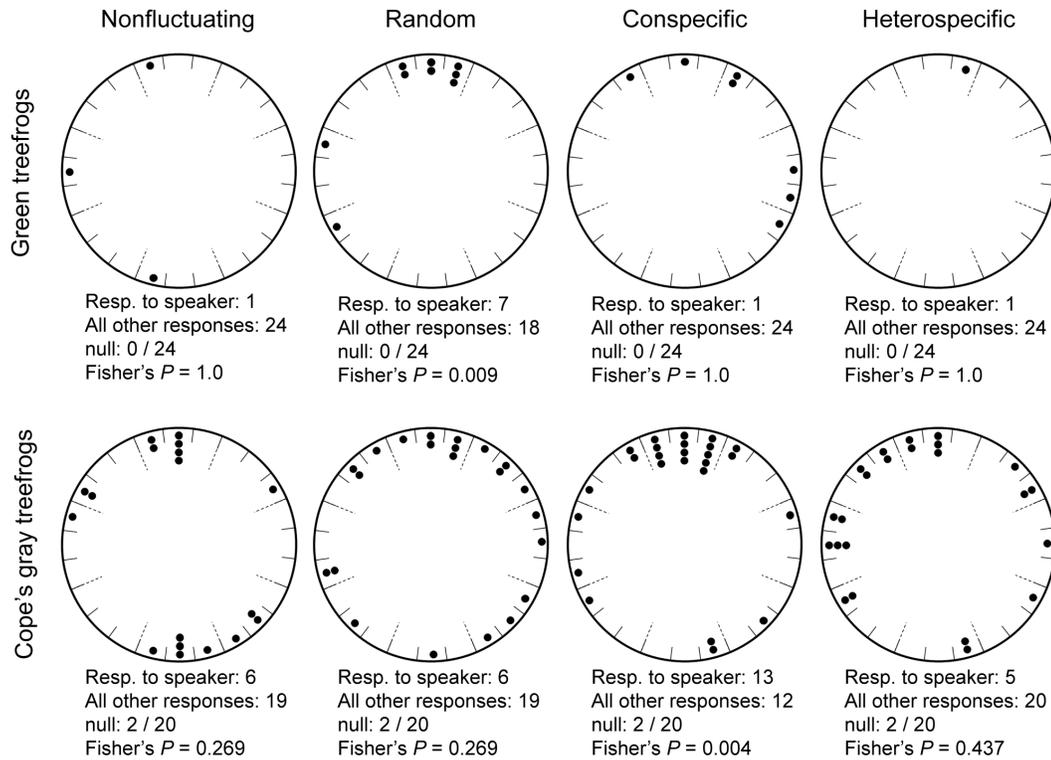


Figure 4-6. Angles at which subjects reached the arena wall in attraction trials for green treefrogs (top row) and Cope's gray treefrogs (bottom row) in the nonfluctuating (far left), random (middle left), conspecific (middle right) and heterospecific (far right) treatments. Points depict number of subjects reaching at each 15° bin (solid, short lines) relative to the position of the target speaker (top of each circle). Dotted lines represent 45° bins. See text for additional details.

subjects that reached the target speaker was significantly higher than the null expectation ($P = 0.004$).

Cope's gray treefrogs were attracted to noises with conspecific level fluctuations and green treefrogs were attracted to those with random level fluctuations. Consequently, these maskers had the potential to 'compete' with the target signal during test trials of our signal recognition experiment. For instance, at low signal-to-noise ratios, females may have recognized the target signal but not responded to it because they were more attracted to the maskers than the signals. However, the degree of 'competition' for subjects'

responses between these maskers and the target signals was likely small. The proportion of subjects that respond to these noises is higher than expected by chance (Fig. 4-6), but it is lower than the proportion of subjects that respond to conspecific calls broadcast at lower levels and in the absence of noise. In green treefrogs, 21 out of 24 females reached the wall inside the 45° bin in front of the speaker at signal levels between 42 dB and 75 dB (mean signal level = 55.3 dB; Vélez et al. in review). Out of 20 Cope's gray treefrog females tested in the absence of noise and at a signal level of 61 dB SPL, 19 of them met or approached the target speaker (Vélez and Bee 2010). Hence, the attractive maskers in the present study should be regarded as marginally attractive at best.

Together, our results from experiments two and three suggest that the potential confounding effects of the attractiveness of the conspecific and random maskers in Cope's gray treefrogs and green treefrogs, respectively, were probably small. More importantly, we can exclude the potentially confounding treatments and still test the specific hypotheses of this study. With green treefrogs, we can test both the species-specific advantage hypothesis and the dip listening hypothesis after excluding the random treatment. With Cope's gray treefrogs we can test the natural-properties advantage hypothesis and the dip listening hypothesis after excluding the conspecific treatment. The results of these analyses parallel the ones reported above (Appendix 4). Therefore, our study offers strong evidence to reject both the natural-soundscapes advantage and the species-specific advantage hypotheses. Our study also offers strong evidence for the operation of dip-listening mechanisms in Cope's gray treefrogs under some noise conditions.

General Discussion

Our main objective was to test three hypotheses about signal recognition in the presence of fluctuating backgrounds. The dip-listening hypothesis predicts lower signal recognition thresholds in the presence of fluctuating maskers, compared to those in a non-fluctuating control. We only found evidence for dip listening in the random treatment with Cope's gray treefrogs. These results suggest that female Cope's gray treefrogs, but

not green treefrogs, listen in the dips of some fluctuating maskers. The natural-soundscapes advantage hypothesis predicts lower thresholds in naturally fluctuating treatments, compared to those in the random treatment. The evidence led us to reject this hypothesis. While these results are robust, it is important to consider that our random maskers represent only a subset of all the possible artificially fluctuating maskers. Therefore, it would be inappropriate to conclude that the auditory system performs equally well, or better, in artificial soundscapes than in natural soundscapes. Based on the species-specific advantage hypothesis, we predicted lower thresholds in the conspecific treatment than in the heterospecific treatment. We found no evidence to support the species-specific advantage hypothesis. Together, results from this study suggest that female green treefrogs and Cope's gray treefrogs may not be adapted to exploit level fluctuations of the chorus noise to recognize male advertisement calls. While Cope's gray treefrog females possess an ability to listen in the dips of fluctuating backgrounds, this ability appears not to be tuned to the physical properties of the natural soundscape.

