



HAL
open science

Source Distance Perception with Reverberant Spatial Audio Object Reproduction of Real Rooms

Sandeep Chitreddy, Philip Jackson

► **To cite this version:**

Sandeep Chitreddy, Philip Jackson. Source Distance Perception with Reverberant Spatial Audio Object Reproduction of Real Rooms. Forum Acusticum, Dec 2020, Lyon, France. pp.2079-2086, <10.48465/fa.2020.0883>. <hal-03235362>

HAL Id: hal-03235362

<https://hal.science/hal-03235362v1>

Submitted on 27 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

SOURCE DISTANCE PERCEPTION WITH REVERBERANT SPATIAL AUDIO OBJECT REPRODUCTION OF REAL ROOMS

Sandeep Reddy Chitreddy¹

Philip J.B. Jackson²

¹ Department of Electronics and Communication Engg., Koneru Lakshmaiah Education Foundation, Hyderabad, India

² Centre for Vision, Speech and Signal Processing, University of Surrey, United Kingdom

csreddyitk@klh.edu.in, p.jackson@surrey.ac.uk

ABSTRACT

A sound source's apparent distance provides information about spatial relations that can have primary salience relative to other dimensions, including azimuth and elevation in the scene of a story told in 3D. Having methods to create a sense of presence and to control this attribute in immersive content are therefore valuable capabilities in sound design, particularly in object-based audio. This paper examines the ability of the reverberant spatial audio object (RSAO) to cue apparent source distance, using acoustic parameters extracted from publicly-available spatial room impulse responses (RIRs) of real environments measured at a range of source-receiver distances. The RSAO's spatio-temporal reverb representation derived from the directional B-format RIRs encoded the timing, direction and timbre of early reflections, as well as the onset, colouration and decay times of the late reverberation, which were rendered over a 42.1 setup in an acoustically-treated listening room to provide a quasi-transparent pipeline for the reproduced room impression, from recording to listening. An objective analysis of re-synthesised RIRs and re-estimated parameters demonstrated the pipeline's transparency for the majority of parameters, but spectral leakage of band-pass filters in the late tail promoted reverb time convergence across bands. Formal listening tests evaluated the apparent source distance that participants perceived via a multi-stimulus rating method. Statistical analysis indicates participants perceived reproduced distance changes, with logarithmic distance resolution inside the rooms' critical distance. Beyond this, ratings tended to saturate. These effects were clearer in the large hall than in the classroom, and for the voice source than the percussion. Overall, results suggest that the RSAO can provide appropriate cues for source distance perception with resolution comparable to natural sound fields. Further work will investigate how distance perception performs across reproduction setups and develop methods to extrapolate source distance by adapting the RSAO parameters.

1. INTRODUCTION

Distance perception is one of the natural capabilities of the human auditory system. The human ability to perceive

sound source distance is not as sensitive as directional perception of the sound source. Nevertheless, it offers complementary information for localisation, enabling the 3D spatial position of a source to be perceived.

Distance perception depends on the environment in which the source signal measurement is performed. Signal energy concentrates more in closed environments like rooms, concert halls or auditorium, as compared to open external environments. Nevertheless, there was a study on sound source distance perception in open environments ranging from 25m to 800m [1]. This study subjectively investigated the distances above 25m may be under- or over-estimated as compared to the physical distance. But distance perception study in indoor environments is interesting and has many applications. Particularly in entertainment applications, source distance control enables a producer and a listener (through interaction) to create an impression of a source being shallower or deeper in an acoustic scene.

One of the strongest cues for distance perception is the sound intensity. But intensity is functions primarily as a relative cue [2]. Most of the studies agree that the reverberation of the environment is as an absolute cue in the distance perception [3,4]. Apart from the sound intensity and reverberation, spectral cues particularly high frequency attenuation is observed with the increase in source distance. For distances greater than 15m, air absorption modifies sound significantly and can be viewed as low pass filtering effect [5]. There are other cues like binaural cues for nearby peri-personal sounds, and dynamic cues that have also been found to influence distance perception [5].

In this work, we mainly focus on the effect of reverberation on source distance perception in the context of object-based sound reproduction. The Reverberant Spatial Audio Object (RSAO) [6] is a method to parameterise the spatial room acoustic information for reproduction in the form of object-based spatial audio [7]. The object-based spatial reverb has been earlier studied and evaluated for different perceptual attributes [8]. This work investigates the RSAO's relation to the source distance attribute and conducts a reverb parameter analysis across different source-receiver distances. Subsequently, formal subjective listening tests are conducted to study the efficacy of RSAO

in being able to reproduce sound that creates the impression of source distance. We also study the effect of source type and room on the apparent source distance.

2. RSAO PARAMETER ANALYSIS FOR A SET OF SOURCE-RECEIVER DISTANCES

In this section, RSAO parameter analysis for a set of source receiver distances is discussed. First, we present the RSAO parameters that are extracted from the measured B-format RIRs. Subsequently, the spatial correspondence of these parameters is illustrated by visualising the parameters. Finally, the accuracy of the parameters is objectively verified using a synthesis-resynthesis approach.

