Published in final edited form as:

Adv Exp Med Biol. 2013 ; 787: 255–262. doi:10.1007/978-1-4614-1590-9_29.

Neural correlates of the perception of sound source separation

Mitchell L. Day^{1,2,*} and Bertrand Delgutte^{1,2,3}

¹Eaton-Peabody Laboratories, Massachusetts Eye and Ear, Boston, MA 02114, USA

²Department of Otology and Laryngology, Harvard Medical School, Boston, MA 02115, USA

³Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Abstract

As two sound sources become spatially separated in the horizontal plane, the binaural cues used for sound localization become distorted from their values for each sound in isolation. Because firing rates of most neurons in the inferior colliculus (IC) are sensitive to these binaural cues, we hypothesized that these neurons would be sensitive to source separation. We examined changes in the target azimuth tuning functions of IC neurons in unanesthetized rabbits caused by the concurrent presentation of an interferer at a fixed spatial location. Both target and interferer were broadband noise bursts, uncorrelated with each other. Signal detection analysis of firing rates of individual IC neurons shows that responses are correlated with psychophysical performance on segregation of spatially-separated sources. The analysis also highlights the role of neural sensitivity to interaural time differences of cochlea-induced envelopes in performing this task. Psychophysical performance on source segregation was also compared to the performance of two contrasting maximum-likelihood classifiers operating on the firing rates of the population of IC neurons. The “population-pattern” classifier had access to the firing rates of every neuron in the population, while the “two-channel” classifier operated on the summed firing rates from each side of the brain. Unlike the two-channel classifier, the population-pattern classifier could segregate the sources accurately, suggesting that some of the information contained in the heterogeneity of azimuth tuning functions across IC neurons is used to segregate sources.

Keywords

interaural time difference; interaural level difference; interaural correlation; sound localization; neural decoding; inferior colliculus; rabbit

1. Introduction

Sound reaching the two ears exhibits both an interaural time difference (ITD) and an interaural level difference (ILD) which are used to determine the lateral position of the sound source. Both binaural cues become distorted in the presence of a spatially-separated interferer, and these changes are likely used by humans and other species to perceive spatially-separated sources.

Best et al. (2004) investigated the ability of human listeners to perceptually segregate two broadband noise sources, uncorrelated with each other, at various spatial separation. The two sources could be distinguished from a single source with nearly perfect accuracy when separated by as little as 15° when one of the sources was fixed at 0° (straight-ahead). A

*day@meei.harvard.edu.

greater spatial separation was required to achieve perceptual segregation when one of the sources was located more laterally.

We recorded from single neurons in the IC of unanesthetized rabbits to search for neural correlates of human perception of source separation. A “target” broadband noise source was presented at various azimuths in the frontal horizontal plane in the presence of a concurrent, spatially-separated noise source, uncorrelated with the target. We analyzed IC responses in two ways to compare to psychophysical results. First, we characterized neural signal detection based on changes in firing rate with spatial separation for each individual IC neuron. Next, we performed a classification of single sources versus two spatially-separated sources based on the population neural activity using either of two neural decoding strategies that represent opposite extremes of how information may be combined across IC neurons.

2. Methods

Methods for surgery, acoustic stimulus delivery, and single-unit electrophysiology for the unanesthetized, head-fixed rabbit preparation were essentially the same as described previously (Devore and Delgutte, 2010).

Sound stimuli in virtual acoustic space were presented binaurally through ear inserts and were first filtered with rabbit directional transfer functions (DTFs) associated with each azimuthal location in the horizontal plane. Two additional sets of DTFs with altered binaural cues were created for comparison to the “standard” condition in which ITD and ILD naturally co-vary with azimuth. For the “ITD-only” condition, all magnitude spectra were fixed to the spectrum at 0°, allowing ITD to vary naturally with azimuth while ILD and monaural levels remained fixed. Similarly, for the “fixed-ITD” condition, all phase spectra were fixed to the spectrum at 0°, allowing ILD and monaural levels to vary naturally with azimuth while ITD remained fixed.

Target and interferer stimuli were both frozen, broadband (0.1–18 kHz), 300-ms noise bursts, uncorrelated with each other, presented concurrently every 500 ms. Target and interferer were each presented at a sound level 20 dB above the neural threshold at 0°. For each IC neuron, target azimuth tuning functions (average firing rate over the stimulus duration vs. target azimuth) were collected in 15° steps with the interferer either co-located with the target (equivalent to a “single-source” condition) or fixed at 0° (“central interferer”). Azimuth tuning functions were measured under standard, ITD-only, and fixed-ITD conditions. An additional set of responses was collected under the standard condition with both target and interferer azimuths independently varied in 30° steps covering every possible spatial combination of target and interferer in the frontal hemifield. Eight stimulus trials were collected for each spatial combination.

