Behavioral measures of signal recognition thresholds in frogs in the presence and absence of chorus-shaped noise

Mark A. Bee[a)]
Department of Ecology, Evolution, and Behavior, University of Minnesota, 100 Ecology, 1987 Upper Buford Circle, St. Paul, Minnesota 55108

Joshua J. Schwartz
Department of Biology and Health Sciences, Pace University, Pleasantville, New York 10570

(Received 4 June 2009; revised 11 August 2009; accepted 12 August 2009)

Anuran amphibians are superb animal models for investigating the mechanisms underlying acoustic signal perception amid high levels of background noise generated by large social aggregations of vocalizing individuals. Yet there are not well-established methods for quantifying a number of key measures of auditory perception in frogs, in part, because frogs are notoriously difficult subjects for traditional psychoacoustic experiments based on classical or operant conditioning. A common experimental approach for studying frog hearing and acoustic communication involves behavioral phonotaxis experiments, in which patterns of movement directed toward sound sources indicate the subjects' perceptual experiences. In this study, three different phonotaxis experiments were conducted using the same target signals and noise maskers to compare different experimental methods and analytical tools for deriving estimates of signal recognition thresholds in the presence or absence of "chorus-shaped noise" (i.e., artificial noise with a spectrum similar to that of real breeding choruses). Estimates of recognition thresholds based on measures of angular orientation, response probabilities, and response latencies were quite similar in both two-choice and no-choice phonotaxis tests. These results establish important baselines for comparing different methods of estimating signal recognition thresholds in frogs tested in various masking noise conditions.
[DOI: 10.1121/1.3224707]

PACS number(s): 43.80.Nd, 43.80.Lb [JAS] Pages: 2788–2801

I. INTRODUCTION

Animals that signal acoustically in large social aggregations, such as choruses (e.g., Gerhardt and Huber, 2002) and communal crèches (e.g., Aubin and Jouventin, 2002), represent ideal model systems for investigating how animals cope with the problems of noise (Schwartz and Freeberg, 2008). For such species, the potential impacts of auditory masking and interference should be especially severe because of the high degree of overlap among the spectral and temporal properties of the concurrent signals of other conspecific individuals. In humans, these impacts give rise to the so-called "cocktail-party problem," which refers to our difficulty perceiving speech in noisy social settings (Cherry, 1953; Bee and Micheyl, 2008). Anuran amphibians (frogs and toads) represent one taxonomic group of non-human animals for which cocktail-party-like problems are likely to be pronounced in acoustic communication (Narins and Zelick, 1988; Feng and Schul, 2007). In many anurans, reproduction takes place in large and often dense breeding aggregations in which males produce loud advertisement calls that are both necessary and sufficient to attract gravid females (Gerhardt and Huber, 2002).

Here, we report results from the latest in a series of studies aimed at understanding how females of Cope's gray treefrog (*Hyla chrysoscelis*) perceive the acoustic signals of sexually advertising males amid the high levels of noise present in breeding choruses (see Bee, 2007, 2008a, 2008b; Bee and Swanson, 2007; Swanson *et al.*, 2007). The primary goal of this study was to evaluate various behavioral and analytical methods for estimating "signal recognition thresholds" in the presence and absence of "chorus-shaped noise" using phonotaxis experiments. We used both no-choice and two-choice phonotaxis tests (Gerhardt, 1995; Ryan and Rand, 2001) to measure as a function of signal level the responses of females to a conspecific advertisement call presented in the presence or absence of noise, and we explored a range of threshold criteria for estimating signal recognition thresholds. Our study was aimed at evaluating procedures that could allow researchers to use phonotaxis experiments as a tool to investigate more systematically a number of empirical questions concerning how frogs communicate in noisy social environments (reviewed in Bee and Micheyl, 2008). Our approach is one inspired by studies of human speech perception in noise, which commonly measure the "speech reception threshold" (SRT) (Plomp, 1978; Plomp and Mimpen, 1979a, 1979b). The SRT corresponds to the sound level at which a segment of speech must be presented to listeners in order for intelligibility to reach some predetermined threshold level as measured by the percentage of target words correctly repeated. Framed more broadly, the SRT relies on a correct response from a listener in the form of a species-typical behavior that is used to measure thresholds

---

[a)]Author to whom correspondence should be addressed. Electronic mail: mbee@umn.edu

for recognizing conspecific vocal signals. In this study, we operationally defined signal recognition as occurring when a female exhibited a correct response (phonotaxis) with respect to a conspecific signal (see discussion of terminology in Gerhardt and Huber, 2002 and Sec. VI A.); we operationally defined the signal recognition threshold as the minimum signal level required to elicit phonotaxis behavior exceeding a pre-determined criterion level of correct responses.

II. METHODS

A. The study system

Gray treefrogs represent a cryptic species complex comprising Cope's gray treefrog (*H. chrysoscelis*, a diploid) and the eastern gray treefrog (*H. versicolor*, a tetraploid) (Holloway *et al.*, 2006). Both species range widely throughout much of eastern North America and have been the subjects of extensive and detailed investigations of hearing and acoustic communication, species recognition, reproductive behavior, and sexual selection (reviewed in Gerhardt, 2001). Like many other frogs, male gray treefrogs form spring breeding choruses in which they produce advertisement calls with amplitudes ranging from about 85 to 93 dB sound pressure level (SPL) rms (96–104 dB peak) measured at a distance of 1 m (Gerhardt, 1975). The nearly continuous background noise levels in dense gray treefrog choruses can be quite intense, reaching sustained levels of 70–80 dB SPL (Schwartz *et al.*, 2001; Swanson *et al.*, 2007; Vélez and Bee, unpublished data). Gray treefrog advertisement calls comprise a series of discrete pulses produced at species-specific rates; in *H. chrysoscelis*, the pulse rate (about 40–50 pulses/s) is an important species recognition cue (Gerhardt, 2001; Schul and Bush, 2002). The call has a bimodal frequency spectrum, with a fundamental frequency in *H. chrysoscelis* of about 1200–1400 Hz that has an amplitude of -5 to -10 dB relative to the dominant frequency of about 2400–2800 Hz (Gerhardt, 2001). The spectrum of the background noise generated in gray treefrog choruses is similar to that of the advertisement call (Swanson *et al.*, 2007).

B. Subjects

We collected 132 pairs of *H. chrysoscelis* in amplexus between 2100 and 0100 h during the 2008 breeding season (May and June) from wetlands in the Carver Park Reserve (Carver Co., Chaska, MN). Pairs were returned to the laboratory and kept at 2 °C to delay egg deposition until the females were tested (usually within 24 h). Pairs were returned to their original location of capture after testing. Additional details about collections and handling of frogs have been published elsewhere (Bee, 2007, 2008a, 2008b; Bee and Swanson, 2007). Of the 132 females collected and tested for this study, 120 females met an inclusion criterion that required them to finish a designated series of tests to be included as subjects in statistical analyses.

C. General testing procedures

On the day of testing, we transferred pairs to a 20 °C incubator where they remained at least 1 h prior to testing until their body temperature reached 20 °C (±1 °C). Phonotaxis experiments were conducted in two temperature-controlled, hemi-anechoic sound chambers [Industrial Acoustics Co., (IAC), Bronx, NY; inside dimensions $L \times W \times H$: $300 \times 280 \times 216$ cm^3 and $220 \times 280 \times 216$ cm^3]. The inside walls and ceiling of the chambers were painted dark gray and treated with IAC's Planarchoic™ treatment to reduce reverberation. The chambers had vibration-isolation floors that were covered in dark gray carpet. We controlled the temperature inside the chambers at 20 °C (±2 °C), which is a typical temperature at which gray treefrogs breed. With their ventilation units running, the SPL of each chamber's ambient noise floor ranged between 2 and 12 dB SPL (fast rms, flat weighting) in the 1/3-octave bands between 500 and 4000 Hz, which spans the frequency range of interest in this study. The frequency responses of the playback systems in the two chambers were flat (±3 dB) over the same frequency range.

We used ADOBE AUDITION v1.5 (Adobe Systems Inc., San Jose, CA) to broadcast digital acoustic stimuli (20 kHz sampling rate, 16-bit resolution) from a Dell Computer (Optiplex GX620 or GX745; Dell Computer Corp., Round Rock, TX) located outside each chamber. Each computer was interfaced with an M-Audio FireWire 410 multichannel soundcard (M-Audio USA, Irwindale, CA), and the output of the soundcard was amplified using either a Sonamp 1230 (Sonance, San Clemente, CA) or HTD 1235 (Home Theater Direct Inc., Plano, TX) multichannel amplifier.

All behavioral tests were conducted under infrared (IR) illumination provided by two IR light sources (Noldus Information Technology Inc., Leesburg, VA) in each sound chamber that were mounted near the ceilings on opposite walls. We monitored behavioral responses in real time from outside the chamber and recorded them direct to digital video using real-time MPEG encoders (MVR1000SX or MPEGPRO EMR, Canopus, San Jose, CA) interfaced with an overhead, IR-sensitive Panasonic WV-BP334 video camera (Panasonic Corporation of North America, Secaucus, NJ) mounted from the center of each sound chamber's ceiling.

