User Expectation of Room Acoustic Parameters in Virtual Reality Environments

Benjamin Burnett
Institute of Sound Recording
University of Surrey
Guildford, UK
benb2514@outlook.com

Annika Neidhardt
Institute of Media Technology
Technische Universität Ilmenau
Ilmenau, Germany
annika.neidhardt@tu-ilmenau.de

Zoran Cvetković
Department of Engineering
King's College London
London, UK
zoran.cvetkovic@kcl.ac.uk

Hüseyin Hacıhabiboğlu
Graduate School of Informatics
METU
Ankara, Turkiye
hhuseyin@metu.edu.tr

Enzo De Sena
Institute of Sound Recording
University of Surrey
Guildford, UK
e.desena@surrey.ac.uk*

***Abstract*—This paper explores how visual attributes of a VR scene affect user expectations of room reverberation. A psychoacoustic experiment was run wherein subjects wore a VR headset and adjusted two unlabelled sliders controlling the reverberation time (T60) and the acoustic room size until the reverberant response was closest to their expectation of how the room they were seeing should sound. Different visual characteristics, in particular, room type and size, surface material, and furnishing were modified to determine how these might affect their expectations of the reverberant response. Results showed that visual room size had a significant effect on both the expected T60, in agreement with previous literature, and on the expected acoustic room size. Both relations seem to be well-described by a simple sublinear power law model, which could be used, for instance, to design reverberation time (T60) and acoustic room size values that align well with listeners' expectation for a given visual room volume. Differences in visual surface materials were found to have a statistically significant effect on the expected T60. The level of visual furnishing, on the other hand, only had a marginally significant effect on the expected T60. The results also indicate considerable subjective differences in individual expectations.**

***Index Terms*—virtual reality, room acoustic modelling, artificial reverberation, perceptual evaluation, reverberation time, room size**

I. Introduction

The use of audio in virtual reality (VR) applications is shown to increase the user's sense of presence, a desirable quality defined as a sense of 'being there' [1], [2]. Moreover, a higher sense of presence is achieved when the auditory and visual domains are perceptually matched [3], [4]. To this end, perceptual acoustic matching, in terms of audio design congruent with the visual environment, is an important aspect of sound design for VR applications. Consistency between the senses also contributes to the scene's plausibility, a quality whereby a stimulus is in agreement with a user's expectations based on experience of equivalent real-world events [5]. To assess perceptual acoustic matching, one can therefore consider how 'plausible' the sensory experience is. Optimum plausibility is achieved by fulfilling the listener's expectations, even though such expectations might not accurately reflect reality. Previous work has shown that a larger visual room size leads to the expectation of a longer reverberation time (T60) and vice versa [6], but it remains unknown how other visual cues may influence the successful perceptual matching of artificial reverberation.

This work considers two other visual cues within VR that are hypothesised to influence users' expectation of reverberation: surface materials and the degree of furnishing within the space. As it is also not known whether and how reverberation qualities other than T60 require perceptual matching in VR, this work further investigates the relevance of 'acoustic size' as a parameter in the task of perceptual acoustic matching, where *acoustic size* is defined as the size of the modelled space. To that end, this study employs a scattering delay network (SDN) [7] room acoustics simulator.

SDN is chosen for this purpose as it is a real-time acoustic modelling technique that produces spatio-temporally accurate early reflections, which are known to be relevant for the perception of room size [8], [9], whilst its late reverberation provides a close approximation of the image method [10] according to high level perceptual attributes. This allowed for real-time rendering of audio within an accurate room model, as is required in a six-degrees-of-freedom (6DOF) VR environment. SDN has been shown to result in (a) a higher "naturalness" than feedback delay networks (FDNs) [11], convolution methods and ray-tracing [12], (b) a greater sense of externalisation than higher-order ambisonics measurements [13], and (c) a greater sense of immersion in VR environments [14]. Furthermore, it was successfully incorporated in various real-time binaural rendering applications [13], [15], [16], and it has since been extended to (a) model outdoor scenes [17], (b) coupled enclosures [18], and (c) exact modelling of higher-order reflections [19]. Although the results presented in this paper may only be applicable to SDN modelling, it is possible that the findings extend to other reverberation techniques.

The current exploratory study aims to identify how the expectation of the two reverberation parameters, T60 and acoustic room size, changes with the visual room size, surface materials, and room furnishing. These variables are not an exhaustive list, but investigation into these aims to provide insight into how the visual room impression modulates user expectations towards reverberation.

The paper is organised as follows. Section II reviews the high-level physical features of reverberation and their perception, also considering visual cues, and provides a brief overview of the SDN room simulator. Section III details the perceptual experiment, and section IV presents the results. Section V discusses these results and Section VI concludes the paper.

II. Background

To assess how visual cues might affect the expectation of reverberation in VR, it is important to understand principles of room acoustics, in particular, how physical attributes of a room affect its reverberation. Should these physical attributes be in some way visible to the listener, then they have a potential to modulate the user's expectation of reverberation. Hence, the first subsection reviews relevant basic principles of room acoustics, in particular, the relationship between physical attributes of a space and high level attributes of its reverberation. Furthermore, towards investigating expectations of the listener regarding reverberation in VR, one needs to understand the human perception of room reverberation; relevant principles are therefore outlined in the second subsection. Thirdly, room perception is said to be a function of both the auditory and visual senses, and as such the influence of visual and audio-visual cues on spatial perception are discussed in the third subsection. Finally, it is necessary to understand the principles of SDN reverberation in order to interpret the results of the perceptual experiment conducted in this study. An overview of SDN is thus given in the fourth subsection.

A. Room properties that affect reverberation

The primary physical attributes contributing to a room's reverberation are its size, geometry, and surface absorption. According to Norris-Eyring's formula, the reverberation time (T60), is a function of volume V, surface area S, and average absorption α [20]:

$$T60 = \frac{0.161V}{-S \times \ln(1-\alpha)} \ . \tag{1}$$

It can be observed that T60 increases linearly with room size (volume), and decreases inversely to its surface area. T60 also decreases with an increase in absorption coefficients [20].

Room geometry determines the profile of early reflections, which are relevant for the perception of room size [8], [9], as well as the spectral character of reverberation, as determined by modal density, *i.e.* the average number of resonant frequencies per Hz [21] [7] [22].

Another significant factor contributing to reverberation is furnishing, as an empty room sounds very different to a furnished room. The presence of objects like beds, curtains, or shelves may absorb, diffuse, or reflect sound, and hence affect parameters such as T60 [23]. Burgess and Utley [23] used absorption coefficients of light and heavy furnishings to estimate the difference in reverberation time for a $39\,\mathrm{m}^3$ living room, finding a difference of $0.09\,\mathrm{s}$ ($0.35\,\mathrm{s}$ for the lightly furnished room, and $0.26\,\mathrm{s}$ for the heavily furnished room). Furthermore, in [24] it is shown that the arrangement of furniture also has an effect on reverberation time.

Other factors that affect reverberation include for instance temperature [25] and the rigidity of the walls [10], however these particular factors are ordinarily not visible and so not relevant when investigating visual VR cues.

