Effects of Interchannel Crosstalk in Multichannel Microphone Technique

Hyun-Kook Lee

Department of Music and Sound Recording

School of Arts, Communication and Humanities

University of Surrey

February 2006

Thesis submitted in fulfilment for the requirement of the degree of

Doctor of Philosophy

Effects of Interchannel Crosstalk in Multichannel Microphone Technique

Hyun-Kook Lee

Department of Sound Recording

School of Performing Arts

University of Surrey

To my parents

ABSTRACT

Even though the significance of interchannel crosstalk in multichannel microphone technique has been an issue of much debate in the field of sound recording, any effects on the perception of reproduced phantom images have not been investigated systematically. There is consequently no experimental data to which sound engineers can refer when attempting to control interchannel crosstalk in the design and application of multichannel microphone technique. It was therefore necessary to investigate the effects of such interchannel crosstalk in both the perceptual and the physical domains.

Extant multichannel microphone techniques were reviewed, concentrating on their crosstalk characteristics. Findings from concert hall and room acoustics studies relating to the effects of early reflections, which might be the basis for understanding the perceptual effects of interchannel crosstalk, were also studied.

The effects of interchannel time and intensity relationship and sound source type on the perception of stereophonic phantom image attributes were first examined in the context of two-channel stereophonic reproduction. The perceptual attributes of phantom sources affected by interchannel crosstalk in three-channel microphone technique were then elicited, and the effects of interchannel time and intensity relationship, sound source type and acoustic condition on the perception of those attributes were investigated. The effects of interchannel crosstalk on sound quality preference were also examined in both controlled and practical manners. Finally, following objective measurements of experimental stimuli, relationships were established between the perceptual and objectively measured effects of interchannel crosstalk.

It was found that the most salient perceptual effects of interchannel crosstalk were an increase in source width and a decrease in locatedness. The relationship between interchannel time and intensity differences involved in the crosstalk signal was significant for both effects. The type of sound source was significant only for the source width effect whereas the acoustic condition was significant only for the locatedness effect. The source width increase was mainly influenced by the middle frequencies of crosstalk signals in a region of the spectrum around 1000Hz, at the onsets of the signal envelopes. The results of listener preference experiments suggested that the preference for interchannel crosstalk would depend on the spectral and temporal characteristics of sound source to be recorded rather than on the magnitude of interchannel crosstalk.

ACKNOWLEDGEMENTS

The author would like to thank the following people that contributed to the completion of this thesis. Francis Rumsey for all his supervision, encouragement, guidance and inspiration throughout this research; Dave Fisher for his encouragement and discussions on the work; Slawomir Zielinski for his help with SPSS and tutorials on statistical analysis; Tobias Neher for his help with MAX-MSP; Russell Mason for the tutorials on physical measurements using his objective model; Rafael Kassier for the comments on the design of listening test; Ben Supper for his help with calculating interchannel relationships of microphone techniques; my colleagues at the Institute of Sound Recording for discussions on the work; my listening test subjects for their discussions on the grouping of common attributes; Eddie Proud for his technical support; Jorg Wuttke from SCHOEPS for providing CCM microphones; Michael Williams for his discussions and advice on the microphone technique experiments; Haeri Lee for her proofreading and encouragement; Pastors Yongbok Kim, Byungsan Chung and Seungbok Lee for their spiritual support and prayer; Seungha Shin for his friendship and prayer; my friends at the London Full Gospel Church and KOSTU for their encouragement and prayer; my sister Hyunae Lee for her love, encouragement and prayer; my parents for their love, encouragement, prayer and support during all my years in England; and finally Lord Jesus Christ for being my salvation and solid rock in all things I do.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

0 INTRODUCTION

0.1 Background to the Research

As multichannel stereophonic audio systems have become popular in recent years, a number of multichannel microphone techniques for classical music recording have been proposed corresponding to the requirement of the new reproduction configuration. The reproduction configuration that is most widely used for the current multichannel sound recording for classical music employs three front and two rear loudspeakers as recommended in ITU-R BS.775-1 [1994] (see **Figure 0.1**).

