# THE ROLE OF INTERAURAL TIME AND LEVEL DIFFERENCES WHEN LOCALIZING NOISE IN NOISE

PACS: 43.66.PN

Jonas Braasch
Institut für Kommunikationsakustik
Ruhr-Universität Bochum
Universitätsstr. 150
44787 Bochum, Germany
Tel.: ++49-234-3225387
Fax.: ++49-234-3214165
Email: braasch@ika.ruhr-uni-bochum.de

ABSTRACT

The reliability of ITD and ILD cues of a broadband-noise target within a broadband-noise masker was investigated utilizing two localization algorithms: (i) an interaural cross-correlation algorithm (ICC), and, (ii) an algorithm which estimates the ILD using excitation-inhibition cells (EI). The EI-model performance turned out to be inferior to the performance of both humans and the ICC-model. One difficulty was to estimate the sound pressure level of the contralateral channel of the target, another was the high sensitivity of the EI-model to the statistical fluctuations of both the target and the masker in the single frequency bands.

## INTRODUCTION

The aim of this investigation was to better understand the human's ability to localize a noise target within a noise distracter. For the single target condition, Wightman and Kistler (1997a) suggested that the ITDs are the dominant cues. They derived their hypothesis by conducting a localization test, using head-related transfer functions (HRTFs), whose ILDs were set equal to zero throughout the whole frequency range. Interestingly, this manipulation was found to have almost no effect on the listeners' performance to localize the target in the left/right dimension. In the same year, however, Wightman and Kistler (1997b) conducted a second localization test with manipulated HRTFs which results were in contradiction to their previous findings. In respect to this study, the left channel of the HRTFs was attenuated unilaterally by several decibels. Adjusting the attenuation to 10 dB the listeners were able to localize the target fairly well regarding the left/right dimension, but when the attenuation was set to 20 dB their localization performance was strongly affected.

In the present study, a localization test was designed to investigate, if human listeners base their judgements on ITDs or rather on ILDs in those cases where the target is partly masked. Six listeners were instructed to localize unilaterally attenuated random noise bursts (20 dB attenuation in the left channel) in an auditory virtual environment using individual HRTFs. In the first condition the target was presented in an isolated way and in the second condition in presence of a broadband-noise distracter that was manipulated in the same way as the test signal. The results have been previously presented at the joint EAA/ASA/DAGA meeting in Berlin (Braasch and Hartung, 1999).

To evaluate the influence of ITDs and ILDs when localizing a unilaterally attenuated sound source, model simulations were carried out employing the test signals of the psychoacoustic experiment. The ICCD model (Braasch 2001) was used to process ITD cues using a cross-correlation algorithm. The ILD cues were evaluated with a model based on excitation-inhibition cells, which is described further below. The importance of both ITD and ILD cues in the psychoacoustic localization task was then analyzed by comparing the results of both models.

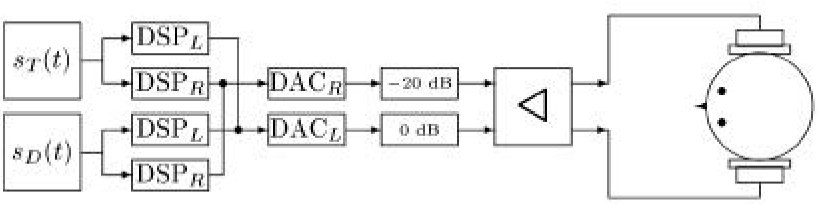

*Figure 1: Virtual auditory environment to create unilaterally attenuated sound sources. $s_T(t)$: target signal, $s_D(t)$ distracter signal (DSP: digital signal processor for HRTF filtering, DAC: digital analog converter).*

## LOCALIZATION OF UNILATERALLY ATTENUATED SOUNDS

### Methods

The localization test was carried out in a virtual environment using individual HRTFs. The measurement of the HRTFs is described by Djelani et al. (2000). The six listeners (one female, five males), who took part voluntarily in the experiment, showed normal localization abilities with their HRTFs in an evaluation test that was conducted beforehand. In this evaluation test a broadband-noise burst had to be localized from various directions within the whole sphere. The listeners' age ranged from 23 to 33 years. All the listeners had normal hearing (hearing loss < 20 dB at octave frequencies between 125 and 8000 Hz).