From visible room properties that are known to affect reverberation, this study investigates effects of: (a) room size, (b) surface materials, and (c) furnishing on the expectation of reverberation properties.

B. Perception of reverberation

There are a number of established objective metrics of reverberation properties, such as reverberation time (T60), early decay time, sound strength, early energy fraction, late lateral sound level, interaural cross-correlation (IACC), and clarity index [26]. The relationship between these objective metrics and perception of room reverberation is not well understood, but T60 is considered the most perceptually significant room acoustics attribute [26]. In fact, a study by Zahorik [27] suggests that virtual room acoustics is perceived along only two perceptual axes relating to T60 and IACC.

Of key consideration when assessing reverberation is the aspect of room size perception, which reverberation cues are understood to contribute to [28]. T60 has been shown to strongly affect room size perception, as does source-receiver distance, more so than the physical volume of a modeled or measured room [29]. Room size perception has been suggested to be the result of relations between direct sound, and early and late reflections and accordingly, using only reverberation time to intentionally control the auditory impression of room size would not be sufficient, as this only affects the late reverberant energy [30].

Colouration is another perceptual feature of reverberation, which arises as a result of modal distribution, particularly in the case of artificial reverberators where reflection density is less than in natural rooms [31].

This study therefore investigates the perceptual expectation of T60 and acoustic room size. The latter was chosen for its effect on early reflection profile, and therefore also direct/early/late energy ratios, as well as for affecting the spectral character of the reverberation.

C. Visual and audiovisual room perception

In the visual perception of a room environment, horizontal area, height, and colour are known to have an impact on the perceived spaciousness [32]. Furthermore, shaping room geometry towards a perfect cuboid while keeping the volume constant, increases the visually perceived room size [33].

When asked to adjust the size of a virtual room to match it to the visual size of a previously seen virtual room, users were very accurate in estimating the relative room size [34]. Therefore, the method of adjusting the relative size of the room is an appropriate method to study perceived room size.

Spatial perception is thought to be dependent on both auditory and visual domains, and as such the perception of reverberation may be affected by visual information. It was shown to that an image of a reverberant space provides enough information to synthesise the acoustic character of a given audio input successfully [35]. When participants had to adjust the reverberation based on a 2D room image, the result depended heavily on the visual room impression [36]. However, the adjusted reverberation times often differed from the actual ones. Moreover, in an evaluation of audiovisual perception of the source distance and the size of six performance rooms, visual cues were dominant over auditory cues [37]. In contrast, Schutte et al. [38] reported that the visual room impression did not have a significant impact on perceived reverberation, when participants rated it on a scale from 1 to 10.

Gil-Carvajal et al. [39] reported that when listening to binaural recordings in real rooms, the auditory image was not affected in its perceived distance, azimuthal direction or compactness, when the listening room differed visually from the recorded room. On the contrary, deviating acoustic properties had an effect. For example, the perceived distance decreased, when the listening room was more reverberant than the recorded room.

Hence, visual room impression seems to play a minor role with regard to the expected reverberation, if other room cues are available in addition. In VR, this multi-modal mix will be created from scratch. Therefore, it is important to study the role of the visual room impression on reverberation expectation in VR environments.

D. Scattering Delay Networks

The SDN reverberator [7] models an enclosure using a minimal topology of scattering nodes connected via bi-directional delay lines. One node per boundary is placed at the exact location of first-order reflections for a given source and listener position, such that first-order reflections are rendered accurately. Subsequent reflections are calculated using the same network of delay lines, which results in an approximation that becomes progressively coarser with reflection order. Scattering nodes are each characterised by a reflectiveness coefficient, which determines the ratio of energy reflected at that node. SDN modelling has been shown to provide a reverberation quality close to that of the image method (IM), but with a significantly lower computational load. This allows for real-time rendering in a 6DOF environment without significantly compromising the reverberation quality.

III. Experiment

The aim of the experiment was to identify how the visual appearance of room size, room furnishing, and surface materials in a VR room contributes to the expectation of two acoustic reverberation parameters: T60 and ‘acoustic size’ of the room, that is the geometric size of the SDN room model.

A. Apparatus

The stereoscopic headset used was the Oculus Quest 2, locked to 90 Hz refresh rate in order to reduce the likelihood of motion sickness [40]. No drops in frame rate were reported and no participant reported any motion sickness. Unity version 2021.3.16f1 was running on a Windows 10 PC with a GTX960 graphics card.

The audio material consisted of speech [41] and percussive [42] audio samples rendered binaurally in six degrees-of-freedom using the MIT Kemar head-related transfer function (HRTF) and SDN. A Unity implementation of the SDN reverberator was used that allowed real-time adjustment of the reverberation parameters. The audio was played back through a pair of Beyerdynamic DT770 pro closed-back headphones, which were set to the same listening level for each subject.

B. Stimuli generation

The study included a total of twenty-six unique stimuli, which varied in room type, size, furnishing, surface materials and audio sample. Original scenes were sourced from the Unity asset store [43]–[47], but some were altered slightly to make the scale of objects relative to the user realistic. Any intrusive furniture items were removed to avoid expectation of significant occlusion. Nine room sizes were used, varying from approximately $3\,\mathrm{m} \times 4.5\,\mathrm{m} \times 2.2\,\mathrm{m}$ (length $\times$ width $\times$ height) to $16\,\mathrm{m} \times 12.8\,\mathrm{m} \times 3.2\,\mathrm{m}$.

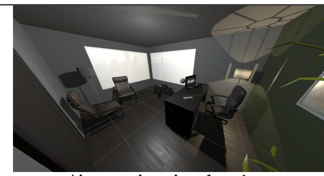
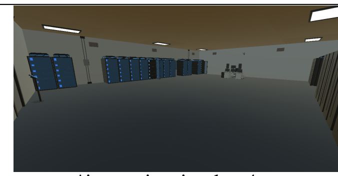
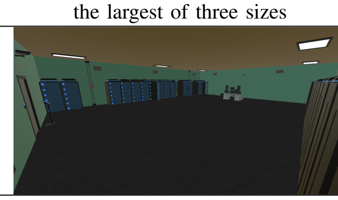
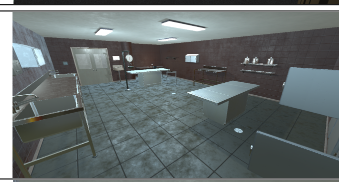
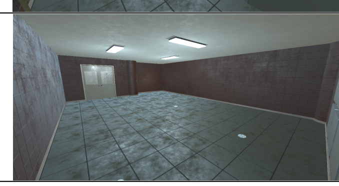
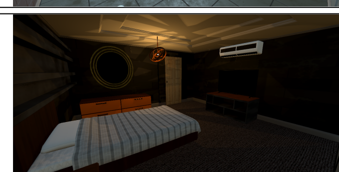
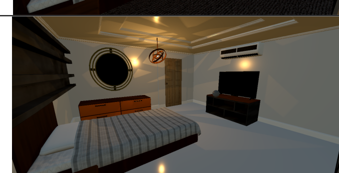
Two of the room types, office and server, were presented in three sizes, while two other rooms, morgue and bedroom, were presented with and without furniture. Additionally, two rooms, server and bedroom, were presented with two different surface materials. The alternative surface materials of the server scene were intended to be of a higher average absorption coefficient (carpet instead of ambiguous flat texture), whereas the alternative materials for the bedroom were intended to be less absorbent (smooth, slightly reflective surfaces as opposed to the original carpet/panelled walls). Each of the 13 unique visual combinations was played twice – once for each of the two audio samples, with the exception of the garage scene which was played three times for each audio sample to check for intra-subject consistency. The stimuli are summarised in Table I.

