nature neuroscience

Cortical interference effects in the cocktail party problem

Rajiv Narayan[1,2], Virginia Best[1,3], Erol Ozmeral[1,3], Elizabeth McClaine[1,2], Micheal Dent[4],
Barbara Shinn-Cunningham[1–3] & Kamal Sen[1,2]

Humans and animals must often discriminate between complex natural sounds in the presence of competing sounds (maskers). Although the auditory cortex is thought to be important in this task, the impact of maskers on cortical discrimination remains poorly understood. We examined neural responses in zebra finch (*Taeniopygia guttata*) field L (homologous to primary auditory cortex) to target birdsongs that were embedded in three different maskers (broadband noise, modulated noise and birdsong chorus). We found two distinct forms of interference in the neural responses: the addition of spurious spikes occurring primarily during the silent gaps between song syllables and the suppression of informative spikes occurring primarily during the syllables. Both effects systematically degraded neural discrimination as the target intensity decreased relative to that of the masker. The behavioral performance of songbirds degraded in a parallel manner. Our results identify neural interference that could explain the perceptual interference at the heart of the cocktail party problem.

In everyday settings, our ears are constantly bombarded by a mixture of sounds coming from multiple, simultaneous sources. Many species of animals, including birds[1], frogs[2] and mammals[3], are adept at analyzing the acoustic mixture to determine which sound sources are present, an ability that is important for procreation and survival. In humans, the problem of identifying speech in a background of competing sounds is often described as the cocktail party problem (CPP)[3,4]. The CPP is especially challenging for hearing-impaired listeners and artificial speech-recognition systems[5–7]. Normal-hearing listeners also have relative difficulty solving this problem under certain conditions; for example, when the target and maskers are spectro-temporally similar and/or spatially colocated[8,9]. Examining neural responses to natural stimuli in the presence of maskers may reveal interference that could explain difficulties in perceptual discrimination under adverse conditions. The auditory cortex is thought to be critically involved in processing complex natural sounds[10–13]. Although broadband noise has been shown to produce nonlinear effects on cortical responses to natural sounds[10], the impact of spectro-temporally complex maskers on cortical discrimination remains unclear.

Here we studied the effects of different maskers on cortical discrimination in the songbird, a model system that shows similarities to humans in the context of vocal communication[14]. Songbirds are attractive for studying the CPP because they communicate using complex vocal communication sounds in complex acoustic environments (for example, in the presence of many other birds vocalizing in the background[1]). In this study, we recorded neural responses from zebra finch field L, the avian homolog of the mammalian auditory cortex, which is thought to be involved in the processing of conspecific songs[15,16]. In previous work, we have shown that the responses of single units in field L provide sufficient information to discriminate between conspecific songs[11,13]. Here we extended this work to examine how response patterns are affected by different kinds of simultaneous maskers that cause distinct patterns of behavioral disruption in human listeners[17]. Our aim was to identify how the different maskers affect neural responses and to determine how this interference translates into changes in neural discrimination.

RESULTS

Neural responses to targets and maskers

We examined neural responses to target songs presented in quiet or embedded in three different maskers: broadband noise (also referred to simply as noise), modulated noise and birdsong chorus (see **Fig. 1** and Methods). Target intensity was varied systematically (with the masker held at a fixed intensity) to give five different target-to-masker ratios (TMRs). In general, the neural sites showed strong, phase-locked responses to individual syllables of the target (**Fig. 2**). Response sensitivity to masked targets varied across sites, but response patterns deviated further from the unmasked responses as the TMR decreased (**Fig. 2**). We observed two types of neural interference. There were additions of spikes (observed particularly during the gaps between the target song syllables) and suppression of spikes (observed particularly during the syllables of the target). The relative contributions of these two effects depended on the particular site, the type of masker and the TMR. Spike suppression was particularly visible at low TMRs, whereas spike additions appeared to be largely independent of TMR.

---

[1]Hearing Research Center and [2]Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215, USA. [3]Department of Cognitive and Neural Systems, Boston University, 677 Beacon Street, Boston, Massachusetts 02215, USA. [4]Department of Psychology, University at Buffalo, The State University of New York, Park Hall Room 206, Buffalo, New York 14260-4110, USA. Correspondence should be addressed to K.S. (kamalsen@bu.edu).

Received 29 August; accepted 10 October; published online 11 November 2007; doi:10.1038/nn2009

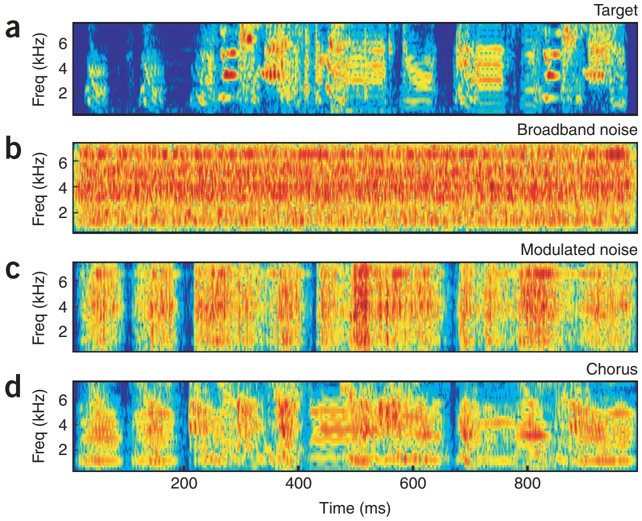


Figure 1 Target and masker stimuli. (**a**–**d**) Exemplar spectrograms of a target song (**a**) and each of the three types of maskers: broadband noise (**b**, steady-state noise with spectral content matching the chorus) modulated noise (**c**, broadband noise multiplied by the envelope of a chorus), and chorus (**d**, random combination of three non-target birdsongs).

Although unmasked responses to the targets were obtained for all sites at a single intensity (0 dB relative to the fixed masker intensity), responses for the full range of target intensities used in the masked trials were obtained from a subset of 22 sites. In this subset, we found that the rates were relatively unchanged with varying target intensity (**Fig. 2b**). This suggests that the changes in firing-rate patterns that were observed in the masked conditions (spike suppression and spike additions) were a result of the presence of the masker and not the absolute intensity of the target.

To quantify the interference attributable to additions and deletions of spikes, target stimuli were segmented into syllables and temporal gaps, and mean firing rates in these regions were compared for responses to unmasked and masked stimuli at the 0-dB TMR across all sites (see **Fig. 2c** and Methods). Overall, within syllables, the mean firing rate was reduced by noise and chorus maskers, but was increased by the modulated noise masker. Changes in within-syllable firing rates were significantly different for the different masker types ($F_{2,136} = 50.04$, $P < 0.001$, one-way repeated measures ANOVA) and Tukey *post hoc* comparisons indicated that there were significant differences between modulated noise and noise ($P < 0.001$) and between modulated noise and chorus ($P < 0.001$), but not between noise and chorus ($P = 0.903$). Between syllables, the mean firing rate was increased in the presence of the different maskers. Changes in between-syllable firing rates were significantly different across masker types ($F_{2,136} = 39.75$, $P < 0.001$, one-way repeated measures ANOVA) and Tukey *post hoc* comparisons indicated that the effects of the three maskers were all significantly different from each other ($P < 0.001$).

