Proceedings of the International Symposium on Room Acoustics, ISRA 2010

29–31 August 2010, Melbourne, Australia

An artificial listener for assessing content-specific objective parameters related to room acoustical quality

J. van Dorp Schuitman (1) and D. de Vries (1)

(1) Delft University of Technology, Faculty of Applied Sciences, Laboratory of Acoustical Imaging and Sound Control, Lorentzweg 1, 2628 CJ Delft, The Netherlands

PACS: 43.66.Ba, 43.55.Mc, 43.55.Hy, 43.66.Pn

ABSTRACT

A nonlinear, binaural auditory model was developed which is able to act as an artificial listener for assessing features related to the quality of room acoustics. The model is able to derive objective parameters for reverberance, clarity, apparent source width (ASW) and listener envelopment (LEV). In order to derive these parameters, the model splits the input audio into two streams: one related to the source (direct sound) and one related to the environment (room). In order to derive these two streams, the splitting algorithm makes use of the nonlinear properties of the auditory model. The whole algorithm works on arbitrary binaural recordings and therefore the parameters can for example be determined in a concert situation using an artificial head. This way the signal type is automatically taken into account, meaning that the acoustics of a room can be tested for multiple types of stimuli. Listening tests show that the resulting objective parameters correlate highly with perceptual results, even in cases where conventional objective parameters show low correlation coefficients.

INTRODUCTION

Generally, when acousticians want to assess the acoustical qualities of a room, objective parameters are evaluated which are determined from measured (or sometimes simulated) room impulse responses. Important objective parameters like reverberation time and clarity index are specified in ISO standard 3382 [1].

However, this way of assessing room acoustics has two major shortcomings:

- Because of practical limitations impulse responses are mostly measured in empty rooms. However, it is known that the values for most acoustical parameters depend on whether a room is filled with people or not [2].
- The method of calculating objective parameters from impulse responses does not take into account the fact that we do not listen to the responses directly, but to the audio of the source(s) convolved with these responses. The temporal and spectral features of the source signal are important for the perception of room acoustics [3–5].

In this paper a more novel method is proposed for assessing quality features of room acoustics, based on a binaural, nonlinear model of the human auditory system. When taking into account features of the auditory system, parameters might be obtained which are closer to human perception. Furthermore, the model accepts arbitrary binaural recordings as input, meaning that it is relatively easy to measure the acoustics of a room in a concert situation using a dummy head, for example. The temporal and spectral features of the source signal will automatically be taken into account.

THE BINAURAL MODEL

The auditory model used is based on the work by Breebaart [6] who developed a binaural extension of the model as developed by Dau *et al.* [7, 8]. A schematic version of the model is shown in figure 1. Below the various stages of the model will be discussed briefly.

- **Peripheral processor:** In this stage the ear canal and cochlea are modeled using a bandpass filter and a Gammatone filter bank, respectively. The hair cells are modeled using half-wave rectification and neural adaptation is modeled using a series of five feedback loops with different time constants. The absolute threshold of hearing is modeled by applying a frequency-dependent threshold right before the adaptation loops. Note that this stage is carried out for the left and right ear signals seperately.
- **Binaural processor:** This processor takes the outputs Ψ of the peripheral processor for the left and right ear signals and determines the Interaural Time Difference ITD as a function of time for each band of the Gammatone filter bank.
- **Central processor:** Here the monaural outputs Ψ' (which are low-pass filtered version of the outputs of the peripheral processor) and the running ITD values are combined to calculate objective parameters related to room acoustics. These parameters will be discussed in the next section.

Figure 2 shows example outputs of the binaural model. A dry, male speech signal was convolved with two different simulated binaural room impulse responses; one for a room with a short reverberation time (0.39 s, left) and one for a room with a longer reverberation time (1.98 s, right). Both input signals were normalized to the same RMS level. Figure 2 shows the monaural outputs Ψ' for the left channel as well as the ITD and ILD values as a function of time.

