

A cat's cocktail party: Psychophysical, neurophysiological, and computational studies of spatial release from masking

Courtney C. Lane¹, Norbert Kopco², Bertrand Delgutte¹, Barbara G. Shinn-Cunningham², and H. Steven Colburn²

1 Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary, Boston, MA, USA
{court, bard}@epl.meei.harvard.edu

2 Hearing Research Center, Boston University, Boston, MA, USA {kopco, shinn, colburn}@bu.edu

1 Introduction

Masked thresholds can improve substantially when a signal is spatially separated from a noise masker (Saber et al. 1991). This phenomenon, termed “spatial release from masking” (SRM), may contribute to the cocktail party effect, in which a listener can hear a talker in a noisy environment. The purpose of this study is to explore the underlying neural mechanisms of SRM.

Previous psychophysical studies (Good, Gilkey, and Ball 1997) have shown that for high-frequency stimuli, SRM was due primarily to energetic effects related to the head shadow, but for low-frequency stimuli, both binaural processing (presumably ITD processing) and energetic effects contributed to SRM. The relative contributions of these two factors were not studied for broadband stimuli.

Previous physiology studies have identified possible neural substrates for both the energetic and ITD-processing components of SRM. For the energetic component, our group has shown that some inferior colliculus units, “SNR units,” have masked thresholds that are predicted by the signal-to-noise ratio (SNR) in a narrowband filter centered at the unit’s CF (Litovsky et al. 2001). For the ITD component, a series of studies (e.g. Jiang, McAlpine, and Palmer 1997) shows that ITD-sensitive units can exploit the differences between the interaural phase difference (IPD) of a tone and masker to improve the neural population masked thresholds. These studies did not describe how the units’ masked thresholds change when a broadband signal and masker are placed at different azimuths.

Here, we examine the contributions of energetic effects and binaural processing for broadband and low-frequency SRM using psychophysical experiments and an idealized population of SNR units. We also show that a population of ITD-sensitive units in the auditory midbrain exhibits a correlate of SRM. Finally, a model of ITD-sensitive units reveals that the signal’s temporal envelope influences the single-unit masked thresholds.

2 Psychophysics and modeling of SRM in humans

2.1 Methods

SRM was measured for three female and two male normal-hearing human subjects using lowpass and broadband stimuli. Azimuth was simulated using non-individualized head-related transfer functions (Brown 2000). Stimuli consisted of a 200-ms 40-Hz chirp train (broadband: 300-12,000 Hz; lowpass: 200-1500 Hz) masked by noise (broadband: 200-14,000 Hz, lowpass: 200-2000 Hz). The spectrum-level for the signal was fixed at 14 dB re 20 μ Pa/ Hz (56 dB SPL for the broadband signal). The masker level was adaptively varied using a 3-down, 1-up procedure to estimate the signal-to-noise ratio (SNR) yielding 79.4% correct detection performance. Stimuli were delivered via insert earphones to subjects in a sound-treated booth.

Inspired by the SNR units described above, predictions from a simple, "single-best-filter" model were used to evaluate if the SNR in the best narrow-frequency band can explain how masked threshold varies with signal and noise locations. The model analyzes SNR as a function of frequency, but does not allow for any across-frequency integration of information or any binaural processing. The model consists of a bank of 60 log-spaced gammatone filters (Johannesma 1972) for each ear. For each spatial configuration, the root-mean-squared energy at the output of every filter is separately computed for the signal and noise. The model assumes that the filter with the largest SNR (over the set of 120) determines threshold. The only free parameter in the model, the SNR yielding 79.4% correct performance, was fit to match the measured threshold when signal and noise were at the same location.

2.2 Results

Figure 1 shows measured (solid lines) and predicted (broken lines) thresholds as a function of noise azimuth for three signal azimuths (arrows). Two sets of model predictions are shown. Dash-dot lines show both lowpass and broadband predictions generated jointly for the model parameter fit to the broadband threshold measured with signal and masker co-located. Dotted lines show lowpass predictions generated with the model parameter fit to the measured lowpass threshold separately. Overall, performance is better for broadband (BB) stimuli than for lowpass (LP) stimuli (BB thresholds are always lower than LP). Further, the amount of SRM, the improvement in threshold SNR compared to the thresholds when signal and noise are co-located, is larger for broadband than lowpass stimuli (30 dB and 12 dB, respectively).

