Deriving Room Acoustical Parameters Using Arrays and Hearing Models

Jasper van Dorp Schuitman[1], Diemer de Vries[2]

[1] *Delft University of Technology, The Netherlands, Email: j.vandorpschuitman@tudelft.nl*

[2] *Delft University of Technology, The Netherlands, Email: d.devries@tudelft.nl*

Introduction

Acousticians use acoustical parameters to express the acoustical qualities of concert halls, or any other room. These parameters are generally determined from impulse responses of the room, either measured or simulated. Common examples are the reverberation time RT, the clarity index C_{80} and the interaural cross-correlation $IACC$.

In order to be able to compare different rooms, or to make valid predictions when a room is still in the design phase, these parameters should of course be accurate, and more importantly, have a high correlation with the subjective attributes which they should express. However, De Vries *et al.* showed in 2001 using measurements along closely spaced microphone positions (array measurements) that parameters for spaciousness may fluctuate severely as a function of measurement position over small intervals, whereas the perceptual cues remain constant [1]. This shows that these parameters are not reliable for assessing room acoustical qualities. Also other authors reported these unwanted spatial deviations, even when measuring within one seat [2].

To overcome this problem the ISO 3382 standard for room acoustical measurements [3] proposes to average the results over multiple microphone positions. In this paper a more novel method is proposed, where parameters are determined using a model of the human auditory system.

The auditory model

In order to derive room acoustical parameters based on the human auditory system, a model is needed which accurately models various psychoacoustic effects. Such a model is developed by Breebaart [4]. This model is basically a binaural version of the model originally proposed by Dau *et al.* [5, 6]. A slightly modified version of the binaural model is used in this research. A schematic version of the complete model is shown in Figure 1. The various stages are:

1. **Preprocessing**: In the preprocessing stage the outer- and inner ear are modeled using a bandpass filter, a Gammatone filterbank, half-wave rectification, lowpass filtering and thresholding respectively. The output of this stage is a time-frequency representation of the input signal.
2. **Adaptation**: This is a nonlinear part of the model, containing five adaptation loops with various time constants simulating the neural firing.
3. **Binaural processor**: Using an approach called the *Equalization-Cancellation (EC) theory*, the binaural processor estimates the Interaural Time Difference ITD and Interaural Level Difference ILD per frequency band. For more information the reader is referred to [4].
4. **Central processor**: Combines the monaural outputs Ψ_L and Ψ_R together with the results of the binaural processor.

In a masking experiment (where subjects have to detect tones in noise, for example), the model described above can be used as an artificial listener. In that case the central processor compares the internal representations of the masker alone with the masker + signal in an optimal way (*optimal detector*). The model is capable of simulating various monaural and binaural masking effects [6, 4] as well as localization in audio reproduction [7]. It is also successfully applied in audio coding [8].

Room acoustical parameters

In this project the model will be used to evaluate acoustics in an auditorily motivated way: using the monaural outputs Ψ_L and Ψ_R various monaural effects can be simulated, like simultaneous and forward masking, and the apparent ITD and ILD results of the binaural processor can be used to estimate parameters related to spaciousness. Because the model is able to describe various psychoacoustic effects, it will probably result in parameters which are closer to human perception compared with the conventional parameters.

One of the main advantages of this approach is the fact that the model works on real-life audio signals rather than impulse responses. Various authors noted that the perception of room acoustics depends on the source signal. This includes sound level, frequency content, temporal features (for example transients), etc. [9, 10, 11].

An example output of the model is shown in Figure 2. In this case a (male) speech signal was convolved with two different impulse responses, one with a short reverberation time ($RT = 0.39$ s) and one with a longer reverberation time ($RT = 1.98$ s) and used as input for the model. Both the monaural output of one of the channels and the output of the binaural processor (ITD and ILD) are shown.

