Factors affecting auditory estimates of virtual room size: Effects of stimulus, level, and reverberation

Andrew J. Kolarik[1*], Brian C. J. Moore[1, 2], Silvia Cirstea[1, 3], Elena Aggius-Vella[4, 5], Monica Gori[4], Claudio Campus[4], Shahina Pardhan[1]

[1] Vision and Eye Research Institute, School of Medicine, Anglia Ruskin University, Cambridge, United Kingdom

[2] Cambridge Hearing Group, Department of Psychology, University of Cambridge, Cambridge, United Kingdom http://hearing.psychol.cam.ac.uk/

[3] School of Computing and Information Science, Anglia Ruskin University, Cambridge, United Kingdom

[4] Unit for Visually Impaired People (U-VIP), Center for Human Technologies, Fondazione Istituto Italiano di Tecnologia, Genoa, Italy

[5] Institute for Mind, Brain and Technology Ivcher School of Psychology Inter-Disciplinary Center (IDC), Herzeliya, Israel

* Corresponding author

E-mail: andrew.kolarik@aru.ac.uk (AJK)

Abstract

When vision is unavailable, auditory level and reverberation cues provide important spatial information regarding the environment, such as the size of a room. We investigated how room size estimates were affected by stimulus type, level and reverberation. In Experiment 1, fifteen blindfolded participants estimated room size after performing a distance bisection task in virtual rooms that were either anechoic (with level cues only) or reverberant (with level and reverberation cues) with a relatively short reverberation time of T_{60} = 400 ms. Speech, noise, or clicks were presented at distances between 1.9 and 7.1 m. The reverberant room was judged to be significantly larger than the anechoic room ($p<0.05$) for all stimuli. In Experiment 2 only the reverberant room was used and the overall level of all sounds was equalized, so only reverberation cues were available. Ten blindfolded participants took part. Room size estimates were significantly larger for speech than for clicks or noise. The results show that when level and reverberation cues are present, reverberation increases judged room size. Even relatively weak reverberation cues provide room size information, which could potentially be used by blind or visually impaired individuals encountering novel rooms.

Keywords: spatial hearing; auditory distance; sound localization; depth; room size; reverberation

Introduction

Advances in binaural technology allow a wide variety of spatial configurations of sound source and listener to be simulated in virtual environments and provide experimental control over the acoustic characteristics of the simulated rooms. These advances have enabled the use of virtualization techniques to explore a range of issues, including how reverberation affects speech understanding (Ellis and Zahorik 2019), and how the availability of visual depth information increases the accuracy of auditory estimates of distance within a room (Anderson and Zahorik 2014). How audition provides spatial information for judgments of the distance of sound sources (Bidart and Lavandier 2016; Kolarik et al 2013a, b; Zahorik 2002) and room size (Kolarik et al 2013d; Kolarik et al 2020) has also been investigated using virtualization techniques. Audition provides valuable spatial information when vision is unavailable and is critical for spatial awareness and navigation by blind people. Although many studies have investigated the factors affecting auditory judgements of sound azimuth and distance (Ahveninen et al 2014; Kolarik et al 2016a; Moore 2012; Zahorik et al 2005), the factors affecting room size estimates have received little attention. The current study used virtualization techniques to investigate factors affecting auditory judgments of room size.

When first entering a novel room, in the absence of vision, people might use distance cues or spatial information based on reverberation from sound-producing sources to make estimates of room size. This information allows a preliminary internal representation of the room layout to be generated. One possibility is that room size is estimated from the judged distance of the farthest sound source within the room, which is an indicator of the nearest possible distance of the far wall (Calcagno et al 2012). Consistent with this idea, significant positive correlations have been reported between room size estimates and farthest-distance estimates (Kolarik et al 2013d). However, listeners consistently underestimate the distance of remote sound sources (for reviews, see Kolarik et al 2016a; Zahorik et al 2005).

The primary auditory distance cues are level, when the level of the source is fixed (Ashmead et al 1990; Coleman 1963; Gamble 1909; Mershon and King 1975; Strybel and Perrott 1984), and direct-to-reverberant energy ratio (DRR, Bronkhorst and Houtgast 1999; Mershon et al 1989; Mershon and King 1975; Zahorik 2002). The effectiveness of the DRR cue is dependent upon the room acoustic characteristics, which are usually quantified by the reverberation time (T_{60}), which is the time required for the sound level to fall by 60 dB after the source is turned off. The T_{60} value is strongly influenced by the size of the room and the sound absorption characteristics of the walls. In reverberant rooms, either level, or DRR, or both, might be used to make farthest-distance estimates on which room-size estimates could be based, although distance estimates made when level cues are unavailable tend to be much less accurate than when level cues are available (Mershon and Bowers 1979). In anechoic rooms, only level cues are available.

Another possibility is that initial estimates of room size are based on the characteristics of the reverberation, for example the range of time delays of the echoes (a room with a wide range of echo delays will be judged as larger than a room with a small range of echo delays). Rooms with longer reverberation times are estimated to be larger than rooms with shorter reverberation times (Etchemendy et al 2017; Mershon et al 1989), suggesting that listeners use their experience of the association between room size and reverberation time when judging room size.

Room size estimates can be affected by the sound stimulus. For normally sighted participants, estimates of room size were reported to be larger and more veridical for speech sounds than for music or noise bursts, but only when reverberation was present (Kolarik et al

2013d). It is possible that this occurs in rooms with a long reverberation time because reverberation fills in the dips in strongly amplitude-modulated signals such as speech (Bidart and Lavandier 2016), thereby providing information about room size that would not be present for less modulated stimuli, such as noise. For music stimuli, participants may implicitly assume that the reverberation is part of the music recording rather than originating from room acoustics, and thus not use reverberation in their judgments of room size. Familiarity with the acoustic characteristics of speech may also affect the room size estimates (Kolarik et al 2013d), as has previously been shown for distance estimates. Underestimation of distance tends to be greater when listening to sounds with unfamiliar acoustic characteristics, such as noise (Zahorik 2002), than when listening to familiar sounds, such as speech (Brungart and Scott 2001; Cochran et al 1968; Gardner 1969; von Békésy 1949).

Gotoh et al (1977) showed that increasing the time delay of simulated room reflections (i.e. the room reverberation) relative to the leading (direct) part of the sound increased perceived distance and that judged distance increased with increasing number of reflections. Mershon et al (1989) asked blindfolded participants to judge the apparent distance of white noise bursts in a room in which the reverberation time was manipulated by the addition of sound absorbing material. The room was designated as either a “live” reverberant room ($T_{60} \approx 1.7$ s) or a “dead” ($T_{60} \approx 0.4$ s) room. Additional acoustic information about the room was explicitly provided by vocal information from the experimenter and participants’ own vocal responses. The “live” room was judged to be larger than the “dead” room. Etchemendy et al (2017) reported that a highly reverberant room (T_{60} = 3.9 s) was estimated to be significantly larger than a near-anechoic room (T_{60} = 0.1 s) by normally sighted participants, who judged room size after performing a visual absolute distance judgment task using illuminated targets in a dark room. The anechoic room had a larger volume (285 m^3) than the reverberant room (189 m^3). Playback of recorded instructions and a microphone allowing communication between the participant and experimenter provided acoustical information. Using a shorter reverberation time (T_{60} = 700 ms) than for previous studies, Kolarik et al (2013d) reported that for a speech stimulus, a virtual reverberant room was judged to be larger than a virtual anechoic room by blindfolded participants. The room size judgments were made after a distance perception task had been performed. With music and noise stimuli, the anechoic and reverberant rooms were judged to be of similar size. The virtualization methods utilized eliminated additional acoustic information from vocal responses.

