Deriving content-specific measures of room acoustic perception using a binaural, nonlinear auditory model

Jasper van Dorp Schuitman[a] and Diemer de Vries
Delft University of Technology, Faculty of Applied Sciences, Department of Imaging Science and Technology, Laboratory of Acoustical Imaging and Sound Control, P.O. Box 5046, 2600 GA Delft, The Netherlands

Alexander Lindau
Technical University of Berlin, Audio Communication Group, 10587 Berlin, Germany

(Received 20 May 2010; revised 8 January 2013; accepted 10 January 2013)

Acousticians generally assess the acoustic qualities of a concert hall or any other room using impulse response-based measures such as the reverberation time, clarity index, and others. These parameters are used to predict perceptual attributes related to the acoustic qualities of the room. Various studies show that these physical measures are not able to predict the related perceptual attributes sufficiently well under all circumstances. In particular, it has been shown that physical measures are dependent on the state of occupation, are prone to exaggerated spatial fluctuation, and suffer from lacking discrimination regarding the kind of acoustic stimulus being presented. Accordingly, this paper proposes a method for the derivation of signal-based measures aiming at predicting aspects of room acoustic perception from content specific signal representations produced by a binaural, nonlinear model of the human auditory system. Listening tests were performed to test the proposed auditory parameters for both speech and music. The results look promising; the parameters correlate with their corresponding perceptual attributes in most cases.
 [http://dx.doi.org/10.1121/1.4789357]

PACS number(s): 43.66.Ba, 43.55.Mc, 43.55.Hy, 43.66.Pn [LMW] Pages: 1572–1585

I. INTRODUCTION

Even after decades of research, a valid assessment of room acoustic perception is still a discussed topic. As a contribution, this paper proposes a new method for the derivation of signal-based measures aiming at predicting aspects of room acoustic perception from content specific signal representations produced by a binaural, nonlinear model of the human auditory system. To differentiate these measures both from the classic impulse-response-based room acoustical parameters and from the psychological domain (i.e., from the dimensions of room acoustic perception they are thought to predict), they are called auditory parameters throughout the following text.

The ISO 3382-1 standard (ISO, 2009) describes a set of physical measures. Although the standard notes that room acoustic quality involves several "complex layers," its only recommendation for quantification is in regard to the reverberation time. In the appendix of the Standard, however, measures for attributes as loudness, clarity, apparent source width (ASW), and listener envelopment (LEV, the feeling of being surrounded by the sound) are also suggested.

Usually, acousticians determine these room acoustic parameters from room impulse responses measured (or simulated) at one or more positions. However, this method has its shortcomings. First of all, room impulse responses are typically measured in empty rooms. But the values for most acoustic parameters depend on whether a room is occupied or not (Beranek, 1996). Methods for correcting the reverberation time when measured in occupied rooms exist, but they are mostly effective in rooms with well-upholstered seats (Barron, 2005).

Second, the method of calculating parameters from room impulse responses does not take into account the fact that perception of room acoustics might be content-specific. The temporal and spectral features of the source signal are important. For example, depending on the stimulus type, the early arriving energy may mask the first part of the later arriving energy, thus influencing certain aspects such as the perceived amount of reverberance, clarity, and LEV. This effect is usually referred to as post-masking, and it typically depends on frequency and masker duration.

This paper presents a new method for deriving auditory parameters that are presumably relevant for the overall perception of room acoustic quality. The present authors decided to focus on four of the most important attributes of room acoustic perception listed in the previously mentioned ISO 3382-1 standard (ISO, 2009): Reverberance, clarity, ASW, and LEV. Section IV explains these four parameters in more detail. To overcome the mentioned drawbacks of impulse response-based measures, the approach includes binaural auditory modeling. The newly derived auditory parameters are thought to better represent the way in which humans naturally perceive room acoustics. Moreover, the new method has two additional benefits:

(1) Contrary to the conventional methods, the model accepts arbitrary binaural recordings as input. This way the resulting parameters, which are directly related to certain

[a]Author to whom correspondence should be addressed. Present address: Philips Research, Eindhoven, The Netherlands. Electronic mail: jasper.van.dorp@philips.com

perceptual attributes, are content-specific, and the room acoustics can be evaluated under specific cases. For example, the acoustics of a theater can be tested for speech and music applications separately.

(2) When acquiring measurements in the occupied hall, there is no need to disturb the audience with intrusive test signals. Instead, parameters can be calculated directly from recordings, for example, during a live concert situation.

II. PREVIOUS RESEARCH

There is a consensus that room acoustic perception is a multidimensional construct comprising some dominating attributes. It is, therefore, not surprising that several studies have assessed the related dimensions of room acoustic perception using multivariate statistical techniques (e.g., Schroeder *et al*., 1974; Lehmann and Wilkens, 1980; Sotiropoulou *et al*., 1995; Sotiropoulou and Fleming, 1995). According to Barron (2005), such studies mention eight attributes regularly: Reverberance, clarity, intimacy, ASW, LEV, loudness, brilliance, and warmth.

To assess the power of physical measures as predictors of perceptual aspects, researchers also correlated the dominating perceptual attributes with physical quantities describing the rooms and their sound fields. Numerous research results on relations between physical measures and perceptual attributes have been published. For example, Lehmann and Wilkens (1980) found dominating perceptual attributes as loudness, source width, reverberance, clarity, or tone color to be predicted by the respective measures of strength (G), reverberation time, and the center time or reverberation decay time over frequency. According to Ando (1983), four physical measures describe the perception of acoustics: The listening level (related to loudness), the arrival time of the first reflection (related to intimacy), the reverberation time (related to reverberance), and the magnitude of the interaural cross-correlation (related to ASW and LEV).

More recent research tried to overcome shortcomings of previous work by including some kind of auditory modeling in the room acoustical analysis stages. Griesinger (1995), for example, is an example of an early work using auditory modeling for the evaluation of room acoustics. He proposed a model based on a combination of 1/3 octave band filters and onset/offset detectors to predict the amount of reverberance that is being masked by an anechoic input signal. Unfortunately, no further details on the model are given in the paper.

Lokki and Karjalainen (2002) proposed a basic model of auditory signal processing to visualize room impulse responses in a perception-oriented fashion. First, the input signal is filtered with a frequency-weighting filter, followed by a Gammatone filter bank. The absolute values of the filter bank outputs are taken. This is followed by compression and a sliding time window. The authors admit that the model is quite basic but also claim that it is a useful tool to analyze room impulse responses because it better respects the frequency and time resolution of human hearing than a one-third octave band spectrum; no parameters are obtained from the model to analyze room acoustics.

Lately, research on auditory modeling of the perception of room acoustic quality has focused on *spatial* attributes like ASW and LEV. This probably has two reasons. So far, no consensus has been reached on the physical measure that predicts the perception of spatial aspects of a sound field accurately. Second, with the increasing popularity of multichannel audio systems, the need for objective methods to assess the quality of reproduced, spatial audio has gained importance.

An early paper where auditory modeling is applied to the assessment of spatial qualities in room acoustics was presented by Bilsen (1994). Bilsen used a "central spectrum" model, based on the work of Jeffress (1948), to calculate the interaural time difference (ITD) as a function of time. Bilsen (1994) further developed this into a measure for the ASW. Listening test results showed that this measure performed quite well in predicting perceived ASW.

Several studies support the finding that the perception of spaciousness is primarily related to the fluctuations in the interaural time difference (ITD) and interaural level difference (ILD) over time (e.g., see Blauert and Lindemann, 1986a,b; Lindemann, 1986a,b).

Becker (2002) tried to predict ASW through the fluctuations in ITD and proposed a binaural model for this purpose. In this model, the middle ear is simulated using a third order Bessel low-pass filter. Furthermore, the model includes a filter bank of 36 Roex filters (Glasberg and Moore, 1990). It also simulates the transduction from mechanical waves to neural pulses using a model as proposed by Meddis (1988). After applying the model to binaural signals, Becker (2002) determined the ITD as a function of time using two methods: An extended correlation method and a subtraction method. The fluctuations in ITD were used as a prediction for ASW. The results showed good correlation with listening test results, where subjects had to evaluate ASW for white noise stimuli (low-pass filtered at different cutoff frequencies), convolved with measured impulse responses. The results were not compared with conventional measures like IACC.

Mason *et al*. (2004) also proposed the usage of a binaural hearing model when evaluating spaciousness. The binaural model that Mason *et al*. propose is capable of calculating the cross-correlation for recorded binaural signals based on auditory processing. Mason *et al*. (2004) presented no results in the paper, but in his Ph.D. thesis, Mason showed that measures based on ITD fluctuations show good correlation with listening test results (Mason, 2002).

Ando (2007) has developed a "theory of preference" for concert hall acoustics using four physical measures, namely reverberation time, initial time delay, level (binaural listening), and interaural cross-correlation (IACC). The model uses a binaural impulse response-based measurement technique and a computational scheme for preference calculation.

