CONTROL OF PERCEIVED ROOM SIZE USING SIMPLE BINAURAL TECHNOLOGY

Densil Cabrera

Faculty of Architecture, Design and Planning
University of Sydney
NSW 2006, Australia
densil@usyd.edu.au

ABSTRACT

The localization of auditory images and their size forms the bulk of the research literature in spatial auditory perception using binaural technology. Nevertheless, binaural technology conveys many other spatial characteristics of sound environments, and the present paper is concerned with one of these: auditory room size perception. This paper reviews the potential cues to room size perception conveyed through simple binaural technology. Statistical room acoustics is shown to provide indications of room size through energy relations between direct sound, early reflections and late reflections. However, binaural hearing could be important in distinguishing the concept of room size from source distance. These theoretical notions are considered in relation to experimental findings on room size perception using simple binaural technology.

[Keywords: Spatial hearing, Binaural technology]

1. INTRODUCTION

Binaural auditory displays can provide rich spatial information to a listener by reproducing at the listener's ears the sound that would occur in the represented environment. While the sense of a room size conveyed through binaural reproduction has received relatively little attention in the research literature, it has some potential to be used in spatial auditory display. For example, an auditory display could conceivably be designed to independently control the apparent source azimuth angle, source distance and room size if factors that influence these percepts are understood sufficiently. Room size, as an auditory display parameter, might need to be treated differently to azimuth and distance, because real rooms tend to be fixed in size, and a person must move to another room to experience a room of another size. Hence it is conceivable that room size could be used as a parameter that sets a context for other parameters that can change more rapidly (such as distance and azimuth, or indeed non-spatial auditory parameters).

This paper introduces important parts of the acoustic theory that underlies auditory room size perception, and compares the theoretical principles with results from subjective tests. Because simple binaural technology provides a convenient platform for spatial auditory display, this paper considers auditory room size perception in that context – both in terms of the capacity of simple binaural technology to reproduce likely auditory cues, and in terms of subjective experimental data on auditory room size using this technology.

2. SIMPLE BINAURAL TECHNOLOGY

The 'simple binaural technology' that is considered in the present paper tends to fall short of the ideal of virtual reality, but still provides an impression of a space that maintains some realism. It is simple in the sense that an interactive head-tracking system is not used and generic (rather than individual) head-related transfer functions (hrtf) are used. The advantage of this simple approach is that it is quite easy and inexpensive to implement, and so is in much wider use than more accurate binaural systems. Recordings for simple binaural reproduction can be made using a dummy head with microphones in its ears, or using a computer program that models the sound of an acoustic environment at a pair of virtual ears. In practice recordings are often made by convolving an anechoic source recording with binaural impulse responses (either measured or modeled). Many of the principles of binaural technology are discussed by Møller [1].

Even if binaural technology is 'simple', there are some techniques that can be helpful in improving realism. One is to ensure that the reproduction system's response is neutralized through inverse-filtering the transfer function from the system to the ears. In practice this is often done by inverting the measured transfer function (after smoothing) from the system to a dummy head in the listener's position, and can be thought of as avoiding the spectral effect of sound traveling twice through the pinnae (once in the original recording, and again in the reproduction). Another is to reproduce sound with calibrated gain, so that the listener's ears receive the same sound pressure level as they would have in the represented environment. To do this requires knowledge of the original soundfield's sound pressure level, or of the sound source's sound power level.

Well known spatial distortions in simple binaural reproduction systems include head-locking of the soundfield (the soundfield moves with the listener's head, removing dynamic localization cues and reducing the ability to externalize the auditory scene) and vague and inaccurate localization. The most important localization distortion is the rotation of auditory images around the 'cones of confusion'. A cone of confusion occurs at a fixed angle around the interaural axis, and is characterized by approximately constant binaural difference cues (and so the auditory system relies on spectral cues to resolve the image direction around the cone). Front-back confusion is a common instance of cone of confusion error, and simple binaural renderings of frontal sound sources tend to be localized behind or above the listener [2].

While simple binaural reproduction is most often done with headphones, it is also possible to reproduce non-head-tracked generic-hrtf binaural sound using a cross-talk cancelling loudspeaker system in an anechoic room. Conceptually the result should be the same: reproduction of the sound at the entrance of each ear. However, a listener's experience of such a system differs markedly from headphone reproduction, perhaps because the listener is not wearing headphones. The stereo-dipole is an interesting instance of a cross-talk cancelling binaural system, where the high frequency loudspeakers form a relatively narrow angle with respect to the listener. Advantages of this are that the upper frequency limit of cross-talk cancellation is raised so that most of the audio frequency range can be covered, and the system is robust in the face of minor head movements [3]. Perhaps because a pair of loudspeakers is visible in front of the listener, it appears to be possible to have frontally located auditory images using this system.

3. BASIC ACOUSTICS OF ROOM SIZE

There are many acoustical parameters that can vary with room size, and a key question for the control of auditory room size perception is which parameters can reliably control perceived room size. Further questions are if and how auditory room size can be controlled whilst maintaining other spatial percepts constant (such as auditory distance perception) and what the simplest method of auditory room size control is.

