Chapter 4
Auditory versus Visual Spatial Impression[1]

Through separate experiments, this chapter considers how visual and auditory spatial impressions vary within two auditoria, Hall A and C, and hence identifies similarities between these two sensory modes.

In the visual experiment, the 'spaciousness', 'envelopment', 'stage dominance', 'intimacy' and target distance were judged by subjects using greyscale projected photographs, taken from various positions in the audience areas of the two auditoria when a visual target was on stage.

In the auditory experiment, the 'apparent source width', 'envelopment', 'intimacy' and performer distance were judged using an anechoic orchestral recording convolved with binaural impulse responses measured from the same positions in the two auditoria.

Results show target distance to be of primary importance in auditory and visual spatial impression – thereby providing a basis for covariance between some attributes of auditory and visual spatial impression. Nevertheless, some attributes of spatial impression diverge between the senses.

[1] (Cabrera, D., Nguyen, A., & Choi, Y. J., 2004). *Refereed paper, ICAD 04 Tenth Meeting of the International Conference on Auditory Display,* Sydney.

4.1. INTRODUCTION

For many years, researchers in the field of auditorium acoustics have been interested in the auditory spatial impression of auditoria, and its variation within an auditorium. Auditory spatial impression contributes much to the judged acoustical quality of music soundfields. Three of the most commonly studied spatial attributes are apparent source width (commonly abbreviated to ASW), listener envelopment (LEV) and intimacy, with a high level of each of these being desirable for musical acoustics (Beranek, 1996 p23, 36-38).

ASW, an aspect of the more general term 'spaciousness', is used to assess the effect of an auditorium environment on the auditory image size of a musical performance. Strong early lateral reflections are among the factors that can contribute to large ASW. Although ASW disregards expansion in dimensions other than width, width is thought to be the most significant expansion dimension, because of the horizontal separation between the two ears. The absolute maximum interaural cross-correlation coefficient (IACC) taken from the first 80 ms of a binaural impulse response is often used as a predictor of ASW.

LEV refers to the sense of being surrounded that is engendered by a reverberant field. While ASW is affected primarily by the spatial impulse response 0-80 ms after the direct sound, LEV is primarily affected by the late sound, after 80 ms. The strength of the late lateral sound energy is an important predictor of LEV. IACC can also be used to predict LEV, although late lateral sound level measures are more effective.

Intimacy is a sense of closeness to and involvement with a performance – as opposed to a sense of detachment. The ability to discern detail is likely to be important for intimacy. While some researchers relate intimacy to the delay between the direct sound and the first significant reflection in the impulse response (Ando, 1985, p48 & pp50-51), this relationship may be of limited application (Barron, 1993, pp39-40). Perceived distance is not normally considered directly in auditory spatial impression in the auditorium acoustics field, but could plausibly be related to intimacy. In rooms, perceived distance depends on multiple cues, such as sound pressure level, frequency content, familiarity with the source, and the ratio of direct to reverberant sound energy (Zahorik, 2002). In auditory distance estimation experiments, there is a tendency for underestimation, especially in the far field – and there may be a context-dependent horizon beyond which all sound sources are heard as equally distant.

Beyond auditorium acoustics, auditory spatial impression has been studied in audio systems. Soulodre et al (2003) have found that physical acoustic correlates of LEV apply to spatial audio reproduction in much the same way as in auditorium contexts. Rumsey (2002) has identified dimensions of spatial audio reproduction, drawing partly on the auditorium acoustics tradition, as well as audio quality listening tests. His key attributes are grouped into those concerned with width, depth and immersion respectively. The concepts of width, depth and immersion are applied to individual sound sources, the ensemble, and the environment. In fact the model is rather more complex than described here, its complexity suggesting that the three previously mentioned attributes of ASW, LEV and intimacy scarcely account for the variety of impressions available in spatial hearing, especially if contexts beyond auditorium acoustics are considered.

The study of visual space appears not to offer a concentrated body of work on spatial impression equivalent to that offered by auditorium acoustics. The approach of the study in this chapter is to apply visual spatial impression scales based on the three auditory spatial impression scales of ASW, LEV and intimacy.

ASW may not be simply related to an aspect of visual experience, because the visually perceived boundaries of images tend to be better defined than those of auditory images. In an auditorium context, the concept of 'stage dominance' could be a rough visual counterpart of ASW – if the stage is to be thought of as the visual target. The term 'spaciousness', which is often closely related to ASW in the auditorium acoustics literature, pertains more to the perception of a room's field of view in vision (Gibson, 1950b; Glinksy, 1951; Ittelson, 1960 pp63-64) rather than the space occupied by performers. The perceived size of a space depends on the size of the visual solid angle and apparent depth of the visual image. Perspective is a cue for depth (Ittelson, 1960, pp63-64), and the balance of surfaces across visual images is controlled by room form (such as width and height).

Envelopment, in the visual world, is a feeling of being surrounded by objects, surface, people, light, and more. In phenomenological writings, envelopment refers to one's body in the centre with the experiential world surrounding (Bachelard, 1964, pp231-241; Pallasmaa, 1996, p26). According to this theory, we experience the world not only through vision, but also through the body. This seems to suggest that 'envelopment' seems to translate well between modes of experience, such as auditory, olfactory, and kinesthetics (Hall, 1966, pp52). Roundness in an architectural space appears to encourage this sense of surroundedness. Darkness – or the depletion of visual perception – may also provide a sense of envelopment as the person relies on senses closer to the body (Merleu-Ponty).

Intimacy, in the visual world, relates to distance or proximity of the body to the surrounding visual elements such as surfaces, objects, or people. Visual intimacy is also associated with the ability to see clearly facial details or surface textures (Gibson, 1950).

Visual distance cues include monocular accommodation and binocular convergence, perspective, knowledge of the target (especially its size), retinal image size of boundaries and textures, and many more (Gibson, 1950a pp1950b; Glinksy, 1951; Ittelson, 1960 pp63-64). There is some evidence for progressive underestimation of visual distances with increasing physical distance, which may be analogous to the underestimation of auditory distances previously mentioned (Gilinsky, 1951, Heelan, 1983 pp28-35).