When the model parameter is fit separately for broadband and lowpass stimuli, predictions are relatively close to observed thresholds although lowpass predictions consistently underestimate SRM. These results suggest that for the chirp-train signals used, 1) the main factor influencing SRM for both lowpass and broadband stimuli is the change in SNR in narrow frequency bands, and 2) binaural processing increases SRM for lowpass, but not broadband stimuli.

When the same threshold SNR parameter is used to predict broadband and lowpass results (dash-dot lines), predicted thresholds are equal when signal and

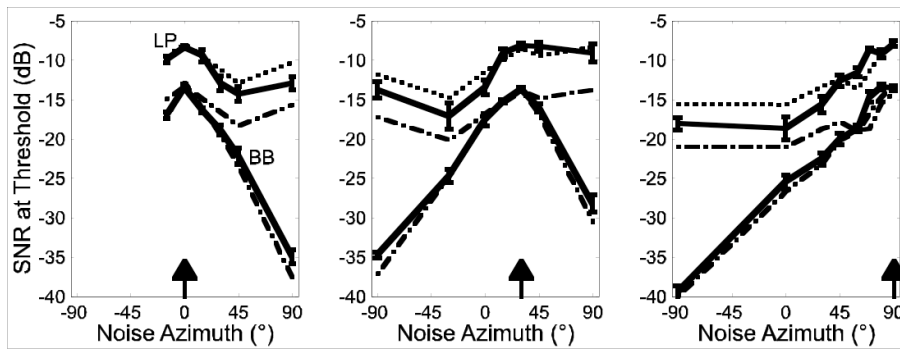


Fig. 1. SRM for human subjects for broadband (BB) and lowpass (LP) stimuli. Measured (subject mean and standard error) and predicted thresholds as a function of noise azimuth for three signal azimuths (arrows). Dash-dot line: lowpass and broadband model fit with same parameter; dotted line: lowpass data fit separately.

noise are co-located, regardless of stimulus bandwidth (because the SNR is constant across frequency when signal and noise are co-located). However, measured performance is always worse for the lowpass stimuli compared to the broadband stimuli. This result suggests that the listener integrates information across frequency, leading to better performance for broadband stimuli.

3 Neural correlates of SRM in the cat auditory midbrain

As shown above, the single-best-filter model underestimates the SRM for low frequencies. Here, thresholds for a population of ITD-sensitive neurons are measured to determine if these units can account for the difference between the single-best-filter model and behavioral thresholds.

3.1 Methods

Responses of single units in the anesthetized cat inferior colliculus were recorded using methods similar to those described in Litovsky and Delgutte (2002). The signal was a 40-Hz, 200-msec chirp train presented in continuous noise; both signal and noise contained energy from 300 Hz to 30 kHz. The chirp train had roughly the same envelope as the one used in the broadband psychophysical experiments. The signal level was fixed near 40 dB SPL, and the noise level was raised to mask the signal response. Results are reported for 22 ITD-sensitive units with characteristic frequencies (CFs) between 200 and 1200 Hz.

3.2 Results

Figure 2A shows the temporal response pattern for a typical ITD-sensitive unit as a function of noise level for the signal in noise (first 200 msec) and the noise alone (second 200 msec). The signal and noise were both placed at +90° (contralateral to the recording site). At low noise levels, the unit produces a synchronized response to the 40-Hz chirp train. As the noise level increases, the response to the signal is

