Weighting of binaural cues in the absence of a reflection

Bernhard U. Seeber

MRC Institute of Hearing Research, University Park, Nottingham, NG7 2RD,
Auditory Perception Lab, Dept. Of Psychology, UC Berkeley, Berkeley, CA, 94720-1650
seeber@ihr.mrc.ac.uk

Summary and Introduction

More than 100 years ago Lord Rayleigh postulated the Duplex Theory of auditory localisation. Low frequency and wide-band sounds are localised based on interaural time differences (ITDs), while high frequency sounds are localised by interaural level differences (ILDs) [1, 2]. In natural situations, the relationship of those binaural cues is provided in the head-related transfer function (HRTF). Hearing devices may alter it and it is important to correctly reproduce those cues in a particular listening situation to afford the full benefit from binaural hearing.

Binaural cue dominance was evaluated by manipulating HRTFs to impose a cue discrepancy such that ITDs stemmed from one direction while ILDs and spectral cues were in accord with another direction. Subjects reported if they heard one or two sound images and the position of the lateralized image was recorded. A cluster analysis was used to assess if responses form two separate images. Results show a generally high weighting for ITDs, and even for a 2 kHz high-pass noise ITDs received almost equal weight to ILDs. Auditory objects split frequently starting at spatial discrepancies between ITDs and ILDs of only 30°. This high sensitivity to binaural discrepancies suggests that image splits might also occur in situations where compression in hearing devices alters binaural cues.

Methods

Virtual acoustics

The relative weighting of binaural cues was studied by varying ITDs and ILDs in virtual acoustical space. First, subjects selected a set of HRTFs from a catalogue of non-individual HRTFs. The selection procedure identifies HRTFs for each subject which improve performance with respect to localization error, variance and externalization [3]. HRTFs from different originating directions were divided into their amplitude and phase components using the Fast-Fourier-Transformation. Amplitude components from one direction were then recombined with phase components from another direction such that ILDs come from, e.g., +30° while ITDs stem from -60° [2]. Sound stimuli were filtered with the recombined HRTFs to yield a virtual acoustics stimulus which was presented through Sennheiser HD580 headphones.

Lateralization method

In most conditions, particularly with discordant ITDs and ILDs, stimuli were perceived within the head. A line-dissection method was used to indicate the position of the lateralized the sound image on the interaural axis. A white line was projected on an otherwise black screen in 2 m distance in front of the subject. It covered a visual angle of approx. ±25° and the endpoints were marked with vertical bars and the words “left ear” and “right ear”. For data analysis, the left ear position was assigned “-1” and the right ear “+1”. The subject adjusted a red bar to the perceived lateralized sound position with a trackball. There was no specific instruction to point to the leftmost or rightmost sound in the case of image splits and subjects indicated the most salient image.

Stimuli

Four different stimuli that should give rise to different weightings of binaural cues were used in the experiments: (1) A burst of wide band noise (WBN, 300 ms duration, 20 ms Gaussian slopes, 20 Hz - 15 kHz), (2) a high-pass noise (HPN, 2 - 15 kHz, 300 ms duration, 20 ms slopes), (3) a harmonic complex tone (HCT, fundamental frequency 200 Hz, 300 ms duration, 20 ms slopes, 20 Hz - 10 kHz), and (4), the word “shape” spoken by a female speaker.

Experimental Procedures and Subjects

All combinations of ITDs and ILDs stemming from directions -60, -30, 0, +30, and +60° were tested with each stimulus. Ten trials were collected for each combination (5 ITDs * 5 ILDs * 10 trials) and administered in random order with roving level in ±6 dB. The presentation was divided into 18 runs of about 8 min each and training was given prior to data collection.

In each trial, the adjustable bar appeared in the middle of the projected line 0.5 sec after sound presentation. 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 the right coded hearing two images. The bar then disappeared and after 0.5 sec the next sound was presented [4].

Results of three subjects with normal hearing, one male and two female (age 21, 22 and 30 years) are presented. Subjects were paid and the study protocol was approved by an ethics committee.

Results and Discussion

The first 3 columns of Figure 1 report the relative influence of ITDs and ILDs on localization computed from the gradient to the surface of mean responses in the ITD-ILD-plane. In column 1 this surface was formed from the mean of all responses. The results for the 300 ms WBN show that ITDs dominated in situations in which they were consistent with ILDs. ITD dominance is also found for the HCT and the word “shape”. For a high-pass noise one would expect ILDs to dominate localization as expressed in the Duplex-theory [5]. However, results show that mean responses are about equally affected by ILDs and ITDs.