Nguyen and Cabrera (2004a) studied the visual spatial impression of three auditoria (two of which are represented in the present study). Subjects (professional architects) made judgements of greyscale photographs using the terms ‘spaciousness’, ‘envelopment’, ‘stage dominance’ and ‘intimacy’, and the subjects also estimated the performer distance. That study (found in Chapter 3) found performer distance to be judged rather accurately across the range from 10 m to 45 m, with a tendency to underestimate. Spaciousness was close to the inverse of intimacy, both being a function of distance from the visual target on the stage. Envelopment was not affected by distance, but instead was affected by the auditorium. Stage dominance was affected by distance and auditorium. However, that study mixed images from the three auditoria in the experiment, thereby focussing the subjects’ attention on differences between the auditoria. In the present study subjects assess one auditorium at a time, and so the emphasis is on differences in spatial impression within each auditorium.

4.2. AIM

This study compares the auditory and visual spatial impression in representations of two auditoria, Hall A & C. As a case study, it investigates the degree of commonality between auditory and visual spatial impression in each auditorium. If matching attributes of spatial impression in the two senses co-vary, then there is some prospect for application where audiovisual interaction are useful.

4.3. METHOD

4.3.1. Auditory experiment

Binaural impulse responses were measured in both halls using a Brüel & Kjær Head and Torso Simulator (HATS), and a loudspeaker in the centre of the stage. The HATS ears were at a height of 1.2 m above the floor, and the loudspeaker at a height of 1.4 m, in the centre of the stage. In Hall A, a Mackie HR824 loudspeaker was used as the measurement source, whilst a Soundsphere 2212-1 loudspeaker on a custom-built subwoofer was used in Hall C (Bassett et al, 2003). The loudspeaker for Hall A is a studio monitor, while the system in Hall C is closer to omnidirectional, more powerful, and covers a broader frequency range (logistics precluded taking this system to New Zealand). In both auditoria, impulse responses were obtained from logarithmic sine sweeps (60 Hz – 18 kHz, 60 s in Hall A; and 20 Hz – 20 kHz, 5.4 s in Hall C). Hall A had 15 receiver positions (galleries and stalls), while Hall C had 9 (stalls only) – refer to Table 4.1. These responses were extracted and analysed using Aurora software (Farina, 2000), yielding 2.5 to 3.5 seconds of impulse response data.

Seat Number	Hall	Distance from Source	Angle from Midline	Gallery or Stalls
23D6	A	12 m	60°	Gallery
24B6	A	13.5 m	35°	Gallery
24F4	A	20 m	35°	Gallery
25D6	A	20 m	25°	Gallery
26C9	A	20 m	10°	Gallery
26H3	A	25 m	0°	Gallery
H32	A	10 m	0°	Stalls
O34	A	15 m	0°	Stalls
O40	A	15 m	15°	Stalls
V26	A	20 m	15°	Stalls
W34	A	20 m	0°	Stalls
W08	A	22 m	45°	Stalls
AA25	A	25 m	15°	Stalls
BB35	A	25 m	0°	Stalls
HH44	A	30 m	15°	Stalls
D12	C	10 m	0°	Stalls
D17	C	10 m	15°	Stalls
D7	C	10 m	15°	Stalls
I12	C	15 m	0°	Stalls
I19	C	15 m	15°	Stalls
I5	C	15 m	15°	Stalls
M21	C	20 m	15°	Stalls
M3	C	20 m	15°	Stalls
N10	C	20 m	0°	Stalls

Table 4.1: Measurement positions represented by the auditory and visual stimuli.

Measurements were calibrated, so that the sensitivity of the two ear channels could be matched, and the variation in sound pressure level through each auditorium could be maintained. However, the measurement system level between auditoria could not be matched meaningfully due to the different loudspeaker systems used.

An anechoic recording of the opening 15 seconds of the Overture to the Marriage of Figaro (Mozart) (Dennon, 1993a) was convolved with the measured impulse responses to create the subjective experiment auditory stimuli. As this experiment used Sennheiser HD600 headphones, the transfer function between these headphones and the HATS ear simulators was compensated for through 1/3-octave band digital equalisation.

Subjects listened to the stimuli in an anechoic room using headphones. Although anechoic conditions were not required, the low noise floor of this room (of 18 dBA) was presumed to be important in allowing subjects to hear the sometimes-subtle differences between the sound stimuli. A black curtain covered the absorptive wall in front of the subjects so as to reduce visual distraction.

The 13 subjects (researchers and students in acoustics and audio) were mostly already familiar with the concepts of spatial impression such as ASW, LEV and intimacy. Nevertheless, these were defined and explained to them and illustrated graphically to ensure that they responded meaningfully to the scales. Stimuli representing the two auditoria were presented in different sessions, so as to emphasise sensitivities to differences within each hall. Subjects listened to all stimuli for an auditorium before beginning their assessment of that auditorium. For each auditorium, four audio compact discs were prepared containing all stimuli, in distinct randomly determined order. Subjects listened to one of these discs for training, and another for the experiment. Half the subjects assessed Hall A first, and the other assessed Hall C first. Thus stimulus order was systematically varied between subjects, such that no two subjects had the same overall stimulus order.

	Hall A	**Hall C**
Reverberation Time RT30	2.3 s (0.05)	2.2 s (0.04)
Early Decay Time	1.7 s (0.24)	2.3 s (0.09)
Clarity Index C50	2.7 dB (1.1)	-6.5 dB (1.3)
Clarity Index C80	4.5 dB (1.3)	-3.5 dB (1.0)
IACC	0.28 (0.09)	0.21 (0.09)
Broadband SPL	76 dB (1.8)	79 dB (1.3)

Table 4.2: Acoustical characteristics of the auditory stimuli.

ASW, LEV and intimacy were rated on a discrete integer scale ranging between 0 (defined as 'none') and 10 (defined as 'maximal'). Source distance was estimated in meters.