We investigated the ability of the rate responses of individual neurons to signal the separation of the target from a central interferer (i.e., neural signal detection). For each spatial separation between target and interferer, we calculated a neural sensitivity index d' to characterize the change in neural firing rate between co-located and separated source conditions. The neural detection threshold was defined as the (interpolated) target azimuth closest to 0° at which $d'=1$ (e.g., Fig. 2B). Neural detection thresholds were calculated from responses collected under standard, ITD-only, and fixed-ITD conditions.

We also investigated the ability of two contrasting neural decoding strategies based on population IC activity to detect the separation of target and interferer anywhere in the frontal hemifield. For both classifiers, a conditional probability density of the population activity was estimated for each target/interferer combination, assuming multivariate Gaussian

densities. “Left” and “right” populations of IC neurons were created by making a mirror twin of every neuron. The “population-pattern” classifier operated on a $2N$ -dimensional vector corresponding to the firing rates of the N neurons in our sample and their mirror twins. The “two-channel” classifier operated on the 2-dimensional vector formed by the summed firing rates across the left and right IC populations. For 500 iterations of a Monte Carlo procedure, a random stimulus trial was removed from each neuron before the mean and variance of each dimension were calculated. The removed trial was then used to test the classifier performance in that iteration by classifying it to the spatial combination having the maximum likelihood. Classifier performance was assessed as the fraction of test trials in which the classifier correctly distinguished between a single source and spatially-separated sources, regardless of localization accuracy.

3. Results

To investigate the physiological correlates of perception of source separation, we measured the responses to concurrent target and interferer stimuli from 99 neurons in the right IC of two female, Dutch-belted rabbits. Best frequencies (BFs) ranged from 0.19 to 17 kHz. A characteristic feature observed in many neurons was a pronounced peak or notch at 0° in the target azimuth tuning function with a central interferer (Fig. 1B, solid line, “notch”), while there was no such feature in the single-source tuning function (Fig. 1A, solid line). In other words, the firing rate was either sharply suppressed or enhanced as the target was separated by just 15° in either direction from the central interferer. Rate suppression or enhancement was only found in neurons with high BFs (>1.5 kHz), and occurred in a majority (72%) of these neurons (Day et al., in press).

The restriction of rate suppression and enhancement to high-BF neurons suggests the involvement of ILDs or envelope ITDs, as these are the binaural cues to which high-BF neurons are sensitive (Delgutte et al., 1995; Joris, 2003). To identify the relevant cues, we compared azimuth tuning functions between altered-cue and standard conditions. For the high-BF neuron in Fig. 1A, the single-source tuning function in the fixed-ITD condition resembles that in the standard condition, while the tuning function in the ITD-only condition is nearly flat, suggesting that tuning is primarily determined by ILD. In contrast, in the presence of a central interferer (Fig. 1B), the fixed-ITD tuning function no longer resembles the standard tuning function, while the ITD-only tuning function shows rate enhancement similar to that observed in the standard condition. Therefore, the rate enhancement with target separation from a central interferer was dependent on sensitivity to ITD, although the single-source tuning function was primarily dependent on ILD. This neuron was representative in that rate suppression or enhancement were consistently observed in the ITD-only condition but not the fixed-ITD condition.

Rate suppression or enhancement can be explained by a cross-correlation model operating on cochlea-induced envelopes of the ear input signals. In this model, left and right sound stimuli were bandpass filtered (0.25-oct bandwidth) to mimic cochlear processing, and the normalized correlation (Trahiotis et al., 2005) was computed between the Hilbert envelopes of the left and right bandpass-filtered waveforms. As the ITD of a broadband source is varied, the interaural cross-correlation coefficient (IACC) decreases *slowly* (Fig. 1C, dashed line) because the envelopes being cross-correlated only contain low frequencies. However, when a central interferer is introduced (Fig. 1C, solid line), the IACC changes more *rapidly* with target ITD, oscillating at the period of the filter center frequency (3.19 kHz in Fig. 1C). While the envelopes in this case are still limited to low frequencies, the rapid oscillation occurs because the waveform interactions between the target and interferer that determine the mixture envelope are dependent on the temporal fine structure after cochlear filtering. The behavior of IC neurons that are sensitive to envelope ITD is consistent with model

predictions. While the firing rates of IC neurons typically change little when a single source moves by 15° (Fig. 1A, dashed line), the azimuth tuning functions with a central interferer show sharp changes in firing rate as the cochlea-induced envelopes become strongly decorrelated when the target is separated by 15° from the interferer.