We evaluated the effect of TMR on the differences in firing rate within and between target syllables among different maskers in the subset of 22 sites where unmasked responses to the targets were recorded for all of the target intensities (**Supplementary Fig. 1** online). Within target syllables, the maskers generally caused suppression; however, modulated noise caused spike additions at TMRs ≤ 0 dB. Between target syllables, all maskers caused spike additions at all TMRs.

Neural discrimination in the presence of a masker

We quantified neural discrimination using a metric-based spike-train classification scheme (see Methods). The accuracy of classifying the target songs embedded in the three masker types at different TMRs

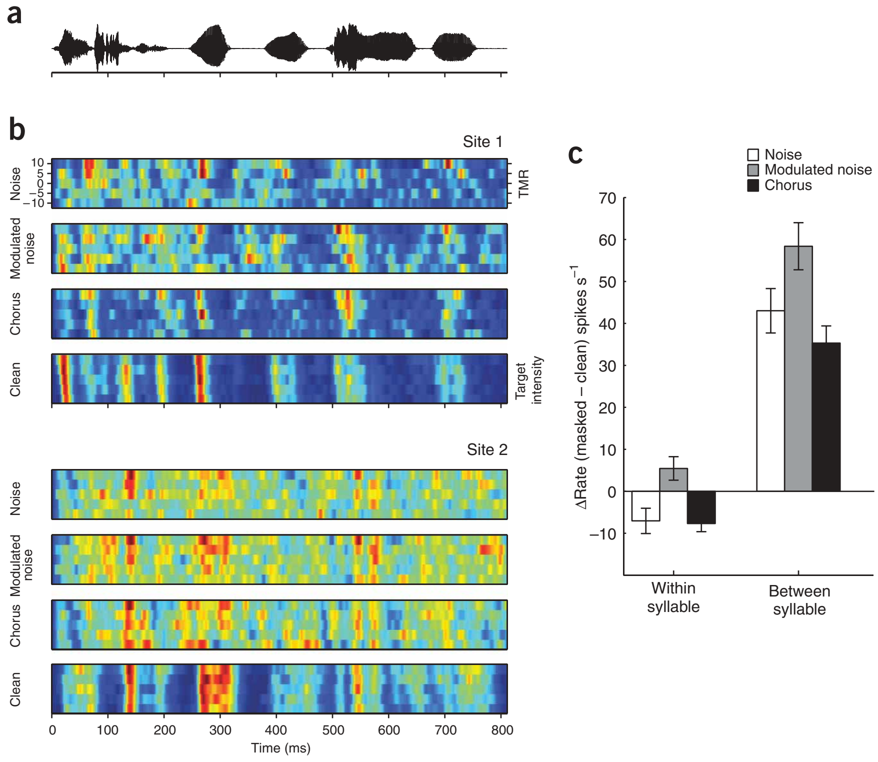


Figure 2 Neural responses to targets in the presence of maskers. (**a**) The sound pressure waveform of a target as a function of time. (**b**) Responses of two representative sites to the target shown in **a**, embedded in three maskers. For each site, the top three panels depict color-coded mean firing rates of the responses to the same target in noise, modulated noise and chorus, respectively. Target intensity was varied systematically with the masker held at a fixed intensity to give five different TMRs. In each panel, the responses are arranged in decreasing TMR order (from top to bottom, +10 dB to –10 dB). The last panel shows the mean firing rate response of each site to tokens of the target presented alone, as a function of decreasing target intensity. The color-coded rates are normalized to the peak rate of the response to the target in the unmasked condition at 0-dB TMR. (**c**) The mean differences in firing rate (masked – unmasked response) both within and between target syllables at 0-dB TMR (mean ± s.e.m., $n = 69$). Positive differences in the mean firing rate reflect a dominance of spike additions and negative differences reflect a dominance of spike suppression. Within syllables, 50 sites showed a reduction in mean firing rate in chorus and noise, whereas 36 sites showed an increase in firing rate in modulated noise. Between syllables, an increase in firing rate was observed in 66, 68 and 63 sites for chorus, modulated noise and noise, respectively.

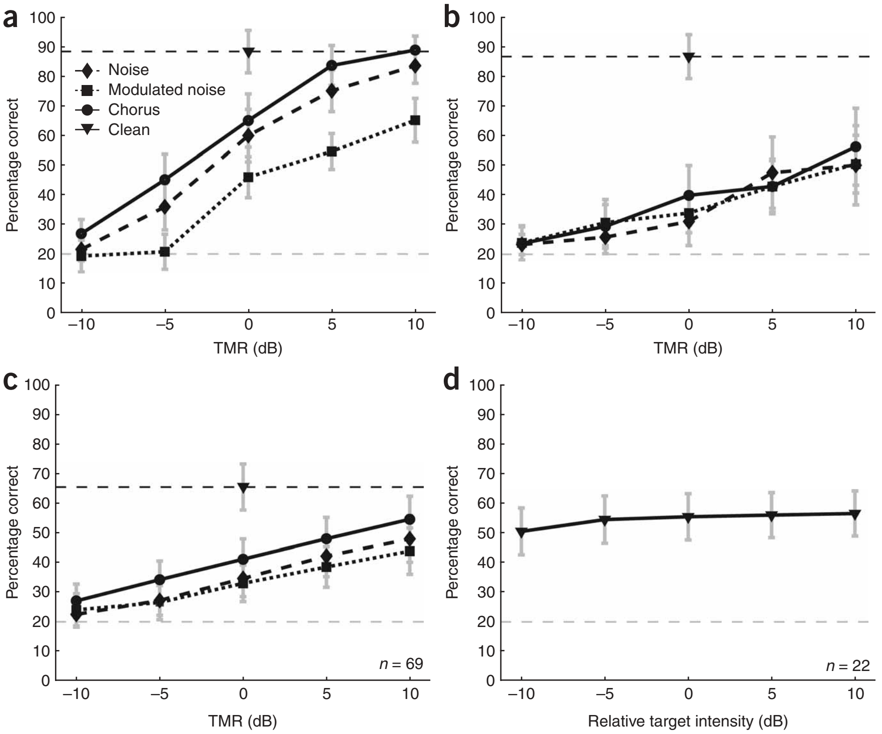


Figure 3 Neural discrimination under different masking conditions. (**a**,**b**) Discrimination accuracy of two representative sites, plotted as a function of TMR (neurometric curves), in the presence of three different maskers (mean ± 1 s.d.). The performance achieved with unmasked targets (at 0-dB level) is also indicated (triangle). Chance performance was 20% (gray horizontal dashed line). (**c**) Mean discrimination in the presence of the three maskers for 69 sites (mean ± s.e.m.). Two-way repeated-measures ANOVA showed that there was a statistically significant interaction between TMR and masker type ($F_{8,544} = 17.89$, $P < 0.001$), in addition to significant main effects of TMR ($F_{4,272} = 152.81$, $P < 0.001$) and masker ($F_{2,136} = 48.17$, $P < 0.001$) on performance. (**d**) Mean discrimination of targets in the absence of maskers for 22 sites, when the target intensity was at a level corresponding to each TMR tested. Although the differences in performance with target intensity were found to be significant ($F_{4,84} = 17.75$, $P < 0.001$, one-way repeated measures ANOVA), Tukey *post hoc* comparisons showed that this effect was driven by the lowest intensity only, where a small, but significant, decrease compared with all other intensities was detected ($P < 0.001$).

