THE PERCEPTUAL RELEVANCE OF EXTANT TECHNIQUES FOR THE OBJECTIVE MEASUREMENT OF SPATIAL IMPRESSION

R Mason Institute of Sound Recording, University of Surrey, Guildford, UK
TS Brookes Institute of Sound Recording, University of Surrey, Guildford, UK
F Rumsey Institute of Sound Recording, University of Surrey, Guildford, UK

1 INTRODUCTION

Research into the subjective attributes that contribute to the perceived spatial impression of concert hall acoustics and the objective measurements that may relate to these has been in progress for a large number of years. From this, it has been broadly agreed that spatial impression consists of apparent source width (ASW) and listener envelopment (LEV)[1]. However, despite the research that has been conducted into objective measurements that relate to these attributes, a consensus has not yet been reached about the most suitable measurement technique to employ. This is indicated by the fact that they are included as an appendix and are not contained in the main body of the current International Organization for Standardization (ISO) document[2]. In addition, the reader is given a number of options to choose from, with no indication of the relative merits of each.

The main measurement types that are suggested are based on either lateral fraction or cross-correlation metrics. Whilst both of these have their uses, the authors believe that the cross-correlation measurements are more universally applicable due to the fact that they quantify the properties of the signals that arrive at the ears[3]. Therefore this paper focuses primarily on measurements that are based on the interaural cross-correlation coefficient.

Although a great deal of research has been undertaken into the interaural cross-correlation coefficient as a predictor of the perceived spatial impression of concert halls, it is apparent that there are still a number of limitations to its application. Firstly, it has been shown that the results of the interaural cross-correlation coefficient vary greatly across the width of a concert hall, as shown by de Vries et al[4] and Okano et al[5]. It was found that for frequencies above 500 Hz, measurements that were made within 1 metre of the centre line were very different to the measurements that were made in the rest of the hall and that, for frequencies down to 125 Hz, the area of anomalous results extended to 5 metres from the centre line. It has also been shown that the interaural cross-correlation coefficient does not always accurately follow the perceived spatial impression in all cases, especially at low frequencies[5].

It is possible that these limitations of the interaural cross-correlation coefficient are not due to inherent problems with the cross-correlation calculation itself, but that they are caused by the manner in which the measurements are applied. The most common methods that are employed for making measurements of the interaural cross-correlation coefficient in concert halls are based on quantifying the properties of a measured impulse response. These are then subdivided into the segments of the impulse response that are thought to affect the apparent source width, and the segments of the impulse response that are thought to affect the listener envelopment.

However, recent research has indicated that this might not be the most appropriate method of applying these measurements. This paper examines this topic in more detail, and gives preliminary results from a subjective experiment to investigate one particular topic.

2 QUANTIFICATION OF THE PROPERTIES OF AN IMPULSE RESPONSE

Objective measurements that relate to the perceived spatial impression of concert hall acoustics are commonly conducted by analysing the properties of measured impulse responses, as recommended by ISO 3382[2]. This is a logical method of making such measurements, as a large number of factors can be derived from such an impulse response.

However, whilst this is an established method, it may not necessarily be relevant for relating objective measurements to a perceivable effect. 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.

The differences in the durations of a transient, wide-band impulse and a relatively continuous tonal musical signal cause the reverberant space to be excited differently. The effect of the duration of source signals on the measured interaural cross-correlation was investigated by Yanagawa, Yamasaki and Itow[6], who found that continuous steady-state signals caused the measured interaural cross-correlation to vary rapidly at first, before reaching a steady state. In contrast, measurements of transient signals never reached a steady state.