Best et al. (2004) showed that the perceptual segregation performance for two broadband noise sources (with one fixed at 0°) remains good when the stimuli are highpass filtered, but is degraded in the fixed-ITD condition, suggesting the use of envelope ITD for segregation. The sharp rate suppression or enhancement seen in high-BF IC neurons is a possible neural correlate of this ability. To test this idea, we calculated a neural detection threshold for the spatial separation of the target from the central interferer for each high-BF neuron in our sample (Fig. 2A and B). The median detection threshold across the population of high-BF neurons was small ($< 8^\circ$) and nearly the same for ITD-only and standard conditions, but was 4 times larger for the fixed-ITD condition (Fig. 2C). Thus, the sensitivity of IC neurons to envelope decorrelation plays a key role in the neural detection of source separation for high-pass sounds.

Best et al. (2004) further showed that the amount of spatial separation between target and interferer necessary for perceptual segregation increases when the interferer is placed more laterally (Fig. 3A). We tested the ability of two neural decoding strategies to segregate two sources anywhere in the frontal hemifield based on the firing rates of the population of IC neurons. The “population-pattern” classifier takes as input the vector of firing rates for every neuron in the population and classifies the population firing pattern into either “single source” or “two sources.” This strategy represents the best performance that can be achieved from the firing rates of IC neurons under the classifier assumptions. In contrast the “two-channel” classifier operates only on the summed rates across IC neurons for each side of the brain. Such a two-channel model of sound localization has been shown to be consistent with some psychophysical results for single sources (Stecker et al., 2005; Devore et al., 2009; Lesica et al., 2010).

The population-pattern classifier was able to perfectly segregate sources with 30° separation when the interferer was fixed at 0° or 30°, but required a wider separation at 60° and 90° to attain perfect segregation (Fig. 3B), similar to human perception. The two-channel classifier, however, failed to correctly identify single sources at 0°, 30° and 60° (Fig. 3C). For example, the two-channel classifier often mistook a single source at 30° for two sources at 30° and 60°, respectively. The classifier made this error because the left and right summed rates for a single source at 30° are very similar to those for 30° and 60° (Fig. 3D). The population-pattern classifier, on the other hand, has much more information available in its inputs. In Fig. 3E, the firing rates of each neuron in our sample in response to a single source at 30° are shown in decreasing order (gray bars). Maintaining the same order, the pattern of firing rates in response to sources at 30° and 60° (black bars) differed greatly from the pattern for 30°. The population-pattern classifier can use information contained in the heterogeneity of azimuth tuning functions across neurons to accurately segregate sources, while this information is averaged away with the two-channel classifier.

4. Discussion

The majority of high-BF neurons in the IC show a sharp rate suppression or enhancement when one broadband source is separated by just 15° from another located at 0°. These sharp rate changes are due to neural sensitivity to the decorrelation of cochlea-induced envelopes that occurs with source separation. We showed that this rate suppression or enhancement is a neural correlate of the perceptual segregation of concurrent, spatially-separated sources. Further, we showed that a maximum-likelihood classifier operating on the pattern of firing

rates across the population of IC neurons can account for psychophysical data on perceptual segregation anywhere in the frontal hemifield by making use of the information contained in the heterogeneity of azimuth tuning functions across neurons.

The success of the population-pattern classifier on source segregation does not prove that this decoding strategy is used in the brain. This classifier assumes that some higher auditory center keeps track of the response of every IC neuron. At minimum, our result is a proof-of-principle that the information contained in rabbit IC firing rates is sufficient to approximate human performance on source segregation. However, the failure of the two-channel classifier does suggest that this particular neural decoding strategy cannot be used to segregate sources, even though it works adequately for the localization of a single source. Our results show that neurons that respond similarly to a single source at a given location can respond in dramatically different ways in the presence of a spatially-separated interferer (Fig. 3E). It is likely that the neural decoding strategy used at higher auditory centers takes into account at least some of this heterogeneous tuning across neurons.

