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

Max Röhrbein¹, Alexander Lindau¹

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.

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).
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 \( t_{\text{mp}} \) [6], the auditory reduction was applied to the early reflections only (first 300 ms of the reflectogram, \( \approx t_{\text{mp}} \) [6], 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 versions. Six subjects assessed all 7 × 3 = 21 test stimuli in a highly sensitive listening test: a train of Dirac pulses, a piece of male speech and an excerpt from a classical piece for string quartet. All subjects were instructed to pay attention to the ‘0’-difference line, despite the presence of musical coloration, spatial impression, localization and loudness. Thus, for critical stimuli our simplifying approach to transfer the mixing 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