It can be hypothesized that auditory room size perception, like other aspects of spatial hearing, is learnt from everyday experience. That is, that the experience of acoustics of everyday rooms provides the basis for interpretation of auditory room size when only sound is present. Therefore this section of the paper considers general relationships between room size and acoustical parameters.

3.1. Diffuse Field Theory

Diffuse field theory is a powerful and simple approach to characterizing room acoustics of medium and large rooms. In its simplest form, the soundfield is conceived of as a direct field (the intensity of which, relative to source power, is related to the distance from the source) and a diffuse field (the energy density of which, relative to source power, is related only to the total absorption in the room). Even at a superficial level, diffuse field theory provides some indicators of room size: (i) reverberation time will tend to increase with room volume; and (ii) the strength of the diffuse field will tend to decrease as room volume increases. Hence, for a given source-receiver distance, a small room will be 'louder' than a large room, but will have a shorter reverberation time.

According to the diffuse field theory of Barron and Lee [4], energy relations in rooms can be calculated from the reverberation time, source-receiver distance and room volume – that is, the energy of the direct sound, early reflections and late reflections (reverberation), as given in equations 1 to 3. Energy values are scaled such that the direct sound is 0 dB at 10 m from the source.

$$E_{direct} = \frac{100}{r^2} \tag{1}$$

$$E_{early} = \left(\frac{31200T}{V}\right) e^{-0.04r/T} \left(1 - e^{-t_{\lim} 6\ln 10/T}\right) \tag{2}$$

$$E_{late} = \left(\frac{31200T}{V}\right) e^{-0.04r/T} e^{-t_{\lim} 6\ln 10/T} \tag{3}$$

Here T is reverberation time (in seconds), V is room volume (cubic metres), and r is source-receiver distance in metres (the source and receiver are omnidirectional) and $t_{\lim}$ is the limiting time that divides early and late energy (in seconds). E_{direct}, E_{early} and E_{late} are the direct, early and late energy of the room impulse response respectively.

Clarity index is the energy ratio of direct and early reflections to late reflections expressed in decibels (equation 4). If 80 ms is taken as the division between early and late energy, this is known as C_{80} (C_{50} is also in common use). Strength factor (G) is the total energy of the impulse response relative to an impulse response from a source of identical power at 10 m in the free field, and so is the sum of E_{direct}, E_{early} and E_{late} expressed in decibels (equation 5).

$$C_{t_{\lim}\times 1000} = 10\log\left(\frac{E_{direct} + E_{early}}{E_{late}}\right) \quad [\mathrm{dB}] \tag{4}$$

$$G = 10\log\left(E_{direct} + E_{early} + E_{late}\right) \quad [\mathrm{dB}] \tag{5}$$

Figure 1 shows these energy relations for a source-receiver distance of 2 m and reverberation time of 1 s as a function of room volume (80 ms is taken as the time dividing early and late reflections in the following figures). This shows that as room volume increases, the influence of the early and late reflections diminishes, leading to an increase in clarity index and a decrease in the strength of the sound.

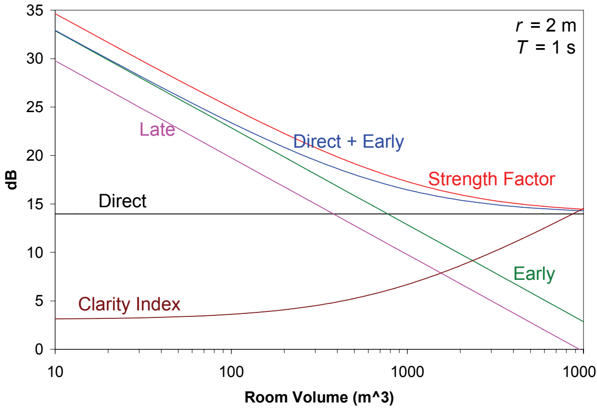

Figure 1. *Energy relations as a function of room volume for a fixed source-receiver distance (2 m) and a fixed reverberation time (1 s).*

Figure 1 maintains a constant reverberation time over a thousand-fold room volume increase, but this is unrealistic. There is a general tendency for reverberation time to increase with room volume, especially if similar room surfaces and furnishings are used. Rather than maintaining a constant reverberation time, Figure 2 maintains a constant average absorption coefficient. This shows that clarity index initially tends to decrease as room volume increases, but increases again for the large volume rooms.

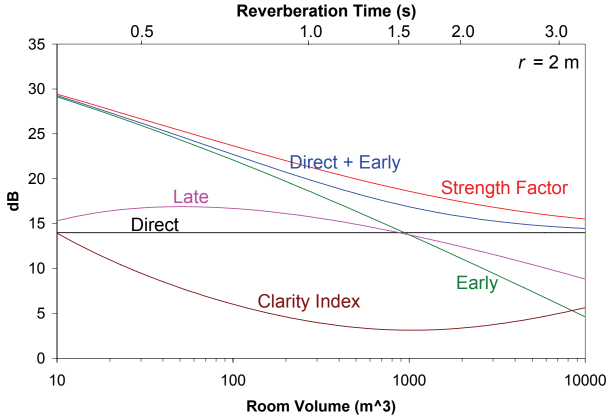

Figure 2. *Energy relations as a function of room volume for a fixed source-receiver distance (2 m) and a variable reverberation time derived from a constant average absorption coefficient.*

A still more realistic approach is to use data based on real rooms. Diaz and Pedrero [5] present regression functions derived from the reverberation time of 8246 furnished bedrooms and 3211 furnished living rooms with room volumes between 10 m^3 and 100 m^3. Reverberation decreases as room volume increases, and the resulting energy relations for a fixed source-receiver distance of 2 m are shown in Figure 3. This shows that the sound pressure level is dominated by the early reflections, which decrease as the room volume increases. In general the result is similar to the equivalent part of Figure 2.

