

Available online at www.sciencedirect.com



Journal of Theoretical Biology 235 (2005) 45-56

Journal of Theoretical Biology

www.elsevier.com/locate/yjtbi

Neural cross-correlation and signal decorrelation: insights into coding of auditory space

Kourosh Saberi*, Agavni Petrosyan

Department of Cognitive Sciences, University of California, Irvine, CA 92697, USA

Received 12 June 2004; received in revised form 30 September 2004; accepted 14 December 2004 Available online 25 February 2005 Communicated by Steven Baer

Abstract

The auditory systems of humans and many other species use the difference in the time of arrival of acoustic signals at the two ears to compute the lateral position of sound sources. This computation is assumed to initially occur in an assembly of neurons organized along a frequency-by-delay surface. Mathematically, the computations are equivalent to a two-dimensional cross-correlation of the input signals at the two ears, with the position of the peak activity along this surface designating the position of the source in space. In this study, partially correlated signals to the two ears are used to probe the mechanisms for encoding spatial cues in stationary or dynamic (moving) signals. It is demonstrated that a cross-correlation model of the auditory periphery coupled with statistical decision theory can predict the patterns of performance by human subjects for both stationary and motion stimuli as a function of stimulus decorrelation. Implications of these findings for the existence of a unique cortical motion system are discussed. © 2005 Elsevier Ltd. All rights reserved.

Keywords: Auditory space; Dynamic signals; Motion; Cross-correlation; Decorrelation

1. Introduction

The aim of this report is to describe a neurocomputational model of auditory-space coding and to experimentally demonstrate how it can predict the ability of humans to detect moving and stationary signals. At the core of the model lies mathematical cross-correlation, an abstraction of the neural operations performed by structures in the auditory periphery on the input signals arriving from the two ears (Carr and Konishi, 1988, 1990; Yin and Chan, 1990; Konishi, 1992, 1993a, 2000). Experiments are described on stationary and moving sound sources. An important parameter of the experiments is the correlation between the signals presented to the left and right ears. Decorrelation results in a decline in the ability to process spatial cues. A cross-correlation

E-mail address: saberi@uci.edu (K. Saberi).

model of the auditory periphery that attempts to correlate these partially decorrelated signals, coupled with a signal-detection theory analysis of the available information at the output of the cross-correlation model, provides insight into the joint mechanisms that underlie motion and stationary sound-source processing. We begin with a brief introduction to the literature on auditory motion detection, followed by a description of our experiments on decorrelated auditory events, and demonstrate that at an early stage of the auditory pathway, moving and stationary sounds may be processed in a similar way (contrary to the vision motion-processing system; Lu and Sperling, 1995; Clifford and Ibbotson, 2002; Vaina and Soloviev, 2003).

Auditory motion perception has been investigated for over a century (Dove, 1839; Mach, 1874; Thompson, 1877, 1878; Rayleigh, 1876, 1877; Peterson, 1916; Valentine, 1928). Psychophysical studies of motion have examined the effects of velocity (Altman and Viskov, 1977; Perrott and Musicant, 1977a; Waugh et al., 1979;

^{*}Corresponding author. Tel.: +19498246310;

fax: +19498242307.

^{0022-5193/\$ -} see front matter © 2005 Elsevier Ltd. All rights reserved. doi:10.1016/j.jtbi.2004.12.018

Grantham, 1986; Saberi and Hafter, 1997), acceleration (Perrott et al., 1993), bandwidth (Perrott and Tucker, 1988; Chandler and Grantham, 1992; Saberi, 1996; Strybel and Menges, 1998), trajectory (Saberi and Perrott, 1990; Saberi et al., 1991), multisource motion (Saberi et al., 2002a), as well as a number of other parameters (Perrott and Nelson, 1969; Perrott and Musicant, 1977b; Grantham, 1986; Strybel et al., 1989, 1998; Strybel and Neale, 1994). These studies have shown that motion detection declines at high velocities. and is better for azimuthal or oblique trajectories than for vertical. Studies that have isolated the role of dynamic interaural cues have revealed that at high velocities, motion detection is less salient when based on a dynamic interaural delay compared to a dynamic interaural level-difference cue. This finding has been referred to as "lag of lateralization" (Blauert, 1972) or "binaural sluggishness" (Grantham, 1984; Grantham and Wightman, 1978) and implies a lowpass filtering of the rate of changing interaural delay. Neurophysiological studies have identified brainstem and cortical neurons that respond preferentially to one direction of motion and are silent, non-responsive, or inhibitory in response to other directions (Spitzer and Semple, 1991; Moiseff and Haresign, 1992; Stumpf et al., 1992; Ahissar et al., 1992). Recent functional neuroimaging studies have examined human cortical activation in response to motion stimuli, with some studies implicating the parietal lobe and planum temporale as uniquely associated with auditory motion and other studies disputing this finding (Warren et al., 2002; Pavani et al., 2002; Smith et al., 2004).

One area of motion processing that has not been investigated concerns signals that are only partially correlated at the two ears. Natural signals are never perfectly interaurally correlated due to differential filtering effects of the pinna (Butler, 1975, 1977; Rayleigh, 1907; Shaw, 1965, 1974; Kuhn, 1987), presence of multiple external sound sources (Yost et al., 1996; Yost, 1997) as well as independent internal neural noise added to the signal prior to binaural convergence in the ascending auditory pathway (Robinson and Jeffress, 1963). Psychophysical research has shown that subjects can localize partially correlated stationary noise for correlation values as low as 0.3 (Jeffress et al., 1962). Consistent with this finding, neurophysiological research has shown that optic tectum and nucleus laminaris neurons have identifiable spatial receptive fields for correlations of 0.3-0.4 (Albeck and Konishi, 1995; Saberi et al., 1998). Interestingly, decorrelation does not significantly affect interaural level coding either neurophysiologically (Egnor, 2001) or behaviorally, even when interaural correlation is zero (Egnor, 2001; Hartmann and Constan, 2002). Here, we examine interaural delay sensitivity for the detection of motion as a function of velocity and interaural correlation (0.1-1.0) and compare these results to those for stationary stimuli obtained from the same observers. It is shown that a cross-correlation model of binaural interaction predicts that changes in the mean and variance of the estimated cross-correlation peak as a function of decorrelation may underlie the observed patterns of performance for both dynamic (motion) and stationary conditions.