2.1 Parameter Extraction

The RSAO specifies a set of parameters to encode room acoustic information and reproduce reverb in an object-based spatial audio format [6]. A Matlab implementation of the RSAO parameterisation methods was used to encode the measured B-format room impulse responses [9]:

- Direct and Early parameters: levels and DOAs of direct and early components, filter coefficients describing the colouration for each early reflection,
- Late parameters: the audio frequency range is divided in to 9 octave bands centered around 62.5Hz, 125Hz, 250Hz, 500Hz, 1kHz, 2kHz, 4kHz, 8kHz, 16kHz. Late reverberations levels, the decay times and the onset times are extracted for each band.

Figures 1 and 2 indicate that RT60 values and late levels extracted from a publicly available B-Format RIR database of Classroom and Octagon Hall [10] for all the nine octave bands. More information about the database can be seen in Section 3.2. Bar plots indicate the values mentioned in the database documentation (navy bars) and the ones in blue indicate the measured values. The box plot indicates the variation of the corresponding late parameter for all the source-receiver distances. The values show that the reverberation times and late levels are tightly grouped across all distances for most of the bands, except the low frequency bands at 125Hz and 250Hz. The green and yellow bars are extracted from synthesised and resynthesised RIRs respectively, which are further discussed in Section 2.3.

2.2 Parameter Visualisation

RSAO parameters of the early reflections can be better visualised in the context of spatial beamformed plots [11], which show the energy and directions of arriving early reflections over time. These plots are extracted by conducting beamforming over the range of azimuthal directions. In this case using B-format microphone signals, steering was achieved by directing a virtual cardioid microphone around the horizontal plane. This process is performed at regularly-spaced azimuths with a spacing of 1° . The beamformed signals obtained for each azimuthal angle are grouped together as each individual column of a matrix

and displayed as an image. The spatial beamformed plots obtained from the Classroom and Octagon Hall RIRs are shown in Figure 4 for a range of distances. These plots are very useful as they can aid in establishing the correspondence between the early parameters and the early reflections of a room. For example, in Figure 4 (left) as the distance between source and receiver increases, direct level decreases, onset time (delay) of direct sound increases, and the time gap between the direct and first reflection decreases. The first dominant reflection from around 0° can indicate either a floor or ceiling reflection of the cuboidal Classroom. Reflections at $-180^\circ/180^\circ$ indicate the reflections from the rear wall of the Classroom. These reflections start as late as 35ms (25ms after the direct sound) for 1m source-receiver distance and slowly approach 25ms (5ms after direct) for 5.5m. In the Octagon Hall (right), reflections are mostly directed to the front, whose effect can be seen in the steered beam plots. The first dominant reflection around 0° azimuth comes from the floor, not from the Hall's domed 21m-high ceiling. With double the Classroom's 0.5-m steps in consecutive measurements, the Octagon Hall shows larger shifts between its 1-m steps.

2.3 Parameter Validation

One way of quantifying the efficacy of the RSAO-based reproduction is to compute the RSAO parameters for synthesised and resynthesised RIRs and to verify the consistency between them. RSAO parameters are initially computed from the measured RIRs for subsequent object based reproduction. Practically synthesised/resynthesised RIRs may be obtained by recording the object-based impulse rendered and reproduced over a large array of loudspeakers. This rendered impulse is referred to as a synthesised RIR. The parameters extracted from these synthesised RIRs can be validated to test the efficacy of RSAO-based rendering and reproduction. The aforementioned cycle can be computed once again up on the synthesised parameters to obtain resynthesised parameters. However, in this work, the whole process of synthesis and resynthesis is simulated via a digital audio workstation (DAW) rather than physical measurements of reproduced sound. In this case, an impulse signal was passed through through a suite of VISR VST plugins within the Reaper DAW, namely the reverb plugin, scene master and the loudspeaker renderer (v0.10.1), all of which were developed as part of the S3A project [12]. In this manner, the late parameters were analysed. Figure 2 and 1 show the late levels and RT60 values computed for all the nine octave bands for both synthesised (green) and resynthesised (yellow) RIRs. The synthesised and resynthesised values approximate those for the measured reverberation times for all bands except the low-frequency 125Hz and 250Hz bands.

3. ASD EXPERIMENTS USING RSAO BASED SOUND REPRODUCTION

In this section, the details of the ASD experiment using RSAO based sound reproduction is discussed. First, ob-