Table 4.2 summarises the characteristics of the auditory stimuli, either in terms of the impulse response properties or binaural measurements of the actual stimuli. Unless otherwise stated, values are for mid frequencies (500 Hz and 1 kHz octave band mean). Low frequency values (mean of 125 Hz and 250 Hz octave bands) and high frequency values (mean of 2 kHz and 4 kHz octave bands) were also included in the analysis, as well as other kinds of acoustical measurement. Exponential means are used for decibels, and standard deviations are parenthesised.

IACC is the early interaural cross-correlation coefficient (Beranek, 1996, pp), and SPL is the equivalent sound pressure level (Leq) measured at the microphones of a dummy head wearing the headphones used in the experiment.

4.3.2. Visual experiment

Greyscale images were taken at a height of 1.2 m at selected seat positions, with a music stand and chair on the stage at the centre of the frame (at the same location as the loudspeaker of the auditory experiment). Camera locations were the same as the dummy head locations of the auditory experiment, except that additional photographs were taken in Hall C, mainly in its gallery (no impulse responses had been recorded in the Hall C gallery for logistical reasons). A Nikon Coolpix 5400 digital camera was used with a lens focal length equivalent to 28 mm (with respect to a 35 mm film camera). Typical minimal performance lighting conditions were used. Fifteen images were taken in each hall. The additional positions used for photographs in Hall C are summarised in Table 4.3.

Seat Number	Hall	Distance from Source	Angle from Midline	Gallery or Stalls
O5	B	20 m	10°	Stalls
O10*	B	20 m	0°	Stalls
GR3	B	10 m	45°	Gallery
GR15	B	25 m	15°	Gallery
GA10	B	20 m	0°	Gallery
GC10	B	22 m	0°	Gallery
GE10	B	25 m	0°	Gallery

Table 4.3*: Additional measurement positions used in the visual but not the auditory experiment. *Seat O10 was used instead of N10 in the visual experiment (these seats are both close to 20 m from the source, on the auditorium's midline).*

The images were presented via a data projector, with the image size approximately 1.5 m across in a small windowless room. There were 13 subjects, 10 of whom had not participated in the auditory test. The distance between the subjects and the screen was from 2 to 3 m. The lights were dimmed in the room. The tests were presented using Microsoft PowerPoint, with the slides being automatically changed in a timed sequence. They began with a preview section, which allowed the subjects to see all 15 images for a hall. The first image was projected for 60 s, each subsequent image having a 2 s reduction, hence the last having a duration of 32 s. The two halls were presented in separate slide shows, so as to emphasise sensitivity to differences between images within each hall. Image order was randomly varied between sessions.

For each image, subjects rated the Sense of Spaciousness (defined as the apparent volume of the room, as perceived from the image); the Sense of Envelopment (defined as the sense of being surrounded by the three-dimensional space); the Stage Spatial Dominance (defined as the degree to which the stage dominates the image); the Sense of Intimacy (defined as the apparent proximity or closeness to the performer); and the estimated distance from the camera to the performer (in meters). The first four subjective parameters were rated using a discrete scale (from 0 to 10, where 0 was defined as 'none', and 10 as 'maximal'). Table 4.4 summarises some characteristics of the images used in the experiment. Note that some images in the stalls of Hall A scarcely had a view of the stage surface due to their angle of view.

	Hall A	Hall C
Image Brightness	78 (16)	91 (8)
Stage Brightness	213 (30)	173 (17)
Non-stage Brightness	75 (18)	81 (14)
Stage:Non-stage area ratio	0.02 (0.02)	0.12 (0.12)

Table 4.4*: Image characteristics. Brightness is stated as an 8-bit number (ranging from 0-255).*

4.4. RESULTS

As the two auditoria were tested in separate sessions in both the auditory and visual experiments, they are separated in the analysis (although they are charted together in the figures). Analysis of variance (ANOVA, factorial) was used to test the significance of differences in subjective ratings for the single independent variable of Seat Number. ANOVA shows significance in both auditoria for the auditory scales of LEV and Intimacy and the visual scales of Spaciousness, Stage Dominance, Intimacy and Estimated Distance. Non-significant auditory results were obtained in Hall A for ASW (P=0.35) and the ratio of perceived to actual distance (P=0.35). A non-significant visual envelopment result was obtained in Hall A (P=0.12). Refer to Appendix 2 for ANOVA tables.

4.4.1 Distance Estimates

Despite their lack of significance (which reflects a wide spread of responses), auditory distance results show some systematic relationships with distance in both auditoria (Figure 4.1). Using trim means (upper and lower quartiles excluded), the stall seats in Hall A match actual distance well (r=0.83, P=0.004), whilst the gallery seats lack such a relationship. In