In a series of papers, Rumsey *et al*. (2008) presented the QESTRAL (quality evaluation of spatial transmission and reproduction using an artificial listener) framework. They developed a model based on the work of Supper (2005). The model includes the division of the input signal into critical

bands, envelope smoothing, calculation of ILD and ITD, loudness weighting, and the combination of ILD and ITD for source localization. From the model outputs, Rumsey *et al.* extracted various metrics like intensity, entropy, ILD and ITD standard deviations, and more (Jackson *et al.*, 2008). A combination of listening test results and a regression model yielded a measure for overall spatial quality. The model's results closely match the listening test results (Dewhirst *et al.*, 2008).

Some relevant dimensions of room acoustic perception could be identified so far (reverberance, loudness, spectral coloration, clarity, ASW, LEV, etc.). Since the early 1990s, auditory modeling has been successfully applied in the prediction of room acoustic perception. However, studies were often lacking in the evaluation of the auditory parameters or did not compare the newly proposed parameters with the conventional ones. Furthermore, most studies focused on aspects of spaciousness or the prediction of evaluative parameters such as preference. This paper introduces four newly developed auditory parameters for predicting relevant perceptual attributes of room acoustic quality.

III. THE AUDITORY MODEL

Various approaches exist for modeling the human auditory system. Because the shortcomings of existing parameters need to be overcome (see Sec. I), an auditory model for assessing the perception of room acoustics should at least have the following features:

(1) Accurate modeling of temporal and spectral masking. Depending on the temporal (post-masking) and spectral (simultaneous masking) content of the stimulus, late parts of the impulse response may be masked, affecting perceptual attributes such as envelopment and reverberance (Griesinger, 1997).
(2) Nonlinearity. The human auditory system behaves as a nonlinear system, which makes the previously mentioned masking effects dependent on the sound pressure level (SPL) (Dau *et al.*, 1996a). As a result, the perceptual attributes will also depend on the SPL. See, for example, Kuttruff (2000), who found that ASW and LEV increase with the SPL.
(3) Binaural interaction. The auditory system assesses spatial aspects of the sound field by analyzing the relation of the two signals arriving at the left and right ear.

A model that features all of these features is the binaural model as proposed by Breebaart (Breebaart, 2001; Breebaart *et al.*, 2001) that is basically a binaural extension of the monaural model by Dau *et al.* (1996a,b). The model has been proven to predict various psychoacoustic effects accurately, like monaural and binaural masking (Breebaart, 2001) and localization in audio reproduction (Nelson *et al.*, 2008). Therefore this paper uses this model as a starting point. A schematic version of the complete model is shown in Fig. 1.

The following sections describe the stages of the model for which the peripheral processor is similar to the one proposed by Breebaart (2001). The binaural and central processors were developed in this research to form the complete model as shown in Fig. 1.

A. The peripheral processor

The first stage models the outer and middle ear, as well as the basilar membrane inside the cochlea, the hair cells, and neural firing. As shown in Fig. 1, all filtering is carried out for both the left and right ear signals separately. Because the binaural model is nonlinear, the input signals should be scaled to the correct level, where an RMS value of 1 for the pressure corresponds to a SPL of 0 dB.

First, a band-pass filter with cutoff frequencies at 1 and 4 kHz models the combined transfer function of the outer ear and the ear canal (Breebaart *et al.*, 2001). This band-pass filter is only applied if the outer ear and ear canal are not already modeled in the measurement process (e.g., by

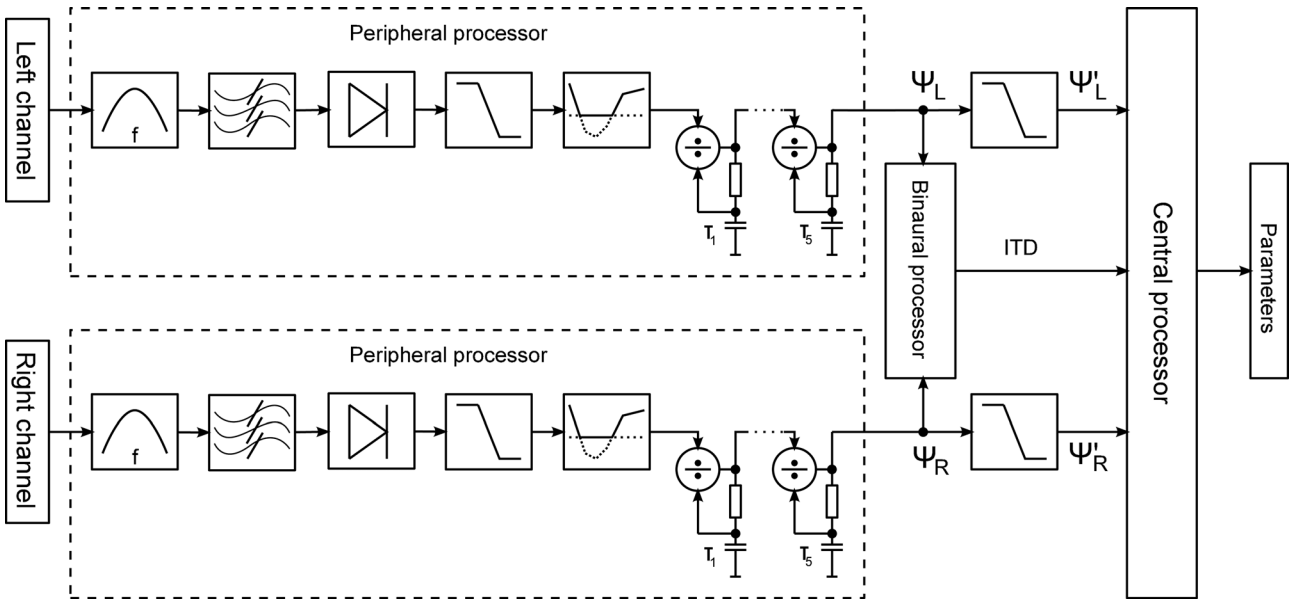

FIG. 1. A schematic version of the binaural auditory model. The full model contains five adaptation loops, of which two are shown in the figure (with time constants τ_1 and τ_5).

measuring using an artificial head). Next, a fourth-order gammatone filter bank consisting of 41 frequency bands with center frequencies from 27 to 20l577 Hz simulates the basilar membrane inside the cochlea (Patterson *et al.*, 1992). The signal processing of the inner hair cells is modeled by a half-wave rectifier, followed by a fifth-order low-pass filter with a cutoff frequency of 770 Hz to simulate phase locking at higher frequencies. Due to this filter, basically all phase information is lost at frequencies above 2000 Hz, and only the signal envelope is preserved (Breebaart *et al.*, 2001).

To incorporate the absolute threshold of hearing (ATH), a lower limit $\epsilon(f_c)$ is then applied to the signals that depends on the center frequency f_c of each band, according to the ATH curve from literature (Terhardt, 1979). Values below this frequency-dependent threshold are set to zero. Note that this approach is different from Breebaart's, who adds a Gaussian noise signal with a frequency-independent RMS level to simulate the ATH (Breebaart *et al.*, 2001).

The neurons in the human auditory system, which transmit electrical signals to the brain, adjust their sensitivity to the input level. To simulate this adaptation effect, the signals pass a chain of five feedback loops (Dau *et al.*, 1996a). The output y_i for each loop i (with $i = 1, \ldots, 5$) equals:

$$y_i[n] = (1 - \mathrm{e}^{-1/f_s\tau_i})\frac{x_i[n]}{y_i[n-1]} + \mathrm{e}^{-1/f_s\tau_i} y_i[n-1], \tag{1}$$

where x_i is the input signal for loop i and τ_i are the time constants (subsequently 5, 50, 129, 253, and 500 ms).

For stationary input signals, the chain of feedback loops acts effectively as a logarithmic compressor. The output level of the adaptation stage for stationary input signals is approximately equal to $y \approx x^{1/32}$ (Dau *et al.*, 1996a). Sudden changes, like onsets and offsets, lead to overshoots and undershoots in the outputs, respectively. Also thanks to the nonlinear adaptation stage, the model is able to simulate post-masking effects, depending on masker level and duration.

After the fifth adaptation loop, the output is scaled according to

$$\Psi[n] = \frac{100(y_5[n] - \epsilon(f_c))}{(10^5)^{1/32} - \epsilon(f_c)}, \tag{2}$$

with $y_5[n]$ the output of the last adaptation loop. Ψ is expressed in model units (MU). Equation (2) normalizes the output such that a stationary input signal of 100 dB SPL for the mid-frequency range results in a steady state output of 100 MU, while silence at the input yields a steady state output of 0 MU. For non-stationary signals, peaks and dips may occur in the output which are well beyond the signal's RMS pressure or even below zero, respectively.

B. The binaural processor

In Breebaart (2001) an equalization-cancellation (EC) approach is used for simulating binaural interaction in the human auditory system. This approach involves so-called excitation-inhibition (EI)-type elements, each with a characteristic ITD and ILD. The output of each EI-type element i for each time sample n and frequency band k is defined as

$$E_i[n,k] = \frac{1}{f_s}\sum_{m=-\infty}^{\infty}\left(10^{(\alpha_i/40)}\Psi_L\left[m + \frac{\tau_i f_s}{2}, k\right] - 10^{(-\alpha_i/40)}\Psi_R\left[m - \frac{\tau_i f_s}{2}, k\right]\right)^2 w(m-n), \tag{3}$$

where α_i and τ_i are the characteristic level and time differences of the EI element, respectively, and Ψ_L and Ψ_R are the left and right ear monaural model outputs (right after the adaptation stage). $w(m-n)$ is a double-sided exponential window with a time constant of $\tau_w = 30$ ms to incorporate a finite binaural temporal resolution (Breebaart, 2001). The outputs of EI-type elements will together form a pattern of the EI activity as a function of characteristic ITD and ILD. The output has its minimum for the EI-type element of which the characteristic ITD and ILD match those of the input signals.