The difference in the frequency spectrum between wide-band impulsive signals and tonal musical signals also causes the reverberant space to be excited differently. This means that the direct sound of the musical signal will interact with the reflections in a manner that depends on the precise frequency content of the signal, as well as the reflection pattern of the acoustical environment. Small variations in either of these parameters may cause a significant change in the measured result, as was found by the authors in an investigation of the fluctuations in interaural time difference that were created by the interaction of a direct sound and a single reflection[7]. In this case, the sound source signal was found to have a large effect on the measured results, and as the fluctuations in interaural time difference influence the resulting interaural cross-correlation coefficient[3], it is apparent that the sound source signal also affects the measured cross-correlation.

These differences between transient and musical signals mean that measurements of the interaural cross-correlation coefficient in concert halls give different results if musical or impulsive signals are employed, as found by Griesinger[8]. He observed that the interaural cross-correlation that was measured by using a musical signal usually gave lower values than if the corresponding impulse response was measured.

The purpose of the measurement is usually to predict the perceived effect that will be caused by musical stimuli that are produced in the concert hall. If the measurement result of a musical stimulus was similar or could be derived directly from the measurement of the impulsive stimulus (i.e. without convolving a musical signal with the impulse response), then a measurement of the impulse response may be sufficient to predict the perceived spatial effect of musical stimuli. However, the differences between wide-band transient impulsive signals and the more continuous tonal musical signals mean that the measurements made of each of these signals can give different results when all other conditions are identical. In view of this, in order to predict the subjective effect of a musical signal in a specific concert hall under measurement, it appears that it would be preferable to use a musical signal for the measurement (or to use specifically created test signals whose salient characteristics are similar to those of a musical signal).

3 DIVISION OF THE SOURCE-RELATED AND ENVIRONMENT-RELATED ASPECTS OF A SOUND

In order to generate objective measurement results that match the established subjective attributes of ASW (which is related to the source) and LEV (which is related to the perceived acoustical environment or reverberation), a value is required for each of these factors. This is usually obtained by dividing the impulse response at 80 ms after the direct sound. From this, the characteristics of the impulse response before 80 ms are related to the perceived attributes of the sound source, and the characteristics of the impulse response after 80 ms are related to the perceived attributes of the reverberant environment.

This cut-off at 80 ms for measurements that relate to spatial impression was proposed by Barron and Marshall, and resulted from their investigation of the spatial impression that was caused by a direct sound and a single reflection[9]. They found that the perceived spatial impression was almost invariant with a delay time of 8 to 90 ms. They hypothesised that a reflection beyond 80 ms may be perceived to be a disturbing echo with a musical stimulus, and therefore they used this value as the maximum limit of the lateral fraction measurement that was later specifically related to the ASW. As a continuation of this research, the work that was conducted by Bradley and Soulodre indicated that the LEV was affected by the later arriving reflections[10]. From this, it was concluded that the late lateral fraction measurement should quantify the properties of the reflections that arrive beyond the 80 ms cut-off that was used for the early lateral fraction measurements.

A similar division was also developed to separate the source and environment related aspects of interaural cross-correlation measurements. Hidaka, Beranek and Okano found that measurements of the interaural cross-correlation coefficient in concert halls gave similar results when measured from the arrival of the direct sound to between 50 and 200 ms later[11]. Therefore, based on these results, they selected a cut-off at 80 ms in order to match the lateral fraction measurements. This resulted in the following two metrics: $IACC_{E3}$ calculated the interaural cross-correlation coefficient of a binaural impulse response from the arrival of the direct sound to 80 ms later and was specifically related to the apparent source width; $IACC_{L3}$ calculated the interaural cross-correlation coefficient of a binaural impulse response from 80 ms after the arrival of the direct sound to 750 ms after the arrival of the direct sound and was related to the diffusivity or envelopment of the reverberation (which may be interpreted as being similar to listener envelopment).

It is apparent from this that the specific value of 80 ms after the direct sound for the division between the source-related and environment-related segments of the impulse response was based on this being the threshold of the perception of disturbing echoes with musical stimuli. However, it must be questioned whether this is actually the case for a wide range of auditory stimuli.

