Exploring Perceptual Similarities in Binaural Reverberation

Thomas McKenzie
Acoustics and Audio Group, Reid School of Music, University of Edinburgh, United Kingdom
Email: thomas.mckenzie@ed.ac.uk

Introduction

Virtual and augmented reality systems require real-time dynamic six degrees-of-freedom audio engines capable of high-fidelity binaural reproduction of a room's acoustics [1, 2]. This is especially important in augmented reality, where both virtual and real-world sounds are present [3]. As each sound source will likely have a different position, the room acoustics reproduction may not need to be authentic, but rather be transfer-plausible [4].

Comparing the perceived reverberance of two binaural signals can be useful in the design stages of virtual and augmented reality systems. However, perceptual listening tests are laborious to run, and each iteration of a design would require a new test. Therefore, having reliable methods for predicting the perceived reverberation similarity in binaural signals is desirable.

Objective metrics for describing the acoustical properties of a room are usually performed on the room impulse response (RIR), and include reverberation time RT60, direct-to-reverberant ratio (DRR), and others such as early decay time and clarity. Metrics used for comparing binaural signals include interaural level and time difference (ILD and ITD, respectively), interaural cross-correlation (IACC), and colouration.

While some work exists on predicting perceived reverberance of binaural signals [5], what is not fully known is the individual relations between simple objective metrics for RIR measurements and the perception of reverberance. Some metrics may be more or less useful, and it may be that using a weighted sum of multiple metrics gives the best overall prediction of similarity.

Method

To investigate the perceived similarity of reverberation in binaural signals, the perceptual listening test results of a previous study on binaural room acoustics similarity [6] were compared to simple objective metrics for reverberation calculated from the RIRs (both spatial and binaural) used to generate the listening test stimuli.

Perceived reverberation similarity

The listening test stimuli were originally created by convolving anechoic speech signals with binaural room impulse responses (BRIRs) rendered from 3rd order Ambisonic spatial room impulse responses (SRIRs) from the variable acoustics 6DoF dataset [7]. Three source-receiver (SR) pairs were used, as illustrated in Figure 1:

SR1: Source 1, receiver 7 → receiver 2.

SR2: Source 2, receiver 5 → receiver 2.

SR3: Source 3, receiver 4 → receiver 1.

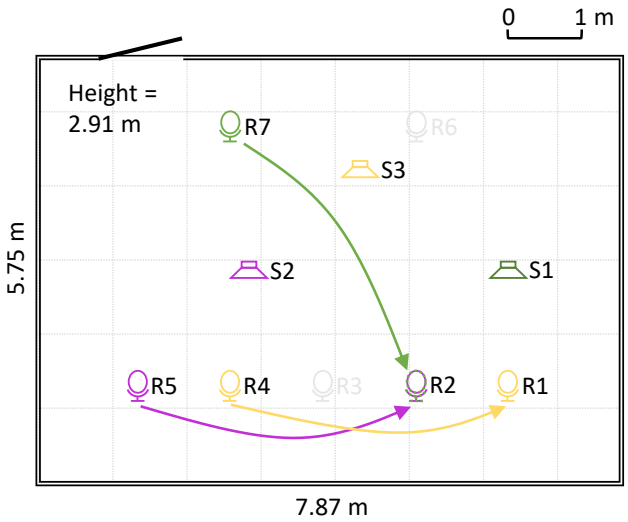

Figure 1: Illustration of the source (S) and receiver (R) positions used to make up the listening test conditions. Microphones and loudspeakers were oriented faced north and south, respectively, according to the illustration orientation. Note that R2 was used both for both SR1 (green) and SR2 (magenta).

For each SR pair, five conditions (C) were created. The conditions were made using hybrid RIRs created from the SR pairs (RIR1 and RIR2) with the same source but a different receiver position as follows:

C1: Entire duration is RIR1.

C2: First 13ms is RIR1, the rest is RIR2.

C3: First 8ms is RIR1, the rest is RIR2.

C4: First 3ms is RIR1, the rest is RIR2.

C5: Entire duration is RIR2.

For example, for SR1, C1 was entirely S1-R7, C2 was the first 13ms of S1-R7 and the rest of S1-R2, and C5 was entirely S1-R2.