2. Stimulus generation and calibration.

All stimuli were generated digitally and presented via 16-bit digital-to-analog converters (Sound Blaster Live, -120 dB noise floor, Milpitas, CA). The sampling rate was 44.1 kHz, and the analog output was filtered through 20-kHz anti-aliasing filters. Stimulus generation and presentation was controlled via software running on a PC workstation. Stimulus levels were calibrated to 70 dB (A-weighting) using a 6cc coupler, 0.5" microphone (Brüel and Kjær, Model 4189) and a modular precision sound analyser (Brüel and Kjær, Type 2260). The waveforms to the two ears had simultaneous onsets, but no rise-decay times to avoid introduction of a potential envelope interaural correlation cue (although this correlation would not carry information for resolving the task). The timing and levels between left and right channels were checked for accuracy using a dual-channel digital storage oscilloscope (Tektronix, Model TDS210). All calibrations were conducted in a double-walled steel acoustically isolated chamber (Industrial Acoustics Company).

Stimuli were Gaussian noise bursts filtered between 0.1 and 10 kHz. Generation of motion stimuli was a two-step process, with both steps occurring in "real-time" between trials. In the first step, a dichotic Gaussian noise waveform was produced with a dynamic linear change in interaural delay, and in the second step, the noise waveforms to the two channels were partially decorrelated (when required) by addition of independent Gaussian noises to each channel.

To generate a noise waveform with a dynamic linear shift in interaural delay (i.e. motion) we first generated a Gaussian noise sample in the frequency domain with amplitudes sampled from a Rayleigh distribution and phases from a uniform $(0, 2\pi)$ distribution. The spacing between frequency components $(\Delta \omega)$ of this waveform is dependent on duration $(\Delta \omega = 1/T)$, and the discrete fourier transform (DFT) will only have measured energy at the harmonics of T (Rabiner and Gold, 1975). To generate the stimulus for the second channel, we selected a duration (T_2) for the waveform to channel 2 that produces a DFT array with component spacing $\Delta \omega_2$, such that each frequency component of channel 2 would be shifted relative to the corresponding component in channel 1 by a proportion required to produce a binaural-beat waveform with a linear shift in interaural delay. This array is then filled with complex numbers (amplitudes and phases) from channel 1. For example, a 0.1s stimulus will have DFT component spacing of 10 Hz at {0,...,500,510,...,1000,1010,...} Hz. To produce a motion velocity of 2000 µs/s, channel 2 will require a frequency spacing of $\Delta \omega_2 = 10.02 \text{ Hz}$, a duration $T_2 = 1/\Delta\omega_2 = 99.800399 \,\mathrm{ms}$, and thus, component frequencies of {0,...,501,511.02,...,1002, 1012.02,...} Hz.¹ Inverse FFT of the stimuli at channels 1 and 2 provides the time waveforms for these channels. We have used this procedure to generate linear interaural-delay-based motion for a number of complex sounds, including noise waveforms, natural sounds, and speech sentences. The perceived image is of a smoothly moving intracranial image along the interaural axis (see Saberi, 2004 for details).

Even though the interaural delay of such waveforms is dynamic, we label this dichotic condition as having an interaural correlation of 1, consistent with binaural studies of stationary sounds (Jeffress et al., 1962; Saberi et al., 1998) and because such a stimulus is perceived as a punctuate auditory image (for a motion velocity of zero, the measured correlation is unity). To produce stimuli with an interaural correlation of r we added independent bursts of Gaussian noise (0.1–10 kHz) to each channel (Jeffress and Robinson, 1962):

$$X_L = kX_1 + X_c,$$

$$X_R = kX_2 + X_c,$$
(1)

where X_L and X_R are the left and right channel waveforms, respectively, $k = \sqrt{(1/r) - 1}$ is a scalar for independent noise samples X_1 and X_2 , and X_c is the noise common to both ears (generated in the previous step). By definition, when r is 0, two-independent noiseburst are used. In the current study, X_c was either interaurally delayed when the task was localization of stationary sounds, or it included a dynamic shift in interaural delay when the task was the detection of motion. The stimulus levels (X_L and X_R) were calibrated to 70 dB SPL.

3. Procedure

Three normal-hearing subjects participated in the experiments. Each subject practiced on the various experimental conditions for 2 h prior to data collection. Subjects listened through Sennheizer HD-400 head-phones in a double-walled steel acoustically isolated chamber (IAC). There were 25 experimental conditions

(5 velocities by 5 correlations) and each subject completed two runs of 100-trials per condition. The five velocity conditions were 0 (stationary), 250, 500, 2500, and 5000 μ s/s. For each velocity, there were five interaural correlation conditions of 0.1, 0.3, 0.5, 0.7, and 1.0. The experiments were run in a block design with correlation and velocity held constant within a run. The order in which subjects ran the various conditions was completely randomized.

3.1. Stationary conditions

The stationary-stimulus conditions were run in a twoalternative forced-choice (2IFC) design. On each trial, a 500-ms dichotic noiseburst was presented with an interaural delay of 250 µs leading to one ear, followed by a 250-ms silent interstimulus interval, followed by a 500-ms dichotic noiseburst with an interaural delay of 250 µs leading to the opposite ear (i.e. a total interaural delay change of $500 \,\mu$ s). In addition, to reduce absolute position cues (which becomes important for motion conditions) a random offset interaural delay was incorporated into the design. The random offset allowed the interaural delays to be selected from a range from -500 to 500 µs, where the negative sign denotes a waveform leading to the left ear. Although the total change in interaural delay, from one interval to the next, was fixed at 500 µs (as described above) this change could occur within any region from -500 to $500 \,\mu s$. For example, on a given trial the interaural delays in the two intervals could have been -400 and +100, or -200 and +300, etc. The ear to which the waveform led in the first interval was randomly selected on each trial with equal a priori probabilities. The subject's task was to indicate which ear received the leading waveform in the first interval by determining if the order of the subjective lateral positions of the two images in the two intervals of the 2IFC was left-then-right or right-then-left. Feedback was provided immediately after each response.

