



19th INTERNATIONAL CONGRESS ON ACOUSTICS
MADRID, 2-7 SEPTEMBER 2007

THE DUPLEX-THEORY OF LOCALIZATION INVESTIGATED UNDER NATURAL CONDITIONS

PACS: 43.66.Pn

Seeber, Bernhard U.

Auditory Perception Lab, Dept. of Psychology, University of California at Berkeley, Berkeley, CA
Now: MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, UK,
seeber@ihr.mrc.ac.uk

ABSTRACT

The duplex theory postulates that low-frequency tones are localized on the basis of interaural time differences (ITDs) while tones above 1.5 kHz are localized by evaluating interaural level differences (ILDs) [1]. Recent research showed that ITDs are also dominant for wide-band sounds [2]. The contribution of envelope-ITDs is supposedly small, but enhanced by temporal modulation [3]. Using tones, previous studies investigated the sensitivity to isolated binaural cues or the dependence of image location on traded cues [4]. The current study investigated lateralization for conflicting, but natural ILDs and ITDs by modifying head-related transfer functions. Preliminary results of 3 subjects showed clear ITD-dominance for almost all tested natural and synthetic sounds. Surprisingly, ITDs dominated for flute sounds while for a 2 kHz-high-pass noise both cues received almost equal weights. The highest relative ILD-weighting was seen when ILDs and ITDs were out of their natural combination, even before two separate images were heard. In case of image splits one image was mostly ILD, the other ITD dominated. Images splits of up to 60% across all tested conditions were reported for a short melody played on a flute while few splits were heard for a single flute tone suggesting that temporal modulation increases the tendency for grouping into two objects. Binaural cue weighting showed large variability between subjects.

INTRODUCTION

The auditory system analyses sounds for their spectral content and temporal information and groups components also with respect to binaural information. In natural situations, binaural cues of interaural time and intensity are present by the relationships provided in head-related transfer functions (HRTFs). The plausibility hypothesis suggests that the auditory system uses the cue that provides a maximum of reliable information [5]. The weighting of this information will be dependent on the stimulus and it might change if HRTFs are used or if lateralization studies are done with unnatural cue combinations. The purpose of the present study was to investigate binaural cue weighting with near-natural binaural cues by using individually-adapted HRTFs. HRTFs were manipulated so that ITDs stemmed from one direction while ILDs and spectral cues were in accord with another direction. Lateralization was investigated and subjective reports of hearing one or two images were recorded. While previous studies used a small set of synthetic stimuli the focus of the present study was to expand the view to a wide array of natural and synthetic stimuli. Specifically, we were interested in how binaural cue weighting changes for various degrees of discordance of binaural cues. The manuscript presents results from a preliminary evaluation of 3 out of 5 subjects.

METHODS

Virtual acoustics

Relative weighting of binaural cues was studied by varying ITDs and ILDs in virtual acoustical space. In a pre-test subjects selected a set of HRTFs from a catalogue of non-individual HRTFs. The selection procedure identifies a set of HRTFs for each subject that improves performance with respect to localization error, variance, front-back-confusions, and inside-the-head

localization [6]. By means of the Fast-Fourier-Transformation (FFT) HRTFs from different originating directions are split into their amplitude and phase components separately for each ear. Amplitude components from one direction are then recombined with phase components from another direction, so that ILDs come from, e.g., +30° while ITDs might stem from -60° [7]. A Gaussian window with 4.4dB damping at the sides was applied to the impulse response of the recombined HRTF to reduce sharp windowing effects introduced by larger phase shifts. Sound stimuli were convolved with the resulting head-related impulse responses (HRIRs) to yield a virtual acoustics stimulus. Stimuli were presented through a calibrated, diffuse-field equalized Sennheiser HD580 headphone.

