Using a combined localization/detection model to simulate human localization performance near the masked threshold level

Jonas Braasch
Institut für Kommunikationsakustik, Ruhr-Universität Bochum

Abstract

Recently, it was shown that the perceptual lateralization of a partly masked target can be simulated using the interaural cross-correlation difference (ICCD) algorithm [1,2]. In reference to this model, the interaural time differences (ITDs) of a target are calculated from the difference in the interaural cross correlation (ICC) of the total sound target+distracter and the ICC of the distracter alone, which is estimated from the part of the distracter that precedes the target. However, in some listening conditions, the target level was below the masked detection threshold level, in which the model could not fully simulate the listeners' responses. To be able to simulate those conditions a detection threshold model was implemented into the localization model. Now, the localization process is only triggered when the target is detected by this stage, and otherwise the position of the sound source is determined according to a behavioral pattern. The detection stage is based on an on- and offset detection algorithm that analyzes the derivative in time. For the simulation of the binaural conditions, the Equalization-Cancellation model (EC) algorithm of Durlach was included into the model. It has been shown that for target levels above the masked detection threshold the on-/offset detection algorithm is accurate enough to trigger the subtraction process of the ICCD algorithm. Furthermore, the simulation of the clustering and scattering effects, which are localization patterns observed for low target-to-distracter ratios, could be improved.

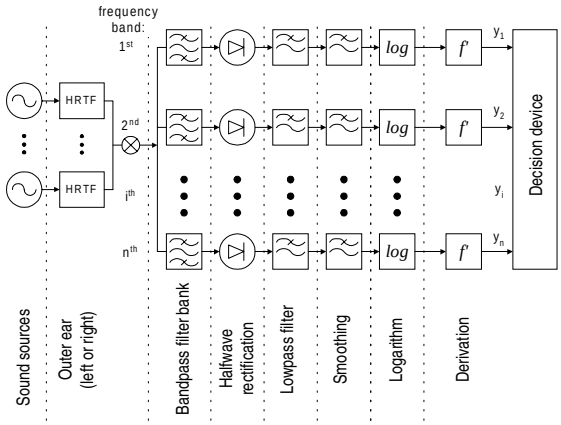

Fig. 1: Model structure of the monaural detection algorithm.

Model structure

An onset detection algorithm (Fig. 1), detecting changes in the envelope of the signal was developed to fulfil the detection task. This specific algorithm had the advantage that it could not only be applied to estimate the detection threshold of the target in noise, but that it could be also applied to estimate the target on- and offset times.

Periphery

The sound sources are filtered with HRTFs of the specific direction and subsequently added separately for the left and right channel. Afterwards, in order to simulate the frequency decoding of the basilarmembrane [3], the signals are processed by a gammatone filter bank with 30 bands at a sampling frequency of 48 kHz, covering a frequency range of 200-12000 Hz. A simple half-wave rectification is implemented to take the inner hair cell response into account. To consider that the fine structure of the signal cannot be resolved in the higher frequency bands, the signals were lowpass filtered after the half-wave rectification (t_g=1 kHz).

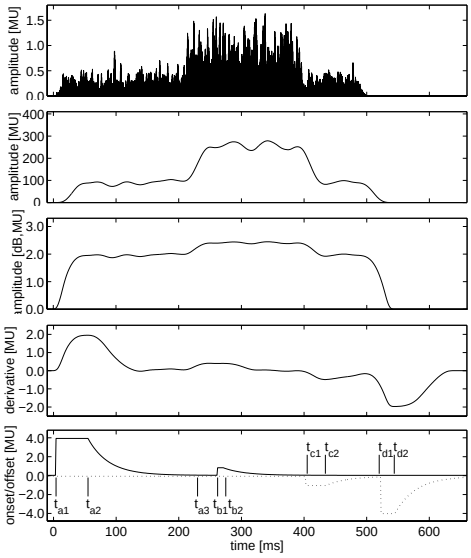

Fig. 2: Stages of the onset detection model.