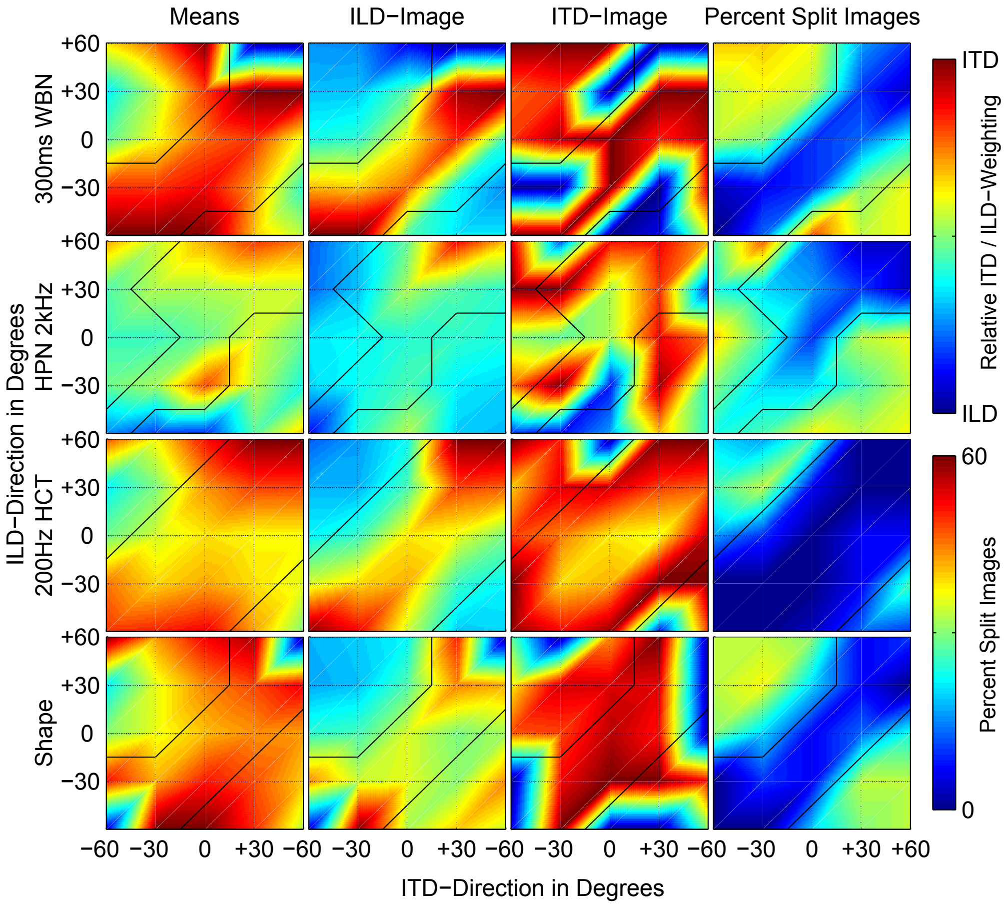

Figure 1: Relative influence of ITDs and ILDs computed from pooled responses (columns 1-3) and the percentage of subjectively reported split images (column 4). Column 1 reports ITD- and ILD-dominance derived from the gradient to the means of pooled responses. Cluster analysis was used to separate responses into two clusters representing ILD- and ITD-images. Columns 2 and 3 give the relative ITD- and ILD-weights from the gradient to the cluster means of the ILD- and ITD-clusters, respectively. The rightmost column lists the percentage of split images reported by the subjects.

Cluster analysis was used to assess the impact of image splits. The expectation-maximization algorithm for Gaussian mixture estimation was used 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 (Chi2, 3 DOF). Additionally, a certain percentage of subjective reports of hearing two images was required.

Columns 2 and 3 of Figure 1 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 clear ITD-dominance or about equal weighting of 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 and the ITD-image retains some sensitivity to ILDs at selected ITD-ILD combinations, e.g. for the HPN or the 300ms WBN.

The last column shows the percentage of split images which already occur for ITD-ILD discrepancies as low as 30°. Image splits appear less often for the steady HCT with the average spectrum of speech compared to the word. It appears as if sensitivity for splits is increased by the temporal modulation inherent in the word or possibly its longer duration.

Acknowledgements

I would like to thank Prof. Ervin Hafter for fruitful discussions. This work was supported by the intramural programme of the MRC and by NIH RO1 DCD 00087.

References

[1] Strutt, J.W.: *On our Perception of Sound Direction.* Philos. Mag., 1907. **13**: 214-232.

[2] Wightman, F.L. and D.J. Kistler: J. Acoust. Soc. Am., 1992. **91**(3): 1648-1661.

[3] Seeber, B.U. and H. Fastl: *Subjective Selection of Non-Individual Head-Related Transfer Functions*, in *Proc. 9th Int. Conf. on Aud. Display*, 2003, pp. 259-262.

[4] Seeber, B.U., S. Kerber, and E.R. Hafter: Hear Res, 2010. **260**(1-2): 1-10.

[5] Macpherson, E.A. and J.C. Middlebrooks: J. Acoust. Soc. Am., 2002. **111**(5): 2219-2236.