Phonotaxis tests were performed in 2-m diameter circular test arenas (one per sound chamber) with walls that were 60-cm high and constructed from acoustically transparent but visually opaque black cloth and hardware cloth. The sound chambers' carpeted floors served as the test arena floors, which were divided into 24 15° arcs along their perimeters. Synthetic advertisement calls (see below) were broadcast through A/D/S L210 speakers (Directed Electronics, Inc., Vista, CA) that were placed in the center of the 15° arcs on the floor just outside the walls of the test arenas and directed toward the arenas' centers. The positions of speakers around an arena's perimeter were varied on a regular basis to eliminate any possibility of a directional response bias in our sound chambers. Masking noise (see below) was broadcast through a Kenwood KFC-1680ie speaker (Kenwood USA Corporation, Long Beach, CA) suspended from the ceiling of each chamber 190 cm above the center of the test arena. The overhead speaker created a uniform (±1.5 dB) noise level across the entire floor of an arena.

At the beginning of each phonotaxis test, a female was placed in a holding cage located on the floor of the sound chamber at the center of the test arena. The holding cage consisted of a shallow, acoustically transparent cup (9-cm diameter; 2-cm height) with a lid that could be removed using a rope and pulley system operated from outside the sound chamber to allow females unrestricted movement within the test arena. Stimulus broadcasts began after a 1.5-min acclimation period and were continued throughout the duration of a test. Females were remotely released from the holding cage after 30 s of signal presentation. Sound levels were measured and calibrated prior to testing by placing the microphone of a Larson-Davis System 834 (Larsen Davis, Depew, NY) or a Brüel & Kjær type 2250 (Brüel & Kjær Sound & Vibration Measurement A/S, Nærum, Denmark) sound level meter at the approximate position of a subject's head at the central release point. Below, we report all signal and noise levels and threshold estimates in units of dB SPL (re 20 μPA, fast rms, *C*-weighted).

III. EXPERIMENT 1: POPULATION-LEVEL RECOGNITION THRESHOLDS IN TWO-CHOICE TESTS

In this experiment, we explored a number of methods to estimate recognition thresholds based on pooling data from a group of subjects tested using a traditional two-choice experimental design (Gerhardt, 1995). Each subject was presented with a target signal (conspecific call) that alternated in time with a non-target signal (heterospecific call). The rms SPLs of the two signals were always equal in each test but were varied systematically between different tests. Subjects experienced each signal level only one time, and subjects were assigned randomly to one of two noise conditions, a "no-noise group" ($N=20$) or a "noise group" ($N=20$). Our estimates of recognition thresholds for each group are based on performance measures that we derived from the behavior of the entire pool of subjects in that group; therefore, we regard these estimates as being "population level" in the sense that we did not attempt to estimate a threshold for each individual as is more common in traditional psychoacoustic experiments (Klump *et al.*, 1995).

A. Methods

1. Experimental design

Experiment 1 was based on a 9 signal level (within subjects) × 2 noise condition (between subjects) factorial design. Across nine "treatment conditions," we varied the SPLs of the target and non-target signals across nine nominal levels ranging from 37 to 85 dB SPL in 6-dB steps. The SPLs of both the target and the non-target signal were adjusted using software control of the M-Audio soundcard to have the nominal signal level at the position of a subject's head at the central release site located 1 m from the speakers. The source level of the signals remained constant during a test, and the order of different signal levels across the nine treatment conditions was determined randomly for each subject.

The acoustic signals were synthetic advertisement calls that were created using custom-made software and modeled after natural calls having spectral and temporal properties close to the averages (corrected to 20 °C) of calls recorded in local Minnesota populations (Bee, unpublished data). The target signal [Fig. 1(a)] was a conspecific (*H. chrysoscelis*) call that was 693 ms in duration and consisted of 32 pulses (11-ms pulse duration) delivered at a rate of 45.5 pulses/s (22-ms pulse period). The non-target (heterospecific) signal was an advertisement call of the closely related eastern gray treefrog (*H. versicolor*), was 690 ms in duration, and consisted of 12 pulses (30-ms pulse duration) delivered at a rate of 16.7 pulses/s (60-ms pulse period). In both signals, each pulse consisted of two harmonically related, phase-locked sinusoids with frequencies of 1.25 and 2.5 kHz and relative amplitudes of −6 and 0 dB, respectively [Fig. 1(b)]. Both signals repeated with a period of 5 s, which is within the range of call periods in local populations (corrected to 20 °C), and alternated so that equal durations of silence preceded and followed each signal. For half of the subjects tested, the alternating stimulus sequence began with the target signal; the other half of subjects heard the non-target signal first. Where possible, we included this "call order"

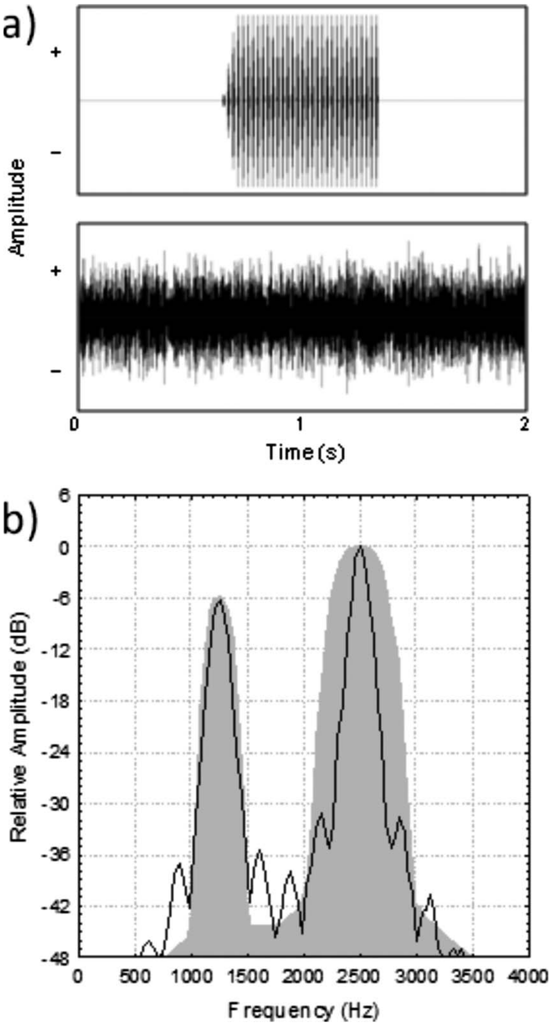

FIG. 1. Target signal and chorus-shaped noise. (a) Waveforms of the synthetic *H. chrysoscelis* call comprising the target signal (top) and a segment of chorus-shaped noise (bottom). (b) Power spectra showing the spectral profile of the target signal (black line) relative to that of the chorus-shaped noise (gray area).

effect in our statistical analyses. The target and non-target signals were broadcast from separate speakers located directly opposite each other around the circular test arena (i.e., 2 m and 180° apart).

All subjects were tested individually in a sequence of 13 phonotaxis tests and were given a 5- to 10-min timeout period inside the incubator between consecutive tests. A test sequence began with a "reference condition" and then alternated between three treatment conditions and another reference condition until all nine treatment conditions had been tested. Each sequence ended with a final test of the reference condition. Hence, each subject was tested a total of 13 times (9 treatment conditions + 4 reference conditions). For all subjects, the reference condition consisted of broadcasting the alternating target and non-target signals at 85 dB SPL in the absence of any broadcast masking noise.

Subjects in the no-noise group always experienced the signals without the broadcast of any additional masking noise. For subjects assigned to the noise group, we broadcast the signals in the presence of a chorus-shaped noise that had a long-term spectrum with acoustic energy at audio frequencies characteristic of gray treefrog choruses (Fig. 1). We used MATLAB v7.6.0 to create five different exemplars of chorus-shaped noise in the following way. For each exemplar, we first filtered two copies of the same 6-min long white noise to create two narrowband noises centered at 1250 and 2500 Hz. The low-frequency band was created using a band-pass finite impulse response (FIR) filter of order 300, pass-band frequencies of 1200 and 1300 Hz, and stop-band frequencies of 1000 and 1500 Hz. The high-frequency bands were created using a band-pass FIR filter of order 150, with pass-band frequencies of 2400 and 2600 Hz and stop-band frequencies of 2000 and 3000 Hz. Both FIR filters had pass-band ripples of 0.1 Hz and stop-band attenuations of 60 dB. The peak amplitude of the low-frequency band was attenuated 6 dB relative to that of the high-frequency band and then both noises were digitally added to create a single chorus-shaped noise (Fig. 1). During behavioral tests with subjects in the noise group, we began broadcasts of the noise 30 s prior to the onset of the alternating signals and the broadcast continued over the duration of the test. The sound level of each noise exemplar was calibrated to be 70 dB SPL (LC_{eq}) at the approximate position of a subject's head at the release site at the center of the test arena.

