Main Manuscript for

Informational Masking Constrains Vocal Communication in Nonhuman Animals

Saumya Gupta[a,1,2], Lata Kalra[a], Gary J. Rose[b], and Mark A. Bee[a,c]

[a] Department of Ecology, Evolution, and Behavior, University of Minnesota, Saint Paul, MN 55108, USA

[b] School of Biological Sciences, University of Utah, Salt Lake City, UT 84112, USA

[c] Graduate Program in Neuroscience, University of Minnesota, Minneapolis, MN 55455, USA

[1] To whom correspondence may be addressed: saumya1692@gmail.com

[2] Current Address: Department of Biology, University of Washington, Seattle, WA 98195, USA

Keywords: animal behavior | auditory perception | frog | hearing | mate choice | vocal signals

Abstract

Noisy social environments constrain human speech communication in two important ways: spectrotemporal overlap between signals and noise can reduce speech audibility ("energetic masking") and noise can interfere with processing informative features of otherwise audible speech ("informational masking"). Whether and how informational masking impacts vocal communication in nonhuman animals remains unknown. In this study of a treefrog, in which females choose mates in noisy breeding choruses, we investigated whether informational masking disrupts the recognition of vocal signals that guide mating decisions. We observed significantly reduced rates of responding to the vocalizations of a potential mate when target signals were broadcast concurrently with other biologically relevant sounds that obscured signal information in the absence of energetic masking. These informational masking effects were evident across a broad range of signal levels and were more pronounced when competing sounds fell in a particularly salient frequency range for processing vocalizations. Our results confirm informational masking is a general communication problem among humans and other animals and identify a crucial yet understudied source of selection on animal communication systems.

Introduction

Humans experience the "cocktail party problem"[1] in noisy social settings when the competing voices of multiple talkers mask speech of interest and make conversations more difficult to follow. Underlying the cocktail party problem are two different forms of auditory masking that interfere with signal processing through different mechanisms[2, 3, 4]. *Energetic masking* generally occurs at the periphery when noise renders a signal inaudible because they are similar in frequency and occur synchronously or nearly so and thus stimulate the same peripheral auditory filter. In contrast, *informational masking* occurs when competing sounds reduce the ability of listeners to correctly interpret otherwise audible signals because they share certain features, such as temporal patterns[5], a common spatial origin[6], or linguistic content[7]. Informational masking is thought to result from breakdowns in auditory stream segregation or impairments of selective attention to audible and well segregated signals[2, 3, 4]. Thus, whereas energetic masking is usually attributed to the limitations of peripheral sensory filters, informational masking is defined in terms of constraints on central processes for selecting and extracting information from signals of interest amid concurrent sounds with similar features. Informational masking is, therefore, an important constraint on speech recognition in noisy social environments. While neurophysiological studies of human cortex have shed some light on the mechanisms underlying informational masking[8], we still lack investigations of its cellular basis in suitable animal models[9], particularly at subcortical levels[10].

As in humans, a diversity of animal species, such as orthopteran insects[11], frogs[12], penguins[13], songbirds[14], and bats[15], communicate in noisy social environments where they experience cocktail-party-like problems because multiple individuals simultaneously produce concurrent acoustic signals[16]. Such environments constrain the detection[17], localization[18], recognition[19], and discrimination[17, 20] of biologically important sounds in a variety of evolutionarily consequential contexts. Ecological and evolutionary studies of animal communication in noise tend to focus on the functional consequences of auditory masking, but do so with little regard for distinguishing between the impacts of energetic versus informational masking[21, 22]. Thus, the extent to which informational masking is a general problem in acoustic communication that also constrains signal recognition in other animals remains unknown[23]. Studies of nonhuman animals are necessary to discover the extent to which informational masking acts as a heretofore unidentified source of selection on animal communication and to identify suitable animal models for investigating its neural basis.

In humans, temporal information conveyed by the speech waveform is critical for speech recognition[24] and is susceptible to informational masking[5]. Temporal information is also key to signal recognition in a diversity of nonhuman animals, from fish[25], to flies[26], to frogs[26]. Here, we tested the hypothesis that informational masking can disrupt the recognition of biologically informative temporal features in the vocal signals of frogs. Specifically, we investigated whether,

in conditions of reduced energetic masking, a potential informational masker could nevertheless disrupt the ability of female frogs to process temporal features of male signals that inform their mating decisions. We tested the informational masking hypothesis using females of Cope's gray treefrog, *Hyla chrysoscelis*, a well-studied anuran species in the contexts of mate choice[27, 28, 29], vocal communication in noise[19, 20, 30], and neural mechanisms for auditory processing[31, 32, 33]. To attract mates, male gray treefrogs aggregate in mixed species breeding choruses where they produce sequences of loud advertisement calls consisting of a short train of 12 to 43 evenly spaced pulses, with an average rate of approximately 50 pulses/s (Fig. 1a)[29, 34]. High levels of concurrent calling by nearby males — both conspecifics and other species that produce pulsatile calls with different spectral frequencies and pulse rates — are commonplace in dense choruses[34, 35]. To initiate mating, gravid female gray treefrogs approach sounds they recognize as the advertisement calls of their own species[27, 28]. Recognition of potential mates in this species is based on the female's ability to correctly process information conveyed by the temporal features of the male's advertisement call[29, 34, 36, 37, 38]. Here, we investigated whether call recognition is susceptible to informational masking due to the presence of other concurrent sounds, as would be encountered frequently in a breeding chorus. We hypothesized that, as for humans at a cocktail party, informational masking can constrain signal recognition, and thus vocal communication. The effects of informational masking on signal recognition were evaluated using female phonotaxis behavior in simulated mate choice trials.

Results

We tested the informational masking hypothesis in a series of two experiments that exploited key features of hearing and communication in gray treefrogs. Each pulse in the male's advertisement call consists of two harmonically related spectral components with frequencies of about 1.25 and 2.5 kHz (Fig. 1a)[29]. Females do not require both frequencies to recognize conspecific calls, and they readily respond to artificial signals with species-specific pulse rates but just one of the two spectral components[34, 39, 40, 41]. Importantly, the two spectral components primarily stimulate two physically distinct sensory organs in the frog inner ear that are sensitive to different frequency ranges of airborne sounds. The amphibian papilla (AP) is tonotopically organized and sensitive to relatively lower frequencies than the basilar papilla (BP), which is broadly tuned to a higher frequency range[28]. In gray treefrogs, the lower (~1.25 kHz) and higher (~2.5 kHz) spectral components of the call primarily stimulate the AP and BP, respectively (Fig. 1a)[42]. Together, these features provided a critical leverage point for testing the informational masking hypothesis: we placed target signals and potential maskers in different frequency ranges in order to stimulate physically distinct peripheral channels (i.e., separate inner ear papillae; Fig. 1b,c), thereby significantly reducing the potential for energetic masking at the periphery.

Informational Masking Reduces Responses to Mating Signals. To assess the impact of informational masking on phonotaxis behavior, we conducted two-alternative choice tests by broadcasting two sequences of identical advertisement calls as target signals (Fig. 2a,b, Movie S1). The two sequences of alternating target signals simulated two antiphonally calling males. Each signal in both sequences consisted of a train of 20 identical pulses, each with a carrier frequency that was fixed at either 1.25 kHz or 2.5 kHz. For one randomly selected sequence of target signals, each signal was presented alone i.e., in the absence of any concurrent sound. In the alternative sequence, each target signal was paired with the simultaneous broadcast of an additional, concurrent sound. In the absence of other sounds, females choose equally (50% each) between two identical signals[39] (*SI Appendix*, Fig. S1); therefore, a statistically significant departure from the expected 50% was taken to indicate a female preference for one stimulus over the other.

In the informational masking (IM) test condition (Fig. 1b), the concurrent sound broadcast with one of the two alternating signals blended elements of informational maskers from classic experiments on humans (e.g., the multitone masking paradigm)[3, 43] with elements of biological realism pertinent to a female gray treefrog choosing a mate (e.g., concurrent calls in a chorus)[34, 44]. This informational masker was a train of random-frequency pulses restricted to the frequency range opposite that of the signal with respect to the tuning of the AP and BP and presented at half the rate (25 pulses/s) of pulses in the signal (50 pulses/s). By separating the signal and masker in both frequency, pulse rate and timing we attempted to eliminate the potential for energetic masking in both the spectral and modulation domains[45]. The frequencies present in the informational masker coincided with frequencies emphasized in the calls of other syntopically breeding frogs [34]. We randomized the carrier frequency of the pulses to introduce stimulus uncertainty, which contributes to informational masking in humans[3, 43]. A pulsatile informational masker was used because pulses are common in animal acoustic signals, including frog calls[28], and thus represent naturalistic sounds for many diverse species. The masker's slower pulse rate was chosen, in part, because other syntopically breeding frog species produce calls with relatively slower pulse rates[34], including the closely-related eastern gray treefrog (*H. versicolor*), whose calls are generally unattractive to females of *H. chrysoscelis*[20, 34, 36, 38].

A key element of our experimental design is that in the IM test condition, consecutive pulses in the informational masker were temporally *interleaved* between consecutive pairs of pulses in the target signal (Figs 1b, 3a). According to the informational masking hypothesis, we predicted that a series of slow-rate, spectrally remote, and interleaved pulses would obscure biologically important temporal patterns encoded in the overall signal envelope and consequently shift female preferences toward identical signals free from overlap. To control for the possibility that the informational masker might shift female preferences because it was inherently aversive, we also tested females in a separate IM control condition, in which the pulses of the informational

masker were made *synchronous* (as opposed to interleaved) with signal pulses. For this IM control condition, we simply shifted the masker in time by 10 ms relative to the target signal such that each of its pulses was synchronous with a pulse in the signal. In contrast to interleaved pulses of the IM test condition, the overall temporal pattern of the overlapped signals was not disrupted when signal pulses were overlapped by synchronous masker pulses in the IM control condition. Under this scenario, no informational masking of temporal information was expected, and hence females were predicted to choose between overlapped and non-overlapped signals in equal proportions. We also tested an energetic masking (EM) control condition (Fig. 1c), in which the concurrent sound that overlapped the signals was a band-limited noise having the same duration, amplitude, and long-term spectrum as the informational masker. The EM control condition was designed to control for any effects of energetic masking in the spectral domain because the band-limited noise stimulated the auditory system with the same overall acoustic energy as the informational masker but lacked information-carrying temporal features. Females tested in the EM control condition were thus predicted to choose between overlapped and non-overlapped signals in equal proportions.