In previous studies that investigated auditory room size judgments, the level cue for sound source distance was always present (Etchemendy et al 2017; Kolarik et al 2013d; Mershon et al 1989). However, level is not always a reliable cue because the level at the source can vary, especially for speech (Zahorik et al 2005), in which case listeners might rely more heavily on reverberation information. To our knowledge, room size estimates based on reverberation information alone have not previously been assessed. Furthermore, in previous studies, participants performed absolute auditory (Kolarik et al 2013d; Mershon et al 1989) or visual (Etchemendy et al 2017) distance judgments of sound sources before estimating room size, making it more likely that they used the farthest judged stimulus distance as an indicator of the nearest possible position of the far wall. No studies have yet assessed room size estimates when absolute distance judgments of the farthest sound source distance are not made.

The aim of the current experiments was to address the following gaps in the literature. Firstly, previous studies have generally compared distance and size estimates for rooms with relatively long room reverberation times (700 ms or more) and rooms with relatively short reverberation times (e.g. 400 ms) (Etchemendy et al 2017; Kolarik et al 2013d; Mershon et al 1989). Whether rooms with short reverberation times are judged to be larger than anechoic rooms is not yet known. It should be noted that the size of a virtual anechoic room is not defined; the signals reaching the virtual listener's ear are independent of the size of the simulated room. Nevertheless, without visual cues, it is likely that anechoic rooms are not perceived to have an infinite size, since listeners can estimate room size based on the judged distances of the farthest sound sources, with distance estimates based on level cues alone. This idea is supported by the findings of previous work for anechoic rooms that were real (Etchemendy et al 2017) or virtual (Kolarik et al 2013d), in which room size estimates made using sound increased with the distance of the farthest source. In experiment 1, we assessed whether a short reverberation time was sufficient to influence room-size judgments. This was done by asking participants to judge the size of an anechoic room and a virtual room with T_{60} = 400 ms. T_{60} values are approximately 200 ms for audiometric test booths, 400-800 ms for offices and living rooms, 400-1200 ms for classrooms, and up to or exceeding 3000 ms for churches and auditoriums (Crukley et al 2011; Nábělek and Nábělek 1994; Smaldino et al 2008).

Secondly, it is not yet known whether room reverberation time affects room size estimates when judgements of the absolute distance of the farthest sound source distance are not made. To avoid absolute distance judgments, participants performed a spatial bisection task before estimating room size. Three sounds (A, B and C) were presented at different virtual distances, with B placed between A and C, and the task was to judge whether B was closer to A or C.

Thirdly, room size estimates made using reverberation information alone have not yet been reported. In experiment 2, the overall level of the sounds at the participant's ears was equalized, in order that room size estimates made on the basis of reverberation information alone could be assessed.

Lastly, the effect of different stimulus types on room size estimates made when level cues only, reverberation cues only, or both types of cue are available in virtual rooms with a relatively short reverberation time has not yet been assessed. In experiments 1 and 2, participants made room size judgements for speech, noise and click stimuli, chosen because they varied in their spectro-temporal characteristics. Clicks were included since click-like stimuli often occur in everyday life, and in principle they provide good information about the pattern of reverberation in a room, but they have not previously been used in experiments assessing room size estimates. Further details regarding the experimental hypotheses are provided below.

General Methods

The simulation methods have been described in our previous studies investigating auditory judgments of distance (Kolarik et al 2013a, b; Kolarik et al 2013c; Kolarik et al 2017a) and room size (Kolarik et al 2013d; Kolarik et al 2020). In the current study, the virtualization was made more realistic by convolving sound reflections with the appropriate head-related transfer functions (HRTFs) in addition to doing this for the direct sound component (Culling et al 2013; Culling 2013; Moore et al 2016). The distance bisection task used (for which the results are described in a paper currently under review; the focus of the current study was to assess estimated room size) was developed from that used for azimuth-bisection studies (Gori et al 2014; Tonelli et al 2015; Vercillo et al 2016; Vercillo et al 2015; Vercillo et al 2018).

Experiment 1

Experiment 1 tested the following hypotheses: (1) participants would judge a virtual room with a shorter reverberation time than used in previous studies (400 ms) to be larger than a virtual anechoic room; (2) Room size estimates would be larger for speech than for noise or clicks for a reverberant virtual room, while for an anechoic room size estimates would be similar for all stimuli, based on the findings of Kolarik et al (2013d).

Methods

Participants

There were 15 participants (7 females, mean age 36 yrs, range 28-50 yrs), with good visual acuities of 6/6 in each eye, equivalent to 20/20 acuity (measured as previous work has shown that visual loss can affect room size estimates (Kolarik et al 2013d)). Audiograms measured following the procedures described by the British Society of Audiology (2011) confirmed that all participants had normal or near-normal hearing, indicated by pure-tone-average (PTA) better-ear hearing thresholds across 0.5, 1, 2, 4, 6, and 8 kHz ≤25 dB HL. Experimental procedures followed the tenets of the Declaration of Helsinki. Informed, written consent was obtained following description of the nature and possible consequences of the study. Experimental approval was granted by the Anglia Ruskin Research Ethics Panel.

Apparatus

The experiment was conducted in a quiet room in Anglia Ruskin University. An Asus AA185 computer with a Realtek High Definition sound card was used to present sounds over

Sennheiser HD 280 PRO headphones. The sample rate was 22.05 kHz. Stimuli were generated using a custom-written MATLAB (Mathworks, Inc.) script, which also created a response interface.

Sounds were presented at a virtual height of 1 m, at 0° elevation and 0° azimuth relative to a virtual participant located at 1 m from the shorter wall at a height of 1 m, facing forward (Figure 1).

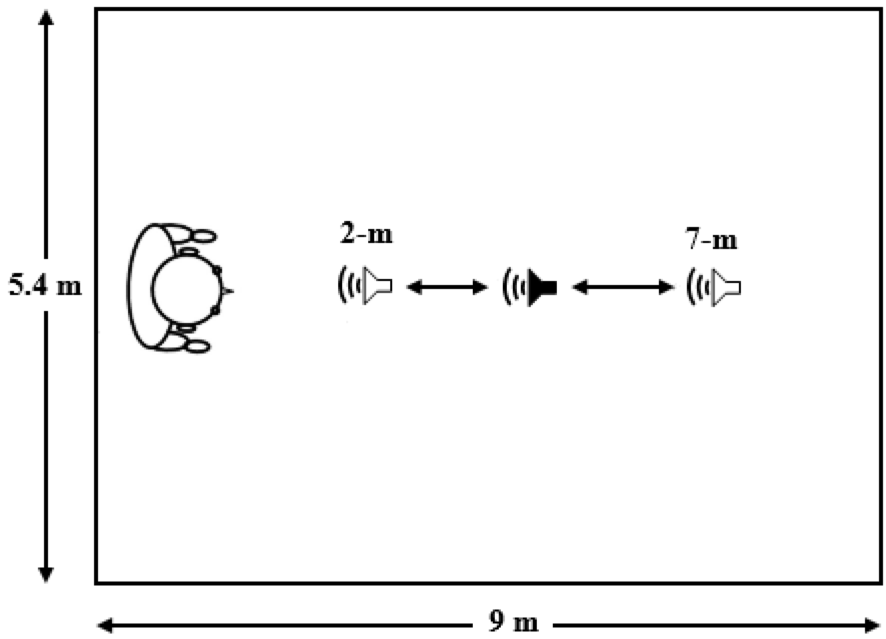

Figure 1. Layout of the virtual room. The position of the participant was simulated to be on the midline of the shorter wall. Loudspeakers show the positions of the virtual sound sources, which were presented in front of the participant. The locations of the reference sound sources are shown by white loudspeakers and the location of the probe sound is shown by the black loudspeaker.