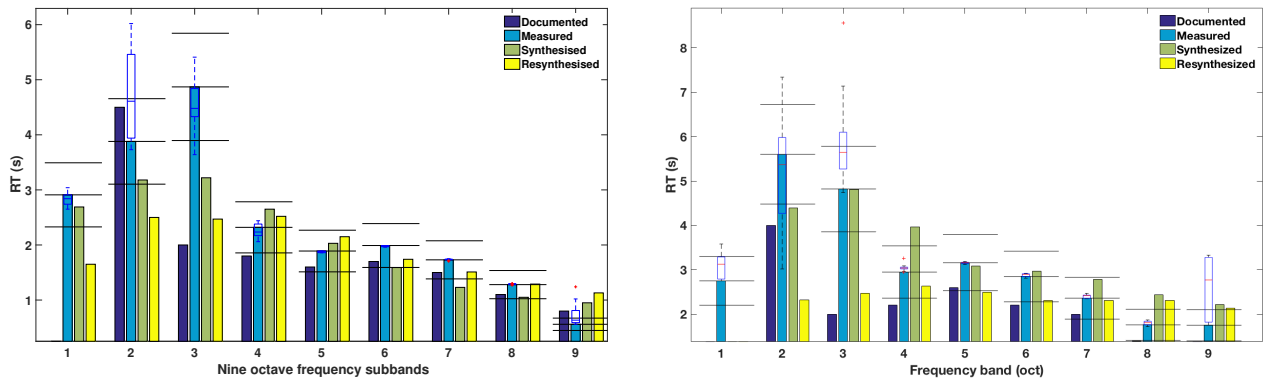


Figure 1. RT60 reverberation times for the Classroom (left) and Octagon Hall (right) for the nine octave bands from the accompanying documentation (Documented, navy), measurements (Measured, light blue), synthesized (Synthesized, green) and re-synthesized (Resynthesized, yellow). The black lines show tolerance ranges about the measured values. The blue box plots show the distribution across measurements.

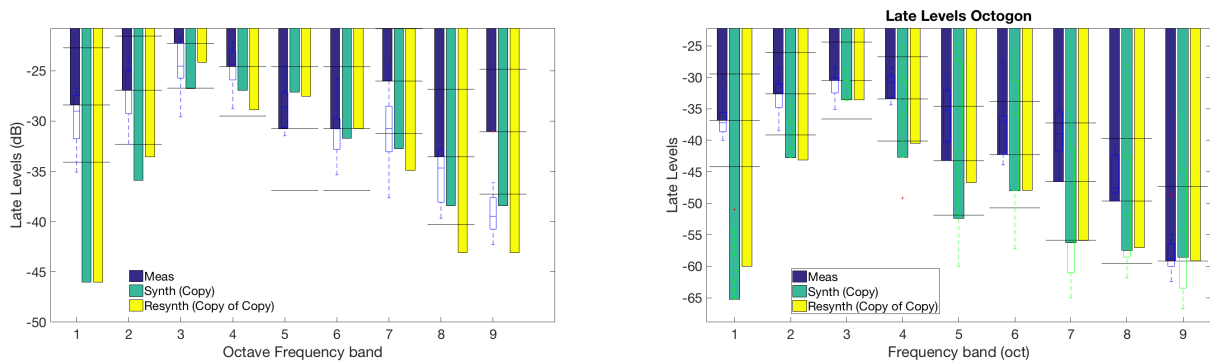


Figure 2. Late reverberation levels for the Classroom (left) and Octagon Hall (right) for the nine octave bands from original measurements (Meas, navy), synthesized (Synth, green) and re-synthesized (Resynth, yellow). The black lines show tolerance ranges about the measured values. The blue box plots show the distribution across measurements.

jectives of the experiments are discussed. Subsequently these objectives are transformed in to an experimental hypothesis. Later the databases used in the experiments are presented. Further experimental conditions, experimental calibration and experimental process are discussed.

3.1 ASD Experimental Hypothesis

The main objective of conducting ASD perceptual tests using RSAO based sound reproduction is, to test the efficacy of RSAO algorithm in being able to reproduce a sound that creates a perception of source distance. This experiment will investigate the resolution with which participants were able to discriminate the parameterised distances, which are called here Perceptually Distinguishable Distances (PDDs). Further the effect of source and room on distance perception are also statistically tested here.

The formulated hypothesis is dependent on the following question: How precisely can participants rate the perceived distance of a source in an unknown room, given reference distances from another room where both are reproduced using RSAO? Considering N stimuli reproduced using RSAO arranged in a random order in terms of their distances, the statistical negation of the null hypothesis that

their means of these distance ratings are equal will determine a PDD. The null hypothesis is formulated as:

$$H_0 : \mu_0 = \mu_1 = \mu_2 = \dots \mu_{N-1} \quad (1)$$

where, μ_i is the mean of the i^{th} distance perceptual scores obtained from K participants. Statistical testing of the above hypothesis determines whether the RSAO is effective in reproducing the perceived distance of a sound. However the resolution of PDD can only be obtained by performing statistical analysis across pairs of distances. Further, in order to understand the effect of source and room, the above hypothesis should be tested for a particular distance with different source and room setting. The next section discusses these issues, but first in the ensuing subsections we outline the other details of the experiment.

3.2 Selection of Reproduced Rooms

There are few spatial RIR databases available in the public domain that have measurements across multiple distances. This work has utilised measurements made in three acoustic environments found in two databases of B-format RIRs: recordings in a small teaching room (*Classroom*) and a large hall (*Octagon Hall*) from Queen Mary, University