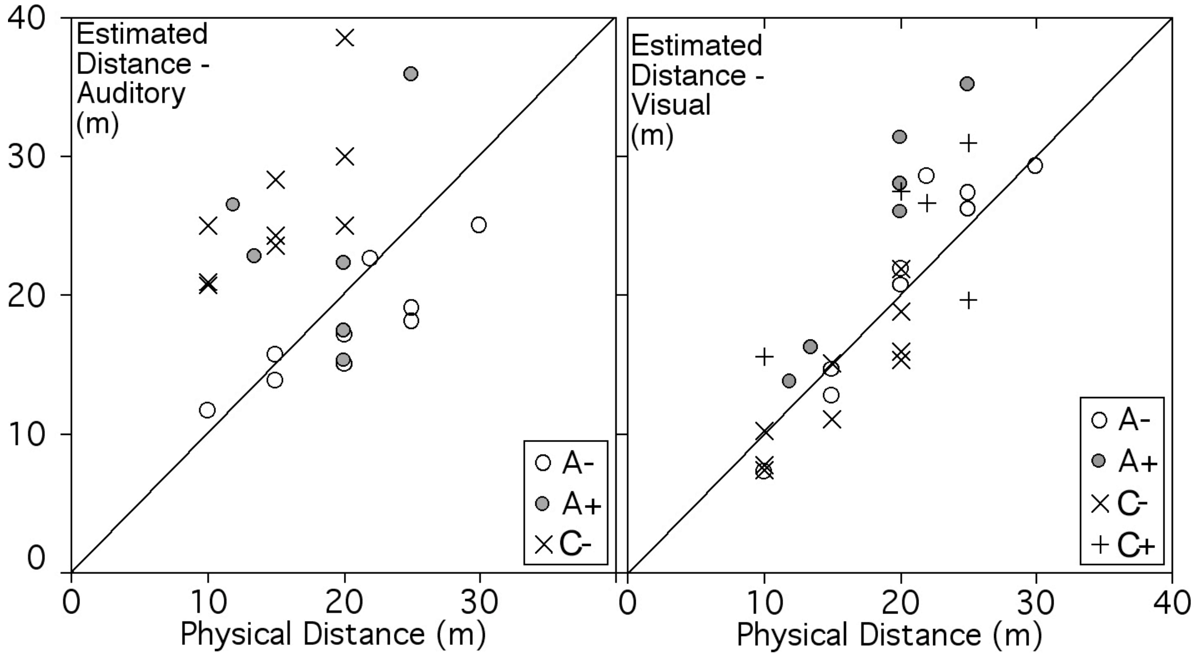

Figure 4.1: *Auditory and visual distance estimates (auditory are trim-mean). 'A' and 'C' are the two halls, with '-' referring to stall positions and '+' to gallery positions*

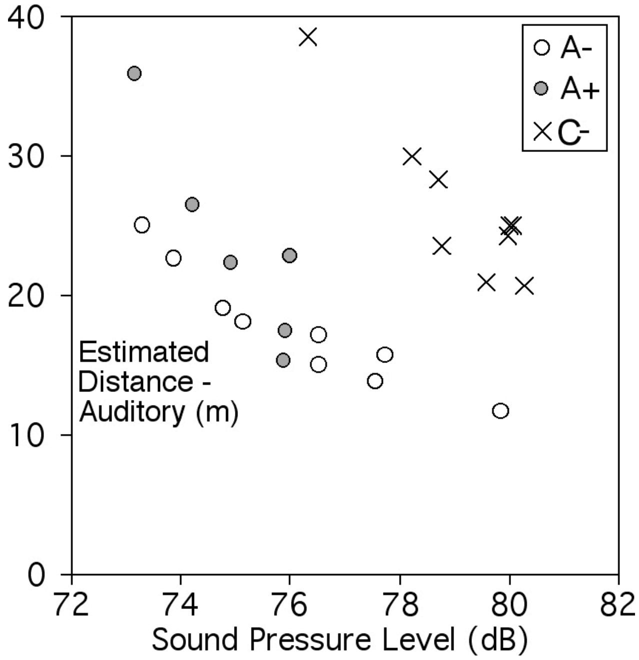

Figure 4.2. *Auditory distance estimate versus music sound pressure level (broadband unweighted L_{eq}) measured binaurally from the experiment stimuli.*

Hall C, distances are overestimated, but still correlate to actual distance (r=0.71, P=0.03). The differences in responses between the two auditoria appear to be partly due to the greater bass level achieved in Hall C, which gave the impression of a larger room (based on informal feedback given by the subjects). The use of an omnidirectional source in Hall C is also likely to have increased its distance estimates by giving a lower direct-to-reverberant sound energy ratio than a directional source. This contrast may have been increased both by the fact that Hall A was designed for greater clarity (Marshall & Hyde, 1979), and that the smaller volume of Hall C for a similar reverberation time gives it a higher reverberant field energy.

Sound pressure level is the strongest acoustical correlate of perceived distance in both auditoria, and it appears to provide some basis for the distance judgements in Hall A's gallery (Figure 4.2).

Individual visual distance estimates are much more reliable than auditory ones – which is understandable considering that subjects were able to use cues such as the number of seat rows in front of the camera. Mean visual distance estimates match real distance quite well in both auditoria for both stall and gallery seats (Figure 4.1). Refer to Appendix 2.1 for the complete statistical analyses.

4.4.2. Spaciousness, ASW and Envelopment

The results for visual spaciousness are consistent with the previous experiment (Chapter 3) in that generally ratings increase with distance, except for seats with a low overhanging ceiling (Figure 4.3). As the photographs were taken with a constant angle of view, the visible room size increases with distance from the stage. Refer to Appendix 2.2 for the complete statistical analysis.

The non-significance of results for ASW may be partly due to the use of non-individualised binaural reproduction – which, to some extent, causes the auditory image to be localised at arbitrary angles around the median saggital plane. Refer to Appendix 2.3 for the complete statistical analyses.

The lack of significance in visual envelopment ratings in either auditorium is consistent with the author's previous experiment (Chapter 3), where visual envelopment varied significantly between auditoria, but not within an auditorium. That study found Hall A to be visually more

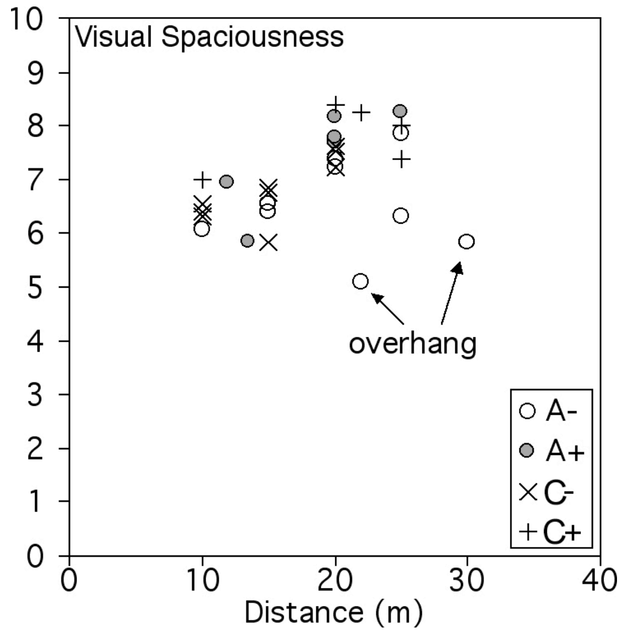

Figure 4.3. *Visual spaciousness ratings, identifying the two images with a low overhanging ceiling.*