2. Data analysis

We scored a "correct response" if the subject touched the arena wall inside the 15° arc in front of the speaker that was broadcasting the target (conspecific) signal within 5 min of being released. In essence, a correct response is one that would likely result in the selection of a conspecific mate under natural conditions, and hence demonstrates correct recognition of a conspecific vocalization. In keeping with our operational definition of recognition (Sec. I), we scored an "incorrect response" if either of the following two conditions were met: (i) a subject touched the arena wall in the 15° arc in front of the speaker broadcasting the non-target (heterospecific) signal; (ii) the subject failed to touch the wall in front of either speaker within 5 min. These two behavioral outcomes might have different underlying causes (e.g., failed recognition versus failed detection, respectively); however, both outcomes are inconsistent with our operational definition of recognition. Although phonotaxis toward the non-target (heterospecific) signal indicates acceptance of the signal as that of an appropriate mate (e.g., Ryan and Rand, 1993), we consider it an incorrect response because such a choice in nature would result in the production of offspring that were inviable, infertile, or of reduced attractiveness (Gerhardt *et al.*, 1994). An incorrect response was only considered legitimate if the subject exhibited a correct response in the next reference condition in sequence. Only subjects that exhibited correct responses in all four reference conditions were included in statistical analyses. We also excluded from the final data set any subject that required more than twice as long to respond in the final reference condition compared with the first reference condition. Such procedures ensure the validity of no responses (or slow responses) in treatment conditions by confirming that subjects remain highly motivated to respond over the duration of the test sequence (Bush *et al.*, 2002; Schul and Bush, 2002). Of the 46 females tested in this experiment, 6 did not meet our response criteria, yielding a final sample size of $N=40$. None of the subjects in this experiment had been tested previously. We used analysis of variance (ANOVA) to assess differences in response latency across the four reference conditions.

We assessed differences in three related response variables as functions of noise condition and signal level.

(a) *Angular orientation.* We assessed the directedness of phonotaxis toward the target signal by measuring a subject's angular orientation relative to the position of the target playback speaker (designated as 0°) when it first reached a distance of 20 cm away from the central release point. We chose a distance of 20 cm as a compromise between analyzing the angles at which females exited the release cage and the angles at which they first touched the arena wall 1 m away. To exit our release cage, females had to climb over a 2-cm high barrier. In our experience, females sometimes climb over this barrier in one direction only to quickly reorient and initiate movement in a different direction while still located immediately adjacent to the release cage. Thus, we believe allowing the females to freely move about on the arena floor (outside of the release cage) is a relatively more accurate measure of orientation behavior. However, we also believe that restricting the measurement distance to 20 cm minimizes any cues related to the variation in signal levels experienced by moving about in the sound field on the arena floor. According to both the inverse square law and our own empirical measurements in the sound chambers, the gain in signal level experienced by moving 20 cm closer to a source originally located 1 m away is less than 2 dB, which is much less than the 6-dB step-size we used between adjacent signal levels. We used V-tests (Zar, 1999) to test the null hypothesis that angles at 20 cm were uniformly distributed against the alternative hypothesis that subjects oriented toward the target signal at 0°. We estimated an upper threshold bound as the lowest signal level at which statistically significant orientation occurred at that level and also at all higher levels; the next lowest signal level was used as an estimate of a lower bound (LB). We then computed a recognition threshold as the average of the upper bound (UB) and LB using the following equation:

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

(b) *Response probabilities.* We used generalized linear models (proc GENMOD in SAS) to examine differences in the probability of a correct response (1=correct response, 0=incorrect response) as functions of noise condition, signal level, and call order. These models are based on the binomial distribution, the logit link function, and the use of generalized estimating equations (GEEs) for estimating within-subjects effects (Horton and Lipsitz, 1999). We explored a range of threshold criteria based on extrapolated and interpolated values along the best-fit logistic regression curves relating response probability to signal level.

We also used data on raw response probabilities to estimate thresholds in two additional ways. First, following Beckers and Schul (2004), we estimated the UB of a recognition threshold as the lowest signal level at which greater than 50% of subjects exhibited correct responses at that level and also at all higher levels. LBs were determined as the next lowest signal level, and the recognition threshold was determined using Eq. (1). Second, we estimated thresholds based on determining as an UB the lowest signal level at which the proportion of females touching the arena wall in front of the target speaker exceeded the chance probability of doing so at that signal level and also at all higher levels. In a series of preliminary experiments, we estimated the false alarm rate of our response criterion by examining the behavior of female gray treefrogs in the test arena when no signals were presented in the presence or absence of chorus-shaped noise. In assessing the proportion of subjects that touched the arena wall in a 15° arc in front of a silent speaker within 5 min, we found that approximately 10%–20% would be expected to do so by chance in the absence of any signal using our protocol (Vélez and Bee, unpublished data). Therefore, we used Eq. (1) to estimate a threshold based on taking as an UB the lowest signal level at which the proportion of subjects exhibiting correct responses was significantly greater than 0.20 (two-tailed binomial tests) at that level and also at all higher levels; the next lowest signal level was taken as the LB.

(c) *Phonotaxis scores.* As a third and final measure of behavioral responsiveness, we calculated "phonotaxis scores," which normalize the latency in a treatment condition by dividing the average latency from the two most temporally proximal reference conditions by the latency in the treatment condition (Bush *et al.*, 2002; Schul and Bush, 2002; Beckers and Schul, 2004). A phonotaxis score of 1.0 thus indicates that the latency in the treatment condition equals that in the temporally proximal reference conditions; scores greater than 1.0 and less than 1.0 indicate latencies that are shorter and longer, respectively, than those in the reference conditions. We assigned a score of 0.0 when subjects exhibited incorrect responses. We analyzed phonotaxis scores using a 9 signal level (within subjects) × 2 noise condition (between subjects) × 2 call order (between subjects) ANOVA and report the Greenhouse and Geisser (1959) corrected P values for tests involving within-subjects effects with more than a single numerator degree of freedom. We used curve fitting procedures to separately compute the best-fit sigmoid function relating mean phonotaxis scores to signal level in the two noise conditions according to the following equation:

$$\text{phonotaxis score} = a/(1.0 + e^{(-(\text{signal level}-b)/c)}) + d, \qquad (2)$$

where a, b, c, and d are the fitted parameters that minimized the sum of the squared absolute error, and e is the base of the natural logarithm. We chose to fit our data with sigmoid curves because such curves often characterize the shapes of both psychometric functions generated in psychophysical experiments and neuronal rate-level functions generated in electrophysiological studies. We used the fitted sigmoid equations to explore a range of criteria for estimating recognition thresholds from phonotaxis scores.

B. Results and discussion

Subjects remained highly motivated to respond over the duration of the test sequence as evidenced by their uniformly strong orientation toward the target signal in the four reference conditions (Table I). Averaged across all four reference conditions and both noise conditions (i.e., the noise and no-noise groups), subjects made their correct responses with a mean (±SD) latency of 76.3 ± 28.6 s. There were no significant differences in latency across the four reference conditions ($F_{3,108}=1.9$, $P=0.1461$). There were also no significant differences in latency between the two noise conditions ($F_{1,36}=4.0$, $P=0.0523$) or according to which signal (target or non-target) was broadcast first ($F_{1,36}=3.5$, $P=0.0695$), nor were there any significant interactions between any of the main effects ($0.1461 < Ps < 0.9702$).

1. Angular orientation

Based on our measures of angular orientation at a distance of 20 cm, the difference in recognition thresholds between the no-noise and noise groups was 30 dB. In the no-noise group, subjects oriented significantly in the direction of the conspecific target signal at signal levels of 43 dB and higher (Table I). Using 43 dB as an UB and 37 dB as the LB, we calculated a recognition threshold of 41 dB in the no-noise group. In the presence of chorus-shaped noise, signal levels of 73 dB and higher elicited significant orientation toward the target signal (Table I). Taking 67 dB as a LB we calculated a recognition threshold of 71 dB in the noise group.

TABLE I. Results of circular statistical analyses for Experiment 1 (two-choice tests).

Noise condition	Signal condition	Mean vector ($\mu°$)	Length of mean vector (r)	Circular SD (deg)	N	V	P
No-noise	Reference 1	−5	0.89	28	20	0.88	<0.0001
	Reference 2	−6	0.90	27	20	0.89	<0.0001
	Reference 3	11	0.93	22	20	0.91	<0.0001
	Reference 4	0	0.84	34	20	0.84	<0.0001
	37 dB	9	0.24	97	16	0.23	0.0940
	43 dB	26	0.51	67	16	0.46	0.0040
	49 dB	−24	0.54	64	18	0.49	0.0010
	55 dB	6	0.54	64	18	0.53	0.0005
	61 dB	20	0.54	64	20	0.51	0.0005
	67 dB	−3	0.84	34	20	0.84	<0.0001
	73 dB	−1	0.78	41	20	0.78	<0.0001
	79 dB	−2	0.89	27	20	0.89	<0.0001
	85 dB	−2	0.95	19	20	0.95	<0.0001
Noise	Reference 1	5	0.98	12	20	0.98	<0.0001
	Reference 2	−5	0.96	17	20	0.96	<0.0001
	Reference 3	1	0.95	18	20	0.95	<0.0001
	Reference 4	2	0.96	16	20	0.96	<0.0001
	37 dB	−164	0.31	87	18	−0.30	0.9650
	43 dB	−175	0.15	113	17	−0.14	0.7980
	49 dB	60	0.31	88	19	0.16	0.1710
	55 dB	−14	0.24	97	18	0.23	0.0810
	61 dB	−75	0.36	82	18	0.09	0.2910
	67 dB	139	0.16	110	18	−0.12	0.7670
	73 dB	−22	0.46	72	20	0.43	0.0030
	79 dB	−2	0.97	14	19	0.97	<0.0001
	85 dB	−3	0.94	19	20	0.94	<0.0001