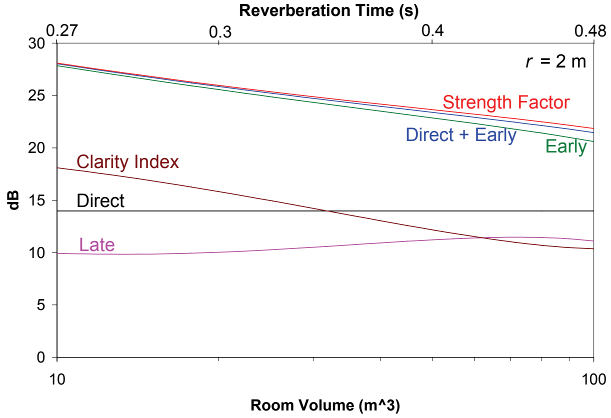

Figure 3. *Energy relations as a function of room volume for a fixed source-receiver distance (2 m) using the 1 kHz reverberation time regression function of Diaz and Pedrero [5].*

A question for auditory display is whether auditory room size could be varied independently of auditory distance. Part of the answer to this question comes from the acoustic differences between variations in room volume and source-receiver distance. Hence Figure 4 shows energy relations as a function of source-receiver distance for a room of constant volume and reverberation time. The most striking difference between this and the previous figures is that the direct sound decreases with increasing source-receiver distance (at -6 dB per doubling of distance) instead of remaining constant. The early and late reflections are quite constant, but decline at the large source-receiver distances. The question, then, is to what extent people can distinguish by ear the pattern of energy relations of Figure 4 from the pattern of energy relations of previous figures in which room volume was varied.

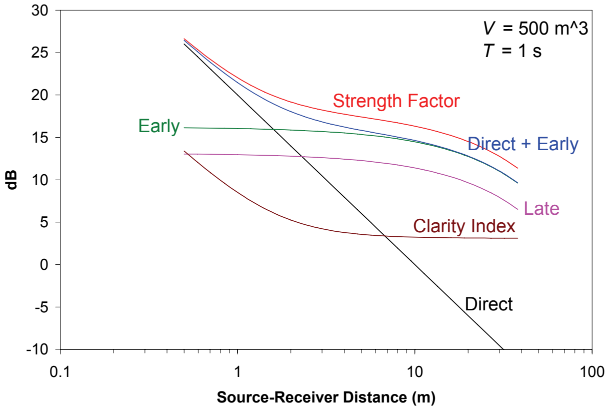

Figure 4. *Energy relations for a room of constant volume (500 m^3) and constant reverberation time (1 s) as a function of source-receiver distance.*

The potential of this theory for auditory room size perception is confirmed in purely physical terms by Kuster [6], who successfully used it for computational estimation of room size based on measured room impulse responses. However, knowing the source-receiver distance (which can be computed precisely from the delay between impulse emission and reception) is necessary to make the computation, implying that auditory distance estimation may contribute to auditory room size perception.

3.2. Binaural Acoustics of Room Size

The rationale for the division between early and late energy is that early reflections tend to be perceptually linked to the direct sound, whereas late reflections are heard as the room reverberation. This is related to the fact that clarity index can be an effective predictor of speech intelligibility, because the early reflections provide useful reinforcement of the direct sound while late reflections degrade intelligibility (50 ms is generally used to divide early and late sound for intelligibility predictions) [7]. The problem for auditory room size perception then is that if early reflections and direct sound are temporally fused, then how might a listener distinguish variable distance from variable room size using the energy relations discussed in the previous section?

Part of the answer of this question is the role of early reflections in auditory spatial perception.

Auditory source width (ASW – also known as 'apparent source width') has been studied extensively in the context of auditorium acoustics [8]. One of the key findings of this field has been that strong early reflections from the side walls of a room create the impression of a sound source that is spread out in space, rather than concentrated in a spot. A common way of assessing this image broadening is to measure the interaural cross-correlation coefficient (IACC), which is the maximum absolute value of the normalized interaural linear cross-correlation function (using time lags of ±1 ms, and the first 80 ms of the binaural impulse response). With the direct sound only, the IACC will be equal to 1, and will decrease to a minimum value approaching 0 as the relative strength of early lateral reflections increases (assuming that the reflected soundfield is diffuse). The implication of this is that IACC could be used by listeners for assessing the balance of direct and early energy, even though the two might be fused in auditory temporal perception. The balance of direct to reflected sound might then assessed by ear through ASW.