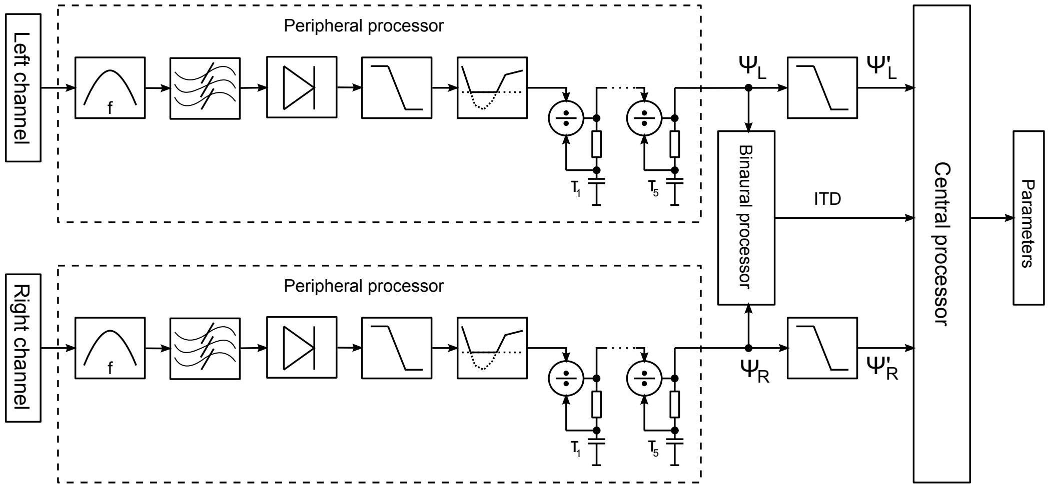

Figure 1: A schematic version of the binaural auditory model.

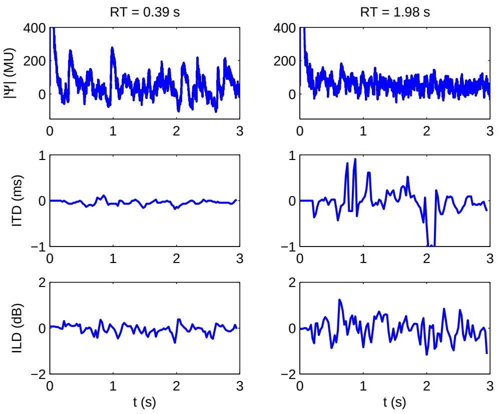

Figure 2: Example output of the binaural model. A speech signal was used, convolved with simulated binaural room impulse responses (BRIRs). The BRIRs were simulated for two virtual halls with reverberation times $RT = 0.39$ s (left) and $RT = 1.98$ s (right). The top figure shows a monaural output Ψ_L 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 265 Hz band.

Some important things can be noticed from the outputs. First, the more reverberant signal results in an overall lower output from the model, even though the RMS levels of the input signals were identical. Second, in the less reverberant case the signal components (phonemes) of the speech signal are more distinct compared with the more reverberant case. They show up as clear peaks in the model output.

Finally, ITD and ILD values fluctuate more as a function of time in the more reverberant case, due to reflections arriving from lateral directions. This will lead to a broadening of the source and a subjective impression of envelopment, as found by Blauert and Lindemann [9, 10]. In a later section this effect will be used to derive objective parameters related to these attributes. Authors who recently published research on this subject include Mason [4], Hess [11] and Rumsey *et al.* [12].

OBJECTIVE PARAMETERS

From the model outputs four different objective parameters will be determined which are related to the following perceptual attributes:

1. Reverbance
2. Clarity
3. ASW
4. LEV

These four attributes are thought of as being relevant for the perception of the acoustics of a room [2]. The attributes will be explained in more detail in the next sections.

