Audio Engineering Society

Convention Paper 5591

Presented at the 112th Convention
2002 May 10–13 Munich, Germany

A comparison of objective measurements for predicting selected subjective spatial attributes

Russell Mason[1] and Francis Rumsey[2]

[1] *Institute of Sound Recording, University of Surrey, Guildford, Surrey, GU2 7XH, UK*
[2] *Institute of Sound Recording, University of Surrey, Guildford, Surrey, GU2 7XH, UK*
Correspondence should be addressed to corresponding author (r.mason@surrey.ac.uk)

ABSTRACT

A controlled subjective experiment was undertaken to evaluate the relative merits of objective measurement techniques for predicting selected perceived spatial attributes of reproduced sound. The stimuli consisted of a number of anechoic recordings of single sound sources that were reproduced in a simulated concert hall and captured using a number of simulated multichannel microphone techniques. These were reproduced in a listening room and the subjects were asked to judge the perceived source width and perceived environment width of each stimulus. A number of objective measurements were made at the listening position and these were then compared with the subjective judgements. The results showed that a perceptually-grouped measurement of the experimental stimuli using a technique based on the interaural cross-correlation coefficient matched the subjective judgements most accurately, though the difference between this measurement and a number of other types was small.

INTRODUCTION

A large amount of the research and testing that is conducted on sound reproduction systems involves subjective evaluations in the form of listening tests, as the ultimate judge of the quality of a product will be a human listener. Whilst subjective tests are a useful method of evaluation, they are both expensive and time consuming to undertake, and they require the training of an expert listening panel in order for them to give consistent and reliable results.

As an alternative to subjective methods of evaluation, objective measures that correlate well with certain subjective parameters would be more accurately repeatable, and could save time and money (1). Therefore it would be useful if subjective assessments could be replaced or complemented by objective measurement methods.

The overall quality of reproduced sound contains a wide range of types of subjective attributes. Factors such as timbre, distortion and dynamic range have been researched in detail, and useful measurements have been developed for each of these. On the other hand, the research into the spatial performance of sound reproduction systems is relatively incomplete, yet this is an important factor that contributes to the perceived quality of reproduced sound (2). This area is of increasing research interest, due to the growing number of sound reproduction systems that can deliver an enhanced spatial auditory experience to the consumer. One such research project that investigated this area, within which the research that is described in this paper was initiated, was Eureka 1653 MEDUSA (Multichannel Enhancement of Domestic User Stereo Applications)[1].

[1] This was a collaborative project with a duration of 3.5 years that involved the British Broadcasting Corporation, the Institute of Sound Recording at the University of Surrey, Nokia Research

Within the field of spatial impression, a great deal of research has been completed into the aspect of localisation in reproduction systems. However, other attributes of spatial impression have so far been left behind (3). Perhaps one reason for this is that localisation can be evaluated in listening tests relatively simply. In contrast, other aspects of spatial impression are much more complicated multidimensional subjective phenomena[2].

One area in which spatial impression has been researched in detail is the perception of concert hall acoustics. Beranek provides a good overview of this (4). Research that has been discussed previously by the authors (5, 6) has indicated that some of these can be applied to sound reproduction in certain situations. However, it is also apparent from existing standards (7) that no decision has yet been made on what would be the most appropriate measurement technique for quantifying the spatial properties of concert halls. In view of this, research is required in this area in order to develop measurements that relate to the perceived spatial impression of concert hall acoustics, and to investigate the most appropriate manner for applying these to reproduced sound.

Limitations of existing measurement procedures

It is apparent from (7) that the most common objective measurement techniques that have been developed for predicting the perceived spatial properties of concert hall acoustics are applied by quantifying the properties of measured impulse responses. Within this, the characteristics of the initial 80 ms of these impulse responses are usually related to the perceived attributes of the sound source, and the characteristics of the remainder of these impulse responses from 80 ms are usually related to the perceived attributes of the reverberant environment.

Whilst this is an established method, it may not necessarily be relevant for relating objective measurements to the perceived effect of musical stimuli. The reason for this is the dissimilarity between an impulsive signal and the majority of the types of programme material that will be produced in a concert hall during performances. This difference is important, as impulsive sounds are rare in the normal experience of concert hall acoustics (8). For the purpose of this paper, an impulsive signal is defined as a transient signal of short duration, with a spectral content that is atonal and that covers a wide frequency range.