The test followed a multiple stimulus paradigm, and participants rated the room acoustic similarity of each test condition (C1 - C5) to those of the reference condition (C1). This was repeated for each SR pair, totalling 15 test conditions. The anechoic speech signal was included as an anchor stimulus. For more information on the listening test methodology and stimuli creation, see [6].

The results showed, as expected, that as more of the RIR used was from a different position, the perceived similarity was reduced. The listening test results data (of the 'same speaker, same sentence' stimuli), is used in

this study. A single result for each condition was taken as the mean result from all participants.

Calculated reverberation similarity

Binaural reverberation similarity was calculated using nine metrics: eight of which used the BRIRs from the listening test; one used the SRIRs. A binaural room impulse response is denoted herein as h, with each sample n up to a total number of samples N.

The first tested metric, simply the difference in root-mean-square (RMS) amplitude between the test and reference BRIRs, ΔRMS was calculated as

$$\Delta\mathrm{RMS} = \frac{|\mathrm{RMS}^{\mathrm{l}}_{\mathrm{test}} - \mathrm{RMS}^{\mathrm{l}}_{\mathrm{ref}}| + |\mathrm{RMS}^{\mathrm{r}}_{\mathrm{test}} - \mathrm{RMS}^{\mathrm{r}}_{\mathrm{ref}}|}{2}. \tag{1}$$

The second metric measured colouration. For this, the composite loudness level (CLL) [8] was first calculated, which uses half-wave rectification, gammatone filterbanks and equal loudness frequency weighting. Colouration (ΔCLL) was then calculated as the mean of the difference between the CLL values for each frequency band of the test and reference BRIRs:

$$\Delta\mathrm{CLL} = \frac{1}{F}\sum_{f=1}^{F} |\mathrm{CLL}_{\mathrm{test}}(f) - \mathrm{CLL}_{\mathrm{ref}}(f)|, \tag{2}$$

where f denotes the index of approximately 1/3 octave frequency bands up to the total of $F = 42$.

The third and fourth metrics measured the change in interaural level difference (ILD) and interaural time difference (ITD), respectively, as ΔILD and ΔITD. ILD and ITD were also calculated in 42 frequency bands, and the difference metric was calculated using the same method as Equation (2).

The fifth metric measured the change in interaural cross-correlation (ΔIACC), which is given as

$$\Delta\mathrm{IACC} = |\max(\mathrm{IACC}_{\mathrm{test}}) - \max(\mathrm{IACC}_{\mathrm{ref}})|. \tag{3}$$

The next metric was the change in reverberation time (ΔRT60), calculated from the T30 ($-$5dB to $-$35dB) of a 1 kHz frequency band using Schroeder integration [9]. RT60 (T30 $\times$ 2) was calculated separately for the left and right of both test and reference BRIRs, and ΔRT60 was calculated as in Equation (1).

Another metric based on reverberation levels of BRIRs is the direct-to-reverberant ratio (DRR). Here, DRR was calculated as the fullband ratio of the first 20ms to the rest of the BRIR,

$$\mathrm{DRR} = 10\log 10\frac{\sum_{n=1}^{0.02 f_s} h[n]^2}{\sum_{n=0.02 f_s+1}^{N} h[n]^2}. \tag{4}$$

Again, DRR was calculated separately for the left and right signals, and change in DRR (ΔDRR) was calculated as in Equation (1).

An auditory model for predicting the perceived externalisation of a static, lateral sound image [10] was included, using the implementation available in the Auditory Modeling Toolbox [11]. The li2020 model estimates externalisation by a combination of monaural spectral cues, ILDs and temporal changes in ILD. The difference in externalisation (Δli2020) was calculated from the externalisation prediction li2020 as $\Delta\mathrm{li2020} = |\mathrm{li2020}_{\mathrm{test}} - \mathrm{li2020}_{\mathrm{ref}}|$.