Our results parallel, to some extent, those from previous studies on signal recognition in fluctuating noise in both species (Vélez and Bee 2011; Vélez et al. in review). In those studies, we investigated the effect of level fluctuation rate on signal recognition using chorus-shaped maskers (i.e., maskers with the long-term frequency spectrum of the sounds of the chorus) that either lacked level fluctuations or were sinusoidally amplitude modulated (SAM) at rates ranging from 0.625 Hz to 80 Hz, in octave steps. Female green treefrogs did not experience dip listening in the presence of any SAM masker (Vélez et al. in review). In Cope's gray treefrogs we found (i) dip listening when the fluctuation rates of the SAM maskers were equal to or slower than 2.5 Hz, (ii) no effect when masker fluctuation rates were between 5 Hz and 20 Hz, and (iii) additional masking when fluctuation rates were 40 Hz and 80 Hz (Vélez and Bee 2011). The additional masking was explained by a process known as 'modulation masking', which describes increased difficulty to detect fluctuating signals in the presence of fluctuating maskers (Bacon and Grantham 1989).

Studies with human listeners indicate that (i) sinusoidal and random amplitude modulations provide greater dip-listening benefits than level fluctuations of multi-talker environments, and (ii) the amount of masking release increases as a function of increasing modulation depth (Gustafsson and Arlinger 1994; Bacon et al. 1998; Jin and Nelson 2006). In Cope's gray treefrogs, dip listening in the random treatment is likely due to the predominance of slow modulation rates and the 'deep' modulation depth of the maskers (Fig. 4-2). The striking threshold differences between the random and the naturally fluctuating treatments in Cope's gray treefrogs stress the importance of studying the physical properties of natural soundscapes and their effects on signal recognition before generalizing from studies using solely artificial stimuli.

Signal recognition in natural settings

Under the conditions tested in this study, we did not find evidence for dip listening in naturally fluctuating backgrounds. The shape and the depth of level fluctuations in our maskers could account for this result. Level fluctuations of the sound of the chorus are not expected to be sinusoidal or to have high values of modulation depth. Randomly fluctuating maskers (this study) and SAM maskers with 100% modulation depth (Vélez and Bee 2011) potentially offered better dip-listening opportunities compared with fluctuations of multi-signaler soundscapes.

The rates of level fluctuation in our maskers could also explain why Cope's gray treefrogs did not experience dip listening in naturally fluctuating treatments. The sounds of green treefrog choruses fluctuate at slow rates, with a predominant peak around 5 Hz (Vélez and Bee 2010; Vélez et al. in review; Fig. 4-2B). The inability of female frogs to listen in the dips of maskers fluctuating at 5 Hz (Vélez and Bee 2011) could potentially explain the lack of dip listening in the heterospecific treatment. Sound levels in Cope's gray treefrog choruses fluctuate both at rates below 3 Hz and at rates between 40 Hz and 60 Hz (Vélez and Bee 2010; Fig. 4-2C). In the conspecific treatment, slow rates of level fluctuation in the maskers could lead to dip listening while fluctuation rates between 40 Hz and 60 Hz could lead to modulation masking (Vélez and Bee 2011). Therefore, the

lack of dip listening in the conspecific treatment could be due to an interaction between dip listening and modulation masking. An interaction of this type has been invoked to explain some observed patterns of speech recognition in the presence of fluctuating backgrounds (Kwon and Turner 2001).

It is important to note that we generated conspecific maskers using chorus recordings in which conspecific males were the only or the most common vocally active individuals. While this is the case for part of the breeding season of both species at our study sites, these species often form mixed-species choruses. At our study sites, green treefrogs can be found vocalizing in sympatry with eastern gray treefrogs (*H. versicolor*), squirrel treefrogs (*H. squirrela*), fowler's toads (*A. fowleri*) and Gulf Coast toads (*Incilius nebulifer*). Cope's gray treefrogs are sometimes active with other anuran species such as eastern gray treefrogs (*H. versicolor*), American toads (*Anaxyrus americanus*) and Boreal chorus frogs (*Pseudacris maculata*). The advertisement calls produced by all of these heterospecific species are pulsed and have frequency spectra that overlap with that of the background noise of green treefrog and Cope's gray treefrog choruses (Rose and Capranica 1984; Gerhardt and Doherty 1988; Sullivan and Wagner 1988; Taylor et al. 2007; Bee et al. 2010). Furthermore, Copes' gray treefrogs and green treefrogs may be found in sympatry in parts of their natural distribution (Conant and Collins 1998). Studies on call interference suggest that heterospecific call overlap can disrupt the perception of important species-recognition cues, affecting a female's ability to recognize appropriate mates (Marshall et al. 2006). Due to partial spectral and temporal overlap, heterospecific signals in mixed-species choruses are therefore likely to have a negative effect on call recognition.

In natural settings, recognition of advertisement calls by female green treefrogs and Cope's gray treefrogs may be influenced by spatial, spectral, and temporal relationships between the target signals and competing sounds. As is the case for receivers in other taxa (reviewed in Bee & Micheyl, 2008), female green treefrogs and Cope's gray treefrogs experience a release from auditory masking when the sources of

signals and noise are spatially separated (Bee, 2007b, 2008a; Schwartz & Gerhardt, 1989; Nityananda & Bee 2012). Cope's gray treefrog females also exploit spectral differences as a cue to perceptually separate signals from distracting sounds (Nityananda & Bee, 2011). Therefore, synergistic interactions between spatial, spectral, and, to a lesser extent, temporal cues might contribute to ameliorating the problem of signal recognition in noisy natural settings.

Between-species differences in dip-listening capabilities

Why Cope's gray treefrogs and not green treefrogs listen in the dips of some fluctuating backgrounds is an open question. The difference between green treefrogs and Cope's gray treefrogs in call duration could account for the differences in dip-listening capabilities. In humans, the ability to detect signals in fluctuating maskers increases with signal duration (Schooneveldt and Moore 1989, but see Klump and Langemann 1995 for a negative result in birds). Another possible explanation for between-species differences in dip-listening capabilities concerns differences in temporal resolution of the auditory system. In humans, hearing-impaired listeners experience little, if any, masking release by means of dip listening (Eisenberg et al. 1995; Bacon et al. 1998; Jin and Nelson 2006). The differences in dip-listening capabilities between normal-hearing and hearing-impaired listeners are due, in part, to differences in temporal resolution of the auditory system (Jin and Nelson 2006). The slow recovery from forward masking and the higher forward-masking thresholds associated with hearing impairment causes the dips in fluctuating backgrounds to be perceptually 'filled in' with noise, providing few dip-listening opportunities (Festen and Plomp 1990). Comparative physiological and behavioral studies on temporal processing in the context of signal recognition in noise are therefore necessary to elucidate the underlying mechanisms for communication in noise. Unfortunately, most physiological studies on temporal processing in anurans have focused on how amplitude modulation is encoded along the auditory pathway and how species-specific features of the signals are processed (reviewed in Feng et al. 1990; Klump et al. 2004; Rose and Gooler 2006).