2. Response probabilities

The proportion of subjects that exhibited correct responses increased as a function of increasing signal level in both the no-noise and noise groups, and this level-dependent increase began at higher signal levels in the presence of chorus-shaped noise [Figs. 2(a)–2(c)]. In the generalized linear model for this experiment, the parameter relating response probability to signal level was significantly different from zero ($\chi^2=29.3$, $P<0.0001$, df=1). The parameters for noise condition ($\chi^2=3.4$, $P=0.0661$, df=1), call order ($\chi^2=0.1$, $P=0.7037$, df=1), and the interaction between signal level and noise condition ($\chi^2=0.03$, $P=0.8545$, df=1) were not different from zero. Subsequent contrast analyses of least-squares means, however, revealed significant differences between the no-noise and noise groups ($\chi^2=23.3$, $P<0.0001$, df=1). The proportions of correct responses in the no-noise group were significantly greater than zero ($\chi^2=25.7$, $P<0.0001$, df=1) while those in the noise group were not ($\chi^2=0.7$, $P=0.4021$, df=1). There was no significant effect according to which signal initiated the alternating broadcasts ($\chi^2=0.2$, $P=0.6934$, df=1). In Table II, we summarize threshold estimates based on different threshold criteria expressed as the probability of a correct response (p') along the fitted logistic regression functions [Fig. 2(c)]. Differences in thresholds between the no-noise and noise groups were consistently close to 20 dB regardless of the threshold criterion (Table II). There was no single threshold criterion along the fitted logistic regression curves that yielded absolute threshold estimates for both the no-noise and noise groups (Table II) that were simultaneously consistent with those based either on angular orientation (Sec. III B 1) or on the raw proportions of 0.50 or 0.20, which we describe next.

Estimates of recognition thresholds based on the raw proportions of females exhibiting correct responses were consistently 30 dB higher in the noise group compared with the no-noise group. The lowest signal levels at which at least 50% of subjects exhibited correct responses were 43 dB (11 of 20 responded) and 73 dB (19 of 20 responded) in the no-noise and noise groups, respectively. These UBs, and their corresponding LBs of 37 and 67 dB, yielded threshold estimates of 41 and 71 dB, respectively. The proportion of females that exhibited correct responses was significantly greater than the expected false alarm rate of 0.20 at signal levels of 67 dB and higher in the noise group, yielding a threshold estimate of 65 dB for this group. In the no-noise group, however, the proportion of females exhibiting correct responses significantly exceeded 0.20 at the lowest signal level of 37 dB and all higher levels. Therefore, to compute a threshold estimate for this group, we assumed that subjects would not have exhibited phonotaxis at the next lowest signal level in series, which would have been 31 dB (i.e., a 6-dB step down from the lowest signal level used). This assumption is reasonable given that Beckers and Schul (2004) reported that 4 of 11 (36.4%) and 5 of 11 (45.5%)

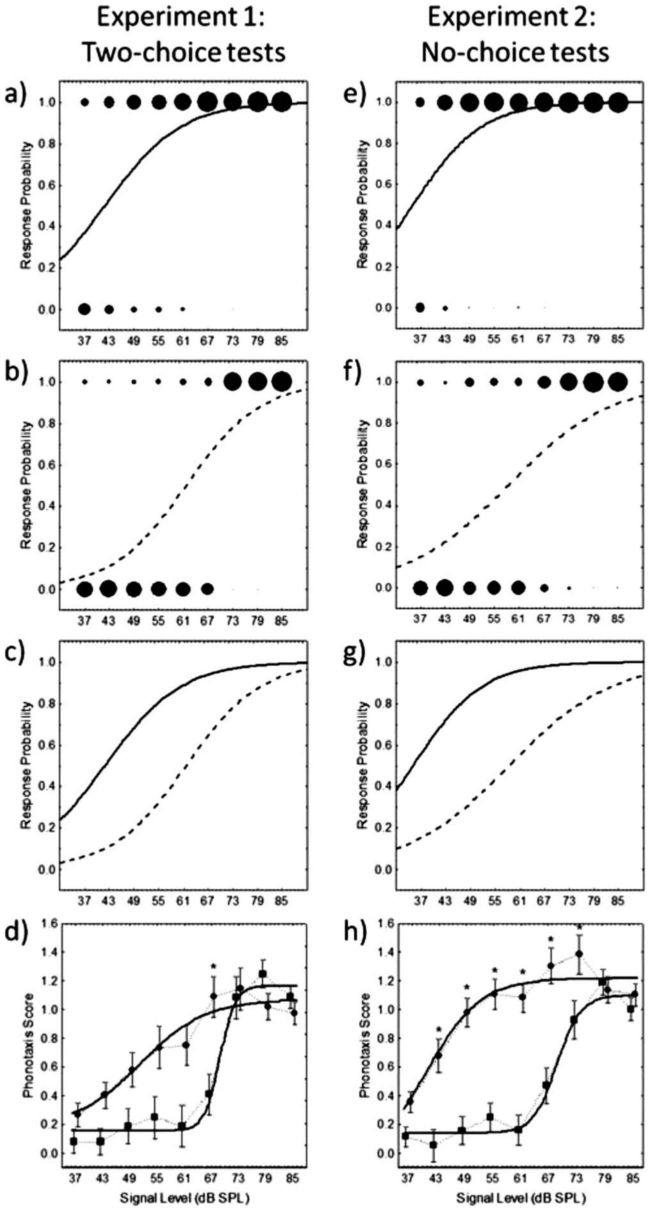

FIG. 2. Response probabilities and phonotaxis scores based on two-choice tests in Experiment 1 (left) and no-choice tests in Experiment 2 (right). [(a) and (e)] Response probabilities in the no-noise groups showing the numbers of individuals (total $N=20$) that exhibited the two types of responses, correct responses (1.0) and incorrect or no responses (0.0). For each response type, the relative sizes of the two paired points at each signal level depict the numbers of individuals exhibiting that response type, with the largest point corresponding to 20 individuals, and the smallest point corresponding to 1 individual (the absence of a point corresponds to zero individuals). The smooth curve represents the best-fit logistic regression function fitted to the response probabilities depicted in the figure. [(b) and (f)] Response probabilities and logistic regression functions (as in panels a and e) for the noise groups. [(c) and (g)] Comparison of the fitted logistic regression functions for the no-noise groups (solid lines) and the noise groups (dashed lines) for each experiment. [(d) and (h)] Mean (±s.e.m.) phonotaxis scores with best-fit sigmoid curves from Eq. (2) for responses in the no-noise groups (circles) and the noise groups (squares) for each experiment. $^{*}P<0.05$ in a Bonferroni *post hoc* test comparing phonotaxis scores in the no-noise and noise groups at each signal level.

TABLE II. Behavioral recognition thresholds in the no-noise and noise groups of Experiment 1 (two-choice tests) as functions of threshold criteria based on fitted response probabilities and phonotaxis scores.

Response variable	Threshold criterion	Estimated threshold signal levels (dB SPL)		Threshold difference (dB)
		No-noise	Noise	
Response probability (p')	0.2	29.2	49.2	20.0
	0.3	34.2	54.1	19.9
	0.4	38.2	58.0	19.8
	0.5	42.0	61.7	19.7
	0.6	45.7	65.3	19.6
	0.7	49.8	69.3	19.5
	0.8	54.8	74.2	19.4
	0.9	62.2	81.5	19.3
Phonotaxis scores (ps')	0.2	23.2	63.6	40.4
	0.3	37.9	65.8	27.9
	0.4	43.4	66.9	23.5
	0.5	47.2	67.7	20.5
	0.6	50.6	68.5	17.9
	0.7	53.8	69.1	15.3
	0.8	57.3	69.8	12.5
	0.9	61.7	70.6	8.9

female *Hyla versicolor* responded to a conspecific call in no-choice tests in the absence of masking noise at signal levels of 31 dB (two-tailed binomial test of $p\geq 0.20$; $P=0.2470$) and 37 dB (two-tailed binomial test of $p\geq 0.20$; $P=0.0504$). At a sample size equivalent to that used in the present study ($N=20$), the proportions reported by Beckers and Schul (2004) would be significantly greater than 0.20 at the 37-dB signal level but not at a level of 31 dB. Therefore, taking 31 and 37 dB as the LB and UB, we estimated a threshold of 35 dB for this group. We acknowledge that this estimate must be accepted with some caution because signal levels below 37 dB were not tested in the present study.

We performed one additional set of analyses based on examining the choices made by those females (out of 20) that touched the wall in front of the speaker broadcasting the target (conspecific) signal or the non-target (heterospecific) signal. In traditional analyses of results from two-choice phonotaxis tests with frogs, the former are typically regarded as a preferential "choice" of the conspecific call over the heterospecific call; the latter are regarded as a choice of the heterospecific call. Considering only those females that made a choice, there was a significant preference for the target (conspecific) signal at all signal levels tested in the no-noise group (Table III). By contrast, a signal level of at least 73 dB was necessary to elicit a significant preference in favor of the target (conspecific) signal in the presence of the chorus-shaped noise. If 37 and 73 dB are taken as UBs of a threshold estimate in the no-noise and noise groups, respectively, and corresponding values of 31 and 67 dB are taken as LBs, then recognition threshold estimates based on measures of behavioral discrimination between conspecific and heterospecific calls would be 35 and 71 dB in the no-noise and noise groups, respectively. The difference between these estimates of recognition threshold (36 dB) based on the responses of females exhibiting choices is therefore greater than those derived from measures of angular orientation (30 dB; Sec. III B 2) and other measures of response probability based on raw or fitted values (≈20–30 dB).