Firstly, it must be considered whether the reflection that arrives beyond 80 ms after the arrival of the direct sound is the first reflection. If the first reflection arrives at the listener 100 ms after the arrival of the direct sound, then it may be perceived to be an echo. However, if there were other earlier reflections, then the particular reflection after 100 ms is unlikely to be perceived to be an echo, unless it has a particularly high amplitude[12].

Secondly, the probability of a reflection being perceived as an echo is dependent on the characteristics of the direct sound[13]. This is mainly a factor of the temporal characteristics of the signal, with reflections being perceivable as echoes as early as 3 ms after the direct sound for clicks[14], and as late as 175 ms for music[15]. These examples relate to a reflection of equal level to the direct sound, whereas in real acoustical environments the reflection will usually be at a lower level than the direct sound[16]. In this case, the delay time of the threshold of the perception of an echo will be longer.

Therefore, it can be seen that the use of the value of 80 ms after the arrival of the direct sound as a division between reflections that may cause a perception of an echo is somewhat arbitrary, as it is dependent on a number of factors.

It has also been shown that the two time segments do not affect the perceived source width and envelopment independently. Bradley and Soulodre found that the apparent source width was also affected by the properties of reflections that arrived beyond 80 ms after the direct sound[10]. The research of Morimoto and his colleagues has also indicated that the reflections that arrive before the 80 ms cut-off can affect the listener envelopment, and those that arrive after the 80 ms cut-off can affect the apparent source width[1]. Therefore it is apparent that a single time-based division of an impulse response at 80 ms after the arrival of the direct sound may not accurately separate the source-related and environment-related aspects of a sound.

4 SUBJECTIVE EXPERIMENT

The research discussed so far indicated that a measurement of the properties of an impulse response and a single time-based division of the source-related and environment-related aspects may not accurately predict the perceived spatial impression of a wide range of musical signals. In order to investigate this further, the authors conducted a controlled subjective experiment. This is summarised here, and more detailed information will be contained in future publications.

For this experiment, it was important to be able to accurately control the signals that reached the ears of the listeners in order to limit extraneous variables as much as possible. For this reason, the most practical method for conducting this experiment was to use headphone-based binaural reproduction. The use of this approach meant that the binaural signals that were measured and the signals that were judged by the listeners were as similar as possible. However, this method of reproduction did not allow for the recreation of cues that may be available by the use of head movement in a real concert hall.

The purpose of the experiment was to judge whether measurements of the properties of the different time segments of an impulse response were highly correlated with the perceived effect of various sound source signals. This required a number of different binaural impulse responses, whose properties varied independently either before or after the 80 ms cut-off. The experiment was also used to investigate problems with cross-correlation measurements that are made on the centre line of an acoustical environment. This required impulse responses from a source to several different receiver positions that were positioned either on or off the centre line of the acoustical environment.

In order to control extraneous variables as much as possible, the concert hall impulse responses were simulated in CATT-Acoustic. This allowed for greater control over the experimental variables, and minimised the risk of problems that could be caused by the impulse response measurement technique[17]. It was found that the objective and subjective results of the impulse responses that were created by this method were strongly dependent on the diffusion coefficient of the surfaces. In view of this, a relatively complex hall model was employed that included detailed estimates of the salient parameters. Using this model, binaural impulse responses were calculated from a single central source position to a number of receiver positions, as shown in Figure 1. In order to limit other variables, all the receiver positions were the same distance from the sound source, and in each case the binaural receiver was oriented to point at the source.

In addition to the factors that were affected by the lateral position of the receivers, the resulting calculated impulse responses were processed to precisely and independently vary the properties of the reflections that arrive up to 80 ms after the direct sound and those that arrive beyond this. For this, either the early or the late segment of the impulse response was manipulated to have an interaural cross-correlation coefficient of 1, by the use of a simple MS (mid / side) process[18]. This meant that the perceived spatial attributes that were affected by the cross-correlation of each of these time segments could be evaluated in a controlled manner.