Griesinger found that concert hall measurements that were made of impulsive signals gave different results to those made of musical signals (9). He observed that the interaural cross-correlation that was measured by using a musical signal usually gave lower values than if the equivalent impulse response was measured. This is evidence that measurements that quantify the properties of impulse responses may not be appropriate for predicting the perceived effect of more continuous signals.

The division of the source-related and environment-related aspects of a signal by the use of a single time division of an impulse response may also not be perceptually relevant for musical signals. Firstly, the use of a single time for the division may not be suitable in all situations. The cut-off at 80 ms was introduced based on the assumption that reflections that arrive beyond this time may be perceived to be a disturbing echo with a musical stimulus (10). However, this is only true if the reflection after 80 ms is the first reflection or if it is particularly prominent (11), and it is not the case for all types of stimuli (12). Secondly, the two time segments do not affect the perceived source width and envelopment independently. Bradley and Soulodre found that the differences in perceived source width that were caused by modifying early reflections were more difficult to perceive in the presence of reverberation (13). Finally, Griesinger argued that the division between the source and environment-related subjective attributes is dependent on perceptual grouping or streaming (9), rather than a time-based process.

Based on this research, a novel procedure for applying the conventional concert hall measurements was formulated. This was based on quantifying the properties of musical signals rather than impulse responses, and the division of the source-related and environment-related aspects of the signal based on perceptual grouping. This involved the selection of segments of a musical signal that contained physical cues that were perceived to be an attribute of the source, and segments that contained physical cues that were perceived to be an attribute of the acoustical environment.

The salience of this procedure then required evaluation to compare this with the existing measurement procedures by attempting to use both techniques to predict the perceived spatial attributes of a number of musical stimuli.

Limitations of existing measurement types

Of those measurement types that are suggested in the existing standards document (7), it is likely that the interaural cross-correlation coefficient is most likely to be successful for application to sound reproduction, as this is a binaural measurement. However, it suffers from a number of limitations. Firstly, the measured cross-correlation varies greatly around the lateral centre line of a room (14, 15). Secondly, as the audio frequency falls from approximately 500 Hz, the resulting interaural cross-correlation coefficient measurements rapidly approach a value of 1 and show much less variation (16, 17). Thirdly, stimuli with a negative cross-correlation coefficient appear to separate into two segments or be perceived to be located within the head (18) which is unlike the trend for stimuli with a positive cross-correlation coefficient that appear to increase in size as the cross-correlation falls (19).

As an alternative to the interaural cross-correlation coefficient, the authors examined the salience of quantifying the properties of the fluctuations in interaural time difference (5, 20, 21, 22, 23), based on the theories of Blauert and Lindemann (24) and Griesinger (25). It was found that altering the magnitude of the fluctuations contained within noise stimuli altered the perceived width of certain components of the sound. When the fluctuations were contained within a signal that was perceived to be a sound source, the perceived source width was altered. On the other hand, when the fluctuations were contained within a signal that was perceived to be reverberation, the perceived width of the acoustical environment was altered.

Based on these results, it is apparent that these spatial attributes may be perceived by the use of a perceptual process that detects the magnitude of the fluctuations in interaural time difference. However, it was also shown that the fluctuations in interaural time difference affect the interaural cross-correlation coefficient, and that the cross-correlation and the fluctuations in interaural time difference are intrinsically linked (6). In addition, it was found that

---

Centre, Genelec Oy and Bang & Olufsen A/S. Topics that were investigated included the subjective attributes of spatial perception and related physical cues, quantification of the requirements of subwoofers in a 5.1 surround sound arrangement, multichannel level alignment, surround sound programme production and virtual surround sound.

[2] In this paper, spatial impression is defined as “the perception of attributes of the audio which are in the first three dimensions; those of height, width and distance, including both perceived sound sources and the acoustic environments in which the sources are located”.

stimuli that manipulate either of these factors result in a similar perceived spatial effect. Due to this fact, it has not yet been possible to ascertain which, if any, of these cues are employed by the human perceptual system to detect certain spatial attributes.