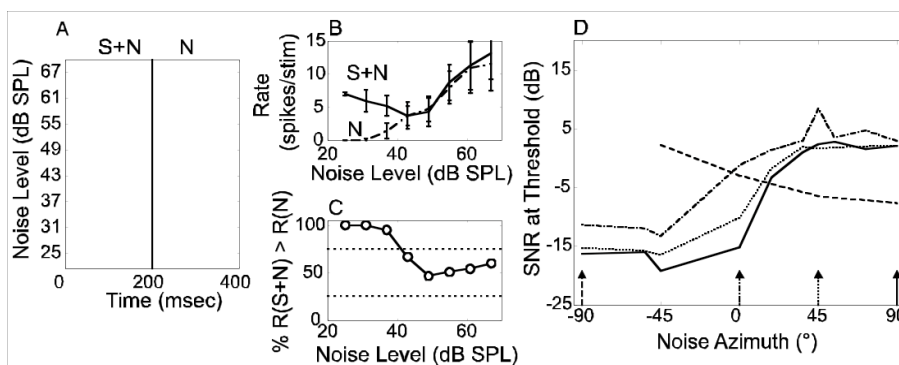


Fig. 2. A: Single-unit response pattern for signal in noise (S+N, 0-200 msec) and noise alone (N, 200-400 msec) for signal and noise at 90°. Signal level is 43 dB SPL. B: Rate-level functions for S+N and N from A. C: Percent of stimulus presentations that have more spikes for S+N compared to N. Threshold is the SNR at 75% or 25% (dotted lines). D: Same unit's masked thresholds as a function of noise azimuth for four signal azimuths (arrows indicate signal azimuth, arrow tail indicates corresponding threshold curve).

overwhelmed by the response to the noise (A, B). For this unit, +90° is a favorable azimuth so both the signal and the noise excite the unit. When placed at an unfavorable azimuth, the signal can suppress the noise response or vice versa.

Threshold is defined for single units as the SNR at which the signal can be detected through a rate increase or decrease for 75% of the stimulus repetitions (75% and 25% lines in Fig. 2C). Thresholds for this unit are shown in D as a function of noise azimuth for four signal azimuths. For three of the signal azimuths (-90°, 45°, and 90°), moving the noise away from the signal can improve thresholds by more than 15 dB. However, when the signal is at 0°, thresholds become slightly worse as the noise moves from the midline to the contralateral (positive azimuth) side. In other words, although some SRM is seen for some signal azimuths, no direct correlate of SRM can be seen in this, or any other, individual unit's responses for all signal and noise configurations.

A simple population threshold is constructed based on the same principle as the single-best-filter model (Section 2). For each signal and noise configuration, the population threshold is the best single-unit threshold in our sample of ITD-sensitive

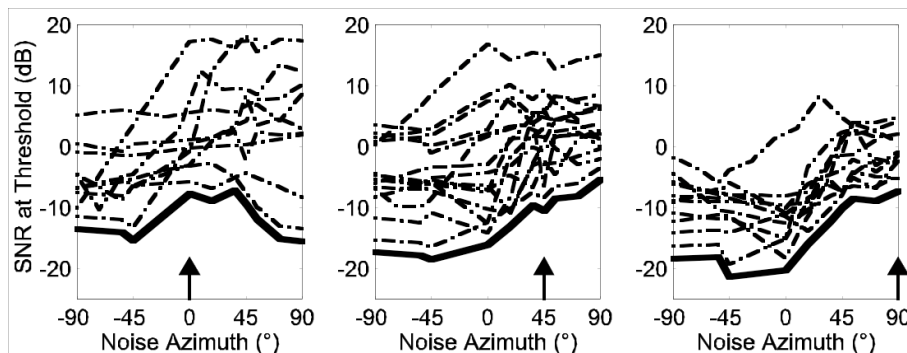


Fig. 3. Neural population thresholds for three signal azimuths (arrow). Dash-dot lines: single unit thresholds; solid lines: population thresholds (offset by 2 dB).

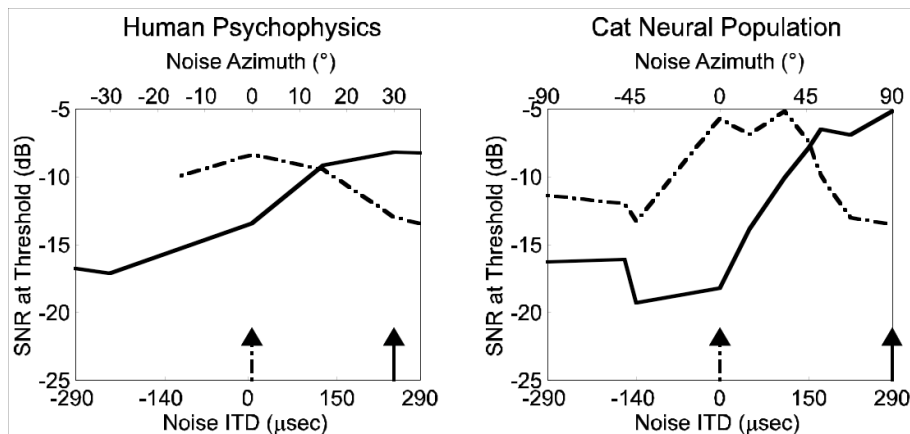


Fig. 4. Human psychophysical thresholds (left) and cat neural population thresholds (right) for two signal azimuths (arrows indicate signal azimuth, arrow tail indicates corresponding threshold curve) as a function of noise ITD (lower axis) and azimuth (upper axis).