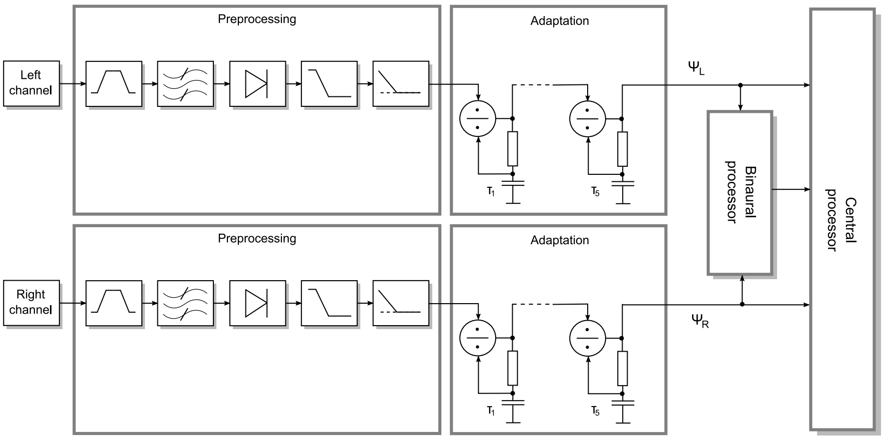

Figure 1: A block diagram of the complete binaural model.

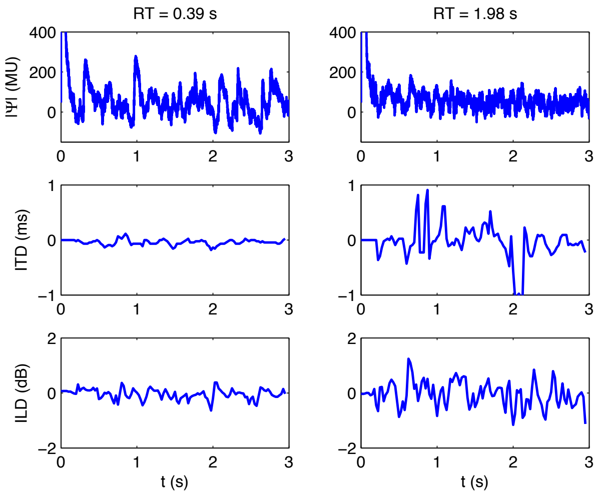

Figure 2: Example output of the binaural model. A speech signal was used, convolved with simulated binaural room impulse responses (BRIRs). The BRIRs were simulated for two virtual halls with reverberation times $RT = 0.39$ s (left) and $RT = 1.98$ s (right). The top figure shows a monaural output Ψ_L in Model Units (MU), the middle figure the interaural time difference ITD and the bottom figure the interaural level difference ILD. All plots are for the 265Hz band.

First, it can be seen from the figure that the more reverberant signal results in a lower output. Furthermore the separate phonemes of the speech signal are less prominent compared with the output for the less reverberant signal. These effects are a result of the nonlinear behavior of the model; because of the adaptation loops the model is less sensitive when the input becomes more stationary [5] as is the case in the more reverberant case. This is exactly as expected, since it is known that generally the ability to here separate components in a signal (*clarity*) is lower in a more reverberant room.

To estimate the clarity as well the level of the reverberant field, an algorithm was developed within this project which separates the monaural model output into two streams: one for the direct sound and one for the reverberant field. An example result of this algorithm is shown in Figure 3.

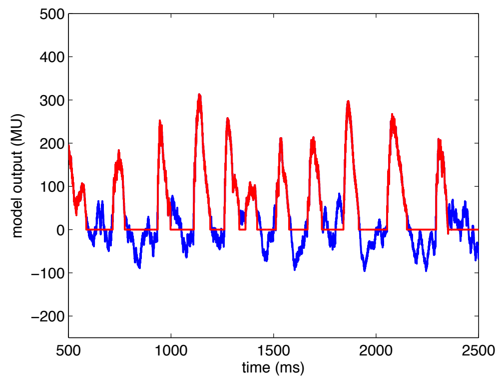

Figure 3: Separation of the model output in direct sound (red) and reverberant sound (blue) streams.

Another conclusion which can be drawn from Figure 2 is that ITD and ILD fluctuate more as a function of time when a room is more reverberant. This will lead to a broadening of the source and a subjective impression of envelopment, as found by Blauert and Lindemann [12, 13]. Authors who recently published research on this subject include Mason [10], Hess [14] and Rumsey *et al.* [15].

Based on the above the following parameters will be extracted from the model output:

- The (average) level of the reverberant stream L_R. This should be related to the impression of *reverberance.*
- The ratio between the levels of the reverberant and direct sound stream $R_{R,D}$, which should be related to *clarity.*
- The standard deviation of the interaural time difference σ_{ITD}, as a measure for the Apparent Source Width ASW.
- The $\sigma_{ITD,R}$ for the reverberant stream, as a measure for the impression of listener envelopment (LEV).