It was considered that it was possible that the perceptual process that is involved in perceiving certain spatial attributes may be similar to either a cross-correlation or time difference fluctuation measurement. If this was the case, it was considered likely that the objective measurement that is most similar to the perceptual process would predict the results of the subjective evaluation more accurately. This was exploited by making measurements of the stimuli using techniques that were based on both the fluctuations in interaural time and level difference and the interaural cross-correlation coefficient, and comparing the objective and subjective results.

Subjective attribute descriptors

The experiments that were conducted by the authors to uncover the subjective spatial attributes that are affected by the fluctuations in interaural time difference indicated that it was the width of the perceived source or the perceived acoustical environment that was affected, as mentioned above. It was also found that this was similar to the perceived effect that was caused by manipulating the interaural cross-correlation of a signal. However, these descriptors were elicited using relatively simple noise stimuli, therefore further investigation was required in order to investigate whether these could be applied to more complex musical stimuli that were reproduced within a reverberant environment.

In addition to this, it was necessary to evaluate whether listening test subjects could relate to the descriptors that had been elicited, and whether they could use scales that were based on these descriptors to grade stimuli in a meaningful and reliable manner. If this was the case, then this would enable evaluations of objective measurements that relate to these attributes to be conducted more simply and without the need for further elicitation exercises.

Aims of the experiment

Based on the research that is summarised above, the experiment that is described in this paper had a number of aims.

The first aim of the experiment was to evaluate the descriptors that had been elicited in the subjective experiments that were described in (20, 21, 22). 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 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 third aim of the experiment was to evaluate the objective measurement techniques that were based on quantifying the properties of musical signals and the division of the source-related and environment-related aspects by the use of perceptual grouping, as summarised above. By comparing the measurements made using these procedures 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 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 was considered likely that the measurement technique that is closest to the perceptual mechanism will predict the subjective results most accurately.

STIMULI

In order for the stimuli that were used in this experiment to be reasonably 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 controlled manipulation of the spatial properties of the stimuli in a manner that would be externally valid was to use a number of different microphone techniques. 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 (26) and shown in Fig. 1.

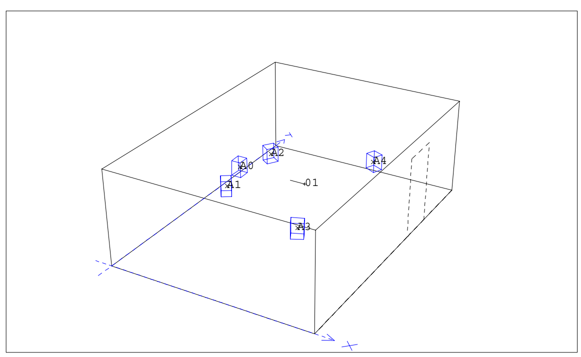

Fig. 1. Three-dimensional representation of the listening room, showing the location of the listening position (labelled 01) and the five loudspeakers in the standard five-channel arrangement, one positioned directly in front of the subject (labelled A0), two positioned at ±30º from directly in front 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 (23). 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 (6), 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 (27). Based on the research into the cross-correlation of signals that are reproduced over loudspeakers (18, 28), 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 (6). These microphone techniques are based on the well-established principles of stereophonic microphone placement (29).

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 ±45º 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 ±135º 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 authors. 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 authors. 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 (30).

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 (3, 31). 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, 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 (32). 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 (33). 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.

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. For this, 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 (34). 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 Fig. 2.

