Acoustics'08 Paris
June 29-July 4, 2008
www.acoustics08-paris.org

Determining acoustical parameters using cochlear modeling and auditory masking

Jasper Van Dorp Schuitman and Diemer De Vries

University of Technology Delft, Lorentzweg 1, 2628 CJ Delft, Netherlands
j.vandorpschuitman@tudelft.nl

The acoustical qualities of a concert hall or any other room are generally expressed using acoustical parameters determined from impulse responses. From microphone array measurements it turned out that these parameters can fluctuate severely over small distances, whereas the perceptual cues for which these parameters are supposed to be a measure remain constant. This means that a local parameter value has a very low predictive value for acoustic quality. In this research, cochlear modelling techniques and simulations of auditory masking effects have been applied to model human hearing. These techniques together model various stages in the auditory path, like the movement of the basilar membrane inside the cochlea and mechanisms inside the brains. It turns out that determining acoustical parameters using this representation leads to results which show much less spatial fluctuations, and are closer to human perception.

1 Introduction

The spatial fluctuations observed when determining acoustical parameters from impulse responses is a common problem when assessing the acoustical qualities of a room. The spatial fluctuations are most noticeable when the parameters are determined from array measurements (closely spaced microphone positions) [1], but also measurements at multiple positions within one seat in a concert hall revealed huge differences [2]. Fig. 1 shows an example of fluctuations in the parameter *lateral fraction* as a function of offset. The results are obtained in the Concertgebouw, along a line array.

These fluctuations do not match with human perception; subjects do not report perceivable differences in acoustical measures like reverberation time and clarity index on closely spaced positions. This means that results for single microphone positions have a very low predicitive value for acoustical quality. The ISO standard therefore proposes a method where results for multiple positions are averaged [3]. In this research a more novel method - based on psychoacoustic principles - is introduced.

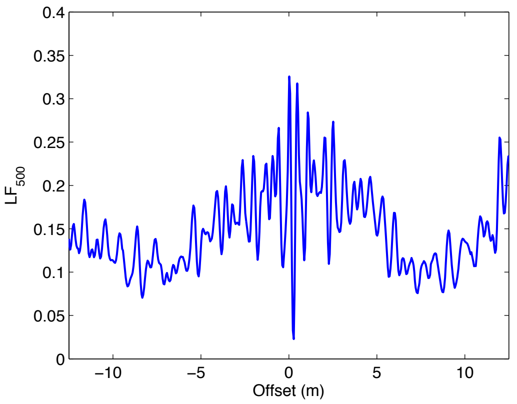

Figure 1: The lateral fraction LF in the 500 Hz octave band as a function of offset. The results are obtained from an array measurement carried out in the Concertgebouw, Amsterdam (NL).

2 The origin of the fluctuations

In [1] it was shown that parameters for spaciousness, like the lateral fraction, fluctuate as a result of wave interferences. These interferences lead to local minima and maxima. Since also monaural parameters like the clarity index and reverberation time suffer from unperceivable fluctuations, we can claim the hypothesis that measured impulse responses have features which influence the results for the determined acoustical parameters while they are not important for human perception.

3 A psychoacoustic model

From psychoacoustics it is known that audio signals may contain components which are not perceivable, because such a component (maskee) is being masked by another component (masker). This masking is a result of mechanisms inside the human auditory system. It is a non-linear effect, which is dependent on various parameters like the type of signal, the length of the signal, the time difference between masker and maskee and the difference in spectral content [4, 5, 6].

Audio coding schemes like MPEG [7] make use of the knowledge on these mechanisms by reducing the amount of bits with which audio is stored, while keeping the resulting quantization noise under the *masking curve*. The masking curve is constructed for each block of audio, using the energy spectrum, the absolute threshold of hearing, the 'tonality' of the signal, spreading of energy over frequencies and results for previous blocks [6].

Based on this MPEG scheme, an algorithm was developed which constructs masking curves for measured impulse responses. The algorithm is shown schematically in Fig. 2.

