Modelling binaural processes involved in simultaneous reflection masking: limitations of current binaural models

Jörg M. Buchholz

Centre for Applied Hearing Research, Ørsted DTU, Ørsteds Plads, 2800 Lyngby, Denmark, jb@oersted.dtu.dk

Introduction

Reflection masking (RM) refers to the auditory masking condition, in which a test reflection is masked by the direct sound (and maybe additional reflections). RM has been typically studied for developing or enhancing technical applications, such as virtual auditory environments or sound reproduction in rooms [1]. Throughout this study, the concept of RM is employed to investigate mechanisms that the auditory system may utilize to process reverberant sounds.

Psychoacoustical Experiments

Methods: Reflection masked thresholds (RMT) were measured for a single test reflection, masked by the direct sound, as a function of the reflection delay. The direct sound was a 200-ms long broadband noise (100-5000Hz) presented via headphones at a sound pressure level of 75dB. The test reflection was realized by a delayed and attenuated copy of this direct sound. Two stimulus conditions were considered: (i) a diotic condition, where direct sound and reflection were presented diotically, and (ii) a dichotic condition, where the test reflection contained an interaural time delay (ITD) of 0.5ms. In order to focus on simultaneous masking effects, the reflection offset was truncated such that the direct sound and the test reflection had a common offset.

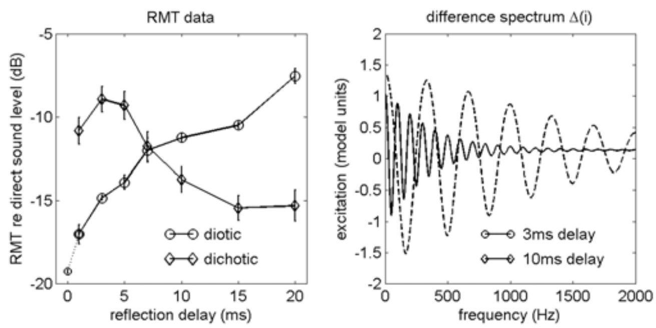

Figure 1: RMT data for diotic and dichotic stimulus presentation and difference-spectrum $\Delta(i)$ of monaural model.

A three-interval, three-alternative forces choice (AFC) procedure was used. All three intervals contained different samples of the direct sound and one randomly chosen interval additionally contained the test reflection. The listener's task was to identify the interval that contained the test reflection. An adaptive three-down one-up procedure was employed to track the 79% correct point on the psychoacoustic function. The starting step size of the test reflection was 4dB, which was reduced to the final step size of 2dB after 6 reversals. Using this final step size, 12 reversals were measured and the mean value and variance over these 12 reversals were calculated. Three well-trained normal-hearing subjects participated in the experiments. At least three measurements were made for each RMT value and subject.

Results: The measured diotic and dichotic RMT data is presented in figure 1. The diotic RMT (circles) increases with increasing reflection delay, while the dichotic RMT (diamonds) decreases with increasing reflection delay. Calculating the difference between both thresholds, i.e., the Binaural Masking Level Difference (BMLD, figure 3), a binaural reflection-suppression effect (negative BMLD) is observed for delays below 7-10ms and a binaural reflection-enhancement effect (positive BMLD) for larger delays.

Masking model framework

In order to identify auditory mechanisms that may be involved in RM, different pre-processing (PP) stages are incorporated into a general masking model framework (figure 2). Following the signal flow through this model framework, the direct sound plus the test reflection $x_{D+T}(n)$ and the direct sound alone $x_D(n)$ are passed into the considered pre-processing stage. The input to this stage can either be monaural or binaural, and the outputs present auditory-internal spectra. The difference between the two output-spectra forms the difference-spectrum $\Delta(i)$. The decision variable $\hat{c}$ is then derived by calculating the RMS value of $\Delta(i)$. The RMT is defined by the test reflection level at which $\hat{c}$ is equal to a pre-defined constant T_c.

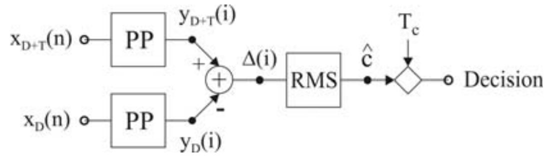

Figure 2: Signal-flow diagram of model framework.

Monaural processing

The monaural pre-processing is realized by a simplified version of the masking model described in [2]. The incoming sound signals are first analyzed by a gammatone filterbank with 51 channels evenly spread over a frequency range of 100-5000Hz (two filters per ERB). The output of each frequency channel is half-wave rectified, low-pass filtered at 1kHz, and amplitude-compressed by a (static) logarithmic function. Then, a non-leaky integration is applied to the entire stimulus interval, forming the auditory-internal output-spectrum. Integrating this pre-processing stage into the proposed model framework (figure 2), a difference-spectrum $\Delta(i)$ is produced, which is exemplarily shown in figure 1 (for a delay of 3ms and 10ms and a reflection gain of -15dB). The difference-spectrum exhibits spectral ripples (or spectral modulations) with a periodicity of $1/\tau$, with τ being the reflection delay. The depth of the spectral ripples decreases with increasing signal-frequency as well as increasing reflection delay, due to the spectral smoothing of the auditory filters. The model predictions of the diotic (or monotic)

RMT are given in figure 3 (diamonds), and are in good general agreement with the according experimental data (circles).

