24th INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
23–27 July 2017, London

ICSV24
23-27JULY 2017
LONDON CALLING

PLANARITY ANALYSIS OF ROOM ACOUSTICS FOR OBJECT-BASED REVERBERATION

Philip Coleman, Philip J. B. Jackson
Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, GU2 7XH, UK
email: p.d.coleman@surrey.ac.uk

Recent work into 3D audio reproduction has considered the definition of a set of parameters to encode reverberation into an object-based audio scene. The reverberant spatial audio object (RSAO) describes the reverberation in terms of a set of localised, delayed and filtered (early) reflections, together with a late energy envelope modelling the diffuse late decay. The planarity metric, originally developed to evaluate the directionality of reproduced sound fields, is used to analyse a set of multichannel room impulse responses (RIRs) recorded at a microphone array. Planarity describes the spatial compactness of incident sound energy, which tends to decrease as the reflection density and diffuseness of the room response develop over time. Accordingly, planarity complements intensity-based diffuseness estimators, which quantify the degree to which the sound field at a discrete frequency within a particular time window is due to an impinging coherent plane wave. In this paper, we use planarity as a tool to analyse the sound field in relation to the RSAO parameters. Specifically, we use planarity to estimate two important properties of the sound field. First, as high planarity identifies the most localised reflections along the RIR, we estimate the most planar portions of the RIR, corresponding to the RSAO early reflection model and increasing the likelihood of detecting prominent specular reflections. Second, as diffuse sound fields give a low planarity score, we investigate planarity for data-based mixing time estimation. Results show that planarity estimates on measured multichannel RIR datasets represent a useful tool for room acoustics analysis and RSAO parameterisation.

Keywords: room acoustics, artificial reverberation

1. Introduction

Analysis of room impulse responses (RIRs) is an important topic in room acoustics, as they contain a complete snapshot of the acoustic environment. Two outcomes of RIR analysis are especially useful: understanding the time of arrival (TOA) and direction of arrival (DOA) of early reflections (which leads to room geometry estimation), and estimating the mixing time. In particular, for virtual acoustics applications such as auralization [1] and reverberation synthesis [2], these aspects are fundamental to assumed models of sound propagation in an enclosed space.

Recent advances in 3D audio rendering present new opportunities to create convincing reverberation based on spatial RIR analysis. At the same time, many different kinds of reproduction setup exist, including traditional channel-based systems, sound bars, and binaural reproduction over headphones. As such, there is a move towards object-based audio [3], wherein a piece of audio content can be produced once and rendered on a wide range of systems [4]. Consequently, there are opportunities to capture, edit, represent and render reverberation with a set of audio object parameters [5]. One recent example of a metadata scheme for object-based reverberation is the reverberant spatial audio object (RSAO) [6], where early reflections are represented as delayed, attenuated and filtered versions

of the direct sound, and the late tail is defined in terms of the mixing time and a set of octave-band exponential decays. These parameters can be estimated from spatial RIR measurements.

Much work has considered estimating the DOAs of reflections along portions of multichannel RIRs [7, 8], especially with applications in room geometry estimation [9, 10, 11, 12], which implies that the first order reflections are detectable in the RIRs. Recently, the topic of reliably estimating peak locations across multichannel RIR measurements was addressed by the clustered dynamic programming phase slope algorithm (C-DYPSA) [9]. However, as the sound propagates around a room, amplitude peaks corresponding to specular reflections become harder to detect. In this case, a measure of the specularity of a peak might help to identify the reflections corresponding to the most prominent (and perhaps perceptually meaningful) image sources.

