Report

Current Biology

Frogs Exploit Statistical Regularities in Noisy Acoustic Scenes to Solve Cocktail-Party-like Problems

Highlights

- Noise induces communication errors that interfere with mating decisions in frogs
- The noise generated in frog breeding choruses exhibits statistical regularities
- Female frogs exploit these regularities to reduce costly errors in mate selection
- Exploiting statistical regularities helps frogs solve a "cocktail party problem"

Authors

Norman Lee, Jessica L. Ward, Alejandro Vélez, Christophe Micheyl, Mark A. Bee

Correspondence

lee33@stolaf.edu

In Brief

Lee et al. integrate modeling and experiments to discover that animals exploit statistical regularities in noisy acoustic scenes to reduce communication errors. They show that the intense noise generated in choruses of calling male frogs has statistical properties that can be exploited by females for improved signal recognition and discrimination.

Lee et al., 2017, Current Biology *27*, 743–750
March 6, 2017
http://dx.doi.org/10.1016/j.cub.2017.01.031

CellPress

Current Biology
Report

Frogs Exploit Statistical Regularities in Noisy Acoustic Scenes to Solve Cocktail-Party-like Problems

Norman Lee,[1,7,8,*] Jessica L. Ward,[1,2] Alejandro Vélez,[1,5] Christophe Micheyl,[3,6] and Mark A. Bee[1,4]
[1]Department of Ecology, Evolution, and Behavior, University of Minnesota, Saint Paul, MN 55108, USA
[2]Department of Fisheries, Wildlife, and Conservation Biology, University of Minnesota, Saint Paul, MN 55108, USA
[3]Department of Psychology, University of Minnesota, Minneapolis, MN 55455, USA
[4]Graduate Program in Neuroscience, University of Minnesota, Minneapolis, MN 55455, USA
[5]Present address: Department of Biology, San Francisco State University, San Francisco, CA 94132, USA
[6]Present address: Starkey Hearing Research Center, Berkeley, CA 94704, USA
[7]Present address: Biology Department, St. Olaf College, Northfield, MN 55057, USA
[8]Lead Contact
*Correspondence: lee33@stolaf.edu
http://dx.doi.org/10.1016/j.cub.2017.01.031

SUMMARY

Noise is a ubiquitous source of errors in all forms of communication [1]. Noise-induced errors in speech communication, for example, make it difficult for humans to converse in noisy social settings, a challenge aptly named the "cocktail party problem" [2]. Many nonhuman animals also communicate acoustically in noisy social groups and thus face biologically analogous problems [3]. However, we know little about how the perceptual systems of receivers are evolutionarily adapted to avoid the costs of noise-induced errors in communication. In this study of Cope's gray treefrog (*Hyla chrysoscelis*; Hylidae), we investigated whether receivers exploit a potential statistical regularity present in noisy acoustic scenes to reduce errors in signal recognition and discrimination. We developed an anatomical/physiological model of the peripheral auditory system to show that temporal correlation in amplitude fluctuations across the frequency spectrum ("comodulation") [4–6] is a feature of the noise generated by large breeding choruses of sexually advertising males. In four psychophysical experiments, we investigated whether females exploit comodulation in background noise to mitigate noise-induced errors in evolutionarily critical mate-choice decisions. Subjects experienced fewer errors in recognizing conspecific calls and in selecting the calls of high-quality mates in the presence of simulated chorus noise that was comodulated. These data show unequivocally, and for the first time, that exploiting statistical regularities present in noisy acoustic scenes is an important biological strategy for solving cocktail-party-like problems in nonhuman animal communication.

RESULTS

In biological systems, noise-induced errors can impose dire fitness consequences for signalers and receivers [7, 8]. Such errors select for the optimization of signal structures [9, 10], signaling strategies [11, 12], and the sensory, perceptual, and cognitive mechanisms for processing signals [9, 10]. The potential for errors in communication is greatest when receivers must respond to signals produced in a complex milieu of competing signals having similar physical properties. Such mixtures of signals constitute significant sources of noise for many animals. The raucous acoustic scenes associated with large groups of conspecifics, such as a human cocktail party [2], a communal songbird roost [6], or choruses of insects [13] and frogs [14], represent social environments where perceptual adaptations for coping with such noise would be particularly advantageous. An emerging view in sensory ecology is that auditory systems, much like visual systems [15, 16], are evolutionarily optimized to process statistical regularities present in natural scenes [4, 17–19]. One statistical regularity of many natural acoustic scenes, like those characteristic of noisy social gatherings, derives from the physical properties of natural sounds, which exhibit slow fluctuations in amplitude through time [19, 20]. In many instances, these fluctuations are correlated across the frequency spectrum (i.e., comodulated) [4–6]. The extent to which comodulation in noisy acoustics scenes is exploited to solve complex communication problems has been a contentious issue in studies of human speech communication [21–23] but remains largely untested in other animals [6, 24].

Natural Statistics of a Noisy Acoustic Scene

Males of Cope's gray treefrog form dense choruses in which they produce loud, pulsatile advertisement calls to attract females (Figure 1). Individual calls (Figure 1A; Audio S1) are produced at high sound pressure levels (SPLs) reaching 85 to 90 dB at 1 m [25], and sustained noise levels in choruses commonly range between 70 and 80 dB SPL. Chorus noise has a frequency spectrum matching that of the call (Figure 1B; Audio S2) and exhibits slow fluctuations in amplitude [26]. Previous studies of auditory

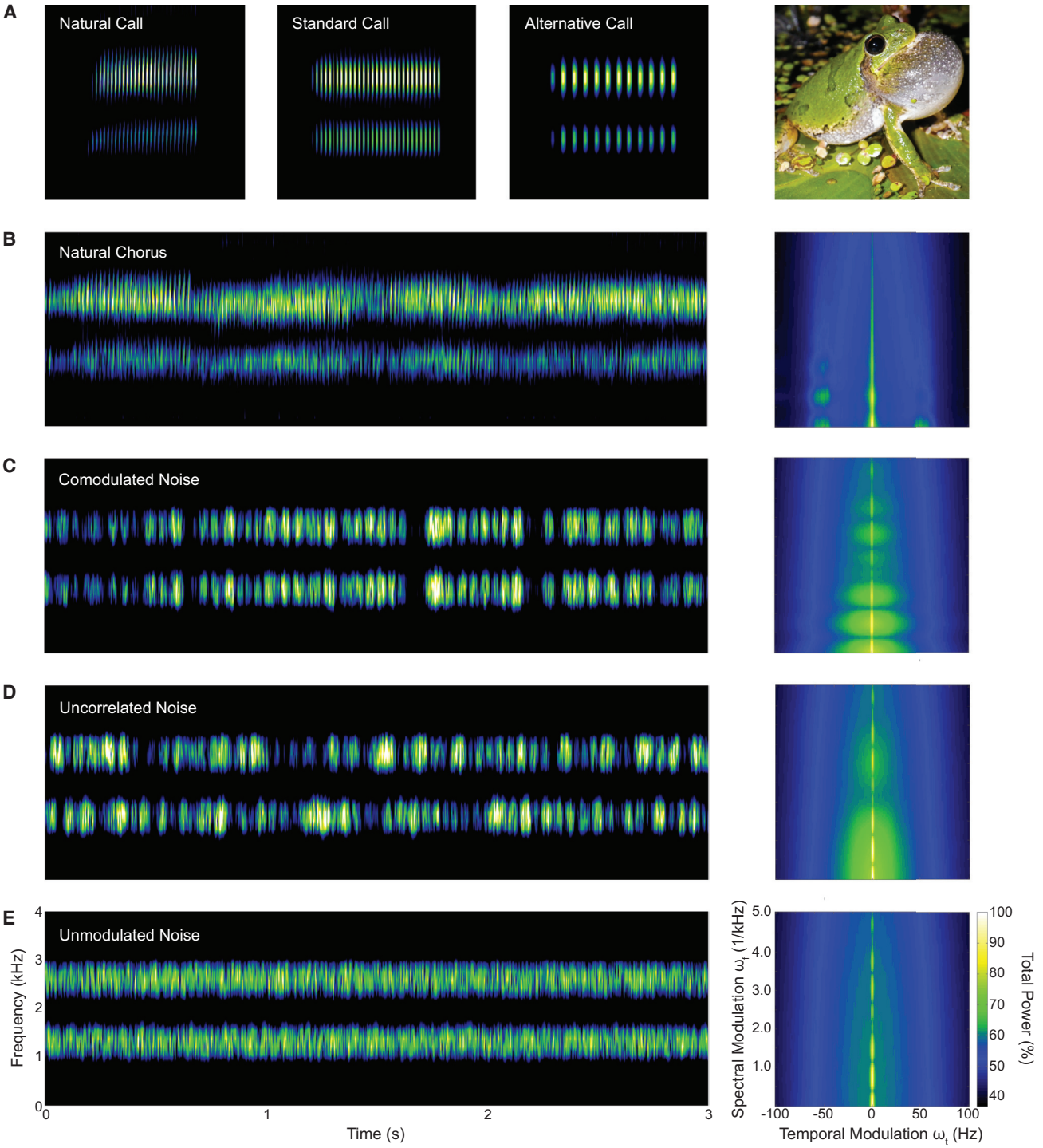

Figure 1. Depictions of Natural and Synthetic Signals and Noise

(A) Spectrograms of a natural advertisement call (left; Audio S1), the synthetic standard call (center), and one of several different synthetic alternative calls (right) used in this study. The depiction of a natural call illustrates two acoustic properties mimicked by synthetic stimulus calls: this signal's pulsatile structure and bimodal frequency spectrum, with spectral peaks near 1.3 and 2.6 kHz. The synthetic standard call was used as a stimulus in experiments 1–4. The alternative call depicted here was used in two-alternative choice tests in experiment 3 and differs from the standard call (50 pulses/s) in having a slower pulse rate (20 pulses/s). Oscillograms of standard and alternative calls used to create differences in call effort in experiment 4 are illustrated in Figure S1A. Also shown here (far right) is a photograph of a calling male of Cope's gray treefrog, used with permission from J.C. Tanner.

(B) An illustrative spectrogram (left) and the mean modulation power spectrum (right) of the natural noise generated by choruses of Cope's gray treefrog (Audio S2). The spectrogram of the natural chorus illustrates the two spectral bands of background noise in choruses arising from the mixture of vocalizations produced

(legend continued on next page)

masking in hylid treefrogs have shown that chorus noise reduces signal active space [27–30], impairs species discrimination [31, 32], interferes with sound localization [33], and constrains choices of preferred mates [29, 34]. Thus, females face an evolutionarily significant "cocktail-party-like problem" in choruses because the large number of signalers in a chorus creates noisy listening conditions that induce costly communication errors.

