Modeling the auditory signal processing in reverberant environments

Jörg Matthias Buchholz
Institut für Kommunikationsakustik, Bochum; Email: buchholz@ika.rub.de

1. Introduction

Cocktail-party processors have often been applied to separate superposed signals emitted from different sound sources. Although such techniques show promising results in simple scenarios, they usually fail in realistic – especially reverberant – environments. However, the human listener is able to perform this signal separation task, by applying a sophisticated monaural and binaural signal processing. Hence, it is very important to understand and to model these auditory processes, in order to employ them in modern cocktail-party processors. In this regard, the phenomenon of room reflection masking is analyzed and a model (termed Room Reflection Masked Model) describing the main underlying auditory processes is proposed.

2. Model Structure

In general, the masked threshold (MT) is defined as the level of a test signal at which the test signal in the background of a masking signal is audible with a certain probability. In the case of room reflection masking, the masker is considered to correspond to the direct sound or/and additional wall reflections, and the test signal to correspond to a single test reflection. The proposed model aims to simulate the "effective" auditory processing of a listener who takes part in a psychoacoustical masking experiment. Following this idea, the listener is assumed to base his decision whether the test reflection is audible or not, on a comparison of an internal signal description of the masker plus test signal and the masker alone. The main processing stages of the auditory model are shown in Figure 1.

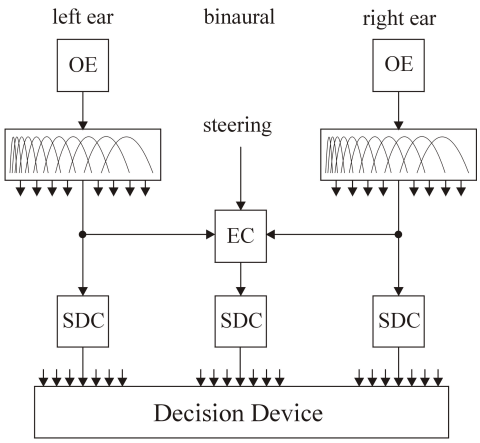

Figure 1: Main signal processing of the masking model.

With respect to Figure 1, three main signal paths can be observed: two monaural signal paths (left and right ear) and a binaural signal path. First, the sound signal is processed by the outer ears (OE), which are sufficiently described by Head-Related Transfer Functions (HRTFs). From a different point of view, the OEs might be considered to be microphones with a certain directivity: omnidirectional at low signal frequencies ($f < 1$kHz) and strongly directional at higher frequencies. After the outer ears, the input signal is decomposed into different frequency channels by utilizing a Gammatone bandpass filterbank. Given that each frequency channel is processed in the same way, only the main signal processing in one frequency channel is further described. At each frequency channel, two monaural and one binaural signal processors are appended. The monaural signal processors are mainly realized by a signal dependent compression (SDC) stage, and the binaural signal processor by an equalization-cancellation (EC) stage followed by a SDC stage. The SDC stage is composed of a static compressive function and an operating point signal which controls the rate of compression dependent on the input signal's evolution. The SDC methodology has already been shown to be very useful in describing various aspects of auditory masking and is described in detail in [1] or [2]. With respect to the EC-stage (see [4]), the two ear signals are first equalized by adjusting their signal level and initial time delay and then, one signal is subtracted from the other. Hence, the EC-stage realizes a variable "null-pointing" antenna, although, the general steering strategy is not known and here, is adjusted by hand to give the best simulation results. The binaural stage is restricted to signal frequencies up to about 1.5 kHz. The outputs of the monaural and binaural signal paths are finally passed into a decision device stage. Inherent in the decision device stage, the different model outputs to the masker signal are subtracted from the corresponding outputs to the masker plus the test signal, weighted by gain factors, added up across frequency, and the result is finally compared to a predefined static threshold. The weighting is realized by simply multiplying the most sensitive channel[1] by one and all others by zero.