Reverberance

Reverberance is the amount of reverberation experienced by listeners. It is related to the reverberation time; the time it takes for the sound pressure level to decay by 60 dB after the sound source stops. Most often the Early Decay Time (EDT) is used, which is obtained from the time in which the sound pressure level decays from 0 to -10 dB, multiplied by six. The EDT has been said to be a better predictor for perceptual reverberance than the reverberation time [13]. In order to determine the amount of reverberation from the model outputs, an algorithm was developed which separates the monaural model output into two streams: one for the direct sound and one for the reverberant field. The splitting procedure is based on a peak detection algorithm, where parts of the signal which are above a threshold for a minimum amount of time are labeled as 'direct sound'. An example result of this algorithm is shown in figure 3.

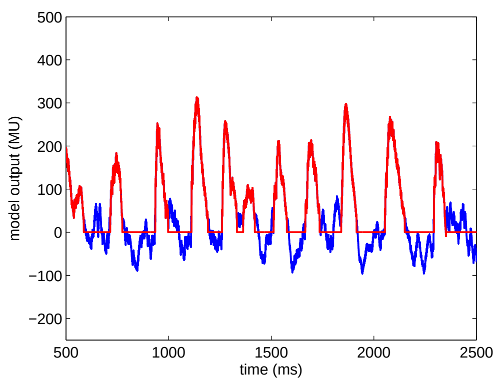

Figure 3: Separation of the model output in direct sound (red) and reverberant sound (blue) streams.

Then, as a predictor for reverberance the average level of the reverberant output stream L_{rev} (in the 250 - 4000 Hz range) is used.

Clarity

Clarity is the degree to which discrete sounds in a signal stand apart in time from one another subjectively. For music, clarity is usually estimated using the clarity index C_{80} which is the ratio between early (< 80 ms) and late (> 80 ms) energy in the impulse response [14]. For speech signals usually C_{50} is used, where the 80 ms time limit is changed to 50 ms ([14]).

Here clarity will be estimated as the ratio between the average level of the direct sound stream over that of the reverberant stream in the 250 - 4000 Hz range ($L_{\text{dir}}/L_{\text{rev}}$).

Apparent Source Width

In a room apparent broadening of a sound source can occur as a result of early lateral reflections, resulting in a certain apparent source width (ASW). ASW is most often assessed using the early interaural cross-correlation ($1-\text{IACC}_{E3}$) which is calculated from binaural impulse reponses as measured using an artificial head [2].

Basically, lateral reflections interfering with the direct sound cause the ITD to fluctuate over time, as discussed in the section describing the binaural model. Therefore the amount of ITD fluctuation is related to ASW, as was also proposed by Blauert and Lindemann [9], Lindemann [15], Griesinger [16], Mason [4], Becker [17] and Hess [11]. Furthermore, [18] showed that the perceived source width is not only related to the interaural decorrelation above 500 Hz, but also depends on the absolute sound pressure level at frequencies below 500 Hz. Therefore in this research the output of the binaural processor and the level in the lower bands are used to estimate ASW using the model:

$$\text{ASW}_{\text{model}} = \alpha_1 L_{0-500} + \log_{10}\left(1+\beta_1\sigma_{\tau,\text{dir}}\right), \qquad (1)$$

where L_{0-500} is the average monaural output level for the gammatone filters with center frequencies between 0 and 500 Hz and $\sigma_{\tau,\text{dir}}$ is the average standard deviation of ITD for the direct sound stream in the 500 - 2000 Hz bands. α_1 and β_1 are constants which can be tuned to yield the best results.

Listener Envelopment

Listener envelopment (LEV) is the second important subjective parameter related to spaciousness and is related to the environment instead of the source. A sound field is called enveloping when a perception of being surrounded by the sound occurs, since the sound seems to originate from all directions. ASW and LEV are considered to be the most important attributes related to spaciousness.