Stimuli were speech, broadband noise, or single clicks. The speech was the British English phrase “Where am I”, spoken by a male at a conversational level, with a duration of 850 ms, as used in previous work studying binaural enhancement processing for hearing aids

(Moore et al 2016). The broadband (0.6-11 kHz) noise had a duration of 500 ms, including rise/fall times of 10 ms. The duration of the click was 3 ms. For a simulated sound source distance 1 m from the participant, the stimulus level was 65 dB SPL (unweighted) at the center of the participant's head. The level of the virtual sound source was fixed, and the level at the center of the participant's head decreased as the virtual distance increased.

An image-source model (ISM) (Allen and Berkley 1979; Lehmann and Johansson 2008) was used to simulate a virtual anechoic room or a reverberant 9 (length) x 5.4 (width) x 2 m (height) (T_{60} = 400 ms) room. The volume of the reverberant virtual room was 97.2 m^3. As noted earlier, the size of the simulated anechoic room does not affect the signals reaching the listener's ears, so the volume of the simulated anechoic room was nominal only. The ISM produced binaural room impulse responses (BRIRs) between the simulated sound source and the simulated participant's head, and calculated ray paths between the virtual sound source and the virtual head. For each individual ray at each ear, the angle of incidence at the virtual head was used to select an appropriate head-related impulse response (HRIR), taken from a database of publicly available recordings made using a KEMAR manikin (Gardner and Martin 1995). Every HRIR was delayed and scaled appropriately, depending on the ray path length and the absorption characteristics of the surfaces within the room that reflected the ray. A BRIR was created by adding the HRIRs. Convolution of the BRIR with a sound stimulus generated a simulation of the sound heard within the virtual room at the set virtual distance.

Externalization of the stimuli (hearing the stimuli outside of the head) or the perceived distance of the simulated sounds might have been affected by employing non-individualized HRIRs in the simulation. However, it has been reported that using non-individualized HRIRs to simulate virtual distance does not adversely affect auditory distance judgements (Prud'homme and Lavandier 2020). Previous work using similar virtualization methods to the current study showed that participants judged sound distance approximately accurately for virtual sounds 1 m away, and made systematic underestimations as virtual distance increased (Kolarik et al 2013b; Kolarik et al 2017a), as has been found for judgments of real sound sources (Coleman 1962; Mershon and Bowers 1979; Zahorik et al 2005), supporting the idea that the virtualization techniques provided an adequate simulation of a real room environment.

On each trial, three sounds were presented. The first and third sounds were references and the second was the probe. The inter-stimulus interval was 500 ms. The mean simulated distances of the reference sounds were 2 and 7 m, and their order (2 then 7, or 7 then 2) was selected randomly at each trial. The two reference sounds were always separated by 5 m, and they were presented either at fixed simulated distances, or at distances that were jittered from trial to trial by ±0.1 m. The simulated distance of the probe was randomly chosen from a number of possible distances: 2.4, 2.8, 3.2, 3.6, 4.0, 4.4, 4.8, 5.2, 5.6, 6, 6.4, and 6.8 m. These were chosen following pilot testing to map out a complete psychometric function ranging from "probe closer to near reference" to "probe closer to far reference."

Procedures

Participants were blindfolded before entering the testing room and escorted to their chair. They were given headphones, and instructed to imagine themselves sitting within a rectangular room of an unspecified size. Loudspeakers positioned at various distances from them would generate three sounds and they should verbally report if the second sound was closer to the first or the third sound. No feedback was given and response time was not constrained. In a given block, a single stimulus type (speech, noise, or click) and a single experimental condition (anechoic or reverberant) were presented. For each block there were 120 trials with 10 repetitions of each probe distance. After a block was completed, participants estimated the room length, width, and height. The experimenter recorded participant judgments. The order of presentation of the six blocks (3 stimulus types and 2 room conditions) was randomized. The experiment was completed in one or two sessions of approximately 2 hours and 30 minutes total with rest breaks.

Results

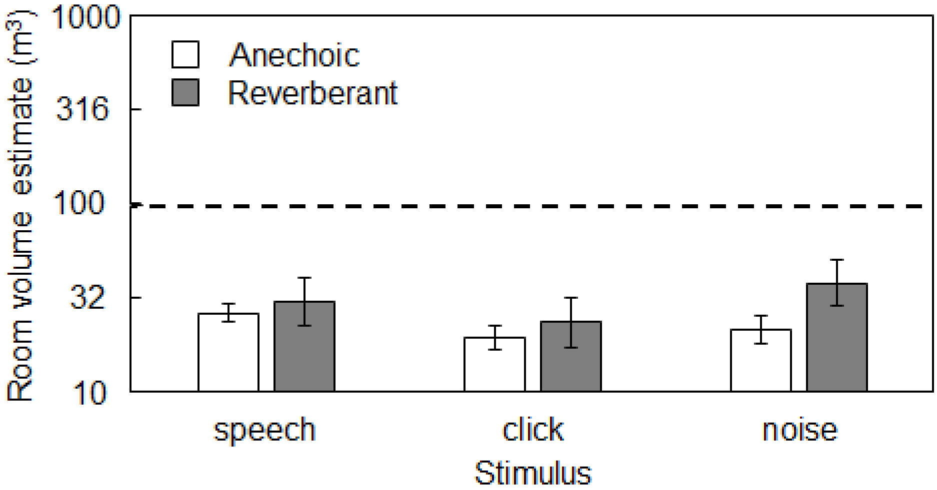

Figure 2. Geometric mean room volume estimates for virtual anechoic (open bars) and reverberant (grey bars) rooms in Experiment 1. In the anechoic room, only level cues for distance were available, while in the reverberant room, both level and reverberation cues were available. Error bars show ±1 standard error of the mean. The dashed line represents veridical performance for the reverberant room (room size = 97.2 m^3). The y axis is logarithmic.

Figure 2 shows geometric mean estimated room volumes for speech, clicks, and noise in the anechoic and reverberant virtual rooms. For all stimuli, participants underestimated the size of the virtual reverberant room and the reverberant room was judged to be larger than the anechoic room. A repeated-measures ANOVA was conducted on the log-transformed volume estimates with factors room reverberation time (anechoic, reverberant), and stimulus (speech, clicks, and noise). There was a main effect of room reverberation time ($F_{1, 14} = 7.32$, $p<0.05$), but not stimulus ($F_{2, 28} = 1.12$, ns), and no interaction between room reverberation time and stimulus ($F_{2, 28} = 0.23$, ns).