The target and the distracter were broadband-noise bursts with 200 ms and 500 ms duration (frequency range: 200 Hz-14 kHz, 20 ms $\cos^2$-ramps). The onset delay between the target and distracter was set to 200 ms (the distracter was partly preceding the target). The signals were generated on a specialized DSP hardware (Tucker-Davis Technologies (TDT), PD1-System, 16 bit resolution, 44.1 kHz sampling rate) and filtered with the individual HRTFs of the listeners. Then, the signals were lowpass filtered at 20 kHz and D/A converted. Next, the right signal was attenuated by 20 dB (TDT, PA4) as shown in Fig. 1. The left signal was sent through a second attenuator (0 dB attenuation) of the same type for symmetry reasons. Afterwards the signals were amplified (TDT, HB6) and presented to the listener through headphones (STAX, SR-Lambda). The direction of the HRTFs for the distracter was set at 0° azimuth and 0° elevation throughout the whole experiment, while the position of the target was varied between 12 equidistant positions in the horizontal plane beginning from 0° azimuth in steps of 30°. Each trial was repeated 10 times. The sound pressure level of the distracter was set at approximately 70 dB SPL (left channel), respectively 50 dB SPL (right channel). The SPL was measured at the ear-canal entrances of a dummy head (head acoustics, KK1412). The T/D-ratio, the power level ratio of the target to the distracter before they are filtered with the HRTFs, was set at 0 dB.
For each listener, the stimuli were presented in a pseudo random order in one session lasting about 14 minutes. The session started with a training phase of 10 training trials. During the experiment each listener was seated on a chair. The listener was asked to keep his/her head at a fixed position during the presentation of the stimuli. After a stimulus had been presented the listener reported the direction of the externalized auditory event of the target using the GELP-method (Gilkey et al., 1995). After the response, the next stimulus was presented with a delay of 2 seconds. No trial-by-trial feedback was provided to the listeners during the training phase or the recording phase.

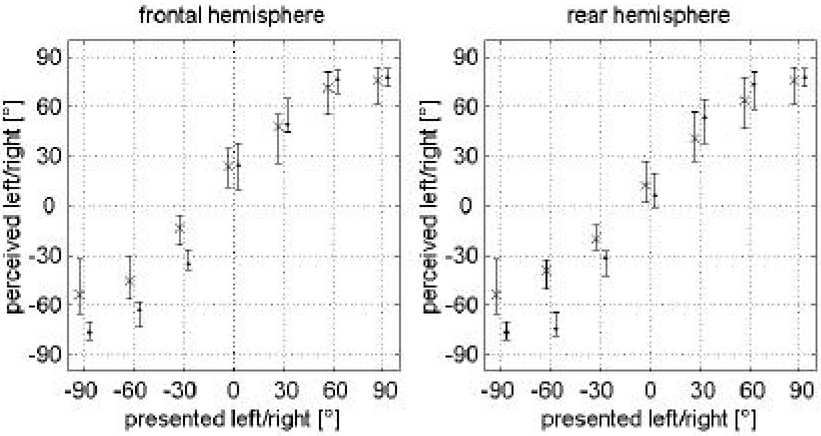

*Figure 2: Median average out from all six listeners localizing an unilaterally attenuated (20 dB, right ear) target. The target alone condition is marked with 'x's, the masked condition with dots. The left graph shows the results for the frontal hemisphere, the right graph for the rear hemisphere. The errorbars show the lower and upper quartiles.*

Results

The results for the single target in the left/right dimension (Wightman and Kistler, 1997a) are shown in Fig. 2 (marked with 'x's). The asymmetric shape of the localization curve is not surprising, because the right channel was attenuated immensely. Especially the 0° azimuth target position and all tested directions in the right hemisphere were perceived more lateralized towards 90°. When the distracter is presented additionally to the target—both signals are attenuated unilaterally in this case—the listeners' responses change significantly (Fig. 2, dots). While the perceptual shift towards 90° azimuth is still observable at the 0° azimuth target position, the shift decreases significantly for the outer angles in the right hemisphere.