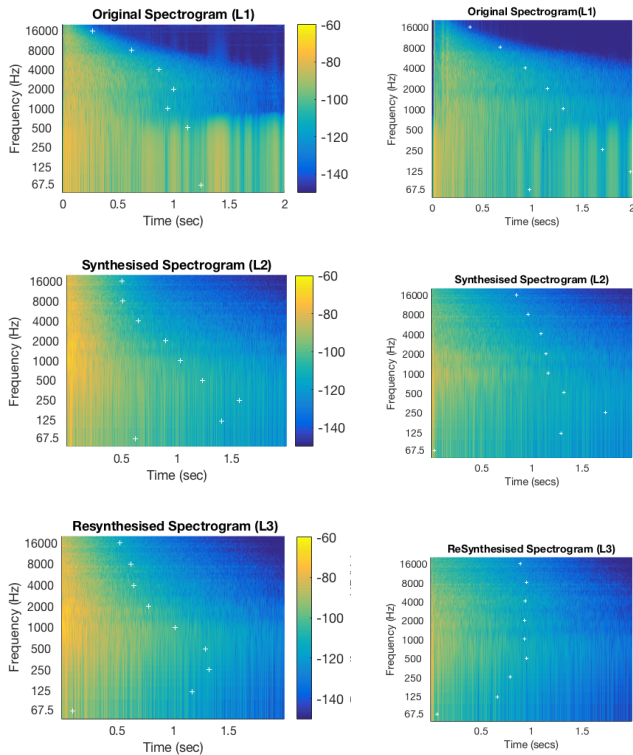


Figure 3. Spectrogram plots of late reverberation: (from top) for measured, synthesized, and re-synthesized RIRs; Classroom (left), Octagon Hall (right).

Room	Vol.	RT60	Crit. dist.
Classroom	240m ³	1.7s	0.7m
Octagon	9500m ³	2.1s	3.8m
Pori Hall	9300m ³	4.6s	2.6m
Audiobooth	43m ³	0.15s	1.0m

Table 1. Room properties of reproduced test and reference rooms, and the listening room: their volume (Vol.), average reverberation time (RT60) across the 500Hz, 1kHz and 2kHz bands, and critical distance (Crit. dist.) estimated with previous columns' values.

of London [10], and a large concert hall (Pori Hall) from Helsinki University of Technology [13].

Classroom has recordings from 1.0m to 5.5m distance with a 0.5m spacing. Octagon Hall has recordings from 2m to 13m with 1m spacing, although measurements only up to 11m were used in this work. Together these spanned more than an order of magnitude in distance for sound sources indoors and gave measurements at 10 distance settings in each environment. As can be seen from Table 1, most of the distances exceeded the critical distances for those rooms (all but two in the Octagon Hall), according to those estimates based on the measured reverberation times. Both the parameter analysis and subjective experiments used RSAO parameters extracted from the B-format RIRs of Classroom and Octagon Hall. The Pori Hall B-format RIRs provided RSAO parameters for the reference distances in the subjective listening-test experiments.

3.3 Experimental Conditions

The subjective listening tests for formal evaluation of ASD required careful design of the experimental conditions, as follows.

Source Type: Stimuli were chosen to include sources of two types, speech and non-speech, representing familiar and less familiar sounds respectively. Distance prediction with less familiar sounds, like a specific type of noise, is found to be harder as compared to familiar sources [14]. The primary objective of our work is not to comprehensively analyse all the types of source, but to validate the performance of RSAO under various conditions. So we sought to select commonly-used sources from the literature for the analysis in this work. For the speech stimuli, a female voice uttering "In 1799" (1.6 sec) was employed from the Archimedes database [15]; for the non-speech stimuli, the percussion sound of drums was chosen (1 sec), which includes a series of impulsive beats. In the analysis below, we refer to these as *Voice* and *Drums* respectively.

Reproduction Setup: A 42.1 loudspeaker setup was used for sound reproduction, which was arranged as a superset of the standard 22.1 setup [16]. A 360° image shows the spherical setup in the Audiobooth in Figure 5, and the configuration angles are marked in Figure 6. This layout was chosen in order to provide high spatial resolution of the sound reproduction and thereby minimize any potential associated artifacts.

Reference Distances: To anchor the rating scale for ASDs across both test rooms, two reference distances were specified from the Pori Hall. Since distance perception tends to show approximately logarithmic compression [5], the full range from 1.0m to 11m was divided into three parts with anchors at 2.5m and 6.0m with the individual room ranges overlapping in the central part.

Recruitment of Participants: Given the need for critical listening, 26 expert subjects working in the domain of speech and audio were recruited to participate in the experiments. Participants were presented with a questionnaire asking their age, gender, prior experience in listening tests, and screening for any hearing impairment. Participants were instructed in the use of the experimental interface which included a training example.

3.4 Experimental Calibration

In this experiment, 42 Genelec 8020A studio monitors were used for sound reproduction along with the subwoofer 7050B. The 42.1 spherical loudspeaker setup was rigged in the audiobooth at University of Surrey, as shown in Figure 5. The audiobooth is an acoustically-treated room with equal reverberation time from 300Hz to 8kHz, complying with ITU-T.P.800 in this frequency range. The loudspeaker setup is depicted in Figure 6. This configuration is a superset of the standard 22.1 setup. Each loudspeaker was positioned at the specified angles using a Bosch PCL 20 Cross Line Laser Level device. All 42 loudspeakers were connected to 2 RME M-32 DA converters via a single port of the MADI interface. These