Similarly, many techniques have been proposed to estimate the mixing time, i.e., when the sound field transitions from coherent early reflections to diffuse late reverberation [13]. In practice, two conditions for a mixed sound field measured at a single microphone array position are usually assumed: (1) that the temporal contributions of individual echoes to the RIRs cannot be distinguished, (2) that the spatial contributions of reflected energy are equally likely to arrive from any direction. Due to the large number of RIRs measured with single microphones, many methods focus purely on the temporal aspects of the mixing time definition [14, 15, 16, 17]. On the other hand, spatial RIRs measured with microphone arrays additionally allow estimation of mixing time based on the spatial properties of the sound field. Spatial sound field analysis for mixing time estimation is usually based on the sound intensity, which can be estimated from first-order [18] or higher-order [13] Ambisonic sound field decomposition, or by steered beamforming [7]. Spatial analysis was combined with temporal analysis in [19], where the variation in DOA between adjacent frames was used to estimate mixing time. The perceptual mixing time was measured in [20], and a linear regression model was subsequently used to predict the perceptual mixing time from an ensemble of model-based and data-based mixing time predictors. Among these predictors, the echo density [14] was found to be the most reliable.

In addition to the above methods to process multichannel RIRs, the metric of planarity [21] is a potentially useful tool for spatial RIR analysis. The planarity is related to the directional diffusion [7], but employs superdirective beamforming in arbitrary steering directions, leading to high resolution sound field analysis. Planarity is also related to the intensity-based diffuseness estimators [18, 19, 13], but is applicable to arbitrary microphone array geometries. The planarity metric and underlying DOA analysis was previously used to evaluate [22] and control [23] sound field reproduction. In this contribution, we use DOA analysis by superdirective beamforming, quantified by planarity, to analyse RIRs measured using a microphone array. The intention is to estimate both TOAs of the most planar reflectors and the mixing time, in order to populate parameters in the RSAO.

In Sec. 2, we introduce the sound field analysis and planarity metric, and in Sec. 3 we describe the experimental setup. Sec. 4 shows the results of the RIR analysis, and we conclude in Sec. 5.

2. Planarity

The planarity metric quantifies the extent to which the sound field measured with a microphone array resembles a plane wave. The RIRs can be written as $\mathbf{h}(n) = [\mathbf{h}_1(n), \mathbf{h}_2(n), \ldots, \mathbf{h}_M(n)]^T$, where there are M microphones and $\mathbf{h}_m(n) = [h(1)h(2)\ldots h(N)]$ is the vector of N RIR samples at the mth microphone. The corresponding frequency domain vector is obtained by applying a fast Fourier Transform (FFT) to the input samples, $\mathbf{h}(k) = \text{FFT}\{\mathbf{h}(n)\}$. The narrowband spatial energy distribution sampled at I angles is given by $w_i(k) = \frac{1}{2}|\psi_i(k)|^2$, where $\mathbf{w}(k) = [w_1(k), w_2(k), \ldots, w_I(k)]^T$ are the energy components at the ith angle and $\psi_i(k)$ is the corresponding plane wave component. A steering matrix $\mathbf{A}(k)$ of dimensions $I \times M$ (i.e., each row is equivalent to the array manifold vector for the ith steering direction) facilitates superdirective spatial analysis of the RIR vector,

$$\mathbf{w}(k) = \frac{1}{2}|\mathbf{A}(k)\mathbf{h}(k)|^2. \tag{1}$$

The elements of the matrix $\mathbf{A}(k)$ were populated by acoustic contrast beamforming, following [21]. An expanded description is given in [24, pp. 65–69]. The spatial spectrum $\mathbf{w}(k)$ is normalised for each frequency by scaling the energy in the maximal direction to unity.

The planarity η is the ratio between the intensity component due to the largest plane wave component and the total energy flux of plane wave components:

$$\eta(k) = \frac{\sum_i w_i(k)\mathbf{u}_i(k) \cdot \mathbf{u}_\alpha(k)}{\sum_i w_i(k)}, \tag{2}$$

where $\mathbf{u}_i(k)$ is the unit vector associated with the ith component's direction, $\mathbf{u}_\alpha(k)$ is the unit vector in the direction $\alpha = \arg\max_i w_i(k)$, and $\cdot$ denotes the inner product. Thus, it gives a measure of the proportion of the plane wave energy at the microphone array that can be attributed to the principal plane wave component. For a single plane wave, all of the energy can be attributed to the largest component and the score approaches 100%. Where a diffuse (or strongly-self cancelling) sound field is measured the score tends towards 0%.

