8 Evaluation of the resulting subjective descriptors and objective measurements

The research and experimentation that has been described so far in the thesis has resulted in the elicitation of a number of descriptors of perceived auditory spatial attributes, and a number of novel objective measurement techniques that may relate to these perceived attributes. It was proposed in Chapter 7 that these findings could be examined in a subjective experiment where the descriptors that had been elicited in the previous subjective experiments were used as judgement scales for grading the spatial attributes of a number of items of programme material, after which the objective measurements could be compared with the subjective results.

This chapter describes the experiment that was undertaken to evaluate the subjective descriptors and the objective measurement techniques that have resulted from the research that is contained in this thesis. The aims of the experiment are summarised, and then the stimuli are discussed in detail. The experimental method that was used is outlined, and the procedure that was used to conduct the experiment is specified. The analysis and results of the subjective judgements are then discussed, and conclusions are drawn from these. The objective measurements are then described, and the results of these measurements are stated. The objective and subjective results are then compared, and conclusions are drawn regarding the success of the different measurement techniques that were employed. Finally, the results of the experiment are discussed, where the merits of the subjective descriptors are evaluated, and the limitations of the objective measurements are considered.

8.1 Aims of the experiment

The first aim of the experiment was to evaluate the descriptors that had been elicited in the subjective experiments that were described in Chapter 4 to Chapter 6. By using these descriptors as grading scales in an experiment to evaluate the spatial attributes of a number of auditory stimuli, it could be ascertained whether the subjects could relate to the terms, whether they found the terms meaningful, and whether they could discriminate between the two spatial attributes that the scales represented.

The second aim of the experiment was to evaluate the developments in the objective measurement techniques that resulted from the research and experimentation that is contained in this thesis. By comparing the measurements that are based on these results with the extant measurement techniques, a judgement could be made on whether the developments resulted in objective measurements that matched the subjective judgements more closely.

The third aim of the experiment was to ascertain whether the perceived width of an attribute of the scene was still related to the physical parameters of the cross-correlation or the time difference fluctuations for stimuli that were more externally valid than the noise signals that were used in the elicitation experiments.

The final aim of the experiment was to gain an indication of the perceptual mechanism that is employed in detecting the spatial attributes of auditory stimuli. It is likely that the measurement technique that is closest to the perceptual mechanism will predict the subjective results most accurately.

8.2 Stimuli

In order for the stimuli that were used in this experiment to be more externally valid, they had to be representative of the type of programme material that would be heard in a concert hall or replayed through a sound reproduction system. The stimuli also had to contain a range of spatial attributes, so that the properties that relate to the two grading scales of perceived source width and perceived environment width were stimulated independently.

The simplest method that was available for manipulating the spatial properties of the stimuli in a manner that would be externally valid was to use a number of different microphone techniques. This would allow the manipulation of the spatial properties of the reproduced sound in a controlled manner, and in a way that is externally valid for surround sound reproduction. However, if pre-recorded stimuli were employed, then this would have introduced a large number of variables that would have made the subjective judgement and objective measurement more difficult. For this reason it was decided to specifically create controlled stimuli for this experiment, so that the microphone technique was the only variable between the stimuli. A number of discrete five-channel microphone techniques were chosen, based on attempting to create stimuli with a wide range of spatial properties. In order to reduce the number of variables in the experiment, these were limited to simple arrays.

It was decided that a single sound source would be most appropriate for this experiment, as it would be difficult for the subjects to determine the exact width of each source if more than one was sounded simultaneously. Also, the measurement techniques that have been developed so far can only quantify the width of a single source. This source was sounded in a single reverberant environment, as the use of a number of environments would have introduced a large number of variables into the experiment. This would have complicated the task and made the experiment longer to undertake.

The stimuli were reproduced over a standard five-channel loudspeaker array, as specified in [ITU-R BS 775 1994] and shown in Figure 8.1.

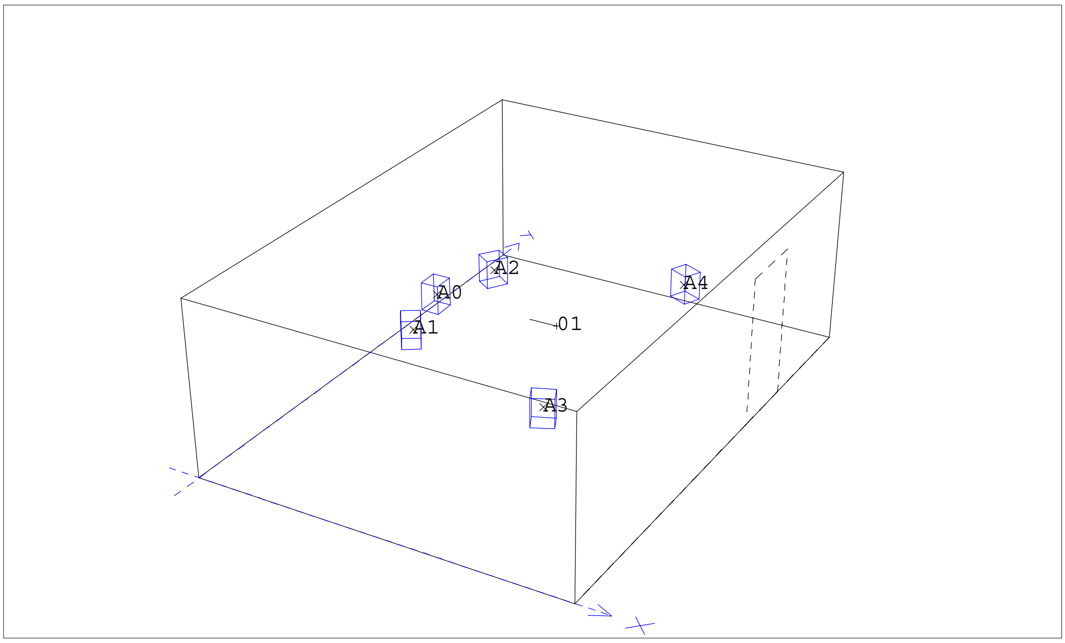

Figure 8.1: Three-dimensional representation of the listening room, showing the location of the listening position (labelled 01) and the five loudspeakers in the standard five-channel arrangement, one positioned directly in front of the subject (labelled A0), two positioned at ±30º from directly in front of the subject (labelled A1 and A2) and two positioned at ±110º from directly in front of the subject (labelled A3 and A4).

In an attempt to independently vary the perceived source width and the perceived environment width, a number of different front and rear microphone arrays were combined to create a number of five-channel microphone techniques. The author conjectured that if there only was a single source signal directly in front of the microphone array, then variations in the microphones that feed the front loudspeakers would principally affect the perceived source width and the microphones that feed the rear loudspeakers would principally affect the perceived acoustical environment, as discussed in [Mason and Rumsey 2001]. Whilst it has been shown that these factors do interact with each other and therefore should not be considered separately in this manner, it is a useful grouping in order to analyse certain attributes of microphone techniques.

In order to create the perception of a narrow image, a single monophonic microphone could be used to feed either a single loudspeaker or a pair of loudspeakers that are positioned symmetrically around the median plane. This would mean that the signals that reach each ear would be similar, giving minimal fluctuations in interaural time and level difference and a high interaural cross-correlation coefficient at the listening position. According to the research that was discussed in this thesis, this should result in a relatively narrow perceived auditory image. In order to create the perception of a wide and diffuse image, a number of widely spaced microphones could be used to feed a number of loudspeakers. This would result in differences in both level and time between the channels that feed the loudspeakers, which causes a low correlation between the signals in each of the channels [Lipshitz 1986]. Based on the research that was discussed in Chapter 1, it is apparent that this would cause a low interaural cross-correlation at the listening position that would result in a relatively wide perceived auditory image. As a medium point between these two extremes, a coincident pair of microphones could be used to feed a pair of loudspeakers. This would cause the direct sound from the source to be highly correlated in the loudspeaker channels, though the early reflections and the reverberation of the recording acoustical environment would be somewhat decorrelated due to the amplitude difference between the channels that is caused by the reflections that arrive from away from the median plane. This would result in the direct sound being perceived as a relatively small auditory image, but with the source broadening that is caused by the addition of the lateral reflections, as discussed in Chapter 1. These microphone techniques are based on the well-established principles of stereophonic microphone placement [Rumsey 2001].

Based on this, the front microphone techniques consisted of the following. For the first front microphone array, a single monophonic omnidirectional microphone was fed solely to the front centre channel. For the second front microphone array, a pair of coincident figure-of-8 microphones that were pointing ±45° from the source in the horizontal plane was fed to the front left and right loudspeakers. For the third front microphone array, three spaced omnidirectional microphones were fed to the front left, centre and right loudspeakers. The microphone arrangements that were used to feed the rear speakers were similar to those that were used for the front. For the first rear microphone array, a single monophonic omnidirectional microphone was fed to both the rear left and right loudspeakers. For the second rear microphone array, a pair of coincident figure-of-8 microphones that were pointing ±135° from the source in the horizontal plane was fed to the rear left and right loudspeakers. For the third rear microphone technique, two spaced omnidirectional microphones were fed to the rear left and right loudspeakers.

The front omnidirectional microphone was positioned to be within the calculated critical distance of the source in the acoustical environment, and was 4 metres directly in front of the source. The rear omnidirectional microphone was positioned to be outside the critical distance, and was 8 metres directly behind the front monophonic omnidirectional microphone. In order to confirm that the front-to-rear separation of these microphones was sufficient so that the source would be clearly perceived to be in front of the subject, the results of this microphone arrangement were auditioned by the author. After this, the other front and rear microphone techniques were positioned in order to create a similar direct to reverberant sound ratio, as determined subjectively by the author. This resulted in three front microphone techniques and three rear microphone techniques. These were combined to create nine different five-channel microphone techniques in total.