When the target signal was presented in an alternating fashion with a target-masker combination, overall subjects were significantly more likely to choose the former in the IM test condition compared with both the IM control condition (GEE: $\beta = 2.00$; $P < 0.001$) and EM control condition (GEE: $\beta = 1.25$; $P < 0.001$); that is, the informational masker appeared to most strongly reduce the attractiveness of the target signal. Choices did not differ significantly based on differences in signal frequency ($\chi^2 = 0.99$; $P = 0.319$). In the IM test condition (Fig. 3a,b), a significant majority of females (72%; two-tailed binomial test: $P < 0.001$) rejected the target signal presented with a concurrent informational masker in favor of an otherwise identical signal presented alone when data were pooled across both test frequencies (Fig. 3b). This trend was more pronounced in tests with the 1.25 kHz target signal, in which 82% of subjects (two-tailed binomial test: $P < 0.001$) chose the signal presented alone vs. only 61% of subjects tested with the 2.5 kHz signal (two-tailed binomial test: $P = 0.281$) (Fig. 3b). Shifting the masker by 10 ms in the IM control condition (Fig. 3a,c) so that its pulses were synchronous with signal pulses caused a reversal of the preference observed in the IM test condition: a significant majority of females (74%; two-tailed binomial test: $P < 0.001$) actually preferred the signal presented concurrently with a masker having synchronous pulses (Fig. 3c). This trend from data pooled over both signal frequencies was also significant for females tested with the 1.25 kHz signal (78%; two-tailed binomial test: $P < 0.001$) and nearly so for those tested with the 2.5 kHz signal (69%; two-tailed binomial test: $P = 0.0501$). In contrast to the IM test and IM control conditions, there was no difference in the percentage of subjects in the EM control condition that chose the target signal presented alone versus with concurrent broadband noise (signal alone: 43%; two-tailed binomial $P = 0.275$; Fig. 3d) when data was pooled over both signal frequencies. Similar results were

obtained when responses to target signals with frequencies of 1.25 kHz (47%; two-tailed binomial $P = 0.868$; Fig. 3d) and 2.5 kHz (38%; two-tailed binomial $P = 0.215$; Fig. 3d) were considered separately.

In this experiment, females preferentially chose signals presented alone over signals presented with concurrent sounds in the IM test condition, but not in the IM control and EM control conditions. The preference observed in the IM test condition, which was somewhat more pronounced (82% vs. 61%) when the target signal was 1.25 kHz, is an important result consistent with an informational masking hypothesis. The observed pattern of preferences is inconsistent with energetic masking. In the IM test condition, pulses in the overlapped signals and the informational maskers were temporally interleaved and primarily transduced by different sensory inner ear structures due to their frequency separation, which substantially reduced the potential for energetic masking. Indeed, females showed no preference for signals alone in the EM control condition, in which we presented band-limited noises lacking temporal features but otherwise matching the informational maskers in overall spectrum and amplitude. The preference for signals alone in the IM test condition also cannot be attributed to a general aversive effect of the informational masker (*SI Appendix*, Fig. S2). When pulses in the overlapped signals and the informational maskers were made synchronous in the IM control condition, the stimulus actually became *more attractive* than a signal presented alone. This result, which is consistent with a previous study showing the importance of temporal coherence as an auditory grouping cue in this species[37], definitively rules out masker aversion as an explanation for the observed preferences for the signal alone in the IM test condition. Together, these results suggest informational masking by concurrent sounds can reduce responses to mating signals if information-carrying temporal features of signals are obscured. Given these findings, we next asked more specifically whether the ability to recognize signals is disrupted under informational masking conditions.

Informational Masking Constrains Signal Recognition. In gray treefrogs, females must hear a threshold number of pulses produced at a conspecific pulse rate before exhibiting the positive phonotaxis that indicates recognition of the call as the sound of a potential mate[30, 31]. In a second experiment, we tested the prediction that concurrent pulsatile sounds at remote frequencies constrain signal recognition, as evidenced by lower response rates, elevated pulse number thresholds, and longer response latencies relative to conditions in which signals were presented alone or with concurrent band-limited noise. Using a series of no-choice tests (Fig. 2c) simulating a single calling male, we presented subjects with a repeating target signal (either 1.25 kHz or 2.5 kHz) in which we varied the number of consecutive pulses in the signal across different tests. Each subject's responses were measured in both a signal alone condition (i.e., in the absence of concurrent sound) and in either the IM test condition or the EM control condition (Fig. 4a,c). Across tests, target signals varied in pulse number from 1 to 20 pulses/call, and an adaptive tracking procedure[31] was used to determine each subject's pulse number thresholds in the signal

alone condition and the assigned masking condition. Stimuli with more than 20 pulses were not tested because 20 pulses exceeds both the minimum number of pulses recorded in local populations (12 pulses)[29] and the upper end of the range of pulse number thresholds for responding to signals with both spectral peaks (16 pulses)[31]. Both signal frequencies and all three conditions were replicated factorially at target-to-masker ratios (TMRs; or equivalent amplitudes in signal alone conditions) of -12 dB, -6 dB, and 0 dB. Difference in responses across the signal alone, IM test, and EM control conditions depended on signal frequency and, similar to the first experiment, were more pronounced with the 1.25 kHz signal (response rates: $\chi^2 = 16.5$, $P < 0.001$; pulse number thresholds: $\chi^2 = 28.9$, $P < 0.001$; response latencies: $\chi^2 = 4.44$, P 0.035); therefore, we analyzed results for each signal frequency separately.

In tests with the 1.25 kHz signal, nearly all subjects (95% pooled across all TMRs) responded to one of the target signals presented in the signal alone condition (Fig. 4b). Consistent with the previous experiment and earlier work[31], this result demonstrates robust recognition of signals with 20 or fewer pulses as potential mate attraction signals in the absence of any other concurrent sounds. Similar to the signal alone condition, a high percentage of subjects (89% pooled across all TMRs) in the EM control condition also responded to a 1.25 kHz target signal (Fig. 4b), confirming robust recognition of stimuli with 20 or fewer pulses in the presence of a concurrent band-limited noise at a remote frequency. While concurrent noise in the EM control conditions reduced response rates significantly ($\beta = -0.97$, $P = 0.032$), the reduction was modest (89% versus 95%; Fig. 4b). In contrast to the signal alone and EM control conditions, only about half of subjects (55% pooled across all TMRs) responded to a 1.25 kHz target signal with 20 or fewer pulses in the IM test condition (Fig. 4b), whereas the other half failed to respond at all, even to a 20-pulse signal, the longest signal used. This result was broadly consistent across a 12-dB range of TMRs. Overall, response rate in the IM test condition was significantly lower than response rates in both the signal alone condition (GEE: $\beta = -2.72$, $P < 0.001$) and the EM control condition (GEE: $\beta = -1.74$, $P < 0.001$). In contrast to tests with the 1.25 kHz signal, nearly all subjects (84% to 100%) responded to a 2.5 kHz target signal having 20 or fewer pulses across all three masking conditions and across all three TMRs (Fig. 4d). These data were not analyzed statistically because of the high and nearly uniform response rates close to 100%, including a 100% response rate in the EM control condition across all three TMRs (Fig. 4d).

To evaluate impacts on pulse number thresholds and response latency, we compared responses across only those subjects that responded to a target signal in both the signal alone condition and the paired masking condition. In tests with the 1.25 kHz signal (Fig. 5a), and at a TMR of -6 dB, pulse number thresholds were significantly elevated in the IM test condition compared with both the signal alone condition and the EM control condition, which themselves did not differ (Fig. 5b; Table 1). Pulse number thresholds in response to 1.25 kHz signals were not significantly different across masking conditions at TMRs of -12 dB and 0 dB (Fig. 5b; Table

1). In responses to the 2.5 kHz signal (Fig. 5d), pulse number thresholds were significantly higher in the IM test condition relative to both the signal alone condition and the EM control condition at a TMR of 0 dB (Fig. 5e, Table 1). Pulse number thresholds were significantly lower in the EM control condition relative to the signal alone condition (Fig. 5e, Table 1). Broadly similar patterns of differences in pulse number thresholds in response to 2.5 kHz signals were also observed at TMRs of -12 dB and -6 dB, but none of these differences was statistically significant after correcting for multiple comparisons (Fig. 5e, Table 1). There were no significant differences in response latency between any of the three conditions for either the 1.25 kHz signal (Fig. 5c, Table S1) or the 2.5 kHz signal (Fig. 5e, Table S1).

Recall that female treefrogs readily exhibit phonotaxis in response to sounds recognized as the vocalizations of a suitable mate. In this experiment, the ability of 1.25 kHz signals to elicit phonotaxis responses was significantly reduced in the IM test conditions compared with the signal alone and EM control conditions. These effects are not easily explained as due to generally aversive effects of concurrent sounds on approach behavior or female preferences for quieter or simpler stimuli. Band-limited noise, for example, had small or negligible effects on phonotaxis responses in the EM control condition compared with the IM test condition. Neither of the concurrent sounds used in the IM test and EM control conditions was aversive when presented in the absence of a signal (*SI Appendix*, Fig. S2; see also Fig. 3c,d), a result that corroborates previous studies showing that female gray treefrogs are not repelled by pulse trains simulating heterospecific frogs[46], nor do they avoid noise[47] or inherently prefer stimuli that are quieter[48] or less complex[49]. We interpret this overall pattern of result as evidence that informational masking by temporally structured sounds at remote frequencies constrained signal recognition in a substantial proportion of subjects (~50%, Fig. 4b). Interestingly, subjects that exhibited phonotaxis had broadly similar pulse number thresholds and response latencies across many conditions, though pulse number thresholds were elevated in some IM test conditions.