C. Methodology

The user interface (UI) shown in Fig. 1 was presented to subjects within the Unity environment in the VR headset. Subjects were instructed to adjust two unlabelled sliders until the reverberant response was closest to their expectation of how the room they were seeing should sound. The sliders controlled the reverberation time (T60) and the geometric size of the SDN

TABLE I

VARIABLE CONFIGURATIONS WITHIN THE 13 UNIQUE SCENES. BOLD FONT INDICATES A VARIATION ON THE STANDARD SCENE PRESENTATION, WHERE THE STANDARD VARIATION IS THE FURNISHED ROOM WITH ORIGINAL MATERIALS AND ORIGINAL SIZE (SCENES 1, 4, 8, 10, AND 13)

Stimulus	Type (repeats per audio item)	Dimensions $L \times W \times H$ [m] (Volume [m^3])	Materials	Furniture	Image
1	Office (1)	$3.1 \times 4.52 \times 2.2$ (30.83)	Wooden floorboards/ green & grey plasterboard-like walls	Yes	 *image is stimulus 1 — the smallest of three sizes
2	Office (1)	$\mathbf{4.65 \times 6.78 \times 2.2}$ **(69.36)**	Wooden floorboards/ green & grey plasterboard-like walls	Yes	
3	Office (1)	$\mathbf{6.2 \times 9.04 \times 2.2}$ **(123.31)**	Wooden floorboards/ green & grey plasterboard-like walls	Yes	
4	Server (1)	$16 \times 12.8 \times 3.2$ (655.36)	Floor/walls, different shades of grey — both flat/ambiguous material	Yes	 *image is stimulus 4 — the largest of three sizes
5	Server (1)	$\mathbf{10.67 \times 8.53 \times 3.2}$ **(291.27)**	Floor/walls, different shades of grey — both flat/ambiguous material	Yes	
6	Server (1)	$\mathbf{8 \times 6.4 \times 3.2}$ **(163.84)**	Floor/walls, different shades of grey — both flat/ambiguous material	Yes	
7	Server (1)	$16 \times 12.8 \times 3.2$ (655.36)	**Carpeted floor/ green walls**	Yes	
8	Morgue (1)	$9.62 \times 6.5 \times 2.5$ (156.33)	Large-tiled floor/ small-tiled walls	Yes	
9	Morgue (1)	$9.62 \times 6.5 \times 2.5$ (156.33)	Large-tiled floor/ small-tiled walls	**No**	
10	Bedroom (1)	$4.6 \times 4.6 \times 2.4$ (50.78)	Carpeted floor/ dark wood-panelled walls	Yes	
11	Bedroom (1)	$4.6 \times 4.6 \times 2.4$ (50.78)	**Shiny, off-white floor/ slightly shiny, light-grey walls**	Yes	
12	Bedroom (1)	$4.6 \times 4.6 \times 2.4$ (50.78)	Carpeted floor/ dark wood-panelled walls	**No**	
13	Garage (3)	$6.33 \times 6 \times 2.4$ (91.15)	Hard, stone-textured floor/ 3× pin-board walls & 1× corrugated metal garage-door	Yes	

reverb model over a continuum in a preset range. Subjects were also asked to rate the 'plausibility' of their answer. This was meant to elicit the degree to which the subjects felt confident that their final choice for the two unlabelled sliders resulted in a plausible acoustic response. However the associated results were not particularly informative and are not reported here.

The two unlabelled sliders were designed as follows. Due to the increase in acoustical mean-free-path, an increase in the room size will produce a greater reverberation time (see Equation (1)). In a preliminary experiment, the two sliders were designed to control the wall absorption coefficients and the acoustic room size, however subjects found it difficult to complete the task, possibly due to the fact that both sliders had an effect on T60. The experiment was therefore revised such that the two unlabelled sliders controlled (a) the reverberation time T60, and (b) the acoustic room size, except now the latter also appropriately adjusted the absorption coefficient α according to Equation (1) such that T60 is kept constant (*i.e.* increasing room size would be accompanied by increasing absorption, such that the T60 stays approximately constant, and vice-versa).

Slider ranges and linearity were calibrated empirically, and initial slider positions were always randomised upon loading a scene. The T60 slider ultimately ranged linearly from $0.05\,\mathrm{s}$ to $1.4\,\mathrm{s}$. The size slider ranged linearly from 0 to 3, and its value x was used to scale the length (L) and width (W) of the considered acoustic room model according to $L \times 2^x$ and $W \times 2^x$. L and W were initialised at a minimum size for each room; the ratio between L and W was thus kept constant and the smaller of the two dimensions was set to $2\,\mathrm{m}$.

Room model height was always kept fixed to the visual height and never scaled. This was decided after preliminary testing that revealed that distances to the floor or ceiling that were particularly large or small led to improper sounding reverberation. It is thought that only scaling length and width was acceptable for the current study as room size will typically vary less in height than in width/length. Adjustments of acoustic room height are left for future research.

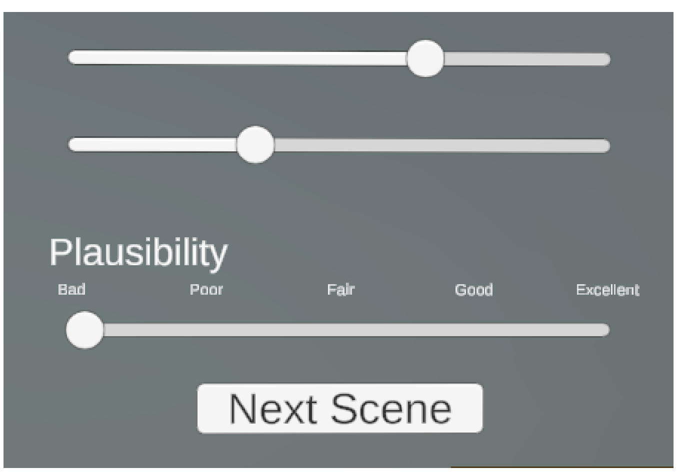


Fig. 1. The user interface presented to subjects within the VR environment, with the two unlabelled sliders controlling the acoustic parameters (T60 and room size), along with the plausibility slider.

This process was repeated for the 30 scenes described in section III-B, and were presented in a randomised order for each participant. Prior to this was a familiarisation stage, in which participants attempted the task once with only the reverb sliders, and then once for the default versions of all five scene types.

D. Subjects

Eighteen trained participants took part in the experiment, all of whom undergraduate students (ages between 18-25) on an audio engineering course. Subjects were not compensated for their time but were given the opportunity to compete in a short VR game with a small prize for the winner.