TABLE III. Number (and percentages) of subjects choosing the target (conspecific) signal and the non-target (heterospecific) signal as a function of signal level in the presence or absence of chorus-shaped noise with results from two-tailed binomial tests of the null hypothesis that $p=0.50$.

Noise condition	Signal level	Target signal	Non-target signal	Total	P
No-noise	37	8 (88.9%)	1 (11.1%)	9	0.0391
	43	11 (92.0%)	1 (8.0%)	12	0.0063
	49	14 (100.0%)	0 (0.0%)	14	0.0001
	55	14 (100.0%)	0 (0.0%)	14	0.0001
	61	17 (100.0%)	0 (0.0%)	17	0.0001
	67	20 (100.0%)	0 (0.0%)	20	<0.0001
	73	19 (95.0%)	1 (5.0%)	20	<0.0001
	79	20 (100.0%)	0 (0.0%)	20	<0.0001
	85	20 (100.0%)	0 (0.0%)	20	<0.0001
Noise	37	4 (50.0%)	4 (50.0%)	8	1.0000
	43	3 (37.5%)	5 (62.5%)	8	0.7266
	49	4 (80.0%)	1 (20.0%)	5	0.3750
	55	5 (62.5%)	3 (37.5%)	8	0.7266
	61	6 (85.7%)	1 (14.3%)	7	0.1250
	67	8 (80.0%)	2 (20.0%)	10	0.1094
	73	19 (100.0%)	0 (0.0%)	19	<0.0001
	79	19 (100.0%)	0 (0.0%)	19	<0.0001
	85	20 (100.0%)	0 (0.0%)	20	<0.0001

3. Phonotaxis scores

An ANOVA revealed significant differences in phonotaxis scores [Fig. 2(d)] across the nine signal levels ($F_{8,288}=27.9$, $P<0.0001$), a significant effect of noise condition ($F_{1,36}=6.2$, $P=0.0173$) and a significant interaction between these two effects ($F_{8,288}=4.8$, $P=0.0002$). No other effects in the model were significant, including the main effect of call order ($F_{1,36}=1.7$, $P=0.2068$) and all interaction terms ($0.3653<Ps<0.6463$). Phonotaxis scores were generally higher for the no-noise group compared with the noise group at signal levels of 67 dB and lower, but only significantly so at the 67-dB signal level [Fig. 2(d)]. The computed sigmoid functions relating mean phonotaxis scores to signal level fit the observed data reasonably well for both the no-noise (adjusted $R^2=0.89$) and noise (adjusted $R^2=0.97$) groups, and most fitted values at a particular signal level fell within one standard error of the observed mean. We summarize in Table II a range of threshold estimates for different threshold criteria expressed as phonotaxis scores (ps') along the fitted sigmoid function. Compared with threshold differences between the no-noise and noise groups based on angular orientation and response probabilities, differences based on phonotaxis scores were much more variable, ranging between 8.9 and 40.4 dB (Table II). In addition, there was no single threshold criterion along the fitted sigmoid curves yielding estimates of threshold for both the no-noise and noise groups (Table II) that were simultaneously consistent with those derived from measures of angular orientation (Sec. III B 1) or from using raw or fitted response probabilities (Sec. III B 2).

IV. EXPERIMENT 2: POPULATION-LEVEL RECOGNITION THRESHOLDS IN NO-CHOICE TESTS

In our second experiment, we used a testing protocol that followed that of Experiment 1 (Sec. III) with one notable exception. In Experiment 2, we used a series of "no-choice" tests in which the same target signal used in Experiment 1 was the only signal presented. Hence, there was no alternating non-target signal, and therefore subjects could not choose between two signals. As in Experiment 1, we regard estimates of recognition thresholds as "population-level" estimates because they are based on the collective responses of a pool of subjects.

A. Methods

1. Experimental design

Aside from the lack of a non-target signal, all experimental details, including the 9 signal level (within subjects) ×2 noise condition (between subjects) factorial design, were the same as those described above for Experiment 1 (Sec. III A 1).

2. Data analyses

Our analyses of the data generally follow those outlined for Experiment 1 (Sec. III A 2). We scored a correct response if subjects touched the arena wall in the 15° arc in front of the speaker broadcasting the target signal. We scored a "no response" if subjects failed to meet this response criterion within 5 min, but responded in the next reference condition. We assessed the directedness of phonotaxis (i.e., angular orientation at a distance of 20 cm from the central release point) using circular statistics (V tests), we compared response probabilities using generalized linear models and nonparametric statistics, and we evaluated differences in phonotaxis scores using ANOVA. We computed estimates of recognition thresholds using the same procedures outlined above for Experiment 1 (Sec. III A 2). Of the 46 females tested in Experiment 2, 6 were excluded from the final data set ($N=40$)

TABLE IV. Results of circular statistical analyses for Experiment 2 (no-choice tests).

Noise condition	Signal condition	Mean vector ($\mu°$)	Length of mean vector (r)	Circular SD (deg)	N	V	P
No-noise	Reference 1	−5	0.86	31	20	0.86	<0.0001
	Reference 2	4	0.91	25	20	0.91	<0.0001
	Reference 3	−2	0.95	18	20	0.95	<0.0001
	Reference 4	0	0.94	20	20	0.94	<0.0001
	37 dB	13	0.27	93	19	0.26	0.0550
	43 dB	44	0.45	72	20	0.32	0.0200
	49 dB	−4	0.72	47	19	0.71	<0.0001
	55 dB	3	0.66	52	20	0.66	<0.0001
	61 dB	0	0.65	53	20	0.65	<0.0001
	67 dB	−6	0.74	44	20	0.74	<0.0001
	73 dB	3	0.90	26	20	0.90	<0.0001
	79 dB	−1	0.93	22	20	0.93	<0.0001
	85 dB	−5	0.89	28	19	0.88	<0.0001
Noise	Reference 1	3	0.91	24	20	0.91	<0.0001
	Reference 2	1	0.97	15	20	0.97	<0.0001
	Reference 3	−1	0.93	23	20	0.93	<0.0001
	Reference 4	−2	0.84	34	20	0.84	<0.0001
	37 dB	−99	0.18	106	20	−0.03	0.5690
	43 dB	−177	0.23	98	20	−0.23	0.9250
	49 dB	−32	0.15	111	20	0.13	0.2060
	55 dB	139	0.03	154	20	−0.02	0.5520
	61 dB	96	0.27	93	20	−0.03	0.5750
	67 dB	154	0.09	126	20	−0.08	0.6930
	73 dB	−19	0.50	68	20	0.47	0.0010
	79 dB	5	0.82	36	20	0.82	<0.0001
	85 dB	−4	0.89	28	20	0.88	<0.0001

because they did not meet our inclusion criteria. None of the subjects in this experiment had been tested previously.

B. Results and discussion

Subjects in the no-noise and noise groups remained similarly motivated to respond to the target signal over the entire duration of the test sequence. Orientation toward the target signal was uniformly strong across all four reference conditions (Table IV). The mean (±SD) latency with which individuals met our response criterion was 71.0 ± 20.6 s, averaged across all four reference conditions and both noise conditions. The mean latencies in the first (65.2 ± 18.7 s), second (73.0 ± 22.3), third (75.7 ± 21.4), and fourth (70.1 ± 19.2) reference conditions differed significantly ($F_{3,114}=4.2$, $P=0.0114$); however, there was no difference in latency between the no-noise and noise groups ($F_{1,38}=1.8$, $P=0.1851$), nor was there an interaction between group membership and the sequential order of the reference conditions ($F_{3,114}=2.1$, $P=0.1176$). In addition, a linear contrast comparing latencies across the sequentially ordered reference conditions was not significant ($F_{1,38}=3.1$, $P=0.0880$).

1. Angular orientation

An analysis of orientation angles (Table IV) revealed results strikingly similar to those reported above for two-choice tests in Experiment 1 (see Table I). Subjects oriented toward the target signal at signal levels of 43 dB and higher in the no-noise group (Table IV). In the presence of the chorus-shaped noise, signal levels of 73 dB and higher elicited significant orientation toward the target signal. Taking LBs of 37 and 67 dB, we estimated recognition thresholds of 41 and 71 dB in the no-noise and noise groups, respectively, which corresponds to a threshold differences of 30 dB between these two groups.

2. Response probabilities

The proportion of subjects that met our response criteria increased as a function of increasing signal level in both the no-noise group [Fig. 2(e)] and in the noise group [Fig. 2(f)]. Again, this level-dependent increase in responsiveness started at higher signal levels in the noise group compared with the no-noise group [Fig. 2(g)]. The model parameter for signal level was significantly different from zero ($\chi^2=18.8$, $P<0.0001$, df=1), but those for noise condition ($\chi^2=0.1$, $P=0.7431$, df=1) and the interaction between signal level and noise condition ($\chi^2=2.0$, $P=0.1620$, df=1) were not. Subsequent contrast analyses based on estimates of least-squares means revealed that response proportions in the no-noise group differed significantly from zero ($\chi^2=31.1$, $P<0.0001$, df=1), while those in the noise group did not ($\chi^2=0.41$, $P=0.5210$, df=1). In addition, there were significant differences between the proportions responding in the no-noise and noise groups ($\chi^2=22.9$, $P<0.0001$, df=1). Table V summarizes threshold estimates as a function of different threshold criteria expressed as response probabilities (p') along best-fit logistic regression curves [Fig. 2(g)]. As in

TABLE V. Behavioral recognition thresholds in the no-noise and noise groups of Experiment 2 (no-choice tests) as functions of threshold criteria based on fitted response probabilities and phonotaxis scores.