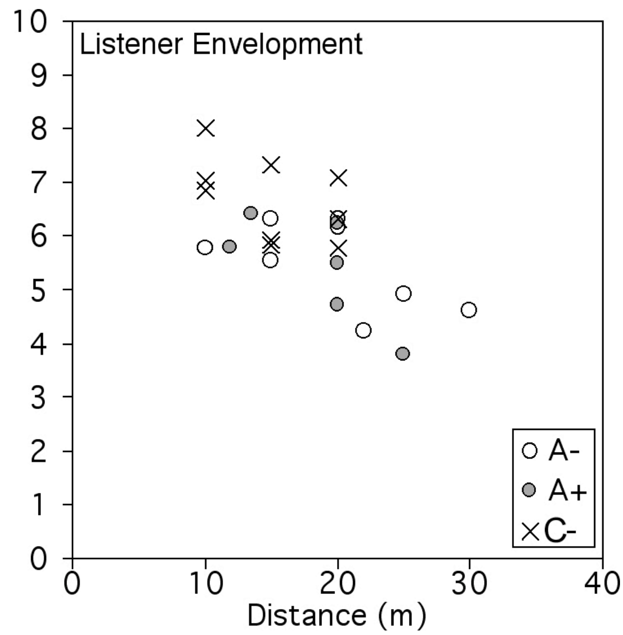

Figure 4.4. *Listener Envelopment (LEV) ratings*

enveloping than Hall B. Refer to Appendix 2.4 for the complete statistical analyses.

The experimental results presented in this chapter for auditory envelopment (LEV) shows a small decline in ratings as distance increases (Figure 4.4). In Hall A, sound pressure level (SPL) weakly correlates with LEV (r=0.68, p<0.01). Using stepwise regression, the residual is best accounted for by low frequency *IACC* – with low *IACC* values associated with increased LEV as might be expected (r=0.83, p<0.001 for the two-independent-variable model). Although LEV in Hall B is not well predicted by acoustic measurements, the Hall B results are not inconsistent with the two-parameter model derived from Hall A & B (Figure 4.5). Refer to Appendix 2.5 for the complete statistical analyses.

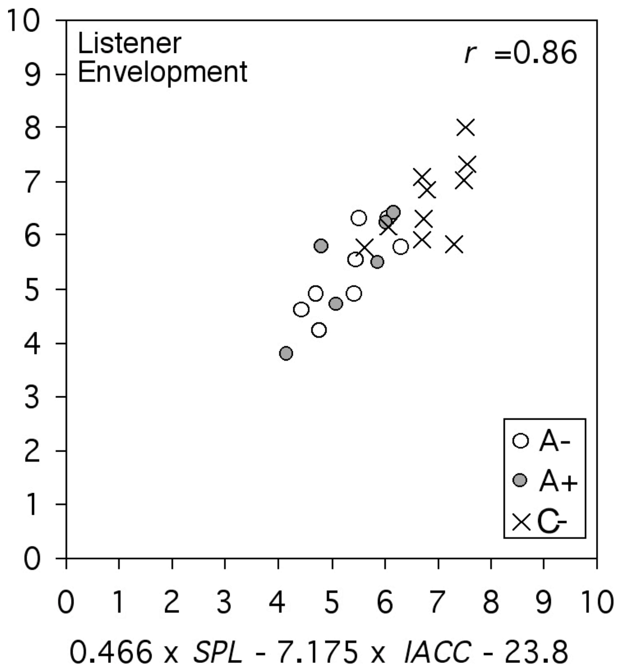

Figure 4.5. LEV regression, using sound pressure level and low frequency IACC.

4.4.3. Stage Dominance and Intimacy

Results for visual stage dominance, intimacy and distance are closely related. However, it appears that stage dominance varies little in gallery positions, while varying systematically with distance in stall positions for both auditoria (Figure 4.6). A similar effect for visual intimacy is found in Hall B, but not Hall A. Some subjects commented that the view of the organ in Hall B emphasises the stage, whereas the choral risers in Hall A distract from the stage. Refer to Appendix 2.6 for the complete statistical analyses.

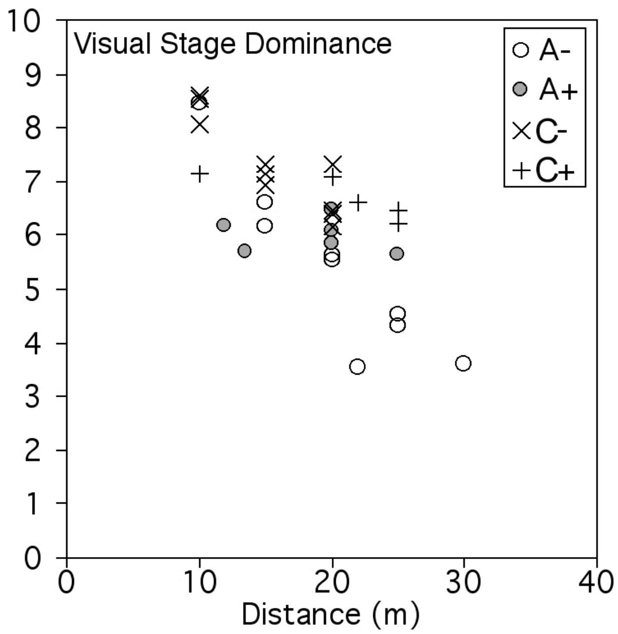

Figure 4.6. Visual stage dominance ratings.

Auditory intimacy decreases with distance. The gallery positions in Hall A exhibit less intimacy than the stall positions, contrasting with the visual intimacy results

(Figure 4.7). In both halls high-frequency (2 kHz and 4 kHz octave bands) SPL is the best acoustical correlate of auditory intimacy (r=0.90, p<0.0001 for Hall A; r=0.85, p<0.01 for Hall B). Refer to Appendix 2.7 for the complete statistical analyses.

In this and the previous study, four subjects commented that the light level contrast seems to increase visual intimacy for the more distant seating positions. However, the intimacy ratings actually decrease as stage to non-stage brightness contrast increases in both auditoria considered separately and together.

