Audio Engineering Society
Convention Paper

Presented at the 128th Convention
2010 May 22–25 London, UK

Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses

Alexander Lindau[1], Linda Kosanke[1], and Stefan Weinzierl[1]

[1] Audio communication group, Technische Universität Berlin, Einsteinufer 17c, 10587 Berlin, Germany

Correspondence should be addressed to Alexander Lindau (alexander.lindau@tu-berlin.de)

ABSTRACT

The mixing time of room impulse responses denotes the moment when the diffuse reverberation tail begins. A diffuse sound field can physically be defined by 1) equidistribution of acoustical energy and 2) a uniform acoustical energy flux over the complete solid angle. Accordingly, the perceptual mixing time is the moment when the diffuse tail cannot be distinguished from that of any other position in the room. This provides an opportunity for reducing the length of binaural impulse responses that are dynamically exchanged in virtual acoustic environments (VAEs). Numerous model parameters and empirical features for the prediction of perceptual mixing time in rooms have been proposed. This study aims at a perceptual evaluation of all potential estimators. Therefore, binaural impulse response data sets were collected with an adjustable head and torso simulator for a representative sample of rectangularly shaped rooms. Prediction performance was evaluated by linear regression using results of a listening test where mixing times could be adaptively altered in real time to determine a just audible transition time into a homogeneous diffuse tail. Regression formulae for the perceptual mixing time are presented, conveniently predicting perceptive mixing times to be used in the context of VAEs.

1. INTRODUCTION

In the context of Virtual Acoustic Environments (VAEs), using binaural room impulse responses to be convolved with anechoic source signals, room impulse responses are usually considered to comprise three successive parts: direct sound, early reflections, and a tail of stochastic reverberation ([1], [2]). The transition point between early reflections and the stochastic reverberation tail is called the physical mixing time (t_m) [3]. There is a consensus in literature that perceptual sensitivity for temporal and spectral fine structure of room impulse responses decreases during the decay process ([4], [5]). Due to increasing reflection density and diffuseness of the decaying sound field individual reflections become perceptively less distinguishable.

Computational demands for VAE systems will be reduced with the amount of early reflections that have to be rendered. An obvious method to achieve this reduction would thus be to replace the individual reverberation tail of the BRIRs – after an instant in time when *perceptual* discrimination is no longer possible – with an arbitrary and constant reverberation tail. In the following, this instant will be referred to as the perceptual mixing time $t_{mp.}$

Already in early publications on dynamic auralization ([4]. [6]) it was proposed to split the convolution process into a time variant convolution of the early impulse response parts and a static convolution with an arbitrary reverberation tail. Thereafter, time aligned outputs of both convolution processes simply would have to be summed up appropriately. In [4], for precalculated BRIRs of a concert hall, this split point was set to 4000 samples resulting in a transition into the static reverberation at 83 ms.

A first formal evaluation of the perceptual mixing time was conducted in [4]. For a small room (V = 185 m³, RT = 0.72 s), t_{mp} was examined for different combinations of early reflections and tails. The authors determined mixing times for a) interchanged tails of the two ears, b) tails from different receiver positions while keeping same source direction and distance, c) tails from the same receiver position but different directions of sound incidence, and finally d) tails from the different receiver positions *and* different directions of sound incidence. Stimuli were convolved with accordingly manipulated binaural impulse responses, resulting in static auralization for presentation. For a listening test, the method of constant stimuli was used. Results lead to the conclusion that – for this room – t_{mp} was about 40 ms and independent from all tested position changes.

Since higher mixing times can be expected for larger rooms, we assessed t_{mp} for a large auditorium (V 8500 m³, RT 2 s), also using static auralization [7]. The perceptual mixing time was indeed higher (up to 140 ms). In accordance with [5] we found no effect of taking a tail from the same receiver position but different direction of sound incidence (i.e. from different head orientations). However, the effect of taking the tail from different *source* position *and* different direction of sound incidence (not tested in [5]) led to considerably increased perceptual mixing times. The different reflection pattern (when stimulated from a different source position) and the different low frequency, modal behavior of the room, included in full range BRIRs, might provide an explanation for this effect. Additionally, it turned out that listeners were most sensitive to a premature transition into the diffuse tail, when a specific drum sample with strong transient behaviour was used.

The aim of our present study was to find the perceptual mixing time for different rooms, its dependence on volume and average absorption, while utilizing state of the art dynamic auralization. Furthermore, by means of regression analysis, several predictors of the physical mixing time – some statistically motivated and some based on analysis of impulse responses – were examined for their ability to predict the perceptual mixing time.

1.1. Diffusion in decaying sound fields

In the sound field of rooms, the transition from early reflections into the stochastic reverberation tail is a gradual one. Every time, a sound wave hits a wall, it is reflected. Depending on the surface properties, this reflection can be specular, partly or fully diffuse. In an ideally diffuse reflecting room, the sound energy continuously spreads about the whole volume in time. At any point in the enclosure, the pulse density grows quadratically with time [8], whereas the energy of the pulses – due to the increasing distances covered – decreases quadratically. While both functions compensate each other, the observable exponential net energy loss is caused by the absorption losses of the enclosure [2]. Finally, the ideal diffuse sound field is characterized by a uniform angular distribution of sound energy flux and a constant acoustical energy density over the whole space [9]. Kuttruff ([8], pp. 120) showed, that complete diffusion can not be reached in a real room with a fraction of specular reflecting surfaces, although, due to even slightest wall anomalies, in reality diffusion must be constantly increasing, because with every specular wall reflection a small amount of diffuse reflection possibly arises, whereas the reverse process does not occur. As shown by Pollack [3], absorbing rooms can never be perfectly diffuse, because due to absorbing walls, there always remains a net energy flow in the direction of the losses (i.e. towards the walls). In summary, it can be stated that perfect diffusion or mixing is an idealization never fully encountered in real rooms.

1.2. The concept of physical mixing time

In terms of particle trajectories 'mixing' means that over time position and direction of two initially adjacent rays have become statistically independent. The duration from the moment of exiting the room with an infinitely short pulse until it is totally mixed is called the (physical) mixing time [3]. A requirement for a room to become mixed is ergodicity, which means that the statistical behavior at all points in the space equals that at one point over time (time average equals ensemble average, [3], [10]). Then, a particle is virtually spending equal time at any point in space while traveling in any direction equally often. Another description is to say, that the sound field has completely lost any memory of its initial state. Ergodicity was shown to be dependent on the shape of the enclosure and the surface reflection properties [11]. Examples for non-ergodic rooms are perfectly rectangular non-diffusing rooms (particle directions remain deterministic) or non-diffusing spherical rooms (due to focusing not all positions will be reached by a particle).

Just like the concept of ideal diffusion (cf. 1.1), the concept of mixing is implicitly confined to a frequency range where the theory of geometrical and statistical acoustics apply. In real rooms, these assumptions are violated most obviously in the low-frequency range. With pronounced modal behavior, the concept of mixing is not appropriate.

The process of diffusion or mixing within the decaying room impulse response may further be disturbed in rooms with non-uniform distribution of large areas with highly varying absorption coefficients (for instance, when large windowpanes are combined with highly absorbing audience seats). Also coupled rooms, highly regular rooms with only little diffuse reflecting boundaries, highly damped rooms, and very small rooms may lack mixing in their decay.

Obviously, the duration of the diffusion process, i.e. the mixing time is maximized for larger rooms, as here, the duration between each reflection is extended due to large free path lengths. This is especially pronounced if the room is lacking any diffusing obstacles [8].

1.3. The relation between physical and perceptual mixing time

Perceptively, a room can be considered to be mixed if the stochastic decay process at one position in the enclosure cannot be distinguished from that of any other position. Ideally, this instant would coincide with the physical mixing time, as the sound field will not contain any position dependent directional or energetic information anymore. However, auditory or cognitive suppression effects as level-, direction-, and duration-dependent instantaneous and noninstantaneous masking or the precedence effect will affect each singular reflection's audibility. Moreover, the audio content was shown to play a vital role in discriminating room reflections [12]. Thus, a relation between perceptual and physical mixing time cannot be trivially inferred.

1.4. Model predictors of mixing time

Several predictors for the perceptual mixing time t_{mp} in rooms have been suggested in literature. Ad hoc values as for instance 50 [13], or 80 ms ([4], [14]) have been proposed regardless of further room properties. From 'practical observation' of reflectograms of concert halls, Reichardt & Lehmann [15] concluded that statistical behavior begins at ca. 150-200 ms. According to Kuttruff [16], for a 'hall of some size' the sound field can be regarded as diffuse after 100-150 ms. According to Hidaka et al. [17] – based on results of two symphony halls – the time separating early and late reflections lies in a range from 50-200 ms.