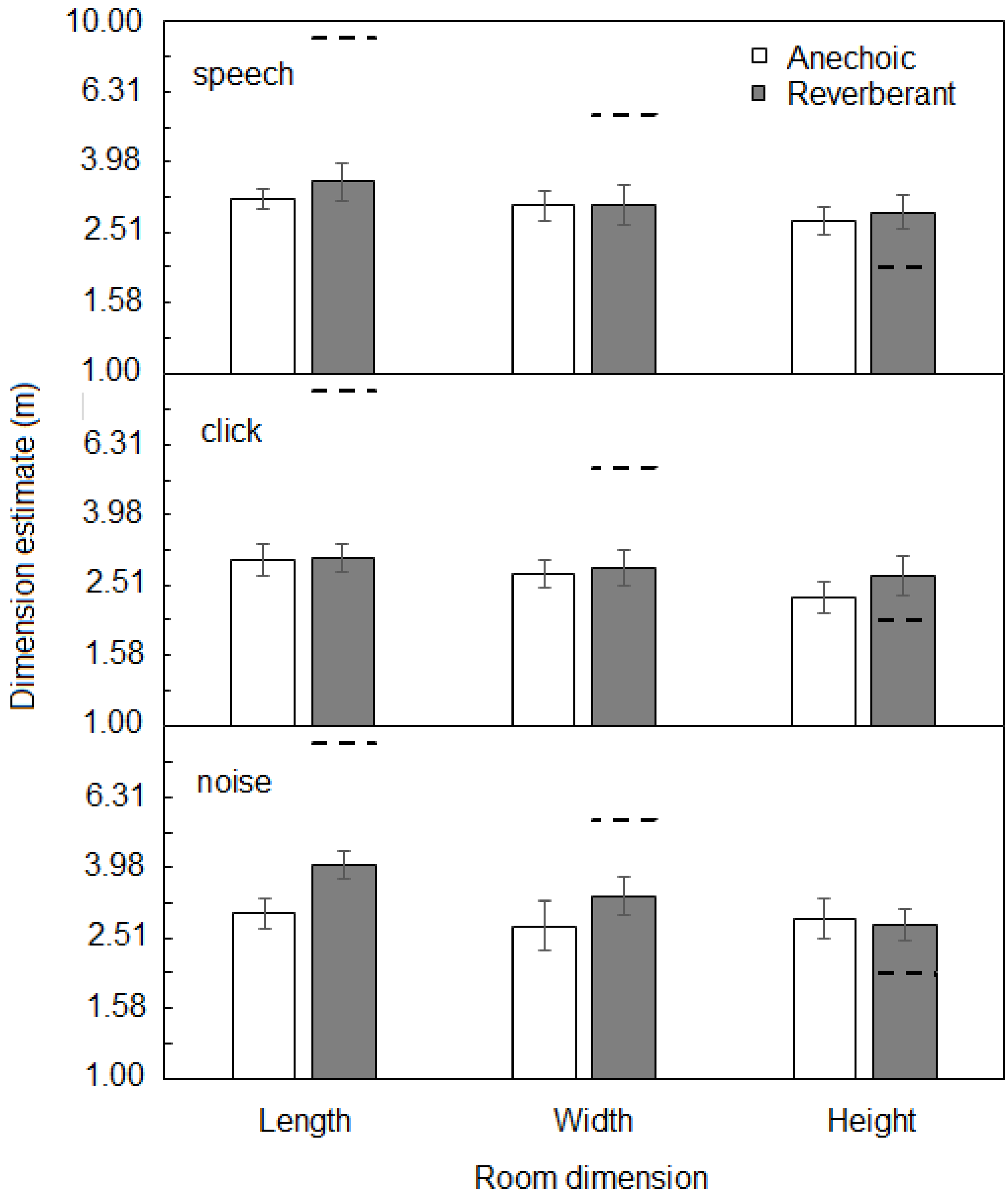

Figure 3. Geometric mean room dimension estimates for virtual anechoic (open bars) and reverberant (grey bars) rooms for Experiment 1. Error bars show ±1 standard error of the mean. The dashed lines represent veridical performance for the reverberant room (length = 9 m, width = 5.4 m, height = 2 m). The y axis is logarithmic.

Figure 3 shows geometric mean estimated room dimensions for speech, clicks, and noise in the anechoic and reverberant virtual rooms. For all reverberant stimuli, participants underestimated length and width, but overestimated height. A repeated-measures ANOVA was conducted on the log-transformed room dimension estimates with factors type of dimension (length, width, height), room reverberation time (anechoic, reverberant), and stimulus (speech, clicks, and noise). There were main effects of room reverberation time ($F_{1, 14} = 7.19$, $p<0.05$) and type of dimension ($F_{2, 28} = 5.00$, $p<0.05$), but not stimulus ($F_{2, 28} = 1.1$, ns), and no significant interactions ($p>0.05$).

Table 1 shows Pearson correlations between room dimension estimates, and between volume estimates and room dimension estimates. For noise and click stimuli, there were significant correlations between each of the room dimensions. For room dimension estimates for speech in an anechoic virtual room, only the correlation between width and length was significant. For room dimension estimates for speech in the reverberant virtual room, correlations between height and width, and between width and length only were significant. Significant correlations were observed between all volume estimates and room dimension estimates.

Stimulus	Room	HxW	HxL	WxL	VxH	VxW	VxL
Speech	Anechoic	0.51	0.39	0.56*	0.73**	0.87**	0.82**
	Reverberant	0.57*	0.27	0.59*	0.76**	0.90**	0.76**
Noise	Anechoic	0.73**	0.67**	0.74**	0.86**	0.94**	0.89**
	Reverberant	0.66**	0.64**	0.84**	0.83**	0.94**	0.92**
Click	Anechoic	0.70**	0.57*	0.81**	0.86**	0.93**	0.89**
	Reverberant	0.75**	0.65**	0.90**	0.84**	0.97**	0.94**

Table 1. Correlations between room dimension estimates for Experiment 1. Correlations are shown between length (L), width (W) and height (H) estimates, and between volume (V) estimates and room dimension estimates, for the three stimulus types and the two room reverberation times. In this and subsequent tables, significant differences are indicated by asterisks: *p<0.05, **p< 0.01.

To investigate whether the distance data were related to the reported estimates of room volume, Pearson correlations were conducted between bisection thresholds/Point of Subjective Equality (PSE) judgements and room volume estimates (Table 2). No significant correlations were observed, with the exception of the reverberant speech and anechoic click thresholds. The finding that no significant correlations were observed in the majority of conditions suggests that bisection judgments were independent of room size judgments.

Stimulus	Room	Threshold	PSE
Speech	Anechoic	0.31	0.18
	Reverberant	0.61*	0.13
Noise	Anechoic	0.28	0.38
	Reverberant	0.31	0.09
Click	Anechoic	0.58*	-0.13
	Reverberant	-0.22	-0.34

Table 2. Correlations between bisection thresholds/PSE judgements and room volume estimates for Experiment 1.

Discussion

The results support the first hypothesis for experiment 1: participants judged the virtual reverberant room to be larger than the anechoic room, even though the reverberant room had a shorter reverberation time than has been studied previously. The second hypothesis was only partially supported by the results. In the anechoic room, room size estimates were similar for all stimuli, as predicted. However, in the reverberant room, size estimates were not larger for speech than for noise or clicks, which differs from the results of previous work (Kolarik et al 2013d). As the room reverberation time was relatively small, participants may have estimated the room size based mainly on farthest-distance estimates using level cues, with only a small contribution from reverberation cues. In order to establish whether stimulus type affected room size estimates when reverberation information alone was present, experiment 2 tested performance in equalized-level conditions.

The results showed that for all reverberant stimuli, participants on average underestimated the room length and width dimensions but overestimated the height, indicating that the underestimation of the volume estimates in the reverberant virtual room was primarily due to underestimation of the length and width. The room height was on average overestimated. This might have been due to the relatively low virtual ceiling height of 2 m used in the experiment and to the low position of the simulated listener. The participants may have been influenced by their expectation that large rooms typically have heights exceeding 2 m. It is possible that the use of larger virtual room heights would result in all room dimensions being underestimated. All of the correlations between room dimensions for noise and clicks were significant, indicating that the judgments were not independent; participants who reported a relatively large estimate for one room dimension also tended to report a relatively large estimate for the other room dimensions. For speech, however, correlations between height estimates and length and width estimates were not significant in the anechoic condition, and the correlation between height and length estimates in the reverberant condition was not significant. It is possible that for speech stimuli, participants expected the height to vary only over a limited range, as is typically the case in real environments, although it is unclear why this would occur for speech but not for noise or clicks. Further investigation of this effect is needed. Significant correlations were observed between all volume estimates and room dimension estimates.