If these stimuli had been recorded using conventional microphones, then there would have been inevitable differences in the frequency and temporal response of the different types of microphone that are required for the different polar patterns. In addition, it is likely that the polar pattern of the microphones would not be ideal at all audio frequencies. This would have meant that the perceived differences between the microphone techniques would not purely be due to the selected positions and polar patterns, and the variation in the frequency response might have made the loudness alignment of the stimuli more difficult. To avoid this problem, the microphone techniques were simulated in CATT-Acoustic. The use of a simulated source, acoustical environment and microphone array also allowed more accurate placement of the source and the microphones. Finally, it helped to eliminate any additional variables, such as background noise and interference that would be caused by the simultaneous placement and use of multiple microphone arrays.

The acoustical environment that was simulated in CATT-Acoustic was a simple shoe-box room with the same overall dimensions and reverberation time as the Vienna Grosser Musikvereinssaal. In order to excite the early reflection pattern as much as possible, the source was simulated as an omnidirectional source. It was positioned on the centre line, 4 metres from the rear wall, and at a height of 1.8 metres. The microphones were placed symmetrically around the centre line and were spaced from the source based on the calculated critical distance, as mentioned above.

The stimuli were created by calculating the impulse response from the source to each of the simulated microphones. To imitate each musical instrument being sounded in the simulated room, these impulse responses were then convolved with the sound source signals.

In order to eliminate any effects that could be caused by spatial or acoustical information that is contained within the source signals, extracts were required that had been recorded monophonically within an anechoic environment. The most readily available source of anechoic recordings was the Bang and Olufsen CD that contains anechoic recordings made for the Archimedes project. The recording of this is well documented in [Hansen and Munch 1994].

In order to test whether the elicited descriptors and objective measurement techniques were applicable to a wide range of programme material, a number of sound sources that contained a variety of properties were selected. These were chosen to include examples such as transients, sustains, wide-band and narrow-band (tuned) signals, a wide range of frequencies, and a human voice. It was also important that there were sufficient gaps in the extracts, so that it was possible to hear the effect of the acoustical environment.

The extracts that were used from the B&O CD contained a cello (sustained, tuned, low frequency) and a trumpet (mixture of transient attacks and sustains, tuned, mid-high frequencies). Two additional extracts were recorded in the free-field room at BBC Research and Development in Kingswood Warren, UK. These contained a snare drum (transient, wide frequency range, separated hits) and a male speaking voice (a mixture of noise and modulated tonal sounds - a popular test item).

The recordings were made in mono with a Brüel and Kjær 4006 omnidirectional microphone that was connected via a custom pre-amp and phantom power supply to a Tascam DA-30 DAT recorder using the internal analogue to digital converters. The aim of the recording was to produce a result which, when replayed over a loudspeaker, would sound as natural as possible. In order to do this, the recording was monitored on a single large loudspeaker and compared with the natural sound from the source.

It has been found that it is easier and more efficient to judge audio signals that are stationary and possibly repetitive [Rumsey 1998] and [Olive et al. 1994]. In view of this, the snare drum and trumpet excerpts were made up of a short loop of a bar or so. To match the duration of the other extracts, this loop was repeated for approximately 60 seconds.

The relative reproduction level of each of the sound sources is also important in recreating them as accurately as possible. Using a Brüel and Kjær SPL meter with a Brüel and Kjær 4145 1-inch capsule, A-weighted SPL measurements with a fast time constant were made of an example of each sound source that was represented. From this, the relative level of each source was calculated, and the level difference was maintained when creating the stimuli. In this way, the stimuli could be reproduced in the listening room at approximately the same level as they would sound in an acoustical environment.

This method of creating the stimuli was necessarily a compromise, as there was no longer a real source sounding in a real acoustical environment. The disadvantages of this approach were due to the artificiality of this simulated environment, source and receiver. For the early part of the reflection pattern of the simulated environment, CATT-Acoustic employs a prediction method that is based on cone tracing [Dalenbäck 1996]. Whilst this is a reasonable method for simulating diffuse reflections, it is not an accurate model, especially when considering higher order diffuse reflections. The reason for this is that the paths of the specularly reflected cones are traced, whilst the diffusely reflected energy is not traced due to the large amount of computation that is required for this task [Dalenbäck et al. 1994]. The late part of the reflection pattern is modelled less accurately in CATT-Acoustic, and is based on a simplified statistical model that takes into account the basic room shape. In view of this, a very large number of cones were employed in the prediction, in order to maximise the accuracy of the simulation. The use of a large number of cones extends the duration of the part of the model that is more accurately simulated, therefore lessening the effect of the part of the model that is less accurately simulated.

The source was simulated with an omnidirectional directivity pattern, which is similar to the directivity of a snare drum, but very different to the directivity of a trumpet. This pattern was used in order to excite the early reflection pattern of the room as much as possible. However, this directivity pattern was not representative of the range of source signals that were employed. The source was also modelled as a perfect omnidirectional source, which means that the frequency response and directivity were perfect, which is not an accurate simulation of practical sound sources. In addition, CATT-Acoustic does not simulate the physical coupling of the source to the air in a manner that accurately represents each of the sound source signals. However, factors such as the timbre, attack, decay and musicality of the sound source signals should be reproduced effectively by this simulated sound source.

The simulated receivers were also unlike real microphones. The elimination of some of the variations between the different microphone types that would have been present with the use of real microphones was deliberate, as mentioned above. However, the polar patterns of the simulated microphones were identical at all frequencies, which is unusual in real microphones, and the frequency and temporal responses of the simulated receivers were ideal and therefore unlike practical microphones.

The stimuli were created by convolving the anechoic sound source signals with the impulse responses that had been calculated using the simulated source, room and receivers. In view of the limitations of the simulation, as discussed above, it is apparent that the experiment would not be applicable as a study of the perceived effect of microphone techniques or the attributes of acoustical environments. However, the aim of this method was to create stimuli that were similar to typical programme material, though with variations solely in the perceived auditory spatial properties of the sound reproduction, whilst the sound source and recorded acoustical environment were kept constant. Audition of stimuli that were created by this method indicated that they were convincing simulations of recordings of a source that was sounded in a reverberant environment, whose spatial properties varied with the different microphone techniques that were simulated. In this way, they were suitable for use in this experiment.

To summarise, nine microphone techniques were used in the creation of the experimental stimuli. They were made up of a combination of three microphone arrays that fed the front channels and three microphone arrays that fed the rear channels. For each of these nine microphone techniques, there were four sound source signals, resulting in a total of 28 stimuli.

8.3 Method

This experiment aimed to evaluate the stimuli by using grading scales that were based on the subjective attributes that had been elicited in the previous subjective experiments. Due to this, the required methodology was different to that which was employed for the elicitation experiments that were described in Chapter 4 to Chapter 6.

For this experiment, the stimuli were presented to the subjects in sets of nine, which included all the microphone techniques for a single sound source signal. This method was used as it is more rapid to undertake than paired comparisons, yet it still allowed the comparison between the stimuli, which enables small differences to be detected and graded in a reliable manner [Bech 1990]. The two grading scales were presented on the same screen, one for source width and one for environment width.

The subjective attribute that was associated with the first grading scale was described to the subjects as follows:

> ‘For source width, you should indicate how wide you perceive the sound source (the musical instrument or voice) to be, measured in degrees from the left of the source to the right of the source at its widest point. Calculate the total angle the sound source appears to cover by adding the angle of the left hand side of the source (in degrees from the centre) to the angle of the right hand side of the source (in degrees from the centre).’

In order to assist the subjects in evaluating the source width, visual angle markers were placed below the loudspeakers, with the azimuth from directly in front of the listening position marked in gradations of 5º.

The subjective attribute that was associated with the second grading scale was described to the subjects as follows:

> ‘For environment width you should indicate how wide or spacious you perceive the reproduction of the acoustical environment or concert hall to be. This scale is from 0 to 100, with a score of 0 indicating that the hall could not sound any narrower and a score of 100 indicating that the hall could not sound any wider.’

Unfortunately it was not possible to give visual markers for the environment width, as it was expected that the perceived width of the reproduced acoustical environment would be larger than the listening room.

The subjective attributes were also described by the use of a graphical depiction. This was based on the results of the sketch-map elicitation methods that were used in the previous experiments, and is shown in Figure 8.2.

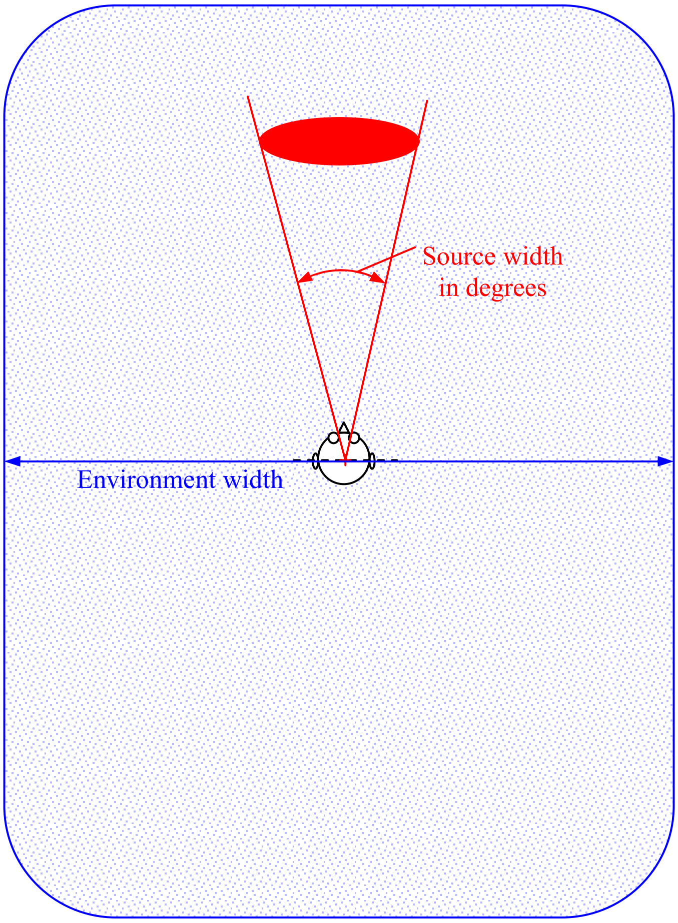

Figure 8.2: The depiction that was given to the subjects as a descriptor to represent the meaning of the two grading scales that were used in the subjective experiment.