3.2. Motion conditions

The motion conditions were run in a single-interval forced-choice design in which the subject had to indicate whether the sound moved or was stationary. On a given trial either a single stationary stimulus (constant interaural delay) or a single motion stimulus was presented with equal prior probabilities. For a motion trial, the direction of motion was randomly selected toward right or left along the interaural axis. The motion stimulus had a maximum change in interaural delay of 500 μ s, and a velocity selected from the set 250, 500, 2500, and 5000 μ s/s (held constant within a run). These velocities were chosen because previous studies have shown that they produce a wide range of motion-detection thresholds (Saberi et al., 2002a, 2003). The

¹The duration is rounded to the nearest digital sampling period, which for the current study is $\sim 23 \,\mu s$. Thus, the maximum quantization error in estimating duration is 11.5 μs , which for a 1 s stimulus results in a maximum frequency shift of less than 0.00002 Hz.

"distance" traveled (500 μ s) was equal to the change in interaural delay for the stationary condition described above (500 μ s between the two intervals of the 2IFC design). As was the case for the stationary conditions a random offset interaural delay was used to reduce absolute position cues. The offset allowed the dynamic interaural delay to cover a 500- μ s distance selected from the range of -500 to 500 μ s. For example, the motion stimulus could have had a starting interaural delay of -300 μ s and a terminating interaural delay of + 200 μ s, or a starting delay of + 50 μ s and an ending delay of -450 μ s. If the trial was a "stationary" trial, a single stimulus was presented which had the same duration as that on a motion trial, and an interaural delay randomly selected from the range of -500 to 500 μ s.

4. Detection of stationary and dynamic signals as a function of interaural decorrelation

Fig. 1 shows results from three subjects (symbols) and their mean performance (solid line). The data are from the same three subjects in all panels. The abscissa represents interaural correlation and the ordinate represents index of detectability d' from statistical decision theory (Green and Swets, 1966). For the 2IFC stationary stimulus condition, d' was calculated from $d' = \sqrt{2}\Phi^{-1}(P(c))$, where P(c) is proportion correct responses and Φ^{-1} is the inverse of the normal cumulative distribution function (Macmillan and Creelman, 1991). A ceiling value of d' = 3.29 was imposed on performance by assuming a 2% inattention rate (P(c) = 0.99) to reduce estimation problems associated with the high variance of large d' values given limited sample sizes on which the d' estimate is based (Macmillan and Creelman, 1991; Saberi and Green, 1997; Watson and Pelli, 1983).

In the stationary condition (upper-left) all subjects easily detected interaural delays in waveforms with interaural correlations as low as 0.3. Near-perfect performance is observed for correlation values of 0.5 and higher. At a correlation of 0.3 all subjects performed above threshold (d' = 1). Even for a correlation value as low as 0.1, one subject performed above threshold while the other two subjects performed above chance (d' = 0). It is clear from these data that the binaural system is extremely robust in detecting the interaural delay of highly degraded acoustic stimuli.

The remaining panels of Fig. 1 show results from the four motion conditions with velocity indicated within each panel. For the motion conditions, d' was calculated from hit and false-alarm rates associated with the 200 trials per velocity, correlation, and subject. A hit (H) was defined as a "motion" response when the stimulus had a dynamic interaural delay, and a false-alarm (FA) was defined as a "motion" response when the stimulus

interaural delay was constant (stationary). Motion detection was then calculated from d' = z(H) - z(FA) where the z-transform is the inverse of the normal distribution function (Macmillan and Creelman, 1991). As in the stationary stimulus condition, a ceiling value of d' = 3.29 was imposed on performance.

The patterns of performance for the 250 and $500 \,\mu s/s$ velocity conditions are similar to that for the stationary stimulus condition with near-perfect performance at correlations of 0.5 and higher (except for one subject at a correlation of 0.5 and velocity of $500 \,\mu s/s$). At these two lower velocities, all subjects showed above threshold performance for an interaural correlation of 0.3, and below threshold, but above-chance detection at a correlation of 0.1. At the two higher velocities of 2500 and 5000 µs/s performance substantially deteriorates at all correlation values. A 5×5 repeated-measures analysis of variance on the data of Fig. 1 showed a statistically significant effect of interaural correlation $(F_{4,8} = 323.35, p < 0.001)$ a significant effect of velocity which includes the stationary conditions as a velocity of 0 ($F_{4,8} = 53.70$, p < 0.001) and a significant interaction effect between these two factors ($F_{16,32} = 6.59$, p < 0.001).

Fig. 2 shows a 3D surface plot of the entire motion data set and a top-down view of the same data with interpolated shading. Color bands represent equal-detectability contours with higher d' values in red and lower values in blue. Note a nearly linear tradeoff between interaural correlation and velocity, particularly for midrange d's. A decrease in correlation must be accompanied by a proportional decrease in velocity to maintain a relatively constant d'. This is less obvious for high d's (dark red) where an asymptotic performance is reached (3.29) or very low values (dark blue) where performance is likely limited by internal neural noise (i.e. a floor effect).

5. Cross-correlation of partially correlated signals

The current study examined how the detection of spatial cues is affected by decreasing the correlation between the waveforms to the right and left ears, as often occurs in natural acoustic signals. The low-velocity motion system was shown to be highly resilient to decorrelation, with above-threshold performance for interaural correlations as low as 0.3. Performance at low velocities was similar to, though slightly below, that for localization of stationary signal. These results are consistent with findings reported in earlier studies of lateralization of interaurally decorrelated stationary sounds (Jeffress et al., 1962). At high velocities, performance deteriorated much faster. We should qualify, however, that our description of low vs. high velocity is relative. A 500 µs/s velocity is roughly



Fig. 1. Observer performance for discrimination of stationary and moving sound stimuli as a function of velocity and interaural correlation. Top-left panel shows localization performance for a stationary source, and the remaining panels display performance for detection of motion as a function of velocity indicated within each panel. Symbols represent different observers and the solid line is average performance.

equivalent to 80° /s along the azimuth in freefield, which is associated with a fast-moving source. High velocities (e.g. 2500 and 5000 µs/s) are not likely to be associated with an auditory source in freefield motion, but rather may be naturally achieved as a result of rapid head rotation in a stationary sound field, and thus a highvelocity motion system may be invoked primarily to maintain perceptual constancy. The current findings are also consistent with a recent neuroimaging study of human cortical activation in response to interaurally decorrelated stationary stimuli (Budd et al., 2003). This study reported a monotonic relationship between local cortical activation levels and interaural correlation in the primary auditory cortex located bilaterally at the lateral extent of Heschl's gyrus, a primarily low-frequency region of the auditory cortex.