Substituting the following:

$$\tilde{\Psi}_L = \Psi_L\left[m + \frac{\tau_i f_s}{2}, k\right] \tag{4}$$

and

$$\tilde{\Psi}_R = \Psi_R\left[m - \frac{\tau_i f_s}{2}, k\right], \tag{5}$$

then Eq. (3) becomes

$$E_i[n,k] = \frac{1}{f_s}\sum_{m=-\infty}^{\infty}(10^{(\alpha_i/20)}\tilde{\Psi}_L^2 + 10^{(-\alpha_i/20)}\tilde{\Psi}_R^2 - 2\tilde{\Psi}_L\tilde{\Psi}_R)w(m-n). \tag{6}$$

The maximum value of the ITD is around $\pm 700\,\mu$s depending on the dimensions of the head (Middlebrooks, 1999). Therefore it is safe to assume that $\tau_i \ll \tau_w$, which leads to the following approximations:

$$\sum_{m=-\infty}^{\infty} 10^{(\alpha_i/20)}\tilde{\Psi}_L^2 w(m-n) \approx \sum_{m=-\infty}^{\infty} 10^{(\alpha_i/20)}\Psi_L[m,k]^2 \times w(m-n) \tag{7}$$

and

$$\sum_{m=-\infty}^{\infty} 10^{(-\alpha_i/20)}\tilde{\Psi}_R^2 w(m-n) \approx \sum_{m=-\infty}^{\infty} 10^{(-\alpha_i/20)}\Psi_R[m,k]^2 \times w(m-n). \tag{8}$$

This means that the first two terms in the summation in Eq. (6) are independent of τ_i while the third term is independent of α_i. From the literature, it is shown that the frequency range 125–1000 Hz is dominant with respect to the perception of spaciousness (Barron and Marshall, 1981). For this frequency range, ITD is the dominant localization cue

(see Wightman and Kistler, 1992; Griesinger, 1992). So within the context of this research, only determining the ITD as a function of time is important, using:

$$\mathrm{ITD}[n,k] = \mathrm{argmin}_{\tau_i}\{E_i[n,k]\} = \mathrm{argmax}_{\tau_i}\left\{\sum_{m=-\infty}^{\infty}\tilde{\Psi}_L\tilde{\Psi}_R w(m-n)\right\}. \quad (9)$$

The term $\tilde{\Psi}_L\tilde{\Psi}_R$ is the running cross correlation between $\tilde{\Psi}_L$ and $\tilde{\Psi}_R$. Thanks to this simplification, the ITD can be calculated much faster than when the full EC algorithm, including ILD, is used.

C. The central processor

The central processor is the final stage of the model. It takes the output of the binaural processor (ITD values) as well as Ψ'_L and Ψ'_R as inputs. The latter two are obtained by applying a first-order low-pass filter with a time constant of 20 ms on the monaural outputs Ψ_L and Ψ_R of the peripheral processor to extract the envelope as proposed by Dau (Dau *et al.*, 1996a). The central processor produces the auditory parameters for assessing perceptual aspects of room acoustics.

Griesinger (1997), Rumsey (2002), and Mason *et al.* (2004) proposed the theory that the human auditory system splits an input stream into two: A direct (foreground) stream, which can be allocated to the source, and a reverberant (background) stream, which can be allocated to the environment. Starting from this, the present authors have developed parameters based on splitting the input stream into two streams. The nonlinear behavior of the model proposed here is exploited to perform the splitting. To illustrate this, examples for two different input signals will follow. Both signals originate from the same excerpt of an anechoic recording of a male speaker. The recording was taken from the EBU SQAM Compact Disk (EBU, 1988). For signal A, the dry signal was convolved with a simulated binaural room impulse response with a short reverberation time of 0.1 s (averaged over the 125–2000 Hz range). For signal B, a longer reverberation time of 0.7 s was used.

Both binaural signals were fed into the model after being scaled to an SPL of 70 dB (relative to 20 μPa). The monaural results Ψ'_L (left channel only) are shown in Fig. 2 for a critical band with a center frequency of 500 Hz. From the plots two important differences between both cases can be seen:

(1) For signal A, the overall model output is higher compared with the result for signal B.
(2) In the less reverberant case (signal A), the individual signal components show up more clearly in the output. The peaks are higher and more distinct.

These two differences are due to the nonlinearity of the model; in the less reverberant case (signal A), the onsets and offsets are more accentuated compared to signal B, where the overall level is more constant. The behavior of the adaptation stage explains this: Sudden changes in the input signal

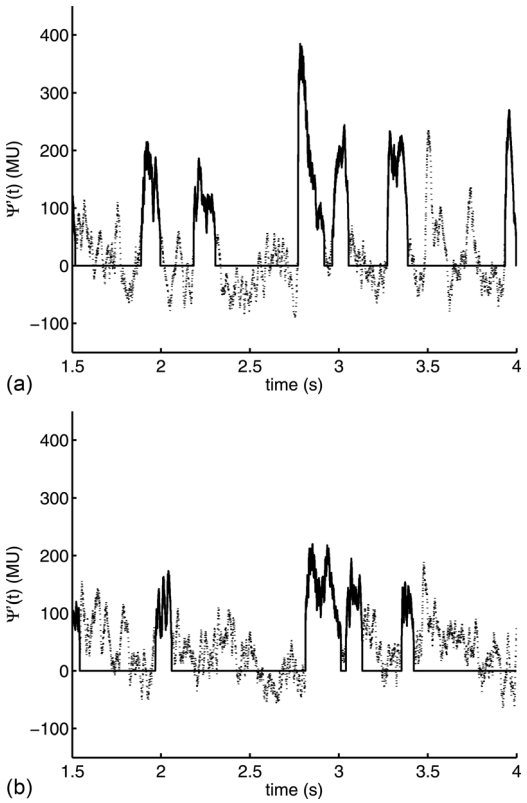

FIG. 2. The left channel monaural output $\Psi'(t)$ for signals A (a) and B (b). These results are for one gammatone filter with a center frequency of 500 Hz. The detected direct (solid line) and reverberant (dotted line) streams are shown separately.

are transformed linearly, whereas stationary parts are compressed (Dau *et al.*, 1996a).

To utilize this feature of the model, this paper proposes an algorithm that splits the model output into a direct and a reverberant stream by detecting peaks and assigning these peaks to the direct stream. The algorithm assigns a peak to the direct stream if the level of the model output stream is above a threshold $\Psi_{\min}$ for a time length of at least $T \geq T_{\min}$. The threshold $\Psi_{\min}$ for each frequency band with center frequency f_c follows from

$$\Psi_{\min}(f_c) = \mu_\Psi L_\Psi(f_c), \quad (10)$$

where $L_\Psi(f_c)$ is the average absolute level of the model output for the frequency band. μ_Ψ is a frequency-independent constant, which should be optimized.

To detect undershoots as a result of offsets in the input signal, the peak detection is repeated in the negative direction (not shown in Fig. 2). Because the undershoots generally have a lower level compared with overshoots, a different, lower threshold is used to detect those.

As an effect of the model's compressive behavior, it will react differently on different signal types. For signals containing more quasi-stationary ("legato") passages, e.g., cello music, longer periods of the streams will be assigned to the direct sound as when compared to more impulsive (e.g., percussive) types of signals. This corresponds to perception; in legato passages of signals, the direct sound will continuously mask the reverberant sound. Because signal offsets do

not occur as much as when compared to impulsive signals, it is more difficult to detect the reverberant decay.

IV. PHYSICAL PARAMETERS AND PERCEPTUAL ATTRIBUTES

The introduction explained that the focus of this paper is on four attributes that are presumably relevant to the perception of room acoustics. The following subsections discuss these attributes in more detail. For each attribute, the commonly used impulse response-based predictor is also discussed. The latter are mostly defined in ISO 3382-1 (ISO, 2009). For each perceptual attribute, a new "auditory parameter" is proposed.

A. Reverberance

Reverberance is the amount of reverberation perceived by listeners and is typically regarded as being closely related to the physical reverberation time; i.e., the time it takes for the sound pressure level to decay by 60 dB after the sound source stops. Because a signal-to-noise ratio (SNR) of 60 dB is hardly achieved in practical room impulse response measurements, ISO 3382-1 defines parameters such as T_{20} and T_{30} for which the reverberation time is estimated from linear fits for the -5 to -25 dB and -5 to -35 dB energy decay, respectively. Often the early decay time (EDT) is measured, which is obtained from extrapolation of the 0 to -10 dB decay (Jordan, 1970). The EDT has been said to be a better predictor for perceptual reverberance than the reverberation time (Soulodre and Bradley, 1995).