(and in quiet) was measured. A comparison of the performance of different sites as a function of TMR (neurometric curves) revealed that there could be marked differences in performance with the masked targets, even for sites that showed comparable performances with unmasked targets. For example, in one site, the discrimination of masked targets was close to chance level (20%) for low TMRs and approached the unmasked performance at high TMRs (**Fig. 3a**). In contrast, another site showed performance close to chance at –10-dB TMR, which improved with increasing TMR, but only up to a level well below the performance for unmasked targets (**Fig. 3b**). There were also

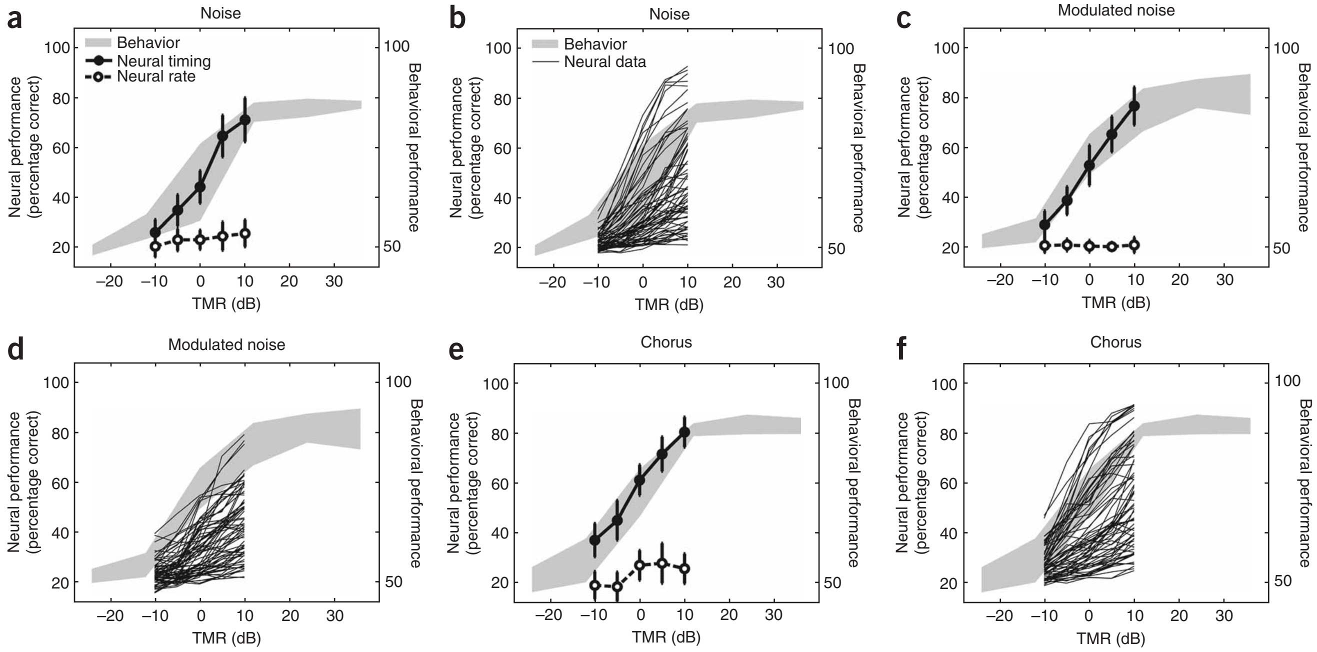


Figure 4 Neural versus behavioral performance. (**a**,**c**,**e**) Individual examples of sites where the neurometric curve (thick solid line) matched the mean psychometric function across the common TMR range (thin solid line; gray region shows ± 1 s.d.). Also shown is the performance for the same sites when the discrimination was based on firing-rate alone (dashed lines). (**b**,**d**,**f**) The neurometric curves for all sites (thin lines) compared with the mean psychometric function (replotted from **a**, **c** and **e**).

site-specific differences in performance across the different maskers. In one instance, the performance was best in chorus, intermediate in noise and worst in modulated noise (**Fig. 3a**), whereas in another, performance was similar across the different maskers (**Fig. 3b**). The mean neurometric performance across 69 sites for masked targets improved as a function of TMR and approached the mean performance obtained using unmasked targets (**Fig. 3c**).

Discrimination for unmasked targets across the full range of target intensities (that is, from –10 dB to +10 dB relative to the fixed masker intensity) was relatively constant in the subset of 22 sites in which responses were obtained (**Fig. 3d**). Overall, the absolute target intensity did not appear to drive the discrimination differences that were seen with the masked targets.

Comparison of neural and behavioral performance

Four zebra finches were tested behaviorally for their ability to discriminate target songs under identical stimulus conditions to those used in the neural experiment (see Methods). Behavioral performance was similar for the three masker types, with scores close to chance (50% for the task used) at the lowest TMR and with an upper asymptote close to the score that was obtained during training with targets in quiet (88%) (**Fig. 4**).

We found sites where the neurometric functions closely matched the psychometric functions across the common TMR range (**Fig. 4a,c,e**). In addition, neural discrimination on the basis of fine timing, rather than mean rate, provided a much better match to the behavioral data for these sites. Logistic fits to individual curves of the 69 recorded sites (see Methods) indicated that the majority of neurometric functions had shallower slopes and higher thresholds than the psychometric functions (**Supplementary Fig. 2** online). Sites that matched the behavioral performance over the entire range of TMRs tended to have relatively high performance levels, overall and were in the top third of the dataset if ranked on the basis of their discrimination of unmasked targets (**Fig. 4b,d,f**).

DISCUSSION

Although human and animal listeners are remarkably adept at understanding communication sounds in noisy or complex backgrounds (solving the CPP), difficulties arise in particularly adverse conditions, such as highly reverberant environments or in the case of hearing impairment. Furthermore, despite a long history of research in artificial speech recognition, no effective computational solution to the CPP is known[6]. Investigating the neural substrates for the CPP in a suitable animal model may help determine the source of the perceptual difficulty posed by a complex listening environment and reveal how biological systems cope with such settings.

