AUDITORY ROOM SIZE PERCEIVED FROM A ROOM ACOUSTIC SIMULATION WITH AUTOPHONIC STIMULI

Manuj Yadav, Densil Cabrera and William. L. Martens
Faculty of Architecture, Design and Planning, The University of Sydney, Australia
manuj.yadav@sydney.edu.au, densil.cabrera@sydney.edu.au, william.martens@sydney.edu.au

By listening to the sound of their own voice in a room, a talking-listener receives useful information about the acoustical characteristics of the enclosed environment. The information they receive about a specific acoustical characteristic is generally supplemented by other sensory, especially visual, stimuli that can influence one's perception of (and in) these environments. One such characteristic is the size of the room perceived through the human auditory apparatus, which can be different from the room's physical size, as well as the visually perceived room size. This paper examines the relationship between judgements of the size of a room environment that is based on auditory stimuli, and relevant room acoustic parameters; where these judgements may contrast with the objective size as indicated by room acoustic theory. The room size judgements were collected from a study conducted in an auditory mixed-reality environment, in which a talking-listener can perceive the sound of his/her own voice in the simulated reverberant conditions of real rooms, while physically being in an anechoic room. In this study, human participants performed talking tasks, and rated the aurally perceived size of each room. The results indicate that the level of the acoustical support provided by the room's environment (quantified here as room gain) accounts for more of the variance in the associated room size judgements than any other predictor.

INTRODUCTION

The size of a room is one of its most basic attributes, and this preliminary study examines the perception of room size using sound alone. Although it can be argued that the most reliable judgement of room size can be arrived at from visual inspection, it is also possible to judge the size of a room using auditory stimuli, without accompanying visual stimulus [1-5]. This involves exciting the room with an appropriate sound source and hearing the characteristics of the acoustic reflections from the walls, furnishings etc. Experimental studies eliciting auditory room size judgements can provide insight to space perception processes of people with a significant visual impairment [1]; contribute to the understanding of reverberance in concert halls [6]; and extend the understanding of psychoacoustics relating to autophonic output [7] (one's own voice) in rooms [8, 9].

In listening to the sound of a room, the sound source can be the listener him/her-self (egocentric stimulus) or there can be a sound source physically distinct from the listener (exocentric stimulus). The scenarios arising from these exocentric and egocentric stimuli constitute exocentric and egocentric tasks, respectively. Previously, in mostly exocentric tasks, auditory room size perception has been shown to be more strongly affected by acoustical parameters (specifically the room's reverberation time, source-receiver distance, interaural cross-correlation and clarity index) than the room's physical volume [10].

This paper investigates auditory room size perception in an egocentric task, based on an auditory mixed-reality (MR) environment, a term consistent with the framework suggested by Milgram and Colquhoun [11] for visual MR, as explained in the following section.

MIXED-REALITY EXPERIMENTAL PROCEDURE

To avoid the complications implicit in conducting *in situ* experiments using different rooms with human participants, the experiment described by this paper employed real-time virtual room acoustic simulations applied to autophonic output, in order to render a MR auditory environment for each of the rooms tested, which has been described elsewhere by the authors [12]. The stages involved in stimulus preparation are briefly described in the following two sections. The subjective test is also described in what follows.

Measurement and processing of room impulse responses

Binaural impulse responses from the mouth to two ears of a head and torso simulator (HATS, Brüel & Kjær Type 4128C) were acquired at positions in six real rooms. In each room, successive measurements were made over a rotational range of -60° to +60° in yaw (by rotating the HATS at 2° increments), in a process described in detail by Cabrera et al. [13]. The measured oral-binaural room impulse responses (OBRIRs) were truncated by removing the first 7.6 ms (comprising the direct sound and first-order floor reflection), for the reasons identified in the next section. The truncated OBRIRs were then subjected to a MATLAB routine to suppress any noise in their tail, by multiplying the noise floor by an exponential decay function that matched the initial noise-free decay rate within each octave band.

As a reliability check, one of the rooms was measured in two conditions, differing only by the presence of a small curtain near the measurement position, leading to a slight change in the acoustical parameters. These two conditions of the same room were included in the current experiment to test the variation in the room size judgements of essentially the same room, leading to a total of seven simulated room conditions.