For the current application of analysing a time-varying RIR, specular components (such as the direct sound and certain early reflections) should therefore give a high planarity, while later reverberation should give a lower planarity score, being increasingly spatially diffuse. The planarity estimation was applied as a piecewise estimate, populating $\mathbf{h}(n)$ with N samples and stepping S samples through the RIR before the next estimate. A Hamming window was applied prior to the FFT.

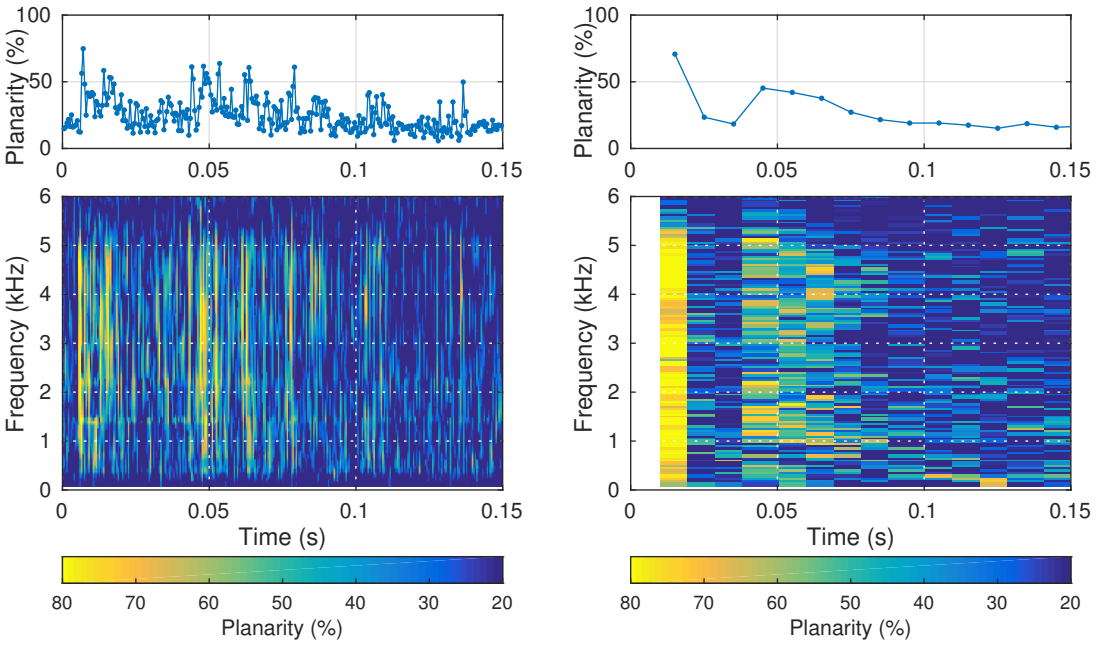
3. Experiment setup

Experiments were conducted on measured spatial RIRs from the S3A Room Impulse Response dataset [25], recorded with a 48 bi-circular channel microphone array [9, 26], with 24 omnidirectional microphones distributed equally around radii of 0.085 m and 0.106 m. Four rooms were tested: *CVSSP Vislab* with dimensions 7.79 × 7.73 × 3.98 m and reverberation time (T30) 0.27 s (averaged over 0.5–2 kHz); *Studio 1*, 14.55 × 17.08 × 6.50 m, T30 1.05 s; *Emmanuel Old Church*, 12.44 × 13.83 × 5.97 m, T30 1.24 s; *Emmanuel Main Church*, 19.68 × 24.32 × 5.97 m, T30 1.14 s. A single test RIR from each dataset was used.

Spatial aliasing effects when beam steering with this microphone array begin above 6.3 kHz, so, for the narrowband processing, the test RIRs were resampled to 16 kHz. The DOA analysis used $I = 360$ steering directions, i.e. one degree resolution in azimuth (the approach could straightforwardly be extended to steer in elevation, with a suitable microphone array geometry). Two sizes of analysis window were tested: $N = 16$, $S = 8$ samples (1 ms, 0.5 ms @ 16 kHz, intended to capture individual reflections), and $N = 480$, $S = 160$ samples (30 ms, 10 ms @ 16 kHz, intended to capture the trend of changing planarity across the RIRs). The FFT length was 2048.