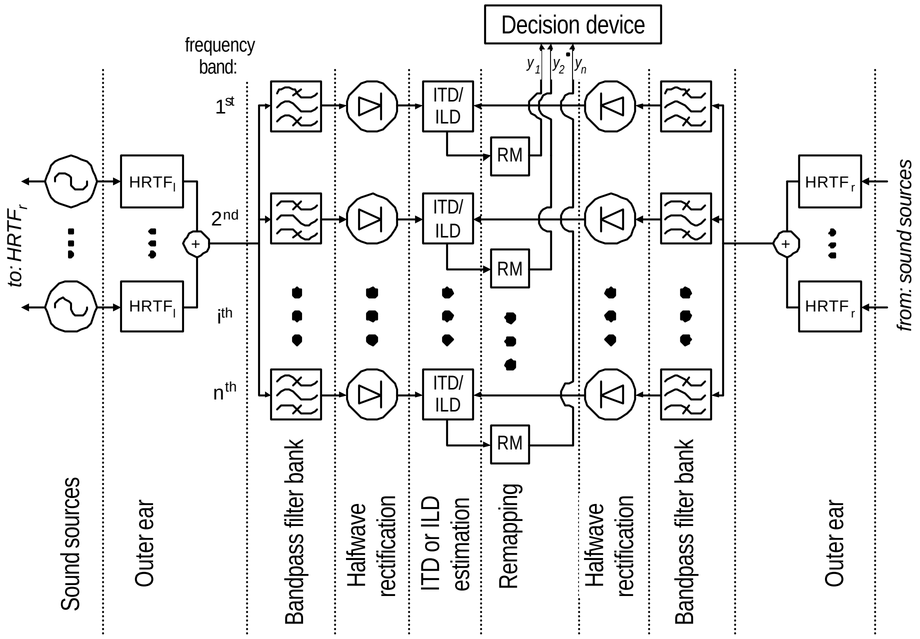

*Figure 3: Structure of the localization algorithms*

## MODEL SIMULATION

Model Structure

The general structure of the model to analyze ILDs or ITDs is shown in Fig. 3. The transformation from the sound source positions (in this case target and distracter) to the eardrums are taken into account by filtering the sounds with the HRTFs of a specific direction.

Afterwards, the outputs for every sound source are added for the left and then right channel. Basilar-membrane and hair-cell behavior are simulated with a gammatone filter bank with 36 bands at a sampling frequency of 48 kHz, as described by Patterson et al. (1995), and a simple half-wave rectification. After the half-wave rectification either the ITDs or the ILDs were estimated within each frequency band over the whole target duration.

The ITDs are estimated using a cross-correlation algorithm, comparing the signals from the left and right channel to each other, in a similar way to Blauert and Cobben (1978). Here, only the frequency bands 3 to 12 (200-1200 Hz) are analyzed to take into account that the human auditory system cannot resolve the temporal fine structure at high frequencies.

The ILDs were estimated using an algorithm based on excitation-inhibition (EI) cells similar to the one proposed by Breebaart et al. (2001). The output $\mathbf{E}(r)$ is the activity of an EI cell that is tuned to a certain ILD $r$. In this model simulation 81 EI-cells were simulated with $r$ between –40 and 40 dB in steps of 1 dB. The activity of the cells $\mathbf{E}(r)$ are calculated as follows:

$$\mathbf{E}(r)=\exp(10^{r/20}\,P_l-10^{-r/20}\,P_r)$$

The variables $P_l$ and $P_r$ are the sound pressure level from the left respectively right channel calculated over the whole target duration. The output of this algorithm has a very similar shape to the output of the ICC function. For the ILDs the whole frequency range of the HRTFs (200-12000 Hz) was analyzed.

Afterwards, the ITDs and ILDs are remapped to azimuth positions. To calculate the ITDs and ILDs of the HRTFs throughout the horizontal plane, the HRTF catalog, which was measured in a resolution of 15° in the horizontal plane, was interpolated to 1° resolution using the spherical spline method (Hartung et al., 1999). Using the described model algorithms a map $g(\alpha,\omega)$ with all ITDs and ILDs values for the frontal horizontal plane (1° resolution) was established for all frequency bands. Each map allows the ICC and EI outputs to be remapped to a basis of azimuth angles $\alpha$.

The decision device in both the ITD and ILD model was established by integrating the values $\Psi_{rm}(\alpha,\omega)$ over the analyzed frequency bands. The sound sources are estimated at the positions of the maximum peak of the integrated function.

Simulating localization in presence of the distracter

Previously, it was reported in several investigations that humans are able to localize sounds quite well, even if the target sound is presented in the presence of a second distracting sound. In order to model this observation, the existing lateralization models focus on the analysis of the test sounds in frequency bands or time gaps that are not disturbed by the distracter. However, human localization also works quite well in cases where the frequency range of test sound and distracter are identical, even if the distracter is present during the complete duration of the test sound. Psychoacoustic test have revealed that the onset of test sound and distracter must be separated temporally in those cases, otherwise human localization performance deteriorates rapidly (Thurlow and Jaques, 1975).