Experiment 2

In experiment 2, level cues were removed by equalizing the overall level of all of the stimuli at the participant's ears. Although not "ecological", equalization has been utilised in some previous studies to isolate the use of reverberation cues for auditory distance perception (Akeroyd et al 2007; Bidart and Lavandier 2016; Kolarik et al 2013a, b; Mershon and Bowers 1979). Participants can make auditory distance estimates in equalized-level conditions, but shorter reverberation times tend to result in greater compression of distance estimates than when level cues are present, with farther distances being strongly underestimated (Bidart and Lavandier 2016; Mershon and Bowers 1979). For speech in a virtual room with $T_{60} \approx 1.5$ s, farthest distances were moderately underestimated (the mean estimated distance was approximately 7 m for a simulated distance of 10 m, see Figure 4 in the control experiment of Bidart and Lavandier 2016). For 200-Hz square-wave signals in a classroom with $T_{60} \approx 700$ ms, underestimation was proportionally greater (the median perceived distance was approximately 3 m for the furthest physical distance of 6 m, see Figure 2 of Mershon and Bowers 1979). Although differences in room size and stimuli may have affected the extent of distance underestimation, it seems likely that for the short reverberation time of 400 ms used in the current equalized-level condition, farthest distances would be even more underestimated. Instead of using the judged distance to the farthest sound to estimate the room size, participants would be more likely to base their estimates of room size on the range of time delays of the echoes, for which stimulus effects are more likely to be apparent (Bidart and Lavandier 2016). It was thus hypothesized that participants would estimate the room to be larger when listening to speech than when listening to noise or clicks.

Methods

Participants, apparatus, and procedures

There were 10 participants (5 females, mean age 25.1 yrs, range 19-35 yrs). None of the participants in Experiment 2 took part in Experiment 1. All participants had good visual acuities and normal or near-normal hearing, based on the methods and criteria described for Experiment 1. Informed consent was obtained for all participants. Apparatus, data acquisition and procedures matched those for Experiment 1. The stimuli were the same as for the reverberant-room condition of Experiment 1, except that the level was equalized for all stimuli to be 65 dB SPL. As before, the spatial-bisection task was performed before the room size estimates were obtained for each stimulus type. The experiment was completed in one or two sessions of approximately 1 hour and 15 minutes total with rest breaks.

Results

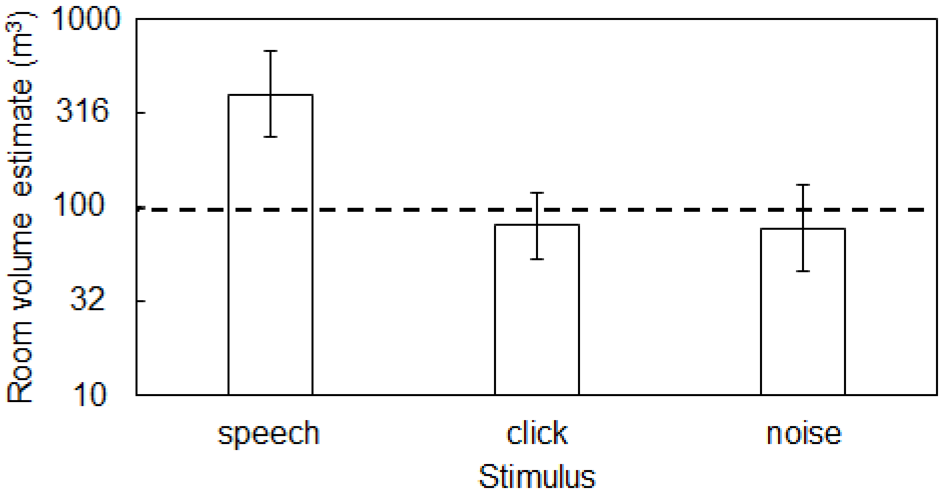

Figure 4. Geometric mean room volume estimates based on equalized-level speech, clicks and noise in Experiment 2 (similar to Figure 2). In this virtual room, only reverberation cues for distance were available. Error bars show ±1 standard error of the mean. The dashed line represents veridical performance (room size = 97.2 m^3). The y axis is logarithmic.

Figure 4 shows geometric mean estimated room volumes for equalized-level speech, clicks, and noise. A repeated-measures ANOVA was conducted on the log-transformed estimates with stimulus (speech, clicks, and noise) as a factor. There was a main effect of stimulus ($F_{2, 18}$ = 6.35, p<0.01). Post hoc paired samples *t*-tests with Bonferroni correction showed that room volume estimates were significantly larger for speech than for clicks (p =

0.015) or noise (p = 0.001). There was no significant difference between room volume estimates for noise and clicks.

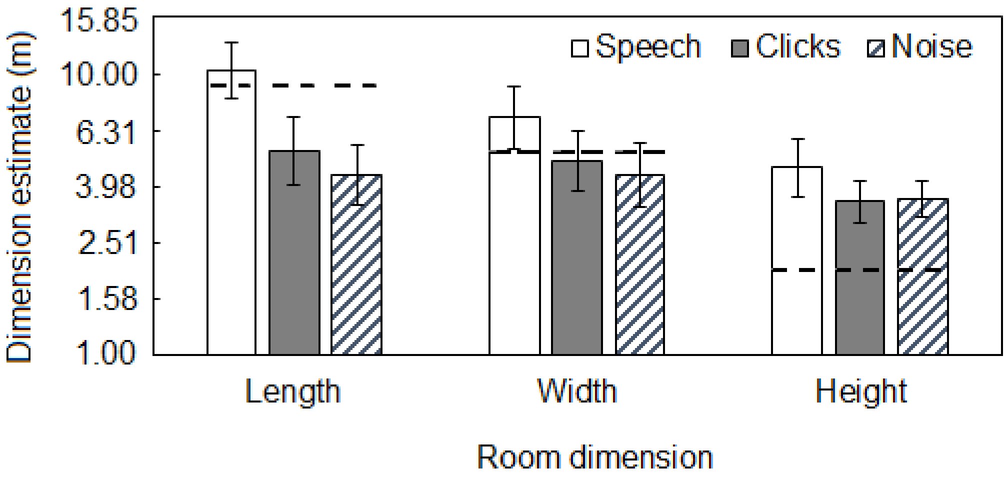

Figure 5. Geometric mean room dimension estimates for Experiment 2. Data for equalized-level speech, clicks and noise are shown by open, grey and diagonal line-filled bars, respectively. Error bars show ±1 standard error of the mean. The dashed lines represent veridical performance (length = 9 m, width = 5.4 m, height = 2 m). The y axis is logarithmic.

Figure 5 shows geometric mean estimated room dimensions for speech (open bars), clicks (grey bars), and noise (diagonal line-filled bars). For speech, all dimensions were overestimated. For clicks and noise, length and width were underestimated and height was overestimated. A repeated-measures ANOVA was conducted on the log-transformed values with factors room dimension type (length, width, height) and stimulus (speech, clicks, and noise). There was no significant main effects of stimulus ($F_{2, 18}$ = 3.24, ns) or room dimension type ($F_{2, 18}$ = 3.20, ns), and no interaction between stimulus and room dimension ($F_{4, 36}$ = 1.31, ns).

Table 3 shows Pearson correlations between room dimension estimates, and between volume estimates and room dimension estimates, for Experiment 2. For room dimension estimates, there were significant correlations between length and width only, for each of the three stimulus types. There were significant correlations between volume estimates and length and width estimates, but not height estimates.

Stimulus	HxW	HxL	WxL	VxH	VxW	VxL
Speech	-0.07	0.13	0.83**	0.46	0.82**	0.90**
Noise	0.44	0.34	0.94**	0.37	0.96**	0.93**
Click	0.15	0.04	0.93**	0.35	0.71*	0.81**

Table 3. Correlations between room dimension estimates made in Experiment 2. Correlations are shown between length (L), width (W) and height (H) estimates, and between volume (V) estimates and room dimensions, for speech, noise and clicks.

To investigate whether the distance data were related to the reported estimates of room volume, Pearson correlations were conducted between the bisection thresholds/Point of Subjective Equality (PSE) judgements and the estimates of room volume (Table 4). With the exception of the speech PSE, no significant correlations were observed, suggesting that bisection judgments were independent of room size judgments.

Stimulus	Threshold	PSE
Speech	0.08	-0.66*
Noise	0.03	-0.22
Click	0.21	0.05

Table 4. Correlations between bisection thresholds/PSE judgements and room volume estimates for Experiment 2.