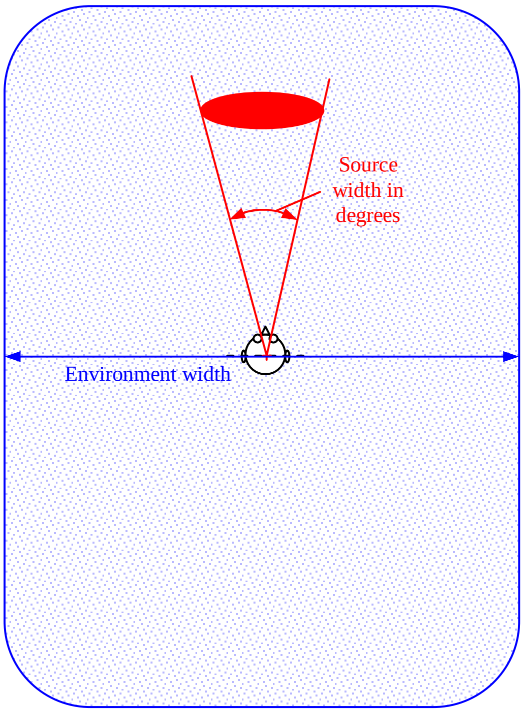

Fig. 2. The depiction that was given to the subjects as a descriptor to represent the meaning of the two grading scales that were used in the subjective experiment.

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 (21). 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 (22). 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 Fig. 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.

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 (35), the stimuli were randomised, which meant that they were presented to each subject in a different random order.

Physical set-up

The experiment was carried out in an ITU-R BS 1116 (36) standard listening room. The loudspeakers were arranged in the conventional five-channel configuration, as specified in (26). 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 Fig. 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 the stimulus creation section, 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.

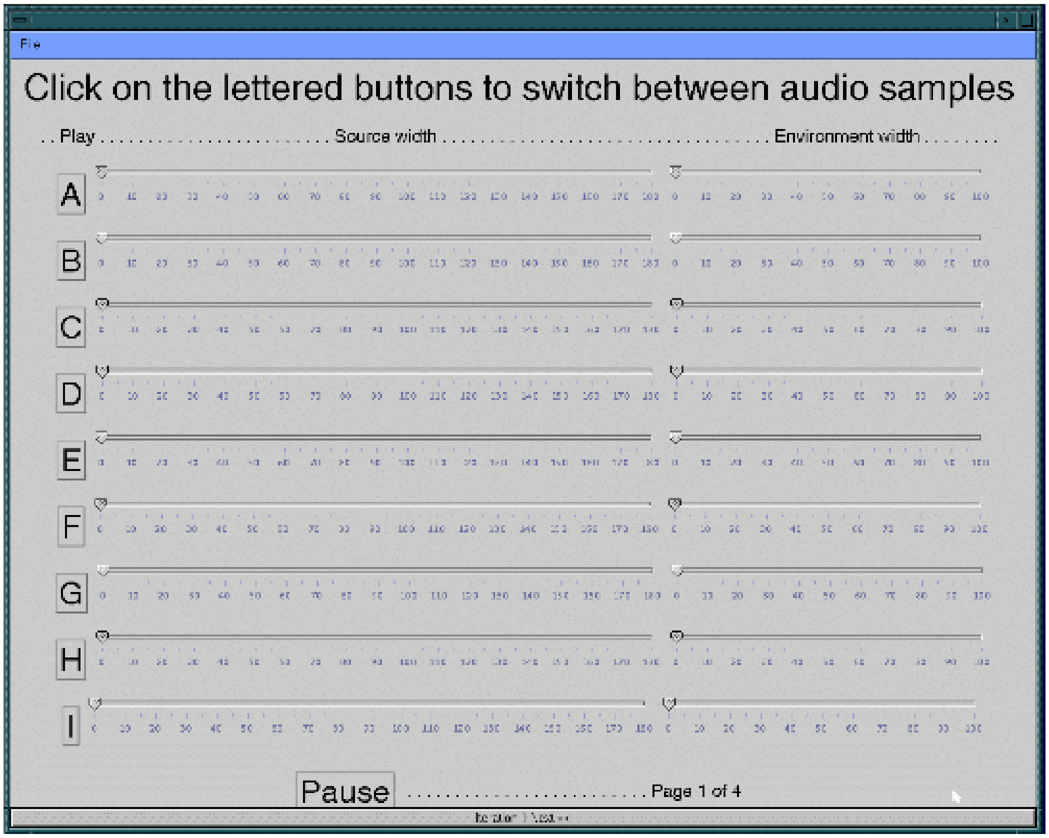

Fig. 3. Screen shot of the user interface that was employed for the subjective experiment, showing the lettered buttons for switching between the stimuli and the two grading scales for each stimulus.

