Auditory Room Size Perception: A Comparison of Real versus Binaural Sound-fields

Densil Cabrera (1), Claudiu Pop (2)* and Daeup Jeong (3)

(1) Faculty of Architecture, University of Sydney, NSW 2006, Australia
(2) Faculty of Architecture, University of Sydney, NSW 2006, Australia
* Currently with Renzo Tonin & Associates, Level 1, 418a Elizabeth Street, Surry Hills, NSW 2010, Australia
(3) Department of Architectural Engineering, Chonbuk National University, Jeonju, Korea

ABSTRACT

Binaural simulation is an important tool in architectural acoustics auralisation, as well as research. The limits of non-individualised and non-head-tracked binaural reproduction have been examined in numerous studies, especially with regard to auditory localisation. This study forms part of an investigation of whether simple binaural reproduction effectively conveys an impression of room size. For our first experiment, blindfolded subjects were led into different rooms, for which they estimated the size based on the sound of reproduced speech. For the second experiment, a different set of subjects listened to the same acoustic environments through binaural reproductions via headphones, again assessing the room size based on the sound. Results show a greater influence of reverberation for the headphone assessments. While only moderate correlations are achieved between the two experiments, the results of both have clarity index as the primary acoustical correlate, consistent with previous experiments by the authors.

INTRODUCTION

In everyday life, people experience rooms of widely varying size, so it is likely that acoustical indicators of room size will be learnt through experience. Such indicators might include reverberation time (long is associated with large rooms, but also with low absorption), the strength of reverberation (strong is associated with small rooms, but also with low absorption), the perceived distance of sound sources (large distances can only exist within large rooms) and patterns of early reflections (dense and strong reflections are associated with small rooms). These are physical cues, and the question of auditory room size perception is if and how such cues are used, and their relative weighting. While there are few studies of auditory room size perception, they do indicate that auditory cues alone do provide useful information on room size. Sandvad (1999) found that subjects could usually correctly identify photographs of rooms that corresponded to binaurally reproduced sound fields representing those rooms. In subsequent experiments, Sandvad found that some listeners used the direct to reverberant energy ratio as cue for room size estimates, while others used the reverberation time. McGrath *et al.* (1996) found that both sighted (but blindfolded) and blind subjects are able to distinguish small and large rooms using the sound of their own speech and other incidental sounds (in actual rooms). Blind subjects evaluated the room acoustical environment more quickly and accurately than sighted subjects. Studies by Mershon *et al.* (1989), Hameed *et al.* (2004) and Sandvad (1999) indicate that reverberation can have a strong effect on the auditory assessment of room size, but show little evidence of an influence of early reflections.

Studying something like auditory room size perception may be greatly facilitated through the simulation of acoustic environments. Experiments in which subjects listen to real rooms tend to be very time-inefficient – yielding very small amounts of data for a large amount of effort. Simulation allows instant switching between acoustic environments, making for an enormous efficiency gain in data acquisition, as well as being much less demanding of subjects. Therefore, the majority of auditory room size perception studies have been conducted using simulations – and the simulation system used has almost always been binaural presented via headphones, without head-tracking or individualisation of head-related transfer functions. Exceptions to this are the studies of Mershon *et al.* (1989) and McGrath *et al.* (1996), where real rooms were used (either one room, in which acoustic conditions were varied, or two rooms). Another exception is the study of Martignon *et al.* (2005) in which four audio systems were compared – binaural headphones, stereo-dipole, double stereo-dipole and conventional stereophony (O.R.T.F.). Stereo-dipole is a binaural reproduction system using loudspeakers with cross-talk cancellation, described by Kirkeby *et al.* (1998). The study of Martignon *et al.* (which was concerned with room size estimation, distance estimation, and realism rating of concert auditoria) found significant differences between the audio systems, with stereo-dipole (single) yielding the most promising results, and binaural headphones yielding poor results. The binaural headphone system yielded poor estimates of source distance, and low ratings of realism, while the stereo-dipole was the opposite in both these respects. The conventional stereophonic system also performed well in these respects, but subjects were unable to separate the concepts of room size from distance perception with stereophony. Nevertheless, since that study did not have room size ratings in real rooms, it was not known what an ideal response would be. Hence the need for the present study, in which auditory room size perception in real rooms and binaural simulations are compared.