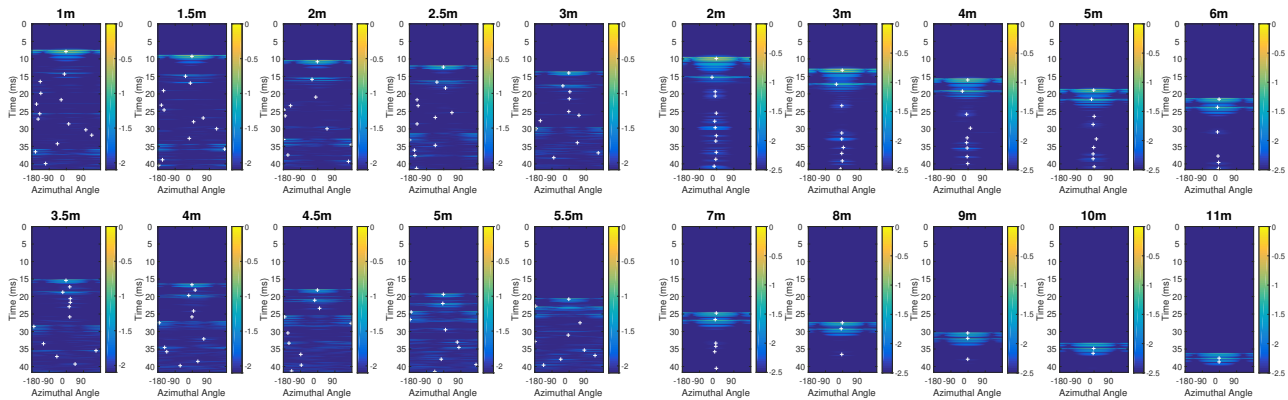


Figure 4. Spatial beamformed plots of steered energy across azimuth and over time: (left) for Classroom from 1.0m to 5.5m in 0.5-m steps and (right) for Octagon Hall from 2m to 11m in 1-m steps.



Figure 5. Panoramic photograph of the 42.1 loudspeaker setup used in the audiobooth at University of Surrey.

loudspeakers were calibrated to equal A-weighted SPL using NTI Minilyser by playing pink noise independently through each loudspeaker and recorded using the omnidirectional microphone attached to the Minilyser. Similarly, the subwoofer was also calibrated at the first octave frequency, i.e., 62.5Hz.

3.5 Experimental Process

Subjective experiments were controlled via a MAX/MSP interface. Figure 7 shows the experimental interface for one page, which presents in a random order 10 test stimuli to be rated and 2 references to anchor the otherwise unlabelled rating scale. Participants played each test stimulus, listened to the reproduction over the 42.1 setup and recorded their assessment of apparent source distance. The references were set at 2.5m and 6.0m in the Pori Hall, whereas the test stimuli were all from either the Classroom or the Octagon Hall at various distances. Participants gave a score 0-100 for each test stimulus with reference to the near (2.5m at 33) and far (6.0m at 67) anchors. Computer keys 1, 2, ... 0 could be used to play the respective test stimuli; keys Z and A for the near and far references. Participants used sliders to rate their scores, and were provided with a sort facility to re-order the 10 test stimuli according to their ratings to check and adjust the scores. Finally participants were stored the scores via the save button.

Four distinct pages presented all combinations of two

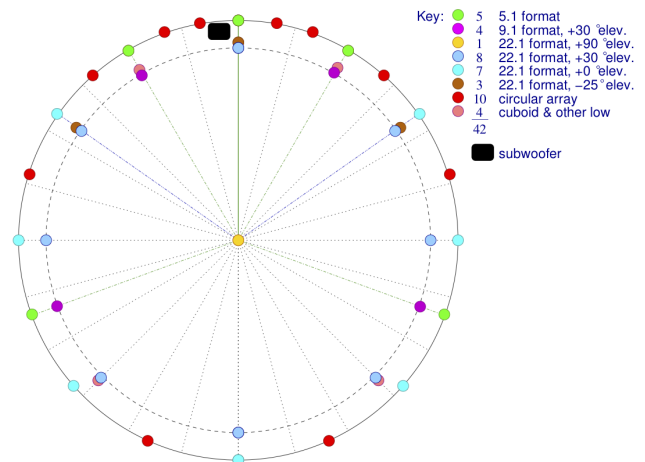


Figure 6. Plan view of the 42.1 loudspeaker setup in which the listener sits in the centre, facing ahead towards the top. The colour of the filled circles denotes loudspeaker position's membership of 5.1, 9.1, 22.1 and cuboid subset configurations.

test rooms with two source types. In total, there were 8 pages to allow two trials per combination. Thus, there were 80 ratings per participant with 10 test stimuli per page. Participants performed the tests in two sessions with a break to reduce fatigue. Each session lasted no more than 45 min.