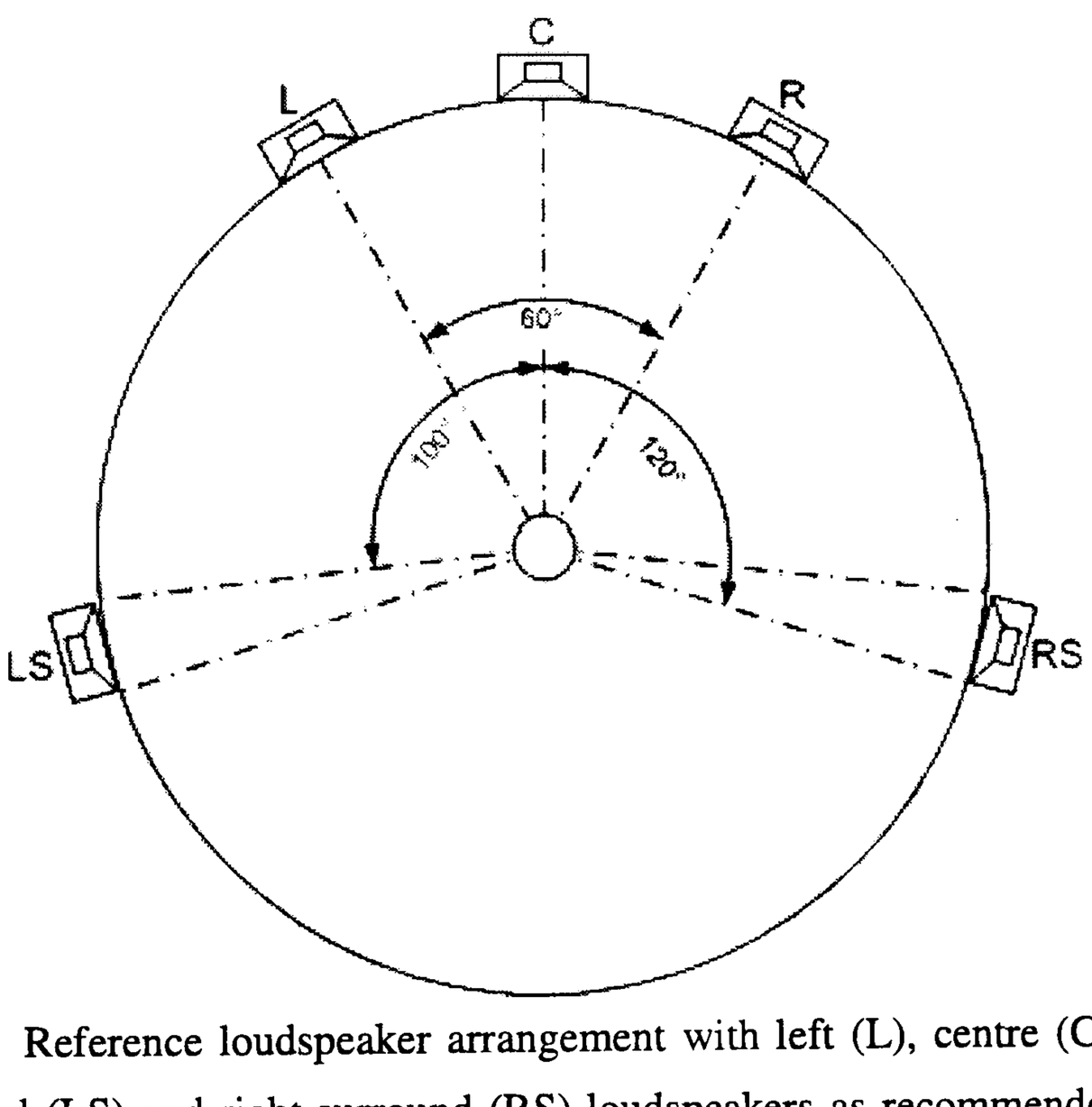


Figure 0.1 Reference loudspeaker arrangement with left (L), centre (C), right (R), left- surround (LS) and right-surround (RS) loudspeakers as recommended in ITU-R BS.775-1 [1994]