We tested the hypothesis that chorus noise is comodulated across frequencies of biological relevance to the frogs themselves. To do so, we passed acoustic recordings of Cope's gray treefrog choruses through an anatomical/physiological model of the species' peripheral auditory system and quantified the degree of comodulation across the frequency spectrum (see Supplemental Experimental Procedures). Acoustic recordings were made from within choruses on nights and at times of high calling activity during the species's breeding season. Our model implemented a bank of gammatone filters (Figure 2) to simulate spectral processing by the two sensory papillae in the frog inner ear that are sensitive to airborne sounds [36]. The tonotopically organized amphibian papilla was modeled as six adjacent filters with center frequencies between 238 Hz and 1.3 kHz. The basilar papilla was modeled as a single filter centered at 2.6 kHz. Thus, two of the filters (1.3 and 2.6 kHz) were centered on the spectral peaks emphasized in conspecific calls (Figure 1A). Filter bandwidths were determined from a meta-analysis of published tuning curves measured electrophysiologically from frog auditory nerve fibers (Figure 2A; see Supplemental Experimental Procedures). Adjacent filters simulating the amphibian papilla were spaced to overlap at frequencies 3 dB above threshold. The relative gain of the center frequency of each filter was adjusted to match the corresponding frequency from the species' audiogram [37]. To simulate the half-wave rectification reflected in the coding of amplitude-modulated sounds in the auditory nerve [38], we analytically determined the Hilbert envelope of the output of each frequency filter. Each Hilbert envelope thus preserved the temporal modulations in amplitude present only in the corresponding range of acoustic frequencies passing through each specified filter.

Comodulation was evaluated by computing the cross-covariance between Hilbert envelopes across all pairwise combinations of frequency filters (Figures 2 and S2A). This cross-covariance procedure allowed us to assess the extent to which sound amplitude in different regions of the frequency spectrum varied together through time (i.e., increasing and decreasing together on a moment-to-moment basis). Consistent with our hypothesis, mean cross-covariance values always significantly exceeded null expectations and were highest for frequencies emphasized in conspecific calls (Figure S2B). The mean value comparing output from the two filters centered on 1.3 and 2.6 kHz was 26.7 standard deviations greater than expected by chance (Figures 2B and S2). Our anatomically and physiologically inspired analyses of natural chorus noise, therefore, confirmed that comodulation is a prominent statistical regularity present in the acoustic scenes of Cope's gray treefrog choruses.

We next tested the hypothesis that receivers exploit comodulation in background noise to improve listening performance in ecologically relevant communication tasks. In four psychophysical experiments, we evaluated female mating decisions by quantifying phonotaxis in response to synthetic advertisement calls (Figure 1A) in the presence of artificial "chorus-shaped" noises (Figures 1C–1E). Each experimental noise was constructed by adding two narrow-band noises centered on the two spectral peaks present in Cope's gray treefrog calls (1.3 and 2.6 kHz). The temporal envelopes of the two noise bands were manipulated so that they were (1) comodulated (Figure 1C; Audio S3), (2) modulated but uncorrelated (Figure 1D; Audio S4), or (3) unmodulated (Figure 1E; Audio S5) (see Supplemental Experimental Procedures). A control experiment confirmed that these noises were behaviorally neutral and did not, by themselves, influence phonotaxis (Figure S3). If comodulation in ambient background noise can be exploited to mitigate noise-induced communication errors, we expected to find relatively better performance in comodulated conditions.

Signal Recognition in Comodulated Noise

In two experiments we estimated "signal recognition thresholds" (SRTs) [30], which are conceptually analogous to the "speech reception threshold" measured in studies of masked speech perception in humans [22]. Compared with quiet, the presence of all three experimental noises introduced errors that were functionally equivalent to missed mating opportunities. That is, in the presence of noise, subjects failed to respond to an attractive signal (the "standard call" in Figure 1A) when it was presented at sound levels that were nevertheless sufficiently high to elicit phonotaxis in quiet. Consistent with our hypothesis, however, subjects responded to signals at lower thresholds in comodulated noise compared to other noise conditions.

In experiment 1 (Figures 3A and S1B), we presented signals at each of five signal-to-noise ratios (SNRs; −12, −6, 0, +6, and +12 dB). The proportion of subjects responding varied significantly as a function of SNR ($\chi^2 = 93.8$, degrees of freedom [df] = 1, $p < 0.001$) and noise condition ($\chi^2 = 7.0$, df = 2, $p = 0.031$) and was significantly higher in the comodulated condition compared with both the uncorrelated ($p = 0.040$) and unmodulated ($p = 0.011$) conditions. We determined SRTs as the lowest signal level at which fitted response proportions exceeded 0.5. In the comodulated condition, SRTs were 2.7 dB and 3.7 dB lower than in the uncorrelated and unmodulated conditions, respectively (Figure 3A).

by calling males. The mean modulation power spectrum [20] illustrates the prominence of temporal fluctuations in amplitude (x axis) occurring at slow rates (e.g., <5–10 Hz). The mean depicted here was determined from an ensemble of 26 chorus recordings (see Supplemental Experimental Procedures). Each recording was truncated into 90 1-s segments, and a Gaussian spectrogram (Gaussian window bandwidth: 32 Hz; window size: 1,316) was computed for each segment. A 2D FFT was computed for each Gaussian spectrogram, and real values were averaged across all segments from all recordings to give an average modulation power spectrum.

(C–E) Spectrograms (left) and modulation power spectra (right) of the three artificial chorus-shaped noises used in experiments 1–4: the comodulated noise (C) (Audio S3), the uncorrelated noise (D) (Audio S4), and the unmodulated noise (E) (Audio S5). During each behavioral test of a subject, a specified noise was broadcast continuously to simulate the ambient background noise of a chorus while one or more specified signals were broadcast periodically to simulate individual calling males.

Additional details on the speaker arrangements used in behavioral experiments are provided in Figure S1B and in the Supplemental Experimental Procedures.

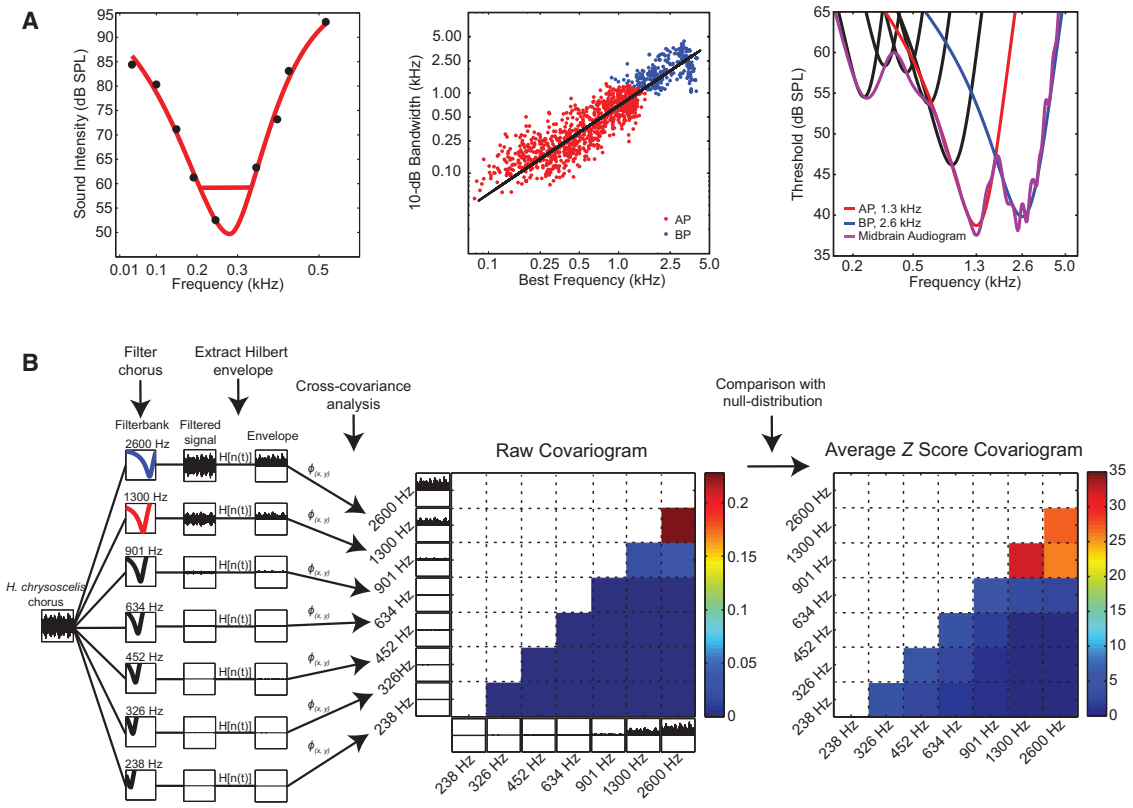

Figure 2. Biologically Inspired Analyses of Chorus Noise Reveal Significant Comodulation

An anatomical/physiological model was used to determine the degree of comodulation present in natural chorus sounds.

(A) The model consisted of a bank of auditory filters fitted to the audiogram of Cope's gray treefrog. Left: each filter was modeled using parameters from VIIIth nerve frequency tuning curves (FTCs) measured in previous studies of frogs (see Supplemental Experimental Procedures). An example of a previously published VIIIth nerve FTC from a frog [35] showing the rounded-exponential function (red curve) used to determine its best frequency (BF), threshold, and bandwidth (BW) 10 dB above threshold (10-dB BW). Center: scatterplot showing the positive relationship between 10-dB BW and BF obtained from a meta-analysis of 1,071 FTCs from seven species of frogs across ten different published studies (see Supplemental Experimental Procedures). Units are classified as innervating either the amphibian papilla (shown in red) or the basilar papilla (shown in blue). Right: diagram showing the model auditory filterbank, with the gain of each filter adjusted to the sensitivity of the midbrain audiogram (purple curve). Filters centered on the 1.3 and 2.6 kHz peaks of the advertisement call are shown in red and blue, respectively.