Real-time room acoustic simulation system

The measured OBRIRs (as described in the previous section) were accessed by a real-time convolver (SIR2 VST plugin), hosted in a Max/MSP patch running on a Windows platform. The Max/MSP patch allowed rooms to be switched in real-time from a selection menu, which would load the corresponding OBRIRs for convolution (with no apparent delay). The AD/DA converter used was a RME ADI-8 QS unit with 48 kHz sampling rate and 32-bit quantization in a 1-in/2-out configuration. The electroacoustic latency of this system was 7.6 ms, which effectively becomes 0 ms as a result of the OBRIR truncation described in the previous section. Essentially, as the truncated OBRIRs contain no samples corresponding to the direct sound and first floor reflection, the system's output smoothly follows the direct sound and the floor reflection; neither of which is simulated because the direct sound is already present with the talking-listener's voice and the floor reflections are provided by adding a carpeted wooden floor to the anechoic room used in the experiment.

The headset microphone used for vocal input was a DPA 4066 and the ear-loudspeakers used for playing back the convolved output (the room reflections corresponding to the current OBRIR) to a talking-listener were a pair of AKG K1000 (loudspeakers near the ears, without any circumaural cushion or contact with the ears). The receiver unit of the headtracker was attached onto the strap of the ear-loudspeakers.

The headset microphone was positioned at a distance of 7 cm from the centre of lips on the right side of the face. This was done to eliminate the detrimental effects associated with plosives and fricatives when the microphone is placed in the direct air-stream from the mouth opening. A similar microphone position has been used in a recent study for egocentric sound in rooms [9]. The simulation system gain was calibrated by measuring its response (with a loaded OBRIR) using a HATS, and gain-adjusting the system response so that it matched the original OBRIR.

The presence or absence of the ear-loudspeakers had a negligible effect on the octave-band oral-binaural gains for microphones (Brüel & Kjær 4101 Binaural Microphone) placed at the entrance of each ear canal for five participants talking (measured separately), and a HATS (Brüel & Kjær 4128C) emitting pink noise [12]. The feedback from the loudspeakers to the headset microphone was also negligible (loop gain < -16 dB) [14].

The simulation system's headtracking (implemented in the Max/MSP patch), follows the yaw angle of the talking-listener's head, ranging from -40° to +40° (i.e., much, but not all, of the measured OBRIR yaw range), and continually selects the OBRIR to be convolved with the current vocal *input*; while the real-time convolution system *outputs* two channels of convolved audio that includes the *output* from the current head position combined with the residual audio generated for any other previous head positions (which may still be following a reverberant decay). This provides an auditory scene that is almost the same as the one that would be produced by vocal transduction in the measured real room for similar head movements.

Subjective room size judgements

Room size judgements were made by 8 participants (ages 23-45; 7 male, 1 female; 4 acoustically knowledgeable and 4 acoustically naïve university students), who were seated on a wooden chair placed on a carpeted floor in an anechoic chamber (with a large wooden board underneath the carpet, as described in the previous section). They were given a few sheets of printed text with the choice that they were free to either read from the text or to use any other speech or vocalisation that would enable them to judge the size of the simulated room, with typical or more exploratory head movements. The participants were tested in the seven room simulations according to a random order, with two trials per room: one with headtracking turned on and the other with headtracking turned off. They gave a room size rating for each trial using a numerical scale ranging from 1 (the size of the anechoic room in which the talking-listeners were physically present) to 10. This scale was merely conceived of as a simple vernacular scale, rather than a precise ratio scale.

DATA PROCESSING

The room size judgements of each participant were centred (by dividing each rating by their mean rating) so that the participants would have equal weight in the analysis of combined results. Following centring, the full set of results has a mean value of 1, and a standard deviation of 0.33. As the room size judgements did not differ significantly between the headtracked and non-headtracked trials, the mean value of these two trails per participant was used in the following analysis. Room size judgements were examined in relation to measures of physical room size (volume, V) and to acoustical parameters derived from the OBRIRs. The acoustic parameters include the following: mid-frequency (500 Hz) reverberation time (RT) with an evaluation range from -10 dB to -30 dB (amended from the more commonly used -5 dB to -25 dB range, to account for the higher gain of the direct sound); room gain (G_{RG}) derived from the amended procedure outlined by Pelegrín-Garcia [15], which was first proposed by Brunskog et al. [16] as a measure of the energy of the room-reflected sound that the talking-listener hears (power average of the two ears, expressed in dB); clarity index (C_{50}) [1,8,10]; and interaural cross-correlation ($IACC_{early}$) [10], using 80 ms as the boundary between early and late. One distinction in the calculation of the room gain values here from the procedure described by Pelegrín-Garcia [15] was the duration of direct sound, which in the current paper was taken as 7.6 ms and corresponded to the duration of the direct sound and first floor reflection of the OBRIR. In the case of the room gain, the values presented here corresponded to the energy summed over the entire duration of the 0° OBRIR starting from 8 ms. The RT, C_{50} and $IACC$ values were the octave band mean values over the entire headtracking range of -40° to +40° yaw as described in Cabrera et al. [17]. Early decay time was not calculated because it is not well-defined for a source very close to a receiver V_{est} is a quasi-acoustical parameter calculated from an empirical function relating room volume to reverberation time ($RT \approx 0.26 \ln(V) - 0.75$) that was derived by Shabtai et al. [18]. The subjective room size ratings and physical parameters are shown in Table 1.