Currently, the objective parameter for envelopment is LEV which can be determined from impulse responses (note that in literature both the subjective and objective parameters for listener envelopment are referred to as LEV). This parameter was proposed by [19] and was recently turned into a more practical form by Beranek [20]:

$$\text{LEV} = 0.5G_{\text{late,mid}} + 10\log\left(1-\text{IACC}_{\text{late,mid}}\right), \qquad (2)$$

where $G_{late,mid}$ is the late (> 80 ms) sound strength, averaged over mid frequencies (500 and 1000 Hz octave bands). $\text{IACC}_{\text{late,mid}}$ is the late (> 80 ms) interaural cross-correlation averaged over those frequency bands.

Following this line of reasoning LEV consists of two elements: the absolute late sound pressure level and a spacious aspect (interaural cross-correlation). This concept is used to obtain an objective parameter predicting LEV from the auditory model outputs:

$$\text{LEV}_{\text{model}} = \alpha_2 L_{\text{rev}} + \log_{10}\left(1+\beta_2\sigma_{\tau,\text{rev}}\right), \qquad (3)$$

where L_{rev} is the mean level of the reverberant stream (500 - 8000 Hz bands) and $\sigma_{\tau,\text{rev}}$ is the mean standard deviation for the ITD values in that stream, averaged over the 500 Hz to 2000 Hz gammatone filter bands. The constants α_2 and β_2 can be tuned to yield the best results.

Table 1 summarizes the different parameters as described in the previous sections.

Subjective param.	Conventional	Model
Reverberance	T_{60}, RT	L_{rev}
Clarity	C_{80}, C_{50}	$L_{\text{dir}}/L_{\text{rev}}$
Apparent source width	$1-\text{IACC}_{E3}$	$\text{ASW}_{\text{model}}$
Listener envelopment	LEV	$\text{LEV}_{\text{model}}$

Table 1: An overview of the subjective parameters and the corresponding objective parameters, both conventional (as determined from impulse responses) as the ones proposed in this paper.

MODEL VALIDATION

To validate the binaural model and the objective parameters proposed in the previous section, listening tests were conducted. In these tests, subjects gave ratings for the four perceptual attributes for different simulated binaural room impulse responses (test I and II) or binaural impulse responses measured in various real rooms (test III). Two different stimuli were used: solo cello music and male speech. Both samples were anechoic recordings. The subjects listened to the signals through headphones (Beyer Dynamic DT-770 Pro), while being able to adjust their ratings on a computer. All orders (samples, rooms, attributes) were randomized, making the test design double-blind. The subjects were also able to sort their ratings from highest to lowest, making it possible to make fine-adjustments. This is an efficient form between direct rating and a paired comparison test, as shown by Chevret and Parizet [21]. The test interface is shown in figure 4.

For tests I and II, the binaural room impulse responses were simulated using custom software which is capable of simulating (microphone and/or dummy head) responses for shoebox-shaped rooms. For dummy head simulations, anechoic measurements from an ITA Head (http://www.akustik.rwth-aachen.de) were used to generate the responses. All samples were normalized to the same loudness level using the Replaygain algorithm (www.replaygain.org). This algorithm estimates the perceived loudness of a sample by evaluating the RMS level in windows of 50 ms length, where frequency weighting is applied according to an approximation of the equal loudness curve. The 95% highest value is considered to be the perceived loudness value.

Listening test I

In a first listening test, a group of five (expert) subjects rated the attributes for nine different virtual rooms. The subjects were considered experts, because they are all working in the field of room acoustics and are familiar with the various attributes. All subjects reported normal hearing. The values for a list of ‘conventional’ objective parameters for these rooms are shown in table 2.