Discussion

This study confirms that informational masking can disrupt the recognition of vocal signals in nonhuman animals. Two important implications stem from this conclusion. First, our results confirm that informational masking is not restricted to human hearing and speech communication. Rather, informational masking should be considered a general acoustic communication problem shared by humans and other animals, particularly those that also communicate in noisy social settings where concurrent sounds may be more likely to share features with signals of biological interest. Second, our results establish that informational masking can constrain signal recognition in evolutionarily important behavioral contexts. Vocal signals often play fundamental roles in the choice of a mate, which is one of the most consequential decisions many animals will ever make. In frogs, as in many other animals[16], mate choice decisions depend on a female's ability, in noisy

social environments, to recognize the vocalizations of a future mate as those of a conspecific and discriminating between his signals and those of his rival courters. By accounting for the potential effects of energetic masking, the present study illustrates how informational masking can disrupt the recognition of vocal signals underlying mate choice. Thus, from an evolutionary perspective, our results suggest that informational masking could act as a previously unrecognized but nevertheless potent source of selection on animal communication systems.

More broadly, this study expands current views on the constraints imposed by noise on animals. Studies of auditory masking in the context of acoustic signaling typically focus (either explicitly or implicitly) on energetic masking, even though informational masking is potentially more detrimental[23]. A few psychoacoustic studies have investigated informational masking and related phenomena using artificial sounds with nonhuman subjects[50, 51, 52, 53], and a few ecological studies have explored the idea that noise causes animals to shift their attention away from behaviorally important tasks[54, 55]. To our knowledge, however, all such studies have been conducted largely outside the naturalistic context of social communication. By demonstrating the potential for informational masking to interfere with vocal signal processing, this study opens several new avenues for future research to further advance our understanding of the mechanisms and evolution of animal hearing and acoustic communication.

At a mechanistic level, informational masking in humans is often attributed to failures of auditory stream segregation or constraints on selective attention[2, 3, 4]. We believe a breakdown in the ability of subjects to perceptually segregate target signals from informational maskers is unlikely to have contributed substantially to our results. Recall that in the IM test condition, the informational masker was designed to be separated in both frequency and time from the target signal to reduce any impacts of energetic masking. Some of the most potent cues promoting the segregation of sounds into different auditory streams in humans and other animals are frequency separation and temporal incoherence (e.g., differences in onset times)[56, 57]. Previous studies have shown that gray treefrogs also perceptually segregate concurrent call-like sounds based on frequency separation[34] and temporal incoherence[37], suggesting the effects of informational masking observed here are unlikely due to a breakdown of perceptual segregation. Instead, the observed effects may reflect constraints on bottom-up mechanisms that contribute to the attentional salience of a signal's informative temporal features.

Bottom-up attention involves stimulus-driven mechanisms by which information is processed involuntarily based on salient stimulus features[58, 59, 60]. In humans, for example, the temporal modulations present in competing speech can informationally mask target speech even when the competing speech is unintelligible[5]. The same effect can also occur when temporal modulations are imposed on broadband noise, even when the modulation rates differ by an octave or more[45]. Such results suggest temporal modulations in competing sounds are salient features that can interfere with bottom-up mechanisms for temporal processing. We found

broadly similar effects in gray treefrogs: temporal modulations in a masker present at a remote spectral frequency and an octave lower in modulation frequency (i.e., pulse rate) interfered with processing a temporally modulated signal. Given the importance of temporally structured signals in animals, informational masking due to constraints on bottom-up temporal processing may be taxonomically widespread, particularly in species that communicate in social environments awash with temporally structured acoustic signals, such as many frogs, insects, and songbirds.

In nonhuman animals with temporally structured acoustic signals, pulse rate is frequently a salient feature mediating signal recognition[28]. In gray treefrogs, the selective phonotaxis that females exhibit toward conspecific pulse rates[36, 38] depends on central processing within the inferior colliculus (IC). A key auditory processing and audio-motor integration area of the vertebrate brain, the IC functions in mediating acoustically guided behaviors with short latencies, such as responding to prey, predators, or conspecifics[61]. Small lesions of the gray treefrog IC nearly eliminate phonotaxis behavior[62]. In stark contrast, even extensive lesions of the gray treefrog thalamus, a brain area implicated in top-down attention to acoustic signals in frogs[62, 63], fail to eliminate selective phonotaxis based on pulse rate[62]. Importantly, single neurons in the gray treefrog IC exhibit stimulus-driven pulse-rate selectivity mirroring that observed in behavioral experiments[31, 32, 33]. This selectivity derives from so-called “interval counting” neurons in the IC, which produce action potentials after a threshold number of pulses are presented with regular temporal intervals that are optimal for each cell and generally characteristic of conspecific calls[31, 33, 64]. Interval-counting and pulse-rate tuning arise from the precise integration of time-dependent excitation and inhibition[65, 66]. Given the known patterns of excitatory and inhibitory frequency tuning in the frog IC[67], we hypothesize that an informational masker composed of slow-rate pulses in the range of the opposite inner ear papilla elicits additional excitation and inhibition that, depending on its temporal relation to the signal, disrupts the precisely timed integration of excitation and inhibition underlying the interval counting process that mediates signal recognition. More broadly, we hypothesize that disruptions of central computations underlying the processing of temporal features of vocalizations and other acoustic signals may be a form of bottom-up informational masking common to frogs, humans, and many other animals.

Another interesting finding from this study, which requires further investigation, is that informational masking in gray treefrogs depended on the spectral relationships between signals and maskers. Informational masking was more pronounced with the 1.25 kHz signal, for which the signal and masker frequencies primarily stimulated the inner ear papillae tuned to low (AP-range) and high (BP-range) frequencies, respectively. Several non-mutually exclusive explanations may account for this asymmetry. First, the AP is tonotopically organized and the BP is not[28]. Thus, the effects of randomizing the frequencies of pulses in the lower-frequency informational masker (0.7 to 1.4 kHz) may have diminished its overall ability to disrupt temporal processing at a central level by distributing its energy across multiple different tonotopic channels.

Second, frequencies below about 1.1 kHz fall outside the natural range of frequencies present in male calls[29] and are less effective at eliciting phonotaxis from females compared with frequencies between 1.1 and 4.0 kHz[34]. In contrast, all pulses in the higher-frequency informational masker (2.1 to 2.8 kHz) stimulated the same single channel of the BP and occurred at frequencies that elicit phonotaxis[34]. Finally, although gray treefrogs can process temporal information related to pulse rate in sounds transduced by either the AP or BP, a frequency-dependent bias often favors signals that primarily stimulate one or the other sensory papilla[34, 39, 40, 41]. In Minnesota, where this study was conducted, female phonotaxis is biased toward frequencies that fall within the higher frequency range of the BP[39, 40]. Together, these considerations suggest a perceptual or attentional bias toward processing temporal information in the BP channel may have contributed to the greater susceptibility to informational masking observed when the informational masker was in the frequency range of the BP. Additional research is needed to resolve these issues.

Future studies will be necessary to investigate differences in the types of constraints energetic and informational masking impose on animals, particularly those that, like humans and frogs, frequently communicate in noisy social environments. It also will be important to determine how such constraints independently and synergistically influence the evolution of signal design and the mechanisms underlying signaling behavior and signal perception. Given the growing awareness of the diverse negative impacts of anthropogenic noise on animals[23, 68], an important next step would be to disentangle the potential impacts of both energetic and informational forms of masking caused by anthropogenic noise. Because frogs are a well-established vertebrate model in research on vocal communication and auditory neuroethology[28], future work on frogs is particularly well suited to address the existing dearth of studies investigating the neural basis of informational masking at a cellular level in subcortical areas of the vertebrate brain[10].

Materials and Methods

Subjects. All subjects were wild-caught gravid females of *H. chrysoscelis* collected from local wetlands. In total, 416 females were collected and tested in our experiments and released at their collection site after the completion of testing. All animal procedures were approved by the University of Minnesota's Institutional Animal Care and Use Committee (1701-34456A and 2001-37746A). See the Supporting Information (*SI Appendix*) for additional details.

Acoustic Stimuli. All signals and maskers were synthesized in MATLAB R2018a (Mathworks, Natick, MA, USA) at a sampling rate of 44.1 kHz (16 bit). Each target signal was a pulse train consisting of between 1 and 30 identical, equally spaced pulses that were modeled after a typical pulse in natural advertisement calls recorded from our study populations (Fig. 1a)[29]. The carrier frequency of each pulse within a target signal was fixed at either 1.25 kHz (AP range) or 2.5 kHz (BP range), which are close to the average values of the two harmonically related spectral components present in natural calls[29]. Pulses within each target signal were shaped to have

species typical rise and fall times (3.1-ms inverse exponential rise time; 5.4-ms exponential fall time) and were delivered at a rate of 50 pulses/s (10-ms pulse duration, 20-ms pulse period, and 50% pulse duty cycle). During a phonotaxis test, each target signal was repeated at a rate of 11 signals/min, which approximates the average call rate in local populations[29].