The final metric used in this study assessed the direction-of-arrival of the SRIRs, prior to the binaural Ambisonic rendering stage. DoA was estimated above 3 kHz using a steered response power method, by directing a fourth-order hyper-cardioid beamformer over the sphere in a 5 degree resolution and calculating the power [12]. The seven directions with the highest power were used, which typically reveals the direct sound and loudest early reflections. The ΔDoA metric was calculated as

$$\Delta\mathrm{DoA} = \frac{1}{I}\sum_{i=1}^{I} \theta(i)_{\mathrm{test}} - \theta(i)_{\mathrm{ref}} \pmod{360}, \tag{5}$$

where i denotes the direction of one of the maximum power values, up to $I = 7$, and θ denotes spherical coordinate.

Results

For ease of comparison with the perceived similarity ratings, the objective metrics were first converted into a percentage of similarity by fitting a linear regression to each metric's results. The a and b values of the regression fittings are shown in Table 1. A visual comparison is presented in Figure 2, whereby a narrow linearly increasing distribution indicates a higher correlation. Data points are labelled 1-15, based on the source-receiver pair and condition number $(5(\mathrm{SR}-1)+\mathrm{C})$, and coloured as in Figure 1. To assess correlation between the predicted and perceived results, the Pearson's correlation coefficient was calculated; the r and p-value results of which are also presented in Table 1.

The results show that many of the tested metrics are somewhat capable of predicting the room acoustic similarity of the tested signals. Two metrics were less success-

Table 1: Linear regression coefficients (a and b) and Pearson's correlation coefficients (r) with respective p-values of the tested metrics against the perceived similarity results.

Metric	a	b	r	p
ΔRMS	-18.88	89.77	-0.880	< 0.001
ΔCLL	-16.04	90.13	-0.789	< 0.001
ΔILD	-13.20	86.79	-0.853	< 0.001
ΔITD	-430600	86.22	-0.586	0.022
ΔIACC	-2.256	88.63	-0.828	< 0.001
ΔRT60	-1043	85.05	-0.595	0.019
ΔDRR	-11.97	90.11	-0.917	< 0.001
Δli2020	-21.61	86.08	-0.857	< 0.001
ΔDoA	-0.4085	100.2	-0.878	< 0.001