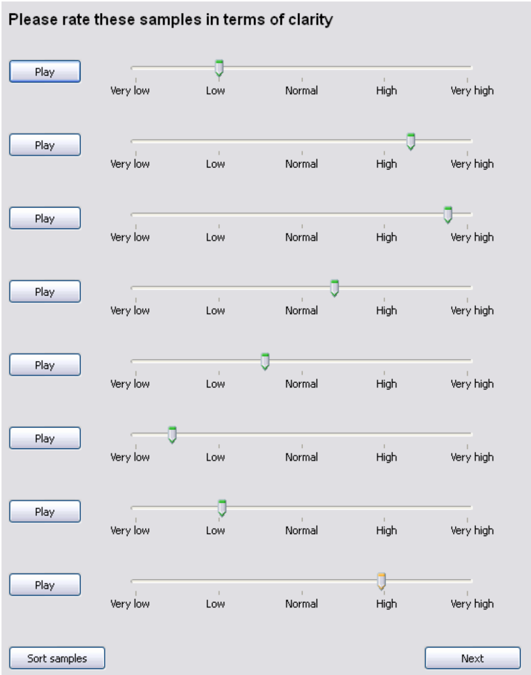

Figure 4: Example of the user interface used in the listening tests.

Room	EDT_{mid} (s)	C_{80} (dB)	C_{50} (dB)	$(1-IACC_{E3})$	LEV (dB)
1	0.01	22.17	15.24	0.00	−31.12
2	0.07	7.78	4.23	0.06	−11.46
3	0.51	1.53	−0.72	0.21	−1.56
4	0.74	−0.43	−3.28	0.59	−1.57
5	1.67	−3.97	−5.91	0.55	−0.04
6	1.79	−4.36	−5.53	0.81	−0.31
7	1.42	−2.23	−8.42	0.79	1.29
8	1.99	−3.60	−7.08	0.58	−0.49
9	6.97	−8.04	−11.82	0.91	1.40

Table 2: The various room acoustical parameters at the receiver's position for the virtual rooms used in listening test I.

As can be seen from the table, the rooms have a wide range for the various conventional objective parameters.

To test the performance of both the conventional room acoustical parameters as the newly proposed parameters the correlation coefficients between these parameters and the subjective results from the listening tests are shown in tables 3 and 4.

	Correlation coefficients r	
Attribute	**Conv. param.**	**Model param.**
Reverberance	EDT (0.79)	L_{rev} (0.94)
Clarity	C_{80} (0.93)	$L_{\text{dir}}/L_{\text{rev}}$ (0.93)
Source width	$IACC$ (0.96)	$\text{ASW}_{\text{model}}$ (0.93)
Envelopment	LEV (0.77)	$\text{LEV}_{\text{model}}$ (0.98)

Table 3: The correlation coefficients beween subjective and objective attributes (conventional and as obtained from the model) for the male speech sample for listening test I.

	Correlation coefficients r	
Attribute	**Conv. param.**	**Model param.**
Reverberance	EDT (0.85)	L_{rev} (0.98)
Clarity	C_{80} (0.83)	$L_{\text{dir}}/L_{\text{rev}}$ (0.94)
Source width	$IACC$ (0.95)	$\text{ASW}_{\text{model}}$ (0.85)
Envelopment	LEV (0.77)	$\text{LEV}_{\text{model}}$ (0.95)

Table 4: The correlation coefficients beween subjective and objective attributes (conventional and as obtained from the model) for the solo sample for listening test I.

Listening test II

Based on the findings of listening test I a second listening test was conducted with eight new virtual rooms, which were simulated using the same procudure as for the previous test. The same five subjects as in test I participated in this test. An attempt was made to make the perceptual attributes more independent, for example by making rooms with side walls with an absorption coefficient of 1, whereas the rest of the boundaries get a very low absorption coefficient. This way the reverberance in the room will be high, but due to a lack of reflections from the side walls the apparent source width will be low.

The conventional room acoustical parameters are listed in table 5. Note that the rooms all have the same geometry, but have different values for the various room acoustical parameters due to differences in absorption coefficient, receiver position, diffusivity, etc.

Room	EDT_{mid} (s)	C_{80} (dB)	C_{50} (dB)	$(1-IACC_{E3})$	LEV (dB)
1	1.07	−1.66	−3.11	0.44	−1.23
2	1.06	−2.00	−2.91	0.47	−1.22
3	1.14	−1.21	−2.11	0.59	−1.38
4	1.64	−3.29	−3.70	0.42	−0.65
5	1.60	−2.97	−4.50	0.55	−0.13
6	1.52	−2.93	−4.09	0.71	0.74
7	1.55	−2.64	−3.15	0.49	−0.05
8	1.56	−4.29	−4.64	0.63	5.16

Table 5: The various room acoustical parameters at the receiver's position for the virtual rooms used in listening test II.