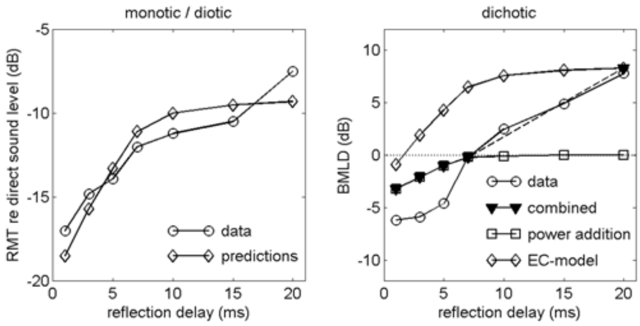

Figure 3: Experimental data and model predictions for the diotic RMT and the BMLD.

Binaural processing

The first binaural model approach is based on the power-addition model given in [3]. The sound signals at the left and right ear are passed through the monaural pre-processing given above. Then the average is calculated of the resulting left and right spectra, forming the output spectrum of the power-addition model. In figure 4, the difference-spectra $\Delta(i)$ for the left ear, the right ear, and the power-addition output are exemplarily shown for a reflection delay of 3ms and a reflection ITD of 0.5ms. Due to the introduced ITD, the monaural spectra exhibit spectral modulations with slightly different modulation frequencies. The average of the monaural spectra results into a reduced output, whereby the amount of reduction depends on the spectral-phase difference of the monaural spectra. Applying this power-addition based pre-processing to the proposed masking model framework, and calculating the difference between the resulting dichotic RMT predictions and the above diotic RMT predictions, the BMLD given in figure 3 (squares) is predicted. For early reflections ($\tau \leq 7$ms), the model predictions are in good qualitative agreement with the according BMLD data (circles). However, the model fails to describe the full amount of binaural suppression observed in the data. Moreover, the model can also not account for the binaural enhancement observed for reflection delays larger than 7ms.

The second binaural model approach is a simplified version of the Equalization Cancellation (EC) based model given in [4]. The model is realized by: (i) passing the left and right ear signals through a gammatone filterbank and (ii) applying half-wave rectification, lowpass filtering at 770Hz, and (static) logarithmic compression to each frequency channel output. Given that the direct sound (i.e., the masker) was always presented diotically, the EC-stage was here simply realized by: (i) calculating the difference between the left and right ear signals for each frequency channel and (ii) calculating the signal power of the EC output over the entire stimulus interval. The subtraction process within the EC operation thereby results in a maximum suppression of the direct sound and thus in an enhancement of the test reflection. The output of the EC-based difference spectrum is exemplarily shown in figure 4 for a delay of 3ms and 20ms and a reflection gain of -10dB. Applying the EC-based pre-processing to the above masking model framework, a dichotic RMT is predicted that is independent of the reflection delay. The difference between the EC-based dichotic RMT predictions and the above diotic RMT predictions -the BMLD- is given in figure 3 (diamonds). The EC-based model predictions can only describe the BMLD data (circles) for a delay of 20 ms (i.e., the point for which the model was optimized) and thus fails to describe the BMLD data.

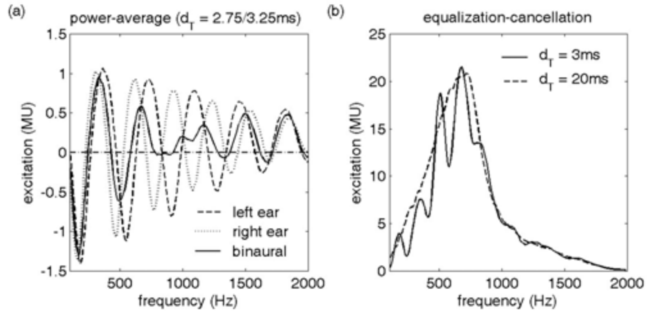

Figure 4: Difference-spectrum ***Δ(i)*** for the power-addition-based model and the EC-based model.

Conclusions

Considering the above BMLD predictions, it might be suggested that, dependent on the reflection delay, the auditory system applies different strategies: for very early reflections, the auditory system might apply a mechanism similar to the power-addition process, and a mechanism similar to the EC-process for late reflections (figure 3, dashed line and triangles). However, it is unclear: (i) how these two mechanisms can be combined and (ii) how the full size of the observed binaural suppression effect can be produced.

The diotic RMT at a delay of $\tau = 0$ ms (figure 1) refers to a JND in level of about 0.85dB, measured with two perfectly correlated noises. The diotic RMT for a 20-ms delayed reflection refers to the same JND, but measured for two uncorrelated noises. Hence, it seems that the monaural auditory system processes the direct sound and the test reflection as two correlated signals for short delays, and as two uncorrelated signals for larger delays. Applying this idea to the BMLD data, it might be similarly assumed that binaural reflection-suppression refers to the binaural processing of two correlated signals, and binaural reflection-enhancement to the binaural processing of two uncorrelated signals.

Literature

[1] Olive, S. E. and Toole, F. E.: The detection of reflections in typical rooms. J. Audio Eng. Soc. 37 (1989), 539-553.

[2] Buchholz, J., M. and Mourjopoulos, J.: A Computational Auditory Masking Model based on Signal Dependent Compression. I. Model Description and Performance analysis. ACUSTICA/acta acustica 90 (2004), 873-886.

[3] Zurek, P. M.: Measurements of binaural echo suppression. J. Acoust. Soc. Am. 66 (1979), 1750-1757.

[4] Breebaart, J., van de Par, S., and Kohlrausch, A.: Binaural processing model based on contralateral inhibition. I. Model structure. J. Acoust. Soc. Am. 110 (2001), 1074-1088.