In terms of distinguishing room size from source-receiver distance, IACC will tend to increase as room volume increases, but will tend to decrease as source-receiver distance increases (notwithstanding local acoustic effects), and since these tendencies are in opposition, IACC could be of assistance in making a distinction between distance and room size, through aiding in the interpretation of the energy relations. However, these tendencies will only occur in situations where the direct sound and early reflections both have some influence on IACC – that is, when one is not much weaker than the other. Figures 1-4 show that it is quite plausible for the direct sound to be much weaker than the early reflections (in small rooms and for long source-receiver distances). The predicted early reflection energy is substantially greater than the direct sound throughout the entire range of rooms characterized by Diaz and Pedrero (although probably the 80 ms early reflection integration window is too long for this context). The relative strength of the early reflection energy also depends on the absorption of surfaces within the early reflection window, which in Barron and Lee's model is derived from reverberation time. In general the result of early reflection dominance will be low IACC values.

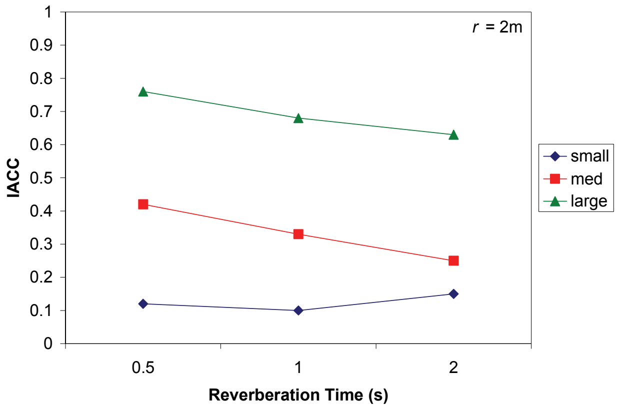

Figure 5. *Interaural cross-correlation coefficient for a source-receiver distance of 2 m in a small, medium and large room with reverberation times of 0.5 s, 1 s and 2 s.*

Figure 5 gives an example of the effect of room volume on IACC values, using simulated rooms (modeled using CATT-Acoustic, which is a computer program that models room acoustics using image-source and ray tracing methods). The room volumes are 31 m^3, 249 m^3 and 1997 m^3 – i.e., the room linear dimension increases by a factor of 2 between the small and medium, and the medium and large, rooms. Three reverberation times are applied to each room (0.5 s, 1 s and 2 s). In the case of the small room, the IACC is very low for all reverberation times, and so the results do not show any systematic effect as reverberation time changes. More generally, the effect of doubling the reverberation time on IACC is substantially less than the effect of doubling the room's linear dimension.

Figure 6 gives an example of the effect of source-receiver distance on IACC values (using two of the simulated rooms of Figure 5). However, the example scarcely agrees with the tendencies discussed in this section – IACC is low for the shortest source-receiver distance, rather than high. The reason for this is that the source position was fixed in the example, so bringing the receiver close to the source also brings it closer to the walls (because the source needed to be near walls in order to achieve a 4 m source-receiver distance in the medium room). Bringing the receiver near to the walls increases the strength of lateral early reflections, and so reduces IACC. The decrease in IACC for the 2 m to 4 m source-receiver distances is consistent with the general tendency that was discussed, but this reduction is less than the increase in IACC for the 1 m to 2 m distance.

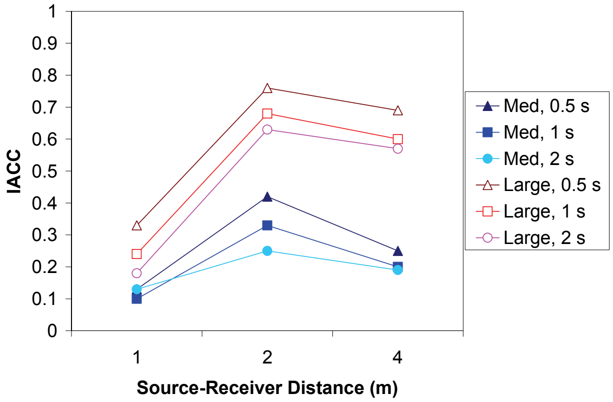

Figure 6. *Interaural cross-correlation coefficient for source-receiver positions in medium and large rooms with reverberation times of 0.5 s, 1 s and 2 s.*

This example illustrates a problem with IACC as a cue for room volume or source-receiver distance: that is, IACC is sensitive to local acoustic features, often more so than to the general tendencies discussed here. The simplifying assumption of a spatially diffuse early reflection soundfield is far from the reality of many acoustic situations. For a given source-receiver position the tendency of IACC increasing as room size increases should hold, but of course people do not normally experience a room that changes volume in reality, so it is not clear whether this theoretical concept can translate well to subjective judgments of room size represented by binaural technology.

Simple binaural technology does convey IACC effectively to a listener because the effect does not depend on individual head and pinna features. This suggests that if IACC contributes to room size judgments, simple binaural technology should provide a significant improvement over simpler forms of presentation in allowing listeners to distinguish room size from source-receiver distance, especially in situations where the direct sound energy is not overwhelmed by early reflections (or *vice versa*).