Multichannel stereophony is able to overcome some of the limitations of conventional two-channel stereophony, by adding a centre channel providing a stable centre image and two rear channels delivering an enhanced sense of spatial impression. However, the addition of extra channels in multichannel microphone techniques gives rise to a question about the effect of interchannel crosstalk, which has been a debating issue between many recording engineers recently. The current three-channel or five-channel main microphone techniques, which are discussed in detail in Chapter 1, are designed so that phantom imaging of a sound source primarily relies on the time and intensity relationship between the signals from the two microphones covering the sector of the stereophonic recording angle in which the source lies. In those types of microphone techniques, therefore, there is the implicit assumption that signals from microphones other than the pair that is primarily responsible for phantom imaging can be treated as unwanted crosstalk. For instance, as illustrated in **Figure 0.2**, if a three-channel microphone array was used for recording a single sound source located in the right recording sector of the array, signals from the microphone pair of C and R, which cover the recording sector where the source lies, would be considered to be 'wanted' while any signal from the contralateral microphone L would be regarded as 'unwanted' crosstalk. The crosstalk channel would have certain time and intensity relationships to the wanted channels depending on the distance and angle between microphones in the array and therefore the presence of the crosstalk would be likely to affect certain aspects of the perception of the phantom image, even if the location of that phantom image could be determined solely by the wanted channels.

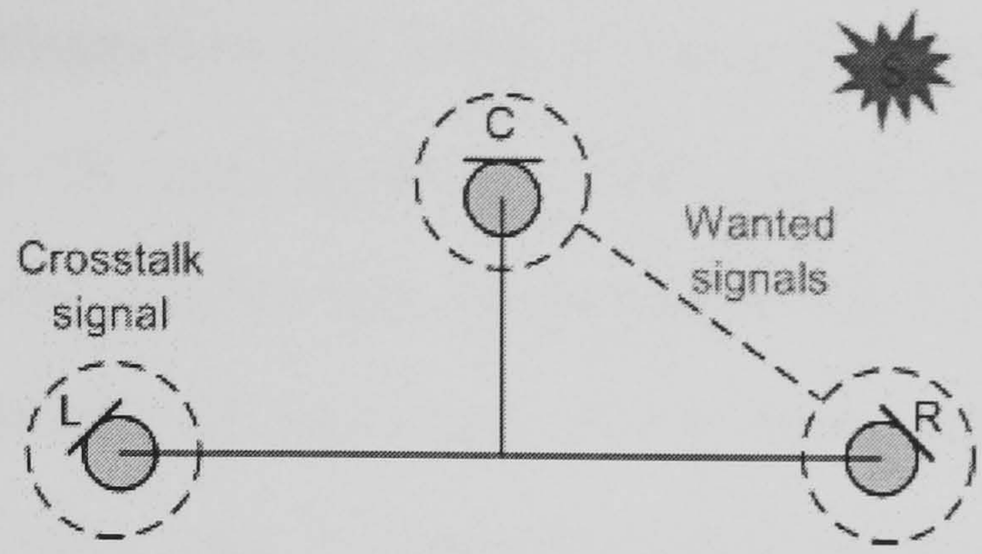


Figure 0.2 Conceptual illustration of interchannel crosstalk in a three-channel microphone array (real source shown at S)

Based on the fact that each pair of microphones (C-L, C-R and L-R) in a three-channel microphone array would pick up the sound with different interchannel time and intensity relationships, Theile [2000] claimed that interchannel crosstalk in a three-channel microphone technique would result in the perception of triple phantom images, thus decreasing the focus and clarity of phantom image localisation. From this, he suggested that in order to achieve the optimum sound image quality, microphone techniques should be designed to reduce the intensity of interchannel crosstalk as much as possible.

Theile's hypothesis concerning the perception of three separate images was questioned by Rumsey [2001]. Rumsey asserted that the listener would be likely to perceive a single fused phantom source whose 'size, stability and position are governed by the relevant intensity and time differences between the signals', and suggested a need for further experiments regarding the perceptual effect of interchannel crosstalk. In fact, there is no experimental evidence available to support the triple phantom image hypothesis.

Williams [2003] disagreed about the perceptual importance of interchannel crosstalk claimed by Theile. He argued that the interchannel crosstalk could be reduced to a great extent using directional microphones, and therefore it would not be particularly consequential. He seemed to suggest that in order to obtain a balanced and accurate localisation performance, it would be more important to link the recording angles of each stereophonic segment without overlap than to achieve the maximum suppression of interchannel crosstalk.