Loudness alignment

It is apparent that the loudness of an auditory signal can have a large effect on how it is perceived (37). If the loudness of the stimuli was not aligned, it would be a confounding variable in the experiment. This means that any perceived differences between two stimuli may be partially due to the differences in loudness, in addition to the controlled changes that are made to the independent variables. 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 (38). 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 of 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 of similar loudness (5). The results of the loudness alignment that was conducted for this experiment were confirmed by audition by the author.

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 the method section.

This experiment was conducted using 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 (39). 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 (40). However, as this experiment was an exercise in detecting and grading the perceived quantitative spatial attributes of the stimuli, the authors 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.

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 (41). 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 (42). The significance value results of the univariate ANOVA and the one-way non-parametric Kruskal-Wallis tests are shown in Table 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 mic technique	0.000	0.000	0.000	0.000
Rear mic technique	0.011	0.010	0.000	0.000

Table 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 2. In order to test how well the ANOVA model fitted the data, the standardised residuals were analysed. This showed that the model

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

Table 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.

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 (43).

It is apparent from the results that are contained in Table 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 (41)) 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 (41).

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 (44). 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 (20, 21, 22). These descriptors were used to describe the judgement scales that were employed in the experiment that is described in this paper.

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

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 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 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.

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 (45) and Bech (39). 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.

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 (41). 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 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 3 for the judgements of source width, and Table 4 for the judgements of environment width.

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 through 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.

ANALYSIS OF THE OBJECTIVE MEASUREMENTS

The objective measurements that were made for the experiment were all based on quantifying the properties of binaural signals, as it was discussed in (6) that it was likely that this type of measurement would be more successful for sound reproduction. 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

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 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.

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.

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 (7), and following the research of Hidaka, Beranek and Okano (17). 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 paper as $IACC_{BSf}$, 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 paper as $IACC_{E3}$. 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 paper as $IACC_{L3}$.

The results for these measurements are shown in Table 5.

Perceptually grouped 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[3] 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 segment[4] 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 (6). 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 (17), 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 $PGIACC_A$ is the active sound source segment and $PGIACC_R$ 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 6.

Microphone technique	$PGIACC_A$ for the cello stimuli	$PGIACC_R$ 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 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.

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

[3] For the purpose of simplicity in this paper, the phrase ‘active sound source segment’ refers to the segment of time when there is a direct sound component reaching the receiver. During this time, the sound that is captured by the receiver is made up of both the direct sound and reverberation and in this case it is the segment of time where the direct sound of the sine tones is reaching the receiver.

[4] Again, for simplicity in this paper, the phrase ‘reverberant segment’ refers to the segment of time where the direct sound has ended and the sound reaching the receiver is purely indirect or reverberant.

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 (6). 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 $IACCFF_A$ is the active sound source segment and $IACCFF_R$ is the reverberant segment.

The measurements of the fluctuations in interaural level difference were made using the process that is described in (6). 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 $ILDFF_A$ is the active sound source segment and $ILDFF_R$ 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 7.

CORRELATION BETWEEN THE OBJECTIVE AND SUBJECTIVE RESULTS

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

The objective measurement data that are contained in Table 5, Table 6 and Table 7 were compared with the means of the subjective data that are contained in Table 3 and Table 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, 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. 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 (7) (denoted as $IACC_{BS}$ in Table 8) provide different results in different frequency bands. Only the measurement that was centred on 2 kHz was highly correlated with

Microphone technique	$IACCFF_A$ for the cello stimuli	$IACCFF_R$ for the snare drum stimuli	$ILDFF_A$ for the cello stimuli	$ILDFF_R$ 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 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.

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) 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) 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) 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) 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) 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 analysis of the subjective results section. 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 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 (7) (denoted as $IACC_{BS}$ in Table 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 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 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. One of these plots is shown in Fig. 4.

It is apparent from Fig. 4 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.