More elaborate predictors for t_{mp} put an estimation of the room's actual reflection density in relation to the time resolution of the auditory system. The well-known reflection density formula derived from the mirror source model of the rectangular room reads [18]:

$$\frac{dN}{dt} = \frac{4\pi \cdot {c_0}^3 \cdot t^2}{V} \qquad (1)$$

with c_0 being the sound velocity in [m/s] and V the room volume in [m³]. According to Schroeder [19], a reflection density of 1000 s^{-1} would be sufficient to generate a perceptively 'flutterfree' reverberation tail. Ruback & Johansen [20] suggested 4000 s^{-1} to provide a good quality, while Griesinger [21] promoted up to 10.000 and more reflections for high quality reverberation algorithms.

Schreiber [22] determined a just distinguishable reflection density of 2000 s^{-1} experimentally. Thus, in setting dN/dt = 2000 s^{-1}, and solving for t in [18] the relation

$$t_{mp} = 2\sqrt{V} \text{ , with } t_{mp} \text{ in [ms]} \quad (2)$$

was derived. According to ([23], [1]) a reflection density providing at least ten reflections within a characteristic time resolution of the auditory system assumed to be 24 ms would be sufficient. Reflection density dN/dt would thus be 400 s^{-1}. Using this value, the perceptual mixing time would approximately be equal to

$$t_{mp} = \sqrt{V} \text{ , with } t_{mp} \text{ in [ms].} \quad (3)$$

Reichardt & Lehmann [15] proposed the same formula, while Schmidt and Ahnert [24] assumed only five reflections within 20 ms (corresponding to dN/dt = 250 s^{-1}) to be perceptively sufficient.

Ruback & Johansen [20] introduced a different view while discussing the instant of perceptual mixing with respect to the concept of mean free path length l_m:

$$l_m = 4\frac{V}{S} \text{ [m],} \quad (4)$$

where S is the total surface area of the enclosure in [m²]. The rationale of this approach is that the sound field is assumed to be virtually diffuse if every sound particle has on average undergone at least some (e.g. [20]: four) reflections. Thus the equation for t_{mp} reads:

$$t_{mp} = 4l_m\frac{10^3}{c_0} = 4\cdot\left(\frac{4V}{S}\right)\cdot\frac{10^3}{c_0} = 47\cdot\frac{V}{S} \text{ , [ms].} \quad (5)$$

Recently, Hidaka et al. [25] proposed a linear regression formula that will fit their results from a larger study on physical mixing times determined empirically from impulse responses. The study included impulse responses from 59 concert halls of different shapes and sizes. The formula predicts the results for the 500 Hz octave band from the room's reverberation time

$$t_{m500Hz} = 80\cdot RT_{500Hz} \text{ [ms]} \quad (6)$$

Obviously, all these equations are depending on solely three room specific quantities: volume, surface area and reverberation time. They can therefore further be generalized. The reflection density equations (2) and (3) simplifiy to

$$t_{mp1} = k_{refl}\cdot\sqrt{V} \text{ ,} \quad (7)$$

the mean free path length estimation similarly reads

$$t_{mp2} = k_{path}\cdot\frac{V}{S} \text{ ,} \quad (8)$$

and the estimation from reverberation time can be rewritten as

$$t_{mp3} = k_{reverb}\cdot RT_i \text{ .} \quad (9)$$

Thus, three relations for the prediction of the mixing times remain, to be subjected to later evaluation.

1.5. Empirical predictors of physical mixing time

Recently, several algorithms have been proposed for calculating the physical mixing time from empirical room impulse responses. Four of these approaches will be shortly introduced, as they will also be evaluated perceptually later on.

Abel & Huang (2006)

Abel & Huang [27] proposed an approach based on the assumption that the sound pressure amplitudes in a reverberant field take on a Gaussian distribution. For determining the mixing time, a so-called 'echo density profile' is calculated. Therefore, with a short sliding rectangular window of 500-2000 samples, the empirical standard deviation of the sound pressure amplitudes is calculated for each sample index. In order to determine, how well the empirical amplitude distribution approximates a Gaussian behavior, the proportion of samples outside the empirical standard deviation is determined and compared to the proportion, which is expected for a Gaussian distribution. With increasing time and diffusion, this echo density profile should increase until it finally – at the instant of complete diffusion – reaches the value of one. With larger window sizes, the gross shape of the echo density profile stays similar, some smoothing can be observed. We chose a window of 2^{10} samples (23 ms) – as suggested by the authors from discussion of auditory time resolution – and of rectangular shape. The mixing time can be defined as the instant where the echo density profile becomes unity for the first time (criterion I). In order to account for minor fluctuations Abel &

Huang further refined this criterion to account for the instant when the reflection density is within 1-σ_{late} (σ_{late} being the standard deviation of the late echo density, criterion II). We evaluated both stopping criteria, while calculating σ_{late} from the last 20% of the impulse responses before reaching the noise floor. The authors did not evaluate the perceptual relevance of this criterion.

Stewart & Sandler (2007)

Following an idea proposed already in [27] Stewart & Sandler [28] suggested measuring the kurtosis of the sound pressure amplitudes and comparing this value to that expected for a Gaussian distribution. As second order cumulant, kurtosis γ_4 is a measure of the "non-Gaussianess" contained in a signal. In the normalized form, γ_{4n} is given by:

$$\gamma_4 = \frac{\mathrm{E}(x-\mu)^4}{\sigma^4} - 3 . \qquad (10)$$

where E() is the expectation operator, μ is the mean, and σ is the standard deviation of the process. For increasingly Gaussian processes, the normalized kurtosis must approach zero. We calculated this instant for using identical settings as for the echo density profile. As it was not clearly stated in [28], from the authors discussion of the figures, we chose as the mixing time the instant when the kurtosis γ_{4n} (normalized to a maximum value of one) reached zero for the first time. So far, authors did not investigate the perceptual relevance of this criterion.

Hidaka et al. (2007)

Hidaka et al. [25] proposed a new approach for the estimation of the instant when a room impulse response has become diffuse. Here, the mixing time is derived in the frequency domain. Therefore, the time-frequency energy distribution of the impulse response p(t) is calculated according to

$$E(t,\omega) = \left| \int_t^\infty p(t) e^{j\omega t} dt \right|^2 . \qquad (11)$$

When averaging over a frequency range $\Delta\omega$, (11) can be shown to be identical to the Schroeder integration [26]. The energy distribution $E(t,\omega)$ is calculated for impulse responses beginning with the direct sound; initial delays have to be removed in advance. With increasing time t, $E(t,\omega)$ will continuously contain less early reflections and increasingly more stochastic reverberation. As a second step, Pearson's product-moment correlation r(t) is calculated as a continuous function of time for $E(0{:}\infty, \Delta\omega)$ and $E(t{:}\infty, \Delta\omega)$ in arbitrary frequency bands. This will describe the similarity between a) the energy decay process including the initial state and b) the energy decay process with beginning from any time t afterwards in one particular frequency band. Hidaka defines the 'transition time' into stochastic reverberation as the instant when r(t) has become sufficiently small. He cites several references supporting that $r = e^{-1} = 0.367$ is a widely used value to categorize low correlations when treating stochastic processes. Thus, we calculated $E(t,\omega)$ and r(t) for octave bands from 125 Hz to 16 kHz, and detected the mixing time at the moment when $r(t) \leq 0.367$ for the first time. For ease of computation we limited the time resolution to 100 samples ($\Delta t = 2.3$ ms).