Fig. 2. Motion detection as a function of velocity and interaural correlation shown in a 3D and top-down view of a surface plot. Colors are interpolated across values. Averaged data from three observers.

Activation patterns for correlation values near unity were more distinct than those near zero.

As noted earlier, a decrease in interaural correlation affects neural (Saberi et al., 1998) as well as behavioral (Jeffress et al., 1962; Saberi et al., 1998) sensitivity to interaural delays, but not to interaural level differences (Egnor, 2001; Hartmann and Constan, 2002). These observations, in addition to the well-established finding that interaural delays are initially coded in the medial superior olivary complex (MSO) prior to convergence with interaural level information at higher nuclei (i.e. inferior colliculus), suggests that decorrelation has its performance-degrading effects primarily at the level of the MSO. Quantitative models (Sayers and Cherry, 1956; Lindemann, 1986; Saberi, 1995; Saberi et al., 2002b) of the MSO have previously been used to study a number of binaural phenomena. These models use mathematical cross-correlation analysis to simulate neural activation patterns resulting from coincidencedetection and delay-line networks in the MSO (Carr and Konishi, 1988, 1990; Yin and Chan, 1990). It is therefore natural to ask if the degradation of cross-correlation patterns may be predictive of psychophysical performance in localization of interaurally decorrelated sounds, and if one can infer, using such a model, the factors that contribute to motion detection as interaural correlation is varied.

Here, we employ a model of cross-correlation analysis that incorporates several stages of peripheral auditory processing, coupled with statistical decision theory to predict localization and motion-detection of interaurally decorrelated signals. The signals to the left and right channels are initially processed through a bank of 30 bandpass fourth-order GammaTone filters (Holdsworth et al., 1988) whose resonant frequencies are spaced logarithmically from 100 to 1200 Hz (the dominant spectral region for processing interaural delays; Raatgever, 1980). The output of each filter is then *v*th-law halfwave rectified (Shear, 1987):

$$z(t) = x(t)^3 \quad \text{for } x > 0, 0 \qquad \text{otherwise}$$
(2)

to simulate peripheral nonlinearities and the resultant waveforms are processed through a running crosscorrelation function

$$\delta(t,\tau) = c(\tau) \int_{f_l}^{f_u} \int_{-\infty}^{T} q(f) x_L(f,t) x_R(f,t-\tau)$$

×w(t) dt df (3)

where x_L and x_R represent the left and right waveforms, respectively, f_l and f_u are the lower and upper frequency limits over which the cross-correlation patterns are integrated, τ is the argument of the cross-correlation function (i.e. lag), and $c(\tau)$ is a centrality weighting function (Shackleton et al., 1992; Saberi, 1996):

$$c(\tau) = \frac{1}{2^{-3}\sqrt{2\pi}} e^{-0.5(\tau/0.002)^2}.$$
 (4)

This function emphasizes cross-correlation lags near zero, consistent with psychophysical findings of a mild auditory foveation in spatial regions directly in front of listeners (Hafter and DeMaio, 1975; Saberi, 1991; Saberi and Petrosyan, 2004; Yost and Gourevitch, 1987). The function q(f) is a frequency-dependent weighting function (Stern et al., 1988; Zurek and Saberi, 2003) that reflects the dominant frequency region in sound localization by humans (Raatgever, 1980) and is of the form

$$q(f) = 10^{-(a_1 f + a_2 f^2 + a_3 f^3)/10},$$
(5)

where $a_1 = (-9.383)(10^{-2})$, $a_2 = (1.126)(10^{-4})$, and $a_3 = (-3.992)(10^{-8})$. This function is valid for frequencies below 1200 Hz, and has a peak near 600 Hz. The function

$$w(t) = e^{-(T-t)/0.01}$$
(6)

represents exponential sensory memory decay with a time constant of 10 ms, and heavily weights recent



Fig. 3. Output of a model of cross-correlation in response to partially interaurally correlated noise waveforms (see text). Each row represents model output in response to one value of interaural correlation, shown in the upper-left corner of left panels. Left panels display cross-correlation activity along a frequency-by-lag surface, with greater activity represented by lighter shades. The straight trajectory across frequencies represents the stimulus interaural delay (waveform group delay). Middle panels show model output integrated across frequency, and right panels show histograms of the lag corresponding to peak activity obtained from 5000 stimulus presentations in model simulations.

cross-correlation activity patterns, i.e. running or shortterm cross-correlation (Sayers and Cherry, 1957; Blauert and Cobben, 1978; Saberi, 1995, 1996; Saberi et al. 1998, 2002b, 2004; Zurek and Saberi, 2003).

The output of this model in response to noise waveforms with varying degrees of interaural correla-

tion, and an interaural delay of $250 \,\mu s$ are shown in the left panels of Fig. 3 prior to integration across frequency channels, and in the middle panels after frequency integration (right panels will be discussed shortly). Each row of panels corresponds to one value of interaural correlation, with the top row corresponding to a



Fig. 4. Spatial receptive fields from an optic tectum neuron as a function of interaural correlation. Note the decline in the dominant receptive field as correlation is reduced. Data are from Saberi et al. (1998).

correlation of 1.0 and the bottom row corresponding to correlation of 0.1. Left panels show cross-correlation activity as a function of lag (τ) and filter center frequency in hertz. The straight trajectory of peak activity across frequency occurs, as expected, at a lag of 250 µs, i.e. the waveform interaural delay, with curved secondary trajectories at other lags. As interaural correlation is reduced from unity to 0.5, peak trajectories become less stable, but still clearly defined. These trajectories become highly distorted when interaural correlation is 0.1.

The middle panels show cross-correlation patterns integrated across frequency channels, as occurs in neurons of higher nuclei of the auditory tract (Konishi, 1993b, 2003). A large peak occurs at the signal interaural delay of 250 µs for a correlation of 1.0, which becomes smaller at correlations of 0.5 and 0.3. At a correlation of 0.1, the pattern is highly variable (the bottom panels correspond to a single instance of signal presentation which was chosen to demonstrate the high variance in cross-correlation peaks at low interaural correlations). Consistent with these patterns, Fig. 4 shows typical receptive fields of an optic tectum neuron as a function of interaural correlation (from Saberi et al., 1998). The abscissa is interaural delay and the ordinate is the averaged firing rate in 100 ms. The tectum, which receives converging frequency information from lower nuclei, clearly shows patterns of neural responses to decorrelated signals in line with our psychophysical findings. Note that the neuron's dominant receptive field is clearly identifiable for and interaural correlation as low as 0.4.