In addition to the verbal and graphical descriptions that are shown above, a number of exemplary stimuli from the previous elicitation experiments were provided to demonstrate the subjective attributes that related to each grading scale. For this, the subjects were given narrow and wide examples of both source width and environment width. The stimuli that demonstrated source width were from the experiment that was described in Chapter 5. They were the continuous noise signals with the lowest and highest magnitudes of inter-loudspeaker time difference fluctuations that were presented over the loudspeakers that were positioned at ±30º. The stimuli that demonstrated environment width were from the experiment that was described in Chapter 6. They were the decaying noise signals with the lowest and highest magnitudes of inter-loudspeaker time difference fluctuations that were presented over the rear loudspeakers in addition to the centre loudspeaker.

The reproduction of the stimuli was controlled by using a computer. The screen simultaneously displayed the play buttons and the two grading scales for each of the nine stimuli, as shown in Figure 8.3. The subjects were free to switch between the stimuli whenever and as often as they required. A selected stimulus would loop continuously until the pause button was clicked. The switching between the stimuli was set to be synchronous, meaning that the stimulus that was selected would not start from the beginning, but would continue from the point that the previous stimulus had reached. The judgements were made on the grading scale by clicking and dragging the marker on the appropriate slider. When the subject had graded all the stimuli and was satisfied with the grades, clicking ‘Next’ would move them to the next set of stimuli.

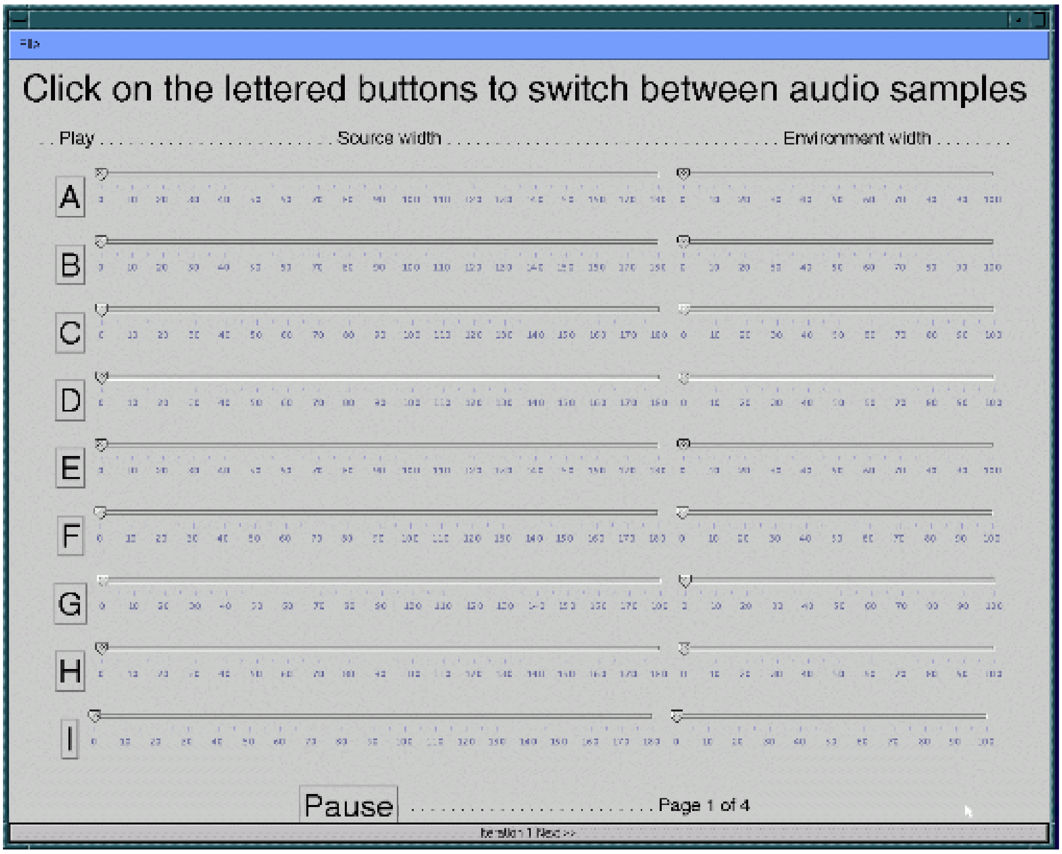

Figure 8.3: Screen shot of the user interface that was employed for the subjective experiment, showing the lettered buttons for switching between the stimuli and the two grading scales for each stimulus.

The subjects were free to grade the stimuli in any order, but they were encouraged to listen to all the stimuli before commencing the grading. In order to eliminate any effects data that may have been caused by using a single order of presentation [Meilgaard et al. 1991], the stimuli were randomised, which meant that they were presented to each subject in a different random order.

8.3.1 Physical set-up

The experiment was carried out in an ITU-R BS 1116 standard listening room. The loudspeakers were arranged in the conventional five-channel configuration, as specified in [ITU-R BS 775]. This meant that they were positioned at 0º, ±30º and ±110º from the frontal median plane, at a distance of 2 metres from the subject, as shown in Figure 8.1.

The computer monitor was positioned on a low stand in front of the subject so that it did not obscure the path of the direct sound from any of the loudspeakers to the subject. A mouse was used to control the replay system, and this was placed on a mouse pad that was attached to the arm of the chair on which the subject was seated.

The reproduction of the experiment stimuli was carried out using custom listening test software that ran on a Silicon Graphics O2. The ADAT output of the Silicon Graphics machine was connected to a Yamaha 02R for routing and digital to analogue conversion, and the analogue outputs were then connected to Genelec 1032A loudspeakers, which were arranged as mentioned above. The loudspeakers were level aligned to within ±0.1 dBA by the use of a pink noise generator and an omnidirectional microphone that was positioned at the centre of the listening position that was connected to a Brüel and Kjær 2123 real-time analyser.

The author subjectively set the reproduction level to a comfortable listening level, where quieter parts of the stimuli could be clearly perceived without the louder parts of the stimuli being uncomfortably loud. The measured level offsets between the individual sound source signals that were taken into account when creating the stimuli, as described in Section 8.2, were maintained in the reproduction. This meant that the relative reproduction level of each of the sound source signals was similar to how it would be perceived in an acoustical environment.

8.3.2 Loudness alignment

As mentioned in Chapter 4, the loudness of an auditory signal can have a large effect on how it is perceived. For this reason, each of the different microphone techniques for each sound source signal was aligned to be of similar loudness using the Moore loudness model [Moore et al. 1997]. However, the loudness offset between the different sound source signals was maintained as mentioned above.

The loudness alignment involved the reproduction of all the stimuli in the listening room, and the measurement the average level over the duration of the each stimulus in 1/3rd octave bands. These data were then entered into the Moore loudness model to calculate the approximate subjective loudness in phons. The reproduction level of the stimuli was then aligned based on these data, after which the loudness was measured again. This procedure was repeated until the stimuli were measured to be within ±0.1 phons.

It must be noted that the Moore loudness model was developed for application to static signals, which means that the use of this model for quantifying the loudness of dynamic stimuli, such as musical programme material, was improper. However, previous experience had shown that, for stimuli with similar characteristics (i.e. the same musical extract) and a similar frequency response (within a few dB in each 1/3rd octave band), this method of alignment results in stimuli that are subjectively the same loudness [Mason and Rumsey 2000]. The results of the loudness alignment that was conducted for this experiment were confirmed by audition by the author.

8.4 Experimental procedure

The training that the subjects were given was deliberately limited in order to examine how intuitive they found the grading scales to be. Therefore the training solely consisted of a printed sheet that contained the description of the experiment and the verbal and graphical descriptions of the meaning of the grading scales, in addition to the exemplary auditory stimuli, as described in Section 7.2.

This experiment was not an expert elicitation exercise, unlike the previous experiments. In order to increase the likelihood of a statistically significant result, and to be able to infer the results to a wider population, a larger panel of subjects was used. The subjects that were used were still expert subjects, as they are more familiar with critical listening and analysing the attributes of auditory stimuli than naïve subjects. This means that it was likely that there would be less error in the results compared to if naïve listeners had been used, as was found by Bech [Bech 1992]. It has been shown that, for experiments that involve qualitative judgements such as preference, there can be large differences between the results of expert and naïve subjects [Kirk 1956]. However, as this experiment was an exercise in detecting and grading the perceived quantitative spatial attributes of the stimuli, the author considered that it was likely that there would be little difference between the mean results of expert listeners and naïve listeners.

27 subjects were used in the experiment. They were either students, graduates, postgraduates or staff in the Department of Sound Recording at the University of Surrey. The experiment took an average of approximately 25 minutes to undertake. The subjects were not informed of the nature of the programme material they were auditioning or whether any processing was involved.

8.5 Analysis of the subjective evaluations

The subjective results from this experiment were judgements of the perceived source width and the perceived acoustical environment width, which were graded on two separate scales. The perceived source width was graded as the subtended angle of the source and was measured in degrees. The perceived acoustical environment width was graded on a scale from 0 to 100.

The first stage of the analysis was to check that the data conformed to the assumptions of the analysis of variance (ANOVA). The application of a Kolmogorov-Smirnov test indicated that the data were not normally distributed. In addition, the use of Levene's test indicated that the variance of the data was not homogeneous. It has been shown that the ANOVA is robust to the violation of the assumptions of normal distribution and equal variance, as long as the samples are of equal sizes [Howell 1997]. However, this must be considered when selecting the most appropriate post hoc tests. As the data did not meet the assumptions of the ANOVA, the validity of applying this analysis was tested by comparing the results of the ANOVA with the results of a non-parametric Kruskal-Wallis test, following the method that was employed by Zacharov and Huopaniemi [Zacharov and Huopaniemi 1999]. The significance value results of the univariate ANOVA and the one-way non-parametric Kruskal-Wallis tests are shown in Table 8.1.

Independent variable	**Source width**		**Environment width**	
	ANOVA	**Kruskal-Wallis**	**ANOVA**	**Kruskal-Wallis**
Sound source signal	0.743	0.888	0.000	0.000
Front microphone technique	0.000	0.000	0.000	0.000
Rear microphone technique	0.011	0.010	0.000	0.000

Table 8.1: Table of the significance value results of both the univariate analysis of variance (ANOVA) and the one-way Kruskal-Wallis tests of the subjective data using the independent variables of sound source signal, front microphone technique and rear microphone technique.