Defrance et al. (2009)

Most recently, Defrance et al. [29] suggested a fourth procedure for estimating the physical mixing time from room impulse responses. The rationale behind this method is explained here in short. After excitation with a pulse, the cumulative number of reflections at the observing point in an enclosure is a cubic function of time. This relation can be derived from integration of equation (1) with respect to time

$$N = \frac{4\pi}{3} \cdot \frac{{c_0}^3}{V} \cdot t^3 . \qquad (12)$$

With time, reflection density becomes so large, that single reflections begin to overlap and cannot be distinguished anymore. This is said to be equivalent to the sound field becoming diffuse. Once this critical density is reached, the cumulative function of arrivals should change from a cubic to a linear increase. In order to detect this instant, Defrance et al. use a decomposition technique ('Mixing Pursuit') somewhat similar to wavelet decomposition to detect every singular reflection (called 'arrivals') in the impulse response. As wavelets (here: 'atoms') only the direct sound itself is used ('mother atom'), as for convenience it is supposed that all reflections are more or less copies of the direct sound impulse. Decomposition is conducted by correlating the impulse response with the direct sound atom while shifting the latter along the

impulse response to all possible instances in time. Correlation is calculated as the magnitude of the inner product. At the instance of maximum correlation, the direct sound atom is subtracted from the IR weighted by the corresponding inner product. Weighting coefficients and corresponding instances are saved forming a time vector of decomposition coefficients. The decomposition process is repeated until the energy ratio SRR (signal residual ratio) of the reconstructed signal (reconstructed from the mother atom and the time vector of coefficients) and remaining impulse response signal (the residuum) rises above a certain value. Care has to be taken, as the energy decay of empirical impulse responses leads to a decomposition that wrongly favors the early parts. As energy density w in a room decays according to

$$w(t) = w_0 e^{\frac{c_0 S}{4V} t \ln(1-\alpha)}, \qquad (13)$$

with the mean absorption coefficient α calculated from Sabine's formula and the reverberation time RT an approximate inverse decay function $w^{-1}(t)$ can be derived. This has to be applied to the impulse responses before running the decomposition. Finally, the mixing time is determined from the reconstructed signal using a reflection distance criterion. Defrance et al. argue, that the instant when the slope of the cumulated arrival function changes from cubic to linear, would be equivalent to the moment were the first two reflections are spaced equal or less than the equivalent duration of the direct sound. This duration could be calculated just like the equivalent statistical bandwidth [30] while replacing the frequency spectrum with the time signal (i.e. with the mother atom). Thanks to support by the authors, we could calculate the Mixing Pursuit decomposition using their original Matlab® code. Additionally, we implemented the energy decay compensation and the calculation of the equivalent duration of the direct sound. According to recommendation in [29] we used a SRR of 5dB as stopping criterion for all decompositions. The authors further stated that a perceptive evaluation of their measure is currently in preparation.

2. METHODS

In section 2.1 the motivation for selecting the rooms for the listening tests will be explained, and the rooms will be introduced. In section 2.2, binaural measurement setups will be described, and in section 2.3 the listening test is explained.

2.1. Room selection

The main purpose of this study was to find reliable predictors for the perceptual mixing times in typical musical performance environments to be used for auralization. Hence, the selected rooms should resemble a representative subset of venues for musical performance. The physical mixing times derived in [25] for a large selection of concert halls were at maximum for shoebox shaped rooms. Moreover, from the theory of mixing rooms (cf. section 1.1 to 1.3), their regular shape and their long unobstructed path lengths suggests them to be most critical in terms of late physical mixing times. Therefore, we confined this study to largely rectangular rooms. Coupled enclosures as commonly encountered in theater or opera houses were avoided. Wall surface materials varied from only little diffusing concrete, glass or gypsum to considerably structured wood panels. Floors were made from linoleum, parquet or granite. The smaller (lecture) rooms were equipped with chairs and tables, whereas all the larger rooms included extended audience seating areas.

We selected nine rooms while aiming at a systematic variation of both volume and average absorption coefficient (α_{avg}), each in three increments (cf. Table 1). This so-called complete variation will permit an independent assessment of both influences through analysis of variances for repeated measures while at the same time reducing the amount of test participants needed. When further assuming perceptual equidistance of the threefold quantization of each parameter, it will allow for testing linear and quadratic relations between the physical parameters and the perceptual mixing time.

Due to physical interrelation, it is difficult to vary room volume independently from *absolute* amount of reverberation. However, while varying the average absorption coefficient, we could at least assess the influence of the *relative* amount of reverberation independently of volume. Step sizes of reverberation were chosen to exceed at least a just noticeable difference of 10%. The most important parameters of the selected rooms are shown in Table 1.

Additionally, Figure 1 shows true to scale floor plans of the rooms. The three small rooms were, in order of increasing reverberation time, the Electronic Studio of

the TUB (room 1), and two small lecture rooms, EN-111 and EN-190 (room 2 and 3).

	small V	medium V	large V	α_{avg} (RT)
large α (RT)	***room 1*** 216 m³ α: 0.36 (0.39 s)	***room 4*** 3300 m³ α: 0.28 (1.15 s)	***room 7*** 8298 m³ α: 0.33 (1.52 s)	**0.32** (1 s)
medium α (RT)	***room 2*** 224 m³ α: 0.26 (0.62 s)	***room 5*** 5179 m³ α: 0.23 (1.67 s)	***room 8*** 8500 m³ α: 0.23 (2.08 s)	**0.24** (1.45 s)
small α (RT)	***room 3*** 182 m³ α: 0.17 (0.79 s)	***room 6*** 3647 m³ α: 0.2 (1.83 s)	***room 9*** 7417 m³ α: 0.23 (2.36 s)	**0.2** (1.66 s)
avg. Vol.	**207 m³**	**4042 m³**	**8072 m³**	

Table 1 Volume, average absorption coefficient, and reverberation time of the nine selected rooms

The medium size rooms were the lecture hall H-104, where the large TUB WFS system is situated (room 4), the TUB lecture hall HE-101 whose acoustic consultant was L. Cremer (room 5) and the large recording room of the Teldex Studio Berlin (room 6). The three large venues were the concert hall of the University of Arts in Berlin (room 7), the auditorium maximum of TUB (room 8), and the Jesus-Christus Church (Berlin-Dahlem). Rooms 4 and 8 are – though primarily used as lecture halls – regularly utilized as performance spaces as well.

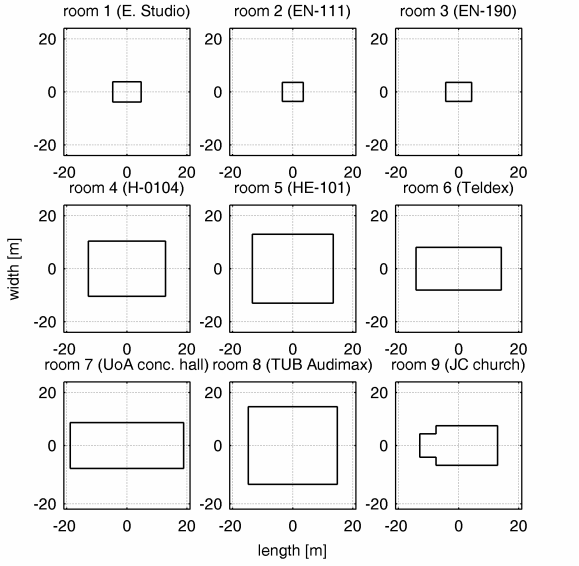

Figure 1 Floor plans of the nine rooms (true scale)

Regarding the positions of all rooms within the 3 x 3 grid of volume and average absorption reverberation time – at least for a selection of real world rooms – the grid is quite regular (cf. Table 1). Only the average absorption coefficient room nine (JC church) was a bit too high, though still exhibiting the largest reverberation time in the field.

2.2. Binaural measurements

In order to provide high quality dynamic binaural simulations of the mentioned environments, we conducted measurements of binaural room impulse responses using the automatic head and torso simulator FABIAN [7] in all nine rooms. As sound source, a 3-way dodecahedron loudspeaker, allowing for high signal to noise ratio and optimal omnidirectional directivity, was placed in the middle of the stage, which was typically located at one of the narrow ends of the rooms. For a wide frequency range of BRIRs, the loudspeaker was equalized to exhibit, within ± 3 dB, an omnidirectional and linear frequency response from 40 Hz to 17 kHz. The three major room dimensions length, width, height were measured using a laser distance meter for calculating room volume. Standard room impulse responses were measured at three different positions in the diffuse field using an omni directional microphone. Reverberation times, as displayed in Table 1, were averaged over octave bands from 125 Hz to 4 kHz, and all three measurement positions. The critical distance was calculated, and FABIAN was seated on a place on the room's longitudinal symmetry axis at twice the critical distance while directly facing the loudspeaker. According to Kuttruff [31] a random sound field can be expected in a distance of at least two times the critical distance. BRIRs were collected for horizontal head orientations within ±80° in angular steps of 1°.

2.3. Listening Test

Perceptual mixing times were determined using an adaptive 3-AFC listening test procedure. Subjects had to discriminate manipulated dynamic binaural simulations from the original ones. In the 'original' simulation, the complete BRIRs were updated in real time according to head movements. In the manipulated simulation, only the early part of the BRIR corresponded to the subject's true head position. The convolution results of the late reverberation tail – taken from the BRIR corresponding to frontal head orientation – were concatenated to the early output but not changed otherwise (cf. Figure 2).