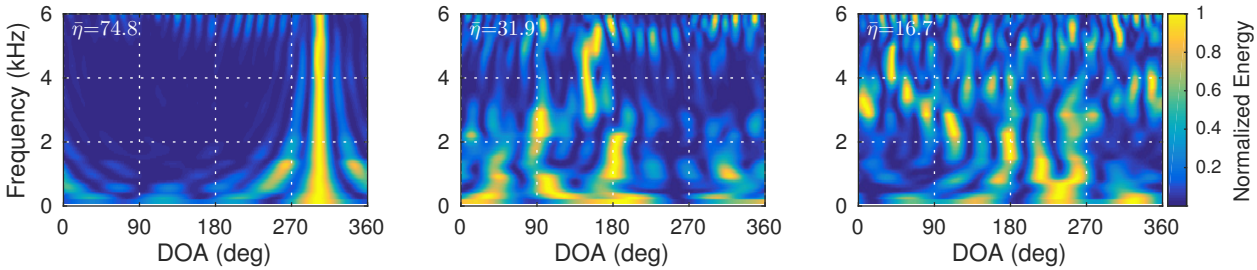
4. Results

The piecewise planarity estimation was applied to the test RIRs. An overview of the information derived from the planarity analysis is shown in Fig. 1, which shows the planarity estimates for each time window and frequency (lower plot), and the frequency-averaged planarity $\bar{\eta}$ over 1–6 kHz (upper plot), for Studio 1 with $N = 16$, $S = 8$ (Fig. 1a) and $N = 480$, $S = 160$ (Fig. 1b). With the 1 ms window (Fig. 1a), the most planar reflections along the RIR can clearly be identified. Some reflections, especially after 100 ms, are less broadband, but still give peaks in the upper (frequency-averaged) curve. With the longer 30 ms window (Fig. 1b) the expected decrease in planarity over time is evident. This might be helpful for estimating mixing time, as smaller peaks are averaged with the neighbouring diffuse energy. There is little planar energy after around 80 ms (compared to the T30 for this room of 1.05 s), which implies that the sound is mixed from a spatial perspective. Application



(a) 1 ms window, 0.5 ms step (b) 30 ms window, 10 ms step

Figure 1: Studio 1 Planarity for with respect to time and frequency, for different window sizes. The upper plot shows the planarity averaged over 1–6 kHz.