It is apparent that the results from the ANOVA and the Kruskal-Wallis test were very similar, and that both had very high levels of statistical significance in most cases. Based on this information it was concluded that the ANOVA could be used to analyse the data further.

The data from the all the subjects were entered into a type III sum of squares general linear model ANOVA, with the fixed factors of sound source signal (SOURCESIG), front microphone technique (FRONT) and rear microphone technique (REAR) and all interactions. The results of the ANOVA are shown in Table 8.2. In order to test how well the ANOVA model fitted the data, the standardised residuals were analysed. This showed that the model was a good representation of the data, as more than 95% of the standardised residuals were between +2 and –2, and more than 99% were between +2.5 and –2.5 [Field 2000].

Source	**Dependent Variable**	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Partial Eta Squared**
Corrected Model	Source width Environment width	111122 270126	35 35	3174 7717	11.0 20.7	0.000 0.000	0.292 0.436
Intercept	Source width Environment width	669559 1902617	1 1	6695 1902617	2327.8 5097.6	0.000 0.000	0.713 0.845
SOURCESIG	Source width Environment width	488 43909	3 3	162 14636	0.6 39.2	0.638 **0.000**	0.002 0.112
FRONT	Source width Environment width	100333 88511	2 2	50166 44255	174.4 118.6	**0.000** **0.000**	0.271 0.202
REAR	Source width Environment width	3498 105924	2 2	1749 52961	6.1 141.9	**0.002** **0.000**	0.013 0.233
SOURCESIG * FRONT	Source width Environment width	3546 1601	6 6	590 266	2.1 0.7	0.056 0.638	0.013 0.005
SOURCESIG * REAR	Source width Environment width	616 26949	6 6	102 4491	0.4 12.0	0.906 **0.000**	0.002 0.072
FRONT * REAR	Source width Environment width	474 78	4 4	118 19	0.4 0.1	0.800 0.995	0.002 0.000
SOURCESIG * FRONT * REAR	Source width Environment width	2167 3154	12 12	180 262	0.6 0.7	0.820 0.749	0.008 0.009
Error	Source width Environment width	269222 349349	936 936	287 373			
Total	Source width Environment width	1049903 2522092	972 972				
Corrected Total	Source width Environment width	380344 619475	971 971				

Table 8.2: Analysis of variance (ANOVA) results table for all listeners, with perceived source width and environment width as dependent variables, and the sound source signal (SOURCESIG), front microphone technique (FRONT), and rear microphone technique (REAR) as fixed factors.

It is apparent from the results that are contained in Table 8.2 that nearly all of the factors and one interaction were statistically significant beyond the p = 0.05 level. It is apparent that the partial eta squared value (an indication of the effect size of each factor and interaction [Howell 1997]) showed that the source width was affected most by the front microphone technique and that the environment width was affected most by the rear microphone technique. However, the front microphone technique also had a large effect on the environment width. There was also a statistically significant interaction between the source signal and the rear microphone technique, though further investigation of this showed that it was an ordinal interaction, which meant that it could reasonably be disregarded [Howell 1997].

In order to determine which of the levels of the independent variables differed significantly from each other, a post hoc comparison was conducted. As mentioned above, the variance in the data was not homogeneous, meaning that a post hoc test that did not assume equal variances had to be employed. In view of this, the data was analysed using Dunnett's C test, as recommended by Green, Salkind and Akey [Green et al. 2000]. A test was done to investigate the effect of the main independent variables of the front microphone technique and the rear microphone technique on the dependent variables of source width and environment width respectively. The results of this test showed that each of the microphone techniques that were employed was statistically significantly different from the other microphone techniques. In other words, for the judgements of source width, the stimuli that had been recorded using the omnidirectional front microphone technique were judged to be significantly wider than the stimuli that had been recorded using the figure-of-8 microphone technique. The stimuli that had been recorded using the figure-of-8 microphone technique were in turn judged to be significantly wider than the stimuli that had been recorded using the monophonic microphone technique. A similar result was also found for the effect of the rear microphone technique on the judgements of environment width.

The results from this experiment were also used to examine the salience of the descriptors that resulted from the elicitation experiments that were described in Chapter 4 to Chapter 6. These descriptors were used to describe the judgement scales that were employed in the experiment that is described in this chapter.

The first test of the judgement scales was to examine the correlation between the two scales. A Pearson analysis of all the results that were given on the two subjective scales showed a correlation of 0.311, which equates to a shared variance of 9.7%. This indicates that the subjects could differentiate between the meaning of the two judgement scales, and could grade using them with reasonable independence.

The second test of the judgement scales involved an examination of the results of the ANOVA. The mean square of the error in the ANOVA was 287 for the source width judgements and 373 for the environment judgements. This appears to be relatively high. However, it must be considered that the judgements were made on a scale of 0 to 180 for the source width and 0 to 100 for the environment width, which is a larger range than the 0 to 10 scale that is usually employed in grading experiments. The data were transformed by the appropriate factors to form scales from 0 to 10, and the resulting mean square error values for the source width and environment width were then 0.888 and 3.732 respectively. This error value for the source width judgement is similar to that found in experiments that used more established judgement scale descriptors, such as those carried out by Rumsey [Rumsey 1999] and Bech [Bech 1992]. This indicates that the subjects could relate to the meaning of the scales, and could use them to grade in a consistent manner. The error value for the environment width was higher than was expected, which indicates that the subjects were not as consistent in grading this attribute. It is possible that this was due to the fact that there were no intermediate anchor points on the scale, unlike the angular divisions of the source width scale. This could be resolved in future experiments by creating audible intermediate anchor points or reference stimuli to demonstrate the meaning of intermediate points on the scale.

8.5.1 Subjective results for comparison with objective measurements

It was also important to generate results that could be compared with the objective measurements that were made of the stimuli. Whilst it was possible to compare the objective measurements with the raw subjective data, this would have resulted in the correlation of 9 data points from the objective data with 972 data points from the subjective data. This comparison between largely differing numbers of data points makes it difficult for the data to meet the assumptions of the correlation tests [Howell 1997]. Therefore it was decided to use the means of the subjective results in the comparisons with the objective data. As it was apparent from the results of the ANOVA that the different sound source signals (SOURCESIG in Table 8.2) caused a statistically significant variation in the perceived environment width, the means were calculated individually for each source signal, as well as for the data that contained all the source signals. The mean values for all the data for each source signal and for all the source signals combined, separated by the five-channel microphone techniques that were employed, are shown in Table 8.3 for the judgements of source width, and Table 8.4 for the judgements of environment width.

Microphone technique	Sound source signal				
	All	Cello	Snare drum	Trumpet	Voice
Mono omni front, Mono omni rear	11.06	7.89	16.00	10.19	10.19
Mono omni front, Figure-of-8s rear	11.59	9.85	11.41	13.56	11.56
Mono omni front, Spaced omnis rear	13.51	10.37	18.89	12.56	12.22
Figure-of-8s front, Mono omni rear	30.40	28.41	30.15	28.33	34.70
Figure-of-8s front, Figure-of-8s rear	29.05	29.07	31.41	25.37	30.33
Figure-of-8s front, Spaced omnis rear	34.72	29.59	35.30	35.19	38.81
Spaced omnis front, Mono omni rear	33.57	36.48	28.37	37.78	31.67
Spaced omnis front, Figure-of-8s rear	33.76	34.74	31.93	35.48	32.89
Spaced omnis front, Spaced omnis rear	38.55	40.07	40.48	37.00	36.63

Table 8.3: Table of the mean values of the subjective judgements of source width, calculated for each source signal and for all the source signals combined, separated by the five-channel microphone technique.

Microphone technique	Sound source signal				
	All	Cello	Snare drum	Trumpet	Voice
Mono omni front, Mono omni rear	16.76	15.85	19.41	18.11	13.67
Mono omni front, Figure-of-8s rear	33.13	27.93	48.37	28.67	27.56
Mono omni front, Spaced omnis rear	42.46	29.22	66.93	37.78	35.93
Figure-of-8s front, Mono omni rear	36.44	39.22	31.37	38.37	36.81
Figure-of-8s front, Figure-of-8s rear	53.06	46.70	70.37	47.93	47.22
Figure-of-8s front, Spaced omnis rear	60.84	50.78	75.81	59.93	56.85
Spaced omnis front, Mono omni rear	37.82	43.44	39.22	38.04	30.59
Spaced omnis front, Figure-of-8s rear	54.47	46.00	68.93	52.22	50.74
Spaced omnis front, Spaced omnis rear	63.19	50.70	80.15	63.56	58.37

Table 8.4: Table of the mean values of the subjective judgements of environment width, calculated for each source signal and for all the source signals combined, separated by the five-channel microphone technique.

A correlation analysis was conducted on the mean values of the subjective judgements of source width and environment width, separated by the five-channel microphone technique and the sound source signal. This analysis indicated that, even though the raw data of the two judgement scales were reasonably uncorrelated, as discussed above, the mean results of the two subjective scales had a correlation coefficient of 0.669. As this is relatively high, it means that if an objective measurement accurately predicts one of the subjective attributes and is therefore highly correlated with the subjective results from one of the judgement scales, it may also be highly correlated with the other judgement scale. This is unfortunate, as it will make it difficult to separate out the measurements that are related to each of the subjective attributes in this experiment.

8.6 Analysis of the objective measurements

8.6.1 Measurement details

The objective measurements that were made for the experiment were all based on quantifying the properties of binaural signals, as it was discussed in Chapter 1 that it was likely that this type of measurement would be more successful. For this, a Neutrik Cortex MK1 head and torso simulator was positioned at the listening position in the listening room and was used to capture the binaural recordings of all the stimuli that were used in the experiment. In addition to this, a maximum length sequence (MLS) signal was convolved with the impulse responses that were used to create the experiment stimuli, and the resulting signals were also reproduced over the loudspeakers and captured using the head and torso simulator in an identical manner to the experiment stimuli. The resulting binaural MLS signals were then analysed using a Maximum Length Sequence System Analyser (MLSSA) to obtain the impulse response of the system from the virtual source in the simulated acoustical environment, via the simulated five-channel microphone arrays and the five-channel reproduction system to the head and torso simulator in the listening room. This allowed measurements to be made of the properties of impulse responses that were processed in an identical way to the sound source signals.