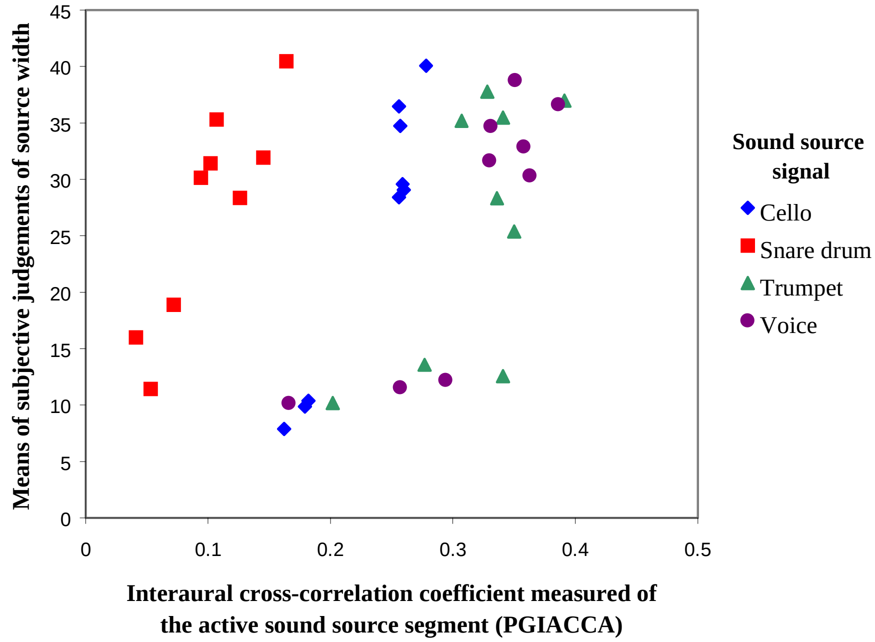

Fig. 4. Scatter plot of the means of the subjective judgements of source width and the interaural cross-correlation coefficients that were measured of the active sound source segment ($PGIACC_A$), calculated using data that was separated by the five-channel microphone technique and the sound source signal.

It is 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 found by Blauert and Lindemann (24).

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 10, and the results for the judgements of environment width are shown in Table 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.

It is apparent from the results that are shown in Table 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

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 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.

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 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.

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 ($ILDFF_{A}$).

From the results that are shown in Table 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, 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 10, and the average correlation between the measurements that related to the reverberant segment and the environment width judgements that are shown in Table 11 were combined. The results of this are shown in Table 12.

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 (6) was as successful as the more established measurements that were made of the impulse response.

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 12: Table of the overall results combining the relevant correlation results from Table 10 and Table 11 to give an overall rating for each type of objective measurement technique.

Further experimentation is needed to evaluate whether there are situations in which either of the techniques fails to predict the subjective results.

DISCUSSION

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 (20, 21, 22) 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 stimulus creation section, 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 the introduction 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 the introduction. 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 similarly. This result is not as conclusive as the authors had hoped, though it is an indication that the novel measurement is as successful as the extant measurement in this case. In addition, the authors believe 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 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.

ACKNOWLEDGEMENTS

The authors would like to thank David Meares and BBC R&D for sponsorship and provision of equipment, the MEDUSA team for their thought-provoking discussions and guidance, the AES Educational Foundation for financial assistance and Lin Oskam for her patient proof reading.

REFERENCES

(1) Grewin, C. 1995: ‘Can objective measures replace subjective assessments?’, **Audio Engineering Society Preprints**, 99th Convention, preprint no. 4067.

(2) Toole, F. 1985: ‘Subjective measurements of loudspeaker sound quality and listener performance’, **Journal of the Audio Engineering Society**, vol. 33, no. 1/2 (January/February) pp. 2-32.

(3) Rumsey, F. 1998: ‘Subjective assessment of the spatial attributes of reproduced sound’, **Proceedings of the 15th International Audio Engineering Society Conference**, Copenhagen, Denmark, pp. 122-135.

(4) Beranek, Leo. 1996: **Concert and Opera Halls – how they sound**, (Woodbury, NY: Acoustical Society of America), ISBN: 1563965205.

(5) Mason, R. and Rumsey, F. 2000: ‘An assessment of the spatial performance of virtual home theatre algorithms by subjective and objective methods’, **Audio Engineering Society Preprint**, 108th AES Convention, preprint no. 5137.

(6) Mason, R. 2002: ‘Elicitation and measurement of auditory spatial attributes in reproduced sound’, **PhD thesis**, University of Surrey.

(7) British Standards Institute. 2000: ‘Acoustics – Measurement of the reverberation time of rooms with reference to other acoustical parameters’, **British Standard**, BS EN ISO 3382.