The reverberant tail of the neutral head orientation was taken for practical reasons. Although a dependence on the source-receiver setups was not found in [5], in [7] at least an effect of different source positions was shown (cf. section 1). Also in the current study, when assessing reverberant tails from different measurement positions in pretests, and apparently due to modal behavior, severe low frequency differences between original and manipulated tails became audible, leading to high perceptual mixing times. As discussed in section 1.2 with modal behavior the concept of mixing is not applicable.

In the listening test, the early and late BRIR parts could be concatenated at arbitrary instants in increments of 5.8 ms (small rooms, no. 1 to 3), and 11.6 ms (medium and large rooms, no. 4 to 9) respectively. The range of different cross fade instants (i.e. mixing times) presentable within the adaptive listening test, was determined in pretests. Mixing times of up to circa 100 ms were found to be sufficient for rooms 1 to 3, and 200 ms for rooms 4 to 9 respectively. Linear cross fading into reverberation tail was realized within the fast convolution engine using a window of a size equal to the step size.

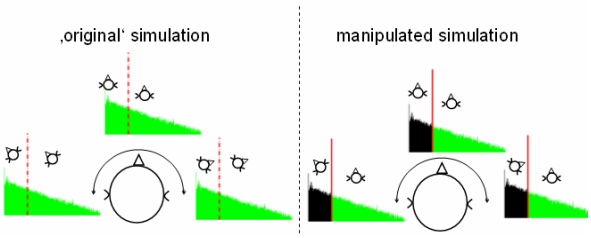

Figure 2 Abstraction of situations to be compared within the adaptive listening test, the concatenation point between early and late parts of the BRIRs (right plot) could be altered adaptively

For each room BRIRs were rendered with different lengths, determined by the rooms' energy decay ratio and the beginning of the noise floor. On average, noise floor was below -80 dB relative to the direct sound level. Nevertheless, we limited the length of BRIRs to about two third of the duration of the decay to noise floor. This was done, because noise floor of different BRIRs exhibited slightly different spectral coloration, which was audible when AB-comparing original and manipulated situation and would thus have biased the detection task. Hence, BRIRs had a length between 14.000 (i.e. 0.375 s for room 1) and 100.000 (i.e. 2.26 s for room 9) samples, thus maintaining at least 60 dB decay for binaural simulations of all rooms.

In order to be able to test medium size first order interaction effects at p=0.05 with 80% test power a sample of 22 different subjects per room would have been needed. Error variance is reduced by using a fully repeated measures design, where each subject is tested under all conditions. Therefore, with same sample size, smaller effects can be tested; alternatively, fewer subjects are needed to test the same effect size. However, for calculation of the reduced number of subjects the mean correlation ρ_{mean} between all subjects has to be known in advance. If ρ_{mean} it is not known, it is usually assumed to be 0.5 [32]. We estimated ρ_{mean} from our pretest results obtained with four subjects to be 0.15. Thus, at least 19 subjects were needed to obtain the above-mentioned test power in a repeated measures design. After the listening test, the mean correlation between the subjects considered for further analysis (cf. section 3) turned out to be $\rho_{mean} = 0.17$, thus proving our estimation to be sufficiently correct.

Twenty-four participants (3 female, 21 male) with an average age of 28.3 took part in the listening test. Most of the subjects had a musical education background, and all had participated in listening tests before. During training subjects were instructed to rotate their head widely for maximizing the difference between original and manipulated reverberation tails. As stimulus, the critical drum set sample from [7] was used again (length: 2.5 s without reverb tail).

Loudness differences between simulated rooms were minimized through normalization of BRIR datasets. Electrostatic headphones (STAX SR-2050II) were frequency compensated using fast frequency deconvolution with high pass regularization from measurements on our dummy head FABIAN [33]. Subjects were allowed to adjust sound pressure during training to a convenient level. This level was then kept constant throughout the listening test.

The listening test was conducted using the WhisPER software [34]. As adaptation method, a Bayesian approach that closely matches the ZEST procedure [35] was chosen due to its unbiased and reliable results. The a-priori probability density function was a Gaussian distribution with its mean in the middle of the stimulus range; the standard deviation was adapted in pretests. WhisPER uses a logistic psychometric model function, which will equal the ZEST approach, originally utilizing a Weibull function, if the increments of the physical stimulus range (i.e. the mixing time steps) can assumed to be perceptively equidistant. Subjects had to listen to

all nine rooms in randomized order. A run was stopped after 20 trials, resulting in a test duration of about 60 minutes per person.

3. LISTENING TEST RESULTS

The listening test was a difficult task for some of the subjects. Not all reached a valid threshold under every tested condition, i.e. some could not even discriminate the manipulated simulation at the lowest mixing time step from the 'original' simulation. Consequently, only 10 out of our 24 subjects were considered as the expert listeners and their results subjected to further analysis. Figure 3 shows the mean perceptual mixing times $t_{mp50\%}$ and confidence intervals ordered according to the two tested conditions volume and average absorption coefficient. Due to a single subject, internal consistency was slightly low (Cronbach's α = 0,635). Nevertheless, we kept this subject within further analysis because of it being a highly sensitive, thus critical, outlier.

As expected, t_{mp} values were found to increase with room volume. As indicated by the growing confidence intervals of rooms 7-9 the subjects' uncertainty increased, too.

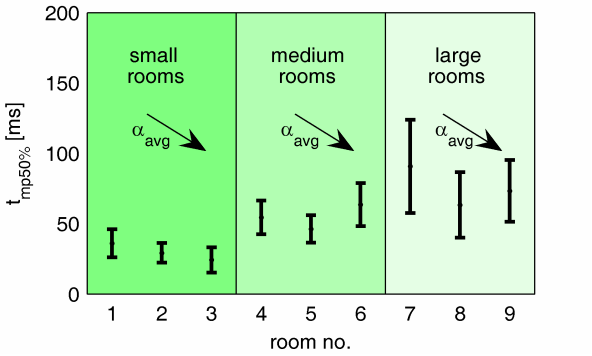

Figure 3 Average perceptual mixing times per room with 95% CIs

The ANOVA for repeated measures proved the volume effect to be significant at p = 0.001. Trend analysis confirmed a significant positive linear relation. It has to be kept in mind, that increasing volume is confounded with increasing reverberation time (cf. column averages of reverberation time in Table 1). However, an effect of the average absorption coefficient (i.e. the relative reverberance independent of volume) could not be found. This is in accordance with theory, as the amount of reverberation is in principle not related to diffusion. However, the reduced sample size allowed only testing a rather large effect (E = 0.34).

4. REGRESSION ANALYSIS

In section 4.1, results of a regression analysis, conducted to test the ability of the three most important model parameter relations (cf. equations 7 to 9) to predict t_{mp}, is presented. Section 4.2 covers some practical problems encountered when using predictors based on empirical room impulse responses (cf. section 1.5). In section 4.3, regression results for the empirical predictors are discussed. In both cases, regression analysis was conducted for a) the mean perceptive mixing time $t_{mp50\%}$, and, in order to receive a more strict criterion for b) the 95% percentile of perceptive mixing time values, i.e. $t_{mp95\%}$. Hence, under most circumstances, the latter regression formulas will guarantee a nearly inaudibly good simulation.

In our study, linear regression is calculated as the least squares fit of our empirical t_{mp} values, derived in the listening test, on t_m as predicted by the discussed predictors. In order to obtain a better linear fit, and for the conduction of inferential analyses, an additional constant term (the intercept) has to be delivered to the regression algorithm. This way, models of the form

$$t_{mp} = b_1 t_m + b_2 \qquad (14)$$

exhibiting a constant term b_2, are derived. However, from theoretical considerations b_2 should be zero as a zero physical mixing time should predict a zero perceptual mixing time.

For practical reasons we had to limit the number of rooms to nine. This number may seem low for conducting linear regression analyses, a fact that is also reflected in the confidence intervals displayed with the models. Nevertheless, this shortcoming is expected to be counterbalanced by the systematic and wide variation applied in selecting the rooms, meant to represent a prototypical subset of reality.

All regression results were evaluated for each predictor by means of explained variance (R^2), and significance of regression, i.e. the possibility of rejecting the null hypothesis of a zero slope value b_1 (at p = 0.05).

4.1. Regression results for model predictors of physical mixing time

Most model predictors of the perceptual mixing time found in literature correspond to a) the square root of volume, b) the mean free path length, being proportional to the ratio of volume and surface area, or c) the reverberation time. Additionally we tested the untransformed volume V, the surface area S (calculated from the three major room dimensions) and the average absorption coefficient α_{mean}.