The properties of the resulting binaural recordings of the experimental stimuli and the binaural impulse responses were then quantified using a number of different measurement techniques.

8.6.2 Conventional interaural cross-correlation coefficient measurements

Measurements were made of the interaural cross-correlation coefficients of the binaural impulse responses in the manner suggested by [BS EN ISO 3382 2000], and following the research of Hidaka, Beranek and Okano [Hidaka et al. 1995]. The measurements that were made according to the first of these were calculated over the entire duration of the impulse responses in one-octave frequency bands, and are denoted in this chapter as $\mathrm{IACC}_{\mathrm{BS}f}$, where f refers to the frequency of the centre of the one-octave bandwidth filter.

The measurements that were made as suggested by Hidaka and his colleagues were similar, but were separated into source-related and environment-related segments and averaged over the results that were measured in one-octave frequency bands centred on 500, 1000 and 2000 Hz. The source-related measurements were made of the early part of the impulse response from the arrival of the direct sound to 80 ms later. This is denoted in this chapter as $\mathrm{IACC}_{\mathrm{E3}}$. The environment-related measurements were made of the late part of the impulse response, from 80ms to 750 ms after the arrival of the direct sound. This is denoted in this chapter as $\mathrm{IACC}_{\mathrm{L3}}$.

The results for these measurements are shown in Table 8.5.

Microphone technique	**$IACC_{BS125}$**	**$IACC_{BS250}$**	**$IACC_{BS500}$**	**$IACC_{BS1000}$**	**$IACC_{BS2000}$**	**$IACC_{BS4000}$**	**$IACC_{E3}$**	**$IACC_{L3}$**
Mono omni front, Mono omni rear	0.928	0.823	0.685	0.808	0.798	0.733	0.769	0.757
Mono omni front, Figure-of-8s rear	0.766	0.668	0.338	0.461	0.645	0.495	0.659	0.237
Mono omni front, Spaced omnis rear	0.512	0.224	0.313	0.240	0.367	0.224	0.333	0.257
Figure-of-8s front, Mono omni rear	0.924	0.869	0.656	0.448	0.211	0.621	0.421	0.617
Figure-of-8s front, Figure-of-8s rear	0.828	0.773	0.441	0.283	0.306	0.330	0.411	0.151
Figure-of-8s front, Spaced omnis rear	0.641	0.348	0.187	0.315	0.192	0.145	0.290	0.231
Spaced omnis front, Mono omni rear	0.929	0.858	0.523	0.279	0.220	0.618	0.301	0.514
Spaced omnis front, Figure-of-8s rear	0.864	0.823	0.410	0.293	0.283	0.401	0.337	0.136
Spaced omnis front, Spaced omnis rear	0.688	0.532	0.075	0.296	0.149	0.221	0.255	0.203

Table 8.5: Table of the results of the interaural cross-correlation measurements made of the binaural impulse responses that were created at the listening position by the use of each of the five-channel microphone techniques.

8.6.3 Novel interaural cross-correlation coefficient measurements

The next set of interaural cross-correlation measurements were made of the properties of the experimental stimuli themselves, rather than the calculated impulse responses that were used for the measurements that are described above. In this case, the source-related and environment-related aspects of the signal were divided based on perceptual grouping. In order to achieve this, segments of the binaural recordings of the experimental stimuli were selected as examples of signals that would be perceived either as a direct sound source or a reverberant decay. For instance, the first long note of the cello extract was selected as the active sound source segment representing that stimulus. This segment was 17500 samples long, and the same segment was edited from each of the nine stimuli that contained the cello sound source signal that employed different microphone techniques. The reverberant decay that was selected was from the end of the cello extract, as this was the only available decay without a direct sound component. This segment was 66000 samples long, and again the same segment was edited from all nine cello stimuli. This process was repeated for the stimuli that contained the trumpet, snare drum and male speaking voice source signals, resulting in 72 edited segments (a direct and a reverberant example from each of the 4 sound source signals and for each of the 9 microphone techniques).

The measurements were made using the process that is described in Appendix A. For this, the audio signals were passed through a gammatone filterbank to mimic the frequency selectivity of the ear, and then a number of consecutive interaural cross-correlation coefficient calculations were made of the resulting narrow-band signals over time. The results of this were then inverted to give a positive correlation with the subjective effect as discussed in Chapter 1, and then filtered to emulate the maximum rate of change of cross-correlation that is perceivable. Finally, the results in each frequency band over time were weighted by the instantaneous amplitude of the signal. From this, the maximum value across the frequency bands at each moment of time was found, and a mean value of these maxima was calculated, which gave the final measured result. This measurement is referred to here as the perceptually grouped interaural cross-correlation coefficient (PGIACC), and the subscript characters of the PGIACC measurement refer to the segment of the signal, where $\mathrm{PGIACC_A}$ is the active sound source segment and $\mathrm{PGIACC_R}$ is the reverberant segment.

Initially, measurements were made of the segments of the stimuli that showed least variation in the subjective results. The cello extract was chosen as the active sound source segment, as this showed least variation in the subjective judgements of source width, and the snare drum was chosen as the reverberant segment, as this showed least variation in the subjective judgements of environment width. The results are shown in Table 8.6.

Microphone technique	**$\mathrm{PGIACC_A}$ for the cello stimuli**	**$\mathrm{PGIACC_R}$ for the snare drum stimuli**
Mono omni front, Mono omni rear	0.162	0.111
Mono omni front, Figure-of-8s rear	0.179	0.242
Mono omni front, Spaced omnis rear	0.182	0.239
Figure-of-8s front, Mono omni rear	0.256	0.141
Figure-of-8s front, Figure-of-8s rear	0.260	0.197
Figure-of-8s front, Spaced omnis rear	0.259	0.252
Spaced omnis front, Mono omni rear	0.256	0.171
Spaced omnis front, Figure-of-8s rear	0.257	0.242
Spaced omnis front, Spaced omnis rear	0.278	0.215

Table 8.6: Table of the results of the perceptually grouped interaural cross-correlation measurements that were made of the stimuli that consist of the cello and snare drum sound source signals that were created at the listening position by the use of each of the five-channel microphone techniques.

8.6.4 Measurements of the fluctuations in interaural time and level difference

Measurements were also made of the fluctuations in interaural time and level difference that were created by the stimuli at the listening position. These were carried out by quantifying the properties of the same active sound source and reverberant segments of the binaural recordings of the stimuli that were used for the perceptually grouped interaural cross-correlation measurements, as described above.

The measurements of the fluctuations in interaural time difference were made using the process that is described in Appendix B. As for the perceptually grouped interaural cross-correlation measurements, the audio signals were passed through a gammatone filterbank, and then the variations over time in the interaural time difference of the resulting narrow-band signals were quantified by the use of a consecutive series of interaural cross-correlation calculations. The results of this were then weighted by the instantaneous amplitude of the signal. From this, the maximum value across the frequency bands at each moment of time was found, and then a mean value of these maxima was calculated, which gave the final measured result. This measurement is referred to here as the interaural cross-correlation fluctuation function (IACCFF). As for the previous measurement, the subscript characters refer to the segment of the signal, where $\mathrm{IACCFF_A}$ is the active sound source segment and $\mathrm{IACCFF_R}$ is the reverberant segment.

The measurements of the fluctuations in interaural level difference were made using the process that is described in Appendix C. The initial stage of the measurement was again to pass the audio signals through a gammatone filterbank. The variations over time in the interaural level difference of the resulting narrow-band signals were then quantified by subtracting the absolute values of the left channel signal from the absolute values of the right channel signal. The results of this were then smoothed by the use of a 3 ms window, in order to reduce errors in the measurement that are caused by the instantaneous signal level and interaural time differences between the signals. From this, the maximum value across the frequency bands at each moment of time was found, and a mean value of these maxima was calculated, which gave the final measured result. This measurement is referred to here as the interaural level difference fluctuation function (ILDFF). As for the previous measurements, the subscript characters refer to the segment of the signal, where $\mathrm{ILDFF_A}$ is the active sound source segment and $\mathrm{ILDFF_R}$ is the reverberant segment.

In the same manner as the perceptually grouped interaural cross-correlation coefficient measurements, the properties of the source-related and environment-related segments were initially measured for representative stimuli that showed least variation in the subjective results. From this, the properties of the active sound source segment of the cello stimuli and the reverberant segment of the snare drum stimuli were quantified. The results are shown in Table 8.7.

Microphone technique	**$\mathrm{IACCFF_A}$ for the cello stimuli**	**$\mathrm{IACCFF_R}$ for the snare drum stimuli**	**$\mathrm{ILDFF_A}$ for the cello stimuli**	**$\mathrm{ILDFF_R}$ for the snare drum stimuli**
Mono omni front, Mono omni rear	34.6	27.0	-31.0	-55.9
Mono omni front, Figure-of-8s rear	34.9	32.1	-32.5	-52.0
Mono omni front, Spaced omnis rear	36.3	32.7	-31.6	-52.6
Figure-of-8s front, Mono omni rear	38.8	26.8	-33.1	-56.2
Figure-of-8s front, Figure-of-8s rear	39.1	30.7	-34.0	-52.3
Figure-of-8s front, Spaced omnis rear	39.0	33.7	-31.9	-52.3
Spaced omnis front, Mono omni rear	39.5	28.7	-30.9	-55.9
Spaced omnis front, Figure-of-8s rear	40.3	32.9	-29.1	-52.4
Spaced omnis front, Spaced omnis rear	39.5	30.7	-29.0	-52.4

Table 8.7: Table of the results of the perceptually grouped measurements of the fluctuations in interaural time and level difference that were made of the stimuli that consist of the cello and snare drum sound source signals that were created at the listening position by the use of each of the five-channel microphone techniques.