Response variable	Threshold criterion	Estimated threshold signal levels (dB SPL)		Threshold difference (dB)
		No-noise	Noise	
Response probability (p')	0.2	23.6	41.2	17.6
	0.3	28.1	47.9	19.8
	0.4	31.8	53.4	21.6
	0.5	35.1	58.4	23.3
	0.6	38.5	63.4	24.9
	0.7	42.2	68.9	26.7
	0.8	46.7	75.6	28.9
	0.9	53.4	85.6	32.2
Phonotaxis scores (ps')	0.2	33.5	61.9	28.4
	0.3	35.9	64.7	28.8
	0.4	37.9	66.2	28.3
	0.5	39.7	67.4	27.7
	0.6	41.5	68.5	27.0
	0.7	43.3	69.6	26.3
	0.8	45.2	70.7	25.5
	0.9	47.4	72.1	24.7

Experiment 1, there was no single criterion that produced threshold estimates for both the no-noise and noise groups that were consistent with those derived above based on angular orientations (Sec. IV B 1) and raw response probability (see below, this section). In contrast to Experiment 1, however, there was considerably more variation in estimates of threshold differences, which ranged between 17.7 and 32.3 dB (Table V; cf Table II).

The lowest signals level at which at least 50% of subjects responded was 43 dB (15 of 20 responded) in the no-noise group and 67 dB (12 of 20 responded) in the noise group, with corresponding LBs of 37 and 61 dB. These UBs and LBs yielded threshold estimates of 41 and 65 dB in the no-noise and noise groups, respectively, representing a threshold difference of 24 dB between the two groups. This magnitude of difference is smaller than the 30-dB difference determined in parallel analyses of results from the two-choice tests of Experiment 1 (Sec. III B 2). In the no-choice tests of Experiment 2, the proportion of subjects in the no-noise group that responded was significantly higher than the expected false alarm rate of 0.20 at the 37-dB signal level (at which 10 of 20 subjects responded) and all higher levels [Fig. 2(e)]. We again assumed 37 dB to be an UB and calculated a threshold of 35 dB for this group. Parallel analyses for the noise group yielded an UB and a LB of 67 and 61 dB, respectively, and a threshold estimate of 65 dB. Thus, estimates based on statistically significant differences from the expected false alarm rate yielded the same absolute thresholds, and thus the same threshold difference (30 dB), as in the two-choice tests described earlier (Sec. III B 2).

3. Phonotaxis scores

As in Experiment 1, phonotaxis scores increased as a function of increasing signal level, and this level-dependent increase began at higher signal levels in the noise group compared with the no-noise group [Fig. 2(h)]. There were significant main effects of signal level ($F_{8,304}=26.4$, $P<0.0001$) and noise condition ($F_{1,38}=45.1$, $P<0.0001$) and a significant interaction between these two effects ($F_{8,304}=7.8$, $P<0.0001$). Phonotaxis scores were similar between the two noise conditions at signal levels of 37, 79, and 85 dB, but they were significantly higher in the no-noise group at all other signal levels [Fig. 2(h)]. The fitted sigmoid relationships between mean phonotaxis scores and signal level explained large portions of the variance in both the no-noise (adjusted $R^2=0.86$) and noise (adjusted $R^2=0.96$) groups. Most fitted values at each nominal signal level fell within one standard error of the actual mean phonotaxis score observed at that level. Table V summarizes threshold estimates as a function of different threshold criteria expressed as phonotaxis scores (ps') along the fitted sigmoid functions. There was generally less variation in the threshold differences between the no-noise and noise groups (24–29 dB, Table V) compared with those from parallel analyses of two-choice tests in Experiment 1 (8–41 dB, Table II). Again, however, no single threshold criterion based on fitted phonotaxis scores yielded threshold estimates for both the no-noise and the noise groups that were entirely consistent with those derived above for angular orientation (Sec. IV B 1) and response probabilities (Sec. IV B 2).

V. EXPERIMENT 3: INDIVIDUAL-LEVEL RECOGNITION THRESHOLDS IN NO-CHOICE TESTS

In Experiment 3, we used no-choice tests and an adaptive tracking method of threshold estimation based loosely on the method of limits commonly used to determine thresholds in more traditional psychoacoustic studies (Klump *et al.*, 1995). Each subject was again tested at several signal levels, but the levels chosen for all but the first test were contingent upon the subject's response in the previous test. In this way, we were able to derive threshold estimates for each individual separately.

A. Methods

The target signal and chorus-shaped noises were the same as those described above for Experiment 1 (Sec. III A 1) and Experiment 2 (Sec. IV A 1). We randomly assigned subjects either to a no-noise group ($N=20$) or to a noise group ($N=20$) for which chorus-shaped noise was broadcast continuously during a test from an overhead speaker at a long-term overall SPL of 70 dB (LC_{eq}). None of the subjects in this experiment had been tested previously.

For each individual subject, a test sequence comprised a variable number of reference conditions and treatment conditions that was determined by the subject's responses. Again, we scored a correct response if subjects touched the wall of the test arena in the 15° arc in front of the target speaker in under 5 min. We recorded a no response when subjects failed to meet this criterion. Each sequence began and ended with the reference condition, which again consisted of the target signal presented alone at 85 dB SPL. We also tested the reference condition after any two consecutive treatment conditions failed to elicit correct responses. The first treatment condition in the sequence always involved presenting the target signal at a level estimated by us to be close to the recognition threshold in the noise condition being tested. Our initial estimates were based on results reported by Beckers and Schul (2004) and Bee and Swanson (2007) and were determined *prior to* any analyses of results from Experiments 1 and 2. For the no-noise group, the initial signal level was 45 dB for the first three subjects tested but was subsequently reduced to 39 dB for the remaining 17 subjects. For the group tested with chorus-shaped noise, the initial signal level was 70 dB for all subjects. Following the first treatment condition, we reduced or increased the signal level by 3 dB in the subsequent treatment condition depending on whether the subject did or did not respond in the previous treatment condition, respectively. We continued either decreasing or increasing the signal level in 3-dB steps in subsequent treatment conditions until the subject's behavior changed (e.g., going from correct response to incorrect response between two consecutive treatment conditions). After the subject's behavior changed, we tested a final treatment condition in which we reversed the direction of signal level change by a reduced step-size of 1.5 dB. If the subject responded in this final treatment condition, the signal level for that treatment condition was used as the UB of a threshold estimate, and the next lowest level tested was used as the LB. If the subject failed to respond in the final treatment condition, the signal level in that condition was taken as the LB of the threshold and the next highest signal level previously eliciting a response was taken as the UB. We computed an estimate of the recognition threshold as the average of the UB and LB using Eq. (1) and compared these between groups using a Mann-Whitney U Test.

B. Results and discussion

The mean (±SD) latencies to respond to the target signal in the first (80.7 ± 24.4 s; $N=40$) and last (86.9 ± 29.4 s; $N=40$) reference conditions did not differ significantly ($F_{1,38}=2.3$, $P=0.1359$). There were no differences in response latency between subjects in the no-noise and noise groups ($F_{1,38}=0.4$, $P=0.5532$), nor was there an interaction between noise condition and reference condition ($F_{1,38}=1.1$, $P=0.2965$). The numbers of signal levels at which individuals in the two noise conditions were tested ranged between 3 and 6 levels and did not differ between groups (Mann-Whitney U Test: $U=190$, $P=0.7868$); the median and the modal number of signal levels tested were 3 levels in both noise groups.

The difference between the median thresholds determined for the no-noise and noise groups was 32.5 dB (Fig. 3) and was statistically significant (Mann-Whitney U Test: $U=0.00$, $P<0.0001$). Across subjects tested in the no-noise group, the LBs of threshold estimates ranged between 31.5 and 42.0 dB and the UBs ranged between 33.0 and 43.5 dB. The median threshold for subjects in the no-noise group was 38.3 dB (Fig. 3). In the noise group, the median threshold was 70.8 dB. Across individuals assigned to the noise group, LBs ranged between 62.5 and 76 dB and UBs ranged between 64.0 and 77.5 dB.

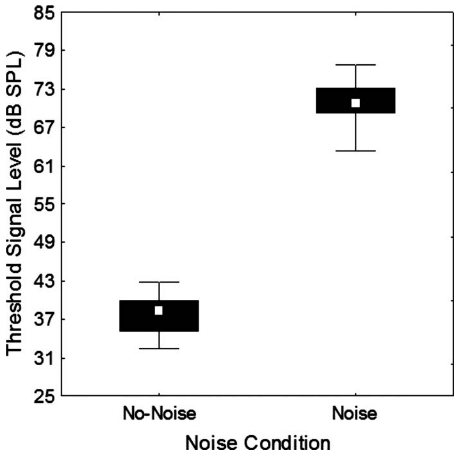

FIG. 3. Individual recognition thresholds based on no-choice tests in the no-noise and noise groups of Experiment 3. Depicted here are the median (point), inter-quartile range (box), and range (whiskers).