(B) Cross-covariance analyses were conducted to quantify the magnitude of comodulation in chorus noise. Left: chorus recordings (n = 26 choruses) of 1.5-min duration were filtered using the model filterbank depicted in the right panel of (A). Center: pairwise comparisons between the Hilbert envelope of the output of each frequency filter were made using cross-covariance, as illustrated here by the raw covariogram for a representative chorus recording. Below the diagonal in the raw covariogram shows peak cross-covariance magnitudes for different envelope comparisons plotted as a heatmap. Right: the mean *Z* score covariogram depicts the mean cross-covariance values, averaged across all 26 chorus recordings, as *Z* scores relative to null distributions based on comparing the envelopes of frequency filter outputs across different choruses (Figure S2; Supplemental Experimental Procedures). Colors indicate the number of standard deviations beyond the mean of the null distribution. The high degree of comodulation revealed by these analyses could not be explained as merely resulting from overlap between adjacent auditory filters in the model (Supplemental Experimental Procedures).

In experiment 2 (Figures 3B and S1B), we estimated SRTs using an adaptive tracking procedure to determine the lowest SNR that reliably elicited phonotaxis. SRTs varied significantly across the three noise conditions ($F_{2,57}$ = 23.9, p < 0.001) and were elevated compared to thresholds measured in quiet. The mean threshold was significantly lower in the comodulated condition compared with both the uncorrelated and unmodulated conditions (Figure 3B). In the comodulated condition, subjects experienced, on average (±SEM), 2.6 ± 1.0 dB and 6.9 ± 4.0 dB of masking release compared with the uncorrelated and unmodulated conditions, respectively. Thresholds were also significantly lower in the uncorrelated condition compared with the unmodulated condition (Figure 3B).

Signal Discrimination in Comodulated Noise

Two additional experiments investigated signal discrimination in the contexts of species discrimination (experiment 3) and sexual selection (experiment 4) using two-alternative choice tests. The presence of noise introduced errors in discrimination performance compared with quiet. Consistent with our hypothesis, however, subjects made fewer discrimination errors in comodulated noise.

In experiment 3, subjects chose between the standard call and an alternative call differing in pulse rate (Figure 1A), which is the primary cue females use to discriminate between conspecific males (40–65 pulses/s) and males of a morphologically indistinguishable sister species, *Hyla versicolor* (17–35 pulses/s) [39]. In quiet, subjects preferentially selected calls with conspecific

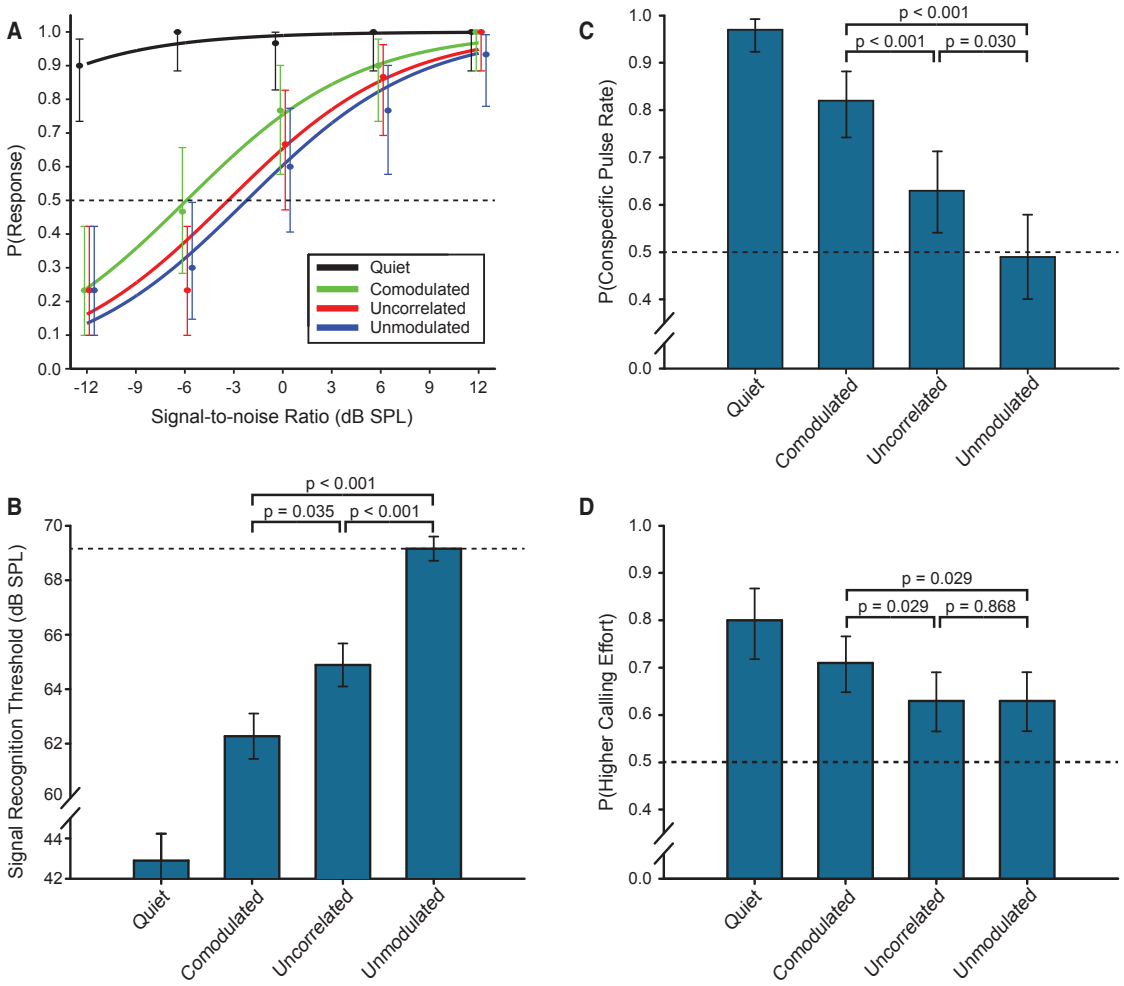

Figure 3. Comodulated Noise Improves Performance in Several Key Communication Tasks Relative to Uncorrelated Noise and Unmodulated Noise

The three artificial chorus-shaped noises used in experiments 1–4 were behaviorally neutral and did not, by themselves, influence phonotaxis (Figure S3).

(A and B) Experiments 1 (A) and 2 (B) consisted of single-stimulus (no-choice) tests and revealed lower signal recognition thresholds (SRTs) in comodulated noise.

(A) Points depict the proportion (±95% exact binomial confidence intervals) of subjects responding at each of five signal-to-noise ratios (−12, −6, 0, +6, and +12 dB, or equivalent signal levels in quiet); solid lines represent fitted functions from generalized estimating equations. The horizontal dashed line represents the criterion (0.5) for determining SRTs.

(B) Bars depict the mean (±SEM) SRTs determined using an adaptive tracking procedure. The horizontal dashed line in (B) indicates the level of performance relative to the condition with the highest threshold.

(C and D) Experiments 3 (C) and 4 (D) consisted of two-alternative choice tests and revealed better discrimination of sound patterns in comodulated noise.

(C) Bars depict the proportion (±95% exact binomial confidence intervals) of subjects choosing stimuli with conspecific pulse rates (P(Conspecific Pulse Rate)). Horizontal dashed line in (C) depicts the level of performance expected by chance (0.5) in a two-alternative choice test.

(D) Bars depict the proportions (±95% exact binomial confidence intervals) of subjects choosing stimuli with relatively higher calling efforts (P(Higher Calling Effort)). Horizontal dashed line in (D) depicts the level of performance expected by chance (0.5) in a two-alternative choice test.

pulse rates (all $p < 0.001$; Figure 3C). Compared with quiet, noise-induced errors in mating decisions were reflected in a reduction in the proportion of subjects selecting calls with conspecific pulse rates. However, error rates differed significantly across noise conditions ($\chi^2 = 26.2$, df = 2, $p < 0.001$). Subjects were significantly more likely to correctly select a conspecific pulse rate in the comodulated condition compared with both the uncorrelated and unmodulated conditions (Figure 3C). They were also better at doing so in the uncorrelated condition compared with the unmodulated condition.

In experiment 4, subjects chose between the standard call (Figure 1A) and an alternative with either a higher or a lower "calling effort," an acoustic property that is a joint function of call rate and call duration (Figure S1A). As in many other animals [40], females of Cope's gray treefrogs prefer males that produce more energetically costly, "high effort" signals [41]. Experiment 4 thus simulated an intraspecific mate choice between two males differing in the quality of their sexual displays. In quiet, females preferentially selected stimuli with relatively higher calling efforts (all $p < 0.001$; Figure 3D). In the presence of noise,

the proportions of subjects choosing stimuli with relatively higher calling efforts were reduced but varied significantly as a function of noise condition (χ^2 = 6.0, df = 2, p = 0.049). Subjects were significantly more likely to correctly choose a simulated caller with a relatively higher calling effort in the comodulated condition compared with both the uncorrelated and unmodulated conditions (Figure 3D). Responses in the uncorrelated and unmodulated conditions were not different.

DISCUSSION

Our results provide strong evidence that receivers can take advantage of statistical regularities in noisy acoustic scenes to solve cocktail-party-like communication problems. An anatomical/physiological model of frog auditory processing revealed that comodulation across biologically relevant frequencies is a prominent statistical regularity in the noise generated by large social aggregations. Behavioral experiments demonstrated that, while noise induces errors in the mating decisions of females compared with those made in quiet, these errors are generally reduced in modulated noise (compared with unmodulated noise) and are further reduced in comodulated noise (compared with modulated but uncorrelated noise). Moreover, improved performance in comodulated noise was not context specific. It extended to recognition of signals at lower SNRs, as well as to better discrimination of sound patterns that mediate species discrimination and assessments of display quality in intraspecific mate choice. These behavioral contexts represent fundamental communication tasks of critical evolutionary importance for frogs and many other animals. Results from this study, therefore, substantially extend earlier work on humans [5, 21–23], frogs [42], and other animals [6, 43, 44] by showing that comodulation is a statistic of natural acoustic scenes that can be exploited to mitigate costly, noise-induced errors in communication.

Our findings cast significant new light on neglected features of environmental noise that are likely of biological relevance in many animal communication systems. Studies of acoustic communication, for example, typically consider the source of noise (e.g., biotic, abiotic, or anthropogenic), as well as its average frequency spectrum, its average amplitude, and how these two static features impact decisions made by signalers [7, 8]. In stark contrast, statistical regularities that describe the dynamic nature of noisy acoustic scenes and how these regularities are exploited by receivers to improve signal reception, as investigated in the present study, have received almost no attention [6, 24]. Given the central importance of noise as a source of selection on animal communication systems [1, 7, 8], we should not be surprised to find in future studies that many animals, similar to humans in some contexts [18], are exquisitely sensitive to the statistical regularities present in natural acoustic scenes. Future empirical studies of natural scene statistics are thus needed to fully and more accurately assess the impacts of noise on the evolution of animal communication systems. More broadly, results from this study have important ramifications for the influential "receiver psychology" paradigm [45], which emphasizes that psychological mechanisms for signal detection and discrimination are potent sources of selection on signal evolution. The prevalence of perceptual adaptations for processing statistical regularities in natural acoustic scenes and how they may ultimately impact the structure of signals remain important and unexplored questions about receiver psychology and signal evolution.