Listening tests

To investigate if the proposed parameters indeed are measures which correlate with subjective attributes, listening tests were conducted. In these tests, five subjects with advanced knowledge on room acoustics participated. The subjects were asked to rate four different subjective attributes on a contineous scale (*very low* to *very high*) on sets of samples: reverberance, clarity, source width and listener envelopment. It was chosen to work with such an "expert panel", since non-experts will need extensive training to get suffiently acquainted with these attributes.

The samples consisted of two different signals: male speech and (solo) cello. These samples were chosen because of their differences in spectral and temporal structure. The samples, which were anechoic, were convolved with simulated different binaural impulse responses. The simulation was performed using models of shoebox-shaped halls with various sizes and other properties (like absorption coefficients), using image source modeling and statistical modeling. In the simulation Head Related Transfer Functions (HRTFs) were used from an ITA artificial head[1]. These HRTFs were measured in an anechoic room for various angles (azimuth and elevation).

The tests were conducted according to the method as proposed by Chevret and Parizet [16]: the subjects are asked to rate the attributes directly using sliders on the computer screen, while listening to the (binaural) samples through headphones. The subjects are also able to sort the samples by rating by pressing a button. After this sorting the participants can easily re-listen to the samples pair-wise and make final adjustments in the ratings. In [16] it was shown that this method is as fast as direct evaluation, while the results are similar to those of the paired comparison method.

The test signals were also used as input for the binaural model to derive the proposed set of parameters. Also, the following conventional acoustic parameters were derived directly from the impulse responses[2]:

- Reverberation time RT, as a measure for reverberance.
- Clarity index C_{80}, as a measure for clarity.
- The early interaural cross-correlation $IACC_{E3}$, as a measure for apparent source width.
- Listener envelopment LEV [17], as a measure for envelopment.

A list of the virtual rooms used in the test, as well as their room acoustical parameters is shown in Table 1. For the rooms with names "Hall1" and "Hall2" the same virtual rooms were used, but the responses were simulated at different positions.

The relation between objective parameters - either the conventional parameters or the output parameters of the proposed model - and the subjective results of the listening tests will be expressed by using the correlation coefficient. The results are shown in Tables 2 and 3 for the speech and cello sample, respectively.

	Correlation coefficients ρ	
Attribute	**Conv. param.**	**Model param.**
Reverberance	RT (0.85)	L_B (0.92)
Clarity	C_{80} (0.86)	$R_{R,D}$ (0.96)
Source width	$IACC$ (0.95)	σ_{ITD} (0.95)
Envelopment	LEV (0.77)	$\sigma_{ITD,R}$ (0.96)

Table 2: The correlation coefficients beween subjective and objective attributes (conventional and as obtained from the model) for the male speech sample.

	Correlation coefficients ρ	
Attribute	**Conv. param.**	**Model param.**
Reverberance	RT (0.80)	L_B (0.85)
Clarity	C_{80} (0.93)	$R_{R,D}$ (0.96)
Source width	$IACC$ (0.95)	σ_{ITD} (0.98)
Envelopment	LEV (0.78)	$\sigma_{ITD,R}$ (0.96)

Table 3: The correlation coefficients beween subjective and objective attributes (conventional and as obtained from the model) for the cello sample.

Discussion

From the high correlation coefficients in Tables 2 and 3 it can be seen that under the test conditions used in this paper, both the conventional room acoustical parameters as the ones obtained using the auditory model perform well in predicting subjective attributes. The conclusion can be drawn that the auditory model used in this paper can indeed be used to predict subjective acoustical parameters.

A big advantage of using the auditory model is the fact that the parameters are determined from running audio. In practice, this will mean that there is no need to use common excitation signals like LMS noise or logarithmic sweeps. Instead, the acoustical quality of a concert hall could be determined during a performance, for example. Moreover the model will take into account the properties of the signal, therefore parameters for different kinds of signals can be obtained.

Currently a second series of listening tests is being conducted, in which a different set of virtual rooms is used. This set of rooms is obtained in the same way as the set in Table 1, but this time it is tried to make the subjective attributes more independent from each other. It is expected that the results from these tests will give better insight in the correlation between subjective and objective parameters. The set also includes three closely spaced measurement positions in the same room, where the conventional parameters differ significantly. This way it can be investigated if the model output parameters suffer less from unwanted spatial fluctuations.