Binaural technology has been adopted widely in architectural acoustics for simulations. This includes auralisation of model rooms to aid in the acoustic design process, or to investigate research questions. Møller (1992) outlines the major issues and techniques pertaining to this type of simulation. While simple binaural reproduction is effective at conveying a somewhat realistic impression of an acoustic environment, it has some well known limitations. One of these is that when headphones are used, the virtual environment becomes locked to the head (i.e., the sound space moves with the listener's head). This means that dynamic localisation cues are much less effective (although they can be present for moving sound sources). The result can be increased front-back confusion errors, as well as a rather vague sensation of space. This is solved through the use of head-tracking (assuming that data are available to make use of head-tracking information). Another limitation is the use of generic head-related transfer functions, as opposed to the listener's own. Again the result is vague and has a problem of false localisation, especially in relation to polar angle (the angle around the inter-aural axis). It is very difficult, if not impossible, to achieve a convincing auditory image directly in front of the listener using simple binaural reproduction via headphones. The situation is somewhat better for loudspeaker-based binaural reproduction, partly because small head movements do not result in major disruptions to the sound-field (for stereo-dipole, as shown by Takeuchi *et al.*(1997)) and head-locking does not occur, and perhaps also because the loudspeaker system provides a visually generated expectation of frontal auditory images.

Most studies of the quality of binaural reproduction are concerned with direction localisation. Some work on auditory distance perception also exists, finding that non-individualised, non-head-tracked binaural reproduction via headphones can be effective for auditory distance perception (Møller *et al.* 1996, Minnaar *et al.* 2001). The authors are not aware of any studies directly assessing the quality of binaural systems in terms of room size perception. One reason for the present study is that the authors have conducted experiments previously on room size perception using binaural headphone reproduction (Cabrera *et al.* 2005, Martignon *et al.* 2005). As mentioned previously, the latter of these experiments raised questions about the quality of this type of reproduction, although there may have been other factors in the experiment method that contributed to the apparently poor result for binaural headphone reproduction.

In the former experiments (Cabrera *et al.* 2005), auditory room size perception was assessed using the method of paired comparisons for binaural simulations of computer modelled rooms (for which physical room size, reverberation time and source-receiver distance were varied), in binaural reproductions of a real room having a fixed volume of 130 m^3 (in which reverberation time and source-receiver distance were varied) and for binaural reproductions of a concert auditorium (the Michael Fowler Centre in Wellington, New Zealand, in which only the listening position was varied). Results showed that clarity index usually exhibited a strong negative correlation to perceived room size. This indicated that subjects may have been assessing room size primarily by comparing the strength of the direct sound and early reflections to the reverberant tail, which could be heard either as a decrease in clarity (eg during speech) or as a full reverberation decay (eg at the end of a speech phrase).

EXPERIMENT 1

This experiment was described previously (Pop and Cabrera 2005), but only incomplete results were presented at the time. The method description is presented again here for convenience, together with results for the complete group of subjects.

Three rooms were prepared for this experiment. Within each room two source positions were chosen, together with one or two listening positions for each source. One of the rooms had adjustable acoustic conditions (curtains along two of the walls), and so the same source and receiver positions were tested for both acoustic conditions. The smallest room (referred to as Room 1) had a volume of 15 m^3, and has mirrors entirely covering the walls (its normal use is as a mirror-chamber sky). It had a mid-frequency reverberation time of 1.0 s, which is very long for such a small room. Room 2 had a volume of 123 m^3, with plasterboard and masonry walls, a concrete ceiling and carpeted floor. The mid-frequency reverberation time was about 0.8 s with all walls exposed, and 0.5 s with the curtains over two walls. Room 3 had a volume of 188 m^3, with a linoleum-covered floor, masonry and plasterboard walls, and a concrete ceiling. A bench ran across the middle of this room, with a curtain around it (normally the room is for photometric measurements). Mid-frequency reverberation time for this room was 1.0 s. Figure 1 shows plans of these rooms.