Lateralization method

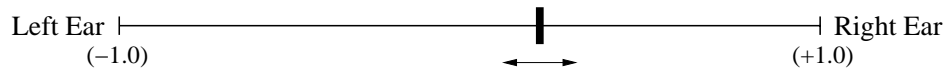


Figure 1: Line dissection method: The line was projected on a screen in 2m distance in front of the subject. The bar could be adjusted with a trackball to the position of the lateralized sound image.

The selection procedure individually adapted HRTFs to present natural values of binaural cues. This resulted in externalization of most stimuli presented with unprocessed HRTFs which contained congruent binaural cues. Most conditions, however, involved stimuli with discordant ITDs and ILDs which were perceived within the head. We decided to use a line-dissection method to assess lateralization of the sound image within the head. If stimuli were perceived outside the head subjects were instructed to project their location onto the interaural axis. The line-dissection method is pictured in Figure 1. The white line was projected on an otherwise black screen in 2m distance in front of the subject. It covered a visual angle of approx. $\pm 25^\circ$ and the endpoints were marked with vertical bars and the words “left ear” or “right ear”. The subject adjusted a red bar to the perceived lateralized sound position with a trackball. For the data analysis in this manuscript the left ear position is assigned -1 and the right ear +1.

Stimuli

Eight different stimuli that should give rise to different weightings of binaural cues were used in the experiments. Focus was set to include natural stimuli with synthetic counterparts. Table 1 gives an overview of stimulus type, bandwidth, envelope structure, duration, level as well as the predicted provision of localization cues according to the duplex theory and its recent extensions.

Table 1: Overview of stimuli and their predicted provision of localization cues.

Stimulus ¹	Bandwidth [Hz]	Envelope slopes [ms] ²	Duration [ms]	Level [dB SPL] ³	Localization cues provision			
					ILDs	Carrier-ITDs	Envelope-ITDs	Spectrum
Burst 3ms	20-15000	0.5	3	65	++	++	++	++
WBN300ms	20-15000	20	300	55	++	++	o	++
LPN 1kHz	20-1000	20	300	60	o	++	o	-
HPN 2kHz	2000 -15000	20	300	60	++	-	o	+
HCT 200Hz	200-10000, CCITT-filter ⁴	20	300	55	++	++	o	++
“Shape”	Speech		800	60	++	++	++	++
Flute 1 Note	440-13000	(20)	180	65	++	o	+	++
Flute 8 Notes	440-13000	(20)	1080	65	++	o	+	++

1 WBN: Wide-band noise (Gaussian noise); LPN: Low-pass noise; HPN: High-pass noise; HCT: Harmonic complex tone.

2 Gaussian slopes applied to envelope, rise-time in ms.

3 Steady-state level of noise before envelope application; Overall level for other sounds.

4 CCITT-filter applied to simulate average spectrum of speech.

Experimental Procedures and Subjects

All combinations of ITDs and ILDs stemming from directions -60, -30, 0, +30, and +60° were tested with each of the eight stimuli. Ten trials were collected for each combination (5 ITDs * 5 ILDs * 10 trials * 8 stimuli = 2000 trials total). The 2000 trials were administered in random order

to prevent adaptation effects to discordant binaural cues. Level was roved randomly in 7 steps within ± 6 dB of the target sound level (Table 1). The presentation was divided into 18 runs. Each run took about 8 min to complete and subjects had to take short breaks between runs. Subjects received at least one training block prior to data collection. Training consisted of 2 runs of 100 trials each in which all direction combinations with all stimuli were presented once.

In a single trial the sound was presented followed by a pause of 0.5 sec after which the adjustable bar appeared in the middle of the line. The subject moved the bar to the lateralized sound position and confirmed this by pressing a button on the trackball. The left button coded hearing a single image while subjects were instructed to press the right button if two or more images were perceived. The bar disappeared and after 0.5 sec the next sound was presented. The line was visible throughout the experiment. At the beginning of each experimental run five uncounted random trails were presented for accommodation. Experiments utilized the Simulated Open-Field Environment in the otherwise darkened anechoic chamber [8].