The right panels of Fig. 3 show histograms of crosscorrelation peaks, $\tau(\max|\delta|)$, based on 5000 estimates per interaural correlation, i.e. the lag at which the cross correlation peak occurs. Note that the estimate of the lag at which cross-correlation peak occurs becomes more variable with increasing interaural decorrelation. Furthermore, the histogram mode shifts slightly toward zero microseconds as interaural correlation decreases. The quantization is due to the waveform sampling frequency of 44.1 kHz ($\sim 23 \,\mu s$). We should note here that for the purpose of visually showing how curved trajectories are arranged relative to the straight trajectory, the centrality weighting function (Eq. (4)) was not imposed on the model output for the left and middle panels of Fig. 3, but was included for estimating the variance and mean of peak estimates when generating the histograms of the right panels of Fig. 3.

6. Detection-theoretic predictions

We used the estimated mean and variance of crosscorrelation peaks to make prediction of psychophysical performance. Top panel of Fig. 5 shows the estimated means (open circles) and standard deviations (asterisks) for the four interaural correlation values of 0.1, 0.3, 0.5, and 1.0 obtained from the cross-correlation model (right panels of Fig. 3). The solid and dashed curves marked μ and σ are fitted modified exponential functions to these estimates. We used these fitted estimates in a detectiontheory model (Green and Swets, 1966; van Trees, 1968) to predict (middle panel of Fig. 5) an index of detectability from mean and variance of cross-correlation peaks and the standard deviation of internal neural noise σ_i^2 estimated from lateralization thresholds of approximately 10 µs for broadband noise stimuli (Klumpp and Eady, 1956):

$$d'_{r} = \frac{\sqrt{2E[\tau(\max|\delta_{r}|)]\beta}}{\sqrt{\sigma_{\tau(\max|\delta_{r}|)}^{2} + \sigma_{i}^{2}}},$$
(7)

where d_r' is the detection index in a 2IFC task corresponding to an interaural correlation of r, $E[\tau(\max|\delta_r|)]$ is the expected value of the estimated interaural delay corresponding to the peak of the

Fig. 5. Top panel shows means (circles) and variances (asterisks) for estimating the lag corresponding to the cross-correlation peak from histograms of Fig. 3. The curves are modified exponential fits to these estimates. Middle panel shows model predictions from Eq. (7) as a function of velocity and interaural correlation. Velocity in μ s/s is indicated next to each curve. Bottom panel shows subject performance as a function of velocity and interaural correlation. Data are averaged from 3 subjects (from Fig. 1).



cross-correlation function at an interaural correlation of r and is derived from simulations, $\sigma_{\tau(\max|\delta_r|)}$ is the corresponding standard deviation, and β scales the peak

interaural-delay excursion such that for motion stimuli, as suggested by psychophysical findings (Perrott and Musicant, 1977a; Grantham and Wightman, 1978; Saberi and Hafter, 1997), perceived spatial excursion is diminished (i.e. a lowpass filter). The functional form of β is an exponential decay that is a declining function of velocity (v), specifically, having the form:

$$\beta = 0.03 + 0.42 \mathrm{e}^{-\nu/850} \tag{8}$$

which best fits the observed data. The variance, and thus slope, of this function is dependent on the velocity constant (850 μ s/s) and is inversely related to the rate at which motion detection declines as a function of velocity. A smaller constant would result in a more significant effect of velocity on motion-detection thresholds. The constant $850\,\mu s/s$ is also in line with a break point reported in a study by Saberi et al. (2003) on detection of rate of change in interaural delays for low-frequency pure tones (500 Hz) for a wide range of velocities $(\sim 3-32\,000\,\mu s/s)$. They showed that for a fixed "distance" of 400 μ s, d' is nearly constant up to a velocity of 800 μ s/s, and then deteriorates with increasing velocity. Furthermore, an exponential form is suggested by the data of Perrott and Musicant (1977a, b) who measured perceived distance traveled by a moving sound source in the freefield. Subjects marked the perceived start and end points of a moving sound source as a function of velocity, from which perceived distance was estimated. Excluding short distances of less than 10° , which produced highly variable estimates, their data shows a negative exponential decrease in perceived distance as velocity is increased from 90 to 600° /s (these values roughly correspond to $650-4300 \,\mu\text{s/s}$ which are in the general range of velocities tested in the current study).²

Eq. (7) is derived from assumptions of detection theory (Green and Swets, 1966; Macmillan and Creelman, 1991; van Trees, 1968) that the observer calculates the distance in two equal-variance distributions in z units. This distance is the difference in the means of the distributions divided by the square-root of the sum of their variances, i.e. $d' = (\mu - (-\mu))/\sqrt{\sigma^2 + \sigma^2} = \sqrt{2\mu}/\sigma$ where σ is itself the sum of the variances of internal noise and the noise introduced by interaural decorrelation as shown in Eq. (7). In the 2IFC task, these two distributions correspond to two signal intervals, and in the motion case the distributions are based on the assumption that the observer compares the beginning and end-points of motion (Perrott and Marlborough, 1989), i.e. perceived peak excursion of the auditory image.

²Note that when v = 0 the value of β is less than unity. This suggests that our estimate of internal noise σ_i derived from the literature for lateralization of stationary broadband noise may be lower than that for our subjects by a constant of proportionality. When we allowed σ_i to freely vary, we obtained a near-perfect fit for v = 0 with β excluded (i.e. set to unity) but at the cost of an additional free parameter which we opted against.

Model predictions are shown in the middle panel of Fig. 5. Empirically measured averaged d's are shown in the bottom panel of Fig. 5 for comparison. Predicted d's were forced to an asymptotic ceiling of 3.29 (2%) inattention rate; see Results) and lowpass filtered to eliminate singularities where it reached asymptotic point at d' = 3.29 (i.e. a smoothing procedure). The model makes good quantitative predictions of the average performance. For low velocities, predicted d's decline only very slightly as r is reduced from 1 to 0.5, and precipitously from r = 0.5 to 0.1, consistent with the data. For high velocities, the model predicts less than asymptotic performance at all correlations, and a monotonic decline in performance as correlation is reduced from unity. Excluding d' values that are at or near asymptotic value, the proportion of variance accounted for by the fit between the 17 non-asymptotic measured and predicted d's is $r^2 = 0.96$. The outlier fit at $v = 5000 \,\mu\text{s/s}$ and interaural correlation of 0.7 was not statistically significantly different than measured d' given intersubject variability at that point, t(2) = 1.85, n.s.