4. AUDITORY ROOM SIZE PERCEPTION: REVIEW OF EXPERIMENTAL RESULTS

4.1. Ability to Perceive Room Size through Sound

While there are few studies of auditory room size perception, they do confirm that auditory cues alone provide useful information on room size. Sandvad [9] found that subjects could usually correctly identify photographs of rooms that corresponded to binaurally reproduced soundfields representing those rooms. In subsequent experiments, Sandvad found that some listeners used the direct to reverberant energy ratio as a cue for room size estimates, while others used the reverberation time. McGrath *et al.* [10] 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. In yet to be published studies, Västfjäll *et al.* [11] and Larsson *et al.* [12] have found that visual impressions affect the perception of room acoustic conditions, including perceived room size. Studies by Mershon *et al.* [13], Hameed *et al.* [14], Sandvad [9], Cabrera *et al.* [15, 16] and Cabrera and Jeong [17] indicate that reverberation can have a strong effect on the auditory assessment of room size, and that while listeners can often judge room size correctly (at least in terms of rank order), reverberation effects can have a stronger influence on judgments than the actual room size. Ueno and Tachibana's [18] study of stage acoustical conditions indicates that a musician's impression of room size (based on the sound of their own instrument) is affected by reverberation time.

4.2. Ability to Distinguish Room Size from Source Distance

The possibility of room size and distance estimates forming independent perceptual dimensions is important if room size is to be used as a parameter in auditory display. Some perceptual studies have included both room size and distance estimates [e.g., 11, 12, 13, 19], but it is difficult to assess the extent to which subjects can make independent judgments of these attributes when both attributes are being assessed at once. A pair of studies by the author and colleagues is interesting in this respect because, for the same set of stimuli, different subject groups assessed one or the other of these auditory attributes. Figure 7 shows the auditory distance estimates of Cabrera and Gilfillan [2] and the perceived room size scale values of Cabrera *et al.* [15] for the same binaural stimuli. Stimuli were generated from a room of fixed volume (130 m^3) with variable reverberation time (0.7 s, 2 s and 5 s) and three source-receiver distances (0.9 m, 2.7 m and 5.1 m). A short speech phrase was used as the stimulus signal, and was reproduced over headphones using constant calibrated system gain. The perceived room size data were obtained through the method of paired comparisons, and the scale values are the probability of selection transformed by the inverse of the normal distribution (hence 0 corresponds to a 0.5 probability of a stimulus being selected as the larger room, 1 to a 0.85 probability of being selected, and 1.5 to a 0.93 selection probability). The distance estimates were absolute estimates in metres.

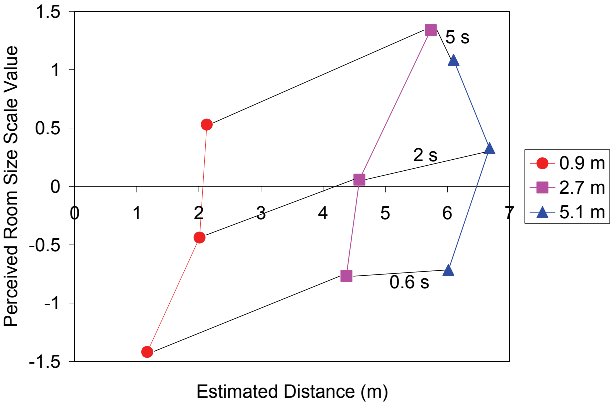

Figure 7. *Auditory distance versus auditory room size perception for a room of fixed volume (130 m^3) with source-receiver distances of 0.9, 2.7 and 5.1 m, and mid-frequency reverberation times of 0.6, 2 and 5 s.*

The relative strengths of direct, early and late energy were derived from room impulse responses, together with the energy sums and ratios discussed earlier. Perceived distance is highly correlated to the direct sound level (r = -0.97), and perceived room size is highly correlated to clarity index (r = -0.97 for C_{50}, r = -0.96 for C_{80}, and r = -0.94 for direct to reverberant energy ratio). For room size perception, there are more modest, but still high, correlations with early energy alone (r = 0.83), late energy alone (r = 0.90), early+late energy (r = 0.89), strength factor (r = 0.82), and reverberation time (r = 0.91). On the other hand, estimated distance does not correlate well with any of the energy components or ratios other than the direct sound. For distance perception, this confirms that simply controlling the direct sound level can strongly influence auditory distance perception (as is implied by Figure 4) even in a wide range of reverberant conditions. As reported by Cabrera and Gilfillan [2], there is an effect of reverberation time on perceived distance (greater reverberation time yields increased perceived distance, which is consistent with findings from other studies), but the effect is not strong enough to be significant in a regression analysis (including multiple stepwise regression) for the nine data-points considered here. In interpreting these results, it is important to note that the stimulus was a speech signal at a realistic sound power level, and this strong first-order relationship between direct sound and perceived distance might be weakened with different signals. As shown by Zahorik [20], speech signals can provide an absolute distance cue because the sound power of unassisted speech is familiar to listeners, whereas arbitrary signals (such as synthetic sounds) do not have this property and so are interpreted with a different cue weighting.

The correlations for room size perception indicate that the reverberant sound level is important for auditory judgments of room size, but that its relation to the direct sound and early reflections provides a stronger cue. Since the room volume did not vary in this experiment, the room size judgments cannot correspond strictly to a physical model. However, other experiments, in which room volume was changed, also have shown that reverberation has a strong effect on room size judgments, stronger than any actual change in room volume [15].