As is shown in Fig. 2, the masking curve is calculated for each small block of the impulse response. In the last step of the algorithm, all energy below the masking curve is removed, leading to 'transformed impulse responses'. This step is not trivial, since the masking curve is constructed using all the energy, including the energy that is about to be removed. However, from informal listening tests it turns out that there is no perceivable difference between signals which are convolved with the original or with the transformed responses.

4 Modified model

A second algorithm was developed, which is a slightly modified version of the MPEG based algorithm of Fig. 2. It included *cochlear modeling*, where the movement of the basilar membrane (BM) inside the human cochlea is simulated. The movement of this membrane can be described by a one-dimensional, linear model as proposed by Duifhuis [8]. The numerical simulation is described by Diependaal et al. [9].

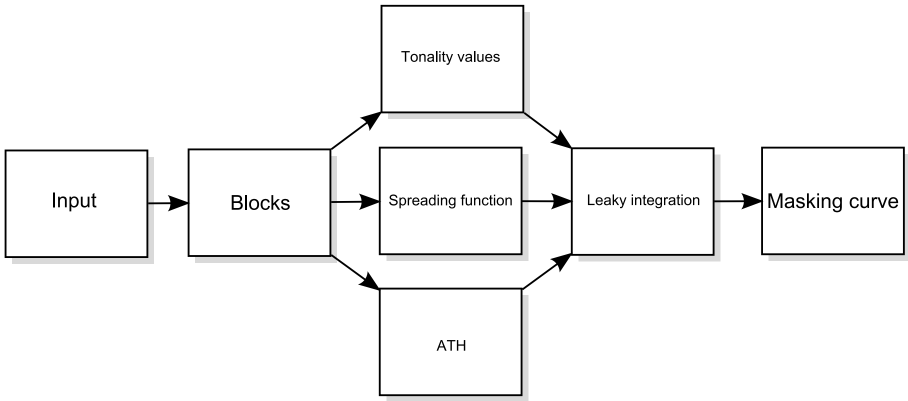

Figure 2: A schematic view of the algorithm which is used to construct the masking curve.

When cochlear modeling is applied on an audio signal, the result will be a so-called "cochleogram"; a plot describing the movement of the basilar membrane as a function of time and frequency [10]. An example of a cochleogram generated for a measured impulse response is shown in Fig. 3.

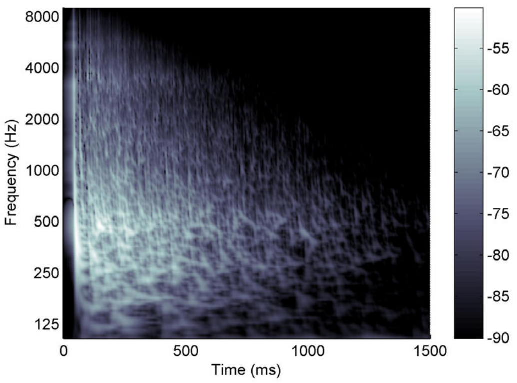

Figure 3: An example of an impulse response, transformed to the cochlear domain. In this cochleogram the velocity of the basilar membrane is plotted in dBs, as function of time and frequency.

The modified algorithm which includes cochlear modeling is shown in Fig. 4. The BM model replaces the estimation of the tonality of the signal and the application of spreading functions which model the width of the auditory filters.

5 A measure for spatial fluctuations

The goal of developing these algorithms, was to apply them to measured impulse responses. If an algorithm removes unwanted components, the resulting acoustical parameters might suffer less from spatial fluctuations compared with the original results. In order to test this we need a novel way of specifying the amount of spatial fluctuation in a parameter. In this research it is chosen to use the *mean absolute derivative*, defined as:

$$\delta_\alpha = \frac{1}{N-1} \sum_{n=0}^{N-2} \frac{|\alpha(n) - \alpha(n+1)|}{\Delta x}, \qquad (1)$$

where N is the number of microphone positions, Δx the distance between the microphones and $\alpha(n)$ the value of parameter α at position n. Note that δ expresses fluctuations in "units per meter". This measure is chosen instead of the more common standard deviation, because δ is less sensitive to general trends in the parameter. Here we are only interested in fluctuations over small intervals.