The auditory parameter related to reverberance is called P_{REV}. This paper proposes to evaluate reverberance using the average level of the reverberant sound stream as obtained by the monaural outputs of the model ($P_{\mathrm{REV}} = L_{\mathrm{REV}}$):

$$L_{\mathrm{REV}} = \frac{1}{N}\frac{1}{K}\sum_{n=0}^{N-1}\sum_{k=k_0}^{k_1}\Psi_{\mathrm{REV}}[n,k], \tag{11}$$

where N is the total number of time samples, k_0 and k_1 are the lowest and highest frequency band involved in the calculation, and $K = k_1 - k_0 + 1$ is the resulting number of frequency bands. k_0 and k_1 should be optimized; Sec. VI will discuss a suitable optimization method. Ψ_{REV} is defined as follows:

$$\Psi_{\mathrm{REV}}[n,k] = \sqrt{\Psi_{L,\mathrm{REV}}[n,k]^2 + \Psi_{R,\mathrm{REV}}[n,k]^2}, \tag{12}$$

with $\Psi_{L,\mathrm{REV}}$ and $\Psi_{R,\mathrm{REV}}$ the detected reverberant streams for the left and right channels.

In contrast to evaluating the physical reverberation time alone, when evaluating the average level of the reverberant stream, two aspects of the perceptual experience of reverberance are automatically taken into account:

(1) The perceived loudness of the reverberation is independent of the loudness of the direct sound (Griesinger, 1997).
(2) Due to masking effects, the content of a signal can potentially mask the reverberation. Because simultaneous and post-masking are accurately simulated by the model, these effects will influence the resulting value for P_{REV}.

B. Clarity

Clarity is the degree to which discrete sounds in a signal stand apart in time from one another subjectively. If clarity is high, it is easy to spot individual notes in a musical piece or individual phonemes in speech. For music, clarity is conventionally estimated using the clarity index C_{80}, which is the ratio between early (<80 ms) and late (>80 ms) energy in the impulse response (Reichardt *et al.*, 1975):

$$C_{80} = 10\log\frac{\int_0^{80\,\mathrm{ms}} p^2(t)\mathrm{d}t}{\int_{80\,\mathrm{ms}}^{\infty} p^2(t)\mathrm{d}t}. \tag{13}$$

For speech purposes, usually C_{50} is used, where the 80 ms time limit in Eq. (13) is changed to 50 ms (Abdel Alim, 1973). Often a 10 ms cosine-shaped window is applied around the transition between early and late energy.

This paper proposes to estimate perceptual clarity using the ratio between the mean direct sound stream level over the mean reverberant level ($P_{\mathrm{CLA}} = L_{\mathrm{DIR}}/L_{\mathrm{REV}}$), where L_{DIR} is calculated for the direct sound stream in a manner similar to Eq. (11):

$$L_{\mathrm{DIR}} = \frac{1}{NK}\sum_{n=0}^{N-1}\sum_{k=k_0}^{k_1}\Psi_{\mathrm{DIR}}[n,k], \tag{14}$$

with

$$\Psi_{\mathrm{DIR}}[n,k] = \sqrt{\Psi_{L,\mathrm{DIR}}[n,k]^2 + \Psi_{R,\mathrm{DIR}}[n,k]^2}. \tag{15}$$

This method will overcome the problems mentioned in Sec. I with energy ratio-based parameters; evaluating the ratio between the two detected streams does not ask for a fixed time limit. The central processor will automatically detect which part of the input stream should be assigned to either of the streams based on which one is psychoacoustically dominant.

C. 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 ASW (Marshall, 1967; Keet, 1968). It is considered to be one of the two most important aspects of acoustic spaciousness together with listener envelopment (see Sec. IV D). ASW is most often assessed using the early interaural cross correlation ($1 - \mathrm{IACC}_{\mathrm{E3}}$) (Schroeder *et al.*, 1974; Bradley and Soulodre, 1995), where

$$\mathrm{IACC}_{\mathrm{E3}} = \max\left|\frac{\int_0^{80\,\mathrm{ms}} p_L(t)p_R(t+\tau)\mathrm{d}t}{\sqrt{\int_0^{80\,\mathrm{ms}} p_L^2(t)\mathrm{d}t\int_0^{80\,\mathrm{ms}} p_R^2(t)\mathrm{d}t}}\right|, \tag{16}$$

where p_L and p_R are the left and right ear pressure responses (measured using a dummy head), respectively, and

$-1 < \tau < +1$ ms. The value is usually averaged over three octave bands: 500, 1000, and 2000 Hz. If the early part of the response is more diffuse due to lateral reflections, the interaural cross-correlation will be lower and the source will sound broader (Damaske and Ando, 1972). Energy that arrives later is thought to contribute to listener envelopment (LEV, see Sec. IV D).

Another commonly used parameter is the (*early*) *lateral energy fraction* LF (Barron and Marshall, 1981):

$$\mathrm{LF} = \frac{\int_{5\,\mathrm{ms}}^{80\,\mathrm{ms}} h_{f8}^2(t)\mathrm{d}t}{\int_{0}^{80\,\mathrm{ms}} h_{SR}^2(t)\mathrm{d}t}, \tag{17}$$

with $h_{f8}(t)$ the impulse response measured using a figure-eight microphone with its null pointed toward the source. LF has values in the range $0, \ldots, 1$, where higher values indicate more lateral energy and thus a broader sounding source. In concert halls for classical music, generally values in the range $0.05, \ldots, 0.35$ are found.

The time limit which separates "early" from "late" energy is set at 80 ms, but there is not much agreement among authors about the optimal value. Other values found in the literature, for example, are 105 ms (Soulodre *et al.*, 2003b) and 150 ms (Griesinger, 1999). In Soulodre (2006), frequency dependent time limits between early and late energy were proposed (from 160 ms at 125 Hz to 45 ms at 8 kHz). In a study dealing with the estimation of the time from which a room impulse response becomes diffuse, Hidaka *et al.* (2007) proposed a variable "transition time" that is dependent on the running correlation.

Morimoto (2002) also stated that the use of a strict time limit between energy contributing to ASW and energy contributing to LEV does not yield accurate predictors. He proposes to include the precedence effect (or "law of the first wave front"). Accordingly, reflections in the impulse response with an energy below a decaying curve defined by this law contribute to the precedence effect (and therefore to ASW), and reflections with energy above this curve contribute to LEV.

The decorrelation between the left and right ear signals, resulting from lateral reflections, leads to fluctuations in ITD and ILD (Blauert and Lindemann, 1986a). As discussed in the introduction, it is well understood that temporal fluctuations of ITD and ILD lead to the perception of spaciousness and that ITD is the dominant cue of these two. These findings can be used to obtain a parameter related to ASW using the model proposed here because ITD as a function of time is one of its outputs. For the two examples discussed in Sec. III C, the ITD as a function of time is plotted in Fig. 3. Because the source is located at 0 deg in this simulation, ITD fluctuates around 0 ms in this figure. The amount of fluctuation in ITD will generally increase when there are more reflections present as is the case here. So this effect can be used to predict perceptual aspects related to spaciousness.

Furthermore, Okano *et al.* (1998) showed that the perceived source width is not only related to the interaural decorrelation but also depends on the absolute sound pressure level at low frequencies. Therefore, in the present research the output of the binaural processor and the level in the lower bands are used to estimate ASW using the model:

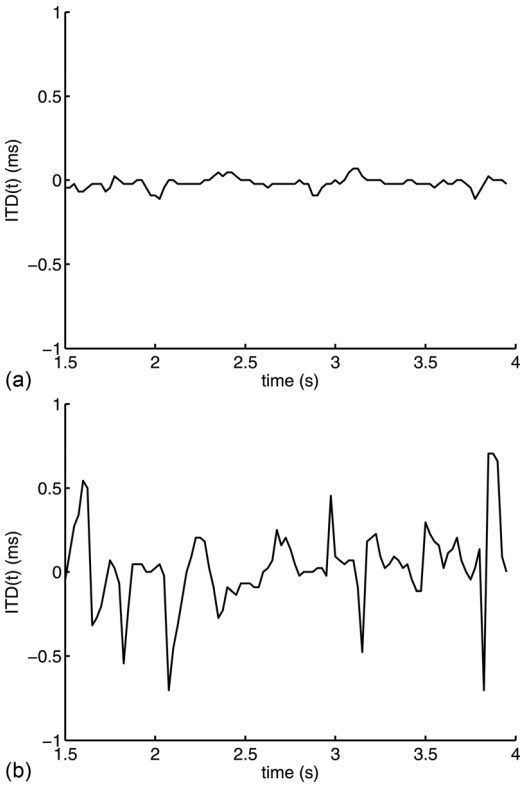

FIG. 3. ITD as a function of time for signal A (a) and B (b). These results are for one critical band with a center frequency of 500 Hz.

$$P_{\mathrm{ASW}} = \alpha_1 L_{\mathrm{LOW}} + \log_{10}(1 + \beta_1 \sigma_{\tau,\mathrm{DIR}} \cdot 10^3), \tag{18}$$

where L_{LOW} is the average monaural output level for the gammatone filters with low frequencies and $\sigma_{\tau,\mathrm{DIR}}$ is the standard deviation of ITD for the direct sound stream. This standard deviation is averaged over a certain frequency range for which the lower and upper limits need to be optimized. α_1 and β_1 are constants that also need to be optimized. Because the fluctuations in ITD in Eq. (18) are evaluated for the direct sound stream only, the model decides which part of the input stream belongs to the source, and thus contributes to ASW, instead of using a fixed time limit.