Conclusions

Our results on signal recognition in the presence of maskers that fluctuate randomly (this study) or sinusoidally (Vélez and Bee 2011; Vélez et al. in review) indicate that Cope's gray treefrogs, but not green treefrogs, exploit dips in temporally structured backgrounds to recognize communication signals. These between-species differences suggest that generalizations about dip-listening capabilities across taxa, even between closely related species, have to be made with caution. Under the conditions tested in this study, the ability of Cope's gray treefrog females to listen in the dips offered no benefit in the presence of naturally fluctuating backgrounds. Furthermore, we show here that receivers do not experience advantages in signal recognition in the presence of fluctuating maskers with species-specific level fluctuations of conspecific soundscapes. These results provide little support for the hypothesis that receivers are adapted to exploit level fluctuations typical of the natural soundscape to recognize communication signals (Langemann & Klump 2005). Together, our results strongly suggest that, in order to understand the evolution of auditory processing mechanisms for communication in noisy environments, it is of great importance to (i) take comparative approaches in behavioral and physiological studies, (ii) investigate the properties of natural soundscapes and how they affect signal recognition, and (iii) use both natural and artificial stimuli to better understand the capabilities and limitations of the auditory system.

References

- Alder TB, Rose GJ (1998) Long-term temporal integration in the anuran auditory system. *Nature Neuroscience* 1:519-523.
- Arch VS, Narins PM (2008) 'Silent' signals: selective forces acting on ultrasonic communication systems in terrestrial vertebrates. *Animal Behaviour* 76:1423-1428.
- Aubin T, Jouventin A (1998) Cocktail-party effect in king penguins. *Proceedings of the Royal Society of London, Series B*, 265, 1665-1673.
- Bacon SP, Grantham DW (1989) Modulation masking: Effects of modulation frequency, depth, and phase. *The Journal of the Acoustical Society of America* 85:2575-2580.
- Bacon SP, Grantham DW (1992) Fringe effects in modulation masking. *The Journal of the Acoustical Society of America* 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. *Journal of Speech Language and Hearing Research* 41:549-563.
- Beckers OM, Schul J (2004) Phonotaxis in *Hyla versicolor* (Anura, Hylidae): the effect of absolute call amplitude. *Journal of Comparative Physiology A* 190:869-876
- Bee MA (2007a) Sound source segregation in grey treefrogs: spatial release from masking by the sound of a chorus. *Animal Behaviour*, 74, 549-558.
- Bee MA (2007b) Selective phonotaxis by male woodfrogs (*Rana sylvatica*) to the sound of a chorus. *Behavioral Ecology and Sociobiology*, 61:955-966.
- Bee MA (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 MA (2008b) Parallel female preferences for call duration in a diploid ancestor of an allotetraploid treefrog. *Animal Behaviour*, 76, 845-853.
- Bee MA, Buschermöhle M, Klump GM (2007) Detecting modulated signals in modulated noise: (II) neural thresholds in the songbird forebrain. *European Journal of Neuroscience*, 26: 1979-1994.
- Bee MA, Cook JM, Love EK, O'Bryan LR, Pettitt BA, Schrode K, and Vélez A (2010) Assessing acoustic signal variability and the potential for sexual selection and social recognition in boreal chorus frogs (*Pseudacris maculata*). *Ethology*, 116: 564-576.
- Bee MA, 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 MA, Riemersma KK (2008) Does common spatial origin promote the auditory grouping of temporally separated signal elements in grey treefrogs? *Animal Behaviour* 76:831-843.
- Bee MA, Schwartz JJ (2009) Behavioral measures of signal recognition thresholds in frogs in the presence and absence of chorus-shaped noise. *The Journal of the Acoustical Society of America* 126:2788-2801.
- Bee MA, Vélez A, Forester JD (in press) Sound level discrimination by Cope's gray treefrog (*Hyla chrysoscelis*) in the presence and absence of chorus-shaped noise. *Journal of the Acoustical Society of America*.
- Branstetter BK, Finneran JJ (2008) Comodulation masking release in bottlenose dolphins (*Tursiops truncatus*). *The Journal of the Acoustical Society of America* 124:625-633.
- 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. *Journal of the Acoustical Society of America* 92:3132-3139
- Brumm H, Slabbekoorn H (2005) Acoustic communication in noise. *Advances in the Study of Behavior* 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. *Animal Behaviour* 63:7-14.
- Buus S (1985) Release from masking caused by envelope fluctuations. *Journal of the Acoustical Society of America* 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 Press, New York, pp 701-730
- Cherry EC (1953) Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America* 25:975-979
- Christie K, Schul J, Feng AS (2010) Phonotaxis to male's calls embedded within a chorus by female gray treefrogs, *Hyla versicolor*. *Journal of Comparative Physiology A* 196:569-579.
- Cohen J (1988) *Statistical power analysis for the behavioral sciences*. 2nd ed. Erlbaum, Hillsdale, New Jersey.
- Conant R, Collins JT (1998) *A field guide to reptiles and amphibians: eastern and central North America*. Houghton Mifflin, Boston, MA.
- Cooke M (2006) A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America* 119:1562-1573.

- Diekamp B, Gerhardt HC (1995) Selective phonotaxis to advertisement calls in the gray treefrog *Hyla versicolor*: behavioral experiments and neurophysiological correlates. *Journal of Comparative Physiology A* 177:173-190.
- Eisenberg LS, Dirks D, Bell TS (1995) Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing. *Journal of Speech and Hearing Research*, 38: 222–233.
- Fay RR (2011) Signal-to-noise ratio for source determination and for a comodulated masker in goldfish, *Carassius auratus*. *The Journal of the Acoustical Society of America* 129: 3367-3372.
- Feng AS, Hall JC, Gooler DM (1990) Neural basis of sound pattern recognition in anurans. *Progress in Neurobiology*, 34:313–329.
- Festen JM, Plomp R (1990) Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America* 88: 1725-1736.
- Füllgrabe C, Berthommier F, Lorenzi C (2006) Masking release for consonant features in temporally fluctuating background noise. *Hearing Research* 211:74-84.
- Gerhardt HC (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 HC (1981) Mating call recognition in the green treefrog (*Hyla cinerea*): Importance of two frequency bands as a function of sound pressure level. *Journal of Comparative Physiology A* 144:9-16.
- Gerhardt HC (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), Birkhäuser Verlag, Basel. pp. 209-220.