IV. Results

The data associated to each attribute consisted of 30 sets, *i.e.* 15 scenes (the 13 scenes summarised in Table I, in addition to two repeats for the garage scene) repeated once for each audio sample.

In order to test for intra-subject consistency, the standard deviation of each subject was calculated for the three data points in the garage scene. The results of which are shown in Fig. 2 for expected reverberation time (T60). Although it can be observed that some subjects had a larger deviation than others, it was decided that there was not sufficient evidence to remove any one of them.

The three sets for the garage scene were then combined into a single one, giving a total of 26 sets per attribute. Table II presents the mean expected T60 and acoustic size responses for each stimulus.

Normality tests (Kolmogorov-Smirnov test with Lilliefors significance correction, and Shapiro-Wilk test) were then run for each of the 26 sets. For the expected T60, two of which failed at the 0.05 significance level. It was still decided to use ANOVA tests, on the basis that they are robust to moderate deviation from normal data [48].

A one-way ANOVA test ($R^2 = 0.472$) run on the expected T60 revealed a statistically significant effect of the visual scene (run across the combined 13 scenes; $p < 0.001; F = 37.110$) and audio sample ($p = 0.002; F = 9.356$), but no interaction between visual scene and audio sample ($p = 0.992; F = 0.282$). Since a statistically significant difference was observed between speech and percussive samples, their corresponding results are presented separately henceforth.

In the case of acoustic room size perception responses, as many as 22 sets out of 26 failed the normality test at the 0.05 significance level, indicating a significant deviation from normal data. For this reason, a non-parametric test was preferred. An independent-sample Kruskal-Wallis test revealed a statistically significant effect of the visual scene (also in this case across the combined 13 scenes; $p < 0.001; \chi^2(12) = 188.530$). No statistically significant difference was observed between the audio samples ($p = 0.847$).

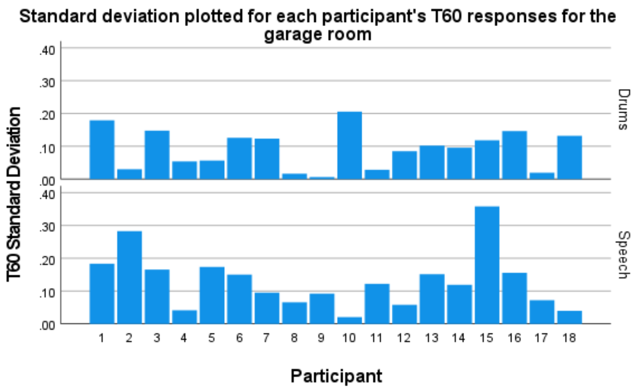


Fig. 2. Bar chart showing Standard deviation of garage scene T60 response for each participant.

TABLE II
AVERAGE T60 AND ACOUSTIC SIZE RESPONSES FOR EACH STIMULUS

Scene (description)	Visual volume	T60 (mean)		Acoustic size (mean)	
		Drums	Speech	Drums	Speech
(1) office (size1)	30.83	0.31	0.25	48.3	58.6
(2) office (size2)	69.36	0.42	0.32	67.4	81.9
(3) office (size3)	123.31	0.46	0.40	84.2	117.7
(4) server (size1)	655.36	0.70	0.70	188.6	239.3
(5) server (size2)	291.27	0.65	0.58	206.4	152.2
(6) server (size3)	163.84	0.61	0.52	134.5	195.3
(7) server (size1 / alternative materials)	655.36	0.58	0.50	295.4	216.1
(8) morgue (furnished)	156.33	0.74	0.66	125.4	129.4
(9) morgue (unfurnished)	156.33	0.78	0.76	110.1	157.3
(10) bedroom (furnished)	50.78	0.25	0.23	94.3	83.8
(11) bedroom (furnished / alternative materials)	50.78	0.33	0.30	69.7	49.1
(12) bedroom (unfurnished)	50.78	0.29	0.28	70.7	62.7
(13) garage (furnished)	91.15	0.63	0.57	55.4	58.3

A. Effect of visual size on expected reverberation time and acoustic size

Two rooms, office and server, were presented in three size variations. The hypothesis was that with visually larger scenes, subjects expect a greater T60 and acoustic size.

Analysis of the office-data using a linear mixed-effects model identifies a statistically significant dependency of the expected T60 on the visual room size with a slope of 0.0016 (95% CI = [0.0012 0.0021], $p < 0.0001$), an average intercept of 0.275 ([0.215 0.335], $p < 0.0001$) and an additional fixed-effect offset of -0.071 ([-0.104 -0.037], $p < 0.0001$) for speech compared to drums. The slope was not affected by the signal ($p = 0.97$). The variation over subjects could be modeled as a random intercept effect with a standard deviation of 0.095 s ([0.065 0.137]) from the estimated average intercept. Considering this random effect increases R^2 from 0.244 to 0.664. Thus, the estimated linear mixed effects model explains about 66% of the changes observed for expected T60.

This means that in the office, listeners expect a reverberance that corresponds to a reverberation time of about 0.27 s ($\pm 0.055\,s$) for very small office versions with a standard deviation of 0.095 s ($\pm 0.035\,s$) between listeners, and it increases by 0.16 s ($\pm 0.065\,s$) per $100\,\text{m}^3$ volume increase with the drums sample. With speech, the expected T60 is about 0.07 s ($\pm 0.03\,s$) lower. Since the visual height remained the same when creating the alternative sizes of the room, the model is equivalently valid for the dependency on visual floor area, if the visual volume is divided by room height (2.2 m). Consequently, expected T60 increases by 0.036 s ($\pm 0.014\,s$) per additional $10\,\text{m}^2$ of floor area.

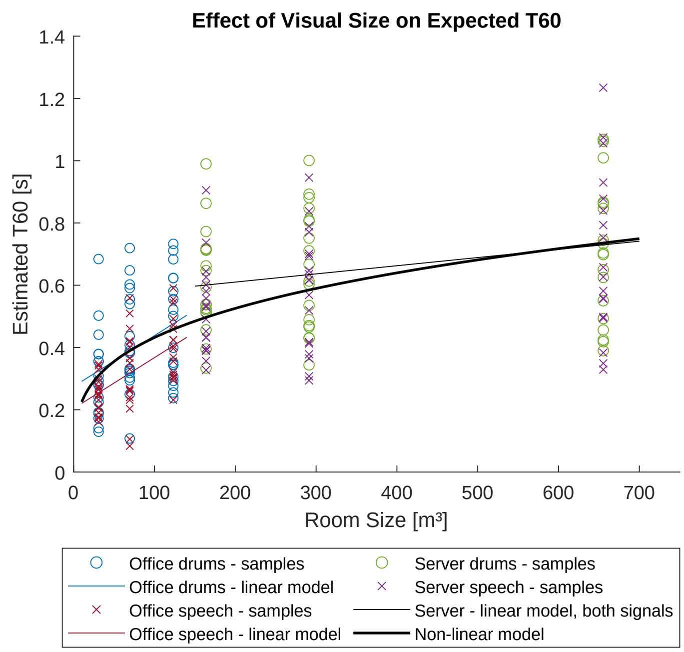


Fig. 3. Linear mixed effects model for the relation between visual room volume and expected T60 in the office and the server room. In addition, the black curve visualizes the fitted perceptual model $T60_{exp} = a\,(V_{visual})^b$ with $a = 0.117 \pm 0.032$, $b = 0.283 \pm 0.050$, and $R^2 = 0.40$.