6 Algorithm results

The two algorithms of Fig. 2 and 4 were tested on a set of impulse responses, measured along a linear microphone array in the Concertgebouw in Amsterdam, The Netherlands. The acoustical parameters reverberation time RT, early decay time EDT and clarity index C_{80} were determined from the impulse responses both with and without simulation of auditory masking.

The results are shown graphically in Fig. 5, and numerically in Table 1.

Table 1: δ values for the parameters RT, EDT and C_{80} (125 - 2000 Hz band), using impulse responses from the Concertgebouw measurement. The results are given without auditory masking simulation (None), with normal auditory masking (Masking) and the new algorithm including a basilar membrane model (Masking+BM).

Method	δ_{RT} (s/m)	δ_{EDT} (s/m)	δ_{C80} (dB/m)
None	10.33	1.56	13.47
Masking	0.90	0.36	2.63
Masking+BM	1.08	0.51	6.23

From the graphs and the table it can be seen that the amount of spatial fluctuations is highly reduced when auditory masking simulation is applied. Also, the "normal" auditory masking simulation gives better results compared with the algorithm where cochlear modeling is included.

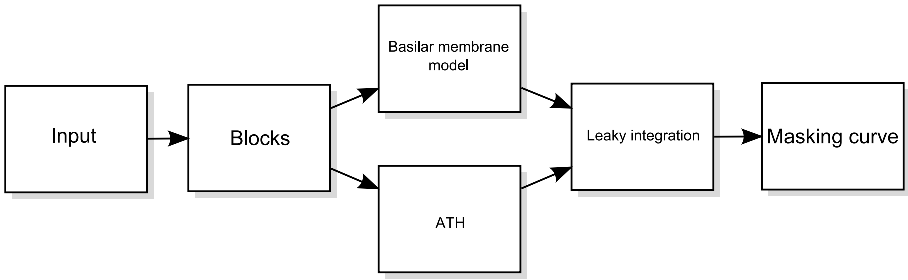

Figure 4: A schematic view of the modified auditory masking algorithm, which includes cochlear modeling.

7 Simulated impulse responses

Based on the results from the previous section, the conclusion can be drawn that auditory masking simulation algorithms indeed are capable of removing components from measured impulse responses which cause spatial fluctuations, but which are not important for perception. To further investigate the results, auditory masking simulation was also applied on artificial impulse responses. The responses were generated by simulating a shoebox-shaped room with the same dimensions and wall absorption coefficients as the Concertgebouw (information was taken from [11]). A "virtual" array measurement was carried out over an offset range of 2.5 m and with a microphone spacing of $\Delta x = 5$ cm. The reverberation time RT was determined with and without auditory masking simulation (MPEG based). The results are shown in Fig. 6.

Remarkably, the originally simulated impulse response already show very little spatial fluctuation. Also applying auditory masking simulation does not improve the results (for both graphs $\delta_{T60} = 0.91$ s/m). This indicates that the cause for the fluctuations probably mostly has to do with measurement noise, and not with wave interferences or masked reflections.

To investigate this further the test was repeated with the simulated impulse responses, but this time the presence of measurement noise is simulated by adding pink noise prior to determining RT. The level of the added noise is such that the resulting signal-to-noise ratio SNR is ~ 35 dB. The results of this test are shown in Fig. 7.

It can be seen in the figure that large spatial fluctuations are introduced. The δ value when no auditory masking simulation is applied is equal to $\delta_{RT,none} = 7.29$ s/m. Applying auditory masking lowers this value considerably: $\delta_{RT,masking} = 1.37$ s/m. The δ values for the other parameters are shown in Table 2.

Table 2: δ values for the parameters RT, EDT and C_{80} (125 - 2000 Hz band) for the simulated responses + pink noise. The results are given with and without auditory masking simulation.