Monaural detection stage

To reduce the effect that temporal statistic fluctuations of the distracter (Fig. 2, top panel) lead to the detection of an onset, the signal in each frequency band was convolved with a hamming window (50 ms window length, 2^{nd} panel from top) and downsampled by a factor of 48. Next, the logarithm to the base 10 was calculated (3^{rd} panel from top); the derivative of the signal was then determined by subtracting the average from the last 50 values from the current value (4^{th} panel from top); internal noise was added to the output of each frequency band; and finally the weighted sum Y over all frequency bands i was calculated ($Y = \sum_{i=1}^{30} \lambda_i y^i$). The amplitude of the internal noise was determined by adjusting the outcome of a loudness discrimination experiment (pair comparison test, 500 Hz sinusoidal tone with 20 ms $\cos^2$-ramps), to a discrimination threshold of 1 dB. All parameters concerning the internal noise were kept constant during the simulation to estimate the detection threshold level.

The onset of the distracter t_{a1} is considered to be detected when its input exceeds the minimal detection threshold (Fig. 2, bottom panel, solid line). Next, the local peak for this specific onset t_{a2} is determined and the new detection threshold is set to twice the maximum value of the peak, plus the minimal threshold value. Afterwards, the detection threshold decays exponentially with time. At the time instant t_{a3} the detection threshold falls below the threshold, the recovery phase of the model is considered to be completed and the algorithm is able to detect onsets again. In this figure, the onset of the target is detected, too (t_{b1},t_{b2}). The offset detection algorithm is very similar except that here the negative values of the input function are considered (Fig. 2, bottom panel, dotted line, target: (t_{c1},t_{c2}), distracter: (t_{d1},t_{d2}).

Binaural detection stage

For the binaural detection tasks, the Equalization-Cancellation (EC) [4] stage (including internal noise at this stage) was implemented into the proposed model, connecting the lowpass filter output of the left and the right processing channel. Afterwards, the signal is processed in

analogy to the monaural detection stage: convolution with a 50 ms hamming window, resampling to a sampling frequency of 1 kHz, mapping by a logarithmic function, addition of internal noise and calculation of the derivative function of the signal.

Procedure

In a pair comparison test a train of three stimuli $n = 1, 2, 3$ was presented to the model algorithm. All three signals consisted of the distracter (white-noise burst, 500 ms duration, 20 ms cos^2-ramp, 0° azimuth, 0° elevation), but only one stimulus, which was chosen by random, contained additionally the target signal (also white-noise burst, 200 ms duration, 20 ms cos^2-ramp, various azimuth angels, 0° elevation, presented with 200 ms delay in regard to the distracter). In order to detect the signal, the model algorithm estimated the maximum values of the model output for each stimulus. Here, only the time interval where the target signal was expected (180 ms–480ms) was analyzed, and it was assumed that the stimulus with the highest output value contained the target signal.

The masked detection thresholds were measured using an adaptive three interval forced-choice (IFC) procedure with adaptive level tracking (two-down one-up rule, initial step size 8 dB, final step size 1 dB, [5]). The detection thresholds were averaged over the medians from the data collected in 50 runs. The initial T/D-ratio was set to 0 dB.

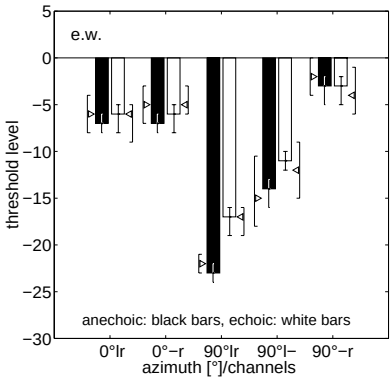

Fig. 3: Detection thresholds.

Results

The average detection thresholds obtained by the model are displayed in Fig. 3. The different conditions are shown in comparison to results of a listening test marked by the triangles (The braces show the interquartile range of the psychoacoustic results, the data is an average over 5 listeners). The following conditions were tested (from left to right): target 0°/binaural condition, target 0°/monaural condition (right ear), target 90°/binaural condition, target 90°/monaural condition (left ear), target 90°/monaural condition (right ear). The black bar of each group shows the masked detection threshold for the anechoic condition, while the white bar shows the masked detection threshold for the reverberant condition. The reverberant environment was created according to [6]. Each bar shows the median averaged over the turning points obtained in all runs. The estimation of each median value is based on 500 measurement points. The upper and lower quartiles of the population are given by the error bars.

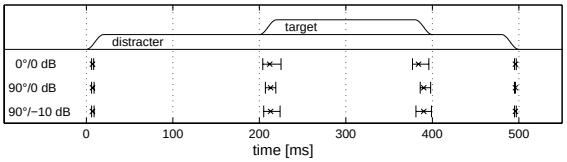

Fig. 4: Estimated on- and offset times.