A separate linear mixed effects model was estimated for the server room-data. It revealed a dependency with an intercept of 0.558 ([0.470 0.644], $p < 0.0001$) with a standard deviation of 0.116 ([0.076 0.178]) for the individual subjects and a slope of 0.00026 ([0.00012 0.00041], $p < 0.0001$). Here, the type of signal did not have a significant effect on the intercept ($p = 0.070$) or the slope ($p = 0.176$). With an R^2 of 0.399, the power of the model, shown in Fig. 3. is considerably lower than for the office. Furthermore, the slope does not even reach one tenth of that estimated for the office. Expected T60 increases by 0.026 s ($\pm 0.015\,s$) with every additional $100\,\mathrm{m}^3$ or, because of the constant room height of 3.2 m, 0.0083 s ($\pm 0.0047\,s$) with every additional $10\,\mathrm{m}^2$ of floor area.

Further two linear mixed-effects models analysed the dependency of expected acoustic room size on visual room size in both rooms. For the office, the model achieved an R^2 of 0.404. It exhibits an intercept of 26.7 ([4.63 48.7], $p = 0.018$) with a standard deviation of 25.11 ([15.62 40.35]) over the subjects and slope of 0.46 ([0.26 0.67], $p < 0.0001$). This means that the expected size of the SDN room model increases with about half the increase of the visual room size. Since visual and acoustic room height were equal, the same relation can be found for the acoustic and the visual floor area.

For the server room, expected acoustic size rises even slower with increasing visual size. The intercept is estimated with 115.84 ([68.70 162.97], $p < 0.0001$) and the slope with 0.10 ([0.027 0.180], $p = 0.008$). In this case, R^2 is 0.384. The signal did not significantly influence expected acoustic room size. Fig. 4 visualises the results.

The three sizes of the server room considered in the experiment were larger than the three office-versions. It is open whether the difference in slope is due to different size or due to other room properties. If the slope decreases with increasing room size, a non-linear compressing model could be interesting.

Therefore, models inspired by Stevens' power law were estimated. The results are represented by the black curve in both figures, and correspond to the following equations, which were obtained using the Matlab Curve Fitting Toolbox:

$$T60_{exp} = 0.114\,(V_{visual})^{0.288} \quad , \tag{2}$$

$$V_{acoustic} = 10.04\,(V_{visual})^{0.475} \quad , \tag{3}$$

where V_{visual} is the visual volume of the scene (measured in m^3), $T60_{exp}$ is the listener expectation of T60 (measured in s), and $V_{acoustic}$ is the expected acoustic room size (measured in m^3).

B. Effect of surface material appearance on expected reverberation time

Two rooms, the bedroom and server room, were presented with two different sets of surface materials, one with visually more reflective materials and the other with visually more absorbent materials (see Table I). The results of this test are shown in Fig. 5.

The combination of surface materials was not the same between the two rooms, so two ANOVA tests were run, one for each room. For the server room, an ANOVA test using audio sample and surface material as factors, revealed that the surface material has a statistically significant effect on the expected T60 ($p = 0.004$; $F = 8.705$; $R^2 = 0.126$)[1]. The same test for the bedroom scene was also significant ($p = 0.002$; $F = 10.156$; $R^2 = 0.146$)[1].

[1]No posthoc test was necessary, since there are only two categories and the pair-wise test returns the same significance value as the omnibus test.

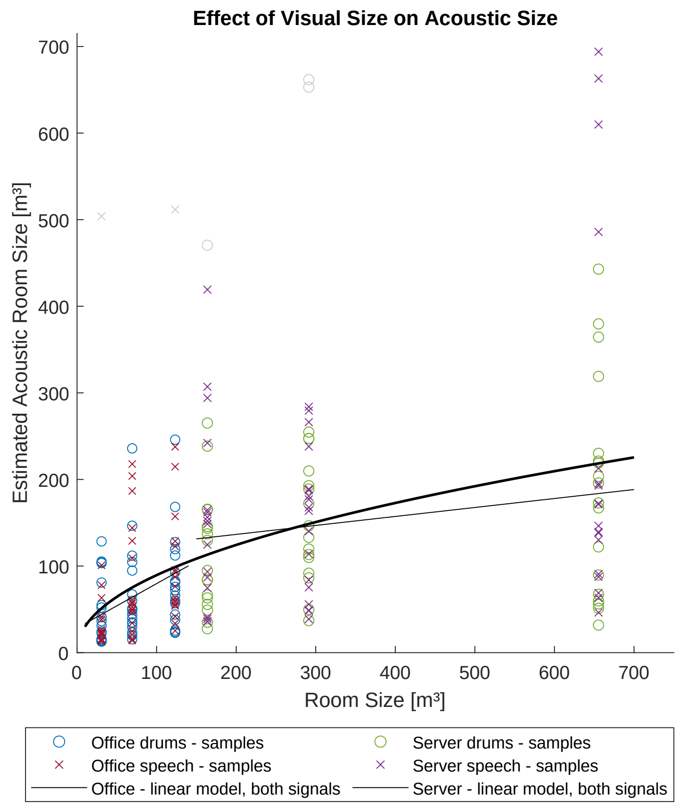


Fig. 4. Linear mixed effects model for the relation between visual room volume and expected acoustic size for the office and the server room. In addition, the black curve visualizes the fitted perceptual model $V_{acoustic} = a\,(V_{visual})^b$ with $a = 10.04 \pm 7.03$, $b = 0.475 \pm 0.12$, and $R^2 = 0.266$. Points in grey indicate outliers which were removed prior to calculation of the models.

C. Effect of scene furnishing on expected reverberation time

Two rooms, morgue and bedroom, were presented with and without furniture. The results of this test are shown in Fig. 6. An ANOVA test restricted to these two rooms was run using room type, audio sample, and furnishing type as factors, and revealed that furnishing had a marginally significant effect on expected T60 ($p = 0.051$; $F = 3.880$; $R^2 = 0.655$)[1]. More specifically, the expected T60 increased from 0.238 ± 0.037 s for the furnished bedroom to 0.285 ± 0.043 s for the unfurnished bedroom and from 0.699 ± 0.063 s for the furnished morgue to 0.770 ± 0.079 s for the unfurnished morgue.

V. Discussion

A. Visual size

The results showed that the expected T60 was significantly affected by the visual size of both the office and the server room. These indications suggest the visual size impression of the VR scene modulates the expectation of T60, supporting previous findings by [6]. However, further work could be done to reveal the extent of this relationship for a greater range of stimuli, particularly in the case of varying audio material.

