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

L. Kosanke, A. Lindau

Audio Communication Group TU Berlin
Email: l-kosanke@mailbox.tu-berlin.de

Introduction

Virtual acoustic environments (VAEs) can be realized by means of dynamic binaural synthesis. Therefore, anechoic audio is convolved with interactively interchangeable binaural room impulse responses (BRIRs) measured for a discrete number of head orientations. In a room impulse response, early reflections gradually pass into a stochastic reverberation tail. The transition point between both domains is called the physical mixing time (t_m). After mixing, the sound field is equivalent to a diffuse field, which is characterized by equidistributed a) energy density and b) directional energy flux within the whole room. Due to increasing diffuseness in the decay process, individual reflections become less perceptively distinguishable. The lesser early reflections have to be rendered in a VAE, the more computational demands will be reduced. An obvious method to achieve this reduction would be to replace the individual reverberation tail of the BRIRs – after an instant when *perceptual* discrimination is no longer possible – with an arbitrary and constant reverberation tail. For a singular small room this instant, i.e. the 'perceptual mixing time' (t_{mp}), was examined in [1]. Several combinations of early reflections and tails from different receiver and source combinations were tested using static auralization, finally leading to the conclusion that – for this room – t_{mp} was about 40 ms and independent of positional changes within the room. As for larger rooms, higher mixing times can be expected, in [2] t_{mp} was assessed for an auditorium, again using static auralization. T_{mp} was indeed higher (140 ms), moreover, and presumably due to modal behavior at low frequencies a positional dependency of t_{mp} was observed. Additionally, listeners were most sensitive to a drum set sample. The aim of our study was to find the perceptual mixing time for several rooms, systematically varied in volume and average absorption, utilizing dynamic auralization. Afterwards, by means of regression analysis, the ability of several model based predictors of t_m (mean free path length ~V/S [3], reflection density ~$\sqrt{V}$ [4], and reverberation time [5]) to indicate the perceptual mixing time t_{mp} was examined.

Method

We selected nine rooms aiming for a systematic variation of volume and the average absorption coefficient (α_{avg}), each in three increments (cf. Table 1). Due to interrelation, it is difficult to vary room volume independent of absolute amount of reverberation. However, varying α_{avg}, we could at least assess the influence of the *relative* amount of reverberation independent from volume. Reverberation time steps were chosen to exceed to at least a just noticeable difference of 10%. All rooms exhibited nearly shoebox-shape as mixing times were expected to be highest in this case [5]. BRIRs were measured using the automatic head and torso simulator FABIAN [2] for horizontal head orientations within ±80° in angular steps of 1°. As source, a frequency compensated 3-way dodecahedron loudspeaker was placed in the middle of the stage, typically located at one narrow end of the room. FABIAN was seated on the longitudinal room axis at twice the critical distance while directly facing the loudspeaker.

Table 1: The nine rooms chosen to exhibit a systematic variation in volume and average absorption coefficient α

	small V	medium V	large V	α_{avg} (avg. RT)
large α (RT)	216 m^3 / 0.36 (0.39 s)	3300 m^3/ 0.28 (1.15 s)	8298 m^3 / 0.33 (1.52 s)	**0.32 (1 s)**
medium α (RT)	224 m^3 / 0.26 (0.62 s)	5179 m^3 / 0.23 / (1.67 s)	8500 m^3 / 0.23 (2.08 s)	**0.24 (1.45 s)**
small α (RT)	182 m^3 / 0.17 (0.79 s)	3647 m^3 / 0.2 (1.83 s)	7417 m^3/ 0.23 (2.36 s)	**0.2 (1.66 s)**
avg. Vol.	**207 m^3**	**4042 m^3**	**8072 m^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 refreshed 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, while the late reverberation tail – for practical reasons taken from the BRIR corresponding to frontal head orientation– was not changed anymore. Therefore, early and late BRIR parts could be concatenated at arbitrary instants in increments of 5.8 ms (small rooms), and 11.6 ms (medium and large rooms) respectively. For concatenation of the reverberation tail, a linear cross fade of a length equal to the step size was used. During training subjects were instructed to rotate their head widely to maximize the difference between original and manipulated simulation. The critical drum set sample from [2] was used as stimulus (length: 2.5 s plus reverb). The listening test was conducted with the WhisPER software [6], using an adaptive method that closely matches the ZEST procedure apart from the a-priori probability density function being a Gaussian distribution. Each of the 24 subjects had to listen to all nine rooms in randomized order; each run was stopped after 20 trials.

Results

To be able to test medium size first order interaction effects at p = 0.05 with 80% power in a repeated measures design at least 19 subjects were needed. Unfortunately, only 10 out of 24 subjects reached a valid threshold under every tested condition. Thus, only results of these expert listeners were considered in further analyses. Figure 1 shows the average

t_{mp} values (i.e. the $t_{mp50\%}$, as distributions could be assumed normal) and confidence intervals ordered according to the two test conditions volume and average absorption coefficient. Due to a single subject, internal consistency was slightly low (Cronbach's alpha α=0,635). Nevertheless, we kept this subject within further analysis because of it being a highly sensitive, thus critical, outlier.