D. Listener envelopment

LEV is the second important perceptual parameter related to spaciousness and refers to the environment instead to the source. A sound field is called an enveloping one when a perception of being surrounded by the sound occurs, because it is coming from all directions (Barron and Marshall, 1981).

In ISO 3382-1, one minus the late interaural cross-correlation $(1 - \mathrm{IACC_L})$ is proposed as a measure for LEV, where:

$$\mathrm{IACC_L} = \max \left| \frac{\int_{80\,\mathrm{ms}}^{\infty} h_L(t) h_R(t+\tau)\mathrm{d}t}{\sqrt{\int_{80\,\mathrm{ms}}^{\infty} p_L^2(t)\mathrm{d}t \int_{80\,\mathrm{ms}}^{\infty} p_R^2(t)\mathrm{d}t}} \right|, \tag{19}$$

with $-1\,\text{ms} < \tau < +1\,\text{ms}$.

Another parameter for envelopment that can be determined from impulse responses was proposed by Soulodre *et al*. (2003a) Recently, Beranek (2008) turned their equation into a more practical form:

$$\text{LEV}_{\text{calc}} = 0.5G_{\text{late,mid}} + 10\log(1 - \text{IACC}_{\text{late,mid}}), \quad (20)$$

where the "late" sound strength G_{late} is calculated as

$$G_{\text{late}} = G - 10\log(1 + \exp(C_{80}/10)), \quad (21)$$

in C_{80} is the clarity index and G is the overall sound strength (Beranek, 1996):

$$G = 10\log\frac{\int_0^\infty p^2(t)\mathrm{d}t}{\int_0^\infty p_A^2(t)\mathrm{d}t}, \quad (22)$$

where $p_A(t)$ is the free-field pressure response as measured at a distance of 10 m from the source.

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

So according to the literature, LEV consists of two elements: The absolute late SPL (i.e., the level in the diffuse part of the impulse response) and a spacious aspect (interaural cross correlation). This concept is translated to the monaural and binaural model outputs to obtain a prediction of LEV, related more closely to auditory impression:

$$P_{\text{LEV}} = \alpha_2 L_{\text{REV}} + \log_{10}(1 + \beta_2\sigma_{\tau,\text{REV}} \cdot 10^3), \quad (23)$$

where L_{REV} is the mean level of the reverberant stream [Eq. (11)], and $\sigma_{\tau,\text{REV}}$ is the mean standard deviation for the ITD values in that stream. α_2 and β_2 are constants that should be optimized and are discussed later.

V. EVALUATION STRATEGY

To evaluate the performance of the method, four listening tests were conducted with varying conditions. In all tests, two different stimuli were used: Male speech and solo cello. In the following text, the four listening tests are briefly discussed:

(a) The first listening test included measurements for various real rooms, and 15 subjects participated. The results from this test were used to optimize the free parameters in the model.
(b) The second listening test included the same rooms as for test a, but this time the samples were normalized to the same loudness to see if this makes a difference perceptually and if the model is able to reproduce this.
(c) Listening test c included five expert subjects and nine "virtual" rooms, simulated using acoustic simulation software. This way, a wide range of acoustic environments can be tested.
(d) The last listening test (d) was performed by the same five expert subjects from test c. Again, simulated rooms were used in the test, but an attempt was made to construct more "acoustically unrealistic" rooms. The goal of this test was to see if the auditory parameters correlate with perceptual attributes, even if the rooms have unrealistic acoustic properties.

Section VI will discuss listening test a and the optimization strategy in more detail. In Sec. VII, tests b to d are discussed.

VI. OPTIMIZATION OF THE MODEL

The model described in the previous sections, as well as the proposed auditory parameters, contain some free parameters that need to be optimized. These free parameters include frequency ranges and the α and β constants in Eqs. (18) and (23), for example. It was chosen to perform this optimization using a genetic algorithm (GA). GAs form a class of algorithms that are capable of optimizing nonlinear, multi-modal problems with numerous free parameters, like the auditory model presented in this paper. These algorithms search the solution space by simulating evolution (survival of the fittest). A complete explanation of the algorithm is out of the scope of this paper; for more information, the reader is referred to Holland (1975).

To obtain a training data set for the optimization, a listening test was conducted ("listening test a"). In this test, subjects were asked to rate the room acoustic impression using the discussed perceptual attributes for different samples in different rooms. For this purpose, binaural room impulse responses were measured in different rooms using the ITA artificial head (Schmitz, 1995). The rooms, as well as values for some conventional room acoustic parameters, are listed in Table I. The SPL values were determined by calibrating the headphones using the ITA dummy head.

The binaural room impulse responses for AUD1, AUD5, AUD6, and AUD8 were all measured in the same room: The auditorium at Delft University of Technology. This room is equipped with a digital electro-acoustic system by Acoustic Control Systems (ACS) (Berkhout, 1988). Different presets of this system were active during the measurements for these four "rooms."

The subjects were asked to rate the four different perceptual attributes discussed in this paper (reverberance, clarity, apparent source width, and listener envelopment) while they listened to samples through headphones. The subjects could do so by moving a slider with the computer mouse to the desired rating on a continuous scale from "very low" to "very high." It was also possible to sort the samples from highest rating to lowest rating, after which fine adjustments could be made in a pair-wise fashion. This test method is capable of delivering results that are of the quality of a paired comparison test, while taking considerably less time (Chevret and Parizet, 2007). Each test included two different stimuli: Male speech and solo cello music. The length of the stimuli was 10 s. The subjects listened to the samples through a pair of Beyer Dynamic DT 770 Pro headphones (unequalized), connected to an RME Fireface 400 soundcard. The room in which the tests were conducted was an acoustically treated listening room with a very short

TABLE I. An overview of the rooms used in listening tests a and b. The first two columns list the room names. The table also lists some common, conventional room acoustic parameter values for each room. The last two columns show the SPL of the resulting audio samples used in the listening test.

Rooms		Room acoustic parameters (conventional)								Sample SPL	
Code	Type	T_{20} (s)	EDT (s)	C_{50} (dB)	C_{80} (dB)	1-IACC $_E$	1-IACC $_L$	LEV_{calc} (dB)	G (dB)	Speech (dB)	Cello (dB)
AR	Anechoic room	0.02	0.02	45.62	45.78	0.01	0.00[a]	−179.42	0.02	47	47
LR	Listening room	0.21	0.13	20.40	30.00	0.13	0.00[a]	−165.05	12.97	67	63
SW	Skyway	0.76	0.79	1.41	4.84	0.77	0.81	1.46	11.21	67	62
HW	Hallway	1.02	0.98	2.76	4.98	0.59	0.89	4.44	16.33	70	66
AUD1	Auditorium (ACS1)	1.21	0.88	3.63	6.66	0.24	0.77	−4.16	1.79	52	50
AUD5	Auditorium (ACS5)	1.67	1.44	5.64	7.03	0.11	0.83	−1.85	5.86	55	54
AUD6	Auditorium (ACS6)	2.29	1.12	3.05	5.89	0.24	0.81	−3.74	1.87	52	50
SC	Staircase	3.94	4.83	−20.61	−15.88	0.73	0.94	5.98	12.68	67	63
AUD8	Auditorium (ACS8)	4.81	2.11	−0.56	0.81	0.25	0.83	−2.50	0.18	53	50
RC	Reverberation chamber	10.12	9.84	−7.26	−6.26	0.71	0.91	11.44	24.78	79	76

[a]For rooms AR and LR, there was too little energy in the late part of the response to calculate $IACC_L$. Therefore 1-$IACC_L$ was set to a value of 0.00 for these rooms.

reverberation time (0.1–0.2 s) and low background noise (<40 dBA).

In test a, 15 subjects participated. The group of subjects consisted mostly of students (male and female) with mixed musical experiences and preferences. None of the subjects reported hearing problems. Before the start of the test, the subjects got instructions (including audio examples) explaining the four attributes that they had to rate.

The rating results of test a were used as input for a custom implementation of the GA to optimize the free parameters. The optimized parameters are shown in Table II.

VII. VALIDATION OF THE MODEL

To validate the model and the optimized parameters, three more listening tests were conducted. In all tests, the task of the subjects was to rate their room acoustic impression using the four selected attributes. The next subsections shortly discuss these tests.

A. Listening test b

The second listening test was conducted similar to test a. Again, a group of 15 subjects participated, different from the ones who participated in test a but with roughly the same demographics. The same rooms were used, although this time, the samples were normalized with respect to their estimated loudness level using the Replaygain algorithm (Robinson, 2001). This algorithm estimates the perceived loudness level 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. Because this algorithm estimates the perceived *loudness* level of a sample, the sound pressure levels of the normalized samples can still differ. The resulting SPL values are shown in Table III. As can be seen, the SPL values are indeed different for the various samples, although the differences are small. Also, the differences are much smaller than in test a (see Table I). The samples were normalized to test if subjects rate differently in this case (test b) compared with the case where loudness differences exist between the samples (test a).