In the IM test conditions (Figs 1b, 3a) and IM control conditions (Fig. 3a), each masker consisted of a repeated sequence of 30 pulses produced at a rate of 25 pulses/s (10-ms pulse duration, 40-ms pulse period, and 25% pulse duty cycle). The temporal features of each pulse were identical to those used in the target signal. The carrier frequency of each pulse in the informational masker was randomly chosen from among 13 frequencies spanning a range of 0.7 kHz. For the 2.5 kHz signal (BP range), the masker's frequencies were spaced one semitone apart between 0.7 kHz and 1.4 kHz (0.7, 0.742, 0.786, 0.832, 0.882, 0.934, 0.99, 1.049, 1.111, 1.177, 1.247, 1.321, and 1.4 kHz). These frequencies were chosen to fall within the frequency range of the AP but below the minimum frequency to which the BP responds (~1.5 to 1.625 kHz)[42]. For the 1.25 kHz signal (AP range), the masker's 13 frequencies had the same absolute spacing (in kHz) but were shifted upward in frequency by 1.4 kHz so that they fell between 2.1 and 2.8 kHz (2.1, 2.142, 2.186, 2.232, 2.282, 2.334, 2.390, 2.449, 2.511, 2.577, 2.647, 2.721, and 2.8 kHz). These masker frequencies fell within the range of the BP but were chosen to be above the maximum frequency to which the AP responds (~1.625 to 1.75 kHz)[42]. In the IM test condition, the pulses of the informational masker were interleaved with those of the target signal such that each masker pulse occurred precisely between consecutive pairs of signal pulses. In the IM control condition, the timing of the masker was shifted by 10 ms relative to the IM test condition so that each masker pulse occurred synchronously with a pulse in the target signal. In the EM control conditions (Figs 1c, 3a), each masker consisted of an independently generated, band-limited noise having the same duration and bandwidth (0.7 kHz to 1.4 kHz or 2.1 kHz to 2.8 kHz) as the corresponding informational masker. In all masking conditions, a masker was gated on and off with each presentation of a concurrent target signal such that each presentation of a target signal was temporally centered in a masker, apart from the 10-ms shift of the masker in the IM control condition. The concurrently occurring signal and masker was presented repeatedly 11 times/min. Each separate rendition of a masker was randomized in the sense that the frequency of its pulses (for the informational maskers) or its temporal envelop (for the bandlimited noise) was determined randomly so that the same combination of target signal and masker never repeated in the same test of a given subject. The sound pressure level (SPL re 20 μPa) of all maskers used in this study were calibrated to 75 dB (LC_{eq}) at 1 m. This value falls within the range of background noise levels measured in gray treefrog choruses[35].

Testing Protocol. We conducted phonotaxis experiments using a circular arena inside an acoustic chamber (Fig. 2, Movie S1). At the beginning of a phonotaxis test, we separated the subject from her mate and placed her in a small acoustically transparent release cage (9-cm

diameter, 2-cm height) located on the center of the arena floor. We allowed the subject to acclimate in the cage for 30 sec, after which we started the playback. At this time, we remotely lifted the lid of the release cage using a rope-and-pulley system and allowed the subjects to move freely in the arena. A test continued until the subject responded to a stimulus by entering a response zone consisting of a 10-cm semi-circle in front of a playback speaker or until the subject was considered to have failed to respond to a stimulus. Our criteria for determining failure to respond were the following: (1) if the subject did not exit the release cage within 3 min; (2) if the subject did not enter a response zone within 5 minutes; (3) if the subject's first physical contact with the arena wall was in the semi-circle opposite the speaker in a single-stimulus, no-choice test or in the quadrant opposite that formed by the two speakers in a two-alternative choice test. Movements of subjects inside the arena were observed under infrared light (Tracksys, Ltd., Nottingham, UK) and scored in real-time by an observer outside the chamber by means of a video camera and monitor. Additional details regarding procedures are reported in the Supplementary Information (see *SI Appendix*).

Statistical Analyses. For all statistical analyses, we used a significance criterion of $\alpha = 0.05$, and within each experiment we used the Holm's Bonferroni sequential correction for multiple comparisons. We analyzed the outcomes of two-alternative choice tests comparing responses to two identical 20-pulse signals, only one of which was presented with a concurrent sound, using two-tailed binomial tests of the null hypothesis that an equal proportion (0.5) of subjects would choose each alternative. To compare differences in choices between masking conditions, Generalized Estimating Equations (GEE) were used. The responses of no-choice tests of signal recognition were analyzed using a combination of GEE and nonparametric statistics. Different GEE models including different sets of predictor variables were compared using ANOVA. Additional details regarding statistical analyses are reported below.

Two-alternative Choice Tests with Identical 20-pulse Signals. We assessed the impacts of concurrent sounds on phonotaxis using a series of two-alternative choice tests[69] (Fig. 2b, Movie S1). Each subject ($N = 73$) was randomly assigned to one of the two signal frequencies (1.25 or 2.5 kHz) and tested in three phonotaxis tests conducted in a randomized order. In these choice tests, subjects chose between two identical target signals, each having 20 pulses. One sequence of target signals was always presented with no additional concurrent sound (signal alone). In the alternative sequence of target signals, each signal was presented concurrently with another sound designed to be a potential masker. In one test (IM test condition), each target signal in the alternative sequence was presented concurrently with the informational masker such that the pulses of the masker were interleaved between consecutive pairs of signal pulses. In a second test (IM control condition), each target signal in the alternative sequence was presented concurrently with the informational masker, the timing of which had been shifted by 10 ms (relative to the IM test condition) such that each pulse of the masker was synchronous with a

pulse in the target signal. In the third test (EM control condition), each target signal in the alternative sequence was presented concurrently with a burst of band-limited noise. For each presentation, signals from the two sequences were presented alternatively in time. Across tests, all signals and maskers were broadcast at 75 dB SPL, which yielded a target-to-masker ratio (TMR) of 0 dB. Tests in which subjects did not make a choice between identical target signals (tests 1-3) were excluded from the final analysis (13 of 219 tests).

Subject choice was scored as a binary response variable (signal alone = 1, signal with concurrent sound = 0) and the proportion choosing a particular stimulus was analyzed using two-tailed binomial tests of the null hypothesis that subjects would choose randomly (expected proportions of 0.5:0.5). For clarity, these proportions are expressed as percentages. To compare stimulus choice between the IM test, IM control, and EM control conditions, we fitted GEE models with logit link functions and exchangeable correlations. We initially fitted a model with and without target signal frequency as a predictor variable. Because there was no significant difference in choice across the two signal frequencies, we pooled data across the two signal frequencies, and fitted a single model that included a single predictor variable of masking condition (IM test, IM control, and EM control), and subject ID was used for clustering to account for repeated measures.

No-choice Tests of Signal Recognition. We assessed signal recognition using a series of single-stimulus, no-choice tests[69] (Fig. 2c) and an adaptive tracking procedure[31] to measure pulse number thresholds, which we operationally defined as the minimum number of pulses required to elicit positive phonotaxis. We measured two pulse number thresholds for each subject, one in a signal alone condition and one in a masking condition, with the order of threshold measurement determined randomly. Subjects were assigned randomly to either the IM test or EM control condition, one of the two signal frequencies (1.25 kHz or 2.5 kHz), and to one of three TMRs (-12 dB, -6 dB, or 0 dB). The amplitude of the masker was fixed at 75 dB SPL, and signal levels of 63 dB, 69 dB, and 75 dB SPL were used, respectively, to achieve nominal TMRs of -12 dB, -6 dB, and 0 dB. Throughout, we use TMRs consistently to describe differences in signal level, even for the signal alone condition, in which there was technically no masker.

We followed procedures described by Gupta et al.[31] to measure pulse number thresholds from 323 subjects. In the first of a series of tests, we presented subjects with a sequence of target signals consisting of 8 pulses. Depending on whether the subject responded to the target signal, we either decreased or increased the number of pulses in the target signal by two pulses in the subsequent test. If the subject responded to the sequence of 8-pulse target signal on this test, they heard a sequence of 6-pulse signals in their next test; subjects that failed to respond to the 8-pulse target signal heard a sequence of 10-pulse target signals in their next test. This process continued until the subject's response changed – from a response to a no response or from no response to a response – between two consecutive tests. After this change in behavior, we

performed a final test in which we reversed the direction of change and either increased or decreased the pulse number by one pulse based on whether it responded to the target signal in the immediately preceding test. The pulse number threshold was calculated by averaging the lowest pulse number that elicited a response and the highest pulse number that did not. The maximum number of pulses used in any target signal was 20 pulses. If the subject did not respond to a target signal with 20 pulses, we did not determine its pulse number threshold. Restricting the maximum number of pulses to 20 was necessary to ensure we could measure two pulse number thresholds per subject, which typically do not respond in more than about 15 to 20 phonotaxis tests before losing response motivation. After determining a subject's first pulse number threshold we immediately restarted the adaptive tracking procedure to determine their second threshold. In addition, we determined the latency to respond to each target signal having the lowest number of pulses that elicited a response. If at any time during testing, a subject failed to respond in three consecutive tests, we performed an additional test in which a sequence of attractive synthetic calls having the average properties of gray treefrog calls were presented. Only subjects that exhibited positive phonotaxis in this test were tested further; two subjects were excluded from the final dataset because they did not meet this inclusion criterion.

Procedures for determining pulse number thresholds in the signal alone condition and in the two masking conditions were the same with the following exception. In the masking condition, subjects experienced a "sham" test prior to commencing the adaptative tracking procedure to determine their masked pulse number threshold. In this sham test, the masker was presented alone without the target signal. That some subjects would respond to the masker alone was not unexpected given that some gravid females will, in the absence of a target signal with conspecific call properties, respond to other types of sounds with some relevant signal features, such as the appropriate spectral content, including bandlimited noise[47], unmodulated tones[70], and signals with slower or faster pulse rate typical of a heterospecific male[27]. To obtain accurate pulse number thresholds in response to target signals using the adaptive tracking procedure, it was, therefore, desirable to exclude subjects that responded to maskers. We excluded 46 subjects from determination of a pulse number threshold because they responded to the masker in the relevant sham condition (Fig. S2); 43 of 186 subjects (23.1%) assigned to the tests with the 1.25 kHz signal (AP range) responded to hearing only the masker (BP range) in the sham condition; 3 of 135 subjects (2.2%) assigned to tests with the 2.5 kHz signal (BP range) responded to the masker (AP range) in the sham condition. Therefore, our final sample size in this experiment was 275 subjects.