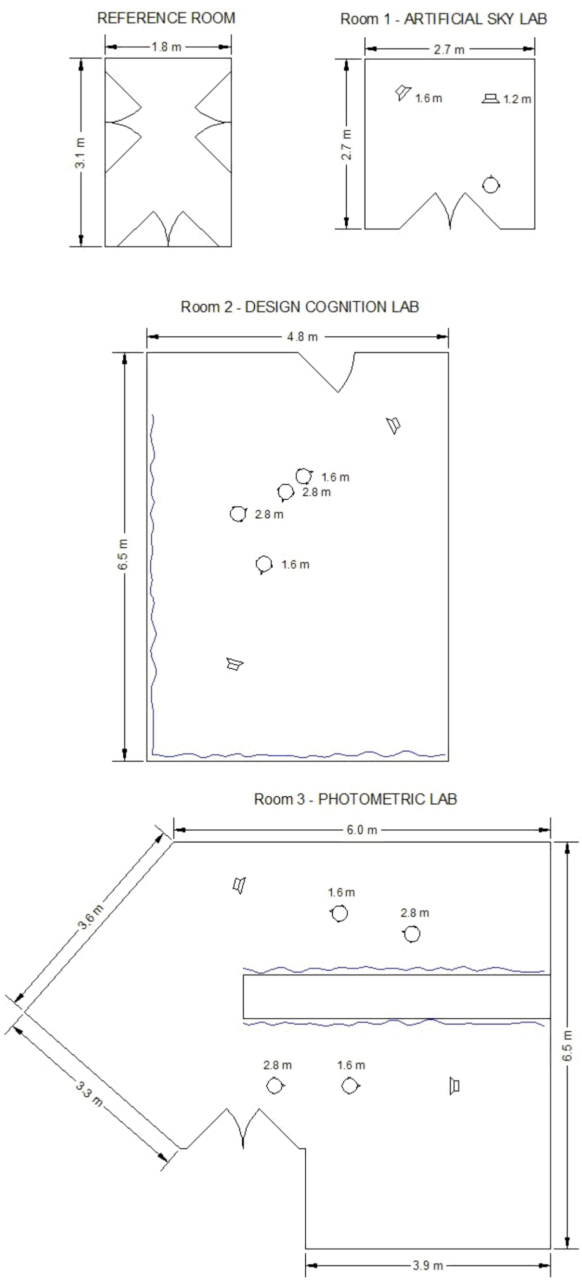

Figure 1. Plans of the reference room and the three test rooms, showing the loudspeaker positions and listening positions.

An anechoic male speech recording of the phrase "I'm speaking from over here" was used in this experiment. Being a calibrated recording, it was possible to reproduce this at approximately the same sound pressure level for a given distance as the original speech (56 dBA at 1 m in free field conditions). More details on the characteristics of this recording are given by Cabrera and Gilfillan (2002). The recording was reproduced using small loudspeakers (Yamaha MSP5A), which are similar in size to a human head. These were supported at a height of 1.2 m on heavy duty microphone stands.

The subject initially talked to the experimenter in the 'reference room', which was a small room of 19 m^3 with a mid-frequency reverberation time of 1.3 s. This room had four walls, each with a door on it, which was intended to give the subject the impression they could be led out any of these doors (although in fact only one of the doors was used). Prior to leaving this room, the subject was blindfolded and earmuffs were worn. Then the subject was led (sometimes circuitously) to one of the test rooms, and seated in a chair at one of the listening positions. The experimenter would then leave the test room and the subject would remove their earmuffs. The speech phrase was then emitted from the appropriate loudspeaker eight times. After this, the subject would put the earmuffs back on and be led back to the reference room, where earmuffs and blindfold would be removed. At this point the subject would be asked to rate the size of the room that they heard, given that the reference room had a size of 10. For example, if the subject's thought the test room was twice the size of the reference room, they would reply '20'. One problem with rating room size is that different types of units can lead to very different responses (eg linear units, square units, or cubic units). Furthermore, most people find it very difficult to relate cubic units to an experience of a room. Therefore the subjects were instructed to not attempt to use physical units in their ratings, but instead to respond using intuitive units of room size, the meaning of which were not defined in any way.

Thirty subjects participated in this experiment. The stimulus order was random and different for each subject. For the first eight subjects, the wrong loudspeaker was activated in Room 1, meaning that the sound was to the side of the direction that the subject was facing. For the remaining subjects this problem was fixed. Only 14 data points were collected from each subject, over a period of about 90 minutes each.

Acoustic measurements were made of each listening situation, using an omnidirectional measurement microphone and a dummy head (B&K 4128C). These included impulse responses using fixed system gain and calibrated recordings of the speech stimuli.

EXPERIMENT 2

The aim of this experiment was simply to emulate Experiment 1 using binaural headphone reproduction.