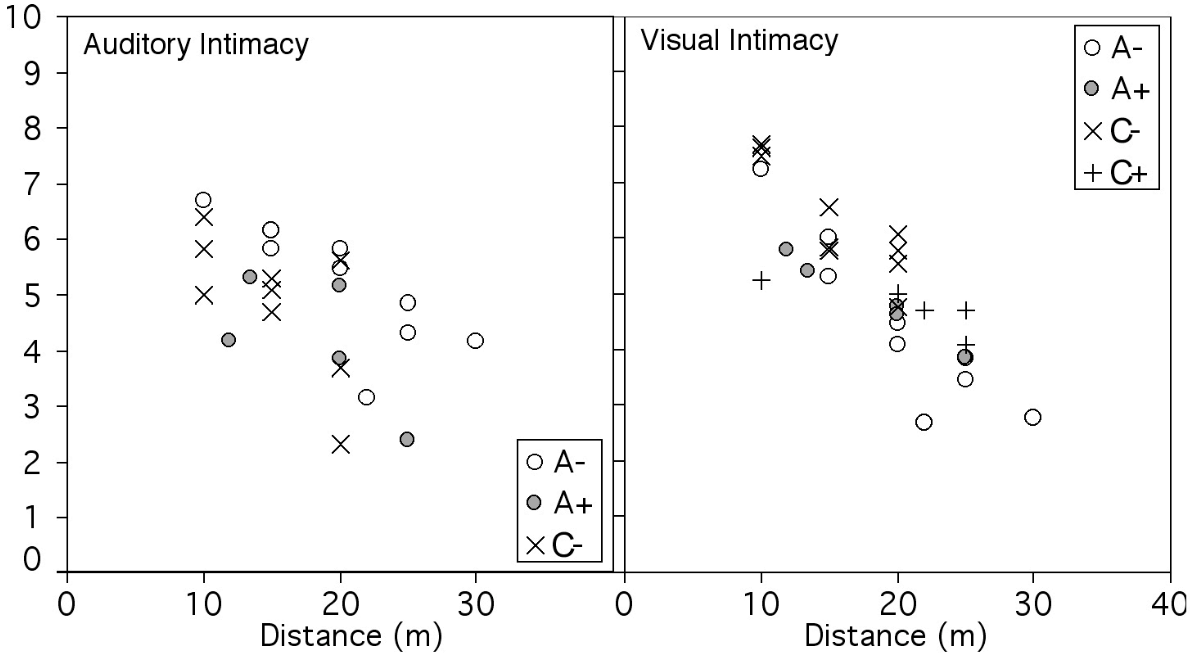

Figure 4.7*: Auditory and visual intimacy ratings*

Seat Number	Hall	Distance from Source	Angle from Midline	Gallery or Stalls
23D6	A	12 m	60°	Gallery
24B6	A	13.5 m	35°	Gallery
24F4	A	20 m	35°	Gallery
25D6	A	20 m	25°	Gallery
26C9	A	20 m	10°	Gallery
26H3	A	25 m	0°	Gallery
H32	A	10 m	0°	Stalls
O34	A	15 m	0°	Stalls
O40	A	15 m	15°	Stalls
V26	A	20 m	15°	Stalls
W34	A	20 m	0°	Stalls
W08	A	22 m	45°	Stalls
AA25	A	25 m	25°	Stalls
BB35	A	25 m	25°	Stalls
HH44	A	30 m	15°	Stalls

Table 4.5: Measurement positions represented by the auditory stimuli.

4.5. AUDITORY PERCEPTION OF ROOM SIZE

Chapter 2, in this study, contains some discussion on visual perception of room size and how it affects the sense of visual intimacy and other visual spatial impressions. A very recent study by Densil Cabrera et al (2005) generated significant relevancy to the topic of this thesis. Cabrera et al (2005) explore the auditory perception of room size and begin partly with the background in distance perception. Similarly, this current study suggests that distance perception is an important subjective aspect of spatial impression in auditoria. They hypothesised that perceived room size and physical room size are likely to be related due to past experiences that enable the observers to interpret auditory room size and perceived distance. Thus, perceived distance is a likely factor affecting judgement of auditory room size. The following is a brief description of their study and some discussion of the relevance and relationship to the auditory and visual spatial impression explored thus far.

4.5.1. Method

Cabrera et al's (2005) study consists of several experiments involving a series of real and computer-modelled rooms. However, for the purpose of relevancy, this study discusses only one of the three experiments. The aim of this particular experiment is to study how subjective judgment of auditory room size may vary between different locations within a concert auditorium, and to consider whether auditory room size perception can be distinguished from auditory distance perception where the reverberation time is approximately constant, and the room volume is constant. This experiment in the auditory perception of room size uses the same binaural impulse responses recorded in Hall A (Wellington, New Zealand). The audio stimuli were created using an anechoic recording of the opening 15 seconds of the *Overture to the Marriage of Figaro* (Mozart) (Dennon, 1993a). 15 different source-receiver positions were used (table 4.5). 14 subjects, all music students in their mid-twenties, participated in this experiment. The concept of this experiment was explained to the subjects to ensure meaningful responses to the scales. Each subject judged each of the 105 pairs once, and the intra-pair order was counter-balanced between subjects. The tests were presented in an anechoic chamber using the convolved anechoic music recording (Mozart) with the impulse responses as mentioned above. Listeners were asked to give their perception of room size given two paired samples. The question for the paired comparison was "Which room is bigger?" Listeners judged the room size wearing the headphones. The stimuli were presented using a pair of Sennheiser HD 600 headphones, and the paired order was presented and chosen randomly by a computer program. It was also used to collect the participants' responses.

4.5.2. Results

Results were scaled using Thurstone's Case V method. The results (figure 4.8) show an increase in room size from the front stalls to the real stall, and perhaps to the gallery. Several results seemed to stand out. H32, which is the closest seat to the sound source (and also directly in front of it), received the smallest room size rating compared to all other stimuli. 23D6, 26H3 and W08 received room size ratings larger than the majority of stimuli. 23D6 and 24F3 are two seats with close proximity, however 23D6 received markedly greater rating compared to 24F3.