8.7 Correlation between the objective and subjective results

8.7.1 Correlation between the objective and subjective results for all sound source signals

The objective measurement data that are contained in Table 8.5, Table 8.6 and Table 8.7 were compared with the means of the subjective data that are contained in Table 8.3 and Table 8.4. The initial analysis compared the objective measurements with the means of the subjective data that contained the judgements for all the sound source signals. The results of the Pearson correlation test are shown in Table 8.8, with correlations of higher than 0.7 denoted in bold text.

Objective measurement	Correlation with means of subjective judgement of source width	Correlation with means of subjective judgement of environment width
IACC_{BS125}	0.117	-0.468
IACC_{BS250}	0.124	-0.407
IACC_{BS500}	-0.336	**-0.784**
IACC_{BS1000}	-0.568	**-0.784**
IACC_{BS2000}	**-0.885**	**-0.794**
IACC_{BS4000}	-0.342	**-0.852**
IACC_{E3}	**-0.791**	**-0.824**
IACC_{L3}	-0.304	**-0.799**
PGIACC_{A}	**0.982**	**0.770**
PGIACC_{R}	0.162	0.681
IACCFF_{A}	**0.958**	**0.752**
IACCFF_{R}	0.095	0.644
ILDFF_{A}	0.301	0.245
ILDFF_{R}	0.059	0.664

Table 8.8: Table of the results of the Pearson correlation analysis between the objective measurements and the mean values of the subjective judgements of source width and environment width including the data from all the sound source signals, separated by the five-channel microphone technique.

A number of factors can be seen from the results of the correlation analysis. Firstly, it appears that the interaural cross-correlation measurements that were made of the entire duration of the impulse responses, as suggested in [BS EN ISO 3382 2000] (denoted as IACC_{BS} in Table 8.8) provide different results in different frequency bands. Only the measurement that was centred on 2 kHz was highly correlated with the subjective judgements of source width. These measurements matched the subjective results of environment width more successfully, though only from 500 Hz and above.

The measurements of the interaural cross-correlation coefficient of the early and late segments of the impulse responses (denoted as IACC_{E3} and IACC_{L3} in Table 8.8) were mostly highly correlated with the subjective data. However, it is apparent that the measurement of the early segment of the impulse response, which is usually related to the perceived properties of the sound source, was more closely correlated with the subjective judgements of the environment width than the source width. In addition, this measurement also matched the subjective judgements of environment width more accurately than the measurement of the late segment of the impulse response that is usually related to the properties of the reverberation.

The perceptually grouped interaural cross-correlation coefficient measurements that were made of the active sound source segment (denoted as PGIACC_{A} in Table 8.8) appeared to match the subjective judgements of the source width most accurately. However, the results of the subjective judgements of environment width were more closely correlated with the active sound source segment measurement (PGIACC_{A}) than the reverberant component (PGIACC_{R}). This may be due to the fact that these measurements were made of specific sound source signals and that the subjective results were the means from the data including all the sound source signals. Whether this assumption is correct could be examined by comparing the objective and subjective results for each sound source signal individually.

The measurements of the magnitude of the fluctuations in interaural time difference that were made of the active sound source segment (denoted as IACCFF_{A} in Table 8.8) showed a high correlation with the subjective judgements of both the source width and the environment width. However, the measurements of the magnitude of the fluctuations in interaural time difference that were made of the reverberant component (denoted as IACCFF_R in Table 8.8) were not highly correlated with the results of either of the subjective judgement scales. The fact that this correlation is poor may also be due to attempting to relate the measurements of a single sound source signal with the subjective results from all the sound source signals.

Finally, the measurements of the magnitude of the fluctuations in interaural level difference that were made of either the active sound source or reverberant segments of the signals (denoted as ILDFF_A and ILDFF_R in Table 8.8) both showed poor correlation with the results of either of the subjective attribute scales.

It is interesting to note that, for a number of the measurement techniques, the measurements that were related to the direct sound scene component were more correlated with the perceived environment width than the measurements that were related to the reverberation. This is partially due to the correlation between the means of the subjective results for the two judgement scales, as discussed in Section 8.5. It is also partially caused by the fact that the results of the environment width judgements were different for some of the sound source signals, as indicated by the statistically significant result for the sound source signal factor in the ANOVA results that are shown in Table 8.2.

8.7.2 Correlation between the objective and subjective results for individual sound source signals

The measurement techniques that are described above that are based on perceptual grouping were implemented by quantifying the properties of stimuli that were used in the experiment. This is dissimilar to the extant measurement techniques that were implemented by quantifying the properties of the measured impulse responses of the entire recording and reproduction chain. In view of this, it was suggested that these techniques might be more closely correlated with the subjective data if the grades for each individual sound source signal were taken into account. In addition to this, the correlation between the objective and subjective results may be lower if the means are taken from all the subjective data, as the different sound source signals cause different perceived widths for the same microphone techniques. In view of this, the data from the subjective evaluations and the objective measurements that were made for each of the sound source signals were compared.

For the objective measurements that were made of the properties of the impulse responses, the same set of measurement results was used to compare with the subjective results for each of the sound source signals. In order to simplify the analysis, the measurements that were made according to [BS ISO EN 3382 2000] (denoted as IACC_{BS} in Table 8.8) that were least consistent in the previous correlation analysis were excluded from this analysis. For the first analysis, the subjective and objective data that contained the results for each of the sound source signals were entered into a single Pearson correlation analysis. The results are shown in Table 8.9, with the correlations of higher than 0.7 denoted in bold text.

Objective measurement	Correlation with means of subjective judgement of source width	Correlation with means of subjective judgement of environment width
$IACC_{E3}$	**-0.767**	-0.699
$IACC_{L3}$	-0.294	-0.678
$PGIACC_{D}$	0.376	0.040
$PGIACC_{R}$	0.322	0.316
$IACCFF_{D}$	0.487	0.431
$IACCFF_{R}$	0.147	0.076
$ILDFF_{D}$	0.116	0.381
$ILDFF_{R}$	0.088	0.108

Table 8.9: Table of the results of the Pearson correlation analysis between the objective measurements and the mean values of the subjective judgements of source width and environment width, that were calculated using data that was separated by the five-channel microphone technique and the sound source signal.

It is apparent from this analysis that the measurements that are based on quantifying the interaural cross-correlation coefficient of the impulse response were most successful in predicting the source width and environment width of the stimuli. In order to investigate the reason why the other measurement techniques were unsuccessful at predicting the perceived spatial effect that was created by the experiment stimuli, the data were analysed further by examining scatter plots of the objective data against the subjective data for each of the sound source signals. Two of these plots are shown in Figure 8.4 and Figure 8.5.

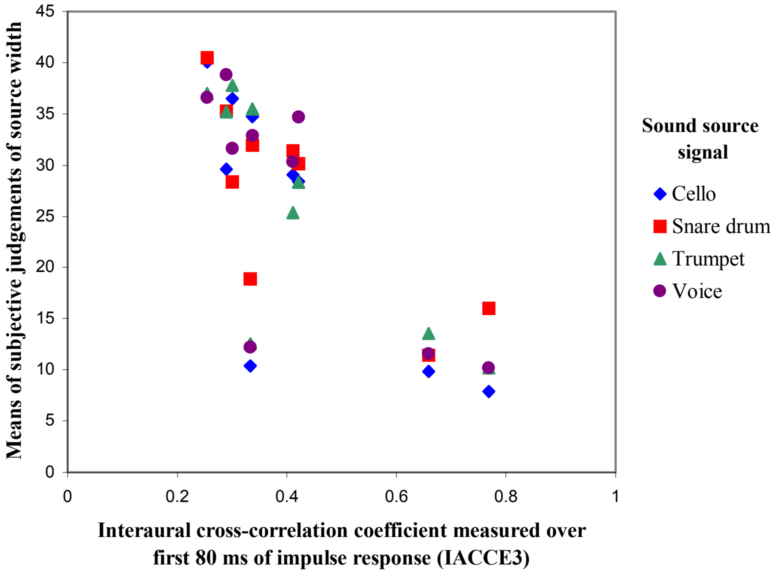

Figure 8.4: Scatter plot of the means of the subjective judgements of source width and the interaural cross-correlation coefficients that were measured over the first 80 ms of the impulse response ($IACC_{E3}$), calculated using data that was separated by the five-channel microphone technique and the sound source signal.

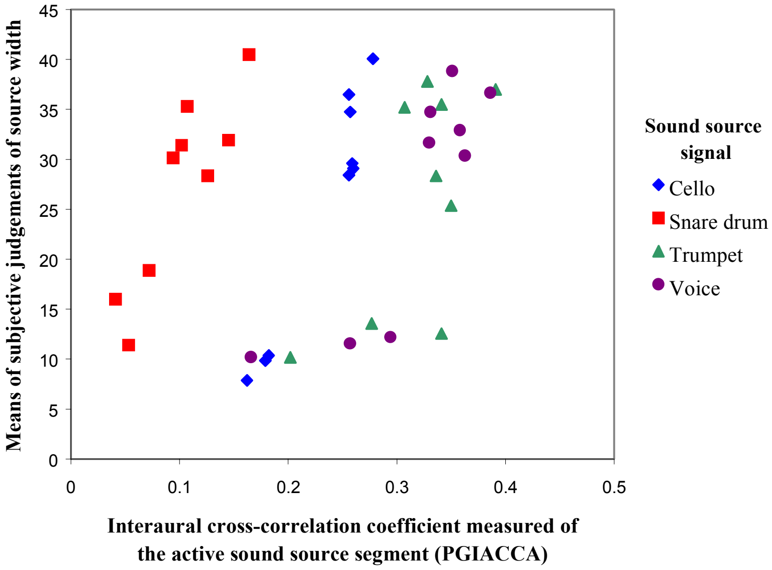

Figure 8.5: Scatter plot of the means of the subjective judgements of source width and the interaural cross-correlation coefficients that were measured of the active sound source segment ($PGIACC_A$), calculated using data that was separated by the five-channel microphone technique and the sound source signal.

It is apparent from Figure 8.5 that the perceptually grouped interaural cross-correlation coefficient that was measured of the active sound source segment of each stimulus was well correlated with the subjective judgements of source width for each of the sound source signals when each sound source signal was considered individually. However, it was apparent that there were differences between the measured results of the different sound source signals, which meant that when the data for all of the sound source signals was entered into a single correlation calculation, the correlation was low.