We determined response rate by coding each subject's response as binary. If a subject's pulse number threshold was 20 or fewer pulses in a particular condition, it was assigned a response score of 1. Otherwise, it was assigned a score of 0. We used GEEs with logit link functions and exchangeable correlations to analyze this binary response variable. An initial

inspection of the data revealed large effects of target signal frequency, which we confirmed by comparing two GEE models with and without signal frequency as a predictor variable for each of the three response variables (response rate, pulse number threshold, and response latency; see Main Text). Because these effects were of secondary interest, we analyzed the effects of masking condition and TMR separately for each signal frequency. For analyses of response rate with the 1.25 kHz signal, the initial GEE model included masking condition (signal alone, IM test, and EM control), TMR (-12 dB, -6 dB, and 0 dB), and the masking condition × TMR interaction as predictor variables. Subject ID was used for clustering to account for repeated measures. We fitted a second GEE model in which we excluded the masking condition × TMR interaction. We compared the two models using ANOVA and found that they were not statistically different. We fitted a third GEE model in which we additionally excluded TMR as a predictor variable and compared this model with the second model using ANOVA and found that they also were not statistically different. For analyses of response rate with the 2.5 kHz signal, we could not fit GEE models because the response rate of subjects was high across all conditions, reaching 100% in seven of the nine factorial combinations of masking condition and TMR (Fig. 4d). The lack of variance in the data led to the problem of complete separation, producing spurious results for the main effects model.

We statistically compared pulse number thresholds and response latencies for each subject that responded in both the signal alone condition and in the presence of their assigned masking condition. This paired data allowed us to evaluate how the presence of each masker influenced responses relative to a signal alone condition in the same group of subjects. However, because these data were not normally distributed and could not be successfully transformed to follow a normal distribution, we used non-parametric tests for hypothesis testing. To make pairwise comparisons, we used a paired-sample Wilcoxon tests at each TMR to compare separately the pulse number thresholds and response latencies in the signal alone condition and the paired masking condition. An independent sample Wilcoxon test was used to compare the pulse number thresholds and response latencies between subjects that responded in the IM test and EM control conditions at each TMR.

Acknowledgments

We thank Paloma Gonzales-Bellido, Katie Krueger, Andrew Oxenham, Annika Rupert, and Emilie Snell-Rood for feedback on the manuscript; Andrew Oxenham for insightful conversations about informational masking; Mridul Yadav for making the 3D schematic of the phonotaxis arena; all members of the Bee Lab 2018-2022 for their help collecting and testing frogs; the Minnesota Department of Natural Resources for permission to collect frogs; and the Three Rivers Park District and the Ramsey County Department of Parks and Recreation for access to field sites. This work was supported by a grant from the U.S. NSF to M.A.B. and G.J.R. (IOS-2022253).

Author Contributions: S.G., G.J.R., and M.A.B. designed research; S.G. and L.K. performed research; S.G. analyzed data; S.G., L.K., G.J.R., and M.A.B. wrote and edited the manuscript.

Competing Interest Statement: The authors declare no competing interest.

References

1. Cherry EC. Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America* **25**, 975-979 (1953).

2. Durlach NI, Mason CR, Kidd G, Arbogast TL, Colburn HS, Shinn-Cunningham BG. Note on informational masking (L). *Journal of the Acoustical Society of America* **113**, 2984-2987 (2003).

3. Kidd G, Mason CR, Richards VM, Gallun FJ, Durlach NI. Informational masking. In: *Auditory Perception of Sound Sources* (ed^(eds Yost WA, Popper AN, Fay RR). Springer (2008).

4. Kidd G, Colburn HS. Informational masking in speech recognition. In: *The Auditory System at the Cocktail Party* (ed^(eds Middlebrooks JC, Simon JZ, Popper AN, Fay RR). Springer (2017).

5. Carlile S, Corkhill C. Selective spatial attention modulates bottom-up informational masking of speech. *Scientific Reports* **5**, 8662 (2015).

6. Ihlefeld A, Shinn-Cunningham B. Spatial release from energetic and informational masking in a selective speech identification task. *Journal of the Acoustical Society of America* **123**, 4369-4379 (2008).

7. Summers RJ, Roberts B. Informational masking of speech by acoustically similar intelligible and unintelligible interferers. *Journal of the Acoustical Society of America* **147**, 1113-1125 (2020).

8. Oxenham AJ. How we hear: The perception and neural coding of sound. *Annual Review of Psychology* **69**, 27-50 (2018).

9. Dollezal LV, Tolnai S, Beutelmann R, Klump GM. Release from informational masking by auditory stream segregation: Perception and its neural correlate. *European Journal of Neuroscience* **51**, 1242-1253 (2020).

10. Snyder JS, Elhilali M. Recent advances in exploring the neural underpinnings of auditory scene perception. *Annals of the New York Academy of Sciences* **1396**, 39-55 (2017).

11. Römer H. Masking by noise in acoustic insects: Problems and solutions. In: *Animal Communication and Noise* (ed^(eds Brumm H). Springer (2013).

12. Bee MA. Treefrogs as animal models for research on auditory scene analysis and the cocktail party problem. *International Journal of Psychophysiology* **95**, 216-237 (2015).

13. Aubin T, Jouventin P. Cocktail-party effect in king penguin colonies. *Proceedings of the Royal Society of London Series B-Biological Sciences* **265**, 1665-1673 (1998).

14. Klump GM. Bird communication in the noisy world. In: *Ecology and Evolution of Acoustic Communication in Birds* (ed^(eds Kroodsma DE, Miller EH). Cornell University Press (1996).

15. Hase K, Kadoya Y, Maitani Y, Miyamoto T, Kobayasi KI, Hiryu S. Bats enhance their call identities to solve the cocktail party problem. *Communications Biology* **1**, 39 (2018).

16. Bee MA, Micheyl C. The cocktail party problem: What is it? How can it be solved? And why should animal behaviorists study it? *Journal of Comparative Psychology* **122**, 235-251 (2008).

17. Lohr B, Wright TF, Dooling RJ. Detection and discrimination of natural calls in masking noise by birds: Estimating the active space of a signal. *Animal Behaviour* **65**, 763-777 (2003).

18. Caldwell MS, Bee MA. Spatial hearing in Cope's gray treefrog: I. Open and closed loop experiments on sound localization in the presence and absence of noise. *Journal of Comparative Physiology A* **200**, 265-284 (2014).

19. Lee N, Ward JL, Vélez A, Micheyl C, Bee MA. Frogs exploit statistical regularities in noisy acoustic scenes to solve cocktail-party-like problems. *Current Biology* **27**, 743-750 (2017).

20. Ward JL, Buerkle NP, Bee MA. Spatial release from masking improves sound pattern discrimination along a biologically relevant pulse-rate continuum in gray treefrogs. *Hearing Research* **306**, 63-75 (2013).

21. Brumm H, Slabbekoorn H. Acoustic communication in noise. *Advances in the Study of Behavior* **35**, 151-209 (2005).

22. Brumm H. Animal Communication and Noise. In: *Animal Signals and Communication* (ed^(eds Janik VM, McGregor PK). Springer (2013).

23. Rosa P, Koper N. Integrating multiple disciplines to understand effects of anthropogenic noise on animal communication. *Ecosphere* **9**, e02127 (2018).

24. Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science* **270**, 303-304 (1995).

25. Hopkins CD, Bass AH. Temporal coding of species recognition signals in an electric fish. *Science* **212**, 85-87 (1981).

26. Lee N, Kirtley AT, Pressman IS, Jirik KJ, Koucoulas D, Mason AC. Developing a phonotaxis performance index to uncover signal selectivity in walking phonotaxis. *Frontiers in Ecology and Evolution* **7**, 334 (2019).

27. Gerhardt HC. Acoustic communication in two groups of closely related treefrogs. *Advances in the Study of Behavior* **30**, 99-167 (2001).

28. Gerhardt HC, Huber F. *Acoustic communication in insects and anurans: Common problems and diverse solutions*. Chicago University Press (2002).

29. Ward JL, Love EK, Vélez A, Buerkle NP, O'Bryan LR, Bee MA. Multitasking males and multiplicative females: Dynamic signalling and receiver preferences in Cope's grey treefrog. *Animal Behaviour* **86**, 231-243 (2013).

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

31. Gupta S, Alluri RK, Rose GJ, Bee MA. Neural basis of acoustic species recognition in a cryptic species complex. *Journal of Experimental Biology* **224**, jeb243405 (2021).

32. Rose GJ, Brenowitz EA, Capranica RR. Species specificity and temperature dependency of temporal processing by the auditory midbrain of two species of treefrogs. *Journal of Comparative Physiology A* **157**, 763-769 (1985).

33. Rose GJ, Hanson JL, Leary CJ, Graham JA, Alluri RK, Vasquez-Opazo GA. Species-specificity of temporal processing in the auditory midbrain of gray treefrogs: Interval-counting neurons. *Journal of Comparative Physiology A* **201**, 485-503 (2015).

34. Nityananda V, Bee MA. Finding your mate at a cocktail party: Frequency separation promotes auditory stream segregation of concurrent voices in multi-species frog choruses. *PLoS ONE* **6**, e21191 (2011).

35. Tanner JC, Bee MA. Within-individual variation in sexual displays: Signal or noise? *Behavioral Ecology* **30**, 80-91 (2019).

36. Bush SL, Gerhardt HC, Schul J. Pattern recognition and call preferences in treefrogs (Anura: Hylidae): A quantitative analysis using a no-choice paradigm. *Animal Behaviour* **63**, 7-14 (2002).

37. Gupta S, Bee MA. Treefrogs exploit temporal coherence to form perceptual objects of communication signals. *Biology Letters* **16**, 20200573 (2020).

38. Schul J, Bush SL. Non-parallel coevolution of sender and receiver in the acoustic communication system of treefrogs. *Proceedings of the Royal Society of London Series B-Biological Sciences* **269**, 1847-1852 (2002).

39. Bee MA. Spectral preferences and the role of spatial coherence in simultaneous integration in gray treefrogs (*Hyla chrysoscelis*). *Journal of Comparative Psychology* **124**, 412-424 (2010).