(8) Barron, M. 1999: ‘Spatial impression and envelopment in concert halls’, **Proceedings of the Institute of Acoustics**, vol. 21, part 6, pp. 163-170.

(9) Griesinger, D. 1997: ‘The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces’, **Acustica**, vol. 83, no. 4 (July/August), pp. 721-731.

(10) Barron, M. and Marshall, A. H. 1981: ‘Spatial impression due to early lateral reflections in concert halls: the derivation of a physical measure’, **Journal of Sound and Vibration**, vol. 77, no. 2, pp. 211-232.

(11) Roy, A. and Gimbott, F. 1993: ‘Automatic echo determination in measured room impulse responses’, **Audio Engineering Society Preprint**, 94th Convention, preprint no. 3573.

(12) Blauert, J. 1997: **Spatial Hearing**, 3rd edition, (Cambridge, USA: MIT Press) , ISBN: 0262024136.

(13) Bradley, J. S. and Soulodre, G. A. 1995: ‘The influence of late arriving energy on spatial impression’, **Journal of the Acoustical Society of America**, vol. 97, no. 4 (April), pp. 2263-2271.

(14) Okano, T., Beranek, L. L. and Hidaka, T. 1998: ‘Relations among interaural cross-correlation coefficient ($IACC_E$), lateral fraction (LF_E) and apparent source width (ASW) in concert halls’, **Journal of the Acoustical Society of America**, vol. 104, no. 1 (July), pp. 255-265.

(15) de Vries, D., Hulsebos, E. and Baan, J. 2001: ‘Spatial fluctuation in measures for spaciousness’, **Journal of the Acoustical Society of America**, vol. 110, no. 2 (August), pp. 947-954.

(16) Tohyama, M. and Suzuki, A. 1989: ‘Interaural cross-correlation coefficients in stereo-reproduced sound fields’, **Journal of the Acoustical Society of America**, vol. 85, no. 2 (February), pp. 780-786.

(17) Hidaka, T., Beranek, L. L. and Okano, T. 1995: ‘Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in

concert halls', **Journal of the Acoustical Society of America**, vol. 98, no. 2 (August), pp. 988-1007.

(18) Plenge, G. 1972: 'Über das problem der im-kopf-lokalisation (On the problem of in-head localisation)', **Acustica**, vol. 26, no. 5, pp. 241-252.

(19) Chernyak, R. I. and Dubrovsky, N. A. 1968: 'Pattern of the noise images and the binaural summation of loudness for the different interaural correlation of noise', **Proceedings of the 6th International Congress on Acoustics**, Tokyo, pp. A53-A56.

(20) Mason, R., Rumsey, F. and de Bruyn, B. 2001: 'An investigation of interaural time difference fluctuations, part 1: the subjective spatial effect of fluctuations delivered over headphones', **Audio Engineering Society Preprint**, 110th Convention, preprint no. 5383.

(21) Mason, R., Rumsey, F. and de Bruyn, B. 2001: 'An investigation of interaural time difference fluctuations, part 3: the subjective effect of fluctuations in continuous stimuli delivered over loudspeakers', **Audio Engineering Society Preprint**, 111th Convention, preprint no. 5457.

(22) Mason, R., Rumsey, F. and de Bruyn, B. 2001: 'An investigation of interaural time difference fluctuations, part 4: the subjective effect of fluctuations in decaying stimuli delivered over loudspeakers', **Audio Engineering Society Preprint**, 111th Convention, preprint no. 5458.

(23) Mason, R. and Rumsey, F. 2001: 'Interaural time difference fluctuations: their measurement, subjective perceptual effect, and application in sound reproduction', **Proceedings of the 19th International Audio Engineering Society Conference**, Bavaria, Germany, pp. 252-271.

(24) Blauert, J. and Lindemann, W. 1986: 'Auditory spaciousness: some further psychoacoustic analyses', **Journal of the Acoustical Society of America**, vol. 80, no. 2 (August), pp. 533-542.

(25) Griesinger, D. 1998: 'General overview of spatial impression, envelopment, localization, and externalization', **Proceedings of the 15th International Audio Engineering Society Conference**, Copenhagen, Denmark, pp. 136-149.