It is interesting to note that the difference between the measured results for each sound source signal seems to be related to the frequency content of the sound source signals. It appears that the sound source signals with more low frequency content create a lower result from the PGIACC measurement than the higher frequency sound source signals. This may be related to the frequency dependency of the interaural cross-correlation coefficient that was discussed in Chapter 1.

In order to investigate the accuracy of each measurement for each individual sound source signal, the correlation between the subjective and objective results was calculated individually for each sound source signal. The results for the judgements of source width are shown in Table 8.10, and the results for the judgements of environment width are shown in Table 8.11. The mean values of the correlation results for each of the sound source signals were also calculated and are shown in the right hand column of the tables. As for the previous results tables, correlations of higher than 0.7 are denoted in bold text.

Objective measurement	**Correlation with means of subjective judgement of source width**				
	Cello	**Snare**	**Trumpet**	**Voice**	**Average**
$IACC_{E3}$	**-0.771**	**-0.791**	**-0.779**	**-0.750**	**-0.773**
$IACC_{L3}$	-0.301	-0.335	-0.286	-0.268	-0.297
$PGIACC_A$	**0.969**	**0.886**	0.636	**0.852**	**0.836**
$PGIACC_R$	0.636	0.158	0.535	0.541	0.468
$IACCFF_A$	**0.957**	0.610	0.365	**0.906**	**0.709**
$IACCFF_R$	0.503	0.119	0.217	0.688	0.382
$ILDFF_A$	0.350	0.177	**0.942**	0.192	0.415
$ILDFF_R$	0.691	0.138	0.559	0.196	0.396

Table 8.10: Table of the results of the Pearson correlation analysis between the objective measurements and the mean values of the subjective judgements of source width including the data from all the sound source signals, calculated using data that was separated by the five-channel microphone technique and the sound source signal.

It is apparent from the results that are shown in Table 8.10 that the measurements that were based on the interaural cross-correlation coefficient were most successful at predicting the perceived source width for the range of sound source signals that were used in this experiment. The measurements that are based on quantifying the fluctuations in interaural time or level difference were only well correlated with the subjective judgements for some of the stimuli.

From this data, it is possible to give a rank order of the success of these measurement techniques at predicting the subjective judgements of source width in this experiment. As all the measurement techniques are divided into active sound source and reverberant segments, it is logical to only consider the active sound source segment in relation to the perceived source width. Based on these results, it appears that the perceptually grouped interaural cross-correlation measurements that were made of the experiment stimuli ($PGIACC_A$) were most highly correlated with the subjective data. This was followed by the interaural cross-correlation measurements that were made of the impulse responses ($IACC_{E3}$) and then the measurements of the fluctuations in interaural time difference ($IACCFF_A$). The only measurement technique of the four that did not result in an average correlation of greater than 0.7 for the four sound source stimuli, was that based on the fluctuations in interaural level difference ($ILDFF_A$).

Objective measurement	Correlation with means of subjective judgement of environment width				
	Cello	Snare	Trumpet	Voice	Average
IACC$_{E3}$	**-0.850**	-0.694	**-0.841**	**-0.803**	**-0.797**
IACC$_{L3}$	-0.583	-0.694	-0.685	**-0.742**	-0.676
PGIACC$_{D}$	**0.954**	0.596	**0.780**	**0.902**	**0.808**
PGIACC$_{R}$	**0.827**	**0.826**	**0.828**	0.541	**0.756**
IACCFF$_{D}$	**0.924**	0.667	0.695	**0.906**	**0.798**
IACCFF$_{R}$	**0.759**	**0.819**	0.616	0.688	**0.720**
ILDFF$_{D}$	0.154	0.625	**0.844**	0.192	0.454
ILDFF$_{R}$	**0.755**	**0.860**	**0.890**	0.196	0.675

Table 8.11: Table of the results of the Pearson correlation analysis between the objective measurements and the mean values of the subjective judgements of environment width including the data from all the sound source signals, calculated using data that was separated by the five-channel microphone technique and the sound source signal.

From the results that are shown in Table 8.11, it appears that the different types of objective measurement technique that were used in the experiment performed similarly in predicting the perceived environment width. It is interesting to note that, in the same way that was apparent for the results in Table 8.8, the judgements of the environment width were more correlated with the measurements that related to the active sound source segment than those that related to the reverberant segment. In this case, this was apparent for three out of the four measurement techniques. This result indicates that the division between the source-related and environment-related aspects of the audio signal may not be clear, either for the measurement or for the perception. Therefore further research is needed to investigate the most appropriate manner of dividing the source-related and environment-related aspects of the signal when using complex stimuli.

As for the previous set of data, it is again possible to give a rank order of the success of these measurement techniques at predicting the subjective judgements, though for the judgements of environment width in this experiment. For this, it is logical to only consider the segments of the measurement techniques that were intended to relate to the reverberation.

The results showed that the type of measurement that was most highly correlated with the subjective data in this case was the perceptually grouped interaural cross-correlation measurements that were made of the experiment stimuli (PGIACC$_{R}$). This was followed by the measurements of the fluctuations in interaural time difference (IACCFF$_{R}$) and the interaural cross-correlation measurements that were made of the impulse responses (IACC$_{L3}$). Again, the measurements of the fluctuations in interaural level difference (ILDFF$_{R}$) did not create an average correlation of greater than 0.7 for the four sound source stimuli.

In order to evaluate the salience of the measurement techniques as a whole, including the calculations of both the segment that is related to the properties of the perceived sound source and the segment that is related to the properties of the perceived reverberation, an average was taken of the correlations between the objective and subjective results for these two segments. In other words, the average correlation between the measurements that related to the active sound source segment and the source width judgements that are shown in Table 8.10, and the average correlation between the measurements that related to the reverberant segment and the environment width judgements that are shown in Table 8.11 were combined. The results of this are shown in Table 8.12.

Objective measurement		**Correlation with means of subjective judgement of environment width**		
		Source width	**Environment width**	**Overall result**
Interaural cross-correlation measurements made of impulse responses	$IACC_{E3}$	**-0.773**		
	$IACC_{L3}$		-0.676	
	Overall			**-0.725**
Interaural cross-correlation measurements made of stimuli	$PGIACC_D$	**0.836**		
	$PGIACC_R$		**0.756**	
	Overall			**0.796**
Interaural time difference fluctuation measurements made of stimuli	$IACCFF_D$	**0.709**		
	$IACCFF_R$		**0.720**	
	Overall			**0.715**
Interaural level difference fluctuation measurements made of stimuli	$ILDFF_D$	0.415		
	$ILDFF_R$		0.675	
	Overall			0.545

Table 8.12: Table of the overall results combining the relevant correlation results from Table 8.10 and Table 8.11 to give an overall rating for each type of objective measurement technique.

The results of this analysis show that the interaural cross-correlation measurements that were made of the experiment stimuli (PGIACC) were most highly correlated with the subjective results when both subjective attributes were taken into account. This was followed closely by the interaural cross-correlation measurements that were made of the impulse responses (IACC), and then the measurements of the fluctuations in interaural time difference (IACCFF). The measurements of the fluctuations in interaural level difference were least correlated with the subjective data (ILDFF).

It must be noted that the difference in the correlation results between the three most highly correlated measurement techniques was relatively small. This means that conclusive results about the most salient measurement type cannot be drawn from this experiment. However, it does indicate that, at least for the stimuli that were used in this experiment, the perceptually grouped cross-correlation measurement technique that was developed from the research that is contained in this thesis was as successful as the more established measurements that were made of the impulse response. Further experimentation is needed to evaluate whether there are situations in which either of the techniques fails to predict the subjective results.

8.8 Discussion

8.8.1 Discussion of the results of the experiment

The purpose of this experiment was to evaluate the descriptors that had been elicited from the subjects in the previous experiments, and to examine the relationship between a number of objective measurements and the subjective results.

The spatial attribute descriptors that had been elicited in the subjective experiments that were described in Chapter 4 to Chapter 6 were evaluated in a number of ways. The first measure of the value of the spatial attribute descriptors was to examine whether they could be employed as descriptors for subjective judgement scales. Anecdotal information that was gained from the subjects indicated that they had no problems with using the judgement scales. In addition, as the results from this grading experiment matched the expectations that were set out in Section 8.2, this showed that the subjects could give useful and meaningful grades using these scales. The error that was apparent in the results was also evaluated to give an indication of whether the subjects could make meaningful grades using judgement scales that were based on these descriptors. This showed that the value for the mean squared error term in the results of the ANOVA for the judgements of source width was typical for this type of experiment. The value for the mean squared error term for the judgements of environment width was higher, though this was likely to be due to the fact that the scale of the environment width had no intermediate anchor points for reference. It was concluded from these results that the subjects could relate to the descriptors that were elicited, and that they could give meaningful results on grading scales whose meaning was described by the elicited descriptors.

The second measure of the value of the spatial attribute descriptors was whether the subjects could discriminate between the meaning of the two scales, and whether they could grade them independently. This was examined by calculating the correlation between the judgements that were given on the two scales. The results of this showed that there was a low correlation between the two sets of results, which indicated that the subjects could differentiate between the meaning of the two scales.

The final measure of the value of the spatial attribute descriptors examined whether these attributes that had been elicited using noise stimuli were relevant for programme material that was more similar to that which may be sounded in a concert hall or through a reproduction system. The fact that statistically significant differences were present in the results indicated that the more externally valid stimuli contained differences that could be indicated using the two spatial attribute scales that had been elicited using the noise signals. Therefore, it appears that the subjective attributes that were elicited using noise signals could be applied to more externally valid music and speech signals.

The salience of the objective measurement techniques that were developed from the research and experimentation that is described in this thesis were also examined in this experiment. This was conducted by making measurements of the experimental stimuli and of impulse responses that had been processed in an identical manner to the experimental stimuli. The measured results were then compared with the subjective judgements of the experimental stimuli. The success of these measurement techniques was then judged, based on how accurately they matched the subjective results for this experiment.