Previous studies have demonstrated that songbirds can detect learned target songs in mixtures of task-irrelevant songs[1,18]. One of these studies also compared the masking effects of distracting songs and level-matched white noise, and found no appreciable differences in performance between the two[18], consistent with our findings. Notably, human listeners often show different amounts of masking for noise and for complex maskers that are spectro-temporally similar to the target[8,19]. Although it is difficult to make detailed comparisons to the human literature, our results suggest that songbirds and humans differ in their relative susceptibility to different masker types. This may be a result of different neural sensitivities (for example, spectral and temporal tuning), attentional strategies or a combination of both. Although the previous studies provided information regarding the influence of maskers on behavioral performance of songbirds, relatively little is known about the effects of maskers at the neural level. To our knowledge, this is the first study to jointly investigate the neural and behavioral discrimination of complex target sounds in the presence of complex maskers.

Several findings point to the involvement of auditory cortex in the CPP in both humans and animals. Auditory cortex is important in the perception of complex sounds such as vocal communication sounds and speech[20,21]. Physiological studies in songbirds suggest a similar role for field L in processing vocal communication sounds [15,22–27]. Different subdivisions of mammalian auditory cortex appear to have differing responses to complex sounds (for example, speech)[21]. Studies in the avian auditory forebrain also suggest distinct response properties to simple versus complex sounds[15,16,22,24,25], although a detailed comparison between the subdivisions in the mammalian and avian systems remains premature. Recent thinking suggests that the auditory cortex may be critical in auditory scene analysis and the formation of auditory objects[12,28]. Although subcortical processing is undoubtedly involved in the analysis of complex sounds[29–38], the auditory cortex is likely to make a critical contribution to solving the CPP. Relatively few studies have investigated the processing of masked targets by cortical neurons[39–44]. In most of those studies, the targets used were simple stimuli such as tones or clicks, and the targets and maskers were dissimilar. Given the significant nonlinearities in cortical responses to complex natural sounds[10,27,45], it is difficult to extrapolate the results of those studies to the CPP. Indeed, in one study, cortical responses to complex targets (bird chirps) in the presence of background noise revealed marked nonlinear interactions[10], highlighting the importance of using complex targets and maskers in the search for cortical substrates of the CPP.

Although we identified interference at the cortical level, it remains possible that some of these effects may be inherited from subcortical levels of processing. Future studies at earlier levels of the auditory system in the songbird will be important to clarify this issue. Our results revealed two forms of interference: the addition of spurious spikes occurring primarily during the gaps between syllables and the suppression of informative spikes occurring primarily during the syllables themselves. The addition of spikes during syllable gaps is expected to be a result of the presence of masker energy. Although this effect was observed for all of the maskers, the number of additions was significantly greater for modulated noise than for broadband noise (which had the same fine structure, but a different envelope) and chorus (which had the same envelope, but a different fine structure). This suggests that both envelope and fine structure contribute to the magnitude of this form of interference. In contrast, the suppression of informative spikes during the syllables was observed primarily for the chorus and noise maskers. Suppression might be expected for the broadband noise masker, as it is known that neural responses adapt to constant stimulation of this kind. However, the fact that suppression was seen for the chorus masker and not for the modulated noise masker indicates nonlinear effects due to the spectral structure of the chorus. Potential mechanisms that could underlie suppression include cortical inhibition and synaptic depression.

We found that the different forms of cortical interference systematically degraded neural discrimination as TMR decreased. Even sites that discriminated very accurately without a masker performed close to chance levels at the lowest TMR (–10 dB). In principle, it is possible to imagine population coding schemes for pooling large populations of neurons that could compensate for the reduction in performance of single sites. However, a comparison of single-site performance and behavioral performance revealed several unexpected results. First, the neurometric functions obtained from some of the sites closely matched the psychometric function over the entire range of TMRs sampled for the neural experiments. Although this result is reminiscent of similar results obtained in visual and somatosensory cortices[46,47], there is an important distinction. In the previous studies, neural discrimination was based on average firing rate. In our data, neural performance determined by average rate was far inferior to behavioral performance, and information present at finer time scales was necessary to match behavioral performance. Second, the distribution of neurometric functions revealed that the sites that matched behavioral performance had relatively high performance levels, ranking approximately in the top third of our dataset on the basis of their performance in the target-alone condition.

Notably, we found several cases in which neural performance was higher than behavioral performance, at least for certain TMRs. Because the neural and behavioral data were collected in different birds, one possible explanation of this result is variability across individuals; that is, the 'best sites' could have come from birds that would have performed at correspondingly higher levels in the behavioral task. Alternative explanations for this result are based on the downstream readout of information. The neural performance that we observed at the level of field L can be thought of as the information available for guiding behavior. However, the specific manner in which this information is read out downstream may impose further limits on performance. One possibility is that additional source(s) of noise at the readout stage could reduce neural performance. Another possibility is that the readout is able to access the best neurons only after repeated exposures via a learning mechanism. Such a scenario suggests that one would observe an improvement in the psychometric function over the course of learning, a common phenomenon in psychophysical experiments.

The neural interference revealed in this study led to a profound reduction in neural discrimination and was accompanied by a parallel reduction in behavioral performance. This kind of interference may lie at the heart of the difficulty of the CPP. In real world settings, however, there are other factors that may influence performance in such tasks. First, the allocation of attention is crucial, and an important future direction will be to examine neural responses and sensitivity to interference in awake animals that are paying attention to the task. Second, behavioral studies in humans have demonstrated that the spatial separation of competing sound sources greatly improves a listener's ability to process a source of interest[4,9,17]. Spatial cues can lead to effective improvements in TMR, which might produce improvements in performance that are predictable from the neurometric functions measured in this study. More importantly, however, spatial separation can aid the selective allocation of attention, and thus may interact with attentional manipulations.

METHODS

Stimuli. Song motifs from five male zebra finches (five renditions or tokens from each bird) were used as target stimuli. Stimuli were band-pass filtered between 250 Hz and 8 kHz. All motifs from a particular bird were highly stereotypical to that bird, but were distinct from those of other birds (see **Supplementary Fig. 3** online).

Three types of maskers were created, all with the same long-term spectral characteristics, but with different short-term statistics. The chorus maskers were generated by adding together three song motifs from nontarget birds. The noise maskers were created by generating broadband noise that had a spectral profile matching that of the average of the set of chorus maskers. The modulated noise maskers were generated by modulating a noise masker with the envelope from a random chorus masker. Six tokens were generated for each type of masker.

Target and masker mixtures were created at five different TMRs. The TMR was defined as the ratio of the target and masker intensities, where intensity was calculated using the broadband root mean squared value over the entire length of each signal. Maskers were always presented at a fixed level of 65-dB SPL (peak, A-weighted intensity measured at 25 cm from the loudspeaker at the position of the birds' ears), and the TMR was set to one of five evenly spaced values between –10 dB and +10 dB by varying the absolute intensity of the target. For each combination of target song (five), masker type (three) and TMR (five), a set of ten stimuli were generated using random selections from the target and masker tokens.