VI. GENERAL DISCUSSION

One goal of previous studies of gray treefrogs (Gerhardt, 2001) and indeed many other anurans (Gerhardt and Huber, 2002) has been to quantify the signal properties that elicit recognition of sound patterns as the sexual advertisement signal of an appropriate mate. Two related issues in the literature concern how sound pattern recognition is defined and experimentally measured (e.g., Ryan and Rand, 1993, 2001; Bush *et al.*, 2002) and how the mechanisms for sound pattern recognition operate in the face of constraints posed by noisy environments (Gerhardt and Klump, 1988; Feng and Ratnam, 2000; Feng and Schul, 2007). Comparatively few studies have explicitly investigated sound pattern recognition as a function of signal level under conditions that also included controlled exposure to natural or artificial sources of noise designed to simulate the acoustic environment of a breeding chorus (but see Ehret and Gerhardt, 1980; Gerhardt and Klump, 1988; Schwartz and Gerhardt, 1998; Wollerman, 1999; Schwartz *et al.*, 2001; Wollerman and Wiley, 2002; Bee, 2007, 2008a, 2008b). Our aim in this study was to evaluate several empirical and analytical methods for estimating sound pattern recognition thresholds in frogs using phonotaxis as a behavioral assay.

A. Phonotaxis as a behavioral response measure

In contrast to studies of most other vertebrates (e.g., Fay and Popper, 1999; Dooling *et al.*, 2000), traditional psychoacoustic approaches based on classical or operant conditioning are notoriously difficult or unsuccessful in studies of anuran hearing (but see Elepfandt *et al.*, 2000). While related experimental procedures, such as reflex modification (reviewed in Simmons and Moss, 1995), have met with some success, they have not been widely adopted. Phonotaxis assays remain the most common experimental approach used to address questions about frog hearing and acoustic communication (reviewed in Gerhardt and Huber, 2002).

Among the advantages of phonotaxis as a behavioral assay are that it can be used to exploit the animal's natural behavioral repertoire to address ecologically valid questions. What is more, many frog species reliably exhibit robust pho-

TABLE VI. Comparison of various methods for estimating call recognition thresholds from this experiment and two other studies.

		Threshold (dB SPL)		Threshold difference (dB)
Experiment or published study	Method of threshold estimation	No-noise	Noise	
Experiment 1 – Two-choice tests	Angular orientation at 20 cm: average of lowest signal level yielding significant orientation and next lowest level	41	71	30
	Response probability ($p > 0.2$): average of lowest signal level yielding significant result in a binomial test of $p > 0.2$ and next lowest level	35	65	30
	Response probability ($p > 0.5$): average of lowest signal level yielding $p > 0.5$ and next lowest level	41	71	30
	Choice results: average of lowest signal level yielding significant preference ($p > 0.5$) for conspecific calls over heterospecific calls	35	71	36
	Response probability ($p' = 0.50$): interpolated value using threshold criterion of $p' = 0.50$ from logistic regression equations	42	62	20
	Phonotaxis scores ($ps' = 0.50$): interpolated value using threshold criterion of 0.50 from fitted sigmoid functions	47	68	21
Experiment 2 – No-choice tests	Angular orientation at 20 cm: average of lowest signal level yielding significant orientation and next lowest level	41	71	30
	Response probability ($p > 0.2$): average of lowest signal level yielding significant result in a binomial test of $p > 0.2$ and next lowest level	35	65	30
	Response probability ($p > 0.5$): average of lowest signal level yielding $p > 0.5$ and next lowest level	41	65	24
	Response probability ($p' = 0.50$): interpolated value using threshold criterion of $p' = 0.50$ from logistic regression equations	35	58	23
	Phonotaxis scores ($ps' = 0.50$): interpolated value using threshold criterion of 0.50 from fitted sigmoid functions	40	67	27
Experiment 3 – No-choice tests	Adaptive tracking	38	71	33
Bee and Swanson (2007) – No-choice tests	Response probability ($p' = 0.50$): interpolated value using threshold criterion of $p' = 0.50$ from logistic regression equation	42	68	26
Beckers and Schul (2004) – No-choice tests	Response probability ($p > 0.5$): average of lowest signal level yielding $p > 0.5$ and next lowest level	41	⋯	⋯

notaxis under highly controlled laboratory conditions. Among the disadvantages of phonotaxis as a behavioral assay is that it does not (and cannot) distinguish between signal detection and signal recognition. This follows because a female might fail to exhibit phonotaxis either because she could not detect the sound or because she detected the sound but did not recognize it as the signal of an appropriate mate. Using phonotaxis as a behavioral measure of sound pattern recognition is further complicated by the fact that recognition in some species is not an "all-or-none" phenomenon, but instead may be a continuous function of variation in one or more signal attributes unrelated to signal amplitude (Bush *et al.*, 2002). In addition, phonotaxis behavior cannot be used to generate robust data on difference limens because *perceptual* discrimination between two stimuli may be possible even though individuals exhibit no *behavioral* discrimination. In summary, phonotaxis is a useful tool for investigating "just meaningful differences" (Nelson and Marler, 1990) but cannot by itself provide information on "just noticeable differences." With these caveats in mind, we evaluated several experimental and analytical approaches for deriving estimates of signal recognition thresholds in the presence and absence of masking noise using phonotaxis as a behavioral assay.

B. Comparing behavioral measures of signal recognition thresholds

In studies of humans, the SRT depends on the rate of correct responses (i.e., correctly recognizing a spoken word in masking noise and not simply recognizing that a word was spoken). Therefore, we similarly limited our analyses to correct responses by operationally defining signal recognition as occurring when females exhibited phonotaxis in response to a conspecific advertisement call. Our most striking finding is the generally high degree of agreement among estimates of both absolute thresholds and relative differences derived using different experimental methods and analytical approaches (Table VI). For example, recognition thresholds derived from measures of angular orientation were exactly the same in Experiment 1 (two-choice tests) and Experiment 2 (no-choice tests). Likewise, threshold estimates based on significant differences from an expected false alarm rate ($p = 0.20$) were identical between Experiments 1 and 2 and were within one step-size in signal level (6 dB) of those based on angular orientation for both noise conditions. Threshold estimates based on the signal levels at which the raw proportion of females exhibiting correct responses exceeded 50% were identical for the no-noise groups of Experiments 1 and 2, and they were within one signal-level step-size (6 dB) of each other in the noise groups of these two experiments. Threshold estimates based on the probability of choosing a conspecific over a heterospecific call were also similar to these other estimates. Perhaps most importantly, all estimates of absolute thresholds in Experiments 1 and 2 that were based on angular orientation or raw response probabilities were within 0–6 dB of those derived in Experiment 3 using smaller step-sizes (1.5 and 3 dB) and an entirely different approach based on adaptive tracking. Moreover, estimates of absolute recognition thresholds are in generally good agreement with those reported in previously published studies of gray treefrogs (Table VI). With only one exception (Table VI), the magnitudes of threshold difference between the two noise conditions based on angular orientation and raw probabilities in Experiments 1 and 2 were within 3 dB of that derived in Experiment 3. In general, threshold estimates based on fitted response probabilities (logistic functions) and phonotaxis scores (sigmoid functions) were similar between Experiments 1 and 2, but they were also much more variable and tended to yield smaller threshold differences (≈20–27 dB; Table VI) between the two noise conditions compared to our other estimates.

Given the general similarity among the results from Experiments 1–3, it is worth considering practical and logistical differences between them. The adaptive tracking procedure we used in Experiment 3 had a number of advantages over the methods we used for Experiments 1 and 2. First, by allowing us to estimate a threshold for each individual, our approach in Experiment 3 allowed us to generate measures of central tendency and variability for each noise condition. These measures, in turn, allowed us to make a straightforward between-groups statistical comparison of recognition thresholds in the two noise conditions. Such a direct comparison was not possible in Experiments 1 and 2. Second, our estimates of threshold in Experiment 3 were less dependent on sample size. Some of the upper and lower threshold bounds in Experiments 1 and 2 depended on cutoffs based on level-dependent patterns of statistical significance, which would vary as a function of statistical power, and hence sample size, for any given effect size. Third, we were able to test smaller gradations in signal level in Experiment 3 by using smaller step-sizes (e.g., 1.5–3 dB) compared with those in Experiments 1 and 2 (6 dB), which, in turn, might provide for better accuracy and precision in estimates of recognition thresholds. Fourth, our approach in Experiment 3 required less time and fewer tests of each subject (e.g., three signal levels) compared with our approach in Experiments 1 and 2, in which each subject was tested at all possible signal levels (e.g., nine signal levels). Finally, our approach in Experiment 3 required relatively fewer subjects; no females failed to complete Experiment 3, whereas 13% (12 of 92) of the females tested in Experiments 1 and 2 failed to complete the whole series of tests. The relative advantages of Experiment 3 over Experiments 1 and 2 in terms of shorter testing times and smaller subject pools might be diminished if the approach used in Experiments 1 and 2 were modified so that a test ended as soon as the subject advanced 20 cm from the release point. Our results suggest that such a modified method would yield results similar to the adaptive tracking procedure of Experiment 3.