Discussion

For experiment 2, on average, room volume estimates based on speech stimuli (for which all room dimensions were on average overestimated) were larger than those using clicks and noise stimuli. This contrasts with the findings of experiment 1, where the level cue was present, which showed that the type of stimulus did not affect room size estimates. The over-estimation of room size for the speech stimuli in experiment 2 may be connected with the expectations of the participants about vocal effort. For a distant talker to produce a sound level of 65 dB SPL at the position of the simulated listener, the sound level near the talker would have to be much higher than 65 dB SPL. This higher level would normally be associated with greater vocal effort, which changes the voice quality, leading to an increase in the ratio of high-frequency to low-frequency energy (Pearsons et al 1976). In our simulation, the spectrum of the simulated source was held constant, i.e. the expected change in spectral shape did not occur. As a result, the ratio of high-frequency to low-frequency energy at the simulated position of the participant was lower than "expected" for distant sources, and this may have led the participants to judge the distant stimuli to be farther away than they actually were, since, for most stimuli, greater distance is associated with a lower ratio of high-frequency to low-frequency energy. This overestimation of the distance of the farthest sources, may have led to the overestimates of room size for the speech stimulus. It should be noted that it is the implicit expectations of the listener that are important here; in practice the ratio of high-frequency to low-frequency energy at the listener's ears changes markedly over talker-listener distances distances from 2 to 7 m only in rooms in which the surfaces absorb more high-frequency than low-frequency energy.

Differences in the spectra or temporal structure of the stimuli may have also contributed to the results. In particular, the effects of reverberation may have been easier to hear for the speech stimulus owing to the temporal dips in the speech, which would be partially filled in by the reverberation.

As mentioned in the Methods section of Experiment 1, evidence from previous work (Kolarik et al 2013b; Kolarik et al 2017a; Prud'homme and Lavandier 2020) suggests that the simulation methods used in the current study provide an adequate simulation of a real room, and as a result it is likely that the finding that room volume estimates were significantly larger for speech than for clicks or noise would hold in a real room with low reverberation.

In experiment 2, significant correlations were observed between length and width only for all stimuli, indicating that although length and width judgments were related, height judgments were independent of length and width judgments. These findings differ from those of Experiment 1, where for noise and clicks all of the correlations between height and other room dimensions were significant, although for speech correlations between height estimates and length and width estimates were not significant in the anechoic room, and the correlation between height and length in the reverberant room was not significant. Taken together, these findings suggest that height judgments are independent of length and width judgements when only reverberation is used to estimate room size, but this is not the case when level is also available (at least for click and noise stimuli).

General Discussion

The representation approach to sensory processing assumes that individuals establish an internal representation of the three-dimensional spatial structure of their surroundings using the available sensory information, and this internal representation is used in navigation or path planning (Frenz and Lappe 2005; Turano et al 2005), for which accurate judgments of room size would be beneficial. Experiment 1 showed that although participants underestimated room volume when both level and reverberation cues were available, a reverberation time of 400 ms was sufficiently long to increase estimates of the volume of a room. Experiment 2 showed that when level cues were unavailable, room volume estimates based on clicks and noise were smaller than estimates based on speech stimuli. Theoretically, there must be a lower limit to T_{60} below which judgments of room size are not affected. The current results suggest that 400 ms falls above this limit. The shortest T_{60} that affects judgments of room size remain to be determined.

Although the bisection task did not require absolute judgments of farthest sound source distance to be made, participants in experiment 1 likely formed an estimate of the farthest source distance, which could be used as an indicator of the nearest possible location of the far wall. Previous work showed that room reverberation increased absolute distance judgments for auditory targets (Brungart and Scott 2001; Mershon et al 1989; Nielsen 1993), and that rooms with longer reverberation times were estimated to be larger than rooms with shorter reverberation times (Etchemendy et al 2017; Mershon et al 1989). The current study showed that a reverberant virtual room was judged to be larger than a virtual anechoic room for a shorter T_{60} than used previously.

For clicks and noise, room volume estimates were larger in experiment 2 when the level cue was absent than in experiment 1 when the level cue was present. It is possible that participants made room volume estimates in experiment 2 by relying primarily on the range of time delays of the echoes, rather than on judged distances, which are generally underestimated. The over-estimation of room size for the speech stimuli in experiment 2 may have been caused by the participants' expectations about the way that the spectral shape of the voice of the talker should change with distance, which may have led to over-estimates of the distance of the farthest sources, as described earlier. Stimulus type may not have had an effect in experiment 1 because the level cue was weighted more highly than the relatively weak reverberation cue and because the level cue for speech varied in a way consistent with a talker speaking with constant vocal effort. A previous study investigating distance discrimination (Kolarik et al 2013a) showed that performance based on level only was better than that based on reverberation only for the same T_{60} as used in this study (400 ms). Thus, participants may rely more on level than reverberation cues when both cues are present.

The current study focussed on room size estimates made when level and/or reverberation cues to distance were available. However, other auditory distance cues are often available in daily life, including spectral and dynamic cues (for reviews, see Kolarik et al 2016a; Zahorik et al 2005). Further work is needed to explore the extent to which these other cues influence room size estimates. The effects of visual loss on acoustic room size estimates also require further study. Despite the potential usefulness of information regarding room dimensions for path planning and navigation by blind people, we are aware of only one study to date that has assessed the effect of blindness on acoustically derived room size estimates (Kolarik et al 2013d). People with full visual loss have been shown to develop improved abilities to extract spatial information from room echoes (Dufour et al 2005; Kolarik et al 2013b), and they might be able to utilize reverberation cues to improve their judgments of room size for shorter reverberation times than for normally sighted people. This has not yet been experimentally tested.

The current experiments focused on factors affecting estimates of room size made on the basis of information provided by sound-producing sources within virtual rooms. Information regarding room size might also be gleaned from self-generated sounds using echolocation, especially for blind individuals (for reviews, see Kolarik et al 2014; Stoffregen and Pittenger 1995; Thaler and Goodale 2016). However, echolocation is restricted in terms of range (Kolarik et al 2016b; Kolarik et al 2017b; Rowan et al 2013; Schenkman and Nilsson 2010), so distance information regarding distant walls might not be obtained. Also, proficiency in the use of echoes from self-generated sounds may require training or experience. The effectiveness of echolocation as a means of obtaining room size information requires further study.

As described above, estimates of the farthest sound source distance can be used as an indicator of the nearest possible distance of the far wall. However, the use of this strategy will result in underestimation of the room size if the far wall is much farther away than the farthest sound source. This can be avoided by having sound sources close to the walls, as was the case in the current experiments. Listening in a room in which multiple cues for azimuth and distance are available might provide information regarding the position of the lateral walls as well as the facing wall, and this might increase the accuracy of room size estimates. We are currently investigating this. Further work is needed to establish the acoustic conditions that result in the most accurate room size judgments.

Further work is also needed to establish how many stimuli have to be presented for room size to be judged consistently. A single stimulus presented at the farthest distance (or indeed at an intermediate distance) would probably be sufficient for the participant to obtain an initial rough approximation of the room size. However, an estimate based on only a single sample is unlikely to be reliable. It is probable that the more samples the participant is able to obtain (the more trials/longer the block), the more reliable the room size estimate will be, as multiple samples can be stored in memory and compared, allowing the estimate to be updated and refined. In addition, there is likely to be some form of adaptation to the acoustic room characteristics, such as reverberation time, that affect distance (Zahorik et al 2005) and room size (Etchemendy et al 2017) estimates. Such a form of adaptation might be disrupted by switching between different rooms from trial to trial, resulting in greater variability of distance and room size estimates. To date, no study has reported room size estimates based on a single sample of sound, whether multiple samples increase the reliability of room size estimates, how many samples are required to get reliable room size estimates, or the effect of switching between rooms on distance and room size estimation. Further experiments are needed to investigate these issues.