Two types of measurement technique were employed. The first type was representative of the extant methods that have been employed to quantify the spatial parameters of the acoustics of concert halls. The second type was based on the developments that have been made through the research and experimentation that are described in this thesis. The predominant difference between the two types of measurement was the following. The extant methods quantified the properties of the measured impulse responses, and divided the source-related and environment-related segments by the use of a single point in time. Conversely, the novel measurements quantified the properties of the stimuli themselves and divided the source-related and environment-related segments by the use of perceptual grouping.

The novel measurement methods also included techniques that were based on both calculations of the interaural cross-correlation coefficients and calculations of the fluctuations in interaural time and level difference. By comparing the accuracy of each of these techniques at matching the subjective judgements, an indication could be gained of which perceptual process may be employed in judging the perceived spatial attributes. This was based on the assumption that if the perceptual process that is employed is similar to one of these measurement techniques, it is likely that this technique will produce results that are more similar to the subjective judgement.

The comparison between the extant measurement techniques and the perceptually grouped measurement techniques indicated that, for the subjective judgements for each of the sound source signals, the perceptually grouped interaural cross-correlation measurement matched the subjective judgements more closely than the extant interaural cross-correlation measurement. However, the difference between the results was small, indicating that the two measurement techniques performed equally well. This result is not as conclusive as the author had hoped, though it is an indication that the novel measurement is as successful as the extant measurement in this case. In addition, the author believes that, as the novel measurement has been developed in a more perceptually valid manner, it may match the subjective judgements in more situations than the extant measurement. However, further investigation is required to confirm this.

This analysis also uncovered the fact that the results of the perceptually grouped measurements could not be directly compared across the different sound source signals. This was indicated by the low correlation between the objective and subjective results when the data from all the sound source signals were entered into a single analysis, as shown in Table 8.9. This is a problem if the measurement technique is required to quantify the properties of any type of stimulus that may be reproduced in a concert hall or through a sound reproduction system. However, this measurement could still be used to compare the results of reproducing identical test stimuli or exemplary items of programme material in different acoustical environments or through different processing or reproduction systems. Nevertheless, this is a limitation to the measurement technique that requires further research to resolve.

The perceptually grouped measurement technique that is based on the interaural cross-correlation and the perceptually grouped measurement techniques that are based on the fluctuations in interaural time and level difference were also compared. The results from all the correlation analyses that were undertaken showed that overall, the measurements of the fluctuations in interaural level difference were poorly correlated with the subjective results. The measurements of the fluctuations in interaural time difference were more successful, and in most cases these were almost as correlated with the subjective results as the interaural cross-correlation measurements. The fact that the measurements based on the interaural cross-correlation coefficient were more closely correlated with the subjective results indicates that the perceptual mechanism that is involved in perceiving these spatial attributes of auditory stimuli may be based on a process that is similar to cross-correlation. However, as the difference between the success of the measurements was small, further research is needed to confirm this.

8.8.2 Limitations of the objective measurements

The analysis of the objective measurement results uncovered a number of limitations in the measurement techniques that need to be addressed. The first of these is that the measurement techniques that were employed in this chapter cannot quantify the properties of more than one simultaneous sound source. As mentioned in Chapter 7, some form of model of auditory streaming that can segregate the physical cues that are associated with different sound sources would be required in order to measure the properties of multiple sound sources.

The second limitation was that the research that was outlined in Chapter 1 indicated that the interaural cross-correlation that is required to create a certain perceived width was dependent on the audio frequency of the signal. Therefore in order for the measurement to be accurate at all audio frequencies, it may be that a frequency-based weighting is required in order to compensate for the variable subjective effect. The measurements of the interaural cross-correlation coefficient that were employed in the experiment that is described in this chapter may have been limited by the fact that this had not been taken into account. The lack of a frequency-based equal perception weighting may not cause a problem for wide-band signals where the perceived width is similar at all frequencies. However, it will be an issue when comparing narrow-band signals of different audio frequencies. It is possible that this contributed to the perceptually grouped interaural cross-correlation measurement being unable to make comparable measurements of the different sound source signals. Research is needed to investigate the magnitude of the weighting that will be required at different frequencies, and whether this needs to be applied to the relevant measurement techniques.

The third limitation of the measurement techniques that were employed in this chapter was that the loudness of the stimuli had not been considered. There is evidence to suggest that the loudness of a stimulus affects the perceived spatial properties, as discussed in Chapter 1. Research is needed to confirm whether this is the case, and to investigate the precise relationship between the loudness and other physical parameters that affect the perceived spatial properties. This research then needs to be applied to further develop the objective measurement techniques.

The fourth limitation of the objective measurements involved the weighting of the results by the audio amplitude. This is included to minimise the effect of noise in the measurements. However, it may also have affected the results in other ways. One parameter that will affect the amplitude-weighted measurement results is the reverberation time of a stimulus. If two stimuli are compared with identical levels of interaural cross-correlation but different reverberation times, and the measurements are made over the same duration, then the amplitude-weighted measurement will show that the stimulus with the longer reverberation time has a lower average interaural cross-correlation coefficient. The reason for this is that the mean amplitude of the stimulus with the longer reverberation time will be higher than the mean level of the stimulus with the shorter reverberation time. The first factor that needs to be investigated is whether this difference is perceptually relevant. The results of the experiment that was described in Chapter 6 indicate that it may not be the case that more reverberant stimuli are perceived to be wider. However, this result is not conclusive for real acoustical environments. The second factor that needs investigation is how this error can be avoided if it is found to be perceptually inaccurate.

This is also related to the fifth limitation of the objective measurements, which is whether a mean value or a maximum value should be quoted as the final result. The measurements do allow the investigation of a wider range of data, including the results in different frequency bands and at different points in time. However, to enable simple comparison with other results, it is preferable for there to be a single figure result. The measurements that were employed in this chapter were mostly calculated by taking the maximum value across the frequency bands at each point in time, and then calculating the mean value of these maxima for the length of the segment of audio. It is logical that if the subjects were judging the maximum width of the stimuli, then a maximum value would have been more appropriate to use as the measurement result. However, a maximum value is more susceptible to error than a mean value, meaning that errors that may be induced by factors such as the lack of the frequency-based weighting that was mentioned earlier, would be greater for a maximum value than a mean value. It may be that if all the segments of the objective measurement technique are correct, then the maximum value can be used as the output of the measurement. However, until that is the case, the mean value may be more appropriate.

The final limitation of the objective measurements that were employed in this chapter was the difficulty of accurately selecting segments of the stimuli that solely contained either direct sound source or reverberant components. Indeed, for some of the stimuli it was difficult to perceive the point in time at which the direct sound ended. Further research is required to determine exactly how the source-related and environment-related aspects of the sound are divided in this case, and whether there is a common approach that can be applied to all source signals. In addition to this, it was difficult to select identical segments from each of the stimuli that contained the same sound source signal but with different microphone techniques. This task could be simplified by using predetermined sound source signals, of which the exact start and end points could be calculated. However, it would be preferable if a technique could be developed to automatically divide the physical parameters that are related to the source and the physical parameters that are related to the acoustical environment. The creation of such an automated technique would require a great deal of further research.

8.9 Summary

This chapter described the experiment that was undertaken to evaluate the results of the research and experimentation that was described in this thesis. For this experiment, the subjective spatial attribute descriptors that had been elicited in the subjective experiments were used as grading scales for judging the spatial attributes of a number of auditory stimuli in order to investigate their value. In addition, the objective measurement techniques that resulted from the research were assessed by comparing the objective results with the subjective judgements.

The method that was used to create the stimuli was specified in Section 8.2, for which a number of anechoic sound sources were convolved with impulse responses of a simulated source, acoustical environment and a number of microphone techniques. The resulting sound files were reproduced over a five-channel loudspeaker arrangement in a standard listening room. The stimuli were intended to be representative of the types of stimuli that may be produced in a concert hall and reproduced over a multichannel reproduction system. However, in order to make the experiment practical, the stimuli only consisted of single sources, the recorded acoustical environment was simulated in CATT-Acoustic, and only one simulated acoustical environment was used.

The spatial attributes of the stimuli were judged using two grading scales that were described using the descriptors that had been elicited from the subjective experiments that were summarised in Chapter 4 to Chapter 6. These results were then analysed using an analysis of variance, which showed that the perceived source width was predominantly altered by the front microphone technique, and that the perceived environment width was predominantly altered by the rear microphone technique. The fact that the experiment resulted in statistically significant results with a relatively low error, and the fact that the results matched the prior expectations, indicated that the subjects could grade the spatial attributes of the stimuli in a meaningful, consistent and reliable way by using the descriptors that had been elicited. This also indicated that the descriptors that had been elicited using noise stimuli could be applied to more externally valid stimuli.

Objective measurements were made of the experimental stimuli or of impulse responses that had been processed in an identical manner to the stimuli, and these were compared with the subjective results. This showed that the novel perceptually grouped interaural cross-correlation measurement matched the subjective judgements most accurately when the results for each individual stimulus were compared. However, the results of the extant interaural cross-correlation method were similarly correlated with the subjective judgements, so it could not be concluded that a significant improvement had been made. Nevertheless, this demonstrated that for this experiment, this novel measurement technique performed at least as well as the more established technique. Novel measurement techniques that were based on quantifying the fluctuations in interaural time and level difference were also evaluated. This showed that the fluctuations in interaural time difference matched the subjective results only slightly less accurately than the measurements that were based on the interaural cross-correlation, but that the fluctuations in interaural level different did not match the subjective results well overall.

The results of the experiment were discussed in Section 8.8, and a number of areas that require further research were highlighted. It was concluded that the spatial attribute descriptors that had been elicited were useful as grading scales for evaluating the perceived spatial attributes of stimuli.

It was also concluded that the novel measurement that was based on the interaural cross-correlation performed at least as well as the extant method, and that it may match the subjective judgements in a greater range of situations than the extant method. In addition, it was considered that the perceptual process that is involved in detecting the perceived spatial attributes of auditory stimuli may be based on a cross-correlation process, as the measurements that were based on this technique matched the subjective results more reliably than the measurements that were based on the fluctuations in interaural time difference. However, as the difference was small, it was concluded that further research was required.

Finally, the limitations of the objective measurement techniques were also discussed, and approaches for avoiding these were considered. It was concluded that further research is required to widen the application of these measurements, and to improve the accuracy of the measurements for a large range of types of stimulus.