Multiple and linear regression analyses were conducted. Depending on the selection method of variables used within multiple regressions, models containing one or three predictors resulted. The latter could be rejected, as the additional linear coefficients were insignificant, exhibited collinearity problems (high intercorrelation), and confidence intervals spanning to zero.

Values of $t_{mp50\%}$ were best predicted by the ratio V/S, the kernel of the mean free path length formula. In this case, the explained variance R^2 reached 81.5% (r = 0.9). Moreover, all models' slopes b_1 were positive and significant, despite those derived for the average absorption coefficient α_{mean}. Regression over $\sqrt{V}$ reached 78.6% explained variance, whereas volume alone achieved an R^2 of 77.4%. Reverberation time appears to be rather unsuitable as predictor of the perceptual mixing time. The explained variance was only 53.4 %, which can be completely attributed to confounded volume variation, since the average absorption coefficient α_{mean} shows nearly no linear relation to $t_{mp50\%}$ (R^2 = 0.8%).

Figure 4 shows $t_{mp50\%}$ values and linear regression models including 95 % confidence intervals of both data and models. The regression formula for the best predictor of average $t_{mp50\%}$ was:

$$t_{mp50\%} = 20 \cdot V/S + 12 \quad [ms] \qquad (15)$$

For the sake of simplicity, we calculated surface area from the three main dimensions length, width and height of the considered ideal shoebox room. Additional surfaces of galleries or furniture were thus neglected.

From the regression formulas of the mean free path length relation and the reflection density kernel $\sqrt{V}$, the underlying just audible quantities, as inferable from the models based relations in equations 7 and 8, can now be deduced on an empirical basis. Thus, when comparing equations 5 and 15 and neglecting the constant term of the linear model, the average number of reflections after which the sound fields were experienced as being diffuse was between two and three. Additionally, and while also neglecting the constant model term, just audible reflection density can be calculated by substituting t in equation 1 with the slope b_1 of the regression formula

$$t_{mp50\%} = 0.58 \cdot \sqrt{V} + 21.2 \quad [ms]. \qquad (16)$$

Hence, the just audible reflection density was always above $dN/dt = 114s^{-1}$. However, it shall be emphasized, that these are not measured physical quantities, but are inferred from the model based relations in equations 7 and 8. Further, by neglecting the constant terms, the inferred just audible quantities might be true only in the case of very large rooms, where the constant term of the linear models becomes more and more irrelevant.

For the conservative estimation of t_{mp}, meant to provide an inaudibly exact simulation, linear regression was calculated over the 95%-percentiles of the t_{mp} distribution. The found prediction models are different from the $t_{mp50\%}$ models due to different amounts of deviation of t_{mp} within each room. Again, all models' slopes where positive and significant, despite that derived from the average absorption coefficient α_{mean}. The perceptual mixing time of the 95% percentile $t_{mp95\%}$ is best predicted by volume (R^2 = 78.7 %):

$$t_{mp95\%} = 0.0117 \cdot V + 50.1 \quad [ms] \qquad (17)$$

Results for further predictors are displayed in Figure 5.

4.2. Practical considerations for assessment of empirical mixing time predictors

Impulse response based predictors were calculated from the dummy head's left and right ears' impulse response of the neutral head orientation (i.e. when facing the source) and from the three measurements collected with the omni directional microphone thus giving five mixing time estimates for each room. All four approaches introduced in section 1.5 were considered. The mixing time according to Abel & Huang [27] was calculated using two different mixing time criteria:

Criterion I) - the first moment when the reflection density profile became equal to one, and

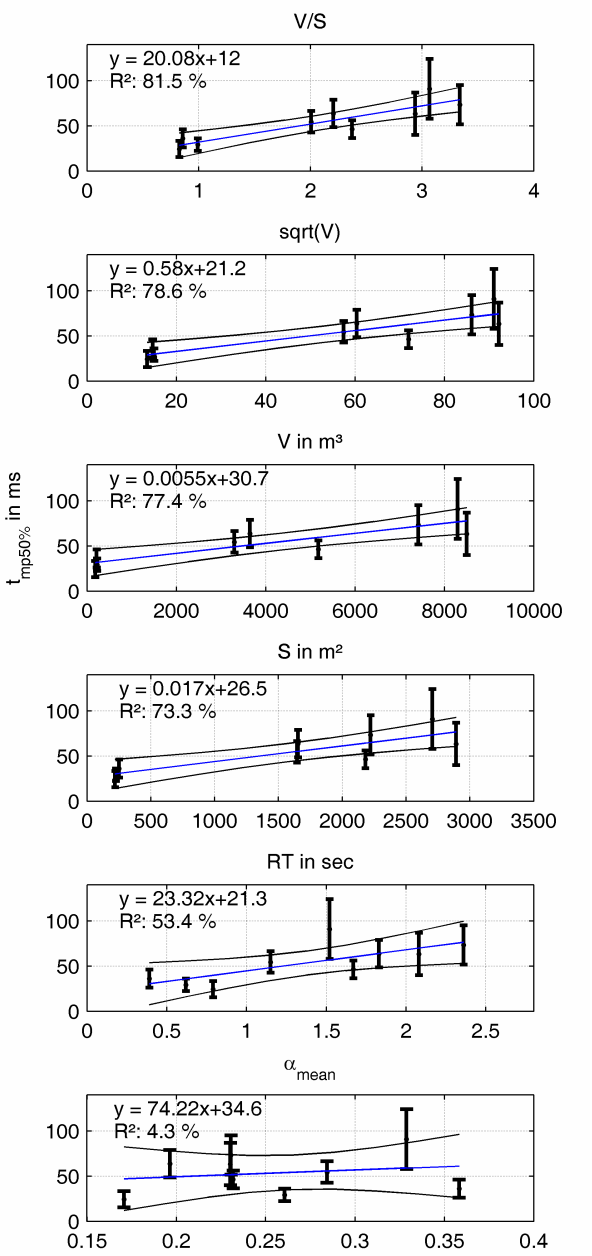

Figure 4 Average perceptual mixing times plotted over model predictors (incl. 95% CIs). Linear model (incl. 95% CIs).

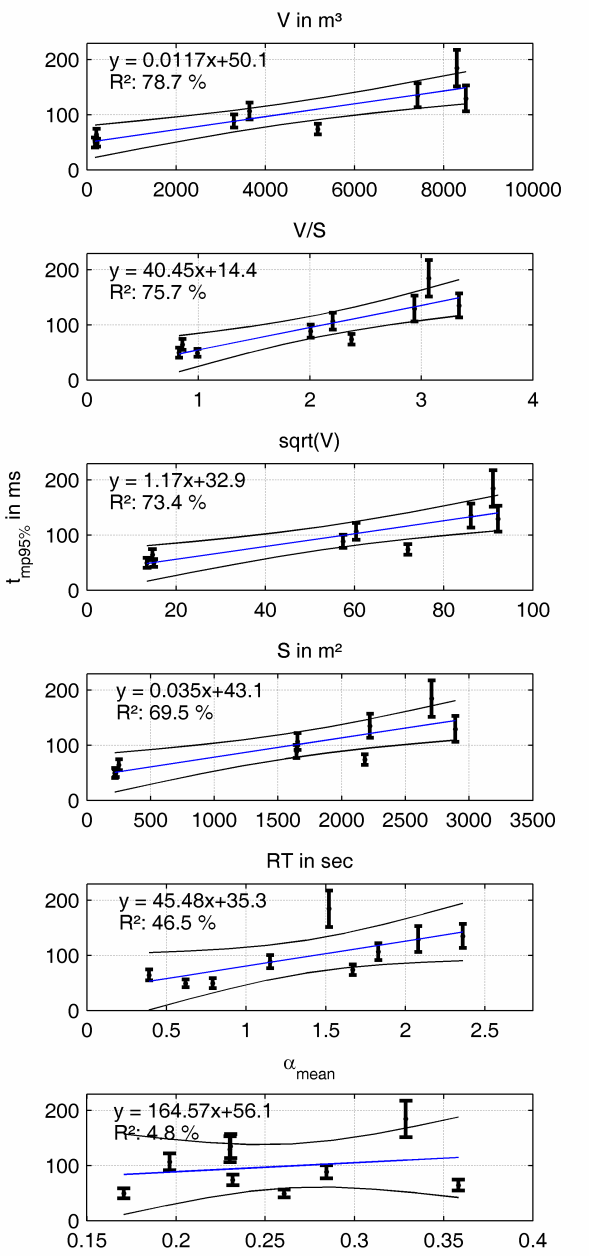

Figure 5 95% percentiles of perceptual mixing times plotted over model predictors (incl. 95% CIs). Linear model (incl. 95% CIs).