3. Model Simulations

The model performance is analyzed by comparing model simulations to corresponding psychoacoustical results on room reflection masking taken from the literature ([3] and [6]). At first, the case in which the test reflection is masked by the direct sound is analyzed and afterwards the case of the test reflection being masked by the direct sound plus an additional reflection is considered.

Direct sound masks test reflection

In the case that the direct sound masks the test reflection, mainly the signal part of the test reflection, which succeeds the direct sound, determines the audibility of the test reflection. Hence, in this simple scenario, mainly forward masking aspects have to be considered. Given that for forward masking situations, the binaural masking level difference (BMLD) is often considered to be zero [5], for the simulations described in the current section, the binaural processing shown in Figure 1 is disregarded. The masked threshold is only determined by the most sensitive monaural processing path, which is here called the "best-ear" approach.

In Figure 2, the MT dependency on the test reflection delay d_t is shown with the direct sound duration D_m being a parameter (stimulus: white noise; direct sound spectrum level: 40dB-SPL). The model simulations (closed symbols) show a good agreement with the psychoacoustics reference data (taken from [3], filled symbols). The MT is decreasing with increasing test reflection delay and furthermore, lower MT values are found for shorter direct sound duration.

[1] A channel is considered to be a frequency band in one of the signal paths (left ear, right ear, binaural).

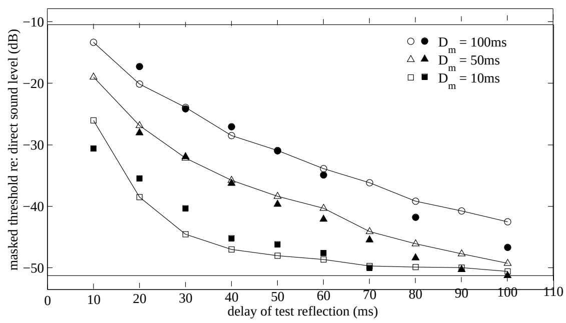

Figure 2: Dependency of the MT on the test reflection delay d_t and the direct sound duration D_m

In Figure 3, the MT dependency on the direct sound level L_m is shown (stimulus: white noise; direct sound duration: 50ms; test reflection delay: 30ms). The model simulations (open symbols) show again a good agreement with the reference data (taken from [3], filled symbols). The MT (note: the MT is shown relative to the direct sound level) decreases with increasing masker level, which implies that a test reflection is better audible for higher direct sound level.

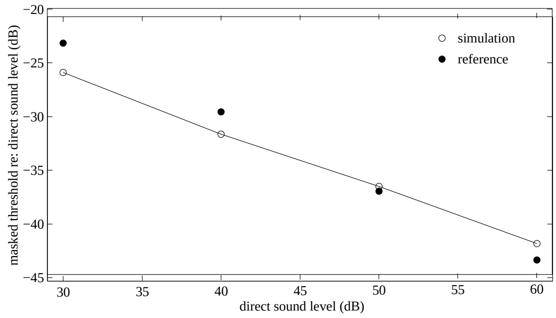

Figure 3: Dependency of the MT on the direct sound level L_m

In Figure 4, the MT dependency on the direction of incidence of the test reflection in the horizontal plane (lateralization) is shown for the case when the direct sound is in front ($\alpha = 0°$) of the listener's head (stimulus: white noise; direct sound spectrum level: 40dB-SPL; direct sound duration: 100ms; test reflection delay: 30ms). The model simulations (open symbols) show a good agreement with the reference data (taken from [3], filled symbols). Given that the monaural "best-ear" approach is sufficient to describe the data in Figure 4, the lateralization dependency of the MT seems to be explained by taking into account only the directivity pattern of the outer ears.