0.2 Aims of the Research

Despite the above debate, to date there seems to be no conclusive answer about the question of whether interchannel crosstalk in multichannel stereo microphone arrays matters or not. In fact, this topic seems to be largely dependent on the recording engineer's personal preference for the resulting sound quality since sound recording is an artistic achievement as well as a technical one. The primary problem, however, is that to date there is no clear information available about the specific influences that interchannel crosstalk has on the perception of the resulting reproduced sound, since no detailed research of which the author is aware has been conducted on this topic. There is therefore no experimental basis for discussing how interchannel crosstalk should be taken into consideration in the design of new multichannel microphone techniques or in the application of existing techniques. The more clearly the perceptual effects of interchannel crosstalk under various recording conditions are understood, the more flexible and successful the design and application of

multichannel microphone techniques will be. Based on this background, the current research was undertaken to provide in-depth experimental data on the perceptual effects of interchannel crosstalk. The specific aims of this research were to answer the following questions.

- What kinds of auditory attributes can be perceived, arising from interchannel crosstalk, and how audible are they?
- What variables in the recording environment affect the perception of crosstalk attributes?
- How are any perceptual effects related to the physical characteristics of the crosstalk signal?
- How does interchannel crosstalk influence the subjective preference for perceived sound quality?

0.3 Theoretical Basis for the Research

Prior to conducting experiments, it was first necessary to understand the psychoacoustic principles of stereophonic phantom imaging as they became the theoretical basis for the creation of the experimental stimuli. It was also important to discuss existing multichannel microphone techniques with regard to the relationships between their crosstalk characteristics and the resulting localisation characteristics, since Theile [2000] originally proposed that interchannel crosstalk would primarily affect localisation accuracy. Then the concert hall and room acoustics research

conducted on the effects of acoustic reflection needed to be reviewed. Since interchannel crosstalk and room reflections both represent secondary delayed signals and most of the reflection studies were conducted in the context of stereophonic reproduction, the perceptual attributes of reflection found in such acoustics research were expected to become the basis for formulating experimental hypotheses, which are presented in each corresponding experimental chapter. Moreover, the reflection studies show the relationships between perceived effects and various physical parameters, which became a useful basis for discussing the results of the current experiments. However, the difference between acoustic reflection and interchannel crosstalk in respect of such experimental parameters as the range of delay time and the type of sound source needed to be taken into consideration when discussing the results.

0.4 General Overview of Experimental Methodology

In order to achieve the above mentioned aims successfully, this research was conducted using a range of appropriate methods. The detailed method employed for each experiment will be described in each corresponding chapter, but this section briefly covers the type of specific technique used to collect data with respect to each research question. Firstly, the extraction of the perceptual attributes of interchannel crosstalk was achieved by analysing descriptive terms that were elicited from listeners. Secondly, the significances of the experimental variables were statistically analysed using the data obtained from a grading experiment. Thirdly, in order to examine the relationship between the physical parameters and perceived results, physical

measurements of the experimental stimuli were made using an appropriate objective model. Finally, when investigating the preference for interchannel crosstalk, subjects were asked to grade the magnitude of preference as well as describing the reasons for their judgments. Therefore, in summary, the current research involved both quantitative and qualitative approaches, incorporating both perceptual experiments and physical measurements, in order to obtain a suitably comprehensive understanding of the effects involved.

0.5 Structure of the Thesis

The remainder of this thesis is divided into six main chapters and three appendices. The outline of each part is as follows.

Chapter 1 covers the psychoacoustic principles of stereophonic sound recording and reproduction. Firstly, the interchannel relationships required for specific phantom image locations in two-channel stereophonic reproduction are discussed, followed by the review of the design principles of two-channel microphone techniques. Then, the features of imaging characteristics in multichannel stereophonic reproduction are described, and the current multichannel microphone techniques are reviewed and discussed with regard to their crosstalk characteristics.

Chapter 2 reviews the previous research relating to the perceptual effects of reflection that have been conducted in the context of concert hall and room acoustics.

Localisation, spatial impression and timbre are described as the main auditory attributes that are influenced by the addition of reflection, but only the first two attributes are considered in this review. The precedence effect is described as the law of auditory localisation in the presence of reflection. The physical parameters required for triggering this effect are discussed, and the cognitive aspects of this effect are examined. Then the conceptual properties of spatial impression are discussed and various perceptual paradigms are introduced. Finally, various objective parameters that can be used for the measurement of spatial impression are discussed.

Chapter 3 describes subjective experiments that were conducted to obtain a useful experimental basis for investigating interchannel crosstalk. The first experiment was to elicit the perceptual attributes of phantom images in two-channel stereophonic reproduction and the second experiment was to grade the magnitudes of the effects of interchannel time and intensity relationship and sound source type on the perception of those attributes. The experimental design including stimuli creation, experimental physical setup and subject selection is described. Then, for each experiment the listening test method is described and the results are discussed. The limitations of these experiments are also considered.