Room volume estimates were larger and generally more accurate for the equalized-level condition used in Experiment 2 than in Experiment 1, where both level and reverberation cues were available. One possible explanation for this result is in terms of cue combination and the possible greater accuracy of reverberation than level for room size judgements. To generate an internal representation of room size, it is likely that information from multiple cues is appropriately weighted and combined in a similar way to that proposed for visual (Landy et al 1995) and auditory (Zahorik 2002) distance perception. Level cues are generally more "dominant" than reverberation cues when estimating distance using auditory cues (Zahorik 2002). However, for room volume judgments, reverberation may provide more accurate information than level, especially when the level at the source is variable and uncertain. In Experiment 1, greater perceptual weight may have been assigned to level than to reveberation cues, leading to smaller estimates of room size than for Experiment 2, where reverberation cues only were available. However, if reverberation is more reliable than level for room size judgments, it is unclear why level would be weighted more heavily than reverberation.

Another potential explanation is in terms of a "specific room size tendency" under conditions of reduced spatial information. In Experiment 2 where reduced spatial information was available, participants may have given estimates close to a default room size (or based on default individual room dimensions). A similar effect, known as specific distance tendency, has been postulated for distance judgments (Gogel 1969; Mershon and King 1975). It may be that the default room size in experiment 2 was close to the actual room size, leading to reasonably accurate room volume estimates for clicks and noise (Fig. 4). Future experiments conducted under reduced cue conditions could be utilized to assess whether there is a specific room size tendency and to determine if there are specific default values of room length, width, height and volume.

Lastly, in the current experiments participants estimated room size by reporting length, width and height. It is not known whether alternative measurement methods might result in more accurate estimates. For example, participants might be asked to adjust the image of a virtual room on a computer screen, to select one of many pictures of rooms to match the estimated size, to throw a ball such that it would land at the perceived far wall distance, or to walk to the estimated wall positions or along the perimeter of the room. For distance estimates, previous work has shown good correspondence between verbal and walking responses, with walking responses showing lower between-subject variability (Loomis et al 1998). It is not known whether a similar pattern of responses would occur for room size estimates.

In summary, the results showed that: (1) When both level and reverberation cues were available, participants judged a virtual room with a relatively short reverberation time of 400 ms to be significantly larger than an anechoic room and room-size estimates did not vary significantly with stimulus type; (2) When level cues were not available, a reverberant room was judged to be larger when listening to speech than when listening to noise or clicks.

Acknowledgments

This research was supported by the Vision and Eye Research Institute, Faculty of Health, Education, Medicine and Social Care, Anglia Ruskin University. We thank the Editor Hiroshi Ashida and two anonymous reviewers whose comments helped to strengthen the manuscript.

Declaration of conflicting interests

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Figure and Table captions

Figure 1. Layout of the virtual room. The position of the participant was simulated to be on the midline of the shorter wall. Loudspeakers show the positions of the virtual sound sources, which were presented in front of the participant. The locations of the reference sound sources are shown by white loudspeakers and the location of the probe sound is shown by the black loudspeaker.

Figure 2. Geometric mean room volume estimates for virtual anechoic (open bars) and reverberant (grey bars) rooms in Experiment 1. In the anechoic room, only level cues for distance were available, while in the reverberant room, both level and reverberation cues were available. Error bars show ±1 standard error of the mean. The dashed line represents veridical performance for the reverberant room (room size = 97.2 m^3). The y axis is logarithmic.

Figure 3. Geometric mean room dimension estimates for virtual anechoic (open bars) and reverberant (grey bars) rooms for Experiment 1. Error bars show ±1 standard error of the mean. The dashed lines represent veridical performance for the reverberant room (length = 9 m, width = 5.4 m, height = 2 m). The y axis is logarithmic.

Figure 4. Geometric mean room volume estimates based on equalized-level speech, clicks and noise in Experiment 2 (similar to Figure 2). In this virtual room, only reverberation cues for distance were available. Error bars show ±1 standard error of the mean. The dashed line represents veridical performance (room size = 97.2 m^3). The y axis is logarithmic.

Figure 5. Geometric mean room dimension estimates for Experiment 2. Data for equalized-level speech, clicks and noise are shown by open, grey and diagonal line-filled bars, respectively. Error bars show ±1 standard error of the mean. The dashed lines represent veridical performance (length = 9 m, width = 5.4 m, height = 2 m). The y axis is logarithmic.

Table 1. Correlations between room dimension estimates for Experiment 1. Correlations are shown between length (L), width (W) and height (H) estimates, and between volume (V) estimates and room dimension estimates, for the three stimulus types and the two room reverberation times. In this and subsequent tables, significant differences are indicated by asterisks: *p<0.05, **p< 0.01.

Table 2. Correlations between bisection thresholds/PSE judgements and room volume estimates for Experiment 1.

Table 3. Correlations between room dimension estimates made in Experiment 2. Correlations are shown between length (L), width (W) and height (H) estimates, and between volume (V) estimates and room dimensions, for speech, noise and clicks.

Table 4. Correlations between bisection thresholds/PSE judgements and room volume estimates for Experiment 2.

References

Ahveninen J, Kopco N, Jääskeläinen I P, 2014 "Psychophysics and neuronal bases of sound localization in humans" *Hearing Research* **307** 86-97

Akeroyd M A, Gatehouse S, Blaschke J, 2007 "The detection of differences in the cues to distance by elderly hearing-impaired listeners" *The Journal of the Acoustical Society of America* **121** 1077-1089

Allen J B, Berkley D A, 1979 "Image method for efficiently simulating small-room acoustics" *The Journal of the Acoustical Society of America* **65** 943-950

Anderson P W, Zahorik P, 2014 "Auditory/visual distance estimation: Accuracy and variability" *Frontiers in Psychology* **5** 1097

Ashmead D H, LeRoy D, Odom R D, 1990 "Perception of the relative distances of nearby sound sources" *Attention and Psychophysics* **47** 326-331

Bidart A, Lavandier M, 2016 "Room-induced cues for the perception of virtual auditory distance with stimuli equalized in level" *Acta Acustica united with Acustica* **102** 159-169

British Society of Audiology, 2011 *Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking* (Reading, UK: British Society of Audiology)

Bronkhorst A W, Houtgast T, 1999 "Auditory distance perception in rooms" *Nature* **397** 517-520

Brungart D S, Scott K R, 2001 "The effects of production and presentation level on the auditory distance perception of speech" *The Journal of the Acoustical Society of America* **110** 425-440

Calcagno E R, Abregú E L, Eguía M C, Vergara R, 2012 "The role of vision in auditory distance perception" *Perception* **41** 175-192

Cochran P, Throop J, Simpson W, 1968 "Estimation of distance of a source of sound" *American Journal of Psychology* **81** 198-206

Coleman P D, 1962 "Failure to localize the source distance of an unfamiliar sound" *The Journal of the Acoustical Society of America* **34** 345-346

Coleman P D, 1963 "An analysis of cues to auditory depth perception in free space" *Psychological Bulletin* **60** 302-315

Crukley J, Scollie S, Parsa V, 2011 "An exploration of non-quiet listening at school" *Journal of Educational Audiology* **17** 23-35

Culling J, Lavandier M, Jelfs S, 2013 "Predicting binaural speech intelligibility in architectural acoustics", in *The technology of binaural listening* pp 427-447

Culling J F, 2013 "Energetic and informational masking in a simulated restaurant environment", in *Basic Aspects of Hearing* pp 511-518

Dufour A, Després O, Candas V, 2005 "Enhanced sensitivity to echo cues in blind subjects" *Experimental Brain Research* **165** 515-519

Ellis G M, Zahorik P, 2019 "A dissociation between speech understanding and perceived reverberation" *Hearing Research* **379** 52-58