The segregation study of Best et al. (2004) required subjects to respond either “one source” or “two sources.” However, a similar study of the localization of a target broadband noise in the presence of an interferer found that human listeners usually perceived a single auditory event and less often two spatially diffuse events when there were no onset or offset differences between the two stimuli (Braasch, 2002). As such, the subjects’ responses in the Best et al. study may be better interpreted as “single source” and “diffuse source,” and similarly the population-pattern classifier may be detecting the diffuseness of the sound percept rather than two distinct source locations. The neural mechanisms involved in source separation are relevant to sound localization in reverberation, which can be interpreted as a superposition of many attenuated, separated (but coherent) sources. The effects of envelope decorrelation have also been observed in the responses of IC neurons under reverberant conditions (Devore and Delgutte, 2010). Therefore it is likely that the neural mechanisms for the perception of source separation are more generally involved in the coding of spatial properties of the surrounding environment.

Acknowledgments

This work was supported by NIDCD (US) grants R01 DC002258 and P30 DC005209.

References

- Best V, van Schaik A, Carlile S. Separation of concurrent broadband sound sources by human listeners. *J Acoust Soc Am*. 2004; 115:324–36. [PubMed: 14759025]
- Braasch J. Localization in the presence of a distracter and reverberation in the frontal horizontal plane: II. Model algorithms *Acta Acustica United with Acustica*. 2002; 88:956–969.
- Day ML, Koka K, Delgutte B. Neural encoding of sound source location in the presence of a concurrent, spatially-separated source. *J Neurophysiol*. in press.
- Delgutte, B.; Joris, PX.; Litovsky, RY.; Yin, TC. Relative importance of different acoustic cues to the directional sensitivity of inferior-colliculus neurons. In: Manley, GA.; Klump, GM.; Koeppl, C.; Fastl, H.; Oeckinghaus, H., editors. *Advances in Hearing Research*. World Scientific Publishing; Singapore: 1995. p. 288-99.
- Devore S, Delgutte B. Effects of reverberation on the directional sensitivity of auditory neurons across the tonotopic axis: influences of interaural time and level differences. *J Neurosci*. 2010; 30:7826–37. [PubMed: 20534831]
- Devore S, Ihlefeld A, Hancock K, Shinn-Cunningham B, Delgutte B. Accurate sound localization in reverberant environments is mediated by robust encoding of spatial cues in the auditory midbrain. *Neuron*. 2009; 62:123–34. [PubMed: 19376072]

- Joris PX. Interaural time sensitivity dominated by cochlea-induced envelope patterns. *J Neurosci.* 2003; 23:6345–50. [PubMed: 12867519]
- Lesica NA, Lingner A, Grothe B. Population coding of interaural time differences in gerbils and barn owls. *J Neurosci.* 2010; 30:11696–702. [PubMed: 20810890]
- Stecker GC, Harrington IA, Middlebrooks JC. Location coding by opponent neural populations in the auditory cortex. *PLoS Biol.* 2005; 3:e78. [PubMed: 15736980]
- Trahiotis, C.; Bernstein, LR.; Stern, RM.; Buell, TN. Interaural correlation as the basis of a working model of binaural processing: an introduction. In: Popper, AN.; Fay, RR., editors. *Sound Source Localization*. Springer; New York: 2005.

Figure 1.

Target azimuth tuning functions of one IC neuron under altered-cue and standard conditions for a single source (A) and with a central interferer (B). (C) Target ITD tuning functions from a cross-correlation model using a cochlear filter with center frequency, $f_{center}=3.19$ kHz. Interferer location (solid triangles); ITD equal to one period of f_{center} (open triangles).

Figure 2.

Neural signal detection. (A) Target azimuth tuning function in the presence of a central interferer for a single IC neuron (bars: 1 SD). Triangle indicates interferer location. (B) d' calculated from the data in (A) showing the detection thresholds (dotted lines) where $d' = 1$ (dashed line). (C) Distribution of detection thresholds across high-BF neurons (N = 58) showing medians (dashed lines) and interquartile ranges (boxes).

Figure 3. Neural decoding. (A) Listener performance on the perception of separate sources as a target was separated from an interferer [modified from Best et al. (2004), Fig. 3e]. Each line indicates performance for an interferer fixed at the location marked by a triangle with matching grayscale. Performance of the population-pattern (B) and two-channel (C) classifiers on the same task. Mean rates of the two-channel (D) and population-pattern (E) classifiers for two spatial combinations (single source at 30° and two sources at 30° and 60°).