ELSEVIER

Available online at www.sciencedirect.com
SCIENCE DIRECT®

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.

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 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.: +1 949 824 6310; fax: +1 949 824 2307.
E-mail address: saberi@uci.edu (K. Saberi).

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