The expectation of acoustic size was also shown to differ significantly with a change in visual room size. Changing

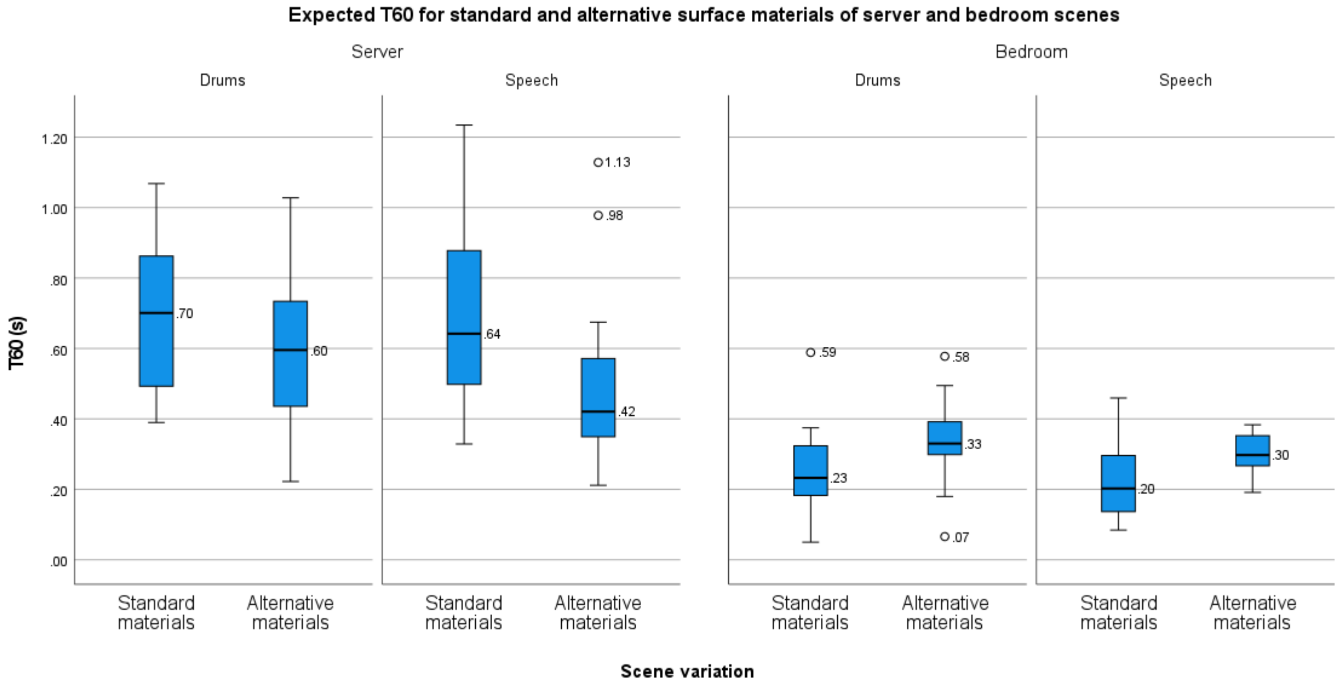


Fig. 5. Boxplots showing T60 expectations for standard and alternative material variations of the server and bedroom scene, with both drums and speech audio samples

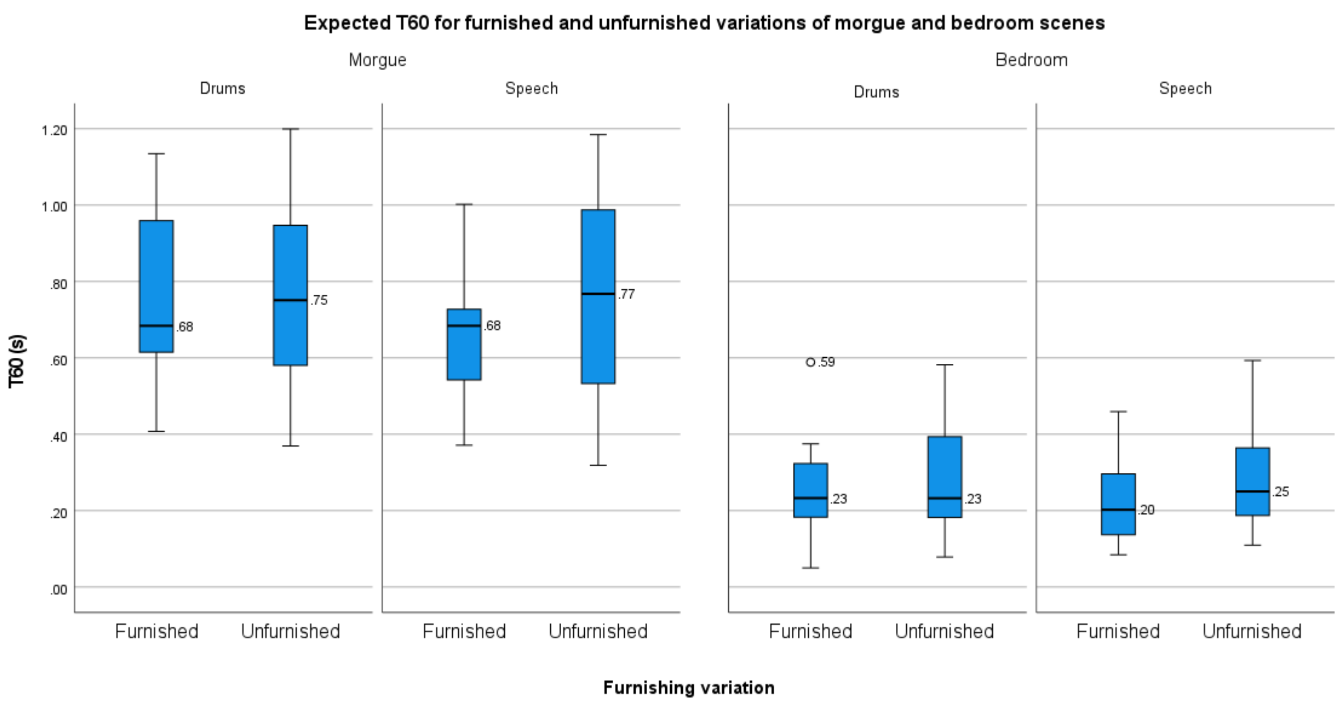


Fig. 6. Boxplots showing T60 expectations for furnished and unfurnished variations of the morgue and bedroom scenes, with both drums and speech audio samples

acoustic room size affects the early reflection profile, as well as the spectral character of the reverberation. So these findings indicate an importance in perceptual matching of these attributes. It is interesting to note that individual responses varied considerably: some subjects expected larger and some smaller than the visual room size. However, for the server room, in particular, the expected size was typically smaller than the visual size. In all cases where significant effects were observed, larger visual room size resulted in greater expected T60 and larger acoustic room size, if measured by the median subject response.