VII. CONCLUSION

Efforts to understand the mechanisms by which humans understand speech in noisy social settings hold a central place in modern hearing research. One prominent research methodology involves estimating SRTs in various masking conditions to understand how the spectral, temporal, and spatial relationships between sources of signals and sources of masking noise influence speech perception. Similar experimental approaches have not been widely adopted in studies aimed at discovering how various nonhuman animals have evolved to solve similar "cocktail-party-like" communication problems. Given the exceptional value of anuran amphibians as model systems for studying the mechanisms of acoustic communication in noisy environments, this study aimed to compare estimates of signal recognition thresholds using several common experimental methods and analytical tools. Our results reveal insights into how phonotaxis experiments might best be used to answer questions concerning how frogs recognize behaviorally relevant sound patterns in high levels of biologically realistic background noise. The main conclusion of this study is that the methods and analyses compared here yielded generally quite similar results; however, they differed in a number of practical ways that will be important to consider in designing future experiments.

ACKNOWLEDGMENTS

We extend special thanks to D. Heil, J. Henderson, J. Henly, M. Kuczynski, J. Lane, A. Leightner, E. Love, K. Riemersma, D. Rittenhouse, K. Speirs, S. Tekmen, A. Thompson, A. Vélez, and J. Walker-Jansen for their assistance in collecting and testing frogs and especially to S. Tekmen for her extensive help compiling data and A. Vélez for assistance with stimulus generation. This work was conducted under Special Use Permit No. 14902 from the Minnesota Department of Natural Resources and with a Special Use Permit issued by M. Linck at the Three Rivers Park District. This research was approved by the University of Minnesota Institutional Animal Care and Use Committee (Protocol No. 0809A46721) and was funded by a Grant-in-Aid of Research from the University of Minnesota Graduate School and NIH Grant No. R03-DC008396 to M.B. and NSF Grant No. IBN-0342183 to J.S.

Aubin, T., and Jouventin, P. (**2002**). "How to vocally identify kin in a crowd: The penguin model," Adv. Study Behav. **31**, 243–277.

Beckers, O. M., and Schul, J. (**2004**). "Phonotaxis in *Hyla versicolor* (Anura, Hylidae): The effect of absolute call amplitude," J. Comp. Physiol. [A] **190**, 869–876.

Bee, M. A. (**2007**). "Sound source segregation in grey treefrogs: Spatial release from masking by the sound of a chorus," Anim. Behav. **74**, 549–558.

Bee, M. A. (**2008a**). "Finding a mate at a cocktail party: Spatial release from masking improves acoustic mate recognition in grey treefrogs," Anim.

Behav. **75**, 1781–1791.

Bee, M. A. (**2008b**). "Parallel female preferences for call duration in a diploid ancestor of an allotetraploid treefrog," Anim. Behav. **76**, 845–853.

Bee, M. A., and Micheyl, C. (**2008**). "The cocktail party problem: What is it? How can it be solved? And why should animal behaviorists study it?," J. Comp. Psychol. **122**, 235–251.

Bee, M. A., and Swanson, E. M. (**2007**). "Auditory masking of anuran advertisement calls by road traffic noise," Anim. Behav. **74**, 1765–1776.

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

Cherry, E. C. (**1953**). "Some experiments on the recognition of speech, with one and with two ears," J. Acoust. Soc. Am. **25**, 975–979.

Dooling, R. J., Fay, R. R., and Popper, A. N. (**2000**). *Comparative Hearing: Birds and Reptiles* (Springer, New York).

Ehret, G., and Gerhardt, H. C. (**1980**). "Auditory masking and effects of noise on responses of the green treefrog (*Hyla cinerea*) to synthetic mating calls," J. Comp. Physiol. [A] **141**, 13–18.

Elepfandt, A., Eistetter, I., Fleig, A., Gunther, E., Hainich, M., Hepperle, S., and Traub, B. (**2000**). "Hearing threshold and frequency discrimination in the purely aquatic frog *Xenopus laevis* (pipidae): Measurement by means of conditioning," J. Exp. Biol. **203**, 3621–3629.

Fay, R. R., and Popper, A. N. (**1999**). *Comparative Hearing: Fish and Amphibians* (Springer, New York).

Feng, A. S., and Ratnam, R. (**2000**). "Neural basis of hearing in real-world situations," Annu. Rev. Psychol. **51**, 699–725.

Feng, A. S., and Schul, J. (**2007**). "Sound processing in real-world environments," in *Hearing and Sound Communication in Amphibians*, edited by P. A. Narins, A. S. Feng, R. R. Fay, and A. N. Popper (Springer, New York), pp. 323–350.

Gerhardt, H. C. (**1975**). "Sound pressure levels and radiation patterns of vocalizations of some North American frogs and toads," J. Comp. Physiol. **102**, 1–12.

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

Gerhardt, H. C. (**2001**). "Acoustic communication in two groups of closely related treefrogs," Adv. Stud. Behav. **30**, 99–167.

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

Gerhardt, H. C., and Klump, G. M. (**1988**). "Masking of acoustic signals by the chorus background noise in the green treefrog: A limitation on mate choice," Anim. Behav. **36**, 1247–1249.

Gerhardt, H. C., Ptacek, M. B., Barnett, L., and Torke, K. G. (**1994**). "Hybridization in the diploid-tetraploid treefrogs *Hyla chrysoscelis* and *Hyla versicolor*," Copeia **1994**, 51–59.

Greenhouse, S. W., and Geisser, S. (**1959**). "On methods in the analysis of profile data," Psychometrika **24**, 95–112.

Holloway, A. K., Cannatella, D. C., Gerhardt, H. C., and Hillis, D. M. (**2006**). "Polyploids with different origins and ancestors form a single sexual polyploid species," Am. Nat. **167**, E88–E101.

Horton, N. J., and Lipsitz, S. R. (**1999**). "Review of software to fit generalized estimating equation regression models," Am. Stat. **53**, 160–169.

Klump, G. M., Dooling, R. J., Fay, R. R., and Stebbins, W. C. (**1995**). *Methods in Comparative Psychoacoustics* (Birkhäuser, Basel).

Narins, P. M., and Zelick, R. (**1988**). "The effects of noise on auditory processing and behavior in amphibians," in *The Evolution of the Amphibian Auditory System*, edited by B. Fritzsch, M. J. Ryan, W. Wilczynski, T. E. Hetherington, and W. Walkowiak (Wiley, New York), pp. 511–536.

Nelson, D. A., and Marler, P. (**1990**). "The perception of birdsong and an ecological concept of signal space," in *Comparative Perception*, edited by M. A. Berkley and W. C. Stebbins (Wiley, New York), Vol. **II**, pp. 443–478.

Plomp, R. (**1978**). "Auditory handicap of hearing impairment and limited benefit of hearing aids," J. Acoust. Soc. Am. **63**, 533–549.

Plomp, R., and Mimpen, A. M. (**1979a**). "Improving the reliability of testing the speech reception threshold for sentences," Audiology **18**, 43–52.

Plomp, R., and Mimpen, A. M. (**1979b**). "Speech reception threshold for sentences as a function of age and noise level," J. Acoust. Soc. Am. **66**, 1333–1342.

Ryan, M. J., and Rand, A. S. (**1993**). "Species recognition and sexual selection as a unitary problem in animal communication," Evolution **47**, 647–657.

Ryan, M. J., and Rand, A. S. (**2001**). "Feature weighting in signal recognition and discrimination by túngara frogs," in *Anuran Communication*, edited by M. J. Ryan (Smithsonian Institution, Washington, DC), pp. 86–101.

Schul, J., and Bush, S. L. (**2002**). "Non-parallel coevolution of sender and receiver in the acoustic communication system of treefrogs," Proc. R. Soc., London, Ser. B **269**, 1847–1852.

Schwartz, J. J., Buchanan, B. W., and Gerhardt, H. C. (**2001**). "Female mate choice in the gray treefrog (*Hyla versicolor*) in three experimental environments," Behav. Ecol. Sociobiol. **49**, 443–455.

Schwartz, J. J., and Freeberg, T. M. (**2008**). "Acoustic interaction in animal groups: Signaling in noisy and social contexts—Introduction," J. Comp. Psychol. **122**, 231–234.

Schwartz, J. J., and Gerhardt, H. C. (**1998**). "The neuroethology of frequency preferences in the spring peeper," Anim. Behav. **56**, 55–69.

Simmons, A. M. and Moss, C. F. (**1995**). "Reflex modification: A tool for assessing basic auditory function in anuran amphibians," in *Methods in Comparative Psychoacoustics*, edited by G. M. Klump, R. J. Dooling, R. R. Fay, and W. C. Stebbins (Birkhäuser, Basel), pp. 197–208.

Swanson, E. M., Tekmen, S. M., and Bee, M. A. (**2007**). "Do female anurans exploit inadvertent social information to locate breeding aggregations?" Can. J. Zool. **85**, 921–932.

Wollerman, L. (**1999**). "Acoustic interference limits call detection in a Neotropical frog *Hyla ebraccata*," Anim. Behav. **57**, 529–536.

Wollerman, L., and Wiley, R. H. (**2002**). "Background noise from a natural chorus alters female discrimination of male calls in a Neotropical frog," Anim. Behav. **63**, 15–22.

Zar, J. H. (**1999**). *Biostatistical Analysis* (Prentice-Hall, Upper Saddle River, NJ).