Neural recordings. Extracellular recordings were made from field L of nine urethane-anesthetized, adult, male zebra finches using techniques described previously[11,26]. The experimental protocol was approved by the Institutional Animal Care and Use Committee, Boston University, Charles River Campus. Neural responses were measured for unmasked targets (presented in quiet) and targets embedded in the three maskers. The unmasked and masked stimuli were randomly interleaved and presented in the free field via a single loudspeaker located 25 cm in front of the bird to obtain ten trials for each stimulus. Stimulus presentation and response recordings were performed via a National Instruments data acquisition board (PCI-6052E) and controlled using custom written software in LabWindows CVI (National Instruments).

For each recording site, spike waveforms were first identified using a window discriminator with a manually set threshold, and then further refined using principal components analysis of the waveform shapes in Matlab (Mathworks).

Of all the sites that were probed, the sites that showed an average firing rate that was significantly different ($P < 0.01$, paired t-test) from the average spontaneous firing rate for at least one song stimulus were included in the analysis ($n = 69$). The dataset was comprised of 18 sites that had well-isolated spike waveforms suggesting a single neuron response and 51 sites with waveforms indicative of responses from small clusters of two to five neurons. Here, we use the term 'site' to refer to both types of responses.

The recording locations were verified using Nissl-stained parasagittal 50-µm sections of the brain to compare the electrode tracks to histological markers that define the boundaries of field L[48]. All sites were confirmed to be in the auditory forebrain on the basis of a combination of histology, stereotaxic coordinates and recording depth. Of these, 60 out of 69 sites were from field L and the remaining nine sites were from the overlying area of the caudal mesopallium, the avian analog of secondary auditory cortex.

Data analysis. We examined the recorded spike trains for additions and deletions of spikes (relative to the response to the target in quiet) by measuring firing rates within and between target song syllables. The analysis was done at 0-dB TMR, as this was the only target level for which unmasked responses were collected for all sites. For a subset of 22 sites in which unmasked responses were collected at target intensities corresponding to the entire TMR range, the firing-rate analysis was carried out at each TMR. Each target song was manually segmented to mark the syllable boundaries and the temporal gaps between syllables (**Supplementary Fig. 3**). The firing rates in the unmasked and masked conditions were calculated at each site separately for the within- and between-target syllable portions of the spike responses and were averaged across targets, trials and syllables. The relative difference between the average masked and unmasked rate was then computed for each masker. To account for the neural transmission time to field L, we analyzed the neural response starting 10 ms after the beginning of each syllable.

We evaluated the ability of sites to discriminate between target songs using a spike distance metric[49]. The metric quantifies the dissimilarity between pairs of spike trains, while permitting the time scale of the analysis to be adjusted using a single free parameter (τ). For small τ, the metric acts as a 'coincidence detector' with small differences in spike timing contributing to the distance, whereas for long τ, it acts as a 'rate-difference counter', where average firing rates contribute to the distance.

Responses to unmasked targets were compared with the spike trains elicited by targets embedded in the maskers using a classification scheme based on the spike distance metric[11,50]. Each masked response was classified into a target song category by selecting the target whose unmasked response was closest to the observed response. Percent correct performance in this one-in-five classification task was computed for each site. For performance based on spike timing, the temporal resolution of the distance metric was set to 10 ms (shown in earlier work to give optimal classification performance[11]). For performance based on spike rate, the temporal resolution was set to 1,000 ms (the length of the stimulus-driven response).

Statistical testing. Differences in mean firing rates and discrimination were tested using one-way and two-way repeated measures ANOVA with Sigma-Stat software (SysStat Software). When the ANOVA showed significant main effects, pair-wise comparisons were carried out using Tukey *post hoc* comparison procedures. We report the ANOVA results as $F_{a,b} = c$, where a and b are the numerator and denominator degrees of freedom of the F statistic c. Friedman's nonparametric tests also yielded similar results. Percent correct scores were arcsine transformed before testing. For all tests, the null hypothesis was rejected at the 0.05 level.

Behavioral experiment. The behavioral experiment was conducted on four zebra finches as part of a larger experiment. All procedures were approved by the Institutional Animal Care and Use Committee, University at Buffalo. The birds were placed in a cage inside a sound-attenuating small-animal chamber lined with acoustic foam. A single loudspeaker located directly behind the bird delivered the acoustic stimuli. The birds were trained to sit on a perch in the front of the cage, facing a food hopper and two microswitch response keys. Stimulus presentation and response recordings of the animals were controlled by Sykofizx software and TDT hardware modules (Tucker-Davis Technologies).

The birds were trained using operant conditioning procedures to perform a categorization task using a set of birdsong targets that were almost identical to those used in the neural study. Six target songs were used (the five from the neural study and one additional song). The birds were trained to peck the left key to initiate a trial and a variable waiting interval (2–7 s). Following that interval, a target song was presented from the loudspeaker. The birds were trained to peck the left key again for three of the songs and to peck the right key for the other three songs. The songs were randomly assigned to a key and each bird had a different combination of left and right songs. If they responded correctly and in 2 s, they were rewarded with a 2-s access to seeds from the hopper. If they responded incorrectly, a house light was extinguished for 5 s. Training continued until performance reached asymptote (around 88% correct after 2,800–4,400 trials).

Following training, the birds were tested on the same categorization task in the presence of one of three colocated masker stimuli. Masker stimuli were identical to those used in the neural study. All birds were tested at TMRs of –24, –12, 0, +12, +24 and +36 dB; some of the birds were also tested at TMRs outside of this range (–48, –36, +48 or +60 dB) in some conditions, but these data are not presented here. Percentage correct values were calculated for each bird, collapsed across the six target songs, from at least 100 trials per TMR.

Psychometric and neurometric functions. Classification performance as a function of TMR was analyzed for each individual site and for each behavioral subject to give neurometric and psychometric functions, respectively. For the extraction of threshold and slope parameters, logistic functions were fit to the raw data using the Matlab toolbox PSIGNIFIT. The lower asymptote of these functions was fixed at the chance performance level. The upper asymptote was fixed on an individual basis at the performance level achieved for the classification of unmasked songs. Threshold was defined as the TMR at the midpoint of a given neurometric or psychometric function.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS
This work was supported by the US National Institute on Deafness and Other Communication Disorders grant 1 RO1 DC–007610–01A1 (K.S.), the Deafness Research Foundation and the National Organization for Hearing Research Foundation (M.D.), and the Air Force Office of Scientific Research grant FA9550–04–1–0260 (B.S.-C.).

AUTHOR CONTRIBUTIONS

K.S., B.S.-C., M.D., R.N., V.B and E.O. designed the research. R.N. carried out the neurophysiology experiments and analyzed the neural data. V.B. analyzed the behavioral data and assisted in the neurophysiology experiments. E.O. generated stimuli and assisted in the neurophysiology experiments. E.M. and M.D. carried out the behavioral experiments. M.D. supervised the behavioral experiments. K.S. supervised the neurophysiology experiments. K.S., V.B. and R.N. wrote the manuscript.