Again, the correlation coefficients between objective and subjective parameters were calculated to test the performance of both the conventional room acoustical parameters and the parameters as obtained from the model. The results are shown in tables 6 and 7.

	Correlation coefficients r	
Attribute	**Conv. param.**	**Model param.**
Reverberance	EDT (-0.49)	L_{rev} (0.85)
Clarity	C_{50} (-0.12)	$L_{\text{dir}}/L_{\text{rev}}$ (0.80)
Source width	$IACC$ (0.56)	$\text{ASW}_{\text{model}}$ (0.65)
Envelopment	LEV (0.16)	$\text{LEV}_{\text{model}}$ (0.88)

Table 6: The correlation coefficients beween subjective and objective attributes (conventional and as obtained from the model) for the male speech sample for listening test II.

	Correlation coefficients r	
Attribute	**Conv. param.**	**Model param.**
Reverberance	*EDT* (0.88)	L_{rev} (0.91)
Clarity	C_{80} (0.91)	$L_{\mathrm{dir}}/L_{\mathrm{rev}}$ (0.95)
Source width	*IACC* (0.74)	$\mathrm{ASW}_{\mathrm{model}}$ (0.78)
Envelopment	*LEV* (0.85)	$\mathrm{LEV}_{\mathrm{model}}$ (0.85)

Table 7: The correlation coefficients beween subjective and objective attributes (conventional and as obtained from the model) for the solo sample for listening test II.

Listening test III

To further validate the model, a third listening test was conducted with a larger group of subjects (15) and ten binaural impulse responses measured in *real* rooms. The group of subjects consisted of expert and naive listeners. The subjects got detailed instructions on the test, including audio examples to make them familiar with the semantics. The conventional room acoustical parameters for the rooms used in this test are listed in table 8.

Room	*EDT* (s)	C_{80} (dB)	C_{50} (dB)	$(1-IACC_{E3})$	*LEV* (dB)
1	0.01	8.77	8.74	0.01	−29.23
2	0.10	9.36	5.37	0.15	−9.70
3	0.76	−1.75	−3.89	0.84	2.08
4	0.88	0.11	−2.42	0.69	4.99
5	0.81	−0.50	−1.74	0.57	−3.37
6	0.81	−1.94	−1.99	0.24	−1.87
7	1.00	−1.37	−2.15	0.58	−3.00
8	5.25	−15.73	−23.89	1.00	7.35
9	1.90	−3.18	−4.87	0.53	−1.91
10	9.36	−7.53	−9.13	0.93	12.50

Table 8: The various room acoustical parameters at the receiver's position for the real rooms used in listening test III.

Also in this test, male speech and solo cello were used as stimuli. Just as in the first two tests, the samples were normalized to the same loudness levels using the Replaygain algorithm. The results for the correlation coefficients between objective and perceptual results are shown in tables 9 and 10.

	Correlation coefficients r	
Attribute	**Conv. param.**	**Model param.**
Reverberance	*EDT* (0.75)	L_{rev} (0.96)
Clarity	C_{50} (0.93)	$L_{\mathrm{dir}}/L_{\mathrm{rev}}$ (0.88)
Source width	*IACC* (0.93)	$\mathrm{ASW}_{\mathrm{model}}$ (0.83)
Envelopment	*LEV* (0.89)	$\mathrm{LEV}_{\mathrm{model}}$ (0.99)

Table 9: The correlation coefficients beween subjective and objective attributes (conventional and as obtained from the model) for the male speech sample for listening test III.