- Gerhardt HC (2001) Acoustic communication in two groups of closely related treefrogs. *Advances in the Study of Behavior* 30: 99-167.
- Gerhardt HC, Doherty JA (1988) Acoustic communication in the gray treefrog, *Hyla versicolor*: evolutionary and neurobiological implications. *Journal of Comparative Physiology A* 162: 261-278.
- 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. *Animal Behaviour* 36:1247-1249.
- Gerhardt HC, Klump GM (1988b) Phonotactic responses and selectivity of barking treefrogs (*Hyla gratiosa*) to chorus sounds. *Journal of Comparative Physiology 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 H.C, Tanner SD, Corrigan CM, Walton HC (2000) Female preference functions based on call duration in the gray tree frog (*Hyla versicolor*). *Behavioral Ecology* 11: 663-669.
- Goense JBM, Feng AS (2012) Effects of noise bandwidth and amplitude modulation on masking in frog auditory midbrain neurons. *PLoS ONE* 7: e31589.
- Gooler DM, Feng AS (1992) Temporal coding in the frog auditory midbrain: the influence of duration and rise-fall time on the processing of complex amplitude-modulated stimuli. *Journal of Neurophysiology* 67:1-22.

- Grafe TU, Dobler S, Linsenmair KE (2002) Frogs flee from the sound of fire. Proceedings of the Royal Society B 269: 999-1003.
- Gustafsson HA, Arlinger SD (1994) Masking of speech by amplitude-modulated noise. The Journal of the Acoustical Society of America 95:518-529.
- Hawley ML, Litovski RY, Culling JF (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.
- von Helversen D, von Helversen O. (1997) Recognition of sex in the acoustic communication of the grasshopper *Chorthippus biguttulus* (Orthoptera, Acrididae). Journal of Comparative Physiology A 180: 373-386.
- Höbel G (2011) Variation in signal timing behavior: Implications for male attractiveness and sexual selection. Behavioral Ecology and Sociobiology 65: 1283-1294.
- Höbel G, Gerhardt HC (2007) Sources of selection on signal timing in a treefrog. Ethology 113: 973-982.
- Hofer SB, Klump GM (2003) Within- and across-channel processing in auditory masking: A physiological study in the songbird forebrain. The Journal of Neuroscience 23: 5732-5739.
- Hulse SH (2002) Auditory scene analysis in animal communication. Advances in the Study of Behavior 31:163-200.
- Jensen KK (2007) Comodulation detection differences in the hooded crow (*Corvus corone cornix*), with direct comparison to human subjects. The Journal of the Acoustical Society of America 121:1783-1789.
- Jin SH, Nelson PB (2006) Speech perception in gated noise: The effects of temporal resolution. The Journal of the Acoustical Society of America, 119: 3097-3108.

- Jones DL, Jones RL, Ratnam R (2009) Dynamical interactions in a green treefrog chorus. *The Journal of the Acoustical Society of America* 126:2270-2270.
- Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. *Physiological Reviews* 84:541-577.
- Klump GM (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, NY: Cornell University Press.
- Klump GM, Benedix Jr JH, Gerhardt HC, Narins PM (2004) AM representation in green treefrog auditory nerve fibers: neuroethological implications for pattern recognition and sound localization. *Journal of Comparative Physiology A* 190:1011-1021.
- Klump GM, Dooling RJ, Fay RR, Stebbins WC (1995) *Methods in Comparative Psychoacoustics*. In. Birkhäuser Verlag, Basel
- Klump GM, Langemann U (1995) Comodulation masking release in a songbird. *Hearing Research*. 87:157-164.
- Kroodsma DE (1989) Suggested experimental designs for song playbacks. *Animal Behaviour* 37: 600-609.
- Kroodsma DE (1990) Using appropriate experimental designs for intended hypotheses in ‘song’ playbacks, with examples for testing effects of song repertoire sizes. *Animal Behaviour* 40: 1138-1150.
- Kwon BJ, Turner CW (2001) Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference? *The Journal of the Acoustical Society of America* 110:1130-1140.

- Langemann U, Klump GM (2001) Signal detection in amplitude-modulated maskers. I. Behavioural auditory thresholds in a songbird. *European Journal of Neuroscience* 13:1025-1032.
- Langemann U, Klump GM (2005) Perception and acoustic communication networks. In: McGregor PK (ed) *Animal Communication Networks*. Cambridge University Press, Cambridge, pp 451-480.
- Langemann U, Klump GM (2007) Detecting modulated signals in modulated noise: (i) behavioural auditory thresholds in a songbird. *European Journal of Neuroscience* 26: 1969-1978.
- Lewicki MS (2002) Efficient coding of natural sounds. *Nature Neuroscience*, 5: 356-363.
- Lombard RE, Bolt JR (1979) Evolution of the tetrapod ear: an analysis and reinterpretation. *Biological Journal of the Linnean Society* 11: 19-76.
- Love EK, Bee MA (2010) An experimental test of noise-dependent voice amplitude regulation in Cope's grey treefrog (*Hyla chrysoscelis*). *Animal Behaviour*, 80, 509-515.
- McDermott JH (2009) The cocktail party problem. *Current Biology*, 19, R1024-R1027.
- Marshall VT, Schwartz JJ, Gerhardt HC (2006) Effects of heterospecific call overlap on the phonotactic behaviour of grey treefrogs. *Animal Behaviour* 72: 449-459.
- Miller CT, Bee MA (2012) Receiver psychology turns 20: Is it time for a broader approach? *Animal Behaviour*, 83, 331-343.
- Millman RE, Lorenzi C, Apoux F, Fullgrabe C, Green GGR, Bacon SP (2002) Effect of duration on amplitude-modulation masking. *The Journal of the Acoustical Society of America* 111:2551-2554.