The result of this processing meant that there was a total of 7 binaural impulse responses. These consisted of the raw impulse responses that were calculated by the CATT-Acoustic model for the three receiver positions, and then processed versions of the impulse responses for the central and furthest off-centre receiver positions that had a high cross-correlation either up to 80 ms or from 80 ms onwards. These are summarised in Table 1, together with the terms by which they are referenced later in the paper. In addition, the values of $IACC_{E3}$ and $IACC_{L3}$ that were measured according to the research of Hidaka, Beranek and Okano[11] are shown as a verification of the processing that was employed.

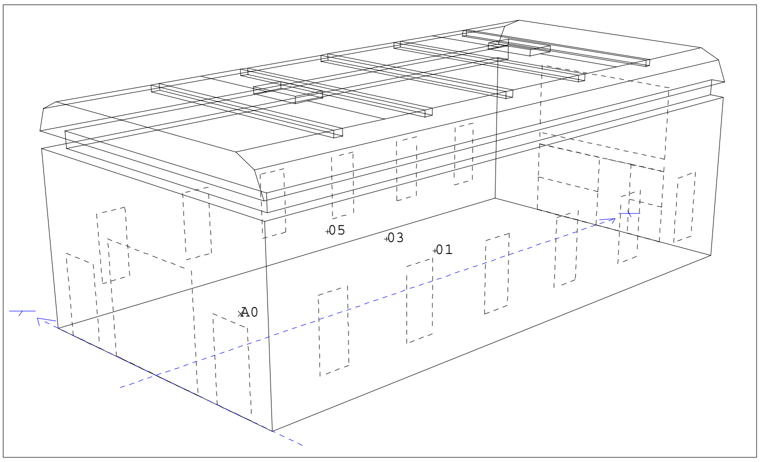

Figure 1: Three-dimensional view of the concert hall model that was used in the simulation, showing the positions of the omnidirectional sound source (labelled A0 and marked with ×) and the binaural receivers (labelled 01, 03 and 05 and marked with +).

Impulse response	Receiver position	Processing	$IACC_{E3}$	$IACC_{L3}$
raw_pos1	Central	Unprocessed	0.51	0.13
raw_pos3	2m off-centre	Unprocessed	0.26	0.10
raw_pos5	4m off-centre	Unprocessed	0.18	0.12
080_pos1	Central	Correlation coefficient of 1 up to 80 ms	1.00	0.13
080_pos5	4m off-centre	Correlation coefficient of 1 up to 80 ms	1.00	0.10
80e_pos1	Central	Correlation coefficient of 1 from 80 ms on	0.52	1.00
80e_pos5	4m off-centre	Correlation coefficient of 1 from 80 ms on	0.18	1.00

Table 1: List of the 7 binaural impulse responses that were used to create the experiment stimuli, including the terms by which they are referenced later in the paper, the receiver position, the processing that was carried out and the measured $IACC_{E3}$ and $IACC_{L3}$ values for each.

In order to investigate how these different impulse responses affected the perceived spatial impression of various types of stimulus, they were convolved with a range of sound source signals. These were selected to comprise both transient and more continuous examples, including both artificial and musical signals. In order to avoid the introduction of additional spatial or reflection-based cues, the sound source signals had to be monophonic and either recorded in an anechoic chamber or created without reverberation. As it has been shown that it is easier and more efficient to judge audio signals that are stationary and possibly repetitive[19], the stimuli consisted of a single event (a note, a hit, or a chord) that was repeated. The transient signals that were chosen were a 50 ms noise burst and a single snare drum hit. The relatively continuous signals were a sine tone chord, an acoustic guitar chord and a trumpet note. As can be seen, there was one artificial signal and at least one musical signal for both the transient and more continuous signal types.