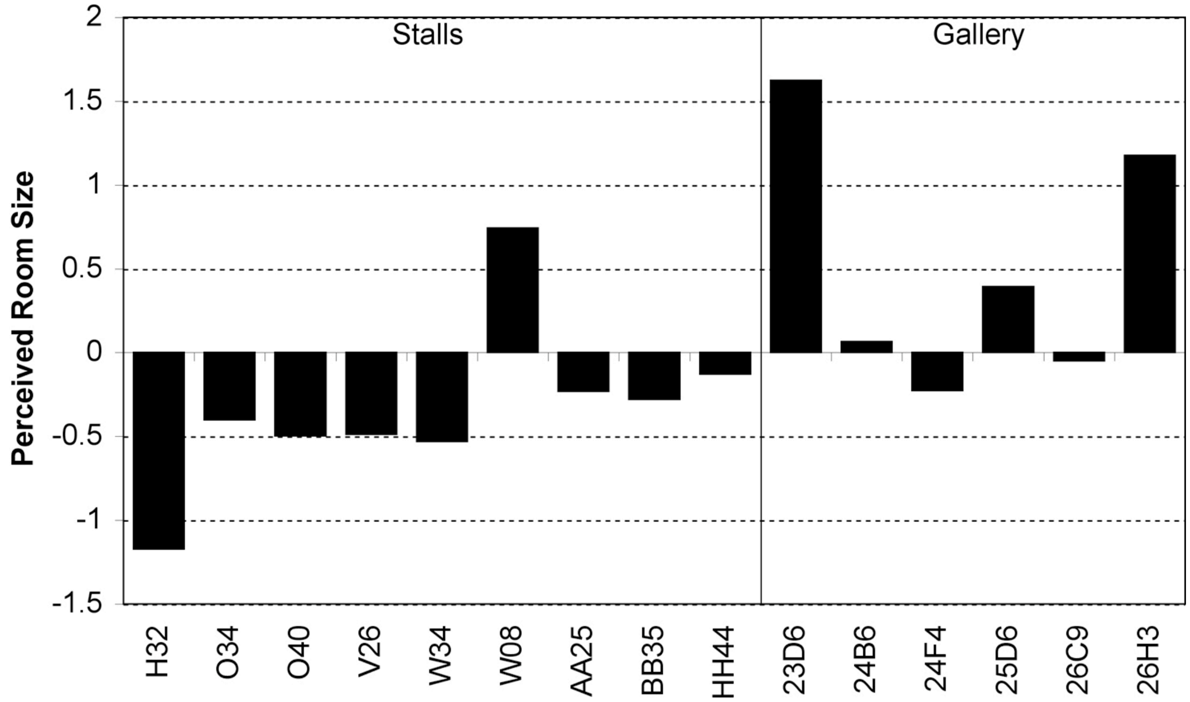

Figure 4.8: *Ratings of room size for the Hall A stimuli. For both the stalls and the gallery, the results are presented from left to right in order of source-receiver distance.*

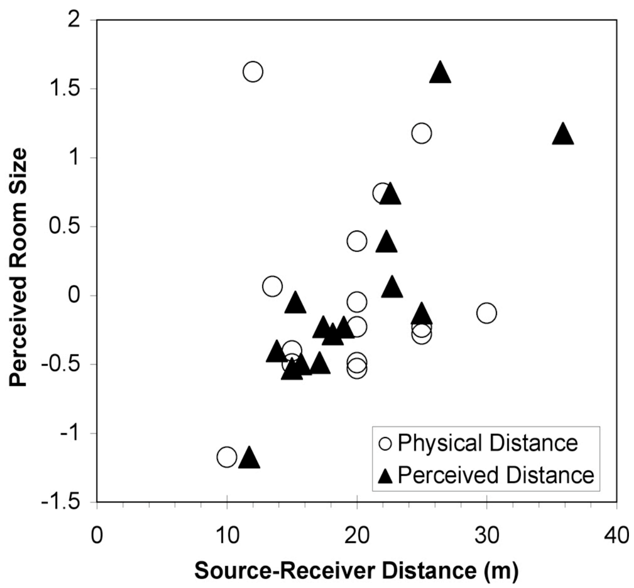

Figure 4.9: *Relationship between perceived room size and both physical and perceived (auditory) distance in Hall A.*

While there appears to be some relationship between source-receiver distance and perceived room size, the result for 23D6 in particular undermines the correlation (the correlation is insignificant, at r=0.11, p=0.69). Figure 4.9 shows that when perceived auditory distance (from the preceding experiment – refer to section 4.4.1) is substituted for physical distance, the relationship with perceived room size strengthens greatly (r=0.83, p<0.0001).

In the preceding experiments discussed in this chapter, auditory intimacy was rated on an integer scale from 0 to 10. The results were closely related to auditory distance estimates (r=-0.88, p<0.0001), and they appear to be very closely related to the perceived room size results (Figure 4.10). The result for seat 23D6 undermines what would otherwise be a correlation of r=-0.92 (p<0.0001), yielding a correlation of r=-0.81 (p<0.0001).

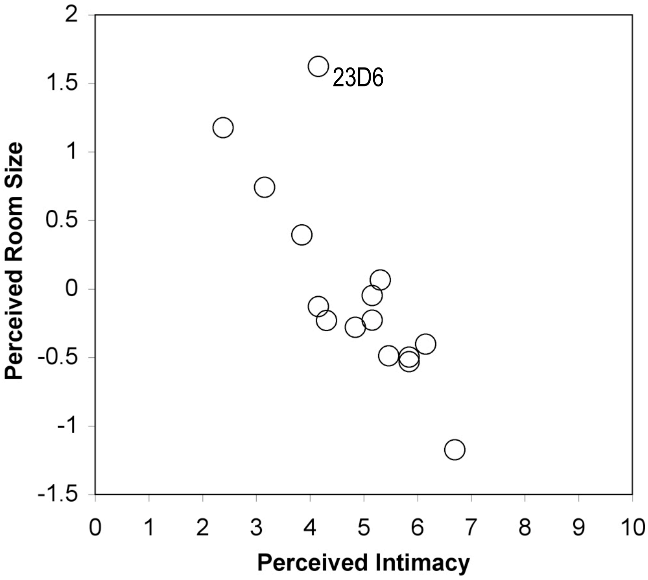

Figure 4.10: *Relationship between perceived room size and perceived (auditory) intimacy in Hall A*

4.6. DISCUSSION

4.6.1. General

There are clear limitations to this study, both with the auditory and visual experiments. The auditory experiment used non-individualised binaural presentation, resulting in localisation artefacts. There are also discrepancies between the stimuli of the two auditoria – with Hall B having a substantially bassier sound quality than Hall A due to the measurement system. In the visual experiment, experience was deliberately restricted to greyscale images to simplify the cues. The images did not use peripheral vision, which makes a substantial contribution to spatial impression (Danahy, 2001). Projected images also exclude cues from binocular convergence and monocular focussing. Both experiments present an unnaturally static environment, de-coupled from the subjects' body movement. Despite these and other limitations, the study does yield results that seem to be reasonable pointers for more general application.