B. Listening test c

In the third listening test, the stimuli were convolved with binaural room impulse responses (BRIRs) or virtual

TABLE II. The result of the free parameter optimization using a genetic algorithm.

Parameter	Description	Value
μ_Ψ	Relative level of the peak threshold	$7.49 \cdot 10^{-3}$
$\mu_{\Psi,dip}$	Relative level of the dip threshold	$-1.33 \cdot 10^{-3}$
T_{min}	Minimum width of a peak or dip	63.1 ms
k_0	Min. frequency band in the stream splitting procedure	5 (168 Hz)
k_1	Max. frequency band in the stream splitting procedure	20 (1.84 kHz)
q_0	Min. frequency band in ITD fluctuation calculation	9 (387 Hz)
q_1	Max. frequency band in ITD fluctuation calculation	20 (1.84 kHz)
z_0	Min. frequency band in low frequency level calculation	5 (168 Hz)
z_1	Max. frequency band in low frequency level calculation	9 (387 kHz)
α_1	Weighting factor used in ASW calculation	$2.00 \cdot 10^{-2}$
β_1	Weighting factor used in ASW calculation	$5.63 \cdot 10^{+2}$
α_2	Weighting factor used in LEV calculation	$2.76 \cdot 10^{-2}$
β_2	Weighting factor used in LEV calculation	$6.80 \cdot 10^{+2}$

TABLE III. The SPL of the different samples used in test b.

	Sample SPL (dB)	
Room	Speech	Cello
AR	70	70
LR	71	70
SW	74	71
HW	74	72
AUD1	72	72
AUD5	73	71
AUD6	72	71
SC	75	72
AUD8	73	72
RC	76	74

rooms, which were generated using simulation software. The software was developed within the scope of this research and is capable of simulating binaural room impulse responses as well as virtual microphone measurements for shoebox-shaped rooms. Convolving the direct sound and room reflections with measured, diffuse-field equalized head-related transfer functions from an ITA artificial head yielded the total BRIR. The first and second order reflections were simulated using image source modeling. The late part of the response was simulated statistically using an exponentially decaying white noise process, obeying the density of reflections, which is expected in the virtual room as a function of time. The boundary absorption coefficients were defined in octave frequency bands and high frequency absorption through air was taken into account.

In total, nine virtual rooms were simulated with a large spread in acoustic features. Table IV shows an overview of the nine rooms, as well as some common, conventional room acoustic parameters. Also here the audio samples were all normalized to the same (estimated) loudness level using the Replaygain algorithm.

In total, five subjects participated in this listening test. This is a rather low number, but it can be justified by the fact that all five subjects can be considered "experts" in the field of acoustics. The subjects were all male and working in the field of acoustics, either as a Ph.D. student or as an associate professor. They had mixed musical experiences and preferences. The participants were all working at the Laboratory of Acoustics of the TU Delft, and they were familiar with the four different attributes.

C. Listening test d

In listening test d, virtual rooms were used to obtain binaural room impulse responses, just as for test c. However, for this test, an attempt was made to make the four room acoustic attributes more independent from each other compared to the previous test. An increased independence of the attributes was attempted by simulating more "unrealistic" rooms. For example, room VR13 has a high reverberation time of $T_{20} = 1.75$ s but also has side walls that are completely absorbing. This resulted in a low value for (1-IACC$_E$) of 0.07 and possibly in a low ASW due to the lack of reflected energy arriving from lateral directions. In turn, the value for LEV$_{calc}$ is rather high. A complete list of the rooms and their corresponding conventional parameters can be found in Table V.

Also in this test, all the signals were normalized to the same estimated loudness level using the Replaygain algorithm. The same group of expert subjects from listening test c participated in test d.

VIII. RESULTS OF THE EVALUATION

The method proposed in this paper was validated by evaluating the correlation coefficients between the listening test results and the auditory parameters P_{REV}, P_{CLA}, P_{ASW}, and P_{LEV} and by comparing their performance to those of the respective conventional room acoustic parameters. The results are shown in Tables VI, VII, VIII, and IX. For each row (test/stimulus combination), the highest correlation coefficient is shown in bold.

TABLE IV. An overview of the (virtual) rooms used in listening test c. The first two columns list the room names as well as the type of room after which they were modeled. The table also lists some common conventional room acoustic parameter values for each room. The last two columns show the SPL of the resulting audio samples used in the listening test after applying level normalization using the Replaygain algorithm.

Rooms		Room acoustic parameters (conventional)									Sample SPL	
Code	Type	T_{20} (s)	EDT (s)	C_{50} (dB)	C_{80} (dB)	1-IACC$_E$	LF	1-IACC$_L$	LEV$_{calc}$[a] (dB)	G[a] (dB)	Speech (dB)	Cello (dB)
VR01	Anechoic room	0.01	0.01	50.91	64.13	0.00	0.04	0.00[b]	−184.61	7.98	68	69
VR02	Office	0.33	0.08	19.03	23.70	0.06	0.04	0.09	−11.39	21.35	69	69
VR03	Auditorium	0.72	0.85	6.43	9.70	0.13	0.06	0.79	−2.04	8.11	69	69
VR04	Auditorium	0.73	0.83	0.85	3.96	0.45	0.28	0.84	−1.66	3.22	70	70
VR05	Concert hall	1.81	2.05	1.15	2.07	0.25	0.12	0.85	−0.22	5.01	71	69
VR06	Concert hall	1.91	1.73	−5.87	−0.54	0.61	0.09	0.76	−0.53	4.82	70	70
VR07	Concert hall	1.40	1.39	−3.49	−0.09	0.66	0.40	0.76	0.97	7.03	70	70
VR08	Concert hall	2.02	2.14	−2.27	−0.34	0.16	0.08	0.83	−0.78	2.91	71	70
VR09	Cathedral	6.92	7.01	−10.16	−9.10	0.49	0.16	0.80	1.38	5.06	71	69

[a]LEV$_{calc}$ and G were calculated before the audio samples were normalized to the same loudness level using the Replaygain algorithm, hence they include the original loudness differences between the rooms.
[b]For room VR01, there was too little energy in the late part of the response to calculate IACCL. Therefore 1-IACCL was set to a value of 0.00 in this case.

TABLE V. An overview of the (virtual) rooms used in listening test d. The first two columns list the room names as well as the type of room after which they were modeled. The table also lists some common, conventional room acoustic parameter values for each room. The last two columns show the SPL of the resulting audio samples used in the listening test, after applying level normalization using the Replaygain algorithm.

Rooms		Room acoustic parameters (conventional)									Sample SPL	
Code	Type	T_{20} (s)	EDT (s)	C_{50} (dB)	C_{80} (dB)	1-IACC$_E$	LF	1-IACC$_L$	LEV$_{calc}$[a] (dB)	G[a] (dB)	Speech (dB)	Cello (dB)
VR10	Concert hall 1 (pos. 1)	1.27	1.42	2.13	3.64	0.19	0.09	0.77	−1.26	4.61	71	70
VR11	Concert hall 1 (pos. 2)	1.21	1.39	1.38	2.79	0.21	0.15	0.72	−1.17	5.16	70	69
VR12	Concert hall 1 (pos. 3)	1.25	1.49	1.03	2.81	0.31	0.16	0.76	−1.23	4.81	71	69
VR13	Concert hall 2	1.75	2.06	0.10	1.19	0.07	0.04	0.65	−0.69	5.97	70	70
VR14	Concert hall 3	1.77	1.83	−0.47	0.54	0.26	0.18	0.69	−0.17	5.89	70	70
VR15	Concert hall 4	1.82	1.72	−0.28	1.77	0.51	0.08	0.95	0.66	5.69	70	70
VR16	Concert hall 5	1.72	1.87	1.39	2.28	0.20	0.15	0.85	−0.06	5.23	70	69
VR17	Concert hall 6	1.98	1.86	−1.37	−0.56	0.25	0.18	0.86	5.21	14.48	70	69

[a]LEV$_{calc}$ and G were calculated before the audio samples were normalized to the same loudness level using the Replaygain algorithm, hence they include the original loudness differences between the rooms.

IX. DISCUSSION

In the previous section, correlation coefficients were presented between listening test results and newly proposed auditory parameters were presented together with data referring to the respective conventional room acoustic parameters. From these correlation coefficients, it seems the new method performs quite well as it often produces the highest correlation coefficient and has less cases of insignificant correlation than the conventional parameters tested. However, these numbers cannot be regarded as a formal proof that the newly proposed parameters outperform the conventional ones, as the number of subjects used in the tests is somewhat low, especially for tests c and d. More tests need to be performed with large number of subjects to get statistically more solid results. However, the results in this paper look promising enough to consider such tests for the future.

The performance of the auditory parameters can be explained by the procedure with which they were obtained, which is closer to the human perception of room acoustics than for the conventional, impulse response-based parameters. To make an accurate prediction of attributes related to room acoustics, it is important to take the properties of the human auditory system—such as auditory filtering and nonlinearities—into account.