- Murphy CG, Gerhardt HC (2002) Mate sampling by female barking treefrogs (*Hyla gratiosa*). *Behavioral Ecology* 13:472-480
- Nelken I, Rotman Y, Bar Yosef O (1999) Responses of auditory-cortex neurons to structural features of natural sounds. *Nature* 397:154-157.
- Narins PM, Zelick R (1988) The effects of noise on auditory processing and behavior in amphibians. In: Fritzsche B, Ryan MJ, Wilczynski W, Hetherington TE, Walkowiak W (eds) *The Evolution of the Amphibian Auditory System*. Wiley & Sons, New York, pp 511-536
- Nelson PB, Jin SH, Carney AE, Nelson DA (2003) Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners. *Journal of the Acoustical Society of America* 113:961-968
- Nieder A, Klump, GM (2001) Signal detection in amplitude-modulated maskers. II. Processing in the songbird's auditory forebrain. *European Journal of Neuroscience* 13:1033-1044.
- Nityananda V, Bee MA (2011) Finding your mate at a cocktail party: frequency separation promotes auditory stream segregation of concurrent voices in multi-species frog choruses. *PLoS ONE* 6:e21191
- Nityananda V, Bee MA (2012) Spatial release from masking in a free-field source identification task by gray treefrogs. *Hearing Research* 285: 86-97.
- Noble W, Perrett S (2002) Hearing speech against spatially separate competing speech versus competing noise. *Perception & Psychophysics* 64:1325-1336.
- Noldus (2005) *EthoVision® Video Tracking System for Automation of Behavioral Experiments: Reference Manual Version 3.1*. Noldus Information Technology, Wageningen, The Netherlands.

- Noldus (2010) EthoVision® XT Reference Manual Version 8. Noldus Information Technology, Wageningen, The Netherlands.
- Oldham RS, Gerhardt HC (1975) Behavioral isolation of the treefrogs *Hyla cinerea* and *Hyla gratiosa*. *Copeia* 1975: 223-231.
- Patterson RD, Moore BCJ (1986) Auditory filters and excitation patterns as representations of frequency resolution. In: Moore, BCJ (ed) *Frequency Selectivity in Hearing*. Academic, London, pp123-177.
- Rheinlaender J, Klump GM (1988) Behavioral aspects of sound localization. In: Fritzsche B, Ryan MJ, Wilczynski W, Hetherington T (eds) *The Evolution of the Amphibian Auditory System*. Wiley & Sons, 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. *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.
- Rieke F, Bodnar DA, Bialek W (1995) Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proceedings of the Royal Society of London, Biological Sciences* 262: 259-265.
- 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.
- Ronacher B, Krahe R (1998) Song recognition in the grasshopper *Chorthippus biguttulus* is not impaired by shortening song signals: implications for neuronal encoding. *Journal of Comparative Physiology A* 183:729-735.

- Rose GJ, Brenowitz EA, Capranica RR (1985) Species specificity and temperature dependency of temporal processing by the auditory midbrain of two species of treefrogs. *Journal of Comparative Physiology A* 157:763-769.
- Rose GJ, Capranica RR (1984) Processing amplitude-modulated sounds in the auditory midbrain of two species of toads: matched temporal filters. *Journal of Comparative Physiology A* 154:211-219.
- Rose GJ, Gooler DM (2006) Function of the Amphibian Central Auditory System. In P. M. Narins et al., eds. *Hearing and Sound Communication in Amphibians*. Springer New York, pp. 250-290.
- Rosenthal R, Rosnow RL (1991) *Essentials of behavioral research: methods and data analysis*. 2nd ed. McGraw-Hill, New York.
- Ryan MJ (2001) *Anuran Communication*. In. Smithsonian Institution Press, Washington D.C.
- Schooneveldt GP, Moore BCJ (1989) Comodulation masking release (CMR) as a function of masker bandwidth, modulator bandwidth, and signal duration. *The Journal of the Acoustical Society of America*, 85: 273-281.
- Schul J, Bush SL (2002) Non-parallel coevolution of sender and receiver in the acoustic communication system of treefrogs. *Proceedings of the Royal Society of London Series B-Biological Sciences* 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, Bee MA (In press) Anuran signal production in noisy environments. In: *Animal Communication and Noise* (Ed. by H. Brumm). New York: Springer.
- Schwartz JJ, Buchanan BW, Gerhardt HC (2001) Female mate choice in the gray treefrog

- (*Hyla versicolor*) in three experimental environments. *Behavioral Ecology and Sociobiology*, 49, 443-455.
- Schwartz JJ, Buchanan BW, Gerhardt HC (2002) Acoustic interaction among male gray treefrogs, *Hyla versicolor*, in a chorus setting. *Behavioral Ecology and Sociobiology*, 53, 9-19.
- Schwartz JJ, 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 JJ, Freeberg TM (2008) Acoustic interaction in animal groups: Signaling in noisy and social contexts - Introduction. *Journal of Comparative Psychology* 122:231-234.
- Schwartz JJ, Gerhardt HC (1989) Spatially mediated release from auditory masking in an anuran amphibian. *Journal of Comparative Physiology A* 166:37-41.
- Schwartz JJ, Huth K, Huncce 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.
- 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.
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelyd M (1995) Speech recognition with primarily temporal cues. *Science* 5234: 303-304.
- Shinn-Cunningham BG, Schickler J, Kopco N, Litovsky R (2001) Spatial unmasking of nearby speech sources in a simulated anechoic environment. *Journal of the Acoustical Society of America* 110:1118-1129

- Smith EC, Lewicki MS (2006) Efficient auditory coding. *Nature*, 439: 978-982.
- Sullivan BK, Wagner WE (1988) Variation in advertisement and release calls, and social influences on calling behavior in the Gulf Coast toad (*Bufo valliceps*). *Copeia*. 4:1014-120.
- Swanson EM, Tekmen SM, Bee MA (2007) Do female anurans exploit inadvertent social information to locate breeding aggregations? *Canadian Journal of Zoology* 85:921-932.
- Takahashi GA, Bacon SP (1992) Modulation detection, modulation masking, and speech understanding in noise in the elderly. *Journal of Speech and Hearing Research* 35:1410-1421
- Taylor RC, Buchanan BW, Doherty JL (2007) Sexual selection in the squirrel treefrog *Hyla squirella*: the role of multimodal cue assessment in female choice. *Animal Behaviour*. 74:1753-1763.
- Vélez A, Bee MA (2010) Signal recognition by frogs in the presence of temporally fluctuating chorus-shaped noise. *Behavioral Ecology and Sociobiology* 64:1695-1709.
- Vélez A, Bee MA (2011) Dip listening and the cocktail party problem in grey treefrogs: Signal recognition in temporally fluctuating noise. *Animal Behaviour* 82:1319-1327.
- Vélez A, Höbel G, Gordon NM, Bee MA (in review) Dip listening or modulation masking? Call recognition by green treefrogs (*Hyla cinerea*) in temporally fluctuating noise. *Journal of Comparative Physiology A*.
- Vélez A, Schwartz JJ, Bee MA (in press). Anuran signal perception in noisy environments. In: *Animal Communication and Noise* (Ed. by H. Brumm) Springer, New York.