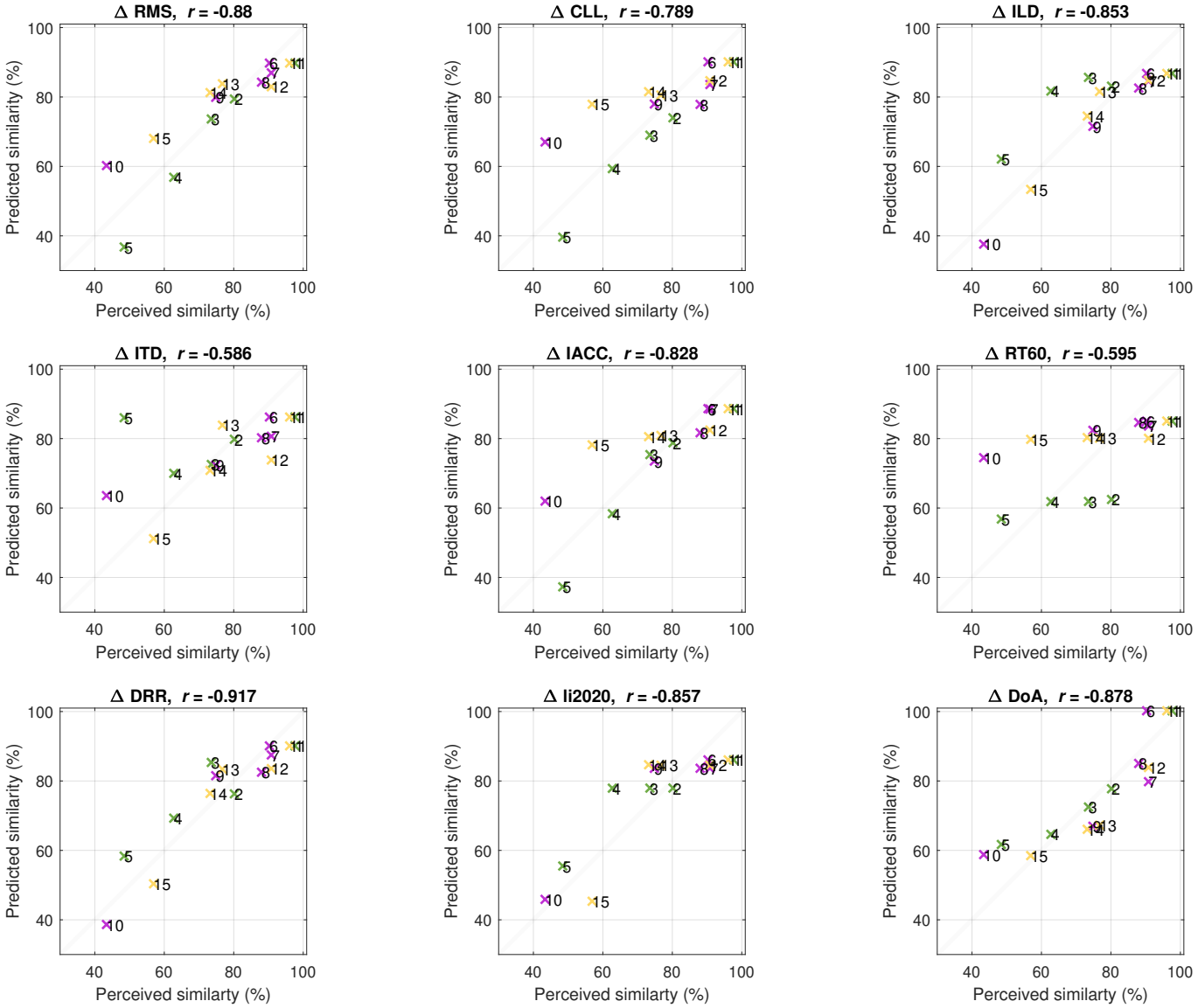

Figure 2: Metric performance by means of the predicted versus perceived similarity, with r denoting correlation coefficient for each model. Data points are labelled 1-15 based on the source-receiver pair and condition number $(5(\mathrm{SR}-1)+\mathrm{C})$, and coloured as in Figure 1.

ful: ΔITD and ΔRT60, with correlation values $r < -0.6$. The ITD calculation was performed in frequency bands, which could be more complicated than necessary. The RT60 metric could be poor due to the approximately uniform absorption of the room, or simply due to the shared reverberation tail for conditions C2-C5.

Looking at Figure 2, it appears a combination of multiple metrics may offer an improved overall correlation with the perceived similarity ratings. For example, SR1 produced lower predicted similarity than SR2 and SR3 for ΔCLL, but higher predicted similarity for ΔILD and ΔDRR.

To test a combination of multiple metrics, a multiple linear regression model was fit to data of 6 of the 9 tested metrics. The three omitted metrics were ΔITD, ΔRT60 and Δli2020: the first two due to their lower correlation values, and the li2020 model due to it using a combination of multiple cues. The results for the fitting are presented in Table 2, and the values of predicted similarity made using the model are shown, along with the Pearson's correlation coefficient, in Figure 3. Data points are again labelled 1-15, based on the source-receiver pair and condition number $(5(\mathrm{SR}-1)+\mathrm{C})$, and coloured as in Figure 1.

The regression model gives an improved prediction than any one metric alone ($r = 0.984$, $p < 0.001$). This seems intuitive considering that reverberation is a multi-faceted percept. The low relative weighting estimate values for ΔCLL and ΔIACC of 0.045 and 0.063, respectively (see Table 2), suggest that the model could produce strong results with even fewer metrics.

Discussion and Outlook

This paper has investigated perceptual similarities in the reverberation of binaural signals. Nine metrics for calculating room acoustic similarity of room impulse responses, including differences in amplitude, temporal, re-

Table 2: Multiple linear regression model fitting results using six of the tested metrics, where β_0 denotes intercept.