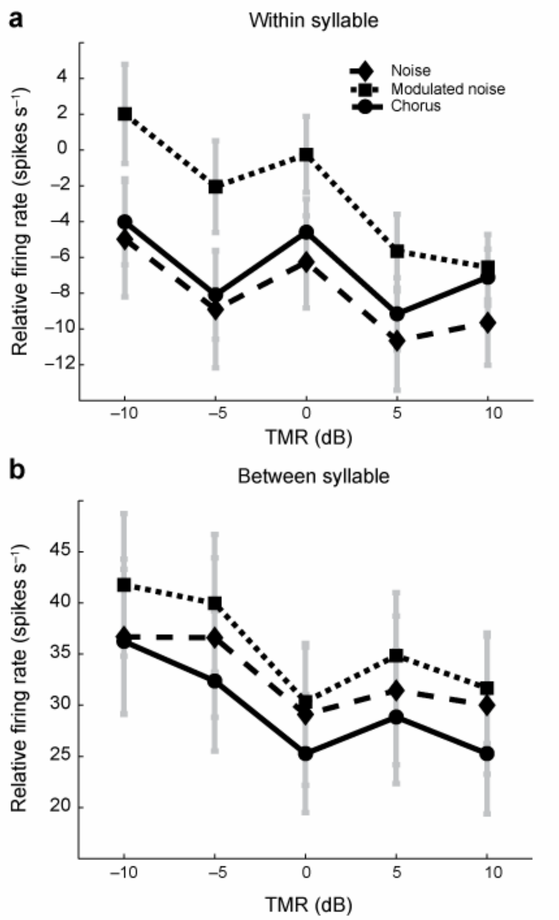
1. Hulse, S.H., MacDougall-Shackleton, S.A. & Wisniewski, A.B. Auditory scene analysis by songbirds: stream segregation of birdsong by European starlings (*Sturnus vulgaris*). *J. Comp. Psychol.* **111**, 3–13 (1997).
2. Endepols, H., Feng, A.S., Gerhardt, H.C., Schul, J. & Walkowiak, W. Roles of the auditory midbrain and thalamus in selective phonotaxis in female gray tree frogs (*Hyla versicolor*). *Behav. Brain Res.* **145**, 63–77 (2003).
3. Cherry, E.C. Some experiments on the recognition of speech, with one and with two Ears. *J. Acoust. Soc. Am.* **25**, 975–979 (1953).
4. Bronkhorst, A.W. The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions. *Acustica* **86**, 694–703 (2000).
5. Asari, H., Pearlmutter, B.A. & Zador, A.M. Sparse representations for the cocktail party problem. *J. Neurosci.* **26**, 7477–7490 (2006).
6. Haykin, S. & Chen, Z. The cocktail party problem. *Neural Comput.* **17**, 1875–1902 (2005).
7. Bronkhorst, A.W. & Plomp, R. Effect of multiple speech-like maskers on binaural speech recognition in normal and impaired hearing. *J. Acoust. Soc. Am.* **92**, 3132–3139 (1992).
8. Brungart, D.S., Simpson, B.D., Ericson, M.A. & Scott, K.R. Informational and energetic masking effects in the perception of multiple simultaneous talkers. *J. Acoust. Soc. Am.* **110**, 2527–2538 (2001).
9. Kidd, G., Jr, Mason, C.R., Rohtla, T.L. & Deliwala, P.S. Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. *J. Acoust. Soc. Am.* **104**, 422–431 (1998).
10. Bar-Yosef, O., Rotman, Y. & Nelken, I. Responses of neurons in cat primary auditory cortex to bird chirps: effects of temporal and spectral context. *J. Neurosci.* **22**, 8619–8632 (2002).
11. Narayan, R., Grana, G. & Sen, K. Distinct time scales in cortical discrimination of natural sounds in songbirds. *J. Neurophysiol.* **96**, 252–258 (2006).
12. Nelken, I. Processing of complex stimuli and natural scenes in the auditory cortex. *Curr. Opin. Neurobiol.* **14**, 474–480 (2004).
13. Wang, L., Narayan, R., Grana, G., Shamir, M. & Sen, K. Cortical discrimination of complex natural stimuli: can single neurons match behavior? *J. Neurosci.* **27**, 582–589 (2007).
14. Doupe, A.J. & Kuhl, P.K. Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22**, 567–631 (1999).
15. Grace, J.A., Amin, N., Singh, N.C. & Theunissen, F.E. Selectivity for conspecific song in the zebra finch auditory forebrain. *J. Neurophysiol.* **89**, 472–487 (2003).
16. Woolley, S.M., Fremouw, T.E., Hsu, A. & Theunissen, F.E. Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nat. Neurosci.* **8**, 1371–1379 (2005).
17. Best, V., Ozmeral, E., Gallun, F.J., Sen, K. & Shinn-Cunningham, B.G. Spatial unmasking of birdsong in human listeners: energetic and informational factors. *J. Acoust. Soc. Am.* **118**, 3766–3773 (2005).
18. Appeltants, D., Gentner, T.Q., Hulse, S.H., Balthazart, J. & Ball, G.F. The effect of auditory distractors on song discrimination in male canaries (*Serinus canaria*). *Behav. Processes* **69**, 331–341 (2005).
19. Carhart, R., Tillman, T.W. & Greetis, E.S. Perceptual masking in multiple sound backgrounds. *J. Acoust. Soc. Am.* **45**, 694–703 (1969).
20. Fitch, R.H., Miller, S. & Tallal, P. Neurobiology of speech perception. *Annu. Rev. Neurosci.* **20**, 331–353 (1997).
21. Rauschecker, J.P. Cortical processing of complex sounds. *Curr. Opin. Neurobiol.* **8**, 516–521 (1998).
22. Langner, G., Bonke, D. & Scheich, H. Neuronal discrimination of natural and synthetic vowels in field L of trained mynah birds. *Exp. Brain Res.* **43**, 11–24 (1981).
23. Lewicki, M.S. & Arthur, B.J. Hierarchical organization of auditory temporal context sensitivity. *J. Neurosci.* **16**, 6987–6998 (1996).
24. Muller, C.M. & Leppelsack, H.J. Feature extraction and tonotopic organization in the avian auditory forebrain. *Exp. Brain Res.* **59**, 587–599 (1985).
25. Scheich, H., Langner, G. & Bonke, D. Responsiveness of units in the auditory neostriatum of the guinea fowl (*Numida meleagris*) to species-specific calls and synthetic stimuli. II. Discrimination of Iambus-like calls. *J. Comp. Physiol. A Neuroethol. Sens. Neural. Behav. Physiol.* **132**, 257–276 (1979b).
26. Sen, K., Theunissen, F.E. & Doupe, A.J. Feature analysis of natural sounds in the songbird auditory forebrain. *J. Neurophysiol.* **86**, 1445–1458 (2001).
27. Theunissen, F.E., Sen, K. & Doupe, A.J. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J. Neurosci.* **20**, 2315–2331 (2000).
28. Nelken, I., Fishbach, A., Las, L., Ulanovsky, N. & Farkas, D. Primary auditory cortex of cats: feature detection or something else? *Biol. Cybern.* **89**, 397–406 (2003).
29. Verhey, J.L., Pressnitzer, D. & Winter, I.M. The psychophysics and physiology of comodulation masking release. *Exp. Brain Res.* **153**, 405–417 (2003).
30. McAlpine, D., Jiang, D. & Palmer, A.R. Binaural masking level differences in the inferior colliculus of the guinea pig. *J. Acoust. Soc. Am.* **100**, 490–503 (1996).
31. Jiang, D., McAlpine, D. & Palmer, A.R. Responses of neurons in the inferior colliculus to binaural masking level difference stimuli measured by rate-versus-level functions. *J. Neurophysiol.* **77**, 3085–3106 (1997).
32. Ratnam, R. & Feng, A.S. Detection of auditory signals by frog inferior collicular neurons in the presence of spatially separated noise. *J. Neurophysiol.* **80**, 2848–2859 (1998).