Among the predicted features of the curves shown in Fig. 5 are: (1) near-perfect asymptotic performance for interaural correlations of 0.7 and 1.0 at low velocities, (2) lower than asymptotic performance at all correlations at high velocities, (3) at a given correlation, monotonic decline in d' as a function of velocity, and (4) as expected, convergence of all curves as interaural correlation approaches zero.

7. Conclusion

Cross-correlation of neural events arriving from the two auditory channels occurs early in the ascending auditory pathway, at the level of the medial superior olivary complex in mammals. The findings described here suggest that one may explain motion thresholds using detection theory coupled with a simple crosscorrelation model without invoking high-order cortical mechanisms. Use of decorrelated signals to examine localization and motion detection revealed parallels in how these two cues are processed. Motion stimuli produced progressively lower performance as a function of increasing velocity consistent with the idea of binaural sluggishness, i.e. a lowpass filtering of dynamic interaural information (Grantham and Wightman, 1978; Saberi and Hafter, 1997). These psychophysical results suggest that by incorporating a simple lowpass filter, motion detection as a function of velocity and interaural correlation may be explained with the same mechanisms that underlie the processing of stationary sounds. Human neuroimaging research, on the other hand, has been equivocal on whether there are unique centers dedicated to auditory motion. For example, some have identified the parietal lobe and planum temporale



Fig. 6. Cortical activation patterns from three subjects (columns) in response to a broadband stationary stimulus (top row), a motion stimulus (middle), and the contrast between the two (bottom). The position of the stationary stimulus was perturbed on each presentation. Data are from Smith et al. (2004).

associated with motion. In our recent neuroimaging work, we contrasted cortical activity in response to motion versus stationary stimuli whose position was perturbed on each presentation (Smith et al., 2004). Fig. 6, which shows a sample set of these data, indicates that a unique cortical motion area could not be identified when stationary cues are controlled for. Our current psychophysical findings, as well as our results from neuroimaging, however, do not necessarily prove that higher-order motion networks do not exist. For example, a network that processes auditory motion may be superimposed on or interleaved with one that processes stationary sounds, and thus occupy nearly identical cortical loci and consequently remain unidentified when contrasting motion to randomly positioned stationary sounds. Furthermore, from a psychophysical standpoint, the predictability of motion thresholds from those of stationary conditions would not necessarily mean that once motion information reaches the cortex, it is not used by other systems, for example, in conjunction with somatosensory and proprioceptive mechanisms in maintaining stable spatial coordinates during head rotation (Perrott, 1989; Perrott et al., 1987, 1990; Lewald and Karnath, 2001).

References

- Ahissar, M., Ahissar, E., Bergman, H., Vaadia, E., 1992. Encoding of sound-source location and movement: activity of single neurons and interactions between adjacent neurons in the monkey auditory cortex. J. Neurophysiol. 67, 203–215.
- Albeck, Y., Konishi, M., 1995. Responses of neurons in the auditory pathway of the barn owl to partially correlated binaural signals. J. Neurophysiol. 74, 1689–1700.

- Altman, J.A., Viskov, D.V., 1977. Discrimination of perceived movement velocity for fused auditory image in dichotic stimulation. J. Acoust. Soc. Am. 61, 816–819.
- Blauert, J., 1972. On the lag of lateralization caused by interaural time and intensity differences. Audiology 11, 265–270.
- Blauert, J., Cobben, W., 1978. Some consideration of binaural crosscorrelation analysis. Acustica 39, 96–104.
- Budd, T.W., Hall, D.A., Goncalves, M.S., Akeroyd, M.A., Foster, J.R., Palmer, A.R., Head, K., Summerfield, A.Q., 2003. Binaural specialisation in human auditory cortex: an fMRI investigation of interaural correlation sensitivity. Neuroimage 20, 1783–1794.
- Butler, R.A., 1975. Influence of the external and middle ear on auditory discrimination. In: Keidel, W.D., Neff, W.D. (Eds.), Handbook of Sensory Physiology. vol. V/2: Auditory System. Springer, Berlin, pp. 247–260.
- Butler, R.A., 1977. Spectral cues utilized in the localization of sound in the median sagittal plane. J. Acoust. Soc. Am. 61, 1264–1269.
- Carr, C.E., Konishi, M., 1988. Axonal delay-lines for time measurement in the owl's brain-stem. Proc. Natl Acad. Sci. USA 21, 8311–8315.
- Carr, C.E., Konishi, M., 1990. A circuit for detection of interaural time differences in the brain stem of the barn owl. J. Neurosci. 10, 3227–3246.
- Chandler, D.W., Grantham, D.W., 1992. Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth, and velocity. J. Acoust. Soc. Am. 91, 1624–1636.
- Clifford, C.W.G., Ibbotson, M.R., 2002. Fundamental mechanisms of visual motion detection: models, cells and functions. Prog. Neurobiol. 68, 409–437.
- Dove, H.W., 1839. Nachtrag zu den Combinationstonen. In: Repertorium der Physik, pp. 404–405.
- Egnor, S.E.R., 2001. Effects of binaural decorrelation on neural and behavioral processing of interaural level differences in the barn owl (*Tyto alba*). J. Comp. Physiol. A 187, 589–595.
- Grantham, D.W., 1984. Discrimination and dynamic interaural intensity differences. J. Acoust. Soc. Am. 76, 71–76.
- Grantham, D.W., 1986. Detection and discrimination of simulated motion of auditory targets in the horizontal plane. J. Acoust. Soc. Am. 79, 1939–1949.
- Grantham, D.W., Wightman, F.L., 1978. Detectability of varying interaural temporal differences. J. Acoust. Soc. Am. 63, 511–523.
- Green, D.M., Swets, J.A., 1966. Signal Detection Theory and Psychophysics. Wiley, New York.
- Hafter, E.R., DeMaio, J., 1975. Difference thresholds for interaural delay. J. Acoust. Soc. Am. 57, 181–187.
- Hartmann, W.M., Constan, Z.A., 2002. Interaural level differences and the level-meter model. J. Acoust. Soc. Am. 112, 1037–1045.
- Holdsworth, J., Nimmo-Smith, I., Patterson, R., Rice, P., 1988. Implementing a Gammatone filterband. IN Annex C of the SVOS Final Report (Part A: The auditory Filter Bank), MRC (Medical Research Council), APU (Applied Psychology Unit) Report 2341, University of Cambridge, Cambridge, United Kingdom.
- Jeffress, L.A., Robinson, D.E., 1962. Formulas for coefficient of interaural correlation for noise. Journal of the Acoustical Society of America 34, 1658–1659.
- Jeffress, L.A., Deatherage, B.H., Blodgett, H.C., 1962. Effect of interaural correlation on precision of centering a noise. J. Acoust. Soc. Am. 34, 1122–1123.
- Klumpp, R.G., Eady, H.R., 1956. Some measurements of interaural time difference thresholds. J. Acoust. Soc. Am. 28, 859–860.
- Konishi, M., 1992. The neural algorithm for sound localization in the owl. Harvey Lect. 86, 47–64.
- Konishi, M., 1993a. Listening with 2 ears. Sci. Am. 268, 66–73.
- Konishi, M., 1993b. Neuroethology of sound localization in the owl. J. Comp. Physiol. A 173, 3–7.