ANALYSIS AND DISCUSSION

To determine whether there was a variation in the rated values of room size with different room conditions, a one-way ANOVA was conducted, and the result indicates a significant effect ($F(6, 49) = 24.63, p < 0.01$). As the near-identical stimuli (rooms 4 and 5) received very similar size ratings (Table 1), this suggests that the participants were consistent in judging the size of the same room in two slightly different conditions. Condition 5 was excluded from further statistical analysis, as its subjective ratings and objective parameters were so similar to condition 4. Also, condition 7, which represented the autophonic perception in a large reverberant room environment (a recital hall) was identified as an outlier and consequently not included in further statistical analysis. In the following analysis, room number 6 in Table 1 will be referred to as room number 5.

Following these changes in the data, correlation analysis showed that none of the parameters are significantly correlated with the physical room volume ($p < 0.05$). However, considering these non-significant correlations for their polarity, a negative sign of the correlation coefficient (r) indicates that as the room volume increases, G_{RG} decreases ($R = -0.31$, $p = 0.30$), and *vice-versa*; whereas a positive sign indicates that as the room volume increases, C_{50} ($R = 0.49$, $p = 0.20$) and $IACC$ increase ($R = 0.59$, $p = 0.14$), and *vice-versa*. These signs are at least partly consistent with expectations from room acoustics theory in that a greater diffuse field strength is expected in smaller rooms (leading to increased G_{RG}, and reduced $IACC$); and the expected relationship between C_{50} and room size is more subtle (see [19]). As an important design feature in this study, it is noteworthy that there is no correlation between reverberation time and room volume for the selection of rooms ($R = 0.01$), although a positive correlation might be expected for a wider selection of rooms (as represented by V_{est}, following [18]).

On the other hand, the room size judgements are significantly correlated with all the parameters that are listed in Table 1, except the room's physical volume and $IACC_{early}$. Figure 1 shows the linear regression model (R^2=0.99, F=220.6, p<0.001) that was yielded by room gain as the independent (predictor) variable, which can be expressed as

$$\text{Predicted room size} = 0.17 + 0.68\ G_{RG} \qquad (1)$$

Compared to G_{RG}, the linear regression models using RT (R^2=0.76, F=13.73, p<0.05), C_{50} (R^2=0.68, F=9.51, p=0.05), and V_{est} (R^2=0.84, F=22.68, p<0.05) as the predictors accounted for lesser variance in the room size judgement values and lower F values.

In recent research, higher room gain values have been shown to be important in providing greater vocal comfort and lesser vocal effort for talking-listeners, and vice-versa [16, 20]. The results of the present research are consistent with these findings, with respect to a negative correlation of physical room volume with room gain, as the strength of the reverberant field in a smaller room is generally higher than bigger rooms. Hence, from an objective perspective, room gain values could serve as an important component in the prediction of the room's size. However, the *positive* correlation of the subjective room size responses with the room gain values, modelled in equation (1) is interesting, as it points towards a conjecture that the strength of the reverberant field in the current experiments was used as an indicator of its reverberance (and that greater reverberance was interpreted as an indicator of greater room size). This conjecture is partly based on the post-experiment interview with the participants, who reported using the reverberation of the rooms as an indicator of their size. Note that the effectiveness of room gain as a predictor in the