The sound source signals were deliberately chosen to be simple single sound sources, in order to limit the variables in the experiment as much as possible, and to enable the listeners to judge the individual source width as accurately as possible.

The combination of the 7 binaural impulse responses and the 5 sound source signals resulted in 35 stimuli to be judged in total in the experiment. The listeners were asked to judge the apparent source width and the apparent environment width on two separate grading scales. The use of the subjective attribute of 'environment width' was different to the subjective attribute of 'envelopment' that is more commonly used in concert hall acoustics. This alternative descriptor was employed in this experiment as the research that has been conducted by the authors indicated that the term 'environment width' is more suitable for describing the attribute that is changed by altering the interaural cross-correlation coefficient of a stimulus that is perceived to be reverberation[3], and that 'envelopment' is also affected by the perceived level of the reverberation[10].

Ten listeners completed the experiment, and these were all staff, postgraduate or final year students in the Department of Sound Recording at the University of Surrey. The experiment was conducted using custom listening test software, which allowed the subjects to switch between the stimuli at any point in time and listen to the stimuli as often as required. Each listener took approximately 30 minutes to complete the experiment.

The initial analysis of the subjective judgements indicated that the data were not normally distributed, which meant that the data did not meet the assumptions of parametric methods of analysis. However, it has been shown that the analysis of variance (ANOVA) is robust to the violation of the assumptions of normal distribution and equal variance, as long as the samples are of equal sizes[20]. In view of this, 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[21]. It was apparent from this analysis that the results from the ANOVA and the Kruskal-Wallis tests 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 further analyse the data.

In order to highlight the perceived differences between the stimuli and to limit the effect of the different usage of the scales by the individual listeners, the data were normalised by the use of a z-transformation, as suggested in ITU-R BS 1116[22]. The ANOVA was then calculated using all the z-transformed subjective data for both judgement scales, and the type of impulse response and the type of sound source signal were entered as the independent variables. The results showed that both independent variables caused a statistically significant change in the judgements of both source and environment width, and that there was also a statistically significant interaction. The statistically significant interaction indicated that the subjective judgements of the different impulse responses needed to be examined individually for each sound source signal. The means and associated 95% confidence intervals of this analysis are shown in Figure 2 and Figure 3.

According to the established objective measurement techniques, the segment of each impulse response up to 80 ms should affect the perceived spatial attributes of the source and the segment of each impulse response after 80 ms should affect the perceived spatial attributes of the acoustical environment. Based on this, it would be expected that the processed impulse responses with a high cross-correlation up to 80 ms (080_pos1 and 080_pos5) would give a low source width, and the processed impulse responses with a high cross-correlation after 80 ms (80e_pos1 and 80e_pos5) would give a low environment width.

It can be seen from the results shown in Figure 2 that this is not the case for the source width for any of the sound source signals. In fact, for most of the sound source signals, it is the processed impulse responses with a high cross-correlation after 80 ms (80e_pos1 and 80e_pos5) that are perceived to have a low source width. In contrast, it can be seen that for the environment width, the processed impulse responses with a high cross-correlation after 80 ms (80e_pos1 and 80e_pos5) were perceived to have a low environment width, as expected.