- Konishi, M., 2000. Study of sound localization by owls and its relevance to humans. Comp. Biochem. Phys. A 126, 459–469.
- Konishi, M., 2003. Coding of auditory space. Annu. Rev. Neurosci. 26, 31–55.
- Kuhn, G.F., 1987. Physical acoustics and measurements pertaining to directional hearing. In: Yost, W.A., Gourevitch, G. (Eds.), Directional Hearing. Springer, New York, pp. 3–25.
- Lewald, J., Karnath, H.O., 2001. Sound lateralization during passive whole-body rotation. Eur. J. Neurosci. 13, 2268–2272.
- Lindemann, W., 1986. Extension of a binaural cross-correlation model by contralateral inhibition: I. Simulation of lateralization for stationary signals. J. Acoust. Soc. Am. 80, 1608–1622.
- Lu, Z.-L., Sperling, G., 1995. The functional architecture of human visual motion perception. Vision Res. 35, 2697–2722.
- Mach, E., 1874. Functions of the pinnae of the ears. Archiv fur Ohrenheilkunde. Cited in S.L. Thompson (1981). Phenomena of binaural audition. Philos. Mag. 12 (5), 351–355.
- Macmillan, N.A., Creelman, C.D., 1991. Detection Theory: A User's Guide. Cambridge University Press, Cambridge.
- Moiseff, A., Haresign, T., 1992. Response of auditory units in the barn owl's inferior colliculus to continuously varying interaural phase differences. J. Neurophysiol. 67, 1428–1436.
- Pavani, F., Macaluso, E., Warren, J.D., Driver, J., Griffiths, T.D., 2002. A common cortical substrate activated by horizontal and vertical sound movement in the human brain. Curr. Biol. 12, 1584–1590.
- Perrott, D.R., 1989. Are there motion detectors in the auditory system? Sound Localization by human observers: Symposium Proceedings. National Academy of Sciences, p. 36.
- Perrott, D.R., Marlborough, K., 1989. Minimum audible movement angle: marking the end points of the path traveled by a moving sound source. J. Acoust. Soc. Am. 85, 1773–1775.
- Perrott, D.R., Musicant, A.D., 1977a. Minimum auditory movement angle: binaural localization of moving sound sources. J. Acoust. Soc. Am. 62, 1463–1466.
- Perrott, D.R., Musicant, A.D., 1977b. Rotating tones and binaural beats. J. Acoust. Soc. Am. 61, 1288–1292.
- Perrott, D.R., Nelson, M.A., 1969. Limits for the detection of binaural beats. J. Acoust. Soc. Am. 46, 1477–1481.
- Perrott, D.R., Tucker, J., 1988. Minimum audible movement angle as a function of signal frequency and the velocity of the source. J. Acoust. Soc. Am. 83, 1522–1527.
- Perrott, D.R., Ambarsoom, H., Tucker, J., 1987. Changes in head position as a measure of auditory localization performance: auditory psychomotor coordination under monaural and binaural listening conditions. J. Acoust. Soc. Am. 82, 1637–1645.
- Perrott, D.R., Saberi, K., Brown, K., Strybel, T.Z., 1990. Auditory psychomotor coordination and visual search performance. Percept. Psychophys. 48, 214–226.
- Perrott, D.R., Costantino, B., Ball, J., 1993. Discrimination of moving events which accelerate or decelerate over the listening interval. J. Acoust. Soc. Am. 93, 1053–1057.
- Peterson, J., 1916. The nature and probable origin of binaural beats. Psychol. Rev. 23, 333–351.
- Raatgever, J., 1980. On the binaural processing of stimuli with different interaural phase relations. Doctoral Dissertation, technische Hogeschool Delft, The Netherlands.
- Rabiner, L.R., Gold, B., 1975. Theory and Application of Digital Signal Processing. Prentice-Hall, Englewood Cliffs, NJ.
- Rayleigh, L., 1876. On our perception of the direction of a source of sound. Nature 14, 32–33.
- Rayleigh, L., 1877. On our perception of the direction of a source of sound. Philos. Mag. 5 (3), 546.
- Rayleigh, L., 1907. On our perception of the direction of a source of sound. Philos. Mag. 8, 214–232.