Criterion II) - the first moment when the reflection density profile became equal to $1-\sigma_{late}$.

Additionally, as proposed by the authors, the correlation between the initial and late energy (Hidaka et al., [25]) was calculated individually for the eight octave bands between 125 Hz - 16 kHz.

With the empirically based predictors, we encountered several problems. A major issue of all these approaches is the variability of t_m values with measurement position. In Figure 6 mixing times from all four approaches are depicted (Abel & Huang, only for criterion 1, Hidaka et al. only for 500 Hz octave band).

As can be seen, values vary by factor two to three, sometime even by factor ten and more. As with BRIRs there are always two measurements available, we tried to reduce this variability by taking the mean of the mixing time values as estimated from both ears' impulse responses.

The methods of Abel & Huang, Stewart & Sandler and Defrance et al. exhibit a mixing time criterion that can also be determined visually. From examination of the plots of echo density profiles (acc. to [27]), normalized kurtosis (acc. to [28]) or cumulated arrival function (acc. to [29], cf. Figure 7) we suspected the deterministic stopping criteria for automatic determination of the mixing time to be responsible for the high positional variability. With echo density profiles and normalized kurtosis, one sometimes observes large jumps, and final values 1 or 0 respectively are approached somewhat slower and later. In reading off mixing times after the last large jump in the profile, we hoped to get more stable results.

Moreover, values determined automatically while applying the equivalent pulse duration on the Mixing Pursuit decomposition results [29] were implausibly low (mostly < 2 ms, cf. Figure 6, bottom plot). As confirmed by the authors, the absolute values and spread are heavily depending on selected signal residual ratio (SRR). A reliable visual determination of the point where the slope of the cumulated arrival function changes from cubic to linear was nearly impossible (cf. Figure 7, bottom plot).

Despite these problems, for these three approaches we additionally examined mixing time values determined by visual inspection of the corresponding curves.

4.3. Regression results for empirical predictors of physical mixing time

Both the mixing time values calculated from the left and right ears' BRIR and their average, determined visually or using the descried deterministic detection criteria were subjected to linear regression analyses. Again, regression analysis was conducted for the mean perceptive mixing time $t_{mp50\%}$, and for the 95% percentile of perceptive mixing time values, i.e. $t_{mp95\%}$.

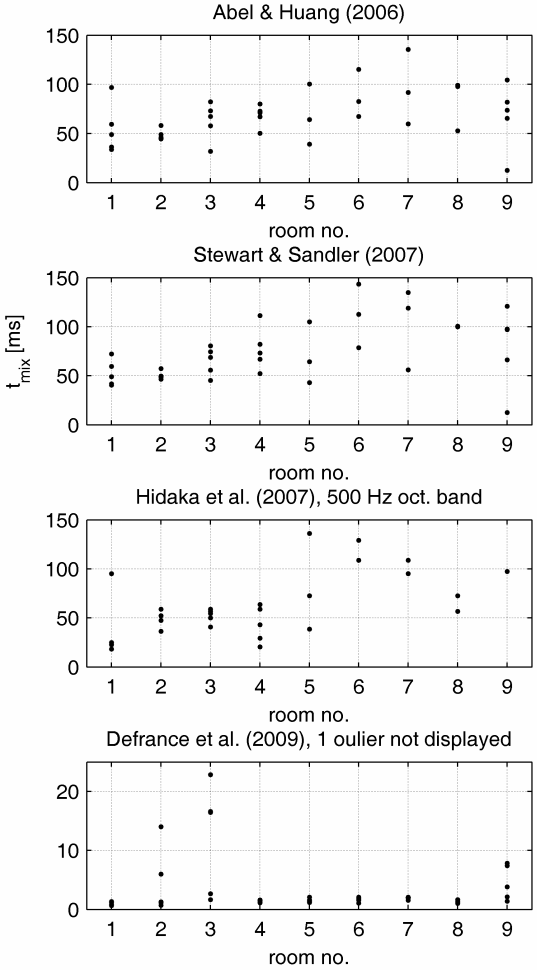

Figure 6 Positional variance obtained with automated t_{mix} detection using impulse response based predictors based (five values per room: left & right ears BRIR, three omnidirectional microphone measurements)

Visually reading off the mixing time values given by reflection density, normalized kurtosis, or cumulative arrival function, lead to considerably less explained variance of the resulting models (10% to 15% less). Moreover, as this procedure is very time consuming it cannot be recommended.

Positional variance was reflected in the regression results too. Although some of the regression models for mixing times derived from a single ears' BRIR reached higher values of explained variance, this happened randomly for the left or the right ear. As a systematic relation could not be found, all further results are solely based on the average mixing time calculated from both ears' BRIRs.

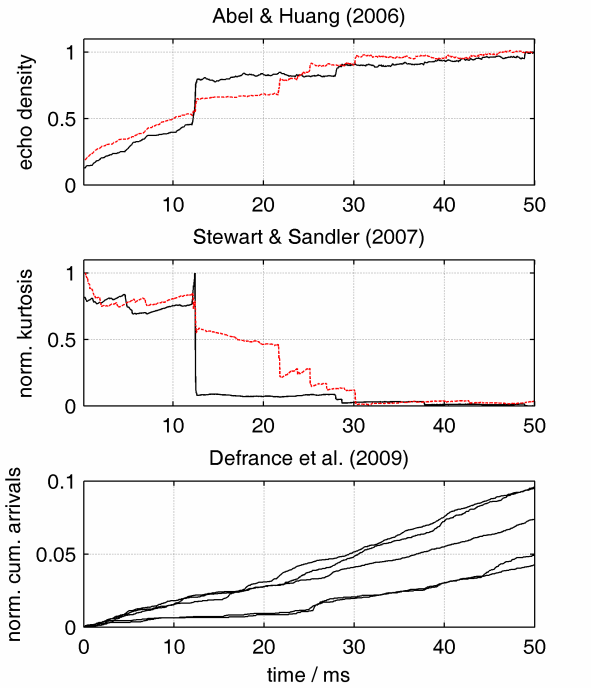

Figure 7 Reliability problems with technical criteria detecting t_{mix} by impulse response based predictors (plots 1 & 2: both ears' impulse responses from room 2; plot 3: all five measurements from room 5)

All models' slopes where positive and significant, except the one derived from physical mixing time estimators using the Stewart & Sandler approach [28]. $T_{mp50\%}$ was best predicted by the algorithm of Defrance et al. (R^2 79.2% cf. Figure 8, plot 2) but only after rejecting results from room three, where the algorithm produced an obviously faulty value. This unreliability limits the algorithms' practical applicability, all the more as it is unclear, how the very low mixing time values detected by the algorithm can be interpreted. From our visual inspections of cumulative arrival functions we suppose, that the equivalent pulse duration criterion does not lead to a detection of the inflection point of the cumulative arrival function as claimed by the authors. So far, we have no explanation for the (potentially) high prediction power of this approach.

In contrast, the echo density approach from [27] reached a R^2 of 74.7% (cf. Figure 8, plot 1) without producing any outliers. The regression formula is

$$t_{mp50\%} = 0.8 \cdot t_{mix\_Abel\_I} - 8 \quad [ms]. \tag{18}$$

It is therefore recommended for a data based determination of $t_{mp50\%}$. The regression formulas and performances of the other approaches can be taken from Figure 8, where results are presented in descending order of performance. The correlation approach from Hidaka et al. [25] reaches minor prediction performance but at least an R^2 of 50.7 % to 56.6 % for the mid frequency octave bands (500 Hz and 1 kHz). The echo density measure using the authors' original detection criterion (echo density >= $1-\sigma_{late,}$ [27]) and normalized kurtosis do not show pronounced prediction power, the latter being already disqualified by lacking any significant linear relation.

Testing the assumption (see 1.3), that the physical mixing time should be the limiting case of perceptual mixing time, we determined whether the models' input values t_m were always larger than the predicted perceptual values $t_{mp50\%}$, i.e. whether

$$t_m \geq t_{m_p} \tag{19}$$

for each linear model. Substituting t_{mp} in equation 19 with the general model function from equation 14, and solving for t_m, and while considering only positive, non-zero slopes b_1 one obtains the solutions

$$t_m \geq \frac{b_1 b_2}{\left(b_1 + b_2^2\right)} \quad \text{for } b_1 \in (0,1), \tag{20}$$

$$t_m \leq \frac{b_1 b_2}{\left(b_1 + b_2^2\right)} \quad \text{for } b_1 > 1, \tag{21}$$

and no solution a for a b_1 equal to one, as in this latter case, inequation 19 holds always or never depending on b_2. For most of our linear models (except of that from the approach of Defrance et al.), b_1 was between zero and one (cf. Figure 8), thus relation 20 holds. It defines a lower limit above which inequation 19 is always fulfilled.