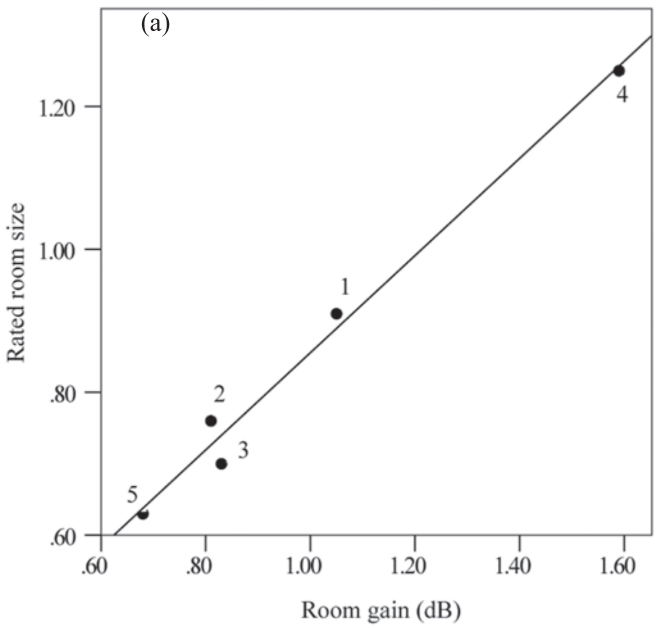
Table 1. The data used for the statistical analysis. The rooms are numbered from 1-7 with a bracketed number showing their index in the paper by Cabrera et al. [17], which characterised the rooms used in this paper in detail. The next columns consecutively show the mean rated room sizes; volumes; mid-frequency reverberation times; early room gains; clarity index; early $IACC$ values; estimated volumes from the linear regression model described by Shabtai et al. [18] Rooms 4 and 5 were the same room measured in two slightly different conditions, but only room 4 is characterized in Cabrera et al. [17]

Room	Rated Size	V(m^3)	RT(s)	G_{RG}(dB)	C_{50}(dB)	$IACC_{early}$	V_{est}(m^3)
1 (3)	0.91	125	0.60	1.05	11.8	0.25	179
2 (6)	0.76	152	0.35	0.81	18.3	0.26	68
3 (7)	0.70	170	0.40	0.83	20.7	0.21	83
4 (8)	1.25	188	0.90	1.59	11.6	0.21	570
5	1.27	188	0.90	1.54	12.5	0.23	570
6 (10)	0.63	310	0.50	0.68	20.5	0.54	122
7 (11)	1.48	7650	1.70	0.29	31.6	0.54	12370

present study might be influenced by the zero correlation between reverberation time and room volume.

Future research should focus on studying the interaction between the strength and temporal aspects of reverberant sound fields with respect to auditory room size judgements, where these two parameters are manipulated within rooms of fixed volumes. Similar to the present study, where the reverberation times of the rooms were uncorrelated with their volumes, various levels of correlation between these parameters may be included as a design feature.

As the room size judgements from the headtracked and non-headtracked trials were not significantly different, it poses a question regarding the usefulness of headtracking in a simulation based room size perception tasks. In a recent study using the same room acoustical simulation system, it was shown that headtracking was detectable by five participants in an ABX task, where the threshold for correct detection was set to be just above chance (0.6) [21]. A study with more participants would be required to address the issue of incorporating headtracking in the present simulation for room size perception task (and perhaps similar tasks).



(b)	*B*	*SE B*	*β*
Constant	0.17	0.05	
Room gain	0.68	0.05	0.99

Figure 1. (a) The room size judgements by the participants, as a function of room gains Table 1, where the rooms are numbered 1-5 in an ascending order of their physical volume. Room number 5 corresponds to room number 6 in Table 1, due to the changes explained in the beginning of the current section. (b) The regression model for predicting perceived room size from room gain (G_{RG}). *B* and *SE B* represent the unstandardised coefficients and their standard error, respectively. β represents the standardized coefficient which gives the number of standard deviations the outcome (predicted room size) will change as a result of one standard deviation in the predictor (G_{RG}).

There is also scope for improving the experimental design of the current study, by including simulated room conditions with a more uniform scale in terms of their physical size and variety in terms of their purpose (e.g., residential rooms). A method more robust than magnitude estimation (e.g., paired-comparison, or photograph-matching [22]) could be employed to validate the findings of this study.

CONCLUSIONS

The work described in this paper shows something of the potential of a real-time simulation system for autophonic room acoustics studies involving human participants. The findings of the experiment point to a possible difference between the perception of room size and physical acoustic correlates of room volume, which raises questions for future study.