33. Palmer, A.R., Jiang, D. & McAlpine, D. Desynchronizing responses to correlated noise: a mechanism for binaural masking level differences at the inferior colliculus. *J. Neurophysiol.* **81**, 722–734 (1999).
34. Lin, W.Y. & Feng, A.S. Free-field unmasking response characteristics of frog auditory nerve fibers: comparison with the responses of midbrain auditory neurons. *J. Comp. Physiol. [A]* **187**, 699–712 (2001).
35. Pressnitzer, D., Meddis, R., Delahaye, R. & Winter, I.M. Physiological correlates of comodulation masking release in the mammalian ventral cochlear nucleus. *J. Neurosci.* **21**, 6377–6386 (2001).
36. Neuert, V., Verhey, J.L. & Winter, I.M. Responses of dorsal cochlear nucleus neurons to signals in the presence of modulated maskers. *J. Neurosci.* **24**, 5789–5797 (2004).
37. Lane, C.C. & Delgutte, B. Neural correlates and mechanisms of spatial release from masking: single-unit and population responses in the inferior colliculus. *J. Neurophysiol.* **94**, 1180–1198 (2005).
38. Ramachandran, R., Davis, K.A. & May, B.J. Rate representation of tones in noise in the inferior colliculus of decerebrate cats. *J. Assoc. Res. Otolaryngol.* **1**, 144–160 (2000).
39. Las, L., Stern, E.A. & Nelken, I. Representation of tone in fluctuating maskers in the ascending auditory system. *J. Neurosci.* **25**, 1503–1513 (2005).
40. Phillips, D.P. & Cynader, M.S. Some neural mechanisms in the cat's auditory cortex underlying sensitivity to combined tone and wide-spectrum noise stimuli. *Hear. Res.* **18**, 87–102 (1985).
41. Nelken, I., Rotman, Y. & Bar Yosef, O. Responses of auditory-cortex neurons to structural features of natural sounds. *Nature* **397**, 154–157 (1999).
42. Furukawa, S., Xu, L. & Middlebrooks, J.C. Coding of sound-source location by ensembles of cortical neurons. *J. Neurosci.* **20**, 1216–1228 (2000).
43. Nieder, A. & Klump, G.M. Signal detection in amplitude-modulated maskers. II. Processing in the songbird's auditory forebrain. *Eur. J. Neurosci.* **13**, 1033–1044 (2001).
44. Hofer, S.B. & Klump, G.M. Within- and across-channel processing in auditory masking: a physiological study in the songbird forebrain. *J. Neurosci.* **23**, 5732–5739 (2003).
45. Machens, C.K., Wehr, M.S. & Zador, A.M. Linearity of cortical receptive fields measured with natural sounds. *J. Neurosci.* **24**, 1089–1100 (2004).
46. Britten, K.H., Shadlen, M.N., Newsome, W.T. & Movshon, J.A. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* **12**, 4745–4765 (1992).
47. Romo, R. & Salinas, E. Flutter discrimination: neural codes, perception, memory and decision making. *Nat. Rev. Neurosci.* **4**, 203–218 (2003).
48. Fortune, E.S. & Margoliash, D. Cytoarchitectonic organization and morphology of cells of the field L complex in male zebra finches (*Taenopygia guttata*). *J. Comp. Neurol.* **325**, 388–404 (1992).
49. van Rossum, M.C. A novel spike distance. *Neural Comput.* **13**, 751–763 (2001).
50. Machens, C.K. *et al.* Single auditory neurons rapidly discriminate conspecific communication signals. *Nat. Neurosci.* **6**, 341–342 (2003).

Cortical Interference Effects in the Cocktail Party Problem.

Rajiv Narayan, Virginia Best, Erol Ozmeral, Elizabeth McClaine, Micheal Dent, Barbara Shinn–Cunningham and Kamal Sen.