Etchemendy P E, Abregú E, Calcagno E R, Eguia M C, Vechiatti N, Iasi F, Vergara R O, 2017 "Auditory environmental context affects visual distance perception" *Scientific Reports* **7** 7189

Frenz H, Lappe M, 2005 "Absolute travel distance from optic flow" *Vision Research* **45** 1679-1692

Gamble E A, 1909 "Intensity as a criterion in estimating the distance of sounds" *Psychological Review* **16** 415-426

Gardner M B, 1969 "Distance estimation of 0° or apparent 0°-oriented speech signals in anechoic space" *The Journal of the Acoustical Society of America* **45** 47-53

Gardner W G, Martin K D, 1995 "HRTF measurements of a KEMAR" *The Journal of the Acoustical Society of America* **97** 3907-3908

Gogel W C, 1969 "The sensing of retinal size" *Vision research* **9** 1079-1094

Gori M, Sandini G, Martinoli C, Burr D C, 2014 "Impairment of auditory spatial localization in congenitally blind human subjects" *Brain* **137** 288–293

Gotoh T, Kimura Y, Kurahashi A, Yamada A, 1977 "A consideration of distance perception in binaural hearing" *Journal of the Acoustical Society of Japan* **33** 667-671

Kolarik A J, Cirstea S, Pardhan S, 2013a "Discrimination of virtual auditory distance using level and direct-to-reverberant ratio cues" *The Journal of the Acoustical Society of America* **134** 3395-3398

Kolarik A J, Cirstea S, Pardhan S, 2013b "Evidence for enhanced discrimination of virtual auditory distance among blind listeners using level and direct-to-reverberant cues" *Experimental Brain Research* **224** 623-633

Kolarik A J, Cirstea S, Pardhan S, Moore B C J, 2013c "An assessment of virtual auditory distance judgements among blind and sighted listeners" *Proceedings of Meetings on Acoustics* **19** 050043

Kolarik A J, Cirstea S, Pardhan S, Moore B C J, 2014 "A summary of research investigating echolocation abilities of blind and sighted humans" *Hearing Research* **310** 60-68

Kolarik A J, Moore B C J, Zahorik P, Cirstea S, Pardhan S, 2016a "Auditory distance perception in humans: A review of cues, development, neuronal bases and effects of sensory loss" *Attention, Perception, & Psychophysics* **78** 373-395

Kolarik A J, Pardhan S, Cirstea S, Moore B C J, 2013d "Using acoustic information to perceive room size: Effects of blindness, room reverberation time, and stimulus" *Perception* **42** 985-990

Kolarik A J, Pardhan S, Cirstea S, Moore B C J, 2017a "Auditory spatial representations of the world are compressed in blind humans" *Experimental Brain Research* **235** 597-606

Kolarik A J, Raman R, Moore B C J, Cirstea S, Gopalakrishnan S, Pardhan S, 2020 "The accuracy of auditory spatial judgments in the visually impaired is dependent on sound source distance" *Scientific Reports* **10** 7169

Kolarik A J, Scarfe A C, Moore B C J, Pardhan S, 2016b "An assessment of auditory-guided locomotion in an obstacle circumvention task" *Experimental Brain Research* **234** 1725-1735

Kolarik A J, Scarfe A C, Moore B C J, Pardhan S, 2017b "Blindness enhances auditory obstacle circumvention: Assessing echolocation, sensory substitution, and visual-based navigation" *PLOS One* **12** e0175750

Landy M S, Maloney L T, Johnston E B, Young M, 1995 "Measurement and modeling of depth cue combination: In defense of weak fusion" *Vision Research* **35** 389-412

Lehmann E A, Johansson A M, 2008 "Prediction of energy decay in room impulse responses simulated with an image-source model" *The Journal of the Acoustical Society of America* **124** 269-277

Loomis J M, Klatzky R L, Philbeck J W, Golledge R G, 1998 "Assessing auditory distance perception using perceptually directed action" *Perception & Psychophysics* **60** 966-980

Mershon D H, Ballenger W L, Little A D, McMurtry P L, Buchanan J L, 1989 "Effects of room reflectance and background noise on perceived auditory distance" *Perception* **18** 403-416

Mershon D H, Bowers J N, 1979 "Absolute and relative cues for the auditory perception of egocentric distance" *Perception* **8** 311-322

Mershon D H, King L E, 1975 "Intensity and reverberation as factors in the auditory perception of egocentric distance" *Attention, Perception, & Psychophysics* **18** 409-415

Moore B C J, 2012 *An Introduction to the Psychology of Hearing.* (Leiden, The Netherlands: Brill)

Moore B C J, Kolarik A, Stone M A, Lee Y-W, 2016 "Evaluation of a method for enhancing interaural level differences at low frequencies" *The Journal of the Acoustical Society of America* **140** 2817-2828

Nábělek A, Nábělek L, 1994 "Room acoustics and speech perception", in *Handbook of clinical audiology* Ed J Katz pp 624–637

Nielsen S H, 1993 "Auditory distance perception in different rooms" *Journal of the Audio Engineering Society* **41** 755-770

Pearsons K, Bennett R, Fidell S, 1976 "Speech levels in various environments. Report to the Office of Resources & Development" *Environmental Protection Agency, BBN Report* **3281**

Prud'homme L, Lavandier M, 2020 "Do we need two ears to perceive the distance of a virtual frontal sound source?" *The Journal of the Acoustical Society of America* **148** 1614-1623

Rowan D, Papadopoulos T, Edwards D, Holmes H, Hollingdale A, Evans L, Allen R, 2013 "Identification of the lateral position of a virtual object based on echoes by humans" *Hearing Research* **300** 56-65

Schenkman B N, Nilsson M E, 2010 "Human echolocation: Blind and sighted persons’ ability to detect sounds recorded in the presence of a reflecting object" *Perception* **39** 483-501

Smaldino J, Crandell C, Brian M, Kreisman A, Kreisman N, 2008 "Room acoustics for listeners with normal hearing and hearing impairment", in *Audiology Treatment* Eds M Valente, H Hosford-Dunn and R J Roeser (New York, NY: New York, NY) pp 418-451

Stoffregen T A, Pittenger J B, 1995 "Human echolocation as a basic form of perception and action" *Ecological Psychology* **7** 181-216

Strybel T Z, Perrott D R, 1984 "Discrimination of relative distance in the auditory modality: The success and failure of the loudness discrimination hypothesis" *The Journal of the Acoustical Society of America* **76** 318-320

Thaler L, Goodale M A, 2016 "Echolocation in humans: An overview" *Wiley Interdisciplinary Reviews: Cognitive Science* **7** 382-393

Tonelli A, Brayda L, Gori M, 2015 "Task-dependent calibration of auditory spatial perception through environmental visual observation" *Frontiers in Systems Neuroscience* **9** 84

Turano K A, Yu D, Hao L, Hicks J C, 2005 "Optic-flow and egocentric-direction strategies in walking: Central vs peripheral visual field" *Vision Research* **45** 3117-3132

Vercillo T, Burr D, Gori M, 2016 "Early visual deprivation severely compromises the auditory sense of space in congenitally blind children" *Developmental Psychology* **52** 847–853

Vercillo T, Milne J L, Gori M, Goodale M A, 2015 "Enhanced auditory spatial localization in blind echolocators" *Neuropsychologia* **67** 35-40

Vercillo T, Tonelli A, Gori M, 2018 "Early visual deprivation prompts the use of body-centered frames of reference for auditory localization" *Cognition* **170** 263-269

von Békésy G, 1949 "The moon illusion and similar auditory phenomena" *American Journal of Psychology* **62** 540-552

Zahorik P, 2002 "Assessing auditory distance perception using virtual acoustics" *The Journal of the Acoustical Society of America* **111** 1832-1846

Zahorik P, Brungart D S, Bronkhorst A W, 2005 "Auditory distance perception in humans: A summary of past and present research" *Acta Acustica united with Acustica* **91** 409-420