- Robinson, D.E., Jeffress, L.A., 1963. Effect of varying the interaural noise correlation on the detectability of tonal signals. J. Acoust. Soc. Am. 35, 1947–1952.
- Saberi, K., 1991. Variations in the speed of sound and interaural differences of time: an additional source of jitter. J. Acoust. Soc. Am. 90, 1677.
- Saberi, K., 1995. Lateralization of comodulated complex waveforms. J. Acoust. Soc. Am. 98, 3146–3156.
- Saberi, K., 1996. An auditory illusion predicted from a weighted crosscorrelation model of binaural interaction. Psychol. Rev. 103, 137–142.
- Saberi, K., 2004. Fast Fourier-based DSP algorithm for auditory motion experiments. Behav. Res. Methods Instrum. Comput. 36, 585–589.
- Saberi, K., Green, D.M., 1997. Evaluation of maximum-likelihood estimators for non-intensive auditory psychophysics. Percept. Psychophys. 59, 867–876.
- Saberi, K., Hafter, E.R., 1997. Experiments on auditory motion discrimination. In: Gilkey, R.H., Anderson, T.R. (Eds.), Binaural and Spatial Hearing in Real and Virtual Environments. Erlbaum, NJ, pp. 315–327.
- Saberi, K., Perrott, D.R., 1990. Minimum audible movement angles as a function of sound-source trajectory. J. Acoust. Soc. Am. 88, 2639–2644.
- Saberi, K., Petrosyan, A., 2004. A detection-theoretic model of echo inhibition. Psychol. Rev. 111, 52–66.
- Saberi, K., Dostal, L., Sadralodabai, T., Perrott, D.R., 1991. Minimum audible angles for horizontal, vertical, and oblique orientations: lateral and dorsal planes. Acustica 75, 57–61.
- Saberi, K., Takahashi, Y., Konishi, M., Albeck, Y., Arthur, B., Farahbod, H., 1998. Effects of interaural decorrelation on neural and behavioral detection of spatial cues. Neuron 21, 789–798.
- Saberi, K., Tirtabudi, P., Petrosyan, A., Perrott, D.R., Strybel, T.Z., 2002a. Concurrent motion detection based on dynamic changes in interaural delay. Hearing Res. 174, 149–157.
- Saberi, K., Takahashi, Y., Egnor, R., Farahbod, H., Mazer, J., Konishi, M., 2002b. Detection of large interaural delays and its implication for models of binaural interaction. J. Assoc. Res. Ottolaryngol. 3, 80–88.
- Saberi, K., Tirtabudi, P., Petrosyan, A., Perrott, D.R., Strybel, T.Z., 2003. Detection of dynamic changes in interaural delay. Acta Acust. Acust. 89, 333–338.
- Saberi, K., Antonio, J.V., Petrosyan, A., 2004. A population study of the precedence effect. Hearing Res. 191, 1–13.
- Sayers, B. McA., Cherry, E.C., 1957. Mechanism of binaural fusion in hearing of speech. J. Acoust. Soc. Am. 29, 973–987.
- Shackleton, T.M., Meddis, R., Hewitt, M.J., 1992. Across frequency integration in a model of lateralization. J. Acoust. Soc. Am. 91, 2276–2279.
- Shaw, E.A.G., 1965. Earcanal pressure generated by a free sound field. J. Acoust. Soc. Am. 22, 465–470.
- Shaw, E.A.G., 1974. Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. J. Acoust. Soc. Am. 56, 1848–1861.

- Shear, G.D., 1987. Modeling the dependence of auditory lateralization on frequency and bandwidth, Master Thesis, Carnegie Mellon University.
- Smith, K.R., Okada, K., Saberi, K., Hickok, G., 2004. Human cortical auditory motion areas are not motion selective. Neuroreport 15, 1523–1526.
- Spitzer, M.W., Semple, M.N., 1991. Interaural phase coding in auditory midbrain: Influence of dynamic stimulus features. Science 254, 721–723.
- Stern, R.M., Zeiberg, A.S., Trahiotis, C., 1988. Lateralization of complex binaural stimuli: a weighted-image model. J. Acoust. Soc. Am. 84, 156–165.
- Strybel, T.Z., Menges, M.L., 1998. Auditory apparent motion between sine waves differing in frequency. Perception 27, 483–495.
- Strybel, T.Z., Neale, W., 1994. The effect of burst duration, interstimulus onset interval, and loudspeaker arrangement on auditory apparent motion in the free-field. J. Acoust. Soc. Am. 3463–3475.
- Strybel, T.Z., Manligas, C., Perrott, D.R., 1989. Auditory apparent motion under binaural and monaural listening conditions. Percept. Psychophys. 45, 371–377.
- Strybel, T.Z., Span, S.A., Witty, A.M., 1998. The effect of timing and spatial separation on the velocity of auditory apparent motion. Percept. Psychophys. 60, 1441–1451.
- Stumpf, E., Toronchuk, J.M., Cynader, M.S., 1992. Neurons in cat primary auditory cortex sensitive to correlates of auditory motion in three-dimensional space. Exp. Brain Res. 88, 158–168.
- Thompson, S.P., 1877. On binaural audition. Philos. Mag. 5 (4), 274–276.
- Thompson, S.P., 1878. Phenomena of binaural audition. Philos. Mag. 5 (6), 383–390.
- van Trees, H.L., 1968. Detection, Estimation, and Modulation Theory, Part I. Wiley, New York.
- Vaina, L.M., Soloviev, S., 2003. First-order and second-order motion: neurological evidence for neuroanatomically distinct systems. Prog. Brain Res. 144, 197–212.
- Valentine, W.L., 1928. Notes on the binaural beat. Comp. Psychol. 7, 357–368.
- Warren, J.D., Zielinski, B.A., Green, G.G.R., Rauschecker, J.P., Griffiths, T.D., 2002. Perception of sound-source motion by the human brain. Neuron 34, 139–148.
- Watson, A.B., Pelli, D.G., 1983. QUEST: a Bayesian adaptive psychometric method. Percept. Psychophys. 33, 113–120.
- Waugh, W., Strybel, T.Z., Perrott, D.R., 1979. Perception of moving sounds: velocity discrimination. J. Aud. Res. 19, 103–110.
- Yin, T.C.T., Chan, J.C.K., 1990. Interaural time sensitivity in medial superior olive of cat. J. Neurophysiol. 64, 465–488.
- Yost, W.A., 1997. The cocktail party effect: 40 years later. In: Gilkey, R., Anderson, T. (Eds.), Binaural and Spatial Hearing in Real and Virtual Environments. Erlbaum Press, NJ, pp. 329–348.
- Yost, W.A., Gourevitch, G., 1987. Directional Hearing. Springer, New York.
- Yost, W.A., Dye, R.H., Sheft, S.A., 1996. Simulated "Cocktail Party" with up to three sound sources. Percept. Psychophys. 58, 1026–1036.
- Zurek, P., Saberi, K., 2003. Lateralization of two-transient stimuli. Percept. Psychophys. 65, 95–106.