Hearing Research 307 (2014) 86–97

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Review

# Psychophysics and neuronal bases of sound localization in humans

Jyrki Ahveninen [a,*], Norbert Kopčo [a,b,c], Iiro P. Jääskeläinen [d]

[a] Harvard Medical School – Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA, USA

[b] Institute of Computer Science, P.J. Šafárik University, Košice, Slovakia

[c] Hearing Research Center, Boston University, Boston, MA, USA

[d] Brain and Mind Laboratory, Department of Biomedical Engineering and Computational Science (BECS), Aalto University, Espoo, Finland

ARTICLE INFO

*Article history:*
Received 15 May 2013
Received in revised form
2 July 2013
Accepted 10 July 2013
Available online 22 July 2013

ABSTRACT

Localization of sound sources is a considerable computational challenge for the human brain. Whereas the visual system can process basic spatial information in parallel, the auditory system lacks a straightforward correspondence between external spatial locations and sensory receptive fields. Consequently, the question how different acoustic features supporting spatial hearing are represented in the central nervous system is still open. Functional neuroimaging studies in humans have provided evidence for a posterior auditory "where" pathway that encompasses non-primary auditory cortex areas, including the planum temporale (PT) and posterior superior temporal gyrus (STG), which are strongly activated by horizontal sound direction changes, distance changes, and movement. However, these areas are also activated by a wide variety of other stimulus features, posing a challenge for the interpretation that the underlying areas are purely spatial. This review discusses behavioral and neuroimaging studies on sound localization, and some of the competing models of representation of auditory space in humans.

*This article is part of a Special Issue entitled <Human Auditory Neuroimaging>.*



## 1. Introduction

Determining the location of perceptual objects in extrapersonal space is essential in many everyday situations. For objects outside the field of vision, hearing is the only sense that provides such information. Thus, spatial hearing is a fundamental prerequisite for our efficient functioning in complex communication environments. For example, consider a person reaching for a ringing phone or a listener using audiospatial information to help focus on one talker in a chattering crowd (Brungart and Simpson, 2002; Gilkey and Anderson, 1997; Middlebrooks and Green, 1991; Shinn-Cunningham et al., 2001). Spatial hearing has two main functions: it enables the listener to localize sound sources and to separate sounds based on their spatial locations (Blauert, 1997). While the spatial resolution is higher in vision (Adler, 1959; Recanzone, 2009; Recanzone et al., 1998), the auditory modality allows us to monitor objects located anywhere around us. The ability to separate sounds based on their location makes spatial auditory processing an important factor in auditory scene analysis (Bregman, 1990), a process of creating individual auditory objects, or streams, and separating from background noise (Moore, 1997). Auditory localization mechanisms can be different in humans compared to other species utilized in animal neurophysiological studies. For example, in contrast to cats, we cannot move our ears toward the sound sources. Further, unlike in barn owls, our ears are at symmetrical locations on our heads, and sound elevation needs to be determined based on pinna-related spectral cues, which is less accurate than comparing the sounds received at the two asymmetric ears (Knudsen, 1979; Rakerd et al., 1999). In the following, we will review key findings that have elucidated the psychophysics and neural basis of audiospatial processing in humans.

## 2. Psychophysics of auditory spatial perception

### 2.1. Sound localization cues in different spatial dimensions

Spatial hearing is based on "binaural" and "monaural" cues (Yost and Gourevitch, 1987). The two main binaural cues are differences

*Abbreviations:* 3D, three dimensional; AC, auditory cortex; AM, amplitude modulation; BAEP, brainstem auditory evoked potentials; BRIR, binaural room impulse response; DRR, direct-to-reverberant ratio; DTI, diffusion-tensor imaging; EEG, electroencephalography; ERP, event-related potential; FM, frequency modulation; fMRI, functional magnetic resonance imaging; HG, Heschl's gyrus; HRTF, head-related transfer function; ILD, interaural level difference; IPD, interaural phase difference; ITD, interaural time difference; MAEP, middle-latency auditory evoked potential; MEG, magnetoencephalography; MRI, magnetic resonance imaging; PET, positron emission tomography; PT, planum temporale; PP, planum polare; SSR, steady-state response; STG, superior temporal gyrus

* Corresponding author. MGH/MIT/HMS-Martinos Center, Bldg. 149, 13th Street, Charlestown, MA 02129, USA. Tel.: +1 617 726 6584; fax: +1 617 726 7422.

*E-mail address:* jyrki@nmr.mgh.harvard.edu (J. Ahveninen).

0378-5955/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.heares.2013.07.008

in the time of arrival (the interaural time difference, ITD, or interaural phase difference, IPD) and differences in the received intensity (the interaural level difference, ILD) (Middlebrooks and Green, 1991). The most important monaural localization cue is the change in the magnitude spectrum of the sound caused by the interaction of the sound with the head, body, and pinna before entering the ear (Blauert, 1997; Macpherson and Sabin, 2007; Middlebrooks and Green, 1991; Shaw, 1966; Wightman and Kistler, 1989). Another monaural cue is the direct-to-reverberant energy ratio (DRR), which expresses the amount of sound energy that reaches our ears directly from the source *vs*. the amount that is reflected off the walls in enclosed spaces (Larsen et al., 2008). In general, monaural cues are more ambiguous than binaural cues because the auditory system must make *a priori* assumptions about the acoustic features of the original sound in order to estimate the filtering effects corresponding to the monaural spatial cues.