4. LISTENING TEST RESULTS

In this section, statistical results obtained from the subjective experiments are presented. First, the consistency of the data are inspected for screening setting giving rise to peculiar outliers or participants who gave unreliable ratings, to ensure that the observed trends are representative. Next, we examine summaries of the data to characterise the main trends, particularly in terms of ASD ratings with respect to distance and the statistically significant differences. Further analysis makes a comparison of data for the reproduced test rooms to explore any observed similarities and differences. Results were obtained from 24 participants for the Voice source type across both test rooms and

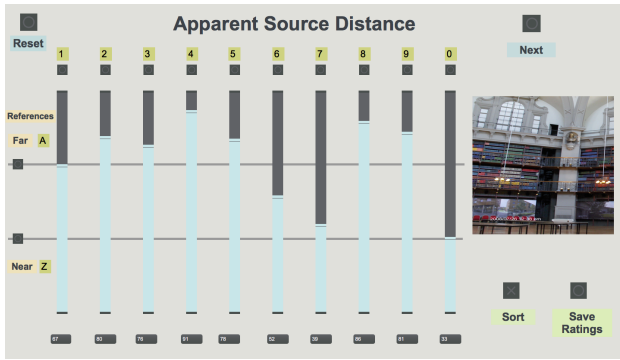


Figure 7. Octagon Hall page of the Max/MSP interface for subjective ASD ratings, showing rating scales for 10 test stimuli (1..0), far/near references (A/Z), and buttons to reset, move to next page, sort and save.

all distances, including repetitions; 26 participants' scores were obtained for the Drums. However, for the sake of consistency, balance and comparisons across both source types, only 24 participants' scores are analysed.

4.1 Data Integrity Analysis

Comparing the repeated measurements from the two trials of identical stimuli by each participant, a measure of the measurements' variability is obtained, which is expressed here as root-mean-square (RMS) repetition error. All participants' ratings at least spanned the range of the anchors, on average using over 80% of the scale. Across all participants and test conditions, the overall RMS error was 17.5 points, which is approximately half the interval between the reference anchors and between the ends of the scale and the anchors. This suggests that the task was fairly taxing for most participants, and yet that they were able to provide relevant information in their ratings.

Looking at the experimental variables, the errors were similar for Classroom and Octagon Hall. No systematic pattern was discerned with respect to distance, and the size of the errors was largely consistent across all distances. Similar errors were found between Voice and Drums, within 2 points across conditions.

The RMS repetition error being skewed by a small number of relative outliers, the substantial majority of participants had an average repetition error below 20 points. Comparing the performance of individual participants, a few exhibited somewhat higher RMS errors. No participant was universally poor in providing ratings, yet some exhibited difficulties with certain source-room combinations. However, removing this subset of scores had little influence on the analysis and left the key findings unchanged, so their data were included in the following analysis.

4.2 ANOVA Findings

Three sets of Analysis of Variance (ANOVA) were conducted with multi-factor mixed-effects models for repeated measures, using the `fitrm` and `ranova` functions in Matlab statistics toolbox (R2019a). First, separate anal-

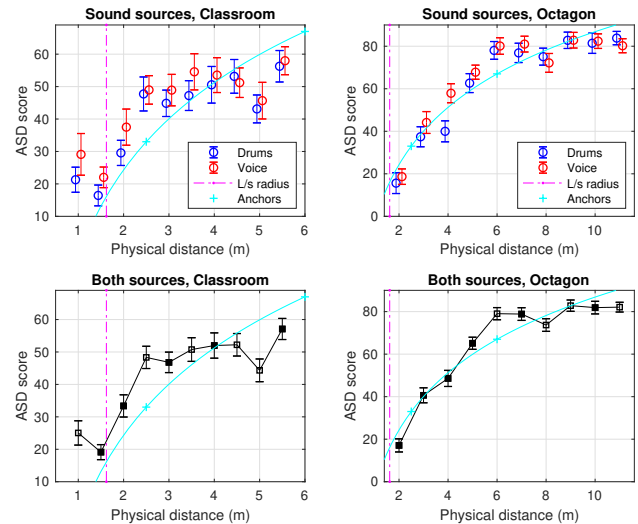


Figure 8. ASD ratings across distance for each room, averaged across participants and repetitions: Classroom (left) and Octagon Hall (right). The upper plots show results for each source: Drums (blue) and Voice (red); results in the lower plots are combined (black). Error bars indicate 95% confidence intervals of ± 2 standard errors. The distance of the loudspeakers in the reproduction environment is marked as a vertical magenta dash-dot line. The reference anchors from the Pori Hall are shown as cyan crosses (+) lying on the log-distance function.

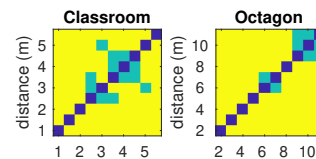


Figure 9. Statistically significant differences (yellow) in ASD ratings amongst all distance pairs for each room, $\alpha = 0.05$: Classroom (left) and Octagon Hall (right).

yses were performed for each test room, Classroom and Octagon Hall. These involved 4 factors: 1 between-subjects [Participant], 3 within-subjects [Distance, Source, Trial]. Then, for the common distances that they shared (i.e., {2m, 3m, 4m, 5m}), a combined analysis was done. This involved 5 factors: 1 between-subjects [Participant], 4 within-subjects [Distance, Source, Trial, Room].

All the analyses showed significant effects for Distance, but the other identified effects differed. In the separate room analyses, they both identified the Participant*Source interaction and one other factor: Source for Classroom, Trial for Octagon Hall. The Participant*Source interaction may have been due to some participants' difficulty with certain source-room combinations. However, in the combined-rooms analysis, the only additional effect was the Distance*Room interaction. Thus, the main effect that was seen consistently across the data was the desired effect for Distance. The details of this will be studied in the following subsections, and a final analysis in Section 4.5 sheds light on the Distance*Room interaction effect.