The headphone stimuli for this experiment were prepared by convolving the original anechoic speech recording with binaural impulse responses that had been recorded in each of the situations of Experiment 1. Since these impulse responses had been recorded with constant system gain, the relative sound levels characteristic of each situation were maintained. Convolution with impulse responses was preferred to direct recordings of the stimuli because it eliminated any variable background noise from the binaural stimuli. Impulse responses were also measured between the headphones (Sennheiser HD600) and the dummy head microphones, and an inverse filter was derived from this, then applied to the stimuli. Once convolved with the anechoic speech, the reproduction system gain was set such that the same sound pressure level was received by the dummy head microphones using headphones as had been measured in the real situations. In addition to the listening situations of Experiment 1, a reference stimulus was prepared, using an impulse response from the reference room convolved with the anechoic speech phrase. Due to the problem with lateral stimuli in Room 1 for the first eight subjects of Experiment 1, both lateral and frontal stimuli were presented for this room in Experiment 2, making a total of 16 stimuli to assess (not including the reference stimulus).

Computers were used for the binaural experiment, with stimuli presented to the subjects using Microsoft PowerPoint. On each slide, the subject could play the reference stimulus and the stimulus to be assessed. The task was the same as Experiment 1 – i.e. to intuitively rate the room size given that the size of the reference room was 10. However, in this case the subject was asked to listen to all stimuli prior to making any assessment, and then to work through the experiment, and finally to check their results by listening again. One difference in the stimuli was that the speech phrase was only played once (rather than eight times) when the subject clicked on its on-screen icon – although the subject was free to play the stimulus as many times as they wished. The slide order was random, and different for each subject, and 30 subjects participated. The experiment typically took 10 to 15 minutes for a subject to complete. Subjects were tested in various office environments in Sydney and Jeonju (Korea), and the subject group was different to that of Experiment 1.

RESULTS

Each subject's raw responses were scaled by dividing each response by that subject's mean response. This scaling was done to focus on the ratios of room size ratings of stimuli, rather than values relative to the reference room (for which the subjects' results diverged widely). However, it is worth mentioning that the raw results for the real room experiment were much larger than for the headphone experiment: the overall mean rating was 33 for Experiment 1, and 9.8 for Experiment 2. The reason for this is not simple to interpret because the subjects perceived the reference room interactively and with all their senses in the real room experiment, but only through listening (non-interactively) in the headphone experiment. However, one factor that is likely to have influenced this is that the reference room was small but reverberant, and the smallness of this room may not have been so obvious when only experienced through sound.

Parallel statistical analyses of the two experiments were done, using factorial analysis of variance (ANOVA) to examine the significance and strength of independent variable effects, and the Tuckey-Kramer *post hoc* test (Tuckey's HSD) to test for the significance of mean differences between states of independent variables.

Mean results in terms of the three rooms (with the medium room split between its two acoustic conditions) are shown in Figure 2. For the real room experiment, there is a clear correspondence between the physical room size and the perceived room size. Furthermore, the change in reverberation time in the medium room causes a change in perceived room size – longer reverberation is associated with greater size ratings. For the real rooms, mean differences between all mean ratings shown in Figure 2 are significant ($p \leq 0.05$). These results contrast with the headphone experiment ratings, for which the smallest room receives the largest room size rating, and the largest room receives a rating that is not significantly different to the medium room without curtains. While, the effect of curtains appears to be maintained in the headphone experiment, the associated mean difference is non-significant.

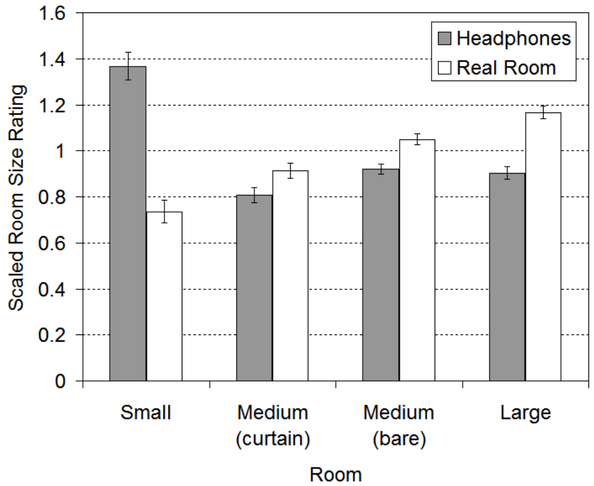

Figure 2. Mean scaled ratings of room size of the four room conditions for the two experiments (±1 standard error).