Findings from this study also have implications beyond animal communication. In many frogs, such as Cope's gray treefrog, the separate high-frequency and low-frequency spectral peaks present in signals and noise are transduced by physically distinct sensory papillae in the inner ear [36], potentially after taking different biophysical routes to the middle ear [46]. Thus, the frog auditory system processes comodulation not only across frequency channels but also across separate peripheral end organs and physical transmission pathways. Hence, the biophysical and neurosensory mechanisms frogs use to exploit comodulated noise are potentially distinct from those operating in other vertebrates. Efforts to discover the biological strategies used by a diversity of species to extract communication signals from noise have potential to uncover evolutionarily novel mechanisms that might be harnessed to improve hearing prosthetics and speech recognition systems. Compared to people with healthy auditory systems, people with hearing loss experience added difficulty communicating in a crowd, and hearing aids and cochlear implants provide limited benefits in such environments [47]. Computer algorithms for automated speech recognition also yield higher error rates in the presence of noise generated by competing speech [48]. However, evolution by natural selection has solved biological analogs of the human cocktail party problem numerous times [3, 8]. Moreover, the sense of hearing had multiple evolutionary origins [49], and even within vertebrates, key auditory mechanisms have arisen multiple times independently and differ among lineages [50]. Consequently, there is almost certainly diversity in evolved solutions to cocktail-party-like problems. Deeper knowledge of the potential diversity in hearing mechanisms could shed light on how evolution has attempted to solve complex communication problems that continue to challenge biomedical and computer engineers.

In summary, our data indicate that exploiting statistical regularities in natural acoustic scenes may be a common signal-processing strategy that has evolved to mitigate noise-induced errors in animal communication. Yet, the mechanisms underlying this strategy may differ across species given the fascinating evolutionary history of hearing. Detailed investigation into these strategies at the perceptual, biophysical, and neuronal levels in a diversity of animals not only will deepen our understanding of the mechanisms and evolution of animal communication, but could ultimately help to improve human health and technology.

ACCESSION NUMBERS

The raw data and MATLAB scripts for the model of the frog auditory periphery have been deposited in the University of Minnesota Digital Conservancy and can be accessed through the following DOI: https://doi.org/10.13020/D6ZP4H.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and five audio clips and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2017.01.031.

AUTHOR CONTRIBUTIONS

N.L., J.L.W., A.V., and M.A.B. designed experiments and collected psychophysical data. N.L., C.M., and M.A.B. designed the model of the auditory periphery. N.L. and M.A.B. wrote the paper. All authors discussed results and commented on the manuscript.

ACKNOWLEDGMENTS

We thank C. Miller, K. Schrode, J. Tumulty, J. Tanner, and three anonymous reviewers for feedback on earlier versions of the manuscript. This research was funded by a grant from the National Institute on Deafness and Other Communication Disorders (R01 DC 009582) to M.A.B. Animals were treated according to protocols reviewed and approved by the University of Minnesota's Institutional Animal Care and Use Committee (1202A10178 and 1401-31258A).

Received: November 8, 2016
Revised: January 13, 2017
Accepted: January 18, 2017
Published: February 23, 2017

REFERENCES

1. Wiley, R.H. (2015). Noise Matters: The Evolution of Communication (Harvard University Press).
2. McDermott, J.H. (2009). The cocktail party problem. Curr. Biol. *19*, R1024–R1027.
3. 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.
4. Nelken, I., Rotman, Y., and Bar Yosef, O. (1999). Responses of auditory-cortex neurons to structural features of natural sounds. Nature *397*, 154–157.
5. Verhey, J.L., Pressnitzer, D., and Winter, I.M. (2003). The psychophysics and physiology of comodulation masking release. Exp. Brain Res. *153*, 405–417.
6. Klump, G.M. (2016). Perceptual and neural mechanisms of auditory scene analysis in the European starling. In Psychological Mechanisms in Animal Communication, M.A. Bee, and C.T. Miller, eds. (Springer), pp. 57–88.
7. Brumm, H., and Slabbekoorn, H. (2005). Acoustic communication in noise. Adv. Stud. Behav. *35*, 151–209.
8. Brumm, H., ed. (2013). Animal Communication and Noise (Springer).
9. Endler, J.A. (1993). Some general comments on the evolution and design of animal communication systems. Philos. Trans. R. Soc. Lond. B Biol. Sci. *340*, 215–225.
10. Feng, A.S., Narins, P.M., Xu, C.H., Lin, W.Y., Yu, Z.L., Qiu, Q., Xu, Z.M., and Shen, J.X. (2006). Ultrasonic communication in frogs. Nature *440*, 333–336.
11. Peters, R.A., Hemmi, J.M., and Zeil, J. (2007). Signaling against the wind: modifying motion-signal structure in response to increased noise. Curr. Biol. *17*, 1231–1234.
12. Hotchkin, C., and Parks, S. (2013). The Lombard effect and other noise-induced vocal modifications: insight from mammalian communication systems. Biol. Rev. Camb. Philos. Soc. *88*, 809–824.
13. Römer, H. (2013). Masking by noise in acoustic insects: problems and solutions. In Animal Communication and Noise, H. Brumm, ed. (Springer), pp. 33–63.
14. Vélez, A., Schwartz, J.J., and Bee, M.A. (2013). Anuran acoustic signal perception in noisy environments. In Animal Communication and Noise, H. Brumm, ed. (Springer), pp. 133–185.
15. Simoncelli, E.P., and Olshausen, B.A. (2001). Natural image statistics and neural representation. Annu. Rev. Neurosci. *24*, 1193–1216.
16. Geisler, W.S. (2008). Visual perception and the statistical properties of natural scenes. Annu. Rev. Psychol. *59*, 167–192.
17. Garcia-Lazaro, J.A., Ahmed, B., and Schnupp, J.W. (2006). Tuning to natural stimulus dynamics in primary auditory cortex. Curr. Biol. *16*, 264–271.
18. McDermott, J.H., Schemitsch, M., and Simoncelli, E.P. (2013). Summary statistics in auditory perception. Nat. Neurosci. *16*, 493–498.
19. Theunissen, F.E., and Elie, J.E. (2014). Neural processing of natural sounds. Nat. Rev. Neurosci. *15*, 355–366.
20. Singh, N.C., and Theunissen, F.E. (2003). Modulation spectra of natural sounds and ethological theories of auditory processing. J. Acoust. Soc. Am. *114*, 3394–3411.
21. Grose, J.H., and Hall, J.W., 3rd. (1992). Comodulation masking release for speech stimuli. J. Acoust. Soc. Am. *91*, 1042–1050.
22. Festen, J.M. (1993). Contributions of comodulation masking release and temporal resolution to the speech-reception threshold masked by an interfering voice. J. Acoust. Soc. Am. *94*, 1295–1300.
23. Kwon, B.J. (2002). Comodulation masking release in consonant recognition. J. Acoust. Soc. Am. *112*, 634–641.
24. Ronacher, B., and Hoffmann, C. (2003). Influence of amplitude modulated noise on the recognition of communication signals in the grasshopper *Chorthippus biguttulus*. J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol. *189*, 419–425.
25. 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.
26. Vélez, A., and Bee, M.A. (2011). Dip listening and the cocktail party problem in grey treefrogs: signal recognition in temporally fluctuating noise. Anim. Behav. *82*, 1319–1327.
27. 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.
28. Wollerman, L. (1999). Acoustic interference limits call detection in a Neotropical frog *Hyla ebraccata*. Anim. Behav. *57*, 529–536.
29. 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.
30. Bee, M.A., and Schwartz, J.J. (2009). Behavioral measures of signal recognition thresholds in frogs in the presence and absence of chorus-shaped noise. J. Acoust. Soc. Am. *126*, 2788–2801.
31. Bee, M.A. (2008). Finding a mate at a cocktail party: spatial release from masking improves acoustic mate recognition in grey treefrogs. Anim. Behav. *75*, 1781–1791.
32. Ward, J.L., Buerkle, N.P., and Bee, M.A. (2013). Spatial release from masking improves sound pattern discrimination along a biologically relevant pulse-rate continuum in gray treefrogs. Hear. Res. *306*, 63–75.
33. Caldwell, M.S., and Bee, M.A. (2014). Spatial hearing in Cope's gray treefrog: I. Open and closed loop experiments on sound localization in the presence and absence of noise. J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol. *200*, 265–284.
34. 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.
35. Feng, A.S., Narins, P.M., and Capranica, R.R. (1975). Three populations of primary auditory fibers in the bullfrog (*Rana catesbeiana*): their peripheral origins and frequency sensitivities. J. Comp. Physiol. *100*, 221–229.
36. Simmons, D.D., Meenderink, S.W.F., and Vassilakis, P.N. (2007). Anatomy, physiology, and function of the auditory end-organs in the frog inner ear. In Hearing and Sound Communication in Amphibians, *Volume 29*, P.A. Narins, A.S. Feng, R.R. Fay, and A.N. Popper, eds. (Springer), pp. 184–220.
37. Hillery, C.M. (1984). Detection of amplitude-modulated tones by frogs: implications for temporal processing mechanisms. Hear. Res. *14*, 129–143.

38. Feng, A.S., Hall, J.C., and Siddique, S. (1991). Coding of temporal parameters of complex sounds by frog auditory nerve fibers. J. Neurophysiol. *65*, 424–445.

39. Schul, J., and Bush, S.L. (2002). Non-parallel coevolution of sender and receiver in the acoustic communication system of treefrogs. Proc. Biol. Sci. *269*, 1847–1852.

40. Ryan, M.J., and Keddy-Hector, A. (1992). Directional patterns of female mate choice and the role of sensory biases. Am. Nat. *139*, S4–S35.

41. Ward, J.L., Love, E.K., Vélez, A., Buerkle, N.P., O'Bryan, L.R., and Bee, M.A. (2013). Multitasking males and multiplicative females: dynamic signalling and receiver preferences in Cope's grey treefrog. Anim. Behav. *86*, 231–243.