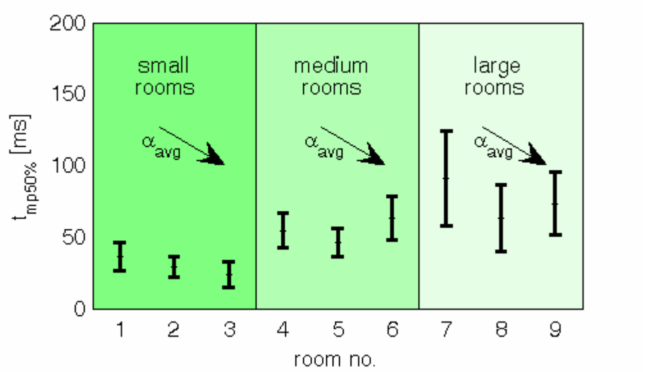

Figure 1: Average perceptual mixing times per room with 95% CIs

Expectedly, because in larger rooms it takes longer for the sound field to become diffuse, t_{mp} values are found to increase with room volume. As indicated by the growing confidence intervals of rooms 7-9 the subjects' uncertainty increased, too. The ANOVA for repeated measures proved the volume effect to be significant at p = 0.001. Trend analysis confirmed a positive linear relation. It has to be kept in mind, that increasing volume is confounded with increasing reverberation time (cf. column averages of RT in Tab. 1). Though, an effect of the average absorption coefficient (i.e. the relative reverberance independent of volume) could not be found. Besides, the reduced sample size allowed only testing a rather large effect (E = 0.34).

Linear Regression Analysis

In order to find a suitable physical predictor for the perceptual mixing time in BRIRs linear regression analysis was conducted. Most model based predictors found in literature can be attributed to a) the square root of volume, b) the reverberation time, or c) the mean free path length, the latter being proportional to the ratio of volume and surface area. Linear regression analysis of average t_{mp} values indicated the ratio V/S, the kernel of the mean free path length formula [3], as the best predictor. In this case, the explained variance R² reached 81.5% (r = 0.9). Regression over $\sqrt{V}$ (from reflection density formula [4]) reached 78.6%, whereas volume alone achieved an R² of 77.4%. Reverberation time (average of octave bands 250 Hz - 4 kHz of three different measurements) appeared to be rather unsuitable as predictor of the perceptual mixing time. The explained variance was only 53.4 %. For comparison, Figure 2 shows the regression model and $t_{mp50\%}$ values including 95 % confidence intervals for data and model for both the ratio V/S and the reverberation time RT. The regression formula for the best predictor of $t_{mp50\%}$, the ratio V/S (with V in m³, S in m²) was:

$$t_{mp50\%} = 20 \cdot V/S + 12 \qquad [\mathrm{ms}] \qquad (1)$$

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.

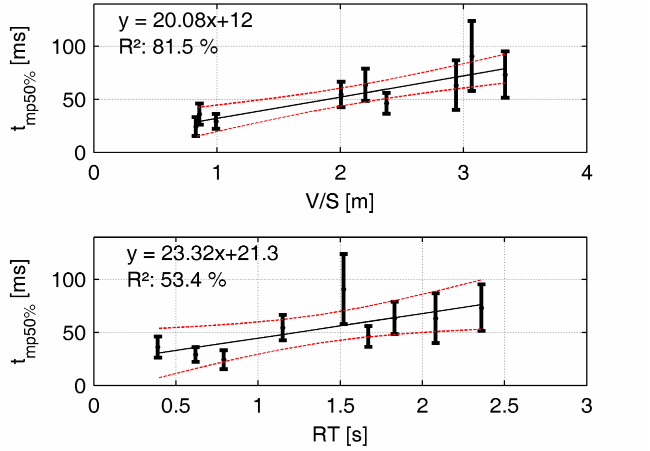

Figure 2: Average perceptual mixing times plotted over V/S and RT (incl. 95% CIs). Linear model (incl. 95% CIs).

In order to find a conservative estimation of t_{mp}, a linear regression over the 95%-percentiles of the t_{mp} distribution was calculated. Hence, under most circumstances, this regression formula will guarantee a nearly inaudibly good simulation. Due to different amounts of deviation of t_{mp} within each room the prediction models are different to that of $t_{mp50\%}$. The perceptual mixing time of the 95% percentile $t_{mp95\%}$ was thus best predicted by volume (R² = 78.7 %, with V in m³):

$$t_{mp95\%} = 0.0117 \cdot V + 50.1 \qquad [\mathrm{ms}] \qquad (2).$$

Conclusion

Perceptual mixing time t_{mp} was assessed for the first time by means of realistic dynamic binaural simulation. Therefore, BRIR data sets of several real rooms, which were systematically varied regarding volume and average absorption, have been acquired. Predictors of the physical mixing time were assessed for their suitability to predict the perceptual results. As a result, linear models for a convenient prediction of $t_{mp50\%}$ and $t_{mp95\%}$ respectively were presented and discussed. Perceptual mixing time appears to be strongly related to room volume. Average absorption, i.e. relative reverberance was not found to have a significant influence. In summary results indicate, that for shoebox shaped rooms which are mostly free from additional diffusing obstacles average perceptual mixing time will be proportional to the size of the enclosure and is predicted best from mean free path length.

Alexander Lindau was supported by a grant from the Deutsche Telekom Laboratories.

References

[1] Meesawat, K.; Hammershøi, D. (2003): "The time when …", In: *Proc. of the 115th AES Conv.*, preprint no. 5859

[2] Lindau, A. et al. (2007): "Binaural resynthesis for comparative …" In: *Proc. of the 122nd AES Conv.*

[3] Rubak, P.; Johansen, L. (1999): "Artificial Reverberation ..." In: *Proc. of the 106th AES Conv.*, preprint no. 4900

[4] Cremer, L. (1978): *"Die wissenschaftlichen …Bd. 1"*, S. Hirzel Verlag, Stuttgart

[5] Hidaka, T. et al (2007): "A new definition of boundary …." *J. Acoust. Soc. Am.*, **122**(1): 326-332

[6] Ciba, S. et al. (2009): "WhisPER …" In: *Proc. of the 126th AES Conv.*, Munich, preprint 7749