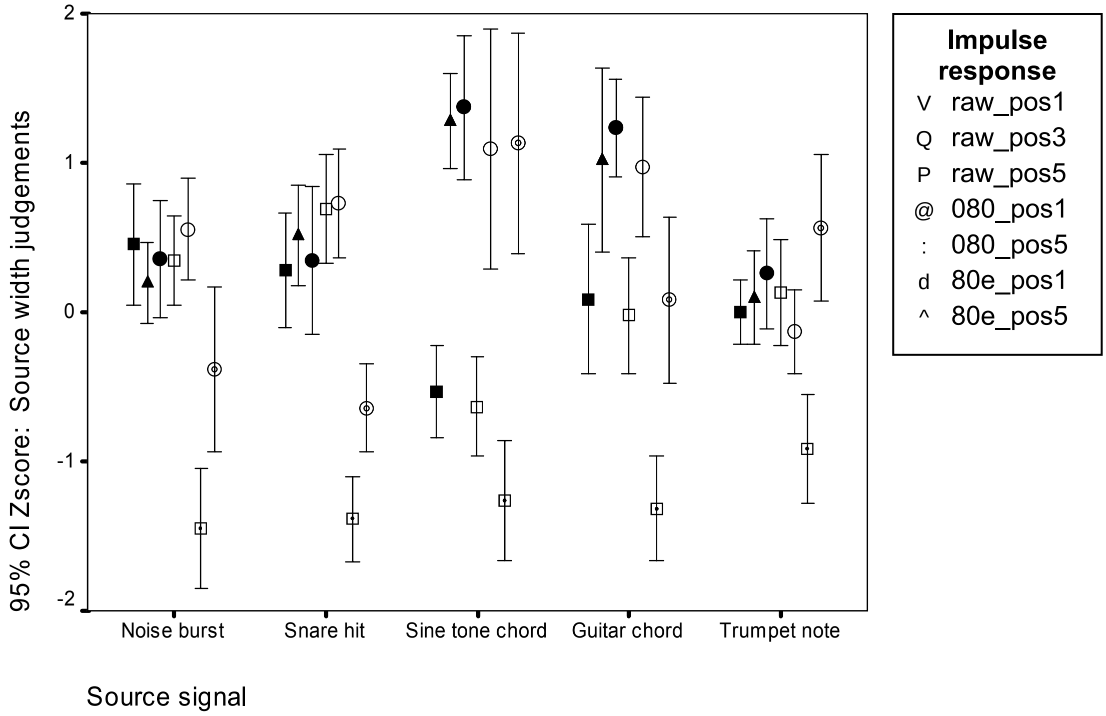

Figure 2: Plot of the means and associated 95% confidence intervals of the z-transformed subjective judgements of source width for each of the experiment stimuli.