40. Gupta S, Bee MA. Female preferences for the spectral content of advertisement calls in Cope's gray treefrog (*Hyla chrysoscelis*). *Journal of Comparative Physiology A*, (2023, in press).

41. Gerhardt HC. Phonotactic selectivity in two cryptic species of gray treefrogs: Effects of differences in pulse rate, carrier frequency and playback level. *Journal of Experimental Biology* **211**, 2609-2616 (2008).

42. Schrode KM, Buerkle NP, Brittan-Powell EF, Bee MA. 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 (2014).

43. Neff DL, Green DM. Masking produced by spectral uncertainty with multicomponent maskers. *Percept Psychophys* **41**, 409-415 (1987).

44. Schwartz JJ, Buchanan BW, Gerhardt HC. Acoustic interactions among male gray treefrogs, *Hyla versicolor*, in a chorus setting. *Behavioral Ecology and Sociobiology* **53**, 9-19 (2002).

45. Conroy C, Kidd Jr G. Informational masking in the modulation domain. *Journal of the Acoustical Society of America* **149**, 3665-3673 (2021).

46. Gerhardt HC, Dyson ML, Tanner SD, Murphy CG. Female treefrogs do not avoid heterospecific calls as they approach conspecific calls: Implications for mechanisms of mate choice. *Animal Behaviour* **47**, 1323-1332 (1994).

47. Vélez A, Gordon NM, Bee MA. The signal in noise: Acoustic information for soundscape orientation in two North American tree frogs. *Behavioral Ecology* **28**, 844-853 (2017).

48. Bee MA, Vélez A, Forester JD. Sound level discrimination by gray treefrogs in the presence and absence of chorus-shaped noise. *Journal of the Acoustical Society of America* **131**, 4188-4195 (2012).

49. Seeba F, Schwartz JJ, Bee MA. Testing an auditory illusion in frogs: Perceptual restoration or sensory bias? *Animal Behaviour* **79**, 1317-1328 (2010).

50. Branstetter BK, Bakhtiari K, Black A, Trickey JS, Finneran JJ, Aihara H. Energetic and informational masking of complex sounds by a bottlenose dolphin (*Tursiops truncatus*). *Journal of the Acoustical Society of America* **140**, 1904-1917 (2016).

51. Eipert L, Klump GM. Interaction of spatial source separation, fundamental frequency, and vowel pairing in a sequential informational masking paradigm in Mongolian gerbils. *Behavioral Neuroscience* **134**, 119-132 (2020).

52. Cai H, Dent ML. Attention capture in birds performing an auditory streaming task. *PLoS ONE* **15**, e0235420 (2020).

53. Cai H, Dent ML. Dimensionally specific attention capture in birds performing auditory streaming task. *Journal of the Association for Research in Otolaryngology* **23**, 241-252 (2022).

54. Allen LC, Hristov NI, Rubin JJ, Lightsey JT, Barber JR. Noise distracts foraging bats. *Proc R Soc B-Biol Sci* **288**, (2021).

55. Chan AAY-H, Giraldo-Perez P, Smith S, Blumstein DT. Anthropogenic noise affects risk assessment and attention: The distracted prey hypothesis. *Biology Letters* **6**, 458-461 (2010).

56. Bregman AS. *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press (1990).

57. Dent ML, Bee MA. Principles of auditory object formation by nonhuman animals. In: *Effects of Anthropogenic Noise on Animals* (ed^(eds Slabbekoorn H, Dooling RJ, Popper AN, Fay RR). Springer (2018).

58. Conway AR, Cowan N, Bunting MF. The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review* **8**, 331-335 (2001).

59. Kaya EM, Elhilali M. Modelling auditory attention. *Philosophical Transactions of the Royal Society B-Biological Sciences* **372**, (2017).

60. Katsuki F, Constantinidis C. Bottom-up and top-down attention: Different processes and overlapping neural systems. *The Neuroscientist* **20**, 509-521 (2014).

61. Casseday JH, Covey E. A neuroethological theory of the operation of the inferior colliculus. *Brain Behav Evol* **47**, 311-336 (1996).

62. Endepols H, Feng AS, Gerhardt HC, Schul J, Walkowiak W. Roles of the auditory midbrain and thalamus in selective phonotaxis in female gray treefrogs (*Hyla versicolor*). *Behav Brain Res* **145**, 63-77 (2003).

63. Ponnath A, Farris HE. Sound-by-sound thalamic stimulation modulates midbrain auditory excitability and relative binaural sensitivity in frogs. *Frontiers in neural circuits* **8**, 85 (2014).

64. Edwards CJ, Alder TB, Rose GJ. Auditory midbrain neurons that count. *Nature Neuroscience* **5**, 934-936 (2002).

65. Edwards CJ, Leary CJ, Rose GJ. Counting on inhibition and rate-dependent excitation in the auditory system. *Journal of Neuroscience* **27**, 13384-13392 (2007).

66. Rose GJ, Leary CJ, Edwards CJ. Interval-counting neurons in the anuran auditory midbrain: Factors underlying diversity of interval tuning. *Journal of Comparative Physiology A* **197**, 97-108 (2011).

67. Hall JC. GABAergic inhibition shapes frequency tuning and modifies response properties in the auditory midbrain of the leopard frog. *Journal of Comparative Physiology A* **185**, 479-491 (1999).

68. Slabbekoorn H. Noise pollution. *Current Biology* **29**, R957-R960 (2019).

69. Gerhardt HC. Phonotaxis in female frogs and toads: Execution and design of experiments. In: *Methods in Comparative Psychoacoustics* (ed^(eds Klump GM, Dooling RJ, Fay RR, Stebbins WC). Birkhäuser Verlag (1995).

70. Kuczynski MC, Vélez A, Schwartz JJ, Bee MA. Sound transmission and the recognition of temporally degraded sexual advertisement signals in Cope's gray treefrog (*Hyla chrysoscelis*). *Journal of Experimental Biology* **213**, 2840-2850 (2010).

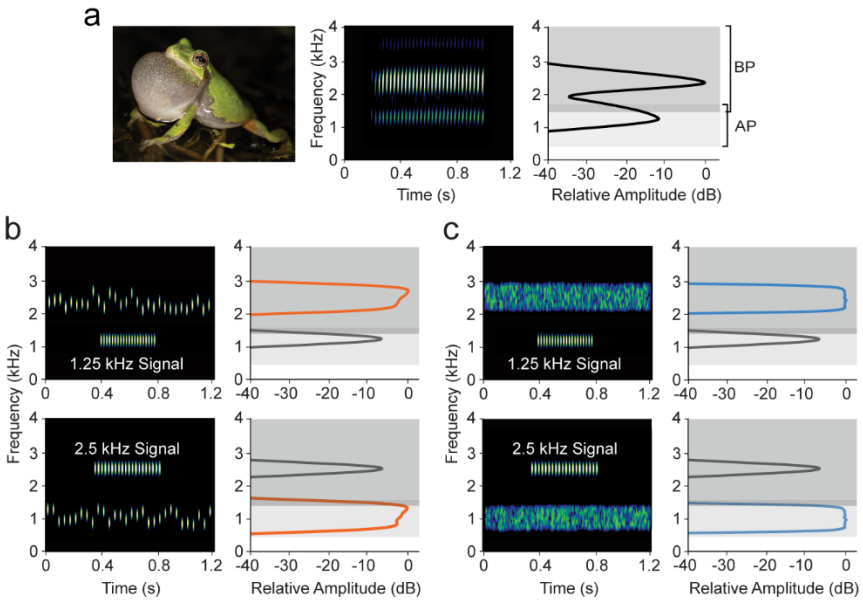

Fig. 1. Acoustic signals and stimuli used to investigate informational masking. (a) Picture (*Left*) of a calling male *Hyla chrysoscelis* (photo credit: Alex Baugh), (*Center*) spectrogram illustrating the pulsatile temporal pattern and frequency composition of a natural advertisement call, and (*Right*) power spectrum illustrating the two spectral peaks of the advertisement call relative to the approximate frequency ranges of the basilar papilla (BP; dark gray) and amphibian papilla (AP; light gray). Based on recordings of the auditory brainstem response in gray treefrogs[42], the lower spectral component (~1.25 kHz) is transduced primarily by the AP (range of sensitivity ~0.5 to 1.75 kHz), and the higher component (~2.5 kHz) is transduced primarily by the BP (range of sensitivity ~1.5 to 4.0 kHz). (b,c) Spectrograms (*Left*) and power spectra (*Right*) of target signals and maskers used in (b) the IM test conditions and (c) the EM control conditions (*Bottom*). The carrier frequency of the target signal was fixed at either (*Top*) 1.25 kHz or (*Bottom*) 2.5 kHz so that it would primarily stimulate only one of the two inner ear papillae. The frequency ranges of maskers were designed to fall within the sensitivity range of the inner ear papilla that was not stimulated by the target signal. In the IM test conditions (b), the masker was a series of random-frequency pulses (25 pulses/s) that were interleaved between every other pulse in a target signal (50 pulses/s). In the EM control condition (c), the masker was a narrowband noise having the same spectral content and intensity as the masker used in the IM test condition. In both masker conditions, target signals were temporally centered within the masker, which was gated on and off around the signal. In signal alone conditions (not shown), target signals were presented without maskers.