- Vestergaard MD, Fyson NRC, Patterson RD (2011). The mutual roles of temporal glimpsing and vocal characteristics in cocktail-party listening. *The Journal of the Acoustical Society of America* 130, 429-439.
- 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. In: *Advances in the Study of Behavior*, Vol 36, pp 217-247
- Wiley RH, Richards DG (1978) Physical constraints on acoustic communication in the atmosphere: Implications for the evolution of animal vocalizations. *Behavioral Ecology and Sociobiology* 3:69-94.
- Wollerman L (1999) Acoustic interference limits call detection in a Neotropical frog *Hyla ebraccata*. *Animal Behaviour*, 57, 529-536.
- Wollerman L, Wiley RH (2002) Background noise from a natural chorus alters female discrimination of male calls in a Neotropical frog. *Animal Behaviour*, 63, 15-22.
- Woolley SMN, Fremouw TE, Hsu, A, Theunissen FE (2005) Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nature Neuroscience*, 8, 1371-1379.
- Zar JH (1999) *Biostatistical Analysis*, 4th edn. Prentice-Hall, Upper Saddle River, NJ

Appendix 1: Assessing the Potential for Dip Listening in Natural Settings

Results from experiment 1 in the main article (Fig. 2-2) suggested that, compared with a non-fluctuating 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. A1-1a; 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. A1-1a; 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 analyzed 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 analyzed 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, non-overlapping 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 dB. 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 non-fluctuating carrier with the same RMS amplitude. Values close to zero 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 Hz – 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

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

modulation spectrum normalized by the DC component, using the following equation:

where $m(Mr)$ is the modulation index as a function of modulation rate, $Fc \text{ envelope } (Mr)$ is the Fourier coefficient for the modulation rate corresponding to each frequency bin of the modulation spectrum, and $Fc \text{ envelope } (0)$ is the Fourier coefficient of the modulation spectrum at DC. We then recorded the maximum modulation index between 40 Hz and 60 Hz. We analyzed 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 Hz – 60 Hz component of the segment's modulation spectrum when data from the 25 recordings were pooled ($r = 0.33$, $P < 0.0001$, $N=7500$ segments, Fi. A1 b). 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. A1-1 c). When each recording was analyzed separately, we found a significant positive correlation for 22 chorus recordings ($0.15 \leq r \leq 0.74$, $P_s < 0.01$, $N = 300$ segments per recording, Fig. A1-1 d, e) and no significant correlation for three chorus recordings ($-0.06 \leq r \leq 0.05$, $P_s > 0.28$, $N = 300$ segments per recording, Fig. A1-1f). 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.

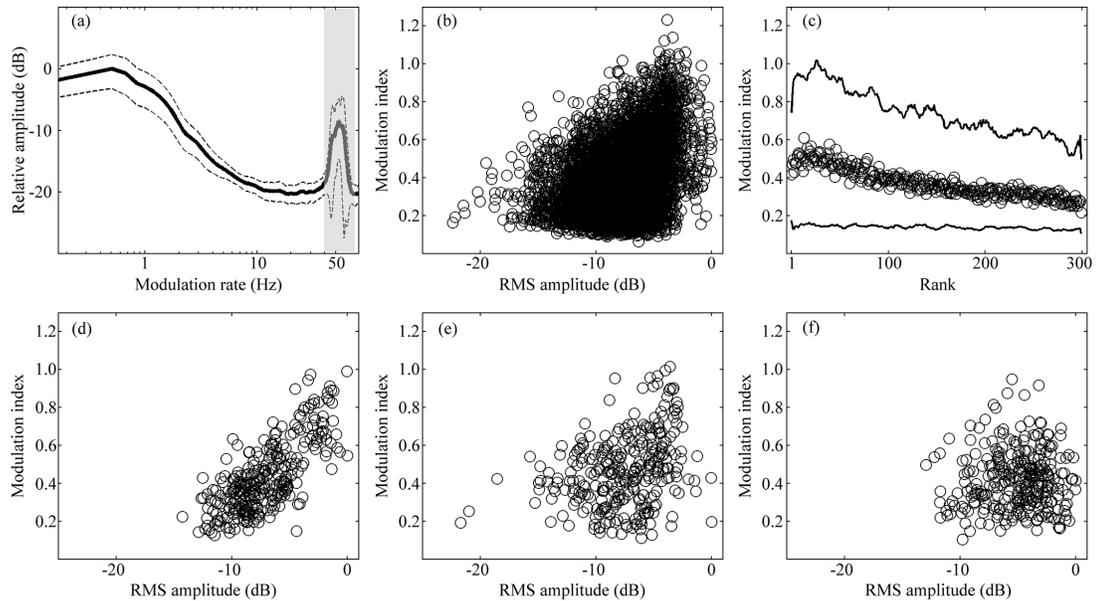


Figure A1-1. Analyses of the temporal level fluctuations in 60-s recordings of 25 Cope's grey treefrog choruses. (a) Mean (bold, solid line) ± 1 standard deviation (thin, dotted lines) modulation spectrum of grey treefrog choruses shown on a logarithmic x-axis. This modulation spectrum depicts the rates 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 figure A1-1a can be found in Vézé & 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 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-s 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. (d-f) Examples of individual recordings ($N=300$ segments per recording) in which the RMS amplitude value and the modulation indices were significantly correlated with a high correlation coefficient ($P < 0.0001$, $r = 0.74$) (d), significantly correlated with a low correlation coefficient ($P < 0.0001$, $r = 0.31$) (e), and not significantly correlated ($P = 0.49$, $r = 0.04$) (f).