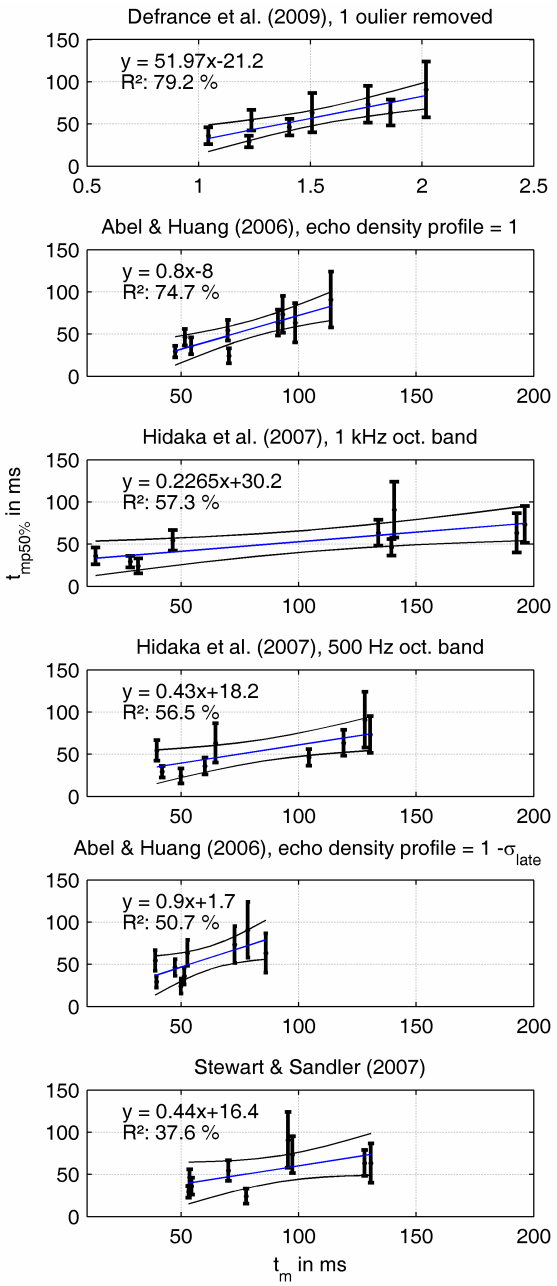

Figure 8 Average perceptual mixing times (mean of left and right ears' BRIR) plotted over impulse response based predictors (incl. 95% CIs). Linear model (incl. 95% CIs).

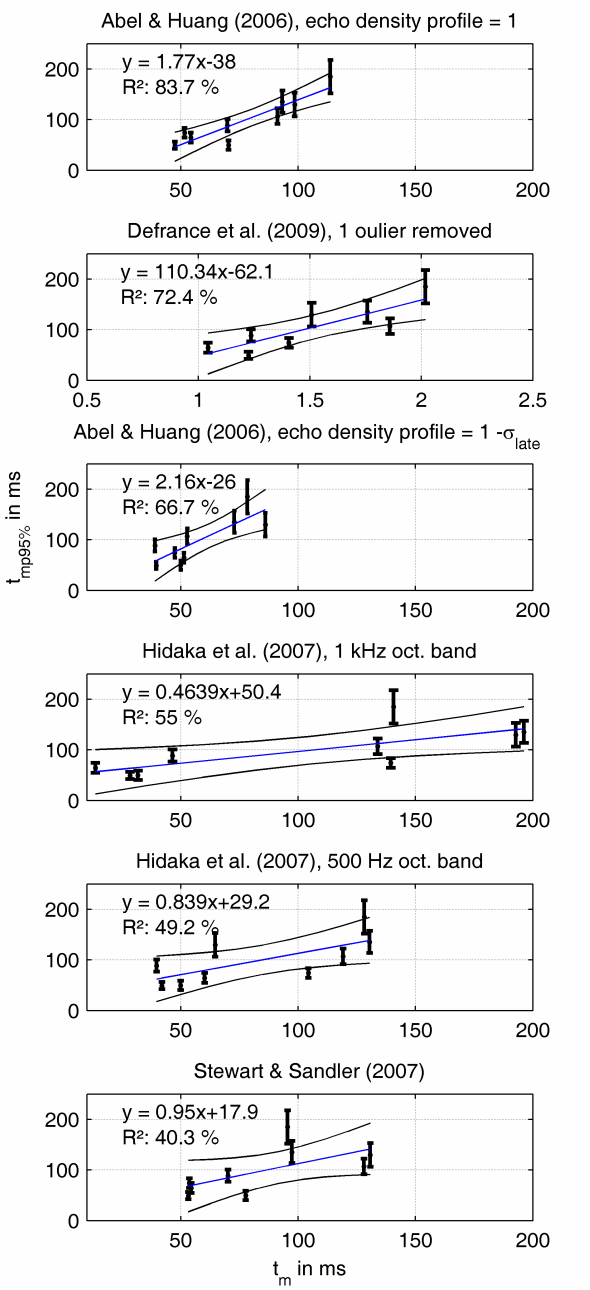

Figure 9 95% percentiles of perceptual mixing times (mean of left and right ears' BRIR) plotted over impulse response based predictors (incl. 95% CIs). Linear model (incl. 95% CIs).

Applying relations 20 and 21 respectively to all models, it was found that only the regression formula of the so far best predictor from Abel & Huang (equation 18) always produces t_{mp} estimates, which are smaller than the estimated physical mixing time. We take this as a further indication for the echo density approach (using criterion I) to be a plausible estimator of the physical mixing time.

Prediction results for $t_{mp95\%}$ are shown in Figure 9, ordered according to descending explained variance R^2. Again, all models' slopes where positive and significant, despite that derived from physical mixing time values estimated with the kurtosis approach [28]. The approach of Abel & Huang repeatedly exhibited superior performance. The explained variance was 83.7%, the regression formula is

$$t_{mp95\%} = 1.8 \cdot t_{mix\_Abel\_I} - 38 \quad [ms]. \qquad (22)$$

The approach of Defrance et al. performed also well, though again only, if results of room three were disregarded. Further measures performed less well, though echo density with criterion II [27] in this case worked better than the correlation measures of Hidaka et al. The differing ranges of predicted values express the difference between the two stopping criteria used with echo density (cf. Figure 8 and Figure 9). With criterion I, a larger spread can be observed; the three volume groups of the rooms seem to be separated better. A possible explanation is, that taking the standard deviation of the last 20% of the impulse response as criterion may introduce additional error. Σ_{late} may already be influenced by noise floor differences that are not related to diffusion.

In Table 2 results from the regression analyses are summarized. Reliable predictors were found for both situations, either 1) if only parameters of the enclosure are known, i.e. in the case of model-based auralization, or 2) if an impulse response is given, i.e. in data based auralization. If all details are known, prediction formulas can also be used for cross checking.

5. CONCLUSIONS

Perceptual mixing time was assessed for the first time by means of a high quality dynamic binaural simulation. BRIR data sets have been acquired for nine acoustical environments, selected to be systematically varied in volume and average absorption. A wide spectrum of a) deterministic model predictors of the perceptual mixing time and b) impulse response based measures for the determination of the physical mixing time were evaluated for their power to predict the listening test results. As a result, linear models for a prediction of a) the average and b) the more critical 95% percentile of the perceptual mixing time were presented.