Positions of objects in three dimensional (3D) space are usually described using either Cartesian ($x$, $y$, $z$) or spherical (azimuth, elevation, distance) coordinates. For studies of spatial hearing, the most natural coordinate system uses bipolar spherical coordinates (similar to the coordinate system used to describe a position on the globe) with the two poles at the two ears and the origin at the middle point between the ears (Duda, 1997). In this coordinate system the azimuth (or horizontal location) of an object is defined by the angle between the source and the interaural axis, the elevation (or vertical location) is defined as the angle around the interaural axis, and distance is measured from the center of the listener's head. Using this coordinate system is natural when discussing spatial hearing because different auditory localization cues map onto these coordinate dimensions in a natural, monotonic manner. However, note that, if the examination is restricted to certain sub-regions of space, the Cartesian and spherical representations can be equivalent. For example, for sources directly ahead of the listener the two representations are very similar.

Binaural cues are the primary cues for perception in the azimuthal dimension. The perceived azimuth of low-frequency sounds (below 1–2 kHz) is dominated by the ITD. For high-frequency stimuli (above 1–2 kHz), the auditory system weights the ILD more when determining the azimuth. This simple dichotomy (ITD for low frequencies, ILD for high frequencies) is referred to as the duplex theory (Strutt, 1907). It can be explained by considering physical and physiological aspects of how these cues change with azimuth and how they are neuronally extracted. However, there are limitations to the applicability of this theory. For example, for nearby sources, the ILD is available even at low frequencies (Shinn-Cunningham, 2000). Similarly, the ITD cue in the envelope of the stimulus, as opposed to the ITD in the fine structure, can be extracted by the auditory system even from high-frequency sounds (van de Par and Kohlrausch, 1997). Finally, in theory, the azimuth of a sound source can be determined also monaurally, because the high-frequency components of the sound are attenuated more compared to low-frequency components as the sound source moves contralaterally (Shub et al., 2008).

The main cue the human auditory system uses to determine the elevation of a sound source is the monaural spectrum determined by the interaction of the sound with the pinnae (Wightman and Kistler, 1997). Specifically, there is a spectral notch that moves in frequency from approximately 5 to 10 kHz as the source moves from 0° (directly ahead of listener) to 90° (above the listener's head), considered to be the main elevation cue (Musicant and Butler, 1985). However, small head asymmetries may provide a weak binaural elevation cue.

The least understood dimension is distance (Zahorik et al., 2005). The basic monaural distance cue is the overall received sound level (Warren, 1999). However, to extract this cue, the listener needs *a priori* knowledge of the emitted level, which can be difficult since the level at which sounds are produced often varies. For nearby, lateral sources the ILD changes with source distance and provides a distance cue (Shinn-Cunningham, 2000). In reverberant rooms, the auditory system uses some aspect of reverberation to determine the source distance (Bronkhorst and Houtgast, 1999). This cue is assumed to be related to DRR (Kopco and Shinn-Cunningham, 2011). Finally, other factors like vocal effort for speech (Brungart and Scott, 2001) can also be used.

To determine which acoustic localization cues are available in the sound produced by a target at a given location in a given environment, head-related transfer functions (HRTFs) and binaural room impulse responses (BRIRs) can be measured and analyzed (Shinn-Cunningham et al., 2005). These functions/responses provide complete acoustic characterization of the spatial information available to the listener for a given target and environment, they vary slightly from listener to listener, and they also can be used to simulate the target in a virtual acoustic environment.

### *2.2. Natural environments: localization in rooms and complex scenes*

While the basic cues and mechanisms of spatial hearing in simple scenarios are well understood, much less is known about natural environments, in which multiple acoustic objects are present in the scene and where room reverberation distorts the cues. When the listener is in a room or other reverberant environment the direct sound received at the ears is combined with multiple copies of the sound reflected off the walls before arriving at the ears. Reverberation alters the monaural spectrum of the sound as well as the ILDs and IPDs of the signals reaching the listener (Shinn-Cunningham et al., 2005). These effects depend on the source position relative to the listener as well as on the listener position in the room. On the other hand, reverberation itself can provide a spatial cue (DRR). Most studies of sound localization were performed in an anechoic chamber (Brungart, 1999; Hofman and Van Opstal, 1998; Makous and Middlebrooks, 1990; Wenzel et al., 1993; Wightman and Kistler, 1997). There are also several early studies of localization in reverberant environments (Hartmann, 1983; Rakerd and Hartmann, 1985, 1986). They show that reverberation causes small degradations in directional localization accuracy. However, performance can improve with practice (Irving and Moore, 2011; Shinn-Cunningham, 2000). In addition, several recent studies have measured the perceived source distance (Bronkhorst and Houtgast, 1999; Kolarik et al., 2013; Kopco and Shinn-Cunningham, 2011; Ronsse and Wang, 2012; Zahorik et al., 2005). These studies show that in reverberant space, distance perception is more accurate, due to additional information provided by the DRR cue.

Processing of spectral and binaural spatial cues becomes particularly critical when multiple sources are presented concurrently. Under such conditions, the auditory system's ability to correctly process spatial information about the auditory scene becomes critical both for speech processing (Best et al., 2008; Brungart and Simpson, 2007) and target localization (Drullman and Bronkhorst, 2000; Hawley et al., 1999; Simpson et al., 2006). However, the strategies and cues the listeners use in complex environments are not well understood. It is clear that factors like the ability to direct selective spatial attention (Sach et al., 2000; Shinn-Cunningham, 2008; Spence and Driver, 1994) or the ability to take advantage of *a priori* information about the distribution of targets in the scene (Kopco et al., 2010) can improve performance. On the other hand, the localization accuracy can be adversely influenced by preceding stimuli (Kashino and Nishida, 1998) or even by the