[1] `http://www.akustik.rwth-aachen.de/`

[2] In order to derive conventional parameters, the omni-directional responses were also simulated.

Room	Name	WxLxH (m)	RT (s)	C_{80} (dB)	LF_E	$(1 - IACC_{E3})$	LEV (dB)
1	Anechoic	10x10x10	0.00	17.70	0.00	0.00	−31.11
2	Office	7x7x3	0.11	7.43	0.03	0.06	−11.45
3	Auditorium1	20x40x10	0.70	1.57	0.04	0.21	−1.56
4	Auditorium2	20x40x10	0.72	−0.36	0.21	0.59	−1.57
5	Hall1	23x39x19	1.82	−3.88	0.11	0.55	−0.04
6	Hall2	23x39x19	1.85	−4.44	0.09	0.81	−0.31
7	Hall3	10x39x19	1.41	−2.29	0.22	0.79	1.29
8	Hall4	39x39x19	2.04	−3.63	0.07	0.58	−0.49
9	Cathedral	20x100x30	6.84	−8.16	0.23	0.91	1.40

Table 1: The dimensions and various room acoustical parameters for the virtual rooms used in the listening tests.

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 Organization for Standardization, 1997.

[4] D. J. Breebaart. *Modeling binaural signal detection.* PhD thesis, Eindhoven University of Technology, 2001.

[5] T. Dau, D. Püschel, and A. Kohlrausch. A quantitative model of the "effective" signal processing in the auditory system. I. Model structure. *J. Acoust. Soc. Am.*, 99(6):3615–3622, June 1996.

[6] T. Dau, D. Püschel, and A. Kohlrausch. A quantitative model of the "effective" signal processing in the auditory system. II. Simulations and measurements. *J. Acoust. Soc. Am.*, 99(6):3623–3631, June 1996.

[7] P. Nelson, M. Park, T. Takeuchi, and F. Fazi. Binaural Hearing and Systems for Sound Reproduction. In *Proceedings of Acoustics '08, Paris*, June 2008.

[8] S. van de Par, J. Koppens, A. Kohlrausch, and W. Oomen. A new perceptual model for audio coding based on spectro-temporal masking. In *Proceedings of the 124th AES Convention, Amsterdam*, May 2008.

[9] D. Griesinger. The psychoacoustics of apparent source width, spaciousness & envelopment in performance spaces. *Acta Acustica united with Acustica*, 83(4):721–731, July 1997.

[10] R. Mason. *Elicitation and measurement of auditory spatial attributes in reproduced sound.* PhD thesis, University of Surrey, 2002.

[11] G. A. Soulodre. Can Reproduced Sound be Evaluated using Measures Designed for Concert Halls? In *Proceedings of the Workshop on Spatial Audio and Sensory Evaluation Techniques, Guildford, UK*, April 2006.

[12] J. Blauert and W. Lindemann. Auditory spaciousness: Some further psychoacoustic analyses. *J. Acoust. Soc. Am.*, 80(2):533–542, August 1986.

[13] J. Blauert and W. Lindemann. Spatial mapping of intracranial auditory events for various degrees of interaural coherence. *J. Acoust. Soc. Am.*, 79(3):806–813, March 1986.

[14] W. Hess. *Time-Variant Binaural-Activity Characteristics as Indicator of Auditory Spatial Attributes.* PhD thesis, Fakultät für Elektrotechnik und Informationstechnik der Ruhr-Universität Bochum, 2006.

[15] F. Rumsey, S. Zielinski, P. Jackson, M. Dewhirst, R. Conetta, S. George, S. Bech, and D. Meares. QESTRAL (Part 1): Quality Evaluation of Spatial Transmission and Reproduction Using an Artificial Listener. In *Proceedings of the 125th AES Convention, San Fransisco*, October 2008.

[16] P. Chevret and E. Parizet. An efficient alternative to the paired comparison method for the subjective evaluation of a large set of sounds. In *Proceedings of the 19th International Congress on Acoustics (ICA 2007), Madrid*, September 2007.

[17] L. L. Beranek. Concert Hall Acoustics - 2008. *J. Audio Eng. Soc.*, 56(7/8):532–544, July 2008.