Chapter 4 contains descriptions of a series of subjective experiments that were conducted to investigate the perceptual effects of interchannel crosstalk in multichannel microphone technique. The first experiment was designed to elicit the relevant attributes and select the most salient of these. The second experiment employed subjective gradings of the magnitudes of perceived effects for the selected

attributes. The third experiment examined the effect of interchannel crosstalk on the subjective preference using the controlled experimental stimuli from the previous experiments. Additionally, the preference for interchannel crosstalk was investigated using practical recordings made with two different microphone techniques having different interchannel crosstalk characteristics. This chapter first discusses the microphone technique and sound source chosen for the experiments, followed by the descriptions of stimuli creation process, experimental physical setup and subject selection. Then, for each experiment the listening test method is described and the results are discussed.

Chapter 5 presents the results of objective measurements made in order to investigate the relationships between the perceived results obtained in the previous grading experiment and their physical causes. The principles of the objective model used for this measurement are summarised. Then, the measured results are compared with the perceived results for each independent variable and for each test attribute. Finally, the effects of frequency and envelope of source signal on the measured results are discussed.

Appendix A describes a two-channel localisation experiment conducted in order to investigate the individual influence of interchannel time and intensity difference on the phantom image localisations of speech and various musical sources. The stimuli and experimental method are described. The results of the experiment are statistically analysed, and the psychoacoustic data obtained for all sound sources are

unified. Finally, a new interchannel time and intensity trade-off function is proposed, and the validity of this function is verified.

Appendix B contains all the figures of the plots obtained from the measurements described in Chapter 5.

0.6 Original Contributions

- Perceptual differences between monophonic source images and the corresponding two-channel stereophonic phantom images, which had not previously been investigated systematically, have been elicited in detail (Chapter 3).
- The effects of interchannel time and intensity relationship and sound source type on the perception of the above differences have been determined (Chapter 3).
- Perceptual attributes arising from interchannel crosstalk in three-channel microphone technique have been elicited (Chapter 4).
- Detailed analysis has been performed on the effects of interchannel time and intensity relationship in microphone technique, sound source type and acoustic condition on the perceived magnitudes of crosstalk attributes (Chapter 4).
- Dependency of the preference for interchannel crosstalk on the type of sound source has been suggested from a systematic subjective comparison between OCT and ICA-3 three-channel microphone techniques, which differ in their interchannel crosstalk characteristics (Chapter 4).

- Dependency of the source-width-increasing effect of interchannel crosstalk on the spectrum and signal envelope of the sound source has been proposed from objective measurements of experimental stimuli that were made using a perceptual model (Chapter 5).
- A novel hypothesis on the mechanism of locatedness perception has been suggested based on the combination of the precedence effect and the localisation lag effect (Chapter 5).
- Original psychoacoustic values of interchannel time and intensity differences required for the localisation of phantom images at 10°, 20° and 30° between loudspeakers in two-channel stereophonic reproduction have been obtained from localisation experiments using speech and various musical sound sources, which had not been used in previous experiments of a similar type (Appendix A).
- Novel interchannel time and intensity trade-off functions for the phantom image shifts of 10°, 20° and 30° have been devised using the psychoacoustic values obtained in the above localisation experiments (Appendix A).

0.7 Summary

This chapter firstly presented the background to the research and determined the aims of the research. Then, the theoretical basis for this research and the general experimental methodology were overviewed. The structure of this thesis was outlined, and finally the original contributions of this research were summarised.

1 PSYCHOACOUSTIC PRINCIPLES OF STEREOPHONIC RECORDING AND REPRODUCTION

This chapter is concerned with the psychoacoustic principles of stereophonic recording and reproduction. Since interchannel crosstalk is a property of multichannel stereophonic microphone technique, it will first be necessary to understand the basic theories of stereophonic phantom imaging, which become the basis for the design of stereophonic microphone technique, and to review the existing multichannel stereophonic microphone techniques concentrating on their crosstalk characteristics. Internationally the configuration of multichannel stereophonic reproduction systems are termed '*n*-*m*' stereo, where *n* is the number of front channels and *m* is the number of rear (surround) channels [Rumsey 2001]. Therefore, the conventional two-channel stereophonic system is called '2-0' stereo whereas the five-channel system is called '3-2' stereo. In the scope of the current study, only the context of classical music recording is considered. Since it is not a usual trend to employ the sub-woofer channel in the multichannel recording and reproduction of classical music, the term '3-2' stereo will be used in this review rather than the popular term '5.1' surround. In this chapter, the aspects of 2-0 and 3-2 stereo will be discussed in turn. For each, the principles of phantom image localisation will be covered first and then the microphone techniques designed on the basis of those principles will be reviewed.