Metric	Estimate	SE	t	p
β_0	−20.64	8.630	−2.392	0.044
ΔRMS	0.329	0.499	0.660	0.528
ΔCLL	0.045	0.436	0.103	0.920
ΔILD	0.291	0.294	0.989	0.352
ΔIACC	0.063	0.477	0.132	0.898
ΔDRR	0.186	0.339	0.550	0.597
ΔDoA	0.356	0.124	2.865	0.021

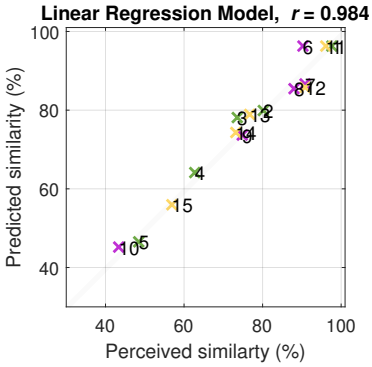

Figure 3: Performance of the multiple linear regression model by means of the predicted versus perceived similarity, with r denoting correlation coefficient.

verberation level and reverberation directionality, have been compared to perceptual listening test results on room acoustic similarity. Some metrics appear to offer a reasonably accurate prediction of the perceived similarity values alone, and a simple multiple linear regression model shows that a weighted combination of multiple metrics can give an even better prediction.

Combinations of metrics could be more nuanced, accounting for human spatial hearing features such as how colouration is more perceptually relevant for ipsilateral signals while reverberance is more relevant for contralateral signals [13]. A model optimised using a regression fitting should also be tested using k-fold cross validation to investigate potential overfitting.

Though the results are promising, the listening test only considered 15 conditions inside one room. These were room impulse responses (both binaural and spatial) and not running signals. To calculate binaural reverberation similarity in real-time, the metrics would need to either work on running signals, or first utilise a blind reverberation estimation method. The rendering was non-individualised and static, and all stimuli used the same source signal. A greater pool of stimuli with more variation is needed, along with a consideration of how individualised binaural hearing affects the perception of reverberation.

Acknowledgments

Thanks to Alec Wright for the insightful discussions on linear regression models.

References

[1] McCormack, L., Politis, A., McKenzie, T., Hold, C., Pulkki, V.: Object-based six-degrees-of-freedom rendering of sound scenes captured with multiple Ambisonic receivers. J. Audio Eng. Soc. 70 (2022), 355–372

[2] McKenzie, T., Meyer-Kahlen, N., Hold, C., Schlecht, S. J., Pulkki, V.: Auralisation of measured room transitions in virtual reality. J. Audio Eng. Soc. 71 (2023), 326–337

[3] Neidhardt, A., Schneiderwind, C., Klein, F.: Perceptual matching of room acoustics for auditory augmented reality in small rooms - literature review and theoretical framework. Trends in Hearing 26 (2022)

[4] Meyer-Kahlen, N., Amengual Garí, S., McKenzie, T., Schlecht, S. J., Lokki, T.: Transfer-plausibility of binaural rendering with different real-world references. DAGA (2022), 154–157

[5] Osses Vecchi, A., Kohlrausch, A., Lachenmayr, W., Mommertz, E.: Predicting the perceived reverberation in different room acoustic environments using a binaural auditory model. J. Acoust. Soc. Am. EL 141 (2017), 381–387

[6] McKenzie, T., Meyer-Kahlen, N., Schlecht, S. J.: The role of source signal similarity in distinguishing between different positions in a room. AES International Conference on Immersive and Spatial Audio (2023)

[7] McKenzie, T., McCormack, L., Hold, C.: Dataset of spatial room impulse responses in a variable acoustics room for six degrees-of-freedom rendering and analysis. arXiv preprint (2021)

[8] Pulkki, V., Karjalainen, M., Huopaniemi, J.: Analyzing virtual sound source attributes using a binaural auditory model. J. Audio Eng. Soc. 47 (1999) 203–217

[9] Schroeder, M. R.: New method of measuring reverberation time. J. Acoust. Soc. Am. 37 (1965) 1187–1188

[10] Li, S., Baumgartner, R., Peissig, J.: Modeling perceived externalization of a static, lateral sound image. Acta Acustica 4 (2020)

[11] Majdak, P., Hollomey, C., Baumgartner, R.: AMT 1.x: A toolbox for reproducible research in auditory modeling. Acta Acustica 6 (2022)

[12] Politis, A.: Microphone array processing for parametric spatial audio techniques. PhD thesis, Aalto University (2016)

[13] Li, S., Schlieper, R., Peissig, J.: The effect of variation of reverberation parameters in contralateral versus ipsilateral ear signals on perceived externalization of a lateral sound source in a listening room. J. Acoust. Soc. Am. 144 (2018) 966–980