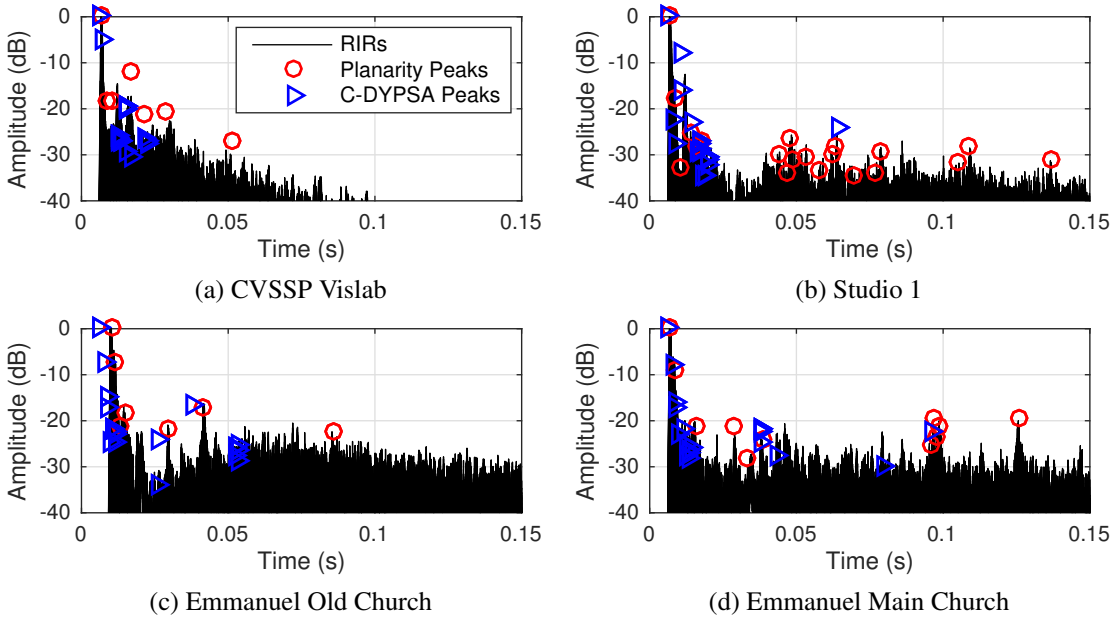


(a) Direct sound, $t = t_d + 0$ ms (b) Early reflections, $t = t_d + 12$ ms (c) Late reverberation, $t = t_d + 88$ ms

Figure 2: DOA vs frequency, showing planarity $\bar{\eta}$ averaged over 1–6 kHz for Studio 1, and indicating the time of the illustrated time window after the direct sound TOA t_d.

of the 1 ms window for early reflection detection and the 30 ms window for mixing time estimation, respectively, are investigated in Secs. 4.1 and 4.2.

In order to interpret the planarity scores, it is also useful to consider the underlying frequency-dependent spatial spectra. These are plotted, for three typical time segments of the RIR (based on 1 ms analysis) in Fig. 2. The frequency-averaged planarity $\bar{\eta}$ is also indicated on each plot. Fig. 2a shows the frequency-dependent DOA for a time segment with high planarity ($\bar{\eta}$=74.8), which corresponds to the direct sound. The direct sound is seen to be well localised over frequency (with the DOA corresponding to the measured loudspeaker position). Fig. 2b illustrates a time segment during the early reflections, 12 ms after the direct sound. The planarity ($\bar{\eta}$=31.9) is lower than expected, given that relatively little time has elapsed. In fact, inspection of the plot suggests that a few (frequency-dependent) planar reflections (e.g., at 95, 153, 185 degrees) may have arrived within the analysis window. For any given frequency, energy belonging to the simultaneous reflections lowers the planarity score. Finally, Fig. 2c shows a time segment later in the RIR (88 ms after the direct sound), with low planarity ($\bar{\eta}$=16.7). Here, it can be seen that there are multiple prominent plane wave components, distributed over space and frequency, which implies that the sound field is diffuse.



(a) CVSSP Vislab

(b) Studio 1

(c) Emmanuel Old Church

(d) Emmanuel Main Church

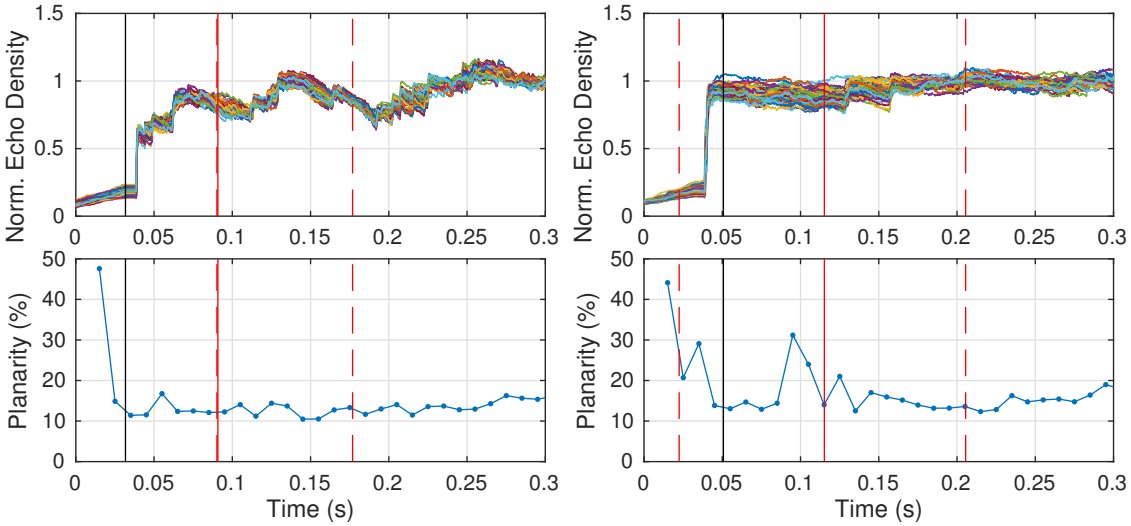
Figure 3: RIRs from the circular array (first 24 channels) overlaid with peak positions estimated from the planarity estimation (○) and C-DYPSA estimation (▷).