1.1 Phantom Imaging Principles for 2-0 Stereophonic Reproduction

The psychoacoustic principles of phantom image localisation for conventional two-channel stereophonic reproduction have been extensively studied in the field of audio engineering since the beginning of stereophonic recording in the 1930s. These principles also become the basis for the design of stereophonic microphone techniques. Localisation in 2-0 stereophonic reproduction is basically governed by the interchannel relationship between the two loudspeaker signals and this should be distinguished from the interaural relationship between the ear input signals. The latter is formed depending on the former through acoustic crosstalk between the ears and this causes the localisation of phantom images to be limited within the spread of the two loudspeakers.

1.1.1 Summing Localisation

In 2-0 stereophonic reproduction, when both loudspeakers radiate coherent signals, the listener will perceive a single phantom image on the median plane between the two loudspeakers. If one of the signals is delayed or attenuated in a small range up to 1.1ms or 15-18dB respectively, the position of the single image will be shifted from the middle toward the earlier or louder loudspeaker [Blauert 1997]. This effect is called 'summing localisation' and it becomes the basis for the phantom image localisation in stereophonic sound reproduction. If the delay time exceeds 1.1ms, the

phantom image will constantly appear at the earlier loudspeaker by virtue of the 'precedence effect', which will be discussed in detail in the next chapter.

Since 1940, a number of researchers carried out subjective experiments based on the summing localisation theory in order to investigate the independent influence of interchannel time difference (ICTD) or interchannel intensity difference (ICID) on the localisation of phantom image (e.g. de Boer [1940], Leakey [1959], Mertens [1965], Simonsen [1984], Wittek [2000]). The data from different researchers vary a lot and this seems to be due to the use of different experimental methods and different sound sources. **Table 1.1** presents a summary of the psychoacoustic data obtained by several researchers who used natural sound sources, including the data obtained from the author's own localisation experiment described in Appendix A. It can be seen firstly that for both ICID and ICTD the values obtained by de Boer [1940] are much greater than the values obtained by the others. de Boer's reports do not indicate the values required for the full image shift. Simonsen [1984]'s data obtained using speech and maracas are arguably the most widely quoted data for the design of two-channel stereophonic microphone techniques, for example the design of Williams [1987]'s near-coincident microphone techniques is based on Simonsen's data. It appears that his ICID values required for the image shifts of 10°, 20° and 30° are approximately 2–3 dB lower than those of Wittek [2000] and the author [2004] (see Appendix A), which is considered to be significant, although there is no such obvious difference between their ICTD values. It is interesting to find that Wittek and the author's data are very similar to each other with regard to both ICID and ICTD. It is not totally clear why there is such a big difference between Simonsen's and Wittek's

or the author's data. However, considering that Simonsen's experiments used only two subjects, it seems unreasonable to apply these values directly without verification. Additionally, it appears that the psychoacoustic value required for the full phantom image shift in summing localisation is approximately double the value required for the full lateral displacement in binaural localisation. This difference in the influences of 'interchannel' and 'interaural' cues is due to the acoustic crosstalk that inevitably arises in stereophonic loudspeaker reproduction.

Researcher		**De Boer** [1940]	**Simonsen** [1984]	**Wittek** [2000]	**Lee (author)** [2004]
Sound source		Speech	Speech / maracas	speech	Speech / various
ICID	10°	5dB	2.5dB	4.4dB	4.0dB
	20°	11dB	5.5dB	8.8dB	8.4dB
	30°	not indicated	15dB	18dB	17.1dB
ICTD	10°	0.7ms	0.20ms	0.23ms	0.27ms
	20°	1.7ms	0.44ms	0.45ms	0.50ms
	30°	not indicated	1.12ms	1.0ms	1.1ms

Table 1.1 Comparisons of interchannel intensity and time differences required for particular phantom image positions in stereophonic loudspeaker reproduction

1.1.2 ICTD and ICID trading in summing localisation

When summing localisation is effective, the direction of a stereophonic phantom image can be determined by a combination of ICTD and ICID. This becomes the basis for the design of near-coincident stereophonic microphone techniques such as

'ORTF' and 'NOS', which will be discussed later. The most widely quoted example of ICTD – ICID trading-off might be the curves that were created by Williams [1987] based on Simonsen's data. As can be seen in **Figure 1.1**, various combinations of ICTD and ICID can cause the phantom image to appear at different positions between loudspeakers in the conventional stereophonic arrangement.