To take this behavior into account in the model algorithms, the ITDs of the test sound are calculated from the difference in the interaural cross correlation (ICC) of the total sound Y (target signal+distracter) and the ICC of the distracter alone ($\Psi=\Psi_Y-\Psi_D$) for each frequency band. The ICC of the distracter is calculated from the part of the distracter that precedes the test sound. Note that the time intervals do not include the 20 ms-cos$^2$ ramps of the target and distracter signal. The time windows, in which the signal was analyzed, were 20 ms–180 ms (preceding part of the distracter) and 220 ms–380 ms (whole signal). The ILD differences are estimated using the difference function in the same manner.

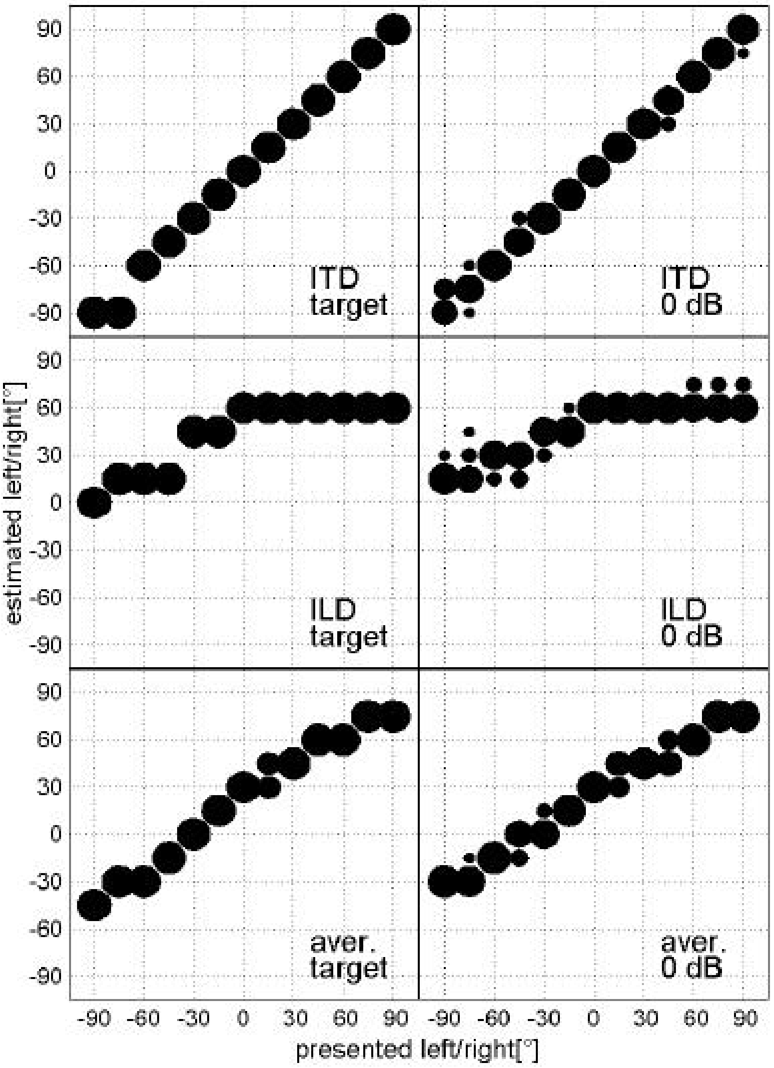

*Figure 4: Model localization performance for 20 dB unilaterally attenuated sounds. Top row: analysis of ITD cues; center row: analysis of ILD cues; and bottom row: average of both ITDs and ILDs. The panels in the left column show the simulation results for an isolated target. The same conditions are shown in the right column for a distracted target (0 dB T/D-ratio).*

Model Results

In Fig.4, the target alone conditions are presented in the left panels, the distracted conditions in the right panel. The results for the ICCD model are given in the upper row. They are identical to the results for broadband-noise sounds that were not attenuated unilaterally. This finding is not surprising, because the normalized ICC algorithm is independent of the absolute input values of the two input functions. The results for the ILDs however are very much influenced by the unilateral attenuation (center row), and the localization curve is more distorted than it was actually observed in the psychoacoustic investigation. Again, there are only minor differences between the target alone and the distracted condition. In the last row, both model approaches are combined, by estimating the target positions as the mean from the target estimations of the ICCD and the EID model. The result is quite a good estimation of the psychoacoustic data for the target alone condition, while the psychoacoustic data of the distracted case is in a better agreement to the output of the ICCD model (ignoring the ILD cues).