4.1 Early reflection estimation

Detecting early reflections is an important problem in room geometry estimation and auralization. Of particular interest to the authors is the problem of selecting the most specular reflections along the RIR, as it closely fits the assumed room model underlying the RSAO [6], which renders early reflections as discrete point sources. In particular, although it is possible to detect amplitude peaks along RIRs, we cannot always assume that the sound field corresponding to the peak is planar.

To detect planar reflections from the planarity analysis, the 20 highest planarity peaks were detected from the frequency-averaged planarity curves (e.g., Fig. 1, upper plots). The corresponding levels were estimated directly from the segmented RIRs, as the mean noise gain over the first 24 microphones, $L = \frac{1}{24}\sum_{m=1}^{24}\sqrt{\sum_N \mathbf{h}_m(n)^2}$. The detected peak times and levels are plotted against the squared RIRs in Fig. 3, for the four rooms tested. Also plotted, for comparison, are the peaks obtained by C-DYPSA [9], following [6]. These peaks are obtained by finding the amplitude peaks along each RIR channel, and clustering over all channels, considering each reflection separately, to reduce the effects of noise and spurious peaks. The levels corresponding to the C-DYPSA-detected peaks are also calculated using the noise gain, but this operates on the (single channel) output of a delay and sum beamformer, which forms an intermediate processing step in the current RSAO implementation (for DOA estimation) [6]. Some planarity-detected peaks, outside the time and amplitude ranges plotted, are not shown. In all cases, the planarity processing identifies peaks later in the RIRs than the C-DYPSA approach. For the Vislab (Fig. 3a) and Emmanuel Old Church (Fig. 3c), these planar reflections are indiscernible from fluctuations in the time domain amplitudes, but the energy in these portions of the RIR is more specular than that in the neighbouring segments. The Studio 1 results (Fig. 3b) are interesting, because there are specular reflections distributed throughout the first 100–150 ms of the RIRs. This implies that modelling the late reverberation as a diffuse, decaying tail might be insufficient. On the other hand, using the planarity estimation for peak detection in Emmanuel Main Church (Fig. 3d) ensures that the prominent peaks visible at around 95 ms and 125 ms are properly detected.

In summary, using peaks in the frequency-averaged planarity to detect early reflections along the RIR represents a spatially-aware approach that is able to detect specular reflections in multichannel RIRs.



(a) Vislab (b) Emmanuel Main Church

Figure 4: Normalised echo density and planarity (30 ms window, 10 ms step, averaged over 1–6 kHz). Perceptual mixing times t_{mp50} also shown from model-based (black) and data-based (red; median (solid), min/max(dashed)) estimators [20].

4.2 Mixing time estimation

Estimating the mixing time is also an important topic in room acoustic auralization and object-based reverberation rendering. The normalised echo density [14] is widely used to predict the mixing time, based on the assumption that reflections arriving during the mixed part of the RIR have a Gaussian distribution. In [20], this method was found to be a reliable predictor of the perceptual mixing time (when listeners could not discern between measurements made at different positions within a room). Examples of normalised echo density profiles are shown alongside the frequency-averaged planarity (30 ms window) in Fig.4, for CVSSP Vislab (Fig. 4a) and Emmanuel Main Church (Fig. 4b). In addition, the predicted perceptual mixing times t_{mp50} (at which 50% of listeners found the test stimuli to be perceptually mixed [20]) are shown with vertical lines. The red lines show the data-based t_{mp50} predictions for all 48 microphone channels (solid line median; dashed lines min/max), while the black lines indicate the equivalent model-based t_{mp50} prediction, based on the room geometry. For all rooms under test, there is a common profile to the echo density. However, the resultant mixing time predictions, which depend on the point at which the normalised echo density reaches unity, vary considerably over the 48 microphones. Although a different threshold on echo density could be considered, it is clear from comparing Fig. 4a and Fig. 4b that this might still be difficult to set.