Results from three subjects are presented in this manuscript while the final study will include five subjects. One male and two female subjects participated, age 21, 22 and 30 years. All subjects had normal hearing thresholds within 300 Hz – 10 kHz as assessed with a Bèkèsy-tracking procedure. Subjects received a payment for their participation. The study protocol was approved by the ethics committee at the University of California at Berkeley.

RESULTS AND DISCUSSION

Figure 2 gives an overview of all subject responses for discrepant and coincident ITD- and ILD-directions. Each row represents results for a different stimulus. Results for the 3 ms wide-band noise burst are given in the top row. Several studies have investigated localization cues for such stimuli and it is thought that ITDs dominate localization through being available in the phase at low frequencies and in the envelope. A pure ITD-dominance would be visible as a diagonal distribution of responses across all panels of a row in Figure 1. Clearly, for the WBN-burst this is not the case. Responses are coarsely aligned along the diagonal; however, for ILDs kept at -60° responses move only slightly to the centre when ITDs are changed to stem from 0° instead of -60° (upper left panel of Figure 1). When ITDs are moved further away than 0° responses split into two separate clusters, indicating perception of two separate images. Similar splits can be seen for most sounds, particularly if discrepancies between ITD- and ILD-directions exceed 60° . Even if no splits are apparent, e.g. for ITDs from 0° and ILDs from $+60^\circ$, the response distribution broadens and response histograms often already show a bimodal distribution.

The second row of Figure 1 shows results from WBN with 300 ms duration. ITD-weighting should be smaller for this stimulus compared to the WBN-burst as envelope-ITDs are not as emphasized. Instead, ITDs might be weighted even more strongly as responses appear more compact, especially for larger ITD-ILD discrepancies in the middle panel (ILDs from 0°).

Results for LPN are strongly ITD-dominated as expected. Responses for HPN, instead, are not purely ILD-dominated. Part of the responses shift according to the ILD-direction, but for large ITD-ILD-discrepancies image splits are visible, i.e. one image is dominated by ITDs, the other by ILDs. This relatively strong ITD-influence on HPN is surprising since envelope cues are not emphasized by modulation. Results for the other two wide-band stimuli, the HCT and the word "Shape", are similar to those of the 300 ms WBN. The two flute-sounds also produced similar responses to the wide-band stimuli, but the audibility of split images differed.

In the following we will evaluate ITD- and ILD-dominance numerically. Figure 3 reports relative ITD- and ILD-influence as well as the percentage of subjectively reported occurrence of two images. The ITD-ILD-weights were computed from the gradient to the surface of the means of subject responses in the ITD-ILD-plane, i.e. the means of the responses in Figure 2. The results in column 1 generally confirm the pattern of binaural cue weights described above. Particularly apparent is the stronger ITD-dominance for the longer duration WBN compared to the burst. Also apparent is that ITD-dominance is restricted to the diagonal – once discrepancies become too large ILD-influence becomes stronger. Eventually responses split in two images which

