Reducing the temporal resolution of spatial impulse responses with an auditory model

Max Röhrbein[1], Alexander Lindau[1]
[1] *TU Berlin, Audio Communication Group, E-Mail: alexander.lindau@tu-berlin.de*

Motivation

The sound field of a source in an enclosed space (a pattern of reflections) can be approximated by a series of plane waves arriving at a receiving point at specific instances in time, from individual directions of incidence, and with specific level and spectral content. Such a 'reflectogram' or room impulse response (RIR) can be used to re-render the sound field by means of spatial sound field synthesis using loudspeaker array based sound reproduction systems. Real time calculation of driving functions for the representation of fully detailed sound fields might become computationally intensive while being unnecessary from a perceptual point of view. Therefore, the aim of this study was to develop and perceptually evaluate an algorithm which reduces reflectograms through modelling temporal, spatial and spectral masking mechanisms of the human auditory system.

Room reflection masking

Masking of reflections has been studied using simple artificial sound fields, typically comprising a direct sound (masker) and a test reflection. An audibility threshold for singular test reflections is typically given as a damping value (in dB) relative to the direct sound level and is called the reflection masking threshold (RMT). RMTs of test reflections have been assessed for instance as a function of delay time, the direct sound level, the spectral difference between direct sound and test reflection, for different directions of incidence of masker and test reflection, under the influence of additional reverberation, and for different audio stimuli (e.g. in [1]). See [2] for a summary of results. In [1] it was shown that reflection masking is especially critical in case of click-like signals. Several approaches to reduce the complexity of sound fields have been proposed. In [2] an auditory model was developed in order to predict reflection masking thresholds in rooms. The model describes the relevant human auditory signal processing and decides about audibility of reflections in the presence of a masker. Following the approach in [2] our aim was to develop an auditory process model of postmasking occurring with impulsive, click-like signals.

Reflection masking model

Since empirical data is available for the masking of pulse pairs (e.g. [1]), we first developed a process model of the masking of two impulses. This model was calibrated using the empirical data from [1] and then extended into an iterative process (cf. Figure 1) in order to be applicable to maskers consisting of impulse patterns as occurring within real RIRs. The model implements the whole auditory pathway of outer and inner ear, the mechano-neural interface of the hair cells, the neural processing and decision stages.

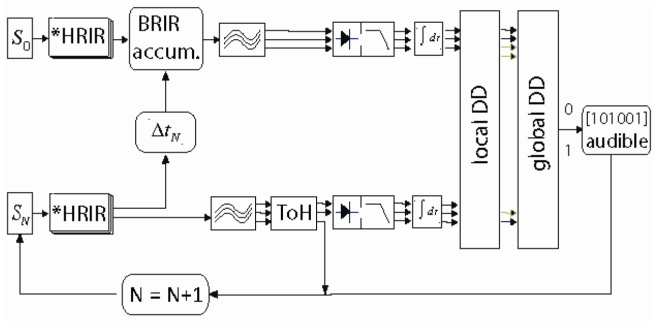

Figure 1: Block diagram of the iterative auditory process model of reflection masking in RIRs (S_0 = direct sound, S_N = current test reflection), see text for further explanation

The frequency dependent directivity of the outer ears is modelled using a data set of head related impulse responses. Frequency analysis as performed by the cochlea is modelled using a 26 band gamma tone auditory filter bank [3]. Each test reflection is compared to the threshold of hearing to assure audibility in at least one auditory band. Therefore, the RMS level of each reflection has been calibrated relatively to a pre-defined direct sound RMS level, which is defined to correspond to a certain absolute sound pressure level (e.g. 80 dB_{SPL}). The inner hair cells are simulated using half-wave rectification and 1 kHz low pass filtering [4]. The temporal resolution of the neural transmission is simulated by averaging over sliding rectangular windows of individual lengths per band [5]. The audibility of a reflection is decided per band in the local decision devices (loc. DD). Here, the amplitude of the auditory signal of the masker is compared to that of the test reflection alone at the instant of the maximum auditory signal amplitude of the test reflection. This approach leads to a simple implementation of postmasking without using adaptive signal processing stages as e.g. in [2]. Until this point, processing was conducted independently for both ears. Now, the global decision device (glob. DD) collects the individual decisions of both ears' local decision devices, and decides on audibility of the test reflection if a certain percentage of positive local decisions is found. The whole process can be repeated iteratively for the complete reflectogram: each reflection is successively tested for audibility against a masker comprising all preceding sound field components. Hence, it is inherently assumed that masking found for pulse pairs can simply be transferred to masking of pulses by pulse patterns.

Calibration of the reflection masking model

As explained above, the process model has two parameters: the threshold used in the local DDs and the percentage of positive local audibility decisions used in the global DD. These two parameters were – in a worst case approach – adjusted, such that the model will react at least as sensitive as according to data from [1] (cf. Figure 2).

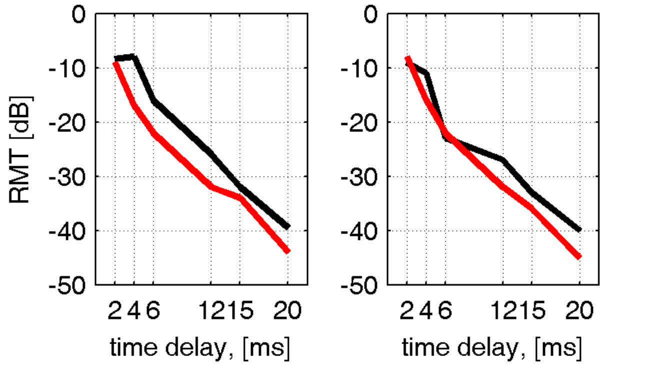

Figure 2: RMT of model (red) compared to empirical RMT data ([1], black, one subject). Left: RMT for a masker from 0°/0° (azimuth/elevation) and test reflection from 65°/0°; right: RMT for a masker from 0°/0° and test reflection from 0°/60°.