A further advantage of the new method is the fact that it can be applied on live recorded data; there is no need to measure impulse responses in empty or even in occupied halls. Measurements can now be performed during a concert by means of dummy head recordings acquired at some selected audience positions. Excerpts of these recordings can be subjected—even continuously—to the auditory model. Thus one could obtain continuous and content-specific representations of the relevant room acoustic attributes.

As a further improvement with respect to conventional impulse response-based parameters, prediction of perceptual attributes can be carried out for different types of programs. For instance, if the hall is unoccupied, and, hence, there will be no natural source available, a conventional loudspeaker can be used to excite the hall with different kinds of music or speech to test the auditorium for different use cases.

Most alternative approaches to room acoustic assessment discussed in the introduction are either focused only on spatial attributes or the visual inspection of processed impulse responses. Moreover, to the knowledge of the authors, the method proposed here is the only auditory

TABLE VI. The correlation coefficients between the listening test results and the conventional parameters for the room acoustic attribute reverberance. Two conventional conventional parameters were evaluated: The reverberation time T_{20} and early decay time EDT. P_{REV} is the auditory parameter related to reverberance as resulting from the method proposed in this paper.

Test (stimulus)	T_{20}	EDT	P_{REV}
a (cello)	0.86	**0.88**	0.76
a (speech)	0.74	0.76	**0.81**
b (cello)	0.79	0.78	**0.95**
b (speech)	0.73	0.73	**0.96**
c (cello)	0.84	0.85	**0.98**
c (speech)	0.79	0.80	**0.96**
d (cello)	**0.96**	0.83	0.87
d (speech)	−0.30[a]	−0.55[a]	**0.88**

[a]Values marked are not significant at the $p < 0.05$ level.

TABLE VII. The correlation coefficients between the listening test results and the conventional parameters for the room acoustic attribute clarity. Two conventional conventional parameters were evaluated: The clarity indices C_{50} and C_{80}. P_{CLA} is the auditory parameter related to clarity as resulting from the auditory method proposed in this paper.

Test (stimulus)	C_{50}	C_{80}	P_{CLA}
a (cello)	0.79	**0.82**	0.79
a (speech)	0.91	**0.94**	0.82
b (cello)	0.91	0.93	**0.96**
b (speech)	0.94	**0.96**	0.87
c (cello)	0.79	0.77	**0.94**
c (speech)	0.87	0.86	**0.90**
d (cello)	0.82	0.88	**0.83**
d (speech)	0.03[a]	0.04[a]	**0.82**

[a]Values marked are not significant at the $p < 0.05$ level.

TABLE VIII. The correlation coefficients between the listening test results and the conventional parameters for the room acoustic attribute ASW. Two conventional conventional parameters were evaluated: One minus the early interaural cross-correlation coefficient ($1 - IACC_B$) and the early lateral energy fraction LF. P_{ASW} is the auditory parameter related to ASW as resulting from the method proposed in this paper.

Test (stimulus)	$1 - IACC_E$	LF	P_{ASW}
a (cello)	0.86	N/A	**0.92**
a (speech)	0.86	N/A	**0.94**
b (cello)	0.67	N/A	**0.86**
b (speech)	0.82	N/A	0.75
c (cello)	0.71	0.33[a]	**0.83**
c (speech)	0.74	0.42[a]	**0.90**
d (cello)	0.40[a]	0.09[a]	**0.64**[a]
d (speech)	0.66[a]	0.21[a]	**0.83**

[a]Values marked are not significant at the $p < 0.05$ level.

motivated methods that predicts *four* of the most important room acoustic attributes. Furthermore, in most former studies, the proposed predictors were not compared to "conventional" impulse response-based parameters, preventing an objective comparison of these approaches to the proposed method.

Besides the field of room acoustics, further applications for the new method are possible. For example, it could be used to assess the quality of holophonic sound field reproduction systems, e.g., for wave field synthesis or (higher order) ambisonics. Instead of measuring impulse responses for each loudspeaker to each receiver location, the simulated sound field can directly be evaluated in terms of its perceptual properties, possibly resembling an approach of higher validity and practical relevance than evaluation using conventional room acoustic parameters. Especially the parameters related to ASW and LEV are of interest because they correspond to spatial aspects of the sound field.

It has to be stressed that because measured parameters are content-specific, documentation of the recording session is crucial. If results for the auditory parameters for a particular room are reported, it is necessary to add the type of signal that was used during the measurement. Preferably, if a room has multiple use cases, it should be tested for different signal types and—if available—for different room acoustic adjustments. For further comparability, e.g., when using electroacoustic stimulation, a common catalog of test signals (solo and orchestra music samples, male and female speech, etc.) should be established.

TABLE IX. The correlation coefficients between the listening test results and the conventional parameters for the room acoustic attribute LEV. Two conventional conventional parameters were evaluated: One minus the late interaural cross-correlation coefficient ($1 - IACC_L$) and Beranek's LEV_{calc}. P_{LEV} is the auditory parameter related to LEV as resulting from the method proposed in this paper.

Test (stimulus)	$1 - IACC_L$	LEV_{calc}	P_{LEV}
a (cello)	0.67	0.64	**0.90**
a (speech)	0.76	0.75	**0.96**
b (cello)	0.84	0.82	**0.94**
b (speech)	0.86	0.84	**0.96**
c (cello)	0.78	0.61[a]	**0.85**
c (speech)	0.78	0.61[a]	**0.93**
d (cello)	0.44[a]	**0.81**	0.69[a]
d (speech)	0.54[a]	0.17[a]	**0.70**[a]

[a]Values marked are not significant at the $p < 0.05$ level.

Finally, more research is needed to test and, possibly, optimize the new method. The test results presented in this paper look promising, but the number of subjects used was small. Future tests should include large numbers of subjects to test the method more formally. For example, a test could be arranged to detect if the newly proposed parameters correlate better with perceptual attributes compared with the conventional ones. The minimum number of subjects needed for such a test could be determined using the G*POWER tool (Faul *et al.*, 2007).

Furthermore, future tests could include various other stimuli compared with the ones used in this paper (male speech and solo cello music). Possible signals are noise, wideband pulses, and other music or speech signals. A variety of signals was already tested informally in (Van Dorp Schuitman, 2011).

X. CONCLUSIONS

In this paper, a binaural auditory model was proposed that can be used to determine auditory parameters to predict relevant attributes of room acoustic perception. The model works on binaural recordings, whereas conventionally room acoustic parameters are determined from room impulse responses. It simulates multiple aspects of human hearing to resemble human auditory perception more closely. Ratings for attributes corresponding to the auditory parameters (reverberance, clarity, apparent source width, and listener envelopment) were collected in four listening tests for various room/stimulus combinations. The results of one test were used for the optimization of the free parameters in the model using a genetic algorithm, and the results of the other three tests served for an informal validation the model. Comparing informal listening test results of conventional and the new auditory parameters show that the new parameters might have the potential to outperform the conventional ones. A future formal proof will require statistically more rigorous listening tests.

ACKNOWLEDGMENTS

This research was supported by grants from the Dutch technology foundation STW, Project No. DTF.7459. The work of A.L. was gratefully supported by a grant from the Deutsche Telekom Laboratories and the Deutsche Forschungsgemeinschaft (Grant No. WE4057/1-1).

Abdel Alim, O. (**1973**). "Abhängigkeit der Zeit- und Registerdurchsichtigkeit von raumakustischen Parametern bei Musikdarbietungen (Dependence of time and register definition of room acoustical parameters with musical performances)," Ph.D. thesis, TU Dresden, Dresden, Germany.

Ando, Y. (**1983**). "Calculation of subjective preference at each seat in a concert hall," J. Acoust. Soc. Am. **74**, 873–887.