	Model based predictors		Empirical predictors	
#	$T_{mp50\%}$	$T_{mp95\%}$	$T_{mp50\%}$	$T_{mp95\%}$
1	V/S R^2 81.5%	V R^2 78.7%	Abel I R^2 74.7%	Abel I R^2 83.7%
2	√V R^2 78.6%	V/S R^2 75.7%	Hidaka 1k R^2 57.3%	Abel II R^2 66.7%
3	V R^2 77.4%	√V R^2 73.4%	Hidaka 500 R^2 56.5%	Hidaka 1k R^2 55%
4	S R^2 73.3%	S R^2 69.5%	Abel II R^2 50.7%	Hidaka 500 R^2 49.2%
5	RT R^2 53.4%	RT R^2 46.5%	Stewart R^2 37.6%	Stewart R^2 40.3%
6	α_{mean} R^2 4.3%	α_{mean} R^2 4.8%	Defrance R^2 79.2%*	Defrance R^2 72.4%*

Table 2 Ranking of model based and empirical predictors of perceptual mixing times $t_{mp50\%}$ and $t_{mp95\%}$, color-code: green R^2 = 100% - 80%, yellow R^2 = 80% - 60%, orange: R^2 = 60% - 40%, red R^2 < 40%, (*disqualified)

Results indicate that for shoebox shaped rooms, which are mostly free from additional diffusing obstacles perceptual mixing time will be proportional to the size of the enclosure and is best predicted from the mean free path length, calculated from volume and the three major room dimensions. Average absorption, i.e. relative reverberance was not found to have a significant influence. If an impulse response is available, perceptual mixing time is predicted best using regression formulae 18 or 22 respectively. For reliable prediction, the input value for these formulae – the instant, when the echo density profile according to Abel & Huang [27] is becoming equal to one – should be averaged over several measurement positions. Regression formulae can now conveniently be applied for reducing the rendering effort in the context of either high quality VAEs or plausible auralizations on limited platforms such as mobile audio devices.

6. ACKNOWLEDGEMENTS

This work of Alexander Lindau was supported by a grant from the Deutsche Telekom Laboratories and by

the Deutsche Forschungsgemeinschaft (DFG), grant WE 4057/1-1.

7. REFERENCES

[1] Jot, J.-M.; Cerveau, L.; Warusfel, O. (1997): "Analysis and Synthesis of Room Reverberation Based on a Statistical Time-Frequency Model." In: *Proc. of the 103rd AES Conv.,* New York, preprint no. 4629

[2] Vorländer, M. (2007): *Auralization. Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. 1st ed., Berlin (a.o.).: Springer

[3] Polack, J.-D. (1992): "Modifying Chambers to play Billiards the Foundations of Reverberation Theory." In: *Acustica*, Vol. 76, pp. 257-272

[4] Reilly, A.; McGrath, D.; Dalenbäck, B.-I. (1995): "Using Auralisation for Creating Animated 3-D Sound Fields Across Multiple Speakers." In: *Proc. of the 99th AES Conv.,* New York, preprint no. 4127

[5] Meesawat, K; Hammershøi, D. (2003): "The time when the reverberant tail in binaural room impulse response begins." In: *Proc. of the 115th AES Conv.,* New York, preprint no. 5859

[6] Reilly, A.; McGrath, D. (1995): "Real-Time Auralization with Head Tracking." In: *Proc. of the 5th Australian Regional AES Conv*., Sydney, preprint no. 4024

[7] Lindau, A.; Hohn, T.; Weinzierl, S. (2007): "Binaural resynthesis for comparative studies of acoustical environments." In: *Proc. of the 122nd AES Conv.,* Vienna, preprint no. 7032

[8] Kuttruff, H. (2000): *Room Acoustics*. 4th ed., new York: Routledge Chapman & Hall

[9] Schroeder, M.R. (1959): "Measurement of Sound Diffusion in Reverberation Chambers." In: *J. Acoust. Soc. Am.*, Vol. 31, No. 11, pp. 1407-1414

[10] Blesser, B. (2001): "An Interdisciplinary Synthesis of Reverberation Viewpoints." In: *J. Audio Eng. Soc.*, Vol. 49, No. 10, pp. 867- 903

[11] Joyce, W. B. (1975): "Sabine's reverberation time and ergodic auditoriums." In: *J. Acoust. Soc. Am.*, Vol. 58, No. 3, pp. 643-655

[12] Olive, S.E.; Toole, F.E. (1989): "The Detection of Reflections in Typical Rooms." In: *J. Audio Eng. Soc.*, Vol. 37, No. 7/8, pp. 539-553

[13] Begault, D. (1992): "Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systems." In: *J. Audio Eng. Soc*., Vol. 40, No. 11, pp. 895-904

[14] Bradley, J.S.; Soulodre, G.A. (1995): "The influence of late arriving energy on spatial impression." In: *J. Acoust. Soc. Am*., Vol. 97, No. 4, pp. 2263-2271

[15] Reichardt, W.; Lehmann, U. (1978): "Raumeindruck als Oberbegriff von Räumlichkeit und Halligkeit, Erläuterungen des Raumeindrucksmaßes R." In: *Acustica,* Vol. 40, No. 5, pp. 277-290

[16] Kuttruff, H. (1993): "Auralization of Impulse Responses Modeled on the Basis of Ray-Tracing Results." In: *J. Audio Eng. Soc*., Vol. 41, No. 11, pp. 876-880

[17] Hidaka, T.; Okano, T.; Beranek, L. L. (1995): "Interaural cross-correlation, lateral fraction, and low- and -high-frequency sound levels as measures of acoustical quality in concert halls ." In: *J. Acoust. Soc. Am*., Vol. 98, No. 2, pp. 988-1007

[18] Cremer, L.; Müller, H. A. (1978): *Die wissenschaftlichen Grundlagen der Raumakustik. Bd. 1: Geometrische Raumakustik. Statistische Raumakustik. Psychologische Raumakustik.* 2nd ed., Stuttgart: Hirzel

[19] Schroeder, M.R. (1962): "Natural sounding artificial reverberation." In: *J. Audio Eng. Soc*., Vol. 10, No. 3, pp. 219-223

[20] Rubak, P.; Johansen, L. G. (1999): "Artificial Reverberation based on a Pseudo-random Impulse Response II." In: *Proc. of the 106th AES Conv.,* München, preprint no. 4900

[21] Griesinger, D. (1989): "Practical Processors and Programs for Digital Reverberation." In: *Proc. of*

the 7th International AES Conference: Audio in Digital Times

[22] Schreiber, L. (1960): "Was empfinden wir als gleichförmiges Rauschen?" In: *Frequenz,* Vol. 14, No. 12, pp. 399 ff.

[23] Polack, J.-D. (1988): *La transmission de l'énergie sonore dans les salles.* Thèse de Doctrorat d'Etat. Le Mans: Université du Maine

[24] Schmidt, W.; Ahnert, W. (1973): "Einfluss der Richtungs- und Zeitdiffusität von Anfangsreflexionen auf den Raumeindruck." In: *Wiss. Zeit. d. TU Dresden,* Vol. 22, pp. 313 ff.

[25] Hidaka, T.; Yamada, Y.; Nakagawa, T. (2007): "A new definition of boundary point between early reflections and late reverberation in room impulse responses." In: *J. Acoust. Soc. Am.*, Vol. 122, No. 1, pp. 326-332

[26] Schroeder, M.R. (1965): "New method of measuring reverberation time." In: *J. Acoust. Soc. Am.,* Vol. 37, pp. 409-412

[27] Abel, J.S.; Huang, P. (2006): "A Simple, Robust Measure of Reverberation Echo Density." In: *Proc. of the 121st AES Conv.,* San Francisco

[28] Stewart, R.; Sandler, M. (2007): "Statistical Measures of Early Reflections of Room Impulse Responses." In: *Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07).* Bordeaux

[29] Defrance, G.; Daudet, L.; Polack, J.-D. (2009): "Using Matched Pursuit for Estimating Mixing Time Within Room Impulse Responses." In: *Acta Acustica united with Acustica,* Vol. 95, No. 6, pp. 1071-1081

[30] Stanley, W. D.; Peterson, S. J. (1979): "Equivalent Statistical Bandwidths of Conventional Low-Pass Filters." In: *IEEE Transactions on Communications,* Vol. 27, No. 10, pp. 1633-1634

[31] Kuttruff, H. (1991): "On the audibility of phase distortion in rooms and its significance for sound reproduction and digital simulation in room acoustics." In: *Acustica,* Vol. 74, pp. 3-7

[32] Bortz, J.; Döring, N. (2006): *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler.* 4. Aufl., Heidelberg: Springer

[33] Schärer, Z.; Lindau, A. (2009): "Evaluation of Equalization Methods for Binaural Signals." In: *Proc. of the 126th AES Conv*., Munich, preprint no. 7721

[34] Ciba, S.; Wlodarski, A.; Maempel, H.-J. (2009): "WhisPER – A new tool for performing listening tests." In: *Proc. of the 126th AES Conv.,* Munich, preprint 7749

[35] King-Smith, P. E. et al. (1994): "Efficient and Unbiased Modifications of the QUEST Threshold Method: Theory, Simulations, Experimental Simulations, Experimental Evaluation and Practical Implementation." In: *Vision Research*, Vol. 34, No. 7, pp. 885-912