units. Figure 3 shows the population thresholds (solid lines) as a function of noise azimuth for three signal azimuths (arrows). Unlike single unit thresholds (dot-dash), the population thresholds show SRM in that thresholds improve when the signal and noise are separated.

Figure 4 compares the low-pass human psychophysical thresholds (left) to the cat neural population thresholds (right). In order to compare the two thresholds despite the difference in species headsize, the axes are matched for noise ITD (lower axis) rather than noise azimuth (upper axis). The neural population thresholds are similar to the human behavioral thresholds, indicating that these ITD-sensitive units could provide a neural substrate for the binaural component of SRM.

3.3 Neural modeling of single-unit thresholds

Because our population consists of ITD-sensitive units, we attempted to model the unit responses using an interaural cross-correlator model similar to Colburn (1977). Figure 5A shows the thresholds for five units for which we measured thresholds for the signal at their best azimuths (+90°, squares) and their worst azimuths (-90°, circles). The noise was placed at the ear opposite the signal. For the data, the best-azimuth thresholds are better or equal to the worst-azimuth thresholds. In contrast, the cross-correlator model predicts that the worst-azimuth thresholds are better (Fig. 5B) because the largest change in interaural correlation occurs when the signal decreases the overall correlation. The cross-correlator, although able to predict the noise-alone response, failed to predict the response to the signal (not shown). The primary difference between the chirp-train signal and the noise is that the signal has a strong 40-Hz amplitude modulation while the noise envelope is relatively flat. Because many units in the IC have enhanced responses to modulated stimuli (Krishna and Semple 2000), we added an envelope processor that changes the rate response in proportion to the energy in the 40-Hz Fourier component of the cross-correlator's output. With envelope processing (Fig. 5C), best-azimuth thresholds are

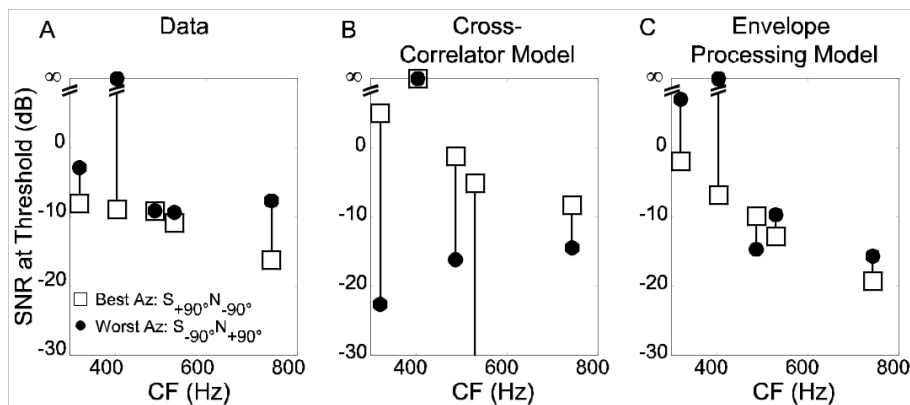


Fig. 5. A: Masked thresholds for 5 units. Best-azimuth thresholds (squares): signal at +90°, noise at -90°; worst-azimuth thresholds (circles): signal at -90°, noise at +90°. B,C: As in A for cross-correlator model (B) and cross-correlator model with envelope processor (C).

about the same or better than worst-azimuth thresholds, consistent with the data, because the envelope processor only changes the responses for favorable azimuths. These results suggest that 1) a traditional cross-correlator model cannot account for neural responses in the IC, 2) the temporal envelope can affect the detectability of signals in inferior colliculus neural responses, and 3) envelope processing is necessary to predict which units are best for signal detection (discussed below).

4 Discussion

Human listeners exhibit a large amount of SRM for both broadband and lowpass 40-Hz chirp-train signals. For broadband stimuli, the SNR in a single high-frequency filter predicts the amount of SRM, indicating high-frequency narrow-band energetic changes determine the SRM. SNR units, which have thresholds that are predicted by the SNR in a narrowband filter, could detect these changes.