	Correlation coefficients r	
Attribute	**Conv. param.**	**Model param.**
Reverberance	*EDT* (0.79)	L_{rev} (0.95)
Clarity	C_{80} (0.92)	$L_{\mathrm{dir}}/L_{\mathrm{rev}}$ (0.89)
Source width	*IACC* (0.85)	$\mathrm{ASW}_{\mathrm{model}}$ (0.86)
Envelopment	*LEV* (0.89)	$\mathrm{LEV}_{\mathrm{model}}$ (0.96)

Table 10: The correlation coefficients beween subjective and objective attributes (conventional and as obtained from the model) for the solo sample for listening test III.

CONCLUSIONS AND DISCUSSION

A nonlinear, binaural auditory model was proposed which is capable of acting as an artificial listener in listening tests related to room acoustics; it outputs predictions for four different perceptual attributes: reverberance, clarity, apparent source width and listener envelopment. Three listening tests were performed to check the vadility of the model. In each test two different stimuli were used: solo cello and male speech.

From the resulting correlation coefficients between the objective parameters and perceptual results, it can be seen that the model is able to predict the perceptual results with a high correlation, even in cases where conventional parameters show low correlation coefficients. Probably due to the fact that aspects of the human perception are taken into account, the results from the model are better overall.

Another advantage of the model is that it accepts arbitrary binaural input, meaning that there is no need to measure sweeps or other artificial signals in empty rooms. Instead, the acoustical features of a room can be assessed in a concert situation, for example. Furthermore, the model takes the spectral and temporal features of the stimulus automatically into account.

REFERENCES

1 ISO. ISO 3382:1997: Acoustics - Measurements of the reverberation times of rooms with reference to other parameters. International Organization for Standardization, 1997.
2 L. L. Beranek. *Concert and opera halls - How they sound*. Acoustical Society of America, 1996.
3 D. Griesinger. The psychoacoustics of apparent source width, spaciousness & envelopment in performance spaces. *Acta Acustica united with Acustica*, 83(4):721–731, July 1997.
4 R. Mason. *Elicitation and measurement of auditory spatial attributes in reproduced sound*. PhD thesis, University of Surrey, 2002.
5 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.
6 D. J. Breebaart. *Modeling binaural signal detection*. PhD thesis, Eindhoven University of Technology, 2001.
7 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.
8 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.
9 J. Blauert and W. Lindemann. Auditory spaciousness: Some further psychoacoustic analyses. *J. Acoust. Soc. Am.*, 80(2): 533–542, August 1986.
10 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.
11 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.
12 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.
13 G. A. Soulodre and J. S. Bradley. Subjective evaluation of new room acoustic measures. *J. Acoust. Soc. Am.*, 98(1): 294 – 301, July 1995.
14 O. Abdel Alim. *Abhängigkeit der Zeit- und Registerdurchsichtigkeit von raumakustischen Parametern bei Musikdar-*

bietungen (Dependence of time and register definition of room acoustical parameters with musical performances). PhD thesis, TU Dresden, 1973.

15 W. Lindemann. Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals. *J. Acoust. Soc. Am.*, 80(6): 1608–1622, December 1986.

16 D. Griesinger. Objective measures of spaciousness and envelopment. In *Proceedings of the 16th International AES Conference*, 1999.

17 J. Becker. Spectral and temporal contribution of different signals to ASW analysed with binaural hearing models. In *Proceedings of the Forum Acusticum 2002, Sevilla*, 2002.

18 T. Okano, L. L. Beranek, and T. Hidaka. Relations among interaural cross-correlation coefficient (IACC_E), lateral fraction (LF_E), and apparent source width (ASW) in concert halls. *J. Acoust. Soc. Am.*, 104(1):255–265, July 1998.

19 G. A. Soulodre, M. C. Lavoie, and S. G. Norcross. Objective measures of listener envelopment in multichannel surround systems. *J. Audio Eng. Soc.*, 51(9):826–840, September 2003.

20 L. L. Beranek. Concert hall acoustics - 2008. *J. Audio Eng. Soc.*, 56(7/8):532–544, July 2008.

21 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.