4.6.2. Auditory and Visual Spatial Impression

This study raises the question of the degree of correspondence between auditory and visual spatial impression. Such correspondence could be necessary (i.e. due to common determinants) or arbitrary, and it could pertain to the structure or the content of spatial impression.

The concept of distance is simple, and easily defined in both auditory and visual modes – even if cues for its perception are complex and subtle. In previous studies, some structural commonality can be seen, with a tendency for underestimation, especially in the far field – implying a context-dependent horizon, beyond which all stimuli are judged equally distant. However in contexts such as that studied here, auditory distance cues are much less robust than visual distance cues. Auditory distance perception of a known source (eg the orchestral sound used in this study) relies primarily on sound intensity, its spectral balance and the direct to reverberant ratio. Visual distance cues available in this study include perspective, the number of objects of known size between the camera and target, the relative size of the target to other elements of the image, and many others.

Within most real contexts, auditory and visual distance should co-vary because, even though the cues are different, they originate from the form of the physical environment. Nevertheless, they can diverge greatly in situations such as whispering walls, or optical magnification.

Visual Spaciousness was considered in this study mainly because 'spaciousness' is sometimes used as a term in auditory spatial impression (auditory spaciousness is related to ASW) (Beranek, 1996, p23, 2004, p29). However, as defined in this study, 'visual spaciousness' refers to impression of room size, which had no direct equivalent in the auditory experiment. There has been little research on the auditory perception of room size, but it seems likely that reverberation is a major contributor (although being confounded by the absorptive qualities of room surfaces). With the direct-to-reverberant sound energy ratio decreasing with distance from a source, the perceived room size may be expected to increase with distance – which was the pattern found for visual spaciousness. However, over the range of possible room surface acoustic absorption conditions, auditory and visual room size should diverge considerably.

A closer visual analogue of ASW is Stage Dominance. This study found neither a necessary nor arbitrary relationship between these, since the results for ASW were not significant. Even in theory, any relationship is likely to be largely arbitrary, as Stage Dominance depends on the proximity and angle of view to the stage, whilst ASW depends more on the strength of lateral reflections at the listener position (and hence is not directly related to the stage).

There seems to be some prospect for correspondence between visual and auditory Envelopment, in that both vary more between auditoria than within an auditorium (based on previous studies). However the physical features that engender envelopment in the two senses seem unrelated. In vision, envelopment may come from roundness – hence the high visual envelopment ratings of Hall A in the previous experiment (Chapter 3). In audition, envelopment comes from the late lateral sound level, which is related in a more complex manner to the auditorium shape, as well as to surface absorption and diffusion. Nevertheless, with envelopment as a desirable characteristic of auditoria, it is possible to design spaces that optimise visual and auditory envelopment.

With a close relationship to perceived distance, auditory and visual Intimacy are likely to co-vary in a given context. This study was not sensitive enough to make explicit the subtle differences between intimacy and perceived distance. This may be pursued in future work.

4.6.3. Auditory spatial impression and perceived room size

Findings in Cabrera et al's study of auditory perception of room size are important in relation to auditory spatial impression discussed in this chapter, as perceived auditory distance, room size, and Intimacy appeared to be related to, and somewhat predictable by distance. Perhaps, a comparative analysis between auditory Spaciousness and perceived auditory room size would be more appropriate, as reverberation time is apt to have a strong effect on both auditory spatial judgment. Since auditory Spaciousness has been a matter of debate in the acoustics, and ASW is used in this thesis, perhaps perceived auditory room size would be an appropriate auditory counterpart to visual spaciousness.

4.7. CONCLUSION

Space perception in vision and audition differ because of the physical characteristics of light and sound, their different interactions with the physical environment, differences between the sensory organs of the two modalities, and differences in the neural processing and cognition of these modes. Physically, the wavelength ranges of light and sound are in sharp contrast – with that of light spanning less than an ‘octave’ at very short wavelengths, and that of sound spanning a ten octave range, between values substantially larger to substantially smaller than the human body. Diffraction and resonance are perhaps of esoteric interest in the optical world pertaining to vision, but are core components of the acoustic world pertaining to audition. If body movement is disregarded, vision is restricted to the view ahead of the person, and foveal vision covers an even smaller angle of view. By contrast, audition has an unlimited ‘angle of view’, and ‘focussing’ is achieved through signal analysis rather than by relying on physical movement. The perception of direction and object boundaries has far greater ‘blur’ in audition than vision. There is some connection between auditory and haptic perception, with certain sounds being felt as well as heard – in this way audition has a stronger somatic association than vision. These and other comparisons might be taken to imply that audition is better at aspects of spatial impression than vision (especially envelopment, and perhaps intimacy), with vision being better at interpreting the geometry of the world around us. Nevertheless, the idea of at least a partial correspondence between the spatial impression gleaned from vision and from audition is appealing, and useful for applications in auditorium acoustics.

This paper presents a limited case study using reduced optical and acoustical representations of two real spaces. Although the concept of a ‘necessary’ relationship between corresponding attributes of auditory and visual spatial impression may overstate any relationship’s basis, there are some attributes that are likely to co-vary in typical real-world contexts. In synthetic contexts, such relationships can be either exploited or discarded. Control of the degree of correspondence and contrast between visual and auditory spatial impression should allow the design of a rich diversity of spatial experience.