Method	δ_{RT} (s/m)	δ_{EDT} (s/m)	δ_{C80} (dB/m)
None	7.29	2.36	30.33
Masking	1.37	1.77	19.92

In [1] it was shown that *wave interferences* are the main cause for spatial fluctuation in measures for spaciousness. From the results in this paper we can draw the conclusion that *measurement noise* can have a major effect on monaural parameters. Apparently the auditory masking simulator works well as a perceptually based noise remover for impulse responses.

8 Conclusions

In this paper it was shown that large spatial fluctuations in acoustical parameters as determined from impulse responses measured on closely spaced microphone positions, can be reduced by applying an algorithm on the measured impulse responses, which is based on auditory masking simulation. Especially when the original responses have a low SNR, large improvements can be observed. Tests with simulated impulse responses led to the conclusion that measurement noise can have a major effect on monaural parameters.

It was also tried to replace parts of this algorithm with a linear, one-dimensional model of the basilar membrane inside the human cochlea, in an attempt to reach results which are even more close to perception. However, although the results from this new algorithm look promising, it does not perform as good as the old algorithm.

9 Future work

In the future it will be investigated if the auditory masking simulation algorithm which includes cochlear modeling can be further improved, to yield results which are better than for the normal, MPEG based algorithm. Furthermore, it should be tested how the algorithm works for measures for spaciousness (like LF and $IACC$).

Besides more research on the theoretical aspects of the acoustical parameters, the results will be linked to perceptive aspects using formal blind listening tests. Subjects will be asked to rate subjective aspects like reverberance and spaciousness while listening to synthesized (binaural) impulse responses convolved with real-life audio signals. The results of this test will be used to couple objective measures to human perception.

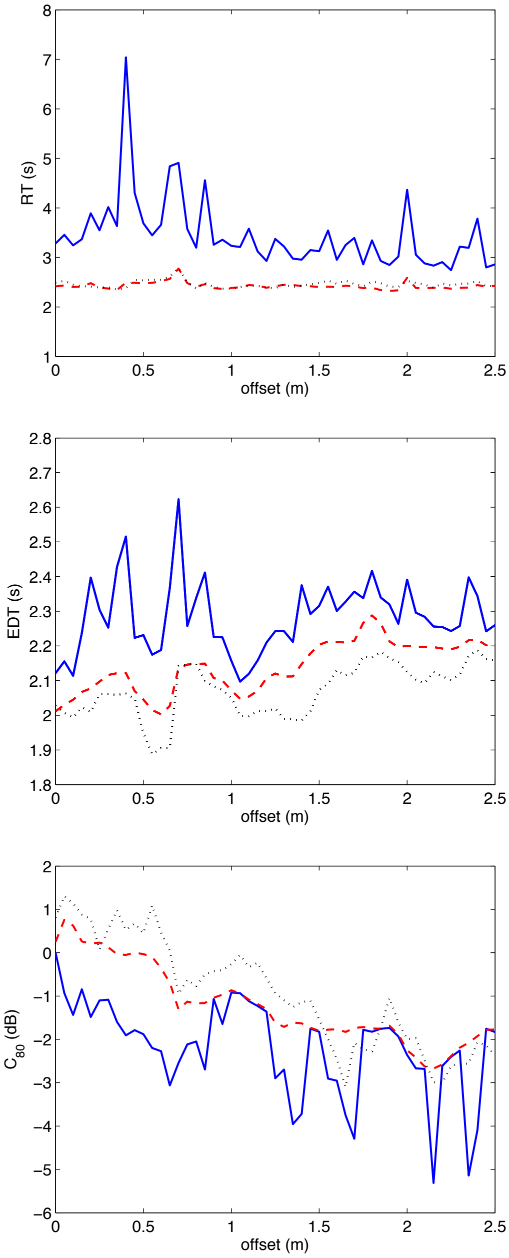

Figure 5: The reverberation time RT, early decay time EDT and clarity index C_{80} in the 125 - 2000 Hz band. The impulse responses are from a measurement carried out in the Concertgebouw. Results are obtained without auditory masking simulation (solid blue line), with normal auditory masking simulation (dashed red line) and with the new algorithm including a basilar membrane model (dotted black line).