If the small room (in which only small source-receiver distances were possible) is excluded from the analysis, the remaining stimuli can be analysed through ANOVA in terms of the independent variables of room condition (three states of: medium room with curtain; without curtain; and large room) and source-receiver distance (two states of 1.6 m and 2.8 m). For the real room experiment, effects are significant and strong for both independent variables (for room condition, F=21, p<0.0001; for source-receiver distance F=27, p<0.0001), and all mean results are significantly different in the *post hoc* test. For the headphone experiment the effect of room condition is significant but much weaker (F=4.4, p=0.01), and the effect of distance is on the threshold of significance (F=3.7, p=0.056). As was noted previously, the mean difference between ratings for the medium room without curtain and the large room is not significant, but the other two mean differences between room conditions are. Mean results for this analysis are shown in Figure 3.

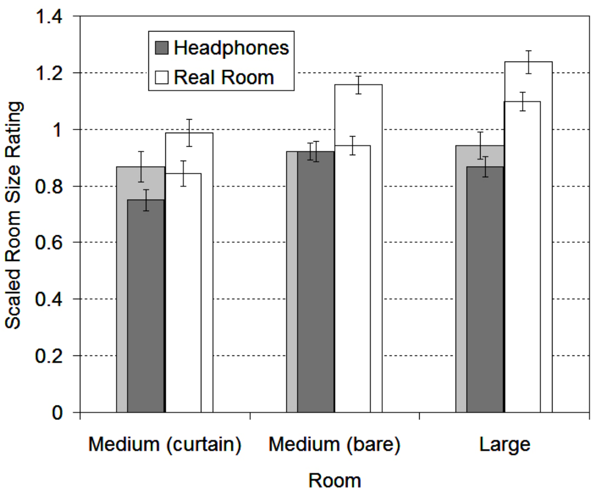

Figure 3. Mean scaled ratings of room size of the medium and large conditions for the two experiments (±1 standard error). Responses for the 1.6 m (narrow bars) and 2.4 m (wide bars) source-receiver positions are superimposed on the chart.

A simple way to examine similarities between the two experiments is to plot the ratings of one against the other, as is done in Figure 4. As is already evident in Figure 2, ratings of the small room received contrasting results in the two experiments, with it being rated small in the real room experiment, and large in the headphone experiment. However, if the small room is discarded, a small positive correlation exists between the two experiments (r=0.71), indicating that 50% of variance is accounted for by this relationship. Another point of interest is that mean ratings in the real room were spread over a wider range than those for headphone stimuli.

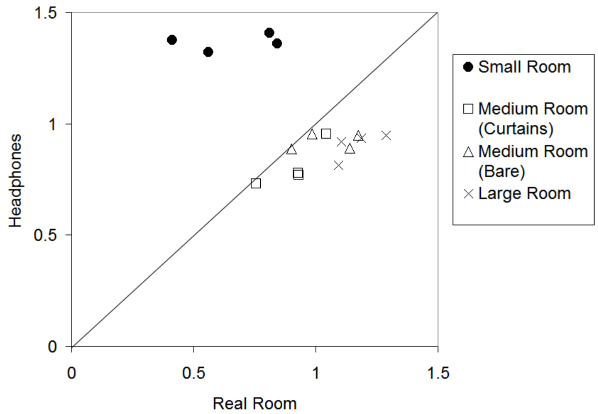

Figure 4. Mean scaled ratings of room size for the two experiments. The four data points for the small room include two with lateral presentation (rather than frontal presentation).

Acoustical parameters were measured in octave bands, but combined into double octave bands of low (125 Hz and 250 Hz), mid (500 Hz and 1 kHz) and high frequency (2 kHz and 4 kHz). These included the sound pressure level of the stimulus, reverberation time, early decay time, centre time, clarity index (C50 and C80), definition, speech transmission index (male), inter-aural cross correlation, inter-aural level difference, and ratios between the frequency bands for the above-mentioned parameters. When correlations with acoustical parameters are examined, clarity index (C80, mid-frequency) provides the best match for both experiments, but for different reasons. For Experiment 1, a high correlation is seen between C80mid and the medium and large room stimuli (leaving the small room stimuli as prominent outliers). On the other hand, for Experiment 2 the small room stimuli reinforce the correlation, which would otherwise be much weaker than that for Experiment 1. These patterns are shown in Figure 5.