For the lowpass condition, the single-best-filter model predicts some SRM, but underestimates the total amount by several dB. A correlate of the lowpass SRM is evident in the population response of ITD-sensitive units in the IC. It is possible, then, that there are two populations of neurons that can give SRM at low frequencies: an ITD-sensitive population and an SNR-unit population. When a listener is able to use the ITD-sensitive population, thresholds should improve by a few dB. When this population cannot be used (such as when the signal and masker are co-located or when listening monaurally), the SNR-unit population would determine performance, resulting in worse masked thresholds for some spatial configurations. These two hypothesized neural populations may respond differently to different stresses. For example, because the SNR population response depends on a neural population with narrow tuning and a wide range of CFs, relying on this population might be especially difficult for listeners with hearing impairment.

The envelope-processing model predicts that different ITD-sensitive populations, in either the left IC or the right IC, will dominate signal detection performance for different stimuli. The best single-unit thresholds for both the data

and the envelope-processing model occur when the chirp-train signal is positioned at a unit's best azimuth. Thus, for modulated signals, the IC contralateral to the signal yields better thresholds than the ipsilateral IC. However, for unmodulated signals, the model predicts that the best thresholds occur for the signal placed at the unit's worst azimuth. This prediction is consistent with previous studies (e.g. Jiang, McAlpine, and Palmer 1997) showing that the best single-unit thresholds for tones in noise occurred when the tone had an unfavorable IPD. Therefore, different ICs seem to be used for signal detection depending on the signal envelope.

Finally, human broadband thresholds are better than lowpass thresholds for all spatial configurations. Because this improvement is evident for co-located signals and maskers, the auditory system seems to integrate information across frequency. Because units in the IC are relatively narrowly tuned, auditory centers above the IC are also likely to be involved in the detection of broadband signals.

In summary, SRM seems to depend on binaural and energetic cues, which may be processed by separate neural populations. Neural processing related to SRM can be observed in the auditory midbrain, but centers higher than the midbrain also seem necessary for the integration of information across frequency.

References

- Brown, T. J. (2000). "Characterization of acoustic head-related transfer functions for nearby sources," unpublished M.Eng. thesis. Electrical Engineering and Computer Science, MIT, Cambridge, MA.
- Colburn, H. S. (1977) Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise. *J. Acoust. Soc. Am.* 61, 525-533.
- Good, M.D., Gilkey, R.H., and Ball, J.M. (1997) The relation between detection in noise and localization in noise in the free field. In R.H. Gilkey and T.R. Anderson (Eds), *Binaural and Spatial Hearing in Real and Virtual Environments*. Lawrence Erlbaum Associates, Mahwah, N.J, pp 349-376.
- Jiang, D., McAlpine, D., and Palmer, A.R. (1997) Detectability index measures of binaural masking level difference across populations of inferior colliculus neurons. *J. Neurosci.* 17, 9331-9339.
- Johannesma, P.I.M. (1972) The pre-response stimulus ensemble of neurons in the cochlear nucleus. In: B.L. Cardozo, E. de Boer, and R. Plomp (Eds.), *IPO Symposium on Hearing Theory*. IPO, Eindhoven, The Netherlands, pp. 58-69.
- Krishna, B.S. and Semple, M.N. (2000) Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. *J. Neurophysiol.* 84, 255-73.
- Litovsky, R.Y. and Delgutte, B. (2002) Neural correlates of the precedence effect in the inferior colliculus: Effect of localization cues. *J. Neurophysiol.* 87, 976-994.
- Litovsky, R.Y., Lane, C.C., Atencio, C., and Delgutte, B. (2001) Physiological measures of the precedence effect and spatial release from masking in the cat inferior colliculus. In: D.J. Breebaart, A.J.M. Houtsma, A. Kohlrausch, V.F. Prijs, and R. Schoonhoven (Eds). *Physiological and Psychophysical Bases of Auditory Function*. Shaker, Maastricht, pp. 221-228.
- Saberi, K., Dostal, L., Sadralodabai, T., Bull, V., and Perrott, D.R. (1991) Free-field release from masking. *J. Acoust. Soc. Am.* 90, 1355-1370.