Appendix 2: Analyses of original movement-pattern variables in sham trials.

During sham trials, we examined female behavior in the test arena in the absence of acoustic stimulation and in the presence of our chorus-shaped maskers broadcast from the overhead speaker. In sham trials, no target signal was broadcast from speakers on the floor. To determine the extent to which female behavior was differentially affected by the different maskers, we compared female movement patterns across treatments. Using the animal tracking software Ethovision v3.1 (Noldus 2005), we measured (i) the total distance moved by the subjects (in cm), (ii) the average velocity of their movements (in cm/s), (iii) the average turn angles (in degrees), and (iv) the meander (in degrees/cm), which quantifies changes in direction of movements relative to the distance moved (see Bee and Riemersma 2008). These variables were correlated (Table A2-1) and the results of principal component analysis (PCA) are reported in the main manuscript. Here, we provide results of the analyses on the raw variables.

Results from analyses of the original movement-pattern variables parallel those on the two factors of the principal component analysis. As stated in the main manuscript, PCA factor 1 loaded with variables describing angular movement: average turn angle and meander. We found no differences across treatments in the average turn angle ($H_{9, 58} = 6.48, P = 0.69$), or meander ($H_{9, 58} = 9.34, P = 0.41$) during sham trials (Figure A2-1a, b). PCA factor 2 loaded with variables describing non-angular movements: distance moved and velocity. There were no differences across treatments in the total distance moved (Kruskal-Wallis test: $H_{9, 58} = 3.27, P = 0.95$) or the average velocity ($H_{9, 58} = 9.54, P = 0.39$) (Figure A2-1c, d). These results suggest that female behavior in the test arena was not differently affected by the type of masker broadcast from the overhead speaker. Furthermore, our results show that female movement patterns in the presence of chorus-shaped noises were not different than those in the absence of sound.

Table A2-1. Correlations between the four original movement-pattern variables.

	Turn angle (degrees)	Meander (degrees/cm)	Distance moved (cm)	Velocity (cm/s)
Turn angle (degrees)	1.00	0.87	0.24	-0.10
Meander (degrees/cm)	0.87	1.00	0.05	-0.27
Distance moved (cm)	0.24	0.05	1.00	0.34
Velocity (cm/s)	-0.10	-0.27	0.34	1.00

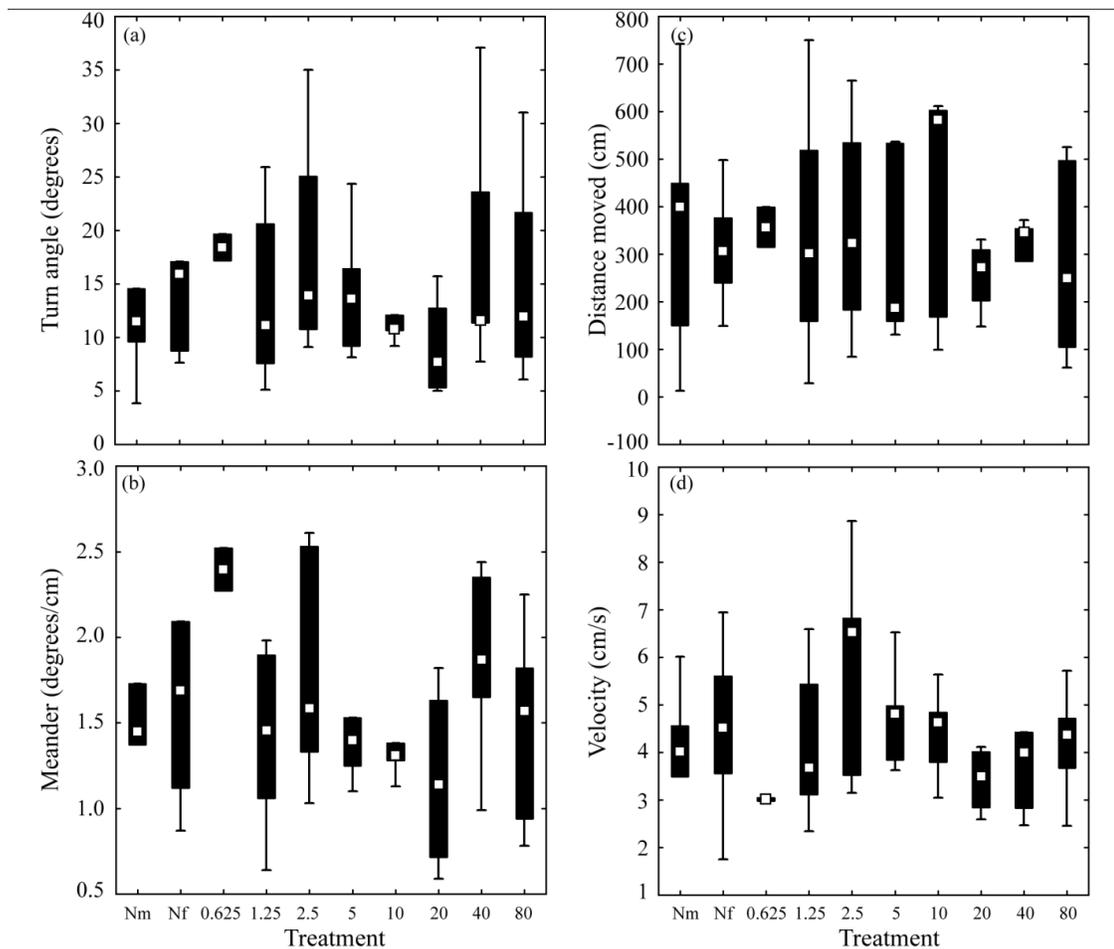


Figure A2-1. Movement patterns in the sham trials. Shown here are the median (point), inter-quartile range (box), and non-outlier range (whiskers) for average turn angle (a), meander (b), total distance moved (c), and average velocity (d) in each treatment. Sample sizes for each treatment are 5 in the Non-fluctuating noise (Nf), 7 in the No-noise (Nm), 2 in the 0.625 Hz SAM noise (0.625), 8 in the 1.25 Hz SAM noise (1.25), 10 in the 2.5 Hz SAM noise (2.5), 5 in the 5 Hz SAM noise (5), 5 in the 10 Hz SAM noise (10), 4 in the 20 Hz SAM noise (20), 5 in the 40 Hz SAM noise (40), and 7 in the 80 Hz SAM noise (80) treatments.