On the other hand, planarity analysis can help identify the time segment where the initial specular reflections have finished arriving. It also reveals that specular reflections may arrive after the perceptual mixing time (e.g., the model-based t_{mp50} for Emmanuel Main Church is 51 ms, but there are a cluster of specular reflections at around 100 ms). Thus, although the sound appears to be mixed temporally, it is not mixed spatially. Further perceptual study is necessary to understand the balance between these definitions of a mixed sound field. If the spatial component of mixed sound field perception is important, the planarity might be a useful tool to help identify a mixing time, assuming that appropriate thresholding or post-processing can be developed.

5. Conclusions and Future Work

Overall, the results presented above have shown planarity analysis to be a useful tool for RIR analysis, with potentially applications in detecting specular peaks in an RIR and estimating a room's

mixing time. Using the planarity approach, specular reflections were detected along the RIR, even when the local peak amplitude did not exceed that of its neighbours. This spatial analysis, utilising signals from all array channels, also ensured that prominent peaks were not discarded in the clustering stage. In particular, specular peaks occurring during the decaying reverberation might lead to more convincing spatial reverberation in the RSAO framework. Furthermore, using a longer time window illustrated the longer term variation of planarity along the RIR. This might be useful, with further processing, to estimate the mixing time based on spatial diffuseness.

Future work will investigate the perceptual differences between early reflections encoded using C-DYPSA compared to using the planarity peaks, when rendered as reverberant spatial audio objects, and investigate further processing of the planarity to estimate the mixing time.

6. Acknowledgements

This work was supported by the EPSRC Programme Grant S3A: Future Spatial Audio for an Immersive Listener Experience at Home (EP/L000539/1) and the BBC as part of the BBC Audio Research Partnership. The authors would like to thank Luca Remaggi for making available the C-DYPSA implementation.

REFERENCES

1. Savioja, L., Huopaniemi, J., Lokki, T. and Väänänen, R. Creating interactive virtual acoustic environments, *J. Audio Eng. Soc.*, **47** (9), 675–705, (1999).

2. Välimäki, V., Parker, J. D., Savioja, L., Smith, J. O. and Abel, J. S. Fifty years of artificial reverberation, *IEEE Trans. Audio Speech Lang. Proc.*, **20** (5), 1421–1448, (2012).

3. Herre, J., Hilpert, J., Kuntz, A. and Plogsties, J. MPEG-H audio — the new standard for universal spatial/3D audio coding, *J. Audio Eng. Soc*, **62** (12), 821–830, (2015).

4. Shirley, B., Oldfield, R., Melchior, F. and Batke, J.-M., (2013), Platform independent audio. *Media Production, Delivery and Interaction for Platform Independent Systems*, pp. 130–165, John Wiley & Sons.

5. Coleman, P., Franck, A., Jackson, P. J. B., Hughes, R., Remaggi, L. and Melchior, F. On object-based audio with reverberation, *Proc. 60th AES Int. Conf., Leuven, Belgium, February*, (2016).

6. Coleman, P., Franck, A., Jackson, P. J. B., Hughes, R., Remaggi, L. and Melchior, F. Object-based reverberation for spatial audio, *J. Audio Eng. Soc.*, **65** (1/2), 66–77, (2017).

7. Gover, B. N., Ryan, J. G. and Stinson, M. R. Microphone array measurement system for analysis of directional and spatial variations of sound fields, *J. Acoust. Soc. Am.*, **112** (5), 1980–1991, (2002).