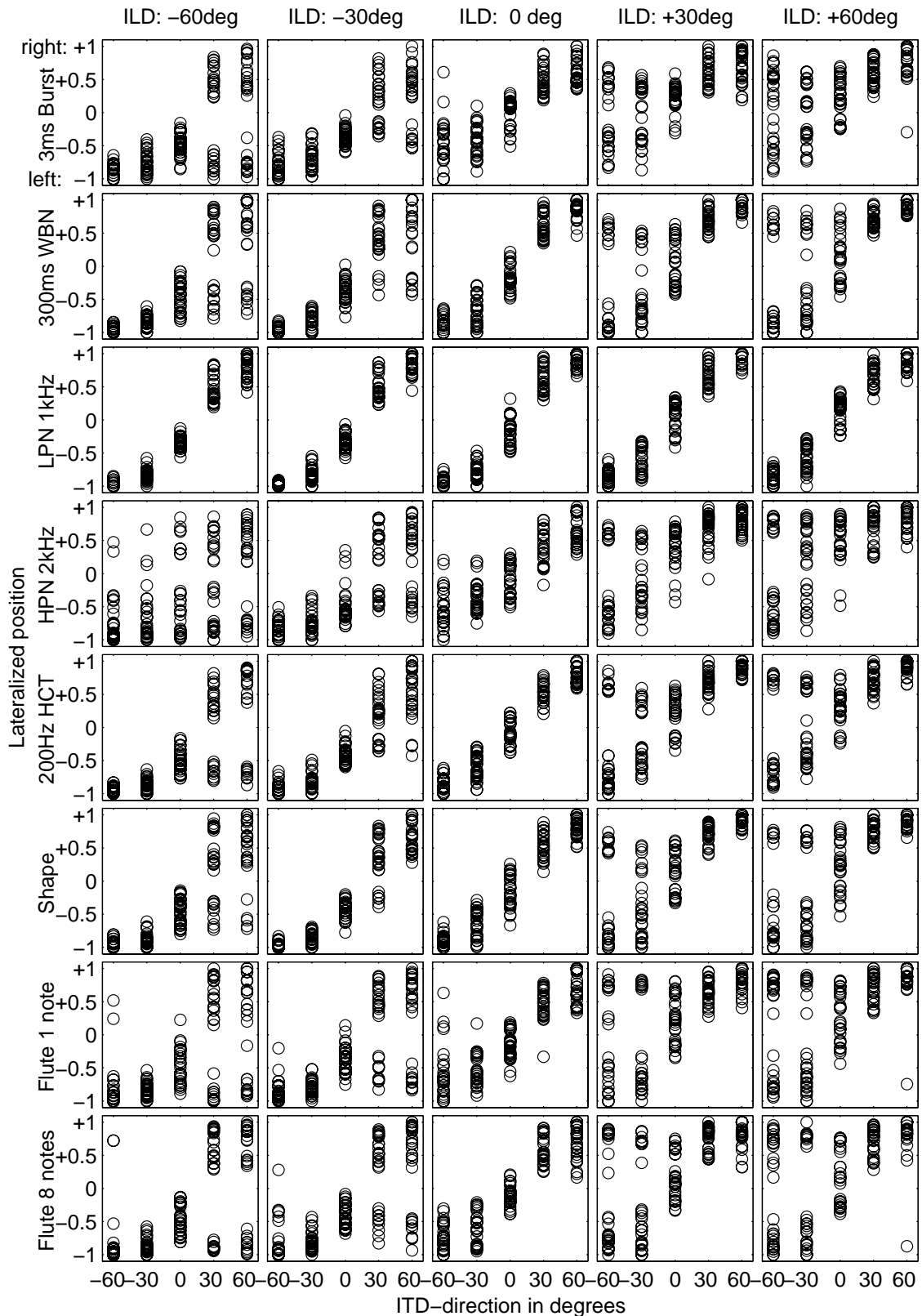
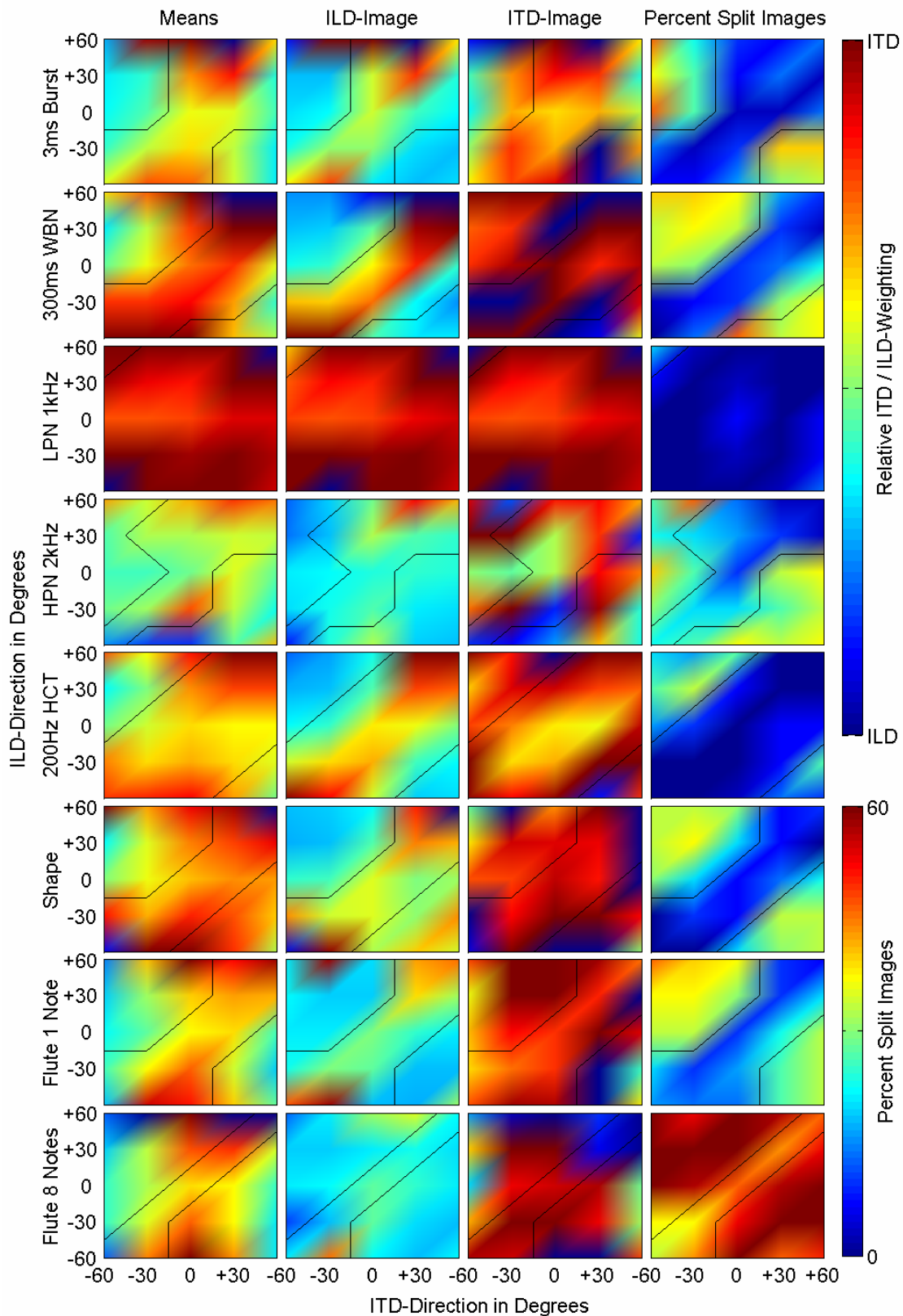