Appendix 3: Signal recognition in fluctating noise: Comparison between Cope's gray treefrogs and green treefrogs

We have recently shown that, depending on the rates at which background noise levels fluctuate, Cope's gray treefrog (*Hyla chrysoscelis*) females experience dip listening, modulation masking, or neither (Vélez and Bee 2011). Following the same protocol, we show in the present study that green treefrog (*H. cinerea*) females experienced neither masking release nor modulation masking in the presence of fluctuating noises. Here, we reanalyzed data from Vélez and Bee (2011) using planned comparisons, as we did in the present study, to make a between-species comparison of the effects of masker level fluctuation rates on signal recognition by female frogs.

In figure A3-1, we show mean (\pm SE) signal recognition thresholds in the presence of the non-fluctuating control treatment and all fluctuating noise treatments for both species. Results from planned comparisons of signal recognition thresholds in fluctuating noise treatments against the non-fluctuating control treatment for both species are summarized in table A3-1. Female green treefrogs do not experience dip listening or modulation masking; mean signal recognition thresholds are not different from that in the fluctuating control. The effect sizes reveal that, in green treefrogs, less than 2% of the variation in signal recognition thresholds can be attributed to differences in the fluctuation rates of the masking noise. In contrast, Cope's gray treefrogs experience dip listening in slowly fluctuating backgrounds and modulation masking in the presence of maskers that fluctuate at faster rates. The effect of masker level fluctuations on signal recognition is small to moderate, accounting for up to almost 10% of the variation in signal recognition thresholds.

Table A3-1. Individual planned comparisons between the non-fluctuating control treatment and all fluctuating treatments for green treefrogs (*H. cinerea*) and Cope’s gray treefrogs (*H. chrysoscelis*). Asterisks represent statistically significant differences.

Treatment	<i>H. cinerea</i> planned comparisons				<i>H. chrysoscelis</i> planned comparisons			
	Contrast estimate	F _{1,207}	P	partial η^2	Contrast estimate	F _{1,207}	P	partial η^2
0.625 Hz SAM	-1.56	1.37	0.2428	0.007	-1.91	2.60	0.1087	0.012
1.25 Hz SAM	0.44	0.11	0.7432	0.001	-3.63	9.33	0.0026*	0.043
2.5 Hz SAM	1.60	1.44	0.2312	0.007	-2.56	4.66	0.0320*	0.022
5 Hz SAM	0.63	0.22	0.6399	0.001	-0.44	0.14	0.7128	0.001
10 Hz SAM	2.63	3.87	0.0504	0.018	0.56	0.22	0.6360	0.001
20 Hz SAM	2.50	3.51	0.0622	0.017	1.31	1.22	0.2700	0.006
40 Hz SAM	2.06	2.39	0.1235	0.011	5.56	21.97	<0.001*	0.096
80 Hz SAM	0.31	0.05	0.8149	<0.001	3.75	9.98	0.0018*	0.046

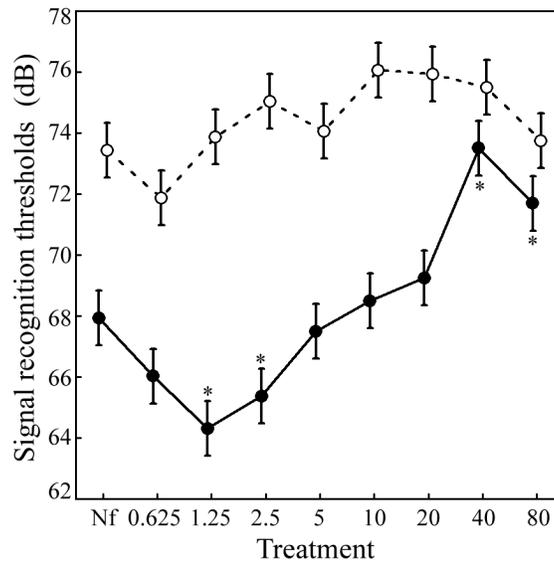


Figure A3-1. Mean (\pm SE) thresholds as a function of masking treatment for green treefrogs (open circles, dotted lines) and gray treefrogs (filled circles, solid lines). Asterisks represent statistically significant differences using planned comparisons against the non-fluctuating control. Nf = Non-fluctuating treatment; 0.625 – 80 represent SAM rates of each fluctuating noise treatment.

Appendix 4: Analyses of signal recognition thresholds excluding attractive maskers

Results from our signal recognition experiment (Experiment 1) could be confounded by the fact that Cope's gray treefrog females were attracted to the conspecific masker and green treefrog females were attracted to the random masker. Here, we reanalyzed signal recognition thresholds and tested our three hypotheses for each species excluding the potentially confounding maskers. Our results show the same pattern as the one reported in the main document including all maskers.

For green treefrogs, we excluded the random treatment from our analysis. Differences in mean signal recognition thresholds across the non-fluctuating, conspecific, and heterospecific treatments were not significant ($F_{2, 72} = 0.34$, $P = 0.711$, partial $\eta^2 = 0.009$). We did not perform planned comparisons to test the dip-listening hypothesis or the species-specific hypotheses for this species. These results parallel the results of our analysis including all masking treatments.

In Cope's gray treefrogs, we excluded the conspecific treatment from the analyses we report below. In contrast to green treefrogs, the effect of masking treatment was significant ($F_{2, 72} = 5.69$, $P = 0.005$, partial $\eta^2 = 0.136$). We used planned comparisons between the non-fluctuating and the fluctuating treatments to test the dip listening hypothesis. Mean signal recognition thresholds in fluctuating maskers were not different from those in the non-fluctuating control treatment ($F_{1, 72} = 1.62$, $P = 0.207$, partial $\eta^2 = 0.022$). Compared to the non-fluctuating treatment, thresholds were not different in the heterospecific treatment ($F_{1, 72} = 0.21$, $P = 0.646$, partial $\eta^2 = 0.003$), and lower in the random treatment ($F_{1, 72} = 7.09$, $P = 0.009$, partial $\eta^2 = 0.090$). We tested the natural-properties advantage hypothesis with planned comparisons between the random treatment and the heterospecific treatment. In contrast to the natural-properties advantage hypothesis, mean signal recognition thresholds were higher in heterospecific treatment than in the random treatment ($F_{1, 72} = 9.76$, $P = 0.002$, partial $\eta^2 = 0.119$).