Ando, Y. (**2007**). "Concert hall acoustics based on subjective preference theory," in *The Springer Handbook of Acoustics* (Springer Science + Business Media, New York), pp. 351–386.
Barron, M. (**2005**). "Using the standard on objective measures for concert auditoria, ISO3382, to give reliable results," Acoust. Sci. Tech. **26**, 162–169.
Barron, M., and Marshall, A. H. (**1981**). "Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure," J. Sound Vib. **77**, 211–232.
Becker, J. (**2002**). "Spectral and temporal contribution of different signals to ASW analysed with binaural hearing models," in *Proceedings of the Forum Acusticum 2002*, Sevilla, pp. 1–6.
Beranek, L. L. (**1996**). *Concert and Opera Halls—How They Sound* (Acoustical Society of America, New York), pp. 643.
Beranek, L. L. (**2008**). "Concert hall acoustics—2008," J. Audio Eng. Soc. **56**, 532–544.
Berkhout, A. J. (**1988**). "A Holographic approach to acoustic control," J. Audio Eng. Soc. **36**, 977–995.
Bilsen, F. A. (**1994**). "Binaural modelling of perceptual qualities in room acoustics," in *Proceedings of the International Conference on Acoustic Quality of Concert Halls*, Madrid, pp. 73–88.
Blauert, J., and Lindemann, W. (**1986a**). "Auditory spaciousness: Some further psychoacoustic analyses," J. Acoust. Soc. Am. **80**, 533–542.
Blauert, J., and Lindemann, W. (**1986b**). "Spatial mapping of intracranial auditory events for various degrees of interaural coherence," J. Acoust. Soc. Am. **79**, 806–813.
Bradley, J. S., and Soulodre, G. A. (**1995**). "The influence of late arriving energy on spatial impression," J. Acoust. Soc. Am. **97**, 2263–2271.
Breebaart, D. J. (**2001**). "Modeling binaural signal detection," Ph.D. thesis, Eindhoven University of Technology, Eindhoven, The Netherlands.
Breebaart, D. J., van de Par, S., and Kohlrausch, A. (**2001**). "Binaural processing model based on contralateral inhibition. I. Model structure," J. Acoust. Soc. Am. **110**, 1074–1088.
Chevret, P., and Parizet, E. (**2007**). "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, pp. 1–5.
Damaske, P., and Ando, Y. (**1972**). "Interaural cross correlation for multichannel loudspeaker reproduction," Acustica **27**, 232–238.
Dau, T., Püschel, D., and Kohlrausch, A. (**1996a**). "A quantitative model of the 'effective' signal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am. **99**, 3615–3622.
Dau, T., Püschel, D., and Kohlrausch, A. (**1996b**). "A quantitative model of the "effective" signal processing in the auditory system. II. Simulations and measurements," J. Acoust. Soc. Am. **99**, 3623–3631.
Dewhirst, M., Conetta, R., Rumsey, F., Jackson, P., Zielinski, S., Meares, D., Bech, S., and George, S. (**2008**). "QESTRAL (Part 4): Test signals, combining metrics and the prediction of overall spatial quality," in *Proceedings of the 125th AES Conference*, San Fransisco, Vol. 7598, pp. 1–8.
EBU (**1988**). "Sound quality assessment material. Recordings for subjective tests," in *Users Handbook for the EBU Demonstration CD, SQAM*, EBU Document Technical Report No. 3253-1988, European Broadcasting Union.
Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (**2007**). "G*POWER 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences," Behav. Res. Methods **39**, 175–191.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Griesinger, D. (**1992**). "Room impression, reverberance, and warmth in rooms and halls," in *Proceedings of the 93th AES Convention*, San Francisco, Vol. 3383, pp. 1–21.
Griesinger, D. (**1995**). "How loud is my reverberation?" in *Proceedings of the 98th AES Convention*, Paris, Vol. 3943, pp. 1–12.
Griesinger, D. (**1997**). "The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces," Acta Acust. Acust. **83**, 721–731.
Griesinger, D. (**1999**). "Objective measures of spaciousness and envelopment," in *Proceedings of the 16th International AES Conference*, Vol. 16–003, pp. 1–15.
Hidaka, T., Yamada, Y., and Nakagawa, T. (**2007**). "A new definition of boundary point between early reflections and late reverberation in room impulse responses," J. Acoust. Soc. Am. **122**, 326–332.
Holland, J. (**1975**). *Adaptation in Natural and Artificial Systems* (The University of Michigan Press, Ann Arbor, MI), pp. 183.
ISO (**2009**). ISO 3382-1:2009, *Acoustics—Measurement of Room Acoustic Parameters—Part 1: Performance Spaces* (International Organization for Standardization, Geneva).
Jackson, P., Dewhirst, M., Conetta, R., Zielinski, S., Rumsey, F., Meares, D., Bech, S., and George, S. (**2008**). "QESTRAL (Part 3): System and metrics for spatial quality prediction," in *Proceedings of the 125th AES Conference*, San Francisco, Vol. 7597, pp. 1–9.
Jeffress, L. A. (**1948**). "A place theory of sound localization," J. Comp. Physiol. Psychol. **41**, 35–39.
Jordan, V. L. (**1970**). "Acoustical criteria for auditoriums and their relation to model techniques," J. Acoust. Soc. Am. **47**, 408–412.
Keet, W. d. V. (**1968**). "The influence of early lateral reflections on spatial impression," in *Proceedings of the 6th International Congress on Acoustics (ICA 68)*, Tokyo, pp. E53–E56.
Kuttruff, H. (**2000**). *Room Acoustics*, 4th ed. (Spon Press, London), 368 pp.
Lehmann, P., and Wilkens, H. (**1980**). "Zusammenhang subjektiver Beurteilung von Konzertsälen mit raumakustischen Kriterien (Relation between subjective assessment of concert halls and room acoustic criteria)," Acustica **45**, 256–268.
Lindemann, W. (**1986a**). "Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals," J. Acoust. Soc. Am. **80**, 1608–1622.
Lindemann, W. (**1986b**). "Extension of a binaural cross-correlation model by contralateral inhibition. II. The law of the first wave front," J. Acoust. Soc. Am. **80**, 1623–1630.
Lokki, T., and Karjalainen, M. (**2002**). "Analysis of room responses, motivated by auditory perception," J. New Music Res. **31**, 163–169.
Marshall, A. H. (**1967**). "A note on the importance of room cross-section in concert halls," J. Sound Vib. **5**, 100–112.
Mason, R. (**2002**). "Elicitation and measurement of auditory spatial attributes in reproduced sound," Ph.D. thesis, University of Surrey, Guildford, U.K.
Mason, R., Brookes, T., and Rumsey, F. (**2004**). "Development of the interaural cross-correlation coefficient into a more complete auditory width prediction model," in *Proceedings of the 18th International Congress on Acoustics (ICA 2004)*, Kyoto, pp. 1–4.
Meddis, R. (**1988**). "Simulation of auditory neural transduction: Further studies," J. Acoust. Soc. Am. **83**, 1056–1063.
Middlebrooks, J. C. (**1999**). "Individual differences in external-ear transfer functions reduced by scaling in frequency," J. Acoust. Soc. Am. **106**, 1480–1492.
Morimoto, M. (**2002**). "The relation between spatial impression and the precedence effect," in *Proceedings of the 8th International Conference on Auditory Display (ICAD2002)*, Kyoto, pp. 1–10.
Nelson, P., Park, M., Takeuchi, T., and Fazi, F. (**2008**). "Binaural hearing and systems for sound reproduction," in *Proceedings of Acoustics'08*, Paris, pp. 3531–3536.
Okano, T., Beranek, L. L., and Hidaka, T. (**1998**). "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**, 255–265.
Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (**1992**). "Complex sounds and auditory images," in *Proceedings of the 9th International Symposium on Hearing*, pp. 429–446.
Reichardt, W., Abdel Alim, O., and Schmidt, W. (**1975**). "Definition and basis of making an objective evaluation to distinguish between useful and useless clarity defining musical performances," Acustica **32**, 126–137.
Robinson, D. (**2001**). "Replay Gain—A proposed standard," www.replaygain.org (Last viewed 5/4/2011).
Rumsey, F. (**2002**). "Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm," J. Audio Eng. Soc. **50**, 651–666.
Rumsey, F., Zielinski, S., Jackson, P., Dewhirst, M., Conetta, R., George, S., Bech, S., and Meares, D. (**2008**). "QESTRAL (Part 1): Quality evaluation of spatial transmission and reproduction using an artificial listener," in *Proceedings of the 125th AES Convention*, San Francisco, Vol. 7595, pp. 1–8.
Schmitz, A. (**1995**). "Ein neues digitales Kunstkopfmeßsystem (A new digital artificial head measuring system)," Acustica **81**, 416–420.
Schroeder, M. R., Gottlob, D., and Siebrasse, K. F. (**1974**). "Comparative study of European concert halls: Correlation of subjective preference with geometric and acoustic parameters," J. Acoust. Soc. Am. **56**, 1195–1201.

Sotiropoulou, A. G., and Fleming, D. B. (**1995**). "Concert hall acoustic evaluation by ordinary concert-goers: II. Physical room acoustic parameters subjectively significant," Acustica **81**, 10–19.
Sotiropoulou, A. G., Hawkes, R. J., and Fleming, D. B. (**1995**). "Concert hall acoustic evaluation by ordinary concert-goers: I. Multi-dimensional description of evaluations," Acustica **81**, 1–9.
Soulodre, G. A. (**2006**). "Can reproduced sound be evaluated using measures designed for concert halls?" in *Proceedings of the Workshop on Spatial Audio and Sensory Evaluation Techniques*, Guildford, U.K., pp. 1–11.
Soulodre, G. A., and Bradley, J. S. (**1995**). "Subjective evaluation of new room acoustic measures," J. Acoust. Soc. Am. **98**, 294–301.
Soulodre, G. A., Lavoie, M. C., and Norcross, S. G. (**2003a**). "Objective measures of listener envelopment in multichannel surround systems," J. Audio Eng. Soc. **51**, 826–840.
Soulodre, G. A., Lavoie, M. C., and Norcross, S. G. (**2003b**). "Temporal aspects of listener envelopment in multichannel surround systems," in *Proceedings of the 114th AES Conference*, Amsterdam, Vol. 5803, pp. 1–9.
Supper, B. (**2005**). "An onset-guided spatial analyser for binaural audio," Ph.D. thesis, Institute of Sound Recording, University of Surrey, Guildford, U.K.
Terhardt, E. (**1979**). "Calculating virtual pitch," Hear. Res. **1**, 155–182.
Van Dorp Schuitman, J. (**2011**). "Auditory modelling for assessing room acoustics," Ph.D. thesis, Delft University of Technology, Delft, The Netherlands.
Wightman, F. L., and Kistler, D. J. (**1992**). "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648–1661.