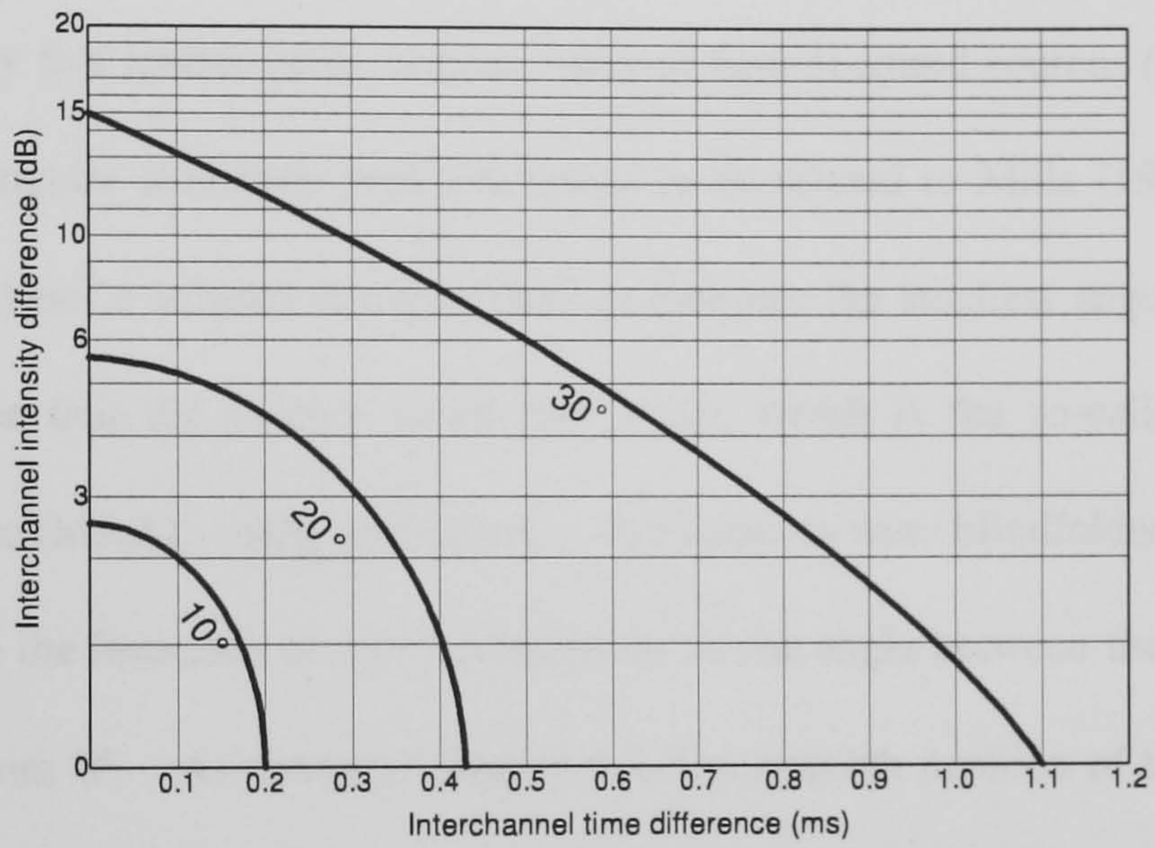


Figure 1.1 Interchannel time and intensity trading in 2-0 stereophonic reproduction [after Williams 1987]

It was proposed by Theile [2001] that the degree of phantom image shift (Ψ) could be calculated simply by the linear combination of ICTD and ICID, as shown below. If the phantom sound source is shifted due to certain ICID and additionally due to certain ICTD, the resulting shift is approximately the sum of both single shifts.