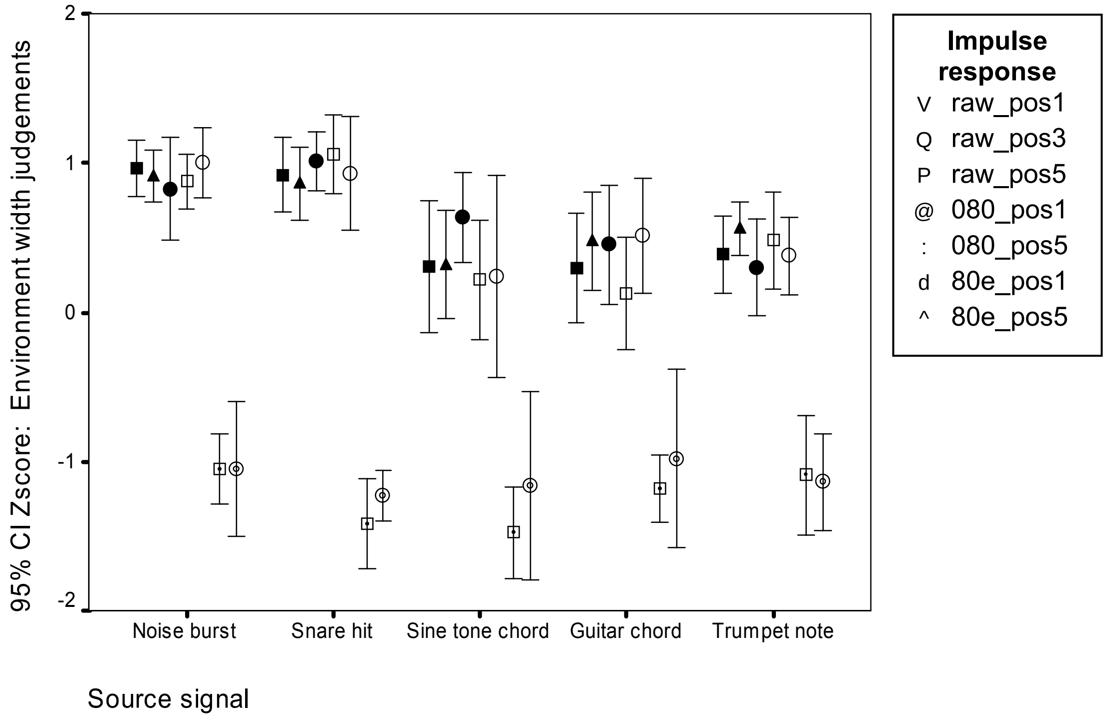

Figure 3: Plot of the means and associated 95% confidence intervals of the z-transformed subjective judgements of environment width for each of the experiment stimuli.

It is also interesting to note the differences in the judgements of source width for the different sound source signals. For the more continuous sound source signals (the sine and guitar chord stimuli), the movement of the binaural receiver from the central position (raw_pos1) to the off-centre position (raw_pos5), which caused a decrease in the $IACC_{E3}$ measurement, caused an increase in the perceived source width. However, this was not the case for the transient sound source signals (the noise burst and snare drum stimuli) where there was no statistically significant difference in the perceived source width for the different receiver positions (raw_pos1, raw_pos3 and raw_pos5). This indicates that objective measurements that accurately predict the perceived spatial attributes of one type of stimulus (either the transient or the more continuous signals) may not necessarily accurately predict the perceived spatial attributes of other types of stimuli. In addition, it must be noted that the results for the trumpet sound source signal were more similar to the transient signals than to the more continuous signals. This indicates that the relationship between the properties of the impulse response and the sound source signal that cause a certain perceived result may be more complex than a simple division between transient and more continuous signals.

Finally, it is apparent from these results that the artificial sound source signals and the musical sound source signals with similar properties were judged to have similar spatial attributes. This means that in subjective tests and objective measurements it may be possible to employ synthesised signals that are representative of the type of programme material that is of interest.

5 CONCLUSION

The results of the subjective experiment that was conducted indicated the following factors. Firstly, the interaural cross-correlation coefficient of the segment of the impulse response from the arrival of the direct sound up to 80 ms later did not appear to be correlated with the perceived source width. Secondly, the segment of the impulse response beyond 80 ms after the direct sound did not solely affect the perceived environment width but also affected the perceived source width. Thirdly, the movement from a central receiver position to an off-centre receiver position affected the perceived source width of only some of the stimuli. Finally, the perceived spatial attributes for a single impulse response were not always the same for the different sound source signals.

These results appear to support the hypothesis that measurements of the properties of impulse responses and the division of the source-related and environment-related aspects of the impulse response at 80 ms are not perceptually relevant for a range of musical sound source signals. It is possible that an approach based on perceptual grouping, as suggested by Griesinger, may be a more successful method[8]. This will be investigated further in due course.

It must be noted that this research does not indicate that the analysis of impulse responses or the division of early and late segments of reflection patterns that are based on a single time constant are never relevant. Indeed, it may be that the 80 ms division is a practical method for implementing measurements that relate to spatial impression in certain types of acoustical environments, as it is likely that the pattern of the early reflections has a strong influence on the characteristics of the later reflections. However, if objective measurements are required that relate accurately to the perceived effect of musical stimuli in a wide range of situations, then it appears that the extant measurement techniques that are most commonly applied are not appropriate, and that a method is required that takes into account the properties of representative sound source signals.

6 ACKNOWLEDGEMENTS

This research was funded by the Engineering and Physical Sciences Research Council (EPSRC) – grant GR/R55528/01. The authors would also like to thank Adrian James Acoustics and Amber Naqvi for the CATT-Acoustic model of the concert hall that was used to create the experiment stimuli, and Lin Oskam for her patient proof reading.