Perceptual evaluation

A listening test was conducted in order to perceptually evaluate the applicability of the model for reducing the amount of early reflections in a RIR. Therefore, a reflectogram was calculated for a single source on the stage at a central listening position in a virtual lecture hall (V: 8500 m³, RT: 2 s) using commercial room acoustic CAD software (EASE 4.3, "Aura Response" module). Since individual reflections are not distinguishable after the so-called perceptual mixing time t_{mp} [6], the auditory reduction was applied to the early reflections only (first 300 ms of the reflectogram, ≈ $t_{mp95\%}$ cf. [6]). The calibrated auditory model found only 3 reflections (the strongest side wall reflection and the two strongest ceiling reflections) to be audible. In a multiple stimulus ABC/HR-test [7] the task of the subjects was – while comparing to a hidden reference sound field – to detect and rate (from 'identical' to 'very different') the similarity of the reference itself and 6 sound field simulations of successively decreasing spatial detail. The fully detailed reference sound field contained 263 early reflections. Additional test sound fields with 76, 19, 10, 4, and 3 (calibrated model state) reflections were generated from the original sound field by continuously changing the model's sensitivity. The direct sound alone was also included in the test sound fields as an additional low-quality anchor. From the seven reflectograms, datasets of binaural room impulse responses (BRIRs) for horizontal head orientations (±80°, 1° step size) were calculated, allowing for a dynamic auralization of the sound fields accounting for head movements. A static dual-channel diffuse reverberation tail, modelled from uncorrelated white Gaussian noise, weighted with decay rates according to the enclosure's original third-octave reverberation times and a diffuse field HRIR was – while establishing the original D/R-ratio – always concatenated to the early reflections using a linear fade-in over the first 100 ms. BRIR data sets were loudness compensated. Using a time variant fast convolution algorithm, BRIRs can be convolved with anechoic audio stimuli in real time. We used three different audio stimuli in the test: a train of dirac pulses, a piece of male speech and an excerpt from a classical piece for string quartet. All subjects assessed all 7 x 3 = 21 test stimuli in a highly sensitive repeated measures design.

Results and discussion

29 subjects (83 % male, ø 31 yrs., ø 12 yrs. of musical education, 76% w. prior listening test experience) took part in the test. Figure 3 shows the results pooled over subjects in terms of difference grades and 95% confidence intervals. Negative difference grades indicate that the manipulated sound field has been properly detected and rated as different from the reference sound field. A difference grade of -4 indicates the bottom end of the scale ('very different').

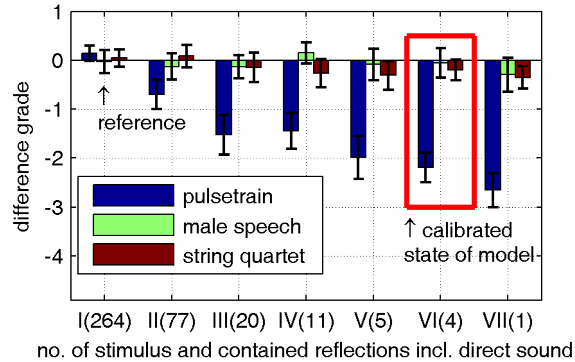

Figure 3: Results from ABC/HR-listening test

In compliance with the prediction of our model, for the natural stimuli speech and music, the reduced sound fields were nearly never distinguishable from the 263-reflection reference sound field. Confidence intervals always include the '0'-difference line, despite in the case of the musical stimulus and the simplest sound field (direct sound plus reverb tail only). Results for the pulse train are very different: None of the stimuli was confused with the reference. Instead, an order of similarity decreasing with the number of reflections was reproduced unexpectedly well. In a questionnaire subjects mentioned difference aspects as coloration, spatial impression, localization and loudness. Thus, for critical stimuli our simplifying approach to transfer the masking behaviour of pulse pairs to that of more complex pulse patterns is not tenable.

Conclusion

An algorithm for the perceptual reduction of RIRs has been presented. It detects strongest and spatially distributed reflections in an apparently plausible manner. Listening test results showed the validity of the prediction to be stimulus dependent. For natural stimuli, the listening test confirms a high potential for perceptual reduction in room sound fields. Results are in agreement with [6], where it was indirectly shown that sound field components arriving after the 1st and 2nd order reflections are hardly detectable.

References

[1] Olive, S.E. et al. (1989): "The Detection of Reflections in Typical Rooms". In: *J. Audio Eng. Soc.*, **37**(7/8), 539-553

[2] Buchholz, J.M. et al. (2001): "Room Masking: Understanding and Modelling the Masking of Room Reflections", In: *Proc. of the 110th AES Conv.*, preprint no. 5312

[3] Slaney, M.(1993): "An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank". Apple Computer Technical Report #35. Online available.

[4] Dau, T; Püschel, D.; Kohlrausch, A. (1996): "A quantitative model of the "effective" signal processing in the auditory system." In: *J. Acoust. Soc. Am.*, **99**(6), pp. 3615-3622

[5] Hulsebos, E. et al. (2002): "Parameterization and reproduction of concert hall acoustics measured with circular microphone array", *Proc. of the 112th AES Conv.*, preprint no. 5579

[6] Lindau, A.; Kosanke, L.; Weinzierl, S. (2010): "Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses." In: *Proc. of the 128th AES Conv.*, preprint no. 8089

[7] ITU (1997): ITU-R Rec. BS.1116-1: *Methods for the subjective Assessment of small Impairments in Audio Systems including Multichannel Sound Systems.* Geneva: International Telecommunication Union