For these stimuli, IACC was also correlated to perceived room size (r = -0.87), but there are not enough data-points to achieve significance in a multiple regression analysis, so second order predictors of perceived room size cannot be identified in this way. Nevertheless it is interesting that the IACC correlation sign is negative, not positive (as discussed earlier, there should tend to be a positive correlation between IACC and actual room size). Whether the IACC correlation reflects an influence on perception is not known – the IACC correlation might be more due to coincidence (because it correlates with reverberant energy parameters) than causation, with the results caused by energy relations.

4.3. Predictors of Perceived Room Size

Every study of auditory room size perception that has included stimuli with a range of reverberation times has found reverberation to affect auditory room size perception (long reverberation time is associated with large room volume). However, the specific acoustical effects that most powerfully influence perceived room size vary between studies. With different methods of analysis in various studies, it is difficult to make direct comparisons between studies. In this subsection, results of two studies are outlined and analysed using the same approach. These are chosen because they include a large range of room volumes, and subjective ratings were found through the method of paired comparisons (which is more robust than direct estimation of magnitudes). The two experiments used simple binaural technology for the stimuli.

Cabrera *et al.* [15] obtained subjective room size values for the three computer modeled rooms given as examples earlier – with volumes of 31 m^3, 249 m^3 and 1997 m^3, three reverberation times applied to each (0.5 s, 1 s and 2 s), and three source-receiver positions in each (distances of 1 m and 2 m in the small room, and 1 m, 2 m and 4 m in the medium and large rooms). Stimuli were binaural, presented via headphones. The signal (convolved with binaural room impulse responses) was the same speech phrase as used by Cabrera and Gilfillan [2]. Results show that a doubling of the room's linear dimension (i.e., increasing its volume by a factor of eight) does yield judgments of increased room size, but that doubling the reverberation time has a much stronger effect. Reverberation time has the strongest correlation with the subjective responses (r = 0.93), and clarity index (C_{80}) is also highly correlated (r = -0.84). A stepwise regression yields a model (r^2 = 0.976) with three independent variables: reverberation time (coefficient of 0.941), stimulus sound pressure level (coefficient of -0.61) and C_{80} (coefficient of -0.107). An alternative model can be constructed from energy parameters alone (r^2 = 0.963): C_{80} (coefficient of -0.212) and E_{early} (coefficient of -0.075). IACC does not make a significant contribution to the results.

In the second study, Cabrera and Jeong [17] obtained subjective room size values for binaural simulations of four Italian concert halls: Parma's *Auditorium Paganini* (780 seats, 48 m long and 17.5 m wide), and the three main halls of Rome's *Parco della Musica* (small – 700 seats, 35 m long and 25 m wide; medium – 1200 seats, 48 m long and 34 m wide; and large – 2800 seats, 56 m long and 32 m wide). In the smallest auditorium of the *Parco della Musica* there were two source-receiver distances (12 m and 24 m), and there were three source-receiver distances in all of the other auditoria (ranging between 10 m and 48 m) – making a total of eleven stimuli. A short fragment of music (piano accordion) was used as the stimulus signal, and was convolved with binaural impulse responses from these auditoria to generate the stimuli (the binaural impulse responses were recorded by Angelo Farina and colleagues [21] using a fixed set of equipment and a constant gain structure). At the time of writing, the experimental data-set is almost complete (18 out of a planned 20 subjects have been tested), and preliminary results are correlated to several room acoustical parameters, including sound pressure level (r = -0.88), clarity index C_{50} (r = -0.86), and many of the other components of the energy relations discussed earlier in this paper. The best correlation is for $E_{direct}+E_{early}$ in the 1000 kHz octave band (r = -0.96) using 50 ms integration.

The correlations between perceived room size and acoustical parameters in these two experiments agree (in sign, if not in magnitude) with the tendencies that would be expected from diffuse field theory (the expected sign of clarity index is discussed below). However, any effect of IACC is not strong enough to contribute to regression analysis. The effect of energy parameters is so strong as to leave little variance to be accounted for by other parameters.

If clarity index is a predictor of auditory room size perception, this raises an interesting problem. For a given reverberation time and source receiver distance, clarity index correlates positively to room volume (Figure 1) – whereas the subjective results have a strong negative correlation. This apparent paradox is examined in some detail elsewhere by the author [22], and is resolved by assuming that rooms of varying volume do not usually have the same reverberation time, and that larger source-receiver distances are only experienced in larger rooms. A theoretical evaluation of this for rooms of constant absorption coefficient with source-receiver distances proportional to the room's linear dimension shows that clarity index is negatively correlated to room volume. This is confirmed as being realistic by considering acoustical data from a large number of real rooms, which show the same negative correlation tendency for source-receiver distances proportional to the room's linear dimension.

4.4. Effectiveness of Binaural Representation

There are many studies on the effectiveness of various binaural techniques for sound localization [e.g., 23, 24], but not on the effectiveness of such systems for room size perception. The author has been involved in two studies that provide a little insight into this question, but further studies are required to clarify issues raised in the results. If room size perception were based entirely on reverberation and energy relations, then perhaps binaural reproduction would be unnecessary to convey a sense of room size. However, the two studies outlined in this section both indicate that auditory room size perception is sensitive to the spatial qualities of the soundfield and/or to confounding factors that are associated with the use of particular spatial audio systems.