A model inspired by Stevens' power law was fitted to the data of both rooms. This model could be used to design reverberation time and acoustic room size values that align well with listeners' expectations for a given visual room volume. The reader should note, however, that the model was fitted using two different rooms with volume ranges that are not overlapping ($31\,\mathrm{m}^3$, $69\,\mathrm{m}^3$ and $123\,\mathrm{m}^3$ for the office; and $163\,\mathrm{m}^3$, $291\,\mathrm{m}^3$ and $655\,\mathrm{m}^3$ for the server room), and the model fitted the data well in part because the slopes were different for each room (higher slope for the office, and lower slope for the server room). Although the obtained model aligns qualitatively with perceptual findings in adjacent fields, it is possible that the different slopes in the two rooms are caused by characteristics of the room other than their visual volume, for example, their furniture, room shape or surface properties. A larger study is required to examine this hypothesis.

B. Surface materials and furnishing

The visual attribute of surface material had a significant effect on the expectation of T60 in both scenes where different surface materials were presented, *i.e.* server and bedroom. These findings are preliminary as a limited set of conditions were included in the study. Further investigations could include a larger set of test conditions, *i.e.* surface materials, room types, and audio material.

A marginally significant difference was also observed for the expected T60 between furnished and unfurnished rooms. The difference was small in absolute terms, with the expected T60 increasing by $0.05\,\mathrm{s}$ (22% increase) from furnished to unfurnished bedroom, and by $0.07\,\mathrm{s}$ (10% increase) from furnished to unfurnished the morgue. This effect is somewhat smaller than the $0.09\,\mathrm{s}$ (34% increase) difference estimated by Burgess and Utley [23] when measuring the reverberation time in an actual room with a reduced furniture level (from $0.26\,\mathrm{s}$ for the heavily furnished room to $0.35\,\mathrm{s}$ for the lightly furnished room).

VI. Conclusions and Future Work

This study identified several relationships between visual cues and the expectation of reverberation within VR rooms using SDN reverberation. It was found that visual room size had a significant effect on expected T60 and acoustic room size. By taking the considerable subjective differences in individual expectations into account as random effects, the models achieved R2 values of 38 to 66 %. The obtained data suggests that the dependency of T60 and acoustic room size on visual room size can be modeled based on Stevens' power law. However, a larger study is required to confirm this.

It was also observed that variation in surface materials led to a change in the expectation of T60. The fact that these findings depend on the room type and audio material in multiple cases would suggest that further work could examine for the extent of the effect of these factors on reverberation expectation. Changes in the amount of furniture in the visual scene led to a difference in expected T60 too, whereby the expected reverberation time increased when no furniture was present. However, the difference was only marginally significant and corresponded to a small change in absolute terms.

The following steps may be taken to extend this exploratory study. Firstly, a more complete set of reverberation parameters should be identified based on their perceptual relevance. For example, lower-level acoustic parameters such as IACC [27] and clarity index [49], [50] have been suggested to be indicative of perceived room size and may be controlled through higher-level parameters such as diffusion [51] or room shape [52]. The experiment presented in this work used a single, frequency-independent absorption coefficient to affect the room's reverberation. However, to mimic real world conditions more realistically, future work could explore the same effect using non-uniform and frequency-dependent absorption coefficients.

A broader variety of visual attributes could be investigated too, e.g. more types of surface material, or non-binary degrees of furnishing. It may also be beneficial to present a full-factorial experiment whereby all visual factors are presented in all combinations with each other, allowing for statistical analysis that considers all factors when explaining the variance in data, and perhaps also confounding factors such as the type of room, e.g. bathroom vs bedroom.

Finally, future work will involve performing the experiment using environments for which there is a known T60 value, allowing for the comparison between expected and actual T60 values. This could perhaps be achieved using 360-degree camera footage of real spaces and comparisons made with measured room responses.

Acknowledgements

The authors would like to thank all the subjects who participated in this study. This work was supported in part by the Engineering and Physical Sciences Research Council (EPSRC) under the 'SCalable Room Acoustics Modelling' grant EP/V002554/1 and Deutsche Forschungsgemeinschaft (DFG) SPP2236 AUDICTIVE APlausE-MR.

References

[1] P. Larsson, D. Västfjäll, and M. Kleiner, "Ecological acoustics and the multi-modal perception of rooms: real and unreal experiences of auditory-visual virtual environments," in *International Conference on Auditory Display, Espoo, Finland*, 2001.

[2] F. Hruby, "The sound of being there: Audiovisual cartography with immersive virtual environments," *Journal of Cartography and Geographic Information 69, pp. 19–28 (2019)*, 2019.

[3] P. Larsson, D. Västfjäll, P. Olsson, and M. Kleiner, "When what you hear is what you see: Presence and auditory-visual integration in virtual environments," in *Proceedings of the 10th Annual International Workshop on Presence*, 2007.

[4] T. Schubert, F. Friedmann, and H. Regenbrecht, "Embodied Presence in Virtual Environments," in *Visual Representations and Interpretations*, R. Paton and I. Neilson, Eds. Springer London, 1999, pp. 269–278.

[5] A. Lindau and S. Weinzierl, "Assessing the Plausibility of Virtual Acoustic Environments," *Acta Acustica united with Acustica*, vol. 98, no. 5, pp. 804–810, 2012.

[6] M. Frank and D. Perinovic, "Matching auditory and visual room size, distance, and source orientation in virtual reality," in *AudioMostly 2022*. St. Pölten Austria: ACM, 2022, pp. 80–83.

[7] E. D. Sena, H. Hacıhabiboğlu, Z. Cvetković, and J. O. Smith III, "Efficient synthesis of room acoustics via scattering delay networks," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, 2015.

[8] F. Rumsey, *Spatial audio*. Oxford: Music technol. series, Focal, 2001.

[9] D. M. Howard and J. A. S. Angus, *Acoustics and psychoacoustics*, 4th ed. Amsterdam Heidelberg: Elsevier, 2009.

[10] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[11] S. J. Schlecht and E. A. Habets, "On lossless feedback delay networks," *IEEE Trans. on Signal Processing*, vol. 65, no. 6, pp. 1554–1564, 2016.

[12] S. Djordjevic, H. Hacihabiboğlu, Z. Cvetković, and E. De Sena, "Evaluation of the perceived naturalness of artificial reverberation algorithms," in *Proceedings of the 148th Audio Engineering Society Convention*, 2020.

[13] M. Geronazzo, J. Y. Tissieres, and S. Serafin, "A minimal personalization of dynamic binaural synthesis with mixed structural modeling and scattering delay networks," in *Proceedings of the 2020 International Conference on Acoustics, Speech and Signal Processing*, 2020.

[14] T. Potter, Z. Cvetković, and E. De Sena, "On the relative importance of visual and spatial audio rendering on VR immersion," *Frontiers in Signal Processing*, vol. 2, 2022.

[15] C. Yeoward, R. Shukla, R. Stewart, M. Sandler, and J. D. Reiss, "Real-time binaural room modelling for augmented reality applications," *Journal of the Audio Engineering Society, 69(11)*, pp. 818–833, 2021.

[16] A. Baldwin, S. Serafin, and C. Erkut, "Towards the design and evaluation of delay-based modeling of acoustic scenes in mobile augmented reality," in *Proceedings of the 4th VR Workshop on Sonic Interactions for Virtual Environments (SIVE-2018)*. IEEE, 2018.

[17] F. Stevens, D. T. Murphy, L. Savioja, and V. Välimäki, "Modeling sparsely reflecting outdoor acoustic scenes using the waveguide web," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 8, pp. 1566–1578, 2017.