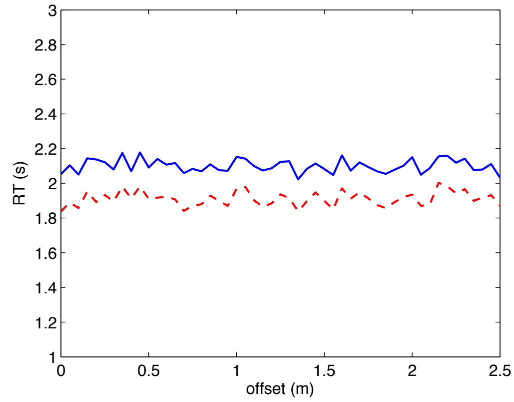

Figure 6: The reverberation time as determined from simulated impulse responses, without (solid blue line) and with (dashed red line) auditory masking simulation.

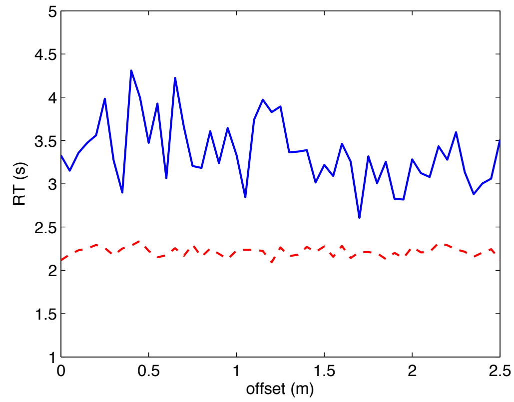

Figure 7: The reverberation time as determined from simulated impulse responses + pink noise, without (solid blue line) and with (dashed red line) auditory masking simulation.

References

[1] D. de Vries, E.M. Hulsebos, and J. Baan. Spatial fluctuations in measures for spaciousness. *J. Acoust. Soc. Am.*, 110(2):947 – 954, 2001.

[2] J.L. Nielsen, M.M. Halstead, and A.H. Marshall. On spatial validity of room acoustic measures. In *Proceedings of the 15th ICA Seattle*, pages 2141 – 2142, 1998.

[3] ISO 3382:1997. Acoustics - measurements of the reverberation times of rooms with reference to other parameters. International Organisation for standardization, Geneva, Switzerland, 1997.

[4] B.C.J. Moore. Characterization of simultaneous, forward an backward masking. In *Proceedings of the AES 12th International Conference*, pages 22 – 33, 2003.

[5] W. Jesteadt, S.P. Bacon, and J.R. Lehman. Forward masking as a function of frequency, masker level and signal delay. *J. Acoust. Soc. Am.*, 71(4):950 – 962, 1982.

[6] B. Lincoln. An experimental high fidelity perceptual audio coder. University of Stanford, 1998.

[7] ISO/IEC 11172:1993. Coding of moving pictures and associated audio for digital storage media up to 1.5 mbit/s. International Organisation for standardization, Geneva, Switzerland, 1993.

[8] H. Duifhuis, H.W. Hoogstraten, S.M. van Netten, R.J. Diependaal, and W. Bialek. Modelling the cochlear partition with coupled van der pol oscillators. In J.B. Allen, J.L. Hall, A. Hubbard, S.T. Neely, and A. Tubis, editors, *Periphal Auditory*

Mechanisms, pages 290 – 297. Springer, New York, 1986.

[9] R.J. Diependaal, H. Duifhuis, H.W. Hoogstraten, and M.A. Viergever. Numerical methods for solving one-dimensional cochlear models in the time domain. *J. Acoust. Soc. Am.*, 82(5):1655 – 1666, november 1987.

[10] T. Andringa. *Continuity preserving signal processing*. PhD thesis, Rijksuniversiteit Groningen, 2002.

[11] L. Beranek. *Concert and opera halls: how they sound*. Acoustical Society of America, 1996.