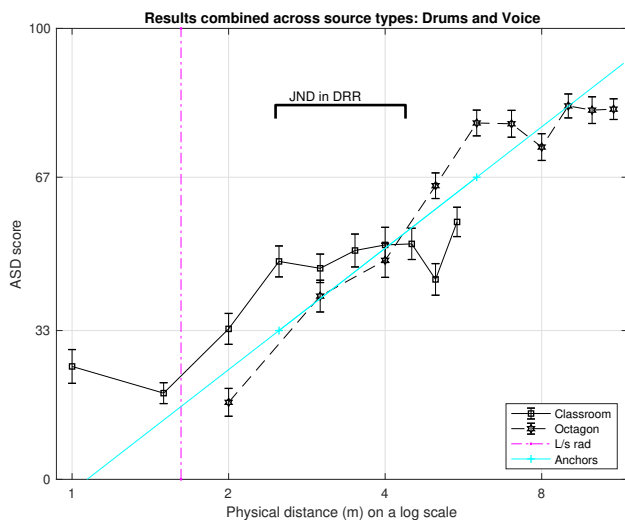


Figure 10. ASD ratings for both test rooms versus log-distance. Error bars indicate 95% confidence intervals of ± 2 standard errors. The distance of the loudspeakers in the reproduction environment is marked as a vertical magenta dash-dot line. The reference anchors from the Pori Hall are shown as cyan crosses (+) lying on the log-distance function. The Just Noticeable Difference (JND) in Direct-to-Reverberant Ratio (DRR) is indicated, based on uniform reverberation and a 5dB difference in the direct sound as if it were a point-source in the free field.

4.3 Main Trends

The results are plotted in terms of the independent variables in Figure 8, averaged over participants and trials. The upper plots provide a break down for each source type in both rooms, while the lower plots combine the Drums and Voice results to summarise the overall effect of distance in each room. In all cases, there is a general increase in ASD with increasing distance, with few exceptions and downward curvature. Drums tended to be perceived 3-4 points closer than Voice, which may be attributed to prior expectations of the loudness and source directivity effects.

In both rooms, there is an initial steep rise and then a tendency to saturate, levelling off the scores albeit at different distances. For Octagon Hall, there is a peak at 6m and an unexpected dip at 8m. For Classroom, there is a peak at 2.5m and an unexpected dip at 5m, and also a higher score at the closest 1.0m distance. This discrepancy may be due to the virtual distance falling inside the loudspeaker radius, creating incongruence with what participants can see in the test environment.

The results of paired t-tests are shown in Figure 9, which indicate significant differences in all but 6 pairs for Classroom and 4 pairs for Octagon Hall. Hence, it is possible to identify 5 or 6 perceptually distinct distances that are monotonically increasing with the physical distance. These are indicated with the filled markers in the lower plots in Figure 8.

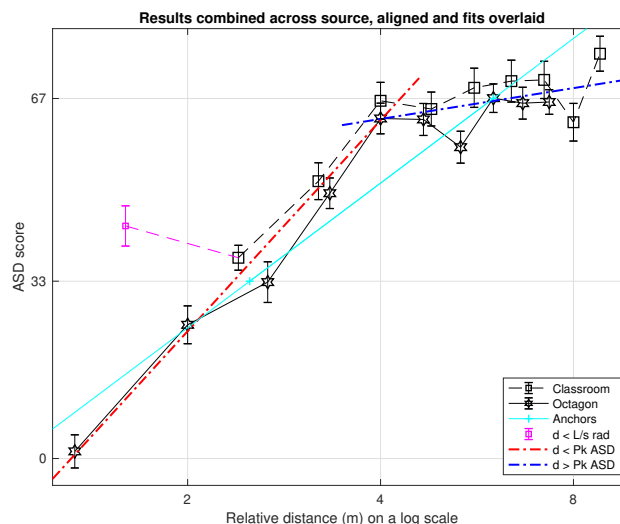


Figure 11. ASD ratings for both test rooms versus log-distance relative to the mid-range ASD peak in each room. Error bars indicate 95% confidence intervals of ± 2 standard errors. The point at a distance inside the loudspeakers (i.e., at 1.0m) in the reproduction environment is marked as a magenta box with a solid line. The reference anchors from the Pori Hall are shown as cyan crosses (+) lying on the log-distance function. Linear fits to the ratings below (red) and above (blue) the ASD peak point are drawn as dash-dot lines.

4.4 Distance Trends

The main purpose of this study was to investigate the ability of the RSAO to cue different settings of apparent source distance from the listener through the object-based rendering and sound reproduction. Thus, it is interesting to see how distances are perceived across the different ranges that were employed for the two test rooms, and to ask whether there is any correspondence to the distance ratings across rooms, by taking advantage of the fixed distances of reference anchors on the rating scale. Figure 10 shows both sets of results plotted on the same axes. In this case, the physical distance is plotted on a logarithmic scale, alongside the line through the reference anchors. With some assumptions of uniform reverberation and point-source-like attenuation of the direct sound over distance with spherical spreading, the straight line through the anchors show the expected trend if ASD were purely explained in terms of the Direct-to-Reverberant Ratio (DRR). In fact, disregarding the point at 1.0m distance which was inside the reproduction space, the main trend is in broad agreement over two and a half octaves, over the range from 1.5m to 10m. However the pattern of deviations from the main trend line appears to scale with the room. So it may be informative to align them for closer inspection.