(26) International Telecommunications Union. 1994: 'Multichannel stereophonic sound system with and without accompanying picture', **International Telecommunications Union - Radiocommunications**, Recommendation ITU-R BS 775.

(27) Lipshitz, S. 1986: 'Stereo microphone techniques: are the purists wrong?', **Journal of the Audio Engineering Society**, vol. 34, no. 9 (September), pp. 716-744.

(28) Keet, W. de V. 1968: 'The influence of early lateral reflections on the spatial impression', **Proceedings of the 6th International Congress on Acoustics**, Tokyo, pp. E53-E56.

(29) Rumsey, F. 2001: **Spatial Audio**, (Oxford: Focal Press), ISBN: 0240516230.

(30) Hansen, V. and Munch, G. 1991: 'Making recordings for simulation tests in the Archimedes project', **Journal of the Audio Engineering Society**, vol. 39, no. 10 (October), pp. 768-774.

(31) Olive, S. E., Schuck, P. L., Sally, S. L. and Bonneville, M. E. 1994: 'The effects of loudspeaker placement on listener preference ratings', **Journal of the Audio Engineering Society**, vol. 42, no.9 (September), pp. 651–669.

(32) Dalenbäck, B.-I. L. 1996: 'Room acoustic prediction based on a unified treatment of diffuse and specular reflection', **Journal of the Acoustical Society of America**, vol. 100, no. 2 (August), pp. 899-909.

(33) Dalenbäck, B.-I., Kleiner, M. and Svensson, P. 1994: 'A macroscopic view of diffuse reflection', **Journal of the Audio Engineering Society**, vol. 42, no. 10 (October), pp. 793-807.

(34) Bech, S. 1990: 'Listening tests on loudspeakers: a discussion of experimental procedures and evaluation of the response data', **Proceedings of the 8th International Audio Engineering Society Conference**, Washington, USA, pp. 101-115.

(35) Meilgaard, M., Civille, G. V. and Carr, B. T. 1991: **Sensory Evaluation Techniques**, 2nd edition, (Boca Raton, FL: CRC Press), ISBN: 0849342805.

(36) International Telecommunications Union. 1997: 'Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems', **International Telecommunications Union - Radiocommunications**, Recommendation ITU-R BS 1116.

(37) Gabrielsson, A., Hagreman, B., Bech-Kristensen, T. and Lundberg, G. 1990: 'Perceived sound quality of reproductions with different frequency responses and sound levels', **Journal of the Acoustical Society of America**, vol. 88, pp. 1359-1366.

(38) Moore, B. C. J., Glasberg, B. R. and Baer, T. 1997: 'A model for the prediction of thresholds, loudness, and partial loudness', **Journal of the Audio Engineering Society**, vol. 45, no. 4 (April), pp. 224-240.

(39) Bech, S. 1992: 'Selection and training of subjects for listening tests on sound-reproducing equipment', **Journal of the Audio Engineering Society**, vol. 40, no. 7/8 (July/August), pp.590-609.

(40) Kirk, R. E. 1956: 'Learning, a major factor influencing preferences for high-fidelity reproducing systems', **Journal of the Acoustical Society of America**, vol. 28, no. 6 (November), pp. 1113-1116.

(41) Howell, D. C. 1997: **Statistical Methods for Psychology**, 4th edition, (Belmont, CA: Wadsworth Publishing), ISBN: 0534519938.

(42) Zacharov, N. and Huopaniemi, J. 1999: 'Results of a round robin subjective evaluation of virtual home theatre sound systems', **Audio Engineering Society Preprint**, 107th AES Convention, preprint no. 5067.

(43) Field, A. 2000: **Discovering Statistics using SPSS for Windows**, (London: SAGE Publications), ISBN: 0761957553.

(44) Green, S. B., Salkind, N. J. and Akey, T. M. 2000: **Using SPSS for Windows: analysing and understanding data**, (New Jersey: Prentice Hall), ISBN: 013020840X.

(45) Rumsey, F. 1999: 'Controlled subjective assessments of two-to-five-channel surround sound processing algorithms', **Journal of the Audio Engineering Society**, vol. 47, no.7/8 (July/August), pp. 563-582.