$$\Psi(\Delta I, \Delta t) = \Psi(\Delta I) + \Psi(\Delta t)$$

However, the above theory of simple linear combination would work only in a limited image shift region since stereophonic reproduction has a problem of angular distortion. Wittek and Theile [2002] pointed out that localisation curves of pure ICTD or ICID that have been introduced in the literature generally show linear progressions up to about 75% (22.5°) of the shift region, and beyond 75% the curves tend to become exponential. A similar tendency was found from the localisation test that was conducted by this author using various types of natural sound sources (see Appendix A). This angular distortion problem seems to be related to Mills [1958]'s finding. Mills carried out a subjective experiment to measure the smallest angular change of sound source that the listener could just detect, which is the so-called 'minimum audible angle (MAA)', using pure tones. The listeners were blindfolded and asked to discriminate the locations of two loudspeakers as the angle between them was varied gradually from 0°. As shown in **Figure 1.2**, the azimuth position of the centre axis of the loudspeaker pair was also varied from 0° to 75°. It was found that the MAA became larger as the loudspeaker pair moved away to the side of the listener. This result seems to suggest that in stereophonic reproduction the listener's sensitivity for localising a phantom source decreases as the direction of the source moves from the front to the side.

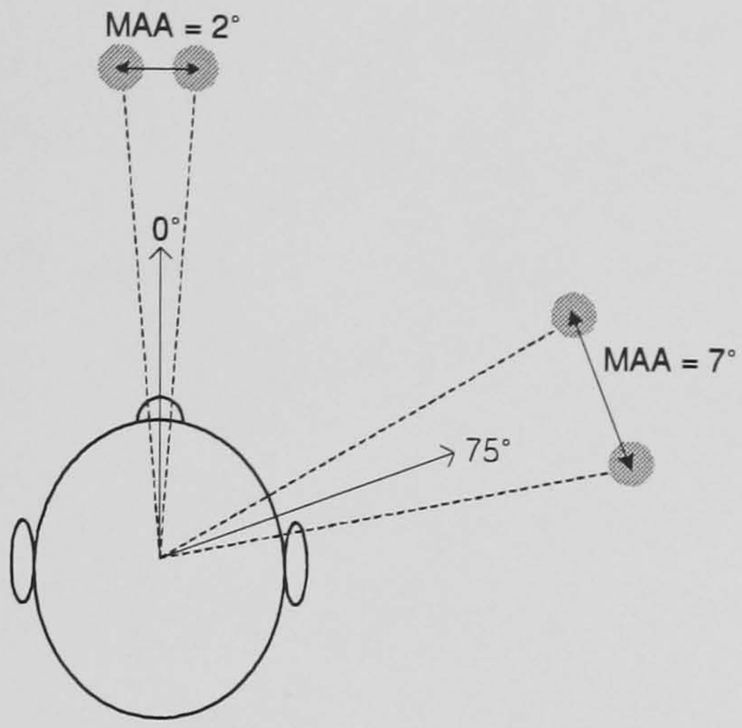


Figure 1.2 Schematic diagram of minimum audible angles (MAA) between two loudspeakers measured directly in front of the listener and at 75° toward one side of the listener [Mills 1958]

There is a report in the context of headphone reproduction that for an auditory image created with a combination of interaural time difference (ITD) and interaural intensity difference (IID), the so-called 'time image' and 'intensity image' can be perceived separately [Whitworth and Jeffress 1961]. It is thought that a similar effect could be observed also in a stereophonic reproduction depending on the combination ratio between ICTD and ICID. If this is the case, this finding of imperfect time-intensity trading might support Theile [2001]'s hypothesis that multiple phantom images could be perceived due to interchannel crosstalk in multichannel microphone technique.

1.2 2-0 Stereophonic Microphone Techniques

The designs of conventional two-channel stereophonic microphone techniques are based on the psychoacoustic principles of stereophonic localisation that were

discussed in the above sections. Conventional two-channel microphone techniques can be divided into three main types by their design concepts: coincident pair technique, spaced pair technique and near-coincident pair technique. As mentioned briefly earlier, for the imaging of a sound source, the coincident pair technique primarily uses the ICID; the spaced pair technique uses the ICTD; and the near-coincident technique uses a combination of the ICTD and ICID. It will be logical to discuss the design principles and operational characteristics applied for these conventional techniques prior to discussing those for the recently developed multichannel microphone technqiues, since the latter is based on the former to a great extent.

1.2.1 Stereophonic