## DISCUSSION

Even though the unilateral manipulation of the signals increases the variance in some of the listeners' responses, it is obvious that there are characteristic differences in the judgements of the listeners between the target alone and the distracted condition. The shift towards the left direction for the target positions in the right hemisphere is only apparent for the target alone condition. The best model simulation for the target alone condition is the combined ITD/ILD model that estimates the sound source positions from the average estimates of the ICCD and the EID model. In the distracted case, however, an estimation based on the ITD cues only is a much better approximation than the combined mode. Nevertheless, in the distracted condition the auditory events are shifted almost 30° azimuth towards the left side, if the target is presented from 0° azimuth. There is an easy explanation for this exception: Only in this position the binaural cues are identical to those of the distracter and, therefore, it can be assumed that the listeners perceived the target coming from the same position as the distracter, because ITDs and ILDs remain constant when the target is presented additionally. The localization of the distracter is a single-source localization task, and therefore comparable to the localization task, when the target is presented in absence of the distracter.

The outcome of the psychoacoustical experiment indicates that the influence of the ITD cues increases when the target is partly masked. This is in good agreement with assumptions that were derived during the model simulations, were it could be shown that the output of the ICC model is more stable when the preceding part of the distracter is subtracted, than it is the case for the ILD model based on an EI algorithm that averages over time. Firstly, the ILDs are determined directly from the amplitudes of target and distracter and therefore the statistical fluctuations cause a greater variance than it is the case for the ITDs. Note that in the ICC algorithm basically only the amplitude of the ICC function is affected but not the position of the peak, which determines the position of the sound source. Secondly, as already discussed in the previous sections, in those cases where a very large ILD occurs, its estimation in a masked condition is problematic, because the resulting signal-to-noise ratio at the contralateral eardrum is very low.

## LITERATURE

BLAUERT, J. and W. COBBEN **(1978)**: *Some consideration of binaural crosscorrelation analysis*, Acustica 39, p. 96-104.

BREEBAART, J., S. VAN DE PAR, S. and A. KOHLRAUSCH **(2001)**: *Binaural processing model based on contralateral inhibition. I. Model structure*, J. Acoust. Soc. Am. 110, p. 1074-1088.

BRAASCH, J. and K. HARTUNG **(1999)**: *Localization of distracted sound sources: Determining the role of binaural cues using unilaterally attenuated and interaurally uncorrelated signals* Acustica/acta acustica, 85, p. S225 (abstr.).

BRAASCH, J. **(2000)**: *A model to simulate human sound localization and detection in presence of a distracter*, J. Acoust. Soc. Am. 108, S. 2597 (abstr.).

DJELANI, T., C. PÖRSCHMANN, J. SAHRHAGE and J. BLAUERT **(2000)**: *An interactive virtual-environment generator for psychoacoustic research II: Collection of head related impulse responses and evaluation of auditory localization*, ACUSTICA/acta acustica 86, p. 1046-1053.

HARTUNG, K., J. BRAASCH and S. STERBING **(1999)**: *Comparison of different methods for the interpolation of head-related transfer functions*, in: AES 16th International Conference on Spatial Sound Reproduction, p. 319-329.

PATTERSON, R.D., M.H. ALLERHAND and C. GIGUÈRE (1995): *Time-domain modeling of periphal auditory processing: A modular architecture and software platform*, J. Acoust. Soc. Am. 98, p. 1890-1894.

THURLOW, W.R. and J.C. JAQUES **(1975)**: *Localization of two noise sources overlapping in time*,
J. of Speech and Hearing Res. 18, p. 663-671.

WIGHTMAN, F.L. and D.J. KISTLER **(1997a)**: *Factors affecting the relative salience of sound localization cues*, in: Binaural and spatial hearing in real and virtual environments (ed. R.H. Gilkey, Robert H. and T.R. Anderson, Lawrence Erlbaum Associates, Mahwah 1997, p. 1-23.

WIGHTMAN, F.L. and D.J. KISTLER **(1997b)**: *Monaural sound localization revisited*, J. Acoust. Soc. Am. 101, p. 1050-1063.