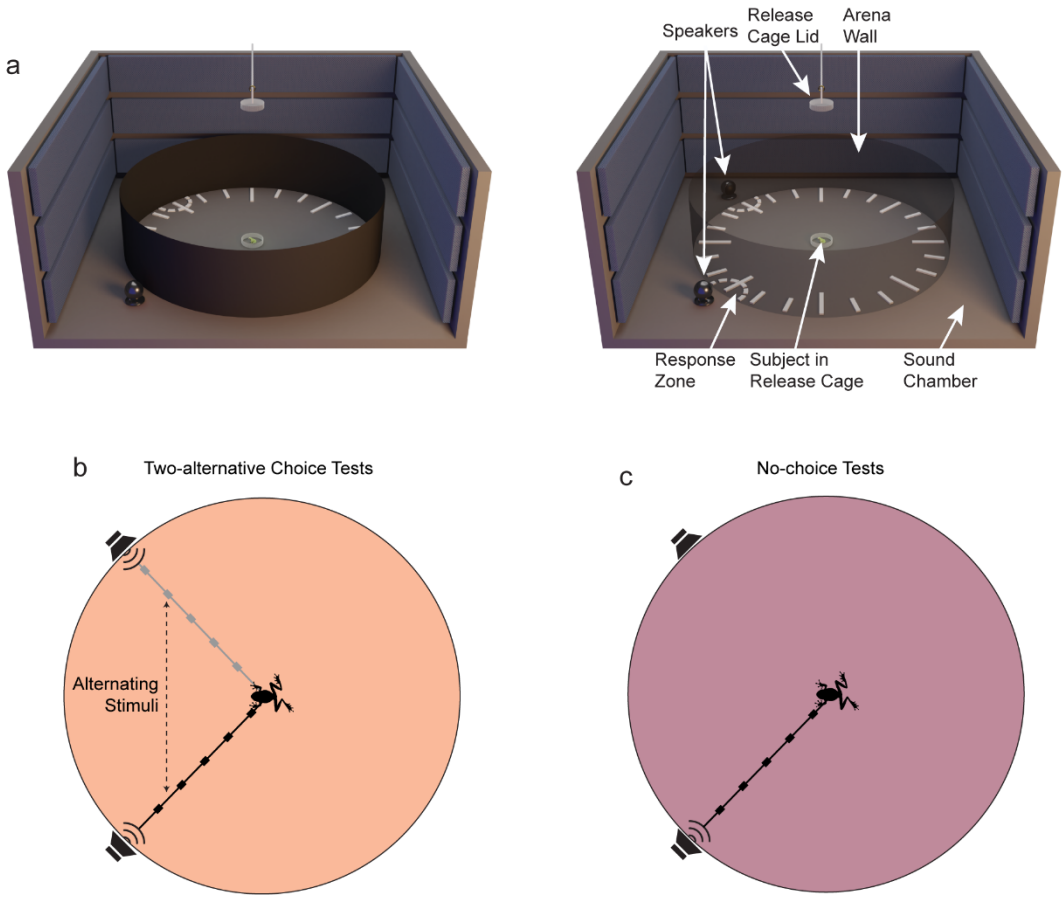

Fig. 2. Schematic illustrations of experimental setups and testing protocols. (a) A 3D schematic of the circular test arena used to conduct phonotaxis tests with female treefrogs shown with visually opaque walls (*Left*) with walls rendered transparent (*Right*) to facilitate visualization of both playback speakers. The test arena (2-m diameter, 60-cm height) was located inside a semi-anechoic chamber and was made from hardware cloth covered in black fabric. It was acoustically transparent but visually opaque. Acoustic stimuli were broadcast from one or two speakers positioned on the floor outside the arena wall and pointed toward a subject release cage at the center of the arena. The lid of the release cage could be remotely lifted from outside the chamber to allow the subject to respond by approaching a response zone. A response was scored when the subject entered a response zone. Responses were recorded using an infrared-sensitive camera mounted above the center of the test arena (not shown) and viewed in real time by one or two experimenters on a monitor located outside the arena. (b,c) Overhead schematics (not to scale) of the testing arena illustrating a two-alternative choice test (*B*) and a single stimulus, no-choice test (c). (b) In two-alternative choice tests, two sequences of repeating stimuli (e.g., two signals, or a signal alone and a signal plus masker) were presented from speakers separated by 90° around the test arena; stimuli in the two sequences alternated in time such that there were equal intervals of silence preceding and following each stimulus presentation (see Movie S1). To minimize any side bias, we randomized which speaker broadcast which stimulus across subjects. Alternating stimulus sequences are illustrated by *gray* and *black* schematic waveforms. (c) In no-choice tests, the stimulus sequence consisted of a single repeated stimulus.

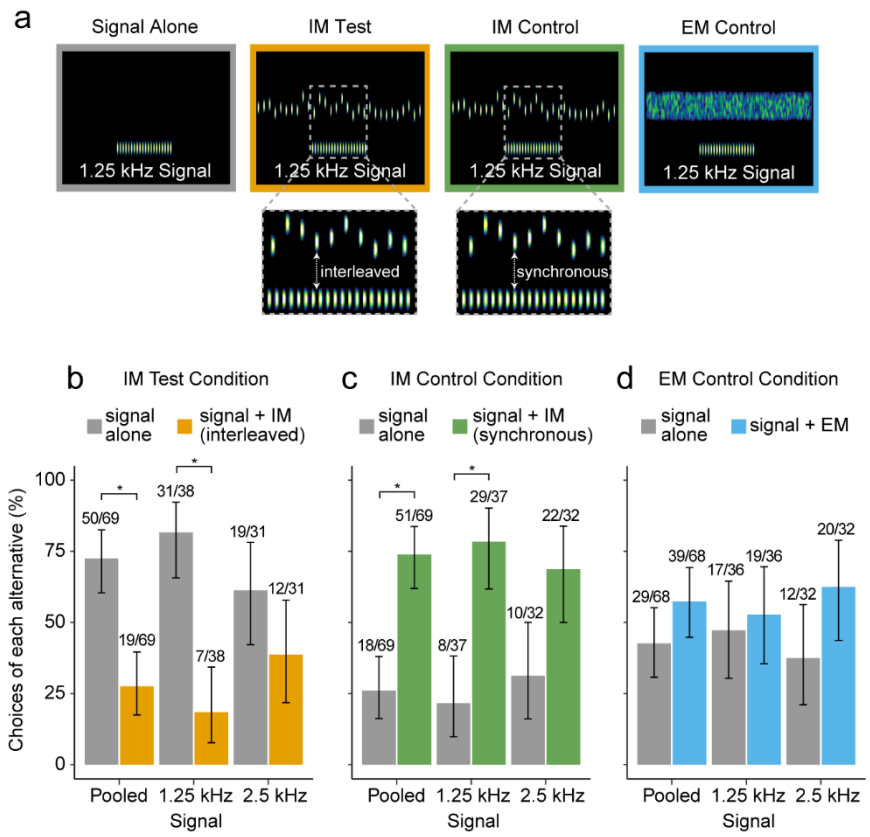

Fig. 3. Informational masking alters choices between identical signals in two-alternative choice tests. (a) Spectrograms illustrating the 1.25 kHz target signal with 20 pulses in the signal alone condition (*Left*), IM test condition (*Center Left*), IM control condition (*Center Right*), and EM control condition (*Right*), depicted by *gray*, *orange*, *green*, and *blue* colors, respectively. All signals and maskers were broadcast at 75 dB SPL for a target-to-masker ratio (TMR) of 0 dB. Across different choice tests, subjects were allowed to choose between two identical, alternating signals, one presented without concurrent sound (signal alone) and the other presented with a concurrent sound in one of the three masking conditions (IM test, IM control, or EM control conditions). (b-d) Percentages (± 95% exact binomial confidence intervals) of subjects that chose each target signal. Data are shown pooled across both signal frequencies and separately for the 1.25 kHz and 2.5 kHz signals. Fractions above each pair of bars show the number of subjects that chose each stimulus over the number of subjects tested. (b) In the IM test condition, in which masker pulses were interleaved between signal pulses, there were significant (*$P < 0.05$) preferences for the signal alone (*gray bars*) over the same signal presented with a concurrent informational masker (*orange bars*) in pooled responses and for the 1.25 kHz signal, but not for the 2.5 kHz signal. (c) In the IM control condition, in which the masker pulses were synchronous with signal pulses, there were significant preferences for the signal presented with a concurrent informational masker (*green bars*) over the same signal presented alone (*gray bars*) in pooled responses and for the 1.25 kHz signal, but not for the 2.5 kHz signal. (d) In the EM control condition, there were no significant differences from chance expectation (50%) in the percentages of subjects that chose the signal alone (*gray bars*) and the signal in the EM control condition (*blue bars*) in which the signal was presented with a concurrent band-limited noise having the same duration, amplitude, and frequency spectrum as the informational masker used in the IM test and IM control conditions.