Cabrera *et al.* [16] investigated whether simple binaural reproduction conveys the same impression of room size as real rooms, using rooms with volumes of 15 m^3, 123 m^3 and 188 m^3, as well as variable absorption in the mid-sized room and different source-receiver distances. Mid-frequency reverberation times for the rooms were 1.0 s (small room), 0.5 s (medium room, curtains), 0.8 s (medium room, bare walls), and 1.0 s (large room). Real room assessments were made with 30 blindfolded subjects (who wore earmuffs to and from the rooms), using recorded speech of fixed sound power (reproduced from a loudspeaker) as the source. Binaural simulations were strictly calibrated and inverse-filtered to match the original signals received at the dummy head's ears in the real rooms, and assessed by a further 30 subjects. Subjects rated the size of the room relative to that of a reference room, and results were scaled by each subject's mean response. Results show a contrasting subjective interpretation of the small room (which had a long reverberation time for its size) – it is heard as small in the real soundfield, but large in the binaural simulation. The medium and large room ratings are similar for the two presentation modes, although there is a wider range of values in the ratings of real rooms. The reason for the contrasting ratings of the small room are not clear, but may be due to its long reverberation time – which could have been easier to interpret in the real room than using a binaural simulation. In the real room, many subtle cues were available that were not available in the binaural simulation (such as feedback from self-made sounds, dynamic localization cues, the feeling of the floor surface (through shoes), and so on).

Martignon *et al.* [19] investigated auditory distance perception, room size perception and realism for simulated soundfields of five concert auditoria (the same auditoria as used by Cabrera and Jeong [17], plus an auditorium in Japan), using three binaural systems and one stereophonic system. The binaural systems employed binaural impulse responses recorded in the auditoria, and the reproduction systems of simple headphones, stereo dipole and double stereo dipole (i.e., a stereo dipole loudspeaker pair both in front of and behind the listener). The stereophonic system employed impulse responses recorded using an O.R.T.F. microphone array (two cardioid microphones 17 cm apart separated by an angle of 110º), and was reproduced using a conventional stereophonic loudspeaker array (with loudspeakers 60º apart). Great care was taken in matching the gain and spectral response of each system, and in reproducing the stimuli at the absolute sound pressure level that would occur for the original source in the auditoria. Thirty subjects used a rating scale to report the room size. While the auditory room size judgments are less reliable than a paired comparisons test, they do reveal differences between audio systems. The stereophonic system yielded room size ratings that were highly correlated with auditory distance ratings, whereas greater divergence between these scales was found for the three binaural systems. Room size ratings for the stereophonic stimuli correlated best to stimulus sound pressure level (r = -0.73), whereas correlations with sound pressure level are weaker for the binaural systems (r = -0.67, -0.54, and -0.43 for double stereo dipole, headphones, and stereo dipole respectively). On the other hand, IACC tended to be correlated to the binaural system responses, but not to the stereophonic system. This relationship is best assessed without the Japanese auditorium because its stimuli had unusually high IACC values (it is not clear whether this is due to measurement error or the auditorium acoustical conditions) [25]. Correlations between IACC and room size ratings were r = -0.11 (stereophony), r = -0.74 (double stereo dipole and headphones), and r = -0.79 (stereo dipole), and IACC was one of the best predictors of auditory room size perception for the binaural systems in this experiment. While this appears to confirm that IACC can contribute to auditory room size perception in binaural systems, the sign of the relationship is negative (whereas the sign based on the tendency in room acoustics as volume is varied is positive). One reason for the negative coefficient could be that larger distances were used in the larger auditoria.

5. CONCLUSIONS

A naïve and straightforward approach to controlling auditory room size in an auditory display might be to adjust reverberation time using a parametric reverberation processor. However, the acoustical relationship between the characteristics of reverberation and room volume is considerably more involved than reverberation time alone: for example, simply increasing reverberation time will increase the late energy level, which in physical acoustics implies that the room size has not changed (merely that the room is less absorptive). Nevertheless, the findings reviewed in this paper indicate that this naïve approach can be very effective: for a given room volume, a change in reverberation time yields a change in perceived room size with a comparatively small change in auditory distance perception. Indeed, changing reverberation time is more powerful than changing the actual room size in binaural simulations.

Based on general acoustic tendencies, binaural factors have a potential role in auditory room size perception, including in helping to assess the relationship between direct sound and early reflections. However, while the reviewed perceptual experiments do sometimes indicate a correlation between IACC and judgments of room size, there are no clear results relating IACC to room size perception in a manner consistent with the theoretical tendency. Instead, some results are even opposite to the acoustic tendency. Experimental work is required to investigate this problem directly.

Binaural technology is effective for reproducing interaural characteristics of the soundfield such as IACC, and the study of Martignon *et al.* [19] indicates that this can affect room size judgments. However, the effectiveness of simple binaural technology in conveying the same impression of room size as auditory perception in real rooms remains open for investigation – while the results of Cabrera *et al.* [16] show a discrepancy for a small room, there may be confounding factors in the experiment that contributed to this.