4.5 Aligned Relative Log-Distance Trends

The results are redrawn in Figure 10 to show the similarity in the pattern between the two rooms. By aligning the two sets of data at their respective mid-range peaks in the

ASD ratings, at 2.5m and 6.0m, the common trends appear more starkly. The steeper initial rise is modelled well by the linear fit (red) for the values up to and including the peak point, disregarding the anomaly at the closest point coloured magenta. Similarly, above the peak, the ASD rating exhibit a saturation behaviour represented by a linear fit (blue) with a much gentler gradient on the log-distance plot. In relation to the anchor line, close sources appear closer, mid-range ones appear farther away, and beyond that the space seems compressed. It is unclear from the present experiments what properties determine the position of the mid-range ASD peak. While the two selected rooms differ in size, neither the ratio of their volumes nor their critical distances matches the ratio of the peak distances. One possible explanation could be found in the ratio of the room lengths, since the Octagon Hall is unusually high. Another explanation could arise from the experimental method and the specific choice of anchor distances. Further listening tests are therefore required in order to disentangle these potential explanations.

5. CONCLUSIONS

The utility of the RSAO for cueing apparent source distance has been investigated objectively and subjectively. Visualisation of the encoded RSAO parameters confirmed the consistency across different octave bands and trends over distance. Formal listening tests showed that these RSAO parameters successfully conveyed effects for distance. The influence of source type was small and did not show a consistent effect, but the interaction of room with distance was significant. Particularly, the perceptual saturation point was higher for the larger room as compared to smaller room.

Overall, these results indicate that the RSAO can be used to cue source distance with resolution comparable to natural listening indoors. Further work will investigate how distance perception performs across reproduction setups and develop methods to continuously vary source distance by adapting the RSAO parameters.

6. ACKNOWLEDGEMENTS

This research was supported in the UK by EPSRC Programme Grant *S3A: Future Spatial Audio for an Immersive Listener Experience at Home* (EP/L000539/1). The authors thank the listening test participants and members of the S3A team who helped in providing software tools, comments on the work and assistance with the set up.

7. REFERENCES

- [1] K. F. Fluitt, T. Mermagen, and T. Letowski, "Auditory perception in open field: Distance estimation," tech. rep., Army Research Lab Aberdeen Proving Ground MD Human Research and engineering, 2013.
- [2] D. H. Mershon and J. N. Bowers, "Absolute and relative cues for the auditory perception of egocentric distance," *Perception*, vol. 8, no. 3, pp. 311–322, 1979.
- [3] D. H. Mershon and L. E. King, "Intensity and reverberation as factors in the auditory perception of egocentric distance," *Perception & Psychophysics*, vol. 18, no. 6, pp. 409–415, 1975.
- [4] A. Kuusinen and T. Lokki, "On studying auditory distance perception in concert halls with multichannel auralizations," in *Proc. of the Int. Conf. on Digital Audio Effects (DAFx-15), Trondheim, Norway*, 2015.
- [5] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst, "Auditory distance perception in humans: A summary of past and present research," *ACTA Acustica united with Acustica*, vol. 91, no. 3, pp. 409–420, 2005.
- [6] L. Remaggi, P. Jackson, and P. Coleman, "Estimation of room reflection parameters for a reverberant spatial audio object," in *Audio Engineering Society Convention 138, Warsaw*, 2015.
- [7] ITU-R BS.2076-2, "Audio Definition Model," 2019. <https://www.itu.int/rec/R-REC-BS.2076-2-201910-I>.
- [8] P. Coleman, A. Franck, P. Jackson, R. J. Hughes, L. Remaggi, F. Melchior, *et al.*, "Object-based reverberation for spatial audio," *Journal of the Audio Engineering Society*, vol. 65, no. 1/2, pp. 66–77, 2017.
- [9] P. Coleman, A. Franck, D. Menzies, and P. J. B. Jackson, "Object-based reverberation encoding from first-order ambisonic RIRs," in *Audio Engineering Society Convention 142, Berlin*, May 2017.
- [10] R. Stewart and M. Sandler, "Database of omnidirectional and b-format room impulse responses," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 165–168, IEEE, 2010.
- [11] L. Remaggi, P. Jackson, P. Coleman, and J. Francombe, "Visualization of compact microphone array room impulse responses," in *Audio Engineering Society Convention 139, New York*, November 2015. eBrief-218.
- [12] S3A project, "VISR production suite," June 2019. <https://www.s3a-spatialaudio.org/plugins>.
- [13] J. Merimaa, T. Peltonen, and T. Lokki, "Concert hall impulse responses pori, finland: Reference," *Tech. Rep.*, 2005.
- [14] J. W. Philbeck and D. H. Mershon, "Knowledge about typical source output influences perceived auditory distance," *The Journal of the Acoustical Society of America*, vol. 111, no. 5, pp. 1980–1983, 2002.
- [15] Bang & Olufsen, "Music for Archimedes," *Compact disc CD B&O 101*, 1992.
- [16] K. Hamasaki, "22.2 multichannel audio format standardization activity," *Broadcast Technology*, vol. 45, pp. 14–19, 2011.