Figure 2: Pooled lateralization responses of three subjects obtained with manipulated HRTFs which contained combinations of ITDs (abscissae) and ILDs (columns) equivalent to directions -60, -30, ..., +60°. Rows identify results separately for each stimulus.

Figure 3 (next page): Relative influence of ITDs and ILDs in pooled responses of three subjects (columns 1-3) and the percentage of subjectively reported split images (column 4). Column 1 reports ITD- and ILD-dominance computed from the gradient to the means of pooled responses.



[Figure 3, cont'd]: As the mean is not representative in the case of split images responses were clustered into separate ILD- and ITD-images. Limits for separation into two clusters are indicated by black lines. Along the diagonal a single cluster was used while split clusters appeared towards larger ITD-ILD-discrepancies. Columns 2 and 3 give relative ITD- and ILD-weights from the gradient to the cluster means of the ILD- and ITD-cluster, respectively. The rightmost column lists the percentage of reported split images. Limits for separation in two clusters are coincident with subjective reports.

requires special analytical treatment. We used the expectation-maximization algorithm for Gaussian mixture estimation to estimate means and weights of two Gaussian distributions fit to the responses. Further analysis was based on two clusters if this Gaussian mixture model represented response data better than the fit of a single Gaussian distribution as assessed with a log-likelihood criterion at the 1%-level (χ^2 , 3 DOF). An additional threshold criterion was derived from the subject's response if one or more images were heard (last column in Figure 3).