8. Sun, H., Mabande, E., Kowalczyk, K. and Kellermann, W. Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing, *J. Acoust. Soc. Am.*, **131** (4), 2828–2840, (2012).

9. Remaggi, L., Jackson, P. J. B., Coleman, P. and Wang, W. Acoustic reflector localization: Novel image source reversion and direct localization methods, *IEEE/ACM Trans. Audio. Speech Lang. Proc.*, **25** (2), 296–309, (2017).

10. Tervo, S. and Tossavainen, T. 3D room geometry estimation from measured impulse responses, *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 513–516, IEEE, (2012).

11. Antonacci, F., Filos, J., Thomas, M. R., Habets, E. A., Sarti, A., Naylor, P. A. and Tubaro, S. Inference of room geometry from acoustic impulse responses, *IEEE Trans. Audio Speech Lang. Proc.*, **20** (10), 2683–2695, (2012).

12. Dokmanić, I., Parhizkar, R., Walther, A., Lu, Y. M. and Vetterli, M. Acoustic echoes reveal room shape, *Proceedings of the National Academy of Sciences*, **110** (30), 12186–12191, (2013).

13. Götz, P., Kowalczyk, K., Silzle, A. and Habets, E. A. P. Mixing time prediction using spherical microphone arrays, *J. Acoust. Soc. Am.*, **137** (2), EL206–EL212, (2015).

14. Abel, J. S. and Huang, P. A simple, robust measure of reverberation echo density, *121st Conv. Audio Eng. Soc., San Francisco, CA, USA*, (2006).

15. Hidaka, T., Yamada, Y. and Nakagawa, T. A new definition of boundary point between early reflections and late reverberation in room impulse responses, *J. Acoust. Soc. Am.*, **122** (1), 326–332, (2007).

16. Stewart, R. and Sandler, M. Statistical measures of early reflections of room impulse responses, *Proc. of the 10th int. conference on digital audio effects (DAFx-07), Bordeaux, France*, pp. 59–62, (2007).

17. Defrance, G., Daudet, L. and Polack, J.-D. Using matching pursuit for estimating mixing time within room impulse responses, *Acta Acustica united with Acustica*, **95** (6), 1071–1081, (2009).

18. Pulkki, V. Spatial sound reproduction with directional audio coding, *J. Audio Eng. Soc.*, **55** (6), 503–516, (2007).

19. Ahonen, J. and Pulkki, V. Diffuseness estimation using temporal variation of intensity vectors, *Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 285–288, IEEE, (2009).

20. Lindau, A., Kosanke, L. and Weinzierl, S. Perceptual evaluation of model-and signal-based predictors of the mixing time in binaural room impulse responses, *J. Audio Eng. Soc.*, **60** (11), 887–898, (2012).

21. Jackson, P. J. B., Jacobsen, F., Coleman, P. and Pedersen, J. A. Sound field planarity characterized by superdirective beamforming, *Proceedings of Meetings on Acoustics*, vol. 19, p. 055056, Montreal, 2-7 June 2013, (2013).

22. Coleman, P., Jackson, P. J. B., Olik, M., Møller, M., Olsen, M. and Pedersen, J. A. Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array a, *J. Acoust. Soc. Am.*, **135** (4), 1929–1940, (2014).

23. Coleman, P., Jackson, P. J. B., Olik, M. and Pedersen, J. A. Personal audio with a planar bright zone, *J. Acoust. Soc. Am.*, **136** (4), 1725–1735, (2014).

24. Coleman, P., *Loudspeaker Array Processing for Personal Sound Zone Reproduction*, Ph.D. thesis, University of Surrey, Guildford, Surrey, GU2 7XH, UK, (2014).

25. Coleman, P., Remaggi, L. and Jackson, P. J. B., (2015), *S3A Room Impulse Responses*. https://doi.org/10.15126/surreydata.00808465.

26. Remaggi, L., Jackson, P. J. B., Coleman, P. and Francombe, J. Visualization of compact microphone array room impulse responses, *AES 139th Conv. (e-Brief), New York, NY, USA*, (2015).