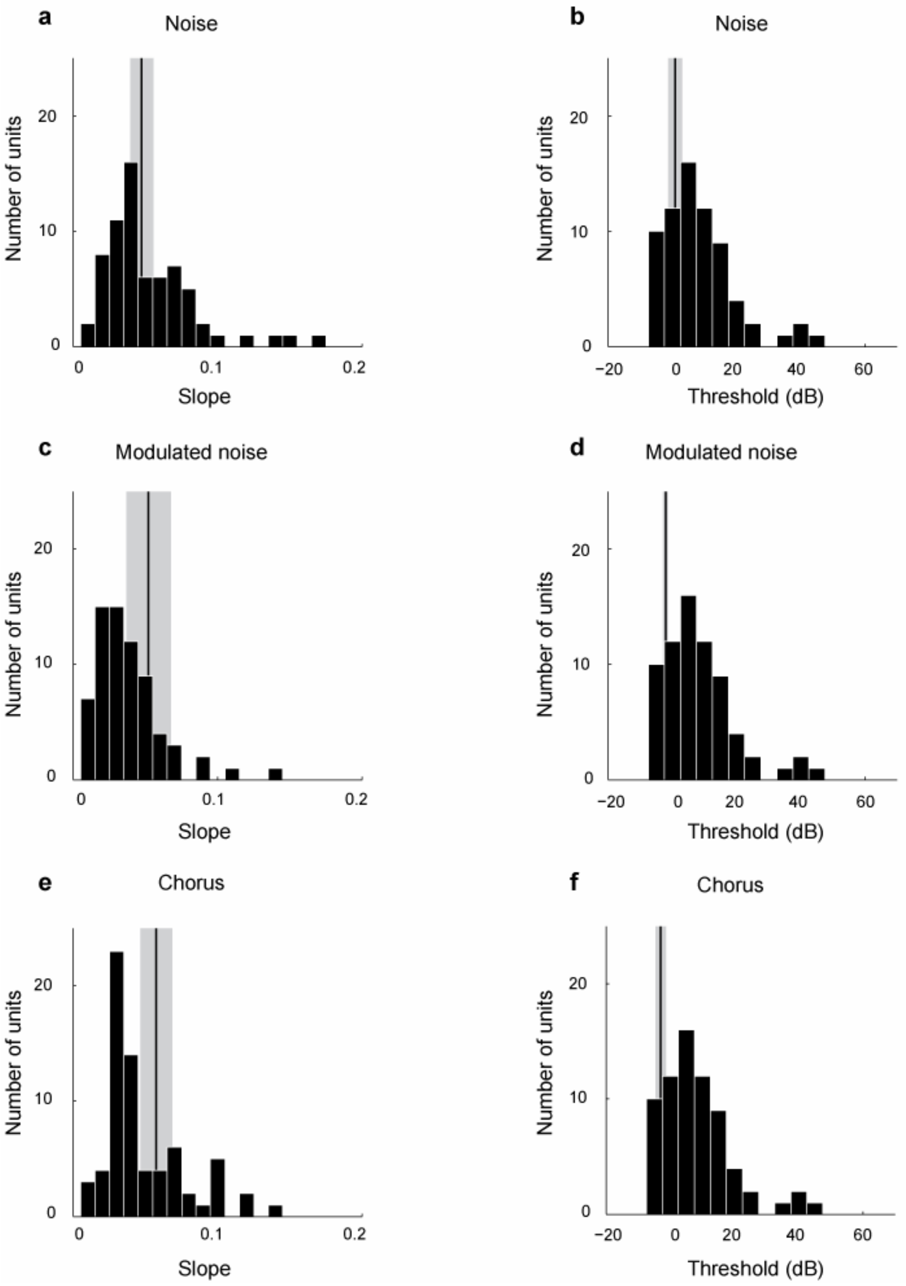
Supplementary figure 1



Firing rates within– and between–target syllables vs. TMR

A comparison of firing rates relative to the unmasked response within (*a*) and between (*b*) target syllables for each masker as a function of TMR in a subset of 22 sites (mean ± 1 s.e.m.). Both within– and between–syllables, there is a significant interaction between masker–type and TMR, implying that the relative effects of the different maskers depend on the TMR (F(8,168)=4.92, $p<0.001$ for within–syllable rate and F(8,168)=2.75, $p=0.007$ for between–syllable rate, two–way repeated measures ANOVA).

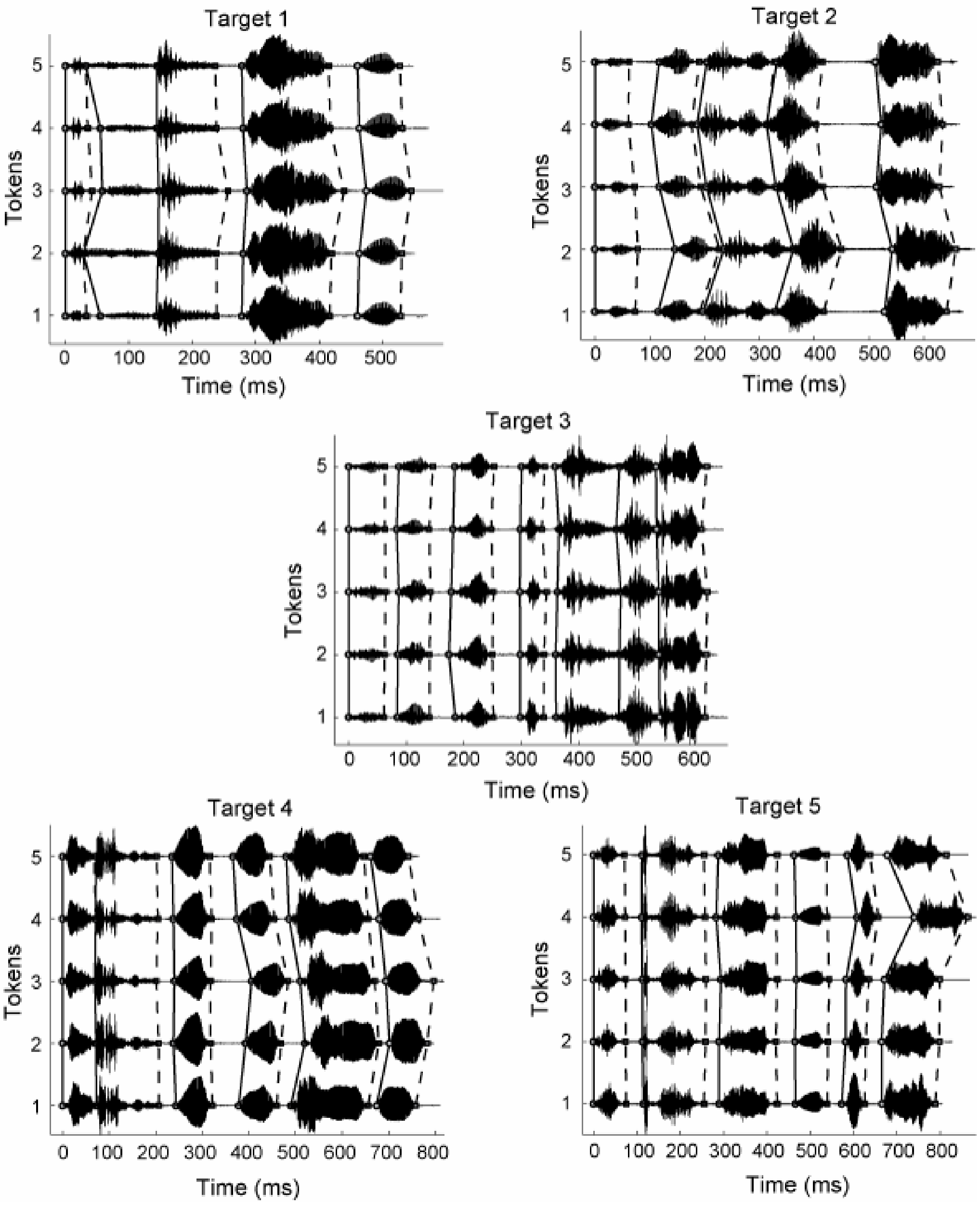
Supplementary figure 2



Slope and threshold parameters

Histograms of slope (*a*,*c*,*e*) and threshold (*b*,*d*,*f*) parameters extracted from logistic fits to the neurometric functions. Background shows equivalent parameters for the psychometric functions (black lines show mean across the four subjects; grey region shows ±1 s.e.m.).

Supplementary figure 3



Target Stimuli

The amplitude waveforms of all the tokens for each target song stimulus used in the experiment. Tokens from a particular target were highly stereotypical to that target but distinct from those of other targets. Each token was manually segmented to mark times with significant energy (within–syllable) and temporal gaps (between–syllable). Vertical lines across the tokens connect the start and stop times for each syllable (solid and dashed lines respectively).