In summary, people can perceive room size through sound alone, but reverberation has a strong effect on auditory room size judgments. Auditory room size and auditory distance perception can be varied with a quite high degree of independence. Judgments of room size appear to be mainly based on reverberation energy parameters, and the role of IACC remains unclear. Nevertheless, binaural technology appears to be helpful in the perceptual distinction between auditory room size and auditory distance.

6. REFERENCES

[1] H. Møller, "Fundamentals of binaural technology," *Applied Acoustics*, vol. 36, no. 3-4, pp. 171-218, 1992.

[2] D. Cabrera and D. Gilfillan, "Auditory distance perception of speech in the presence of noise," in *Proc. 7th International Conference on Auditory Display*, Kyoto, Japan, pp. 431-439, 2002.

[3] T. Takeuchi, P. A. Nelson, O. Kirkeby, and H. Hamada, "Robustness of the performance of the 'stereo dipole' to misalignment of head position," in *Proc. 102nd Audio Eng. Soc. Conv.*, Munich, Germany, 1997.

[4] M. Barron and L.-J. Lee, "Energy relations in concert auditoriums. I," *J. Acoust. Soc. Am.*, vol. 84, pp. 618-628, 1988.

[5] C. Diaz and A. Pedrero, "Reverberation time of furnished rooms in dwellings," *Applied Acoustics*, vol. 66, pp. 945-956, 2005.

[6] M. Kuster, "Room volume estimation from diffuse field theory," in *Proc. 4th Joint Meeting of the Acoust. Soc. Am. and Acoust. Soc. Japan*, Honolulu, USA, December 2006.

[7] J. Bradley, "Predictors of speech intelligibility in rooms," *J. Acoust. Soc. Am.*, vol. 88, pp. 837-854, 1986.

[8] L.L. Beranek, *Concert Halls and Opera Houses: Music, Acoustics, and Architecture*, Springer, New York, NY, 2004.

[9] J. Sandvad, "Auditory perception of reverberant surroundings," *J. Acoust. Soc. Am.*, vol. 105, no. 2, p. 1193.

[10] R. McGrath, T. Waldmann, and M. Fernström, "Listening to rooms and objects," in *Proc. 16th Audio Eng. Soc. Int. Conf.*, Rovaniemi, Finland, 1996.

[11] D. Västfjäll, M. Carlsson, P. Larsson, M. Kleiner and B.-I. Dalenbeck, "Visualization and auralization in merging: Influence of visual cues on room acoustic perception," (submitted for publication).

[12] P. Larsson, D. Västjäll and M. Kleiner, "Auditory-visual interaction in concert halls," (submitted for publication).

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

[14] S. Hameed, J. Pakarinen, K. Valde and V. Pulkki, "Psychoacoustic cues in room size perception," in *Proc. 116th Audio Eng. Soc. Conv.*, Berlin, Germany, 2004.

[15] D. Cabrera, D. Jeong, H.J. Kwak, and J.-Y. Kim, "Auditory room size perception for modeled and measured rooms," in *Proc. Internoise*, Rio de Janeiro, Brazil, 2005.

[16] D. Cabrera, C. Pop, and D. Jeong, "Auditory room size perception: A comparison of real versus binaural sound-fields," in *Proc. 1st Australasian Acoustical Societies' Conference*, Christchurch, New Zealand, November 2006, pp. 417-422.

[17] D. Cabrera and D. Jeong, "Auditory room size perception in concert auditoria," in *Proc. 19th Int. Congress on Acoustics*, Madrid, Spain, September 2007.

[18] K. Ueno and H. Tachibana, "Experimental study on the evaluation of stage acoustics by musicians using a 6-channel sound simulation system," *Acoust. Sci. & Tech.*, vol. 24, no. 3, pp. 130-138, 2003.

[19] P. Martignon, A. Azzali, D. Cabrera, A. Capra, and A. Farina, "Reproduction of auditorium spatial impression with binaural and stereophonic sound systems," in *Proc. 118th Audio Eng. Soc. Conv.*, Barcelona, Spain, May 2005.

[20] P. Zahorik, "Assessing auditory distance perception using virtual acoustics," *J. Acoust. Soc. Am.*, vol. 111, pp. 1832-1846, 2002.

[21] A. Farina and R. Ayalon, "Recording concert hall acoustics for posterity," in *Proc. 24th Int. Audio Eng Soc Conf. on Multichannel Audio*, Banff, Canada, 2003.

[22] D. Cabrera, "Acoustic clarity and auditory room size perception," in *Proc. 14th Int. Conf. Noise and Vibration*, Cairns, Australia, July 2007.

[23] P. Minnaar, S. K. Olsen, F. Christensen, and H. Møller, "Localization with binaural recordings from artificial and human heads," *J. Audio Eng. Soc.*, vol. 49, no. 5, pp. 323-336, 2001.

[24] H. Møller, M. F. Sørensen, C. B., Jensen, and D. Hammershøi, "Binaural technique: do we need individual recordings?" *J. Audio Eng. Soc.*, vol. 44, no. 6, pp. 451-469, 1996.

[25] D. Cabrera, A. Azzali, A. Capra, A. Farina and P. Martignon, "Perceived room size and source distance in five simulated concert auditoria," in *Proc. 12th International Congress on Sound and Vibration*, Lisbon, Portugal, July 2005.