Fig. 4 shows the results for the on- and offset time estimation. The target was always presented from 0° azimuth, 0 dB T/D-ratio (top graph, below the time course), 90° azimuth, 0 dB T/D-ratio (middle graph), 90° azimuth, –10 dB T/D-ratio (bottom graph). For every condition, the stimuli were repeated 1000 times, the distracter was presented from 0° azimuth. The threshold level was set to 0.12 [MU] for the onset

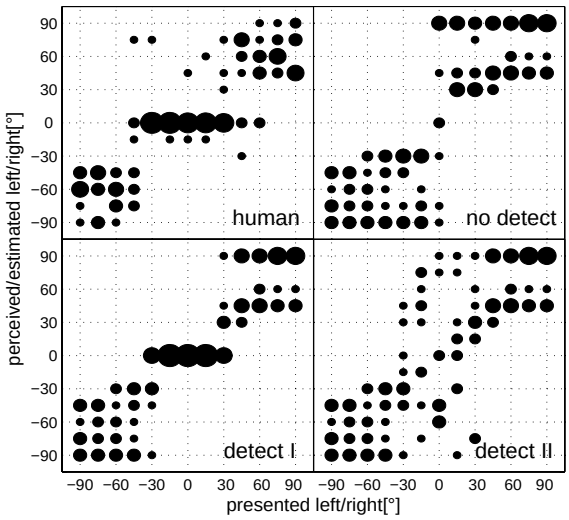

Fig. 5: Localization data.

detection task and –0.12 [MU] for the offset detection task. Below, the median of the estimated on-and offset time values are presented. The error bars show the inter-quartile range. The on- and offset times were estimated at the point where the local maximum magnitude of the input function is measured. The delay of the system was eliminated. At this point, the detection process will be included into the ICCD localization model. It was assumed that the target for the lowest measured T/D-ratio cannot be detected when it is presented from 0° or $|15^\circ|$ azimuth and that it can only be detected in 50% of the cases at an angle of $|30^\circ|$.

The results of the simulation are shown in Fig. 5. The upper left panel shows the psychoacoustic results [6]. The upper right panel displays the results for the localization model without a detection stage. The lower left panel shows the same condition, when the detection stage is integrated. Now, the model indicates the distracter position in those cases where the target can not be detected, forming a third group of clusters and scatterings—which appears for the frontal directions -30° to 30°—as it was found in the localization test [6]. In the investigation of Lorenzi et al. [7] large scatterings in the frontal directions were observed. This can be simulated in the model, if it chooses a direction by random in those cases it cannot detect the target, instead of indicating the direction of the distracter (lower right panel).

Conclusion

The proposed detection model supports the ICCD hypothesis in two ways. Firstly, it could be shown that the on- and offset times can be estimated accurately enough to trigger the subtraction process. Secondly, the integration of a detection algorithm into the ICCD model shows the effects of scattering and clustering as it was found in the localization experiment [6]. Besides supporting the ICCD hypothesis, the detection model is able to simulate masked detection thresholds on a basis of an onset detection algorithm.

Literature

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

[2] J. Braasch **(2001)**: *Ein Lokalisationsmodell zur Simulation des menschlichen Hörens in Mehrschallquellen-Situationen*, in: Fortschr. Akust., DAGA 2001, Hamburg, S. 460–461.

[3] R.D. Patterson and 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, S. 1890–1894.

[4] N.I. Durlach **(1963)**: *Equalization and Cancellation theory of binaural masking-level differences*, J. Acoust. Soc. Am. **(35)**, S. 1206–1218.

[5] H. Levitt **(1995)**: *Transformed up-down methods in psychoacoustics*, J. Acoust. Soc Am. 49, S. 467–477.

[6] J. Braasch und K. Hartung **(1998)**: *Lokalisation von maskierten Breitbandrauschsignalen in reflexionsfreier und reflexionsbehafteter Umgebung*, in: Fortschr. Akust., DAGA 1998, Zürich, S. 500–501.

[7] C. Lorenzi, S. Gatehouse and C. Lever **(1999)**: *Sound localization in noise in normal-hearing listeners*, J. Acoust. Soc Am. 105, S. 1810–1820.