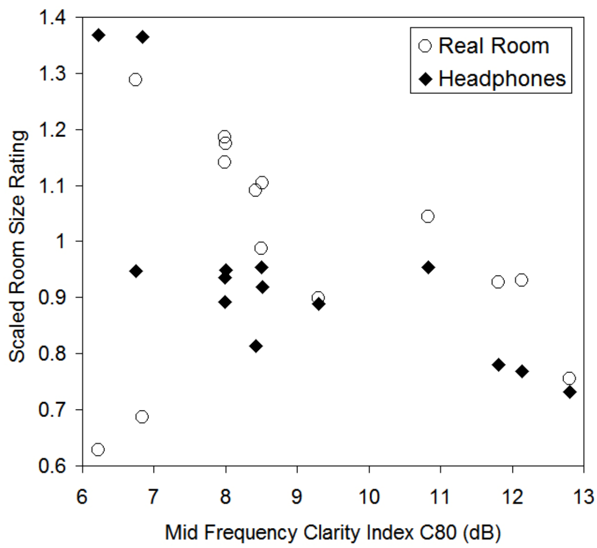

Figure 5. Results for the two experiments plotted against mid-frequency clarity index (C80).

DISCUSSION

The results suggest that simple binaural reproduction, even when performed with calibrated gain and inverse filtering of the reproduction system, does not provide an accurate reproduction for room size perception. In particular, reverberance appears to be interpreted differently for room size judgments in binaural reproduction than for real rooms. Nevertheless, there is some correspondence between the two presentation formats if the extreme situation of the small room with unusually long reverberation time is omitted.

Caution should be applied to the interpretation of the results, because there were many differences between the two experiments. In the real rooms, effort was made to have the subject judge the size simply on the sound of the speech recording. However, in spite of the use of a blindfold and earmuffs (during the walk to and from the listening seat) the subject's experience of each room was much more complex than through the stimulus' sound alone – spatial indicators were also available from the feeling of the floor during the walk, the smell and temperature of the rooms, possible acoustic feedback from self-generated sound, background noises, and so on. By contrast the spatial experience of Experiment 2 involved visual and background noise experiences (of the experiment room) that conflicted with (instead of reinforcing) the represented rooms. Hence, in both experiments, further suppressing cues that were unrelated to the intended stimulus signal might have been helpful in drawing their results closer together.

The reference room itself may provide the most important reason for divergence between the two experiments. This room was quite small but reverberant, giving subjects in Experiment 1 a concrete lesson that reverberance does not necessarily indicate a large room. In Experiment 2, subjects had no indication of the reference room's size other than through listening, and so this lesson was not given. A way of investigating this might be to conduct the binaural experiment again, but in the reference room, asking the subjects to compare a binaurally reproduced recording with the real room that they are in. Alternatively, the real room experiment could be conducted again with a different reference room, or without a reference room.

While the results are not especially encouraging with regard to the use of binaural headphone reproduction for conveying an appropriate impression of room size, the results of both experiments are generally consistent with previous binaural experiments by the authors (Cabrera *et al.* 2005) in that a negative correlation with clarity index provides quite a good predictor of room size ratings. In fact, our previous experiments tended to show correlation patterns with clarity index that were similar to the straight line results of Experiment 1 of the present paper (without the small room outliers). Together these experimental results reinforce the concept of clarity index as a predictor of auditory room size perception. As was mentioned in the introduction, reverberation time and reverberation level are both indicators of room size, but are confounded with room absorption. Large rooms are associated with long reverberation time and low reverberation level, but low absorption is associated with long reverberation and large reverberation level. Hence this double confound should be resolvable to an extent by combining the information from these two aspects of reverberation. However, clarity index values could be lowered by increasing reverberation time or by increasing the level of the reverberant tail (after 80 ms), so the negative correlation between room size perception and clarity index suggests a greater role for reverberation time than reverberation level. Perhaps this competition between reverberation time and level cues is exemplified by the small room in this study, which had a very high reverberation level (but a reverberation time similar to the other rooms), yielding results consistent with a subjective interpretation of reverberance (i.e. the sound of reverberation) more in terms of level for Experiment 1, an more in terms of time for Experiment 2. Another aspect of this is that, for a given room, clarity index is negatively correlated to source-receiver distance (because the direct sound level decreases with distance), so this provides another reason for clarity index to predict auditory room size (since greater distance is associated with greater perceived room size).