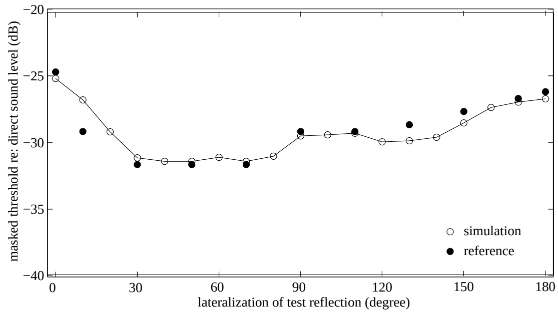

Figure 4: Dependency of the MT on the lateralization of the test reflection. The direct sound is always in front of the listener.

Direct sound and additional reflection mask test reflection

In contrast to the previous signal set-up, in the case that the direct sound plus an additional reflection are masking a test reflection, the test reflection can be completely overlapped in time by the masker. Hence, a strong influence of simultaneous masking effects is expected, which suggest that binaural processes have to be additionally considered (see Figure 1).

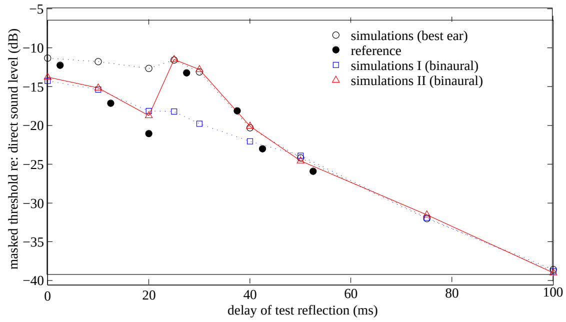

Figure 5: Dependency of the MT on the test reflection delay d_t.

In Figure 5, different model simulations are shown in comparison to the reference data taken from [6] (stimulus: white noise; direct sound duration: 100ms; direct sound spectrum level: 40dB-SPL; delay of add. reflection: 25ms; azimuth of direct sound: 0°, add. reflection: 40°, test reflection: –20°). With respect to Figure 5, the monaural "best-ear" model (open circles) fails to predict the initial decay and the "jump" of the reference MT (filled circles). A cancellation of the additional reflection by the binaural processor (open squares) leads the simulations to show the initial decay but fails to predict the "jump" of the reference MT. Assuming that for short test reflection delays ($d_t < 25$ms) the binaural processor is able to cancel the additional reflection and afterwards is not able to do so, the reference data can be sufficiently predicted (open triangles). However, the general "steering" strategy of the binaural stage has to be further investigated.

Conclusions

A very good agreement between the model simulations and the psychoacoustical reference data could be highlighted, showing that the proposed model structure is able to describe various aspects of room reflection masking.

The work has been accomplished in collaboration with the Institute of Communication Acoustics (Bochum, DE) and the Audio Group (Patras, GR). The work was supported by the European TMR project SPHEAR.

Literature

[1] Buchholz, J. M. (2002). Auditory adaptation effects described by a signal dependent compression, in Fortschr. Akust., DAGA 2002, Deutsche Ges. Akust., Bochum, Germany, pp. 468-469.

[2] Buchholz, J. M. (2003). A computational model of auditory masking based on signal dependent compression. Dissertation, Shaker, Aachen (in press).

[3] Burgtorf, W. (1961). Untersuchung zur Wahrnehmbarkeit verzögerter Schallsignale. Acustica, Vol. 11, pp. 97-111.

[4] Durlach, N. I. (1963). Equalization and cancellation theory of binaural masking-level difference. J. Acoust. Soc. Am., Vol. 35, pp. 1206-1218.

[5] Kohlrausch, A. and Fassel, R. (1997). Binaural masking level difference in nonsimultaneous masking, in: Binaural and Spatial Hearing in Real and Virtual Environments, edited by R. H. Gilkey and T. Anderson, Lawrence Erlbaum Ass., chap. 9, pp. 169-190.

[6] Seraphim, H. P. (1961). Über die Wahrnehmbarkeit mehrerer Rückwürfe von Sprachschall. Acustica, Vol. 11, pp. 80-91.