7 REFERENCES

1. M. Morimoto, How can auditory spatial impression be generated and controlled?, Int. Workshop on Spatial Media, Aizu-Wakamatsu, Japan (2001).
2. ISO 3382, Acoustics – Measurement of the reverberation time of rooms with reference to other acoustical parameters, International Organization for Standardization (1997).
3. R. Mason, Elicitation and measurement of auditory spatial attributes in reproduced sound, PhD thesis, University of Surrey (2002).
4. D. de Vries, E. Hulsebos and J. Baan: Spatial fluctuation in measures for spaciousness, J. Acoust. Soc. Am., vol. 110, no. 2, 947-954 (2001).
5. T. Okano, L.L. Beranek and T. Hidaka, Relations among interaural cross-correlation coefficient ($IACC_E$), lateral fraction (LF_E) and apparent source width (ASW) in concert halls, J. Acoust. Soc. Am., vol. 104, no. 1, 255-265 (1998).
6. H. Yanagawa, Y. Yamasaki and T. Itow, Effect of transient signal length on cross-correlation functions in a room, J. Acoust. Soc. Am., vol. 84, no. 5, 1728-1733 (1988).
7. R. Mason and F. Rumsey, Interaural time difference fluctuations: their measurement, subjective perceptual effect, and application in sound reproduction, Proc. 19th Audio Eng. Soc. Conf., 252-271, Germany (2001).
8. D. Griesinger, The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces, Acustica, vol. 83, no. 4, 721-731 (1997).
9. M. Barron and A.H. Marshall, Spatial impression due to early lateral reflections in concert halls: the derivation of a physical measure, J. Sound and Vibration, vol. 77, no. 2, 211-232 (1981).
10. J.S. Bradley and G.A. Soulodre, The influence of late arriving energy on spatial impression, J. Acoust. Soc. Am., vol. 97, no. 4, 2263-2271 (1995).
11. T. Hidaka, L.L. Beranek and T. Okano, Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in concert halls, J. Acoust. Soc. Am., vol. 98, no. 2, 988-1007 (1995).
12. A. Roy and F. Gimbott, Automatic echo determination in measured room impulse responses, Audio Eng. Soc. Preprint, 94th Convention, no. 3573 (1993).
13. J. Blauert, Spatial Hearing, 3rd ed., MIT Press (1997).
14. H. Babkoff and S. Sutton, End point of lateralization for dichotic clicks, J. Acoust. Soc. Am., vol. 39, no. 1, 87-102 (1996).
15. L. Dietsch and W. Kraak, Ein objektives Kriterium zur Erfassung von Echostörungen bei Musik- und Sprachdarbietungen (An objective criterion for the measurements of echo disturbances during presentation of music and speech), Acustica, vol. 60, 205-216 (1986).
16. D.M. Howard and J. Angus, Acoustics and psychoacoustics, 2nd ed., Focal Press (2001).
17. G.-B. Stan, J.-J. Embrechts and D. Archambeau, Comparison of difference impulse response measurement techniques, J. Audio Eng. Soc., vol. 50, no. 4, 249-262 (2002).
18. A. Nisbett, The use of microphones, 4th ed., Focal Press (1993).
19. S.E. Olive, P.L. Schuck, S.L. Sally and M.E. Bonneville, The effects of loudspeaker placement on listener preference ratings, J. Audio Eng. Soc., vol. 42, no.9, 651–669 (1994).
20. D.C. Howell, Statistical methods for psychology, 4th ed., Wadsworth (1997).
21. N. Zacharov and J. Huopaniemi, Results of a round robin subjective evaluation of virtual home theatre sound systems', Audio Eng. Soc. Preprint, 107th Convention, no. 5067 (1999).
22. ITU-R BS 1116, Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems', International Telecommunications Union – Radiocommunications (1997).