Columns 2 and 3 of Figure 3 show ILD-ITD-weights computed from the ILD- and ITD-images, respectively. Along the diagonal where a single image was heard the pattern shows either ITD-dominance or at least equal weighting for both cues. Off the diagonal, where two images were heard, the ILD-image shows ILD-dominance and the ITD-image cluster ITD-dominance. ITD-dominance, however, is not complete. For nearly every sound the ITD-image retains some sensitivity to ILDs at selected ITD-ILD combinations, visible e.g. for the 8 note flute sound, the HPN or the 300ms WBN.

The last column of Figure 3 shows the percentage of split images. Image splits occur for larger ITD-ILD discrepancies and depend on sound spectrum and duration. Nearly no splits are reported for the low-pass noise, presumably because ILDs are too small at those frequencies. The pattern of image splits is particularly interesting for the flute sound. Most splits were reported for the short flute melody, while few splits were reported just for its first note. Even though the temporal modulation due to blowing is small it seems lead to multiple images. Another example is given by the HCT in comparison to the word "Shape". Image splits occur less often for the steady HCT with the average spectrum of speech compared to the word.

CONCLUSIONS

Relative weighting of binaural cues was studied out of their natural combination for synthetic and real-world sounds. ITDs dominate lateralization especially if introduced cue conflict is small; for larger discrepancies the role of ILDs becomes stronger. When cue conflict was small, ITDs received a high cue weighting even if sounds contained mostly high frequencies. Temporal modulation seems to encourage an increase in segregation of the sound.

ACKNOWLEDGEMENTS

I would like to thank Prof. Ervin Hafter for his extensive support. This work was financed by NIH RO1 DCD 00087. Parts of the data analysis and the write-up of the manuscript occurred after BS joined the MRC.

References:

- [1] J. W. Strutt, "On our Perception of Sound Direction," *Philos. Mag.*, vol. 13, pp. 214-232, 1907.
- [2] F. L. Wightman and D. J. Kistler, "The dominant role of low-frequency interaural time-differences in sound localization," *J. Acoust. Soc. Am.*, vol. 91, pp. 1648-1661, 1992.
- [3] R. L. Freyman and P. M. Zurek, "Onset dominance in lateralization," *J. Acoust. Soc. Am.*, vol. 101, pp. 1649-1659, 1997.
- [4] E. R. Hafter and L. A. Jeffress, "Two-Image Lateralization of Tones and Clicks," *J. Acoust. Soc. Am.*, vol. 44, pp. 563-569, 1968.
- [5] F. L. Wightman and D. J. Kistler, "Factors Affecting the Relative Salience of Sound Localization Cues," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. H. Gilkey and T. R. Anderson, Eds. Mahwah, New Jersey: Lawrence Erlbaum Ass., 1996, pp. 1-23.
- [6] B. U. Seeber and H. Fastl, "Subjective Selection of Non-Individual Head-Related Transfer Functions," in *Proc. 9th Int. Conf. on Aud. Display*, E. Brazil and B. Shinn-Cunningham, Eds. Boston, USA: Boston University Publications Prod. Dept., 2003, pp. 259-262.
- [7] E. A. Macpherson and J. C. Middlebrooks, "Listener weighting of cues for lateral angle: The duplex theory of sound revisited," *J. Acoust. Soc. Am.*, vol. 111, pp. 2219-2236, 2002.
- [8] E. Hafter and B. Seeber, "The Simulated Open Field Environment for auditory localization research," in *Proc. ICA 2004, 18th Int. Congress on Acoustics, Kyoto, Japan, 4.-9.04.2004*, vol. V: Int. Commission on Acoustics, 2004, pp. 3751-3754.