42. Goense, J.B.M., and Feng, A.S. (2012). Effects of noise bandwidth and amplitude modulation on masking in frog auditory midbrain neurons. PLoS ONE *7*, e31589.

43. Branstetter, B.K., and Finneran, J.J. (2008). Comodulation masking release in bottlenose dolphins (*Tursiops truncatus*). J. Acoust. Soc. Am. *124*, 625–633.

44. Fay, R.R. (2011). Signal-to-noise ratio for source determination and for a comodulated masker in goldfish, *Carassius auratus*. J. Acoust. Soc. Am. *129*, 3367–3372.

45. Guilford, T., and Dawkins, M.S. (1991). Receiver psychology and the evolution of animal signals. Anim. Behav. *42*, 1–14.

46. Bee, M.A., and Christensen-Dalsgaard, J. (2016). Sound source localization and segregation with internally coupled ears: the treefrog model. Biol. Cybern. *110*, 271–290.

47. Moore, B.C.J. (2003). Introduction to the Psychology of Hearing, Fifth Edition (Elsevier).

48. Cooke, M., Hershey, J.R., and Rennie, S.J. (2010). Monaural speech separation and recognition challenge. Comput. Speech Lang. *24*, 1–15.

49. Webster, D.B., Fay, R.R., and Popper, A.N., eds. (1992). The Evolutionary Biology of Hearing (Springer).

50. Carr, C.E., and Christensen-Dalsgaard, J. (2015). Sound localization strategies in three predators. Brain Behav. Evol. *86*, 17–27.

Current Biology, Volume *27*

Supplemental Information

Frogs Exploit Statistical Regularities in Noisy Acoustic Scenes to Solve Cocktail-Party-like Problems

Norman Lee, Jessica L. Ward, Alejandro Vélez, Christophe Micheyl, and Mark A. Bee

SUPPLEMENTAL FIGURES

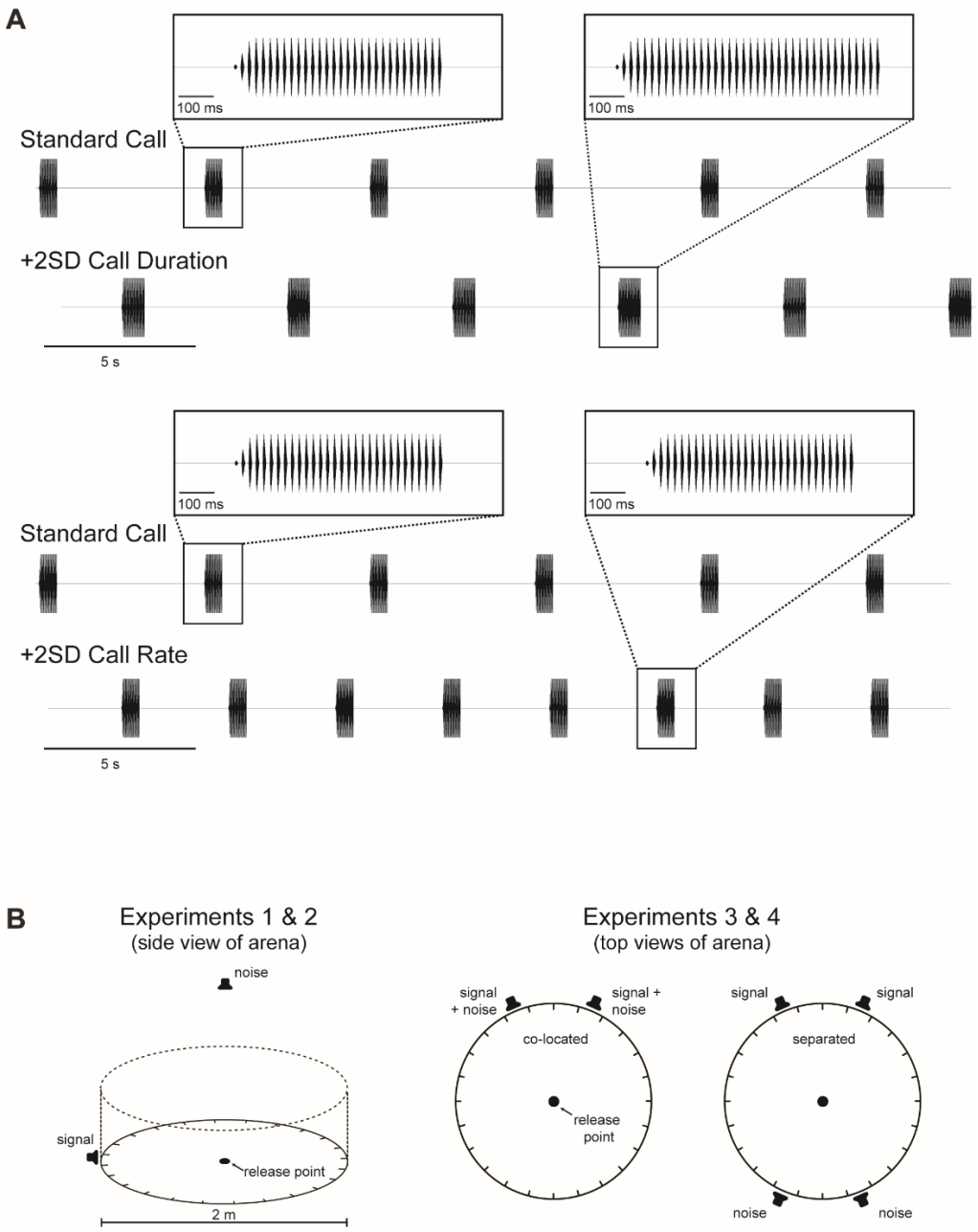

Figure S1. Manipulations of calling effort and speaker arrangements. Related to Figure 1.

(A) Oscillograms illustrating manipulations of calling effort in the two-alternative choice tests of Experiment 4. The top pair of traces illustrates the manipulation of calling effort created by using sequences of two alternative calls having the same call rate but differing in call duration, as reflected by

differences in their number of pulses (standard call versus +2SD call duration; see insets). The bottom pair of traces illustrates the manipulation of calling effort created by using sequences of two alternative calls having different call rates (standard call versus +2SD call rate) but the same call durations (see insets).

(B) Diagrams showing the test arena floor and speaker placements. In Experiments 1 and 2 (*left*), the standard call was presented from a single speaker on the floor, and noise was broadcast from an overhead speaker suspended from the ceiling of the sound chamber 1.9 m above the center of the test arena, the wall of which is indicated by dotted lines. In Experiments 3 and 4 (*right*), two alternative signals were presented from speakers separated by 45°. On 50% of noise trials, noise was broadcast from the same two speakers as the signals in a co-located condition. On the other 50% of noise trials, noise was broadcast from two speakers separated by 45° and positioned on the opposite side of the arena in a separated condition. Nominal SNRs at the subject release point were achieved by calibrating with the two noise sources simultaneously active. In both Experiments 3 and 4, all noise conditions were tested an equal number of times in the co-located and separated conditions. The speaker arrangements in Experiments 3 and 4 were employed to investigate interactions between modulated noise and spatial separation between signals and noise; those results will be reported elsewhere.

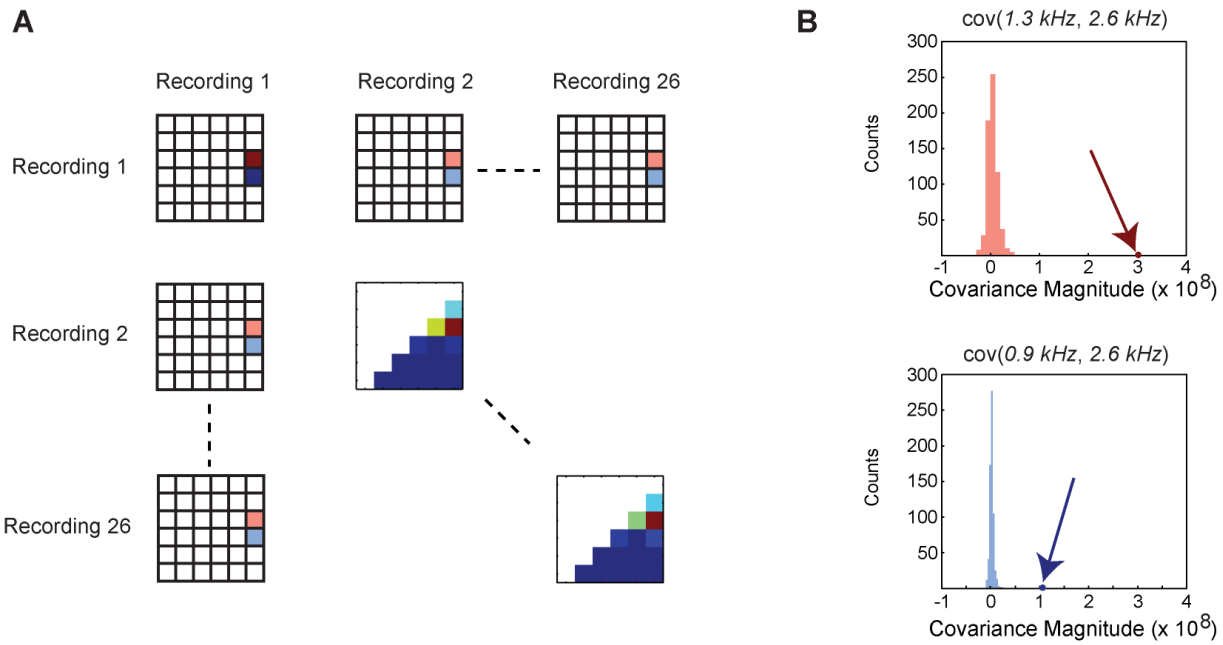

Figure S2. **Quantifying comodulation relative to null expectations. Related to Figure 2**.

(A) Illustrated schematically are computations of cross-covariance scores for each recording and the null distributions of expected values. Analyses of chorus recordings are depicted as 7 × 7 cross-covariograms (based on pairwise comparisons of the outputs of seven auditory filters) that are arranged in a 26 × 26 matrix (based on having 26 recordings of natural choruses). The off-diagonal elements of each 7 × 7 cross-covariogram correspond to the peak cross-covariance magnitude from all pairwise comparisons of different temporal envelopes from the output of the seven auditory filters implemented in the auditory

filterbank model. Each cross-covariogram in the larger matrix illustrates comparisons of envelopes from either within the same recording (along the diagonal) or between two different chorus recordings (the off diagonal). The mean cross-covariogram depicted in Figure 2 of the main text was computed by averaging separately for each pairwise comparison the cross-covariance values along the diagonal (n = 26). Null distributions for each possible pairwise comparison of two filter outputs were determined by averaging all off-diagonal cross-covariograms.