In the physical characterisation of reverberation, we can easily distinguish the concepts of reverberation time and level. However, it is not clear that untrained listeners can make this distinction, and instead may simply get an impression of the quantity of reverberation (reverberance). Examination of individual subject responses suggests this to be the case – that some subjects are mainly listening to 'how much' reverberance is present, and associating more with greater room size. An interesting contrast is that 7 (of 30) subjects gave Room1 a greater mean rating than the remaining room conditions in Experiment 1, whereas 20 (of 30) subjects did this for Experiment 2. In the more artificial situation of headphone listening, a listener's approach to interpreting room size could be biased by cues normally associated with that medium – meaning that the listener might focus more on manipulations of sound that are used conventionally in audio production (which they experience in everyday listening to music, television, radio, etc.) rather than on the more subtle cues experienced in real architectural spaces. If this is so, the bland concept of reverberation quantity would likely play a greater role in headphone listening, consistent with the divergent ratings of Room 1 in the two experiments.

Further investigation of this area may involve a paired comparison test using binaural reproduction (to match the experimental method used in most of our previous experiments on auditory room size perception). That experimental method should be more sensitive and robust than the direct magnitude estimation method applied here. We are also planning to conduct an experiment with a stereo-dipole or similar reproduction system, rather than headphones, because results of Martignon *et al.* (2005) suggest that this should perform better than a headphone based binaural system.

CONCLUSIONS

The results of this study sound a note of caution in relation to representing architectural spaces through the use of non-individualised non-head-tracked binaural reproduction via headphones. It is well known that such systems have direction localisation artefacts, but the present study finds some divergence for room size perception. Nevertheless some correspondence between real rooms and headphone reproduction is seen, and more research is required to clarify reasons for divergence. This study reinforces the concept that clarity index can roughly predict auditory room size perception.

ACKNOWLEDGMENTS

The authors thank the experiment subjects for their voluntary participation in this project. This project was supported by a University of Sydney Research and Development Grant.

REFERENCES

Cabrera, D. and Gilfillan, D., 2002, "Auditory distance perception of speech in the presence of noise," *Proceedings of the International Conference on Auditory Display*, Kyoto, Japan

Cabrera, D., Jeong, D., Kwak, H. J. and Kim, J.-Y., 2005, "Auditory room size perception for modelled and measured rooms," *Proceedings of Internoise*, Rio de Janeiro, Brazil

Hameed, S., Pakarinen, J., Valde, K. and Pulkki, V., 2004, "Psychoacoustic cues in room size perception," *Proceed-*

ings of the 116th Audio Engineering Society Convention, Berlin, Germany

Kirkeby, O., Nelson, P. A. and Hamada, H., 1998, "The 'stereo dipole' – a virtual source imaging system using two closely spaced loudspeakers," *Journal of the Audio Engineering Society*, 46(5), 387-395

Martignon, P., Azzali, A., Cabrera, D., Capra, A. and Farina, A., 2005, "Reproduction of auditorium spatial impression with binaural and stereophonic sound systems," *Proceedings of the 118th Audio Engineering Society Convention*, Barcelona, Spain

McGrath, R., Waldmann, T. and Fernstrőm, 1996, "Listening to rooms and objects," *Proceedings of the 16th Audio Engineering Society International Conference*, Rovaniemi, Finland

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

Minnaar, P., Olesen, S. K., Christensen, F. and Møller, H., 2001, "Localization with binaural recordings from artificial and human heads," *Journal of the Audio Engineering Society*, 49(5), 323-336

Møller, H., 1992, "Fundamentals of binaural technology," *Applied Acoustics* 36(3-4), 171-218

Møller, H., Sørensen, M. F., Jensen, C. B. and Hammershøi, D., 1996, "Binaural technique: do we need individual recordings?" *Journal of the Audio Engineering Society*, 44(6), 451-469

Pop, C. and Cabrera, D., 2005, "Auditory room size perception for real rooms," *Proceedings of the Australian Acoustical Society Conference*, Busselton, Australia

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

Takeuchi, T., Nelson, P. A., Kirkeby, O. and Hamada, H., 1997, "Robustness of the performance of the 'stereo dipole' to misalignment of head position," *Proceedings of the 102nd Audio Engineering Society Convention*, Munich, Germany