[18] T. B. Atalay, Z. S. Gül, E. De Sena, Z. Cvetković, and H. Hacıhabiboğlu, "Scattering delay network simulator of coupled volume acoustics," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 582–593, 2022.

[19] M. Scerbo, O. Das, P. Friend, and E. De Sena, "Higher-order scattering delay networks for artificial reverberation," in *Proceedings of the 25th International Conference on Digital Audio Effects (DAFx-22)*, 2022.

[20] M. Vorländer, *Auralization*. Springer Berlin Heidelberg, 2008.

[21] M. R. Schroeder and K. H. Kuttruff, "On Frequency Response Curves in Rooms. Comparison of Experimental, Theoretical, and Monte Carlo Results for the Average Frequency Spacing between Maxima," *The Journal of the Acoustical Society of America, 34(1)*, pp. 76–80, 1962.

[22] J.-M. Jot and A. Chaigne, "Digital delay networks for designing artificial reverberators," in *90th AES Convention, Paris, France*, Feb. 1991.

[23] M. Burgess and W. Utley, "Reverberation times in British living rooms," *Applied Acoustics*, vol. 18, no. 5, pp. 369–380, 1985.

[24] G. Götz, S. J. Schlecht, and V. Pulkki, "A dataset of higher-order Ambisonic room impulse responses and 3D models measured in a room with varying furniture," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*. IEEE, 2021, pp. 1–8.

[25] A. Nowoświat, "Impact of Temperature and Relative Humidity on Reverberation Time in a Reverberation Room," *Buildings*, vol. 12, no. 8, p. 1282, Aug. 2022.

[26] N. Kaplanis, S. Bech, S. H. Jensen, and T. van Waterschoot, "Perception of Reverberation in Small Rooms: A Literature Study," in *55th AES Int. Conference on Spatial Audio, Helsinki, Finland*, Aug. 2014.

[27] P. Zahorik, "Perceptually relevant parameters for virtual listening simulation of small room acoustics," *The Journal of the Acoustical Society of America*, vol. 126, no. 2, pp. 776–791, Aug. 2009.

[28] A. J. Kolarik, B. C. J. Moore, S. Cirstea, E. Aggius-Vella, M. Gori, C. Campus, and S. Pardhan, "Factors Affecting Auditory Estimates of Virtual Room Size: Effects of Stimulus, Level, and Reverberation," *Perception*, vol. 50, no. 7, pp. 646–663, Jul. 2021.

[29] D. Cabrera, D. Jeong, H. J. Kwak, and J.-Y. Kim, "Auditory Room Size Perception for Modeled and Measured Rooms," *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, vol. 2005, no. 5, pp. 2995–3004, Aug. 2005.

[30] D. Cabrera, "Control of perceived room size using simple binaural technology," in *International Conference on Auditory Display (ICAD), Montreal QC Canada*, 2007.

[31] M. Schroeder and B. Logan, ""Colorless" artificial reverberation," *IRE Transactions on Audio*, vol. AU-9, no. 6, pp. 209–214, Nov. 1961.

[32] A. E. Stamps, "Effects of area, height, elongation, and color on percived spaciousness," *Environment and Behavior, 43(2)*, pp. 252–273, 2011.

[33] E. K. Sadalla and D. Oxley, "The Perception of Room Size: The Rectangularity Illusion," *Environment and Behavior*, vol. 16, no. 3, pp. 394–405, May 1984.

[34] H. Gagnon, S. Creem-Regehr, and J. Stefanucci, "Virtual room recreation: A new measure of room size perception," *SAP '21: ACM Symposium on Applied Perception*, no. 11, p. 1–10, 2021.

[35] C. Chen, R. Gao, P. Calamia, and K. Grauman, "Visual Acoustic Matching," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, LA, USA: IEEE, Jun. 2022, pp. 18 836–18 846.

[36] A. McCreery and P. Calamia, "Cross-modal perception of room acoustics," *J. Acoust. Soc. Am.*, vol. 120, no. 5, p. 3150, 2006.

[37] H.-J. Maempel and M. Horn, "The influences of hearing and vision on egocentric distance and room size perception under rich-cue conditions," in *Advances in Fundamental and Applied Research on Spatial Audio*, B. Katz and P. Majdak, Eds. IntechOpen, 2022.

[38] M. Schutte, S. D. Ewert, and L. Wiegrebe, "The percept of reverberation is not affected by visual room impression in virtual environments," *The Journal of the Acoustical Society of America*, vol. 145, no. 3, p. EL229–EL235, Mar. 2019.

[39] J. C. Gil-Carvajal, J. Cubick, S. Santurette, and T. Dau, "Spatial hearing with incongruent visual or auditory room cues," *Scientific Reports*, vol. 6, no. 37342, 2016.

[40] G. A. Ryu and K.-H. Yoo, "Key factors for reducing motion sickness in 360° virtual reality scene," Republic of Korea, Nov. 2020.

[41] University of York, "Anechoic data - openair," Jan. 2023, accessed: 6/6/2023. [Online]. Available: https://www.openair.hosted.york.ac.uk/?page_id=310

[42] D. Reality, "dearVR UNITY," January 2023, accessed: 6/6/2023. [Online]. Available: https://www.dear-reality.com/products/dearvr-unity

[43] Elcanetay, "Office room furniture," September 2016, accessed: 6/6/2023. [Online]. Available: https://assetstore.unity.com/packages/3d/props/furniture/office-room-furniture-70884

[44] iPoly3D, "Lowpoly server room props," July 2021, accessed: 6/6/2023. [Online]. Available: https://assetstore.unity.com/packages/3d/environments/lowpoly-server-room-props-197268

[45] Rokay3D, "Morgue room pbr," August 2021, accessed: 6/6/2023. [Online]. Available: https://assetstore.unity.com/packages/3d/environments/morgue-room-pbr-65817

[46] Mavi3D, "Basic bedroom starterpack," January 2023, accessed: 6/6/2023. [Online]. Available: https://assetstore.unity.com/packages/3d/props/basic-bedroom-starterpack-215986

[47] AiKodex, "Simple garage," July 2021, accessed: 6/6/2023. [Online]. Available: https://assetstore.unity.com/packages/3d/props/interior/simple-garage-197251

[48] G. V. Glass, P. D. Peckham, and J. R. Sanders, "Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance," *REVIEW OF EDUCATIONAL RESEARCH*, vol. 42, no. 3, pp. 237–288, 1972.

[49] D. Cabrera, C. Pop, and D. Jeong, "Auditory Room Size Perception: A Comparison of Real versus Binaural Sound-fields," *New Zealand*, 2006.

[50] D. Cabrera, "Acoustic Clarity and auditory room size perception," 2007.

[51] T. Schultz, "Diffusion in reverberation rooms," *Journal of Sound and Vibration*, vol. 16, no. 1, pp. 17–28, May 1971.

[52] J. Borish, "Extension of the image model to arbitrary polyhedra," *The Journal of the Acoustical Society of America*, vol. 75, no. 6, pp. 1827–1836, Jun. 1984.