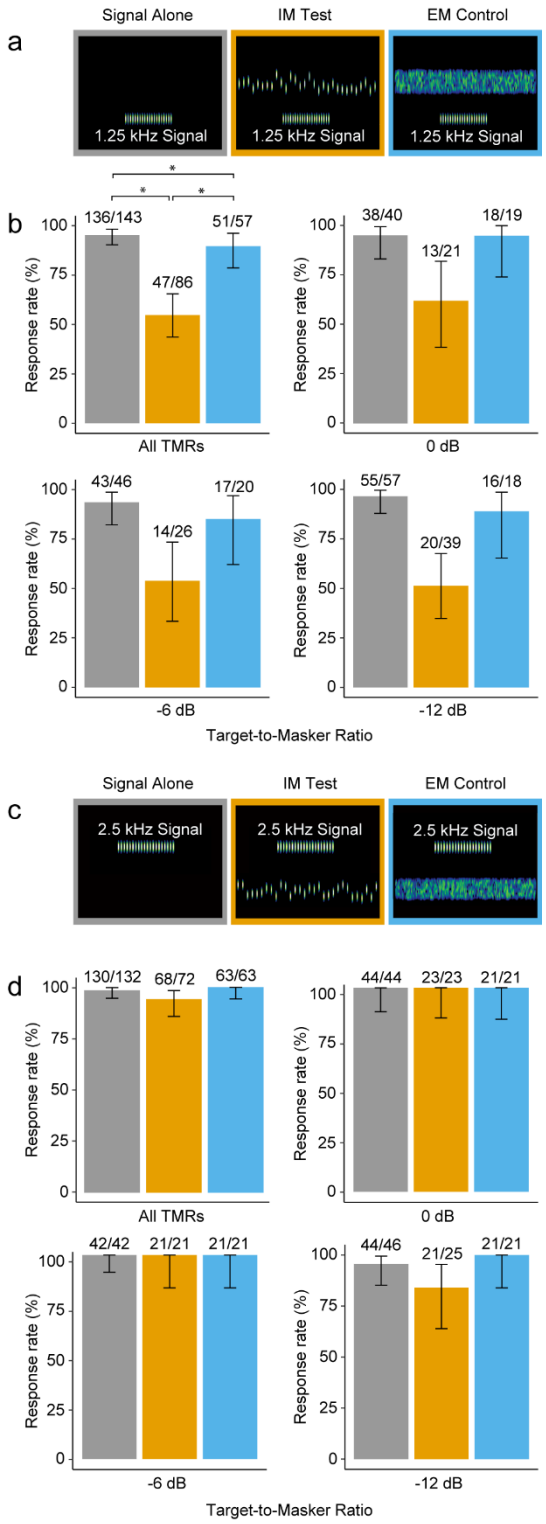

Fig. 4. Informational masking constrains signal recognition. (a) Spectrograms illustrating the 1.25 kHz signal with 20 pulses in the signal alone condition (*Left*), IM test condition (*Center*), and EM control condition (*Right*), depicted by *gray*, *orange*, and *blue* colors, respectively. In the IM test and EM control conditions, the sound pressure level of the masker was set to 75 dB SPL (LC_{eq} re 20 μPa). The sound pressure level of the target signal was varied to achieve the nominal target-to-masker ratio (TMR, or corresponding sound pressure level in the signal alone conditions). (b) Response rate showing the percentages (± 95% exact

binomial confidence intervals) of subjects that responded to 1.25 kHz signals having 20 or fewer pulses pooled over all TMRs and separately for TMRs of -12 dB, -6 dB, and 0 dB. Fractions show the number of tests in which a subject responded over the total number of tests conducted (for pooled data) or the number of subjects that responded over the number of subjects tested (for separate choice tests). Neither TMR (χ^2 = 1.30, P = 0.520) nor the interaction of masking condition and TMR (χ^2 = 1.08, P = 0.900) had significant effects on response. In the pooled data, subjects were significantly (*P < 0.05) less likely to respond in the IM test condition than in the signal alone and EM control conditions, which also differed significantly. (c) Spectrograms illustrating the 2.5 kHz signal in the signal alone condition (*Left*), IM test condition (*Center*), and EM control condition (*Right*). (d) Response rates for 2.5 kHz signals, which were not analyzed statistically because of the high and nearly uniform response rates across conditions.

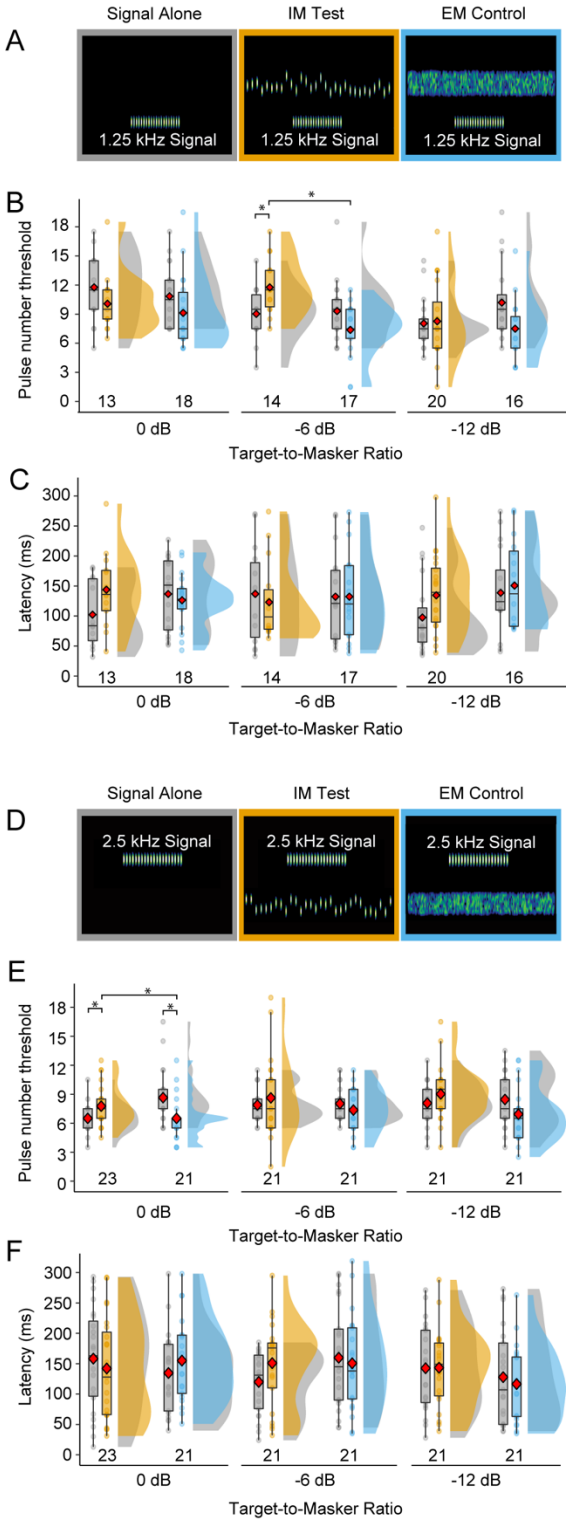

Fig. 5. Informational masking has limited impact on pulse number thresholds and response latencies. (a) Spectrograms illustrating the 1.25 kHz signal with 20 pulses in the signal alone condition (*Left*), IM test condition (*Center*), and EM control condition (*Right*), depicted by *gray*, *orange*, and *blue* colors, respectively. (b) Pulse number thresholds for responding to 1.25 kHz signals measured using an adaptive tracking procedure and defined as the minimum number of pulses required to elicit positive phonotaxis. For each subject tested at each TMR, separate thresholds were determined in the signal alone condition and in either the IM test condition or the EM control condition. For each TMR, the paired box plots on the left depict the first, second (median), and third quartiles, and the mean (red diamond) values in the paired signal alone and IM test conditions, whereas the paired box plots on the right depict these values in the paired signal alone

and EM control conditions. Whiskers depict 1.5 times the interquartile range. Points superimposed on the paired box plots depict individual data points. The two overlapping distributions are kernel density representations of the data that are depicted in the corresponding box plots. Sample sizes are indicated at the base of each dataset. At a TMR of -6 dB, pulse number thresholds were significantly higher (*$P < 0.05$) in the IM test condition compared with the paired signal alone condition and the EM control condition (Table 1). (c) Response latencies for 1.25 kHz signal, computed as the time to respond to the target signal having the lowest number of pulses that elicited positive phonotaxis. There were no significant differences in response latencies. Box plots and distributions as in (b). (d) Spectrograms illustrating the 2.5 kHz signal in the signal alone condition (*Left*), IM test condition (*Center*), and EM control condition (*Right*). (e) Pulse number thresholds for responding to 2.5 kHz signals. At a TMR of 0 dB, pulse number thresholds were significantly higher in the IM test condition compared with both the paired signal alone condition and the EM control condition (Table 1). Pulse number thresholds were significantly lower in the EM control condition compared with the paired signal alone condition (Table 1). (f) Response latencies for 2.5 kHz signal. There were no significant differences in response latencies.

Table 1. Outcomes of Wilcoxon signed rank tests† comparing pulse number thresholds between the signal alone, IM test, and EM control conditions in tests with the 1.25 kHz and 2.5 kHz signals. The signal alone condition is denoted as SA; the IM test and EM control conditions are denoted as IM and EM, respectively.

Signal	TMR	Comparison	Statistic	*P*-value
1.25 kHz	0 dB	SA vs. IM (*N* = 13)	Paired Wilcoxon: V = 65	0.182
		SA vs. EM (*N* = 18)	Paired Wilcoxon: V = 93.5	0.060
		EM vs. IM (*N* = 18,13)	Independent-sample Wilcoxon: W = 148.5	0.211
	-6 dB	SA vs. IM (*N* = 14)	Paired Wilcoxon: V = 4.5	0.021**
		SA vs. EM (*N* = 17)	Paired Wilcoxon: V = 103	0.073
		EM vs. IM (*N* = 17,14)	Independent-sample Wilcoxon: W = 211	< 0.001**
	-12 dB	SA vs. IM (*N* = 20)	Paired Wilcoxon: V = 88.5	0.807
		SA vs. EM (*N* = 16)	Paired Wilcoxon: V = 111	0.027*
		EM vs. IM (*N* = 16, 20)	Independent-sample Wilcoxon: W = 172.5	0.700
2.5 kHz	0 dB	SA vs. IM (*N* = 23)	Paired Wilcoxon: V = 26	0.009**
		SA vs. EM (*N* = 21)	Paired Wilcoxon: V = 180	< 0.001**
		EM vs. IM (*N* = 21,23)	Independent-sample Wilcoxon: W = 329	0.037**
	-6 dB	SA vs. IM (*N* = 21)	Paired Wilcoxon: V = 69.5	0.495
		SA vs. EM (*N* = 21)	Paired Wilcoxon: V = 141	0.376
		EM vs. IM (*N* = 21,21)	Independent-sample Wilcoxon: W = 250	0.463
	-12 dB	SA vs. IM (*N* = 21)	Paired Wilcoxon: V = 40.5	0.158
		SA vs. EM (*N* = 21)	Paired Wilcoxon: V = 147.5	0.035*
		EM vs. IM (*N* = 21,21)	Independent-sample Wilcoxon: W = 314.5	0.018*

† The data were analyzed using a series of paired and independent-sample Wilcoxon tests (with corrections for multiple comparisons) because the distributions of pulse number threshold and response latency were non-normal; paired tests were used to compare responses obtained from the same subject in the signal alone condition and its assigned masking condition.

* P-values below α = 0.05 but not statistically significant following corrections for multiple comparisons.

** P-values below α = 0.05 and statistically significant after corrections for multiple comparisons.