REFERENCES

[1] C.B. Pop and D. Cabrera, "Auditory room size perception for real rooms", *Proceedings of Acoustics 2005*, Busselton, Western Australia, November 2005

[2] R. McGrath, T. Waldmann and M. Fernström, "Listening to rooms and objects", *Proceedings of the 16th Audio Engineering Society International Conference*, Rovaniemi, Finland, April 1999

[3] S. Hameed, J. Pakarinen, K. Valde and V. Pulkki, "Psychoacoustic Cues in Room Size Perception", *Proceedings of the 116th Audio Engineering Society Convention*, Berlin, Germany, May 2004, paper no. 6084

[4] D.H. Mershon and L.E. King, "Intensity and reverberation as factors in the auditory perception of egocentric distance", *Perception & Psychophysics* **18**(6), 409-415 (1975)

[5] A. Tajadura-Jiménez, P. Larsson, A. Väljamäe, D. Västfjäll and M. Kleiner, "When room size matters: Acoustic influences on emotional responses to sounds", *Emotion* **10**(3), 416-422 (2010)

[6] T. Lokki, J. Pätynen, A. Kuusinen, H. Vertanen and S. Tervo, "Concert hall acoustics assessment with individually elicited attributes", *Journal of the Acoustical Society of America* **130**(2), 835-849 (2011)

[7] H. Lane, A. Catania and S. Stevens, "Voice level: Autophonic scale, perceived loudness, and effects of sidetone", *Journal of the Acoustical Society of America* **33**, 160-167 (1961)

[8] D. Cabrera, D. Jeong, H. J. Kwak and J-Y. Kim, "Auditory room size perception for modeled and measured rooms", *Proceedings of Inter-Noise 2005*, Rio de Janeiro, Brazil, August 2005

[9] D. Pelegrín-García, O. Fuentes-Mendizábal, J. Brunskog and C-H Jeong, "Equal autophonic level curves under different room acoustics conditions", *Journal of the Acoustical Society of America* **130**(1), 228-238 (2011)

[10] D. Cabrera, "Control of perceived room size using simple binaural technology", *Proceedings of the 13th International Conference on Auditory Display*, Montréal, Canada, June 2007

[11] P. Milgram and H. Colquhoun, "A taxonomy of real and virtual world display integration", in *Mixed Reality - Merging Real and Virtual Worlds*, edited by Y. Ohta and H. Tamura, Berlin, Springer-Verlag, pp. 1-16 (1999)

[12] M. Yadav, D. Cabrera and W.L. Martens, "A system for simulating room acoustical environments for one's own voice", *Applied Acoustics* doi: 10.1016/j.apacoust.2011.10.001 (2011)

[13] D. Cabrera, H. Sato, W.L. Martens and D. Lee, "Binaural measurement and simulation of the room acoustical response from a person's mouth to their ears", *Acoustics Australia* **37**, 98-103 (2009)

[14] M. Yadav, *A system for room acoustic simulation for one's own voice*, http://hdl.handle.net/2123/7889, Master of Philosophy, The University of Sydney, 2011

[15] D. Pelegrín-García, "Comment on "Increase in voice level and speaker comfort in lecture rooms" [J. Acoust. Soc. Am. 125, 2072–2082 (2009)]", *Journal of the Acoustical Society of America* **129**, 1161–1164 (2011)

[16] J. Brunskog, A.C. Gade, G.P. Bellester and L.R. Calbo, "Increase in voice level and speaker comfort in lecture rooms", *Journal of the Acoustical Society of America* **125**(4), 2072-2082 (2009)

[17] D. Cabrera, D. Lee, R. Collins, B. Hartmann, W.L. Martens and H. Sato, "Variation in oral-binaural room impulse responses for horizontal rotations of a head and torso simulator", *Building Acoustics* **18**(1), 227-252 (2011)

[18] N.R. Shabtai, Y. Zigel and B. Rafaely. "Room volume classification from room impulse response using statistical pattern recognition and feature selection", *Journal of the Acoustical Society of America* **128**(3), 1155-1162 (2010)

[19] D. Cabrera, "Acoustic clarity and auditory room size perception", *Proceedings of the 14th International Congress on Sound and Vibration*, Cairns, Australia, July 2007

[20] D. Pelegrín-García, B. Smits, J. Brunskog and C-H. Jeong, "Vocal effort with changing talker-to-listener distance in different acoustic environments", *Journal of the Acoustical Society of America* **129**(4), 1981-1990 (2011)

[21] M. Yadav, D. Cabrera, R. Collins and W.L. Martens, "Detection of headtracking in room acoustic simulations for one's own voice", *Proceedings of Acoustics 2011*, Gold Coast, Australia, November 2011

[22] J. Sandvad, "Auditory perception of reverberant surroundings", *Journal of the Acoustical Society of America* **105**(2), 1193 (1999)