(B) Null distributions of cross-covariance values are shown for two pairwise comparisons of two frequency channels with different center frequencies, 1.3 versus 2.6 kHz (*upper*) and 0.9 versus 2.6 kHz (*lower*). Arrows depict the corresponding average magnitude of cross-covariance values from within-recording comparisons. Distributions like those depicted here were used to convert the mean cross-covariance for each pairwise envelope comparison into a *Z*-score that expressed the magnitude of cross-covariance in SD units relative to its corresponding null distribution. These *Z*-scores are reported in Figure 2B of the main text.

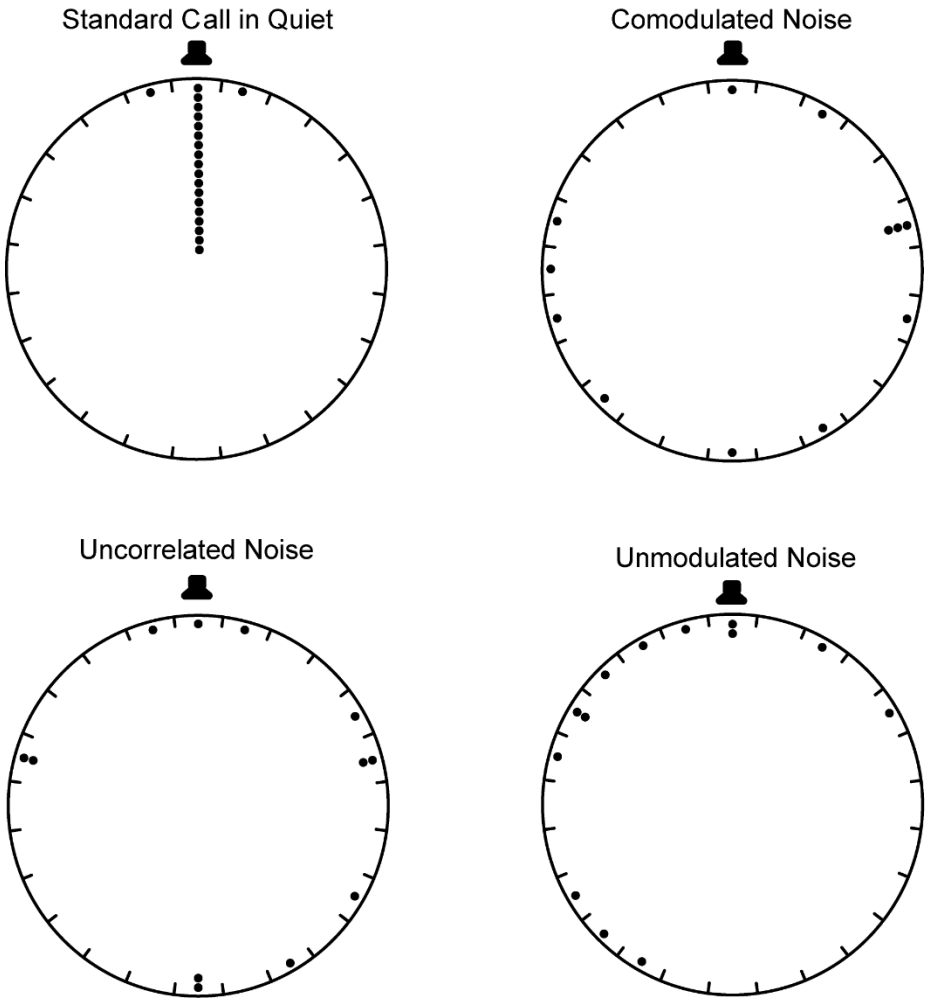

Figure S3. **A control experiment reveals noises were behaviorally neutral. Related to Figure 3**

Circular plots showing an overhead view of the 2-m circular test arena; dashes inside the perimeter of the plot demarcate 15˚ arcs around the wall of the arena. The position of each dot depicts the 15˚ arc in which a subject first touched the arena wall relative to the position of the speaker (indicated by the speaker symbol) in a control experiment (see Supplemental Experimental Procedures). Data are shown separately for subjects that reached the arena wall in response to an attractive standard call presented in quiet, and in response to presentations of comodulated noise, uncorrelated noise, and unmodulated noise. Phonotaxis was not oriented toward or away from any of the experimental noises indicating that the noises were behaviorally neutral.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Acoustic recordings

We made 26 recordings of choruses of Cope's gray treefrog from the western lineage breeding in natural wetlands located within 70 km of the St. Paul campus of the University of Minnesota – Twin Cities. Digital recordings (44.1 kHz, 16-bit) were made in May, June, or July each year between 2007 and 2010 using a Marantz PMD670 recorder and Sennheiser ME62 omnidirectional microphone. The tip of the microphone was placed 5 cm above water or ground level because, in our study populations, females approach choruses, and males usually call, from similar positions. Recordings were obtained at distances between 4 m and 15 m from the nearest calling male to minimize the influence of calls with high signal-to-noise ratios (SNRs) from a single individual. This protocol allowed us to obtain recordings of chorus background noise absent any substantial influence of single calling males. These recordings best represent the background noise present in natural choruses. We obtained chorus recordings on different nights, at different ponds, or at different locations and times within a pond. In our set of recordings, other species were either absent or calling at very low densities and away from the microphone, so that the dominant environmental sound source was always that of the chorus of Cope's gray treefrogs.

Model of the auditory periphery

To characterize comodulation in natural sounds in biologically meaningful ways requires considerations of frequency filtering and envelope extraction by the peripheral auditory system of the relevant animals. Therefore, we used a model of the Cope's gray treefrog auditory periphery to examine comodulation in our chorus recordings. Amphibians are unique among vertebrates in possessing two separate sensory papillae in their inner ears for processing airborne sounds. In species with vocalizations having bimodal spectra with two distinct spectral peaks, such as Cope's gray treefrog (see Natural Call in Figure 1A), each papilla is typically most sensitive to one of the two spectral peaks. The amphibian papilla (AP) is tuned to lower frequencies and is tonotopically organized, while the basilar papilla (BP) is tuned to a broader range of higher frequencies and lacks tonotopy [S1].

We created a model of this peripheral auditory system specific to Cope's gray treefrogs using the following four procedures. First, we determined the general relationship between best frequency (BF) and bandwidth 10 dB above threshold (10-dB BW) for frog auditory nerve fibers. To do so, we conducted a meta-analysis of 1071 VIIIth nerve frequency tuning curves (FTCs) from seven species of frogs across 10 different published studies [S2-11]. (FTCs of auditory nerve fibers have not been measured in Cope's gray treefrog.) We either used values reported in the literature, or we extracted values from published figures that graphically depicted an VIIIth nerve FTC. In the latter cases, individual data points were fitted with a rounded-exponential function (roexp; see Figure 2A, *left*), with BF determined from the fitted curve as the frequency with the lowest threshold and 10-dB BW determined as the bandwidth of the fitted curve between the two frequencies surrounding the BF that were 10 dB above threshold. A cluster analysis [S12] was used to classify units as innervating either the amphibian or basilar papilla (see Figure 2A). An orthogonal regression analysis determined a significant linear relationship (Deming Regression: slope = 1.11, 95% CI [1.109, 1.111], intercept = -0.476) between 10-dB BW and BF on a log-log scale. Second, we generated a bank of seven auditory filters using 4th order gammatone functions. We modeled the AP as a bank of six bandpass gammatone filters and determined the spacing and number of filters by overlapping adjacent filters at their 3-dB down-points. The BP was modeled as a single bandpass

gammatone filter. This procedure produced seven total filters that together spanned a range of center frequencies between 238 Hz and 2600 Hz and encompassed the estimated hearing range of Cope's gray treefrogs [S13, S14]. In Cope's gray treefrogs, the cutoff frequency between the AP and BP is about 1.625 kHz [S13, S14]. Therefore, the highest center frequency of filters in the AP range was set to 1.3 kHz, and the center frequency of the single filter simulating the BP was set to 2.6 kHz. These two frequencies correspond approximately to the frequencies of the lower and upper spectral peak, respectively, present in Cope's gray treefrog advertisement calls (Figure 1A) and chorus noise (Figure 1B). Filter bandwidth for a given center frequency was calculated based on results from our meta-analysis of the relationship between BF and 10-dB BW for frog auditory nerve fibers. Third, the relative gain of each filter was adjusted to match the midbrain audiogram of this species [S13], which is similar in overall shape to an audiogram measured using the auditory brainstem response, which is thought to reflect synchronized activity in the auditory nerve [S13, S14]. The midbrain audiogram of *H. chrysoscelis* from the eastern lineage has previously been described [S13]. This audiogram is expected to differ slightly in the spectral location of the bimodal sensitivity peaks compared to the midbrain audiogram of *H. chrysoscelis* belonging to the western lineage due to overall size differences between the two lineages [S14]. To account for these expected lineage differences, the bimodal sensitivity peaks in the original midbrain audiogram were adjusted so that the lower and higher frequency peaks were centered on 1.3 and 2.6 kHz, respectively, as determined from an audiogram estimated for the western lineage using the auditory brainstem response [S14]. The gain of each filter was adjusted to match threshold at the same frequency from the adjusted midbrain audiogram (see Figure 2A, *right*). Fourth, to simulate half-wave rectification in temporal envelope processing by the frog auditory periphery [S15, S16], the temporal envelope of the output of each of the seven frequency filters was extracted analytically using the Hilbert transformation.

To quantify comodulation within a single chorus recording, we passed it through the simulated filterbank and computed the pairwise cross-covariance (*xcov* in Matlab) between the envelopes of the outputs from all seven filters. We computed cross-covariance, instead of cross-correlation, to preserve variation in the magnitude of filter outputs arising from inherent variation across frequency in both the relative amplitude of chorus sounds and the sensitivity of the auditory system. The peak cross-covariance values across different recordings and filter comparisons ranged from 1.05×10^{9} to 9.51×10^{10}. Relatively higher cross-covariance values indicate relatively greater levels of comodulation across frequency. We averaged the peak cross-covariance values for each pairwise comparison across all 26 chorus recordings and determined whether these mean cross-covariance values were significantly greater than expected by chance. Null distributions of peak cross-covariance values were computed based on comparing the temporal envelope from each filtered frequency channel from each chorus recording to the temporal envelopes from all other frequency channels from the other 25 chorus recordings (Figure S2A). All null-distributions of cross-covariance values followed a Gaussian distribution with means ranging from -1.73×10^{7} to 6.54×10^{8} and variances ranging from 4.43×10^{12} to 2.26×10^{18}. We then expressed the mean cross-covariance value for each pairwise filter comparison as a *Z*-score standardized to the null-mean based on between-recording comparisons (Figure S2B). Z-scores greater than 1.96 are significant for a two-tailed hypothesis and an α of 0.05.

The outputs from adjacent, overlapping filters are generally expected to share greater similarities in envelopes compared to outputs from distant, non-overlapping filters. Therefore, there was some potential

that significant comodulation measured in the output from our biologically inspired auditory filterbank model might result solely from the overlap between adjacent filters rather than genuine comodulation in chorus noise. We ruled out this possibility as follows. We quantified comodulation in our 26 chorus recordings using a separate filterbank model composed of four non-overlapping rectangular filters The bandwidth of each filter was specified as 350 Hz to ensure non-overlap. Filter ripple was reduced using a Chebyshev window. The sensitivity of each filter was adjusted to match threshold at the same frequency from the adjusted midbrain audiogram described above. The 26 chorus recordings were passed through this bank of four rectangular filters and analyzed as described above for the biological filterbank model.

Comparing outputs from the rectangular filters centered on the 1.3 and 2.6 kHz spectral peaks of *H. chrysoscelis* advertisement calls, the mean cross-covariance value was 25.2 standard deviations greater than the null expectation computed using a bank of four rectangular filters. This result is similar in magnitude to the comodulation measured using our biologically inspired filterbank (see main text; Figure 2), indicating that overlap between frequency filters in our biologically-inspired model cannot explain the high level of comodulation measured in natural chorus noise.

Behavioral experiments

Subjects

Animals were collected with permission from the Minnesota Department of Natural Resources (permits 19061 and 19766) and treated according to protocols reviewed and approved by the University of Minnesota's Institutional Animal Care and Use Committee (1202A10178 and 1401-31258A). Female subjects were collected in amplexus at night (2200 – 0200 h) during the breeding seasons of 2007, 2008, 2010, 2012, 2013 and 2014 from the same wetlands in which chorus recordings were made. In total, 414 subjects were tested. Upon completion of behavioral testing, subjects were released unharmed at their location of collection. Subjects were housed in an incubator (20°C) for at least 30 min prior to testing in a temperature-controlled (20°C), hemi-anechoic sound chamber (Industrial Acoustics Corporation) lined with acoustic insulation to reduce reverberation. Testing took place under infrared (IR) light in a 2-m diameter circular arena with acoustically transparent but visually opaque walls (Figure S1B). Phonotaxis behaviors were observed in real time on a video monitor outside the chamber using an IR-sensitive camera mounted over the test arena. A response was considered to have occurred when a subject approached to within 10 cm of a speaker broadcasting a target signal.

Acoustic stimuli

All acoustic stimuli were synthetic sounds (44.1 kHz, 16 bit) generated using custom scripts in Matlab. Unless indicated otherwise, acoustic signals consisted of a repeated "standard call" (Figure 1A and Figure S1A) that was synthesized to have the average properties of *H. chrysoscelis* calls recorded in our study population [S17]. This standard call was delivered in sequence at a rate of approximately 11 calls/min, with each call consisting of 30, 10-ms pulses (10-ms interpulse interval, 0.50 pulse duty cycle, 50 pulses/s) shaped to have conspecific on-ramps and off-ramps. Each pulse consisted of two phase-locked sinusoids (starting phase = 0°) with frequencies (and relative amplitudes) of 1.3 kHz (-11 dB) and 2.6 kHz (0 dB) to simulate the bimodal frequency spectrum of natural calls (Figure 1A). In single-stimulus, no-choice tests (Experiments 1 and 2), this sequence of standard calls was the only signal used. In two-alternative choice tests (Experiments 3 and 4), the two sequences of calls differed in pulse rate

(Experiment 3) or calling effort (Experiment 4). The calls composing the two alternative sequences alternated in time with intervals of equal duration preceding and following each call unless the call rate of one sequence was manipulated (Experiment 4), in which case this antiphonal timing relationship occurred only between the first three broadcast calls (Figure S1A). In manipulations of pulse rate (Experiment 3), we maintained the overall duration of each call as close as possible to the duration of the standard call (590 ms) while manipulating the pulse duration and interpulse interval to achieve the specified pulse rate (with pulse duty cycle, on-ramps, and off-ramps having the same proportions as in the standard call).

Noises consisted of two equal-amplitude narrow bands of filtered white noise with center frequencies of 1.3 and 2.6 kHz, which matched the frequencies of the two spectral peaks in our simulated signals and corresponded to the two spectral bands present in recordings of natural choruses (Figure 1B). In Experiments 1 and 2, the two noise bands used to construct the chorus-shaped noises had the same bandwidth (400 Hz). In Experiments 3 and 4, the lower frequency noise band had a bandwidth of 400 Hz and the upper frequency noise band had a bandwidth of 800 Hz. In all cases, the two noise bands had the same spectrum level. Each noise band was either unmodulated (beyond the inherent temporal fluctuations in narrow band noise; Figure 1E) or modulated by the envelope of a low-pass filtered noise (12.5-Hz cutoff; Figures 1C & 1D). This envelope modulator imparted random, low-frequency fluctuations in the level of the noise band. For the comodulated condition (Figure 1C), we used the same envelope modulator to modulate both noise bands, thereby creating correlated fluctuations in level across the spectrum of both noise bands. To create the modulated but *uncorrelated* noise condition (Figure 1D), we used different envelope modulators for the two noise bands so that they fluctuated independently of one another. Note that modulated noises created in this manner fluctuate in level but are also behaviorally neutral because they are not confounded by containing the specific temporal properties present in individual signals (Figure S3). We used a minimum of four different exemplars of white noise and low-pass noise modulators in each experiment.

All sounds were delivered from a PC computer through Mod1 Orb speakers using an M-Audio Firewire 410 multichannel sound card and HTD DMA-1275 multichannel amplifier. Signal speakers were located just outside the arena wall and aimed toward a subject release point at the center of the arena (Figure S1B). In Experiments 1 and 2, noises were broadcast from an overhead speaker suspended from the ceiling of the sound chamber at a height of 1.9 m directly above the test arena (Figure S1B). In Experiments 3 and 4, the two alternatives were broadcast from speakers separated by 45° around the perimeter of the test arena. Noises in these two experiments were broadcast from speakers located on the floor of the test arena that were either co-located with the two speakers broadcasting signal alternatives or broadcast from two speakers on the opposite side of the arena, with each one separated by 180° from one of the two signal speakers (Figure S1B). Interactions between the type of noise (comodulated, uncorrelated, and unmodulated) and its location of broadcast relative to signals (co-located or separated) were generally small or negligible and are not examined as part of this study. In all cases, signal levels (LCF) and noise levels (LC_{eq}) were calibrated by placing the microphone (Type 4950) of a Brüel & Kjær Type 2250 sound level meter at the subject release point in the center of the circular test arena, where subjects were placed in a small holding cage at the start of each test. In all four experiments, all three types of noise were equalized in software to have the same root-mean-squared (RMS) amplitude and calibrated to 73 dB SPL at the subject release point. This amplitude reflects levels we have measured in

natural choruses. All SNRs are stated based on the calibrated levels of signals and noises at the subject release point.

Experiment 1: SRTs determined with fixed SNRs

We determined signal recognition thresholds (SRTs) in response to the standard call using fixed SNRs. Noise was broadcast at 73 dB SPL and signal amplitude was varied to achieve 5 SNRs (-12, -6, 0, +6, and +12 dB). In a series of signal-stimulus, no-choice phonotaxis tests, each subject was tested in quiet and in comodulated, uncorrelated, and unmodulated noise at a fixed SNR (n = 30 per SNR, total n = 150). In quiet, subjects were tested using the same five signal levels used to achieve the nominal SNRs in the three noise conditions. We used generalized estimating equations (GEE) with the logit link function and an exchangeable correlation matrix to model the proportion of subjects responding as a function of noise type and SNR. In preliminary analyses, all main effects and interaction terms were included in the model. Non-significant interaction terms were removed from the final model and the most parsimonious model was selected based on the Corrected Quasi Likelihood Under Independence Model Criterion (QICC). SRTs were determined from the fitted marginal model at a threshold criterion of 0.5. Subsequent pairwise comparisons were based on the marginal means.

Experiment 2: SRTs determined with adaptive tracking

We determined SRTs for responding to the standard call using an adaptive tracking procedure across a series of single-stimulus, no-choice phonotaxis tests [S18]. This procedure involved varying the level of the signal across different tests conducted in the presence of noise to determine the lowest SNR that reliably elicited positive phonotaxis. Thresholds were determined for different groups of subjects (n = 20 per group) tested with comodulated, uncorrelated, and unmodulated noise (total n = 60). Noise-dependent differences in thresholds were assessed using one-way analysis of variance and planned, pairwise contrasts. In Figure 3B, results from this experiment are graphically compared with the thresholds measured in quiet in a separate study conducted in parallel using the same set up and procedures [S19].

Experiment 3: Pulse rate discrimination

Two-alternative choice tests were conducted to assess pulse rate discrimination in noise. In each test, subjects (total n = 64) chose between two otherwise similar calls differing in pulse rate and thereby simulating a choice between a conspecific call and a heterospecific (*Hyla versicolor*) call (Figure 1A). Each subject was tested in a pair of choice tests replicated in quiet and in the comodulated, uncorrelated, and unmodulated noise conditions (8 tests/subject total). One test in the pair constituted a relatively easy choice between sequences of the standard call (50 pulses/s) and an alternative call with a mean pulse rate simulating *H. versicolor* calls (20 pulses/s; see Figure 1A). The second test in the pair represented a potentially more difficult choice between two alternatives with pulse rates near the most similar extremes of conspecific versus heterospecific males (40 versus 30 pulses/s, respectively) [S20]. Hence in both tests, one alternative simulated a calling conspecific and the other simulated a calling heterospecific. Tests of 50 versus 20 pulses/s were conducted at SNRs of -3 dB and -6 dB (50% of subjects each), and tests of 40 versus 30 pulses/s, because it was expected to be a more difficult test, were conducted at SNRs of 0 dB and -3 dB (50% of subjects each). (The effects of these SNR differences are not examined in this study.) In quiet, 100% of subjects chose 50 pulses/s over 20 pulses/s (one-tailed binomial $p < 0.001$), and 94% chose 40 pulses/s over 30 pulses/s (one-tailed binomial $p < 0.001$). Therefore, in tests with noise, we

considered an approach to within 10 cm of the speaker broadcasting the call alternative with a relatively faster pulse rate to be a correct response to a conspecific call. We evaluated differences in the proportions of subjects choosing the conspecific pulse rate [P(Conspecific Pulse Rate)] as a function of noise type using GEE with a binomial distribution, logit link function, and exchangeable correlation matrix. Subsequent pairwise comparisons of responses across noise types were based on marginal means.

Experiment 4: Calling effort discrimination

Two-alternative choice tests were used to evaluate calling effort discrimination in noise. Calling effort (measured as pulses/min) is the product of call duration (pulses/call) and call rate (calls/min). We manipulated calling effort by varying either the call duration (pulses/call) or the call rate (calls/min) of the two stimulus alternatives in a test (Figure S1A). If call rate was manipulated, call duration was held constant across both alternatives, and if call duration was manipulated, call rate was held constant across both alternatives. Hence, the two males simulated by the two alternatives had either the same call rate or the same call duration, but differed in calling effort. In all tests, one alternative was a sequence of the standard call simulating an average calling effort (330 pulses/min) based on the mean call rate (11 calls/min) and mean call duration (30 pulses/call) for our study population [S17] Across four different two-alternative choice tests (n = 30/test; total n = 120), the alternative to the standard call sequence consisted of a sequence of calls in which the call rate or the call duration was either +2SD or -2SD relative to the population mean values. For manipulations of call rate, the +2SD (17 calls/min) and -2SD (5 calls/min) values corresponded to calling efforts of 510 pulses/min and 150 pulses/min, respectively. In our manipulations of call duration, the +2SD (38 pulses/call) and -2SD (22 pulses/call) values corresponded to calling efforts of 418 pulses/min and 242 pulses/min, respectively. In quiet, females preferred sequences with higher calling efforts (one-tailed binomial tests: +2SD call rate versus the standard call, $p < 0.001$; +2SD call duration versus the standard call, $p = 0.003$; the standard call versus -2SD call rate, $p < 0.001$; the standard call versus -2SD call duration, $p = 0.049$). Therefore, we considered a correct response to be a choice of the alternative with the relatively higher calling effort in each type of test. We used GEE with a binomial distribution, logit link function, and exchangeable correlation matrix to model the proportion of subjects choosing the alternative simulating the relatively higher calling effort [P(Higher Calling Effort)] as a function of noise type. Subsequent pairwise comparisons were based on marginal means.

Control experiment: Sham responses to noise

In some frogs, chorus-shaped noise has potential to act as a biologically informative signal that competes with other target signals (i.e., calls) in ways unrelated to auditory masking [S21-23]. We conducted a control experiment to rule out this possibility (total n = 20). Using single-stimulus, no-choice tests similar to those used in Experiments 1 and 2, we evaluated the possibility that our artificial chorus-shaped noises influenced phonotaxis behavior. In separate tests, either the standard call (presented in quiet) or a different chorus-shaped noise (comodulated, uncorrelated, or unmodulated) was broadcast as if it were a potential target signal from a single speaker located on the floor just outside the arena wall and directed towards the central subject release point. Subjects were given 5 minutes to make contact with the arena wall. (Responses to attractive stimuli typically occur within 1.5 min.) We measured the angle (in 15˚ arcs) at which subjects first touched the arena wall relative to the 15˚ arc in which the speaker was located (designated as 0˚). Noise treatments were tested in a randomized order within subjects. Twenty subjects

were tested in response to each noise, though not all subjects reached the wall within 5 minutes in each test. We used circular statistics (V tests) to determine whether touches of the arena wall were oriented toward the playback speaker.

The control experiment to measure sham responses to comodulated, uncorrelated, and unmodulated noise revealed that these experimental sounds were behaviorally neutral and did not elicit positive phonotaxis. In quiet, subjects exhibited significant orientation toward the standard call ($n = 20$, $V = 19.4$, $p < 0.0001$). Eighteen of 20 subjects (90%) first made contact with the wall of the circular arena in the 15˚ arc centered on the speaker; the remaining 2 subjects made first contact in an adjacent arc (Figure S3). In contrast, there was no indication that frogs oriented toward the chorus-shaped noises (comodulated noise: $n = 12$, $V = -0.2$, $p = 0.53$; uncorrelated noise: $n = 12$, $V = 0.5$, $p = 0.42$; unmodulated noise: $n = 13$, $V = 2.2$, $p = 0.20$) (Figure S3). These results confirmed that the experimental noises used in Experiments 1-4 were behaviorally neutral and did not compete with target signals in ways unrelated to auditory masking. Additionally, data from this control experiment could be used to deduce that the likely false alarm rate was low in Experiments 1 and 2, in which we estimated SRTs. In the absence of a call-like signal (Figure S3), only one or two of 20 subjects (5-10%) made their first contact with the arena wall in the 15° arc in front of a speaker. We previously estimated that the false alarm rate in similar experiments did not exceed 20% [S24].

SUPPLEMENTAL REFERENCES

S1. Simmons, D.D., Meenderink, S.W.F., and Vassilakis, P.N. (2007). Anatomy, physiology, and function of the auditory end-organs in the frog inner ear. In Hearing and Sound Communication in Amphibians, Volume 29, P.A. Narins, A.S. Feng, R.R. Fay and A.N. Popper, eds. (New York: Springer), pp. 184-220.

S2. Feng, A.S., Narins, P.M., and Capranica, R.R. (1975). Three populations of primary auditory fibers in the bullfrog (*Rana catesbeiana*): Their peripheral origins and frequency sensitivities. Journal of Comparative Physiology A *100*, 221-229.

S3. Capranica, R.R., and Moffat, A.J.M. (1975). Selectivity of the peripheral auditory system of spadefoot toads (*Scaphiopus couchi*) for sounds of biological significance. Journal of Comparative Physiology *100*, 231-249.

S4. Frishkopf, L.S., and Goldstein Jr, M.H. (1963). Responses to acoustic stimuli from single units in the eighth nerve of the bullfrog. Journal of the Acoustical Society of America *35*, 1219-1228.

S5. Narins, P.M., and Capranica, R.R. (1980). Neural adaptations for processing the two-note call of the Puerto Rican treefrog, *Eleutherodactylus coqui*. Brain Behav. Evol. *17*, 48-66.

S6. Shofner, W.P., and Feng, A.S. (1981). Post-metamorphic development of the frequency selectivities and sensitivities of the peripheral auditory system of the bullfrog, *Rana catesbeiana*. Journal of Experimental Biology *93*, 181-196.

S7. Zakon, H., and Capranica, R.R. (1981). An anatomical and physiological study of regeneration of the eighth nerve in the leopard frog. Brain Research *209*, 325-338.

S8. Wilczynski, W., Zakon, H., and Brenowitz, E.A. (1984). Acoustic communication in spring peepers: Call characteristics and neurophysiological aspects. Journal of Comparative Physiology A *155*, 577-584.

S9. Zelick, R., and Narins, P.M. (1985). Temporary threshold shift, adaptation, and recovery characteristics of frog auditory nerve fibers. Hearing Research *17*, 161-176.

S10. Narins, P.M., and Wagner, I. (1989). Noise susceptibility and immunity of phase locking in amphibian auditory nerve fibers. Journal of the Acoustical Society of America *85*, 1255-1265.

S11. Ronken, D.A. (1990). Basic properties of auditory-nerve responses from a ‘simple’ ear: The basilar papilla of the frog. Hearing Research *47*, 63-82.

S12. Ronken, D.A. (1991). Spike discharge properties that are related to the characteristic frequency of single units in the frog auditory nerve. Journal of the Acoustical Society of America *90*, 2428-2440.

S13. Hillery, C.M. (1984). Detection of amplitude-modulated tones by frogs: Implications for temporal processing mechanisms. Hearing Research *14*, 129-143.

S14. Schrode, K.M., Buerkle, N.P., Brittan-Powell, E.F., and Bee, M.A. (2014). Auditory brainstem responses in Cope's gray treefrog (*Hyla chrysoscelis*): Effects of frequency, level, sex and size. Journal of Comparative Physiology A *200*, 221-238.

S15. Feng, A.S., Hall, J.C., and Siddique, S. (1991). Coding of temporal parameters of complex sounds by frog auditory nerve fibers. Journal of Neurophysiology *65*, 424-445.

S16. Simmons, A.M., and Ferragamo, M. (1993). Periodicity extraction in the anuran auditory nerve 1. "Pitch-shift" effects. Journal of Comparative Physiology A *172*, 57-69.

S17. Ward, J.L., Love, E.K., Vélez, A., Buerkle, N.P., O'Bryan, L.R., and Bee, M.A. (2013). Multitasking males and multiplicative females: Dynamic signalling and receiver preferences in Cope's grey treefrog. Animal Behaviour *86*, 231-243.

S18. Bee, M.A., and Schwartz, J.J. (2009). Behavioral measures of signal recognition thresholds in frogs in the presence and absence of chorus-shaped noise. Journal of the Acoustical Society of America *126*, 2788-2801.

S19. Nityananda, V., and Bee, M.A. (2012). Spatial release from masking in a free-field source identification task by gray treefrogs. Hearing Research *285*, 86-97.

S20. Ward, J.L., Buerkle, N.P., and Bee, M.A. (2013). Spatial release from masking improves sound pattern discrimination along a biologically relevant pulse-rate continuum in gray treefrogs. Hearing Research *306*, 63-75.

S21. Bee, M.A. (2007). Selective phonotaxis by male wood frogs (*Rana sylvatica*) to the sound of a chorus. Behavioral Ecology and Sociobiology *61*, 955-966.

S22. Swanson, E.M., Tekmen, S.M., and Bee, M.A. (2007). Do female frogs exploit inadvertent social information to locate breeding aggregations? Canadian Journal of Zoology *85*, 921-932.

S23. Christie, K., Schul, J., and Feng, A.S. (2010). Phonotaxis to male's calls embedded within a chorus by female gray treefrogs, *Hyla versicolor*. Journal of Comparative Physiology A *196*, 569-579.

S24. Vélez, A., and Bee, M.A. (2011). Dip listening and the cocktail party problem in grey treefrogs: Signal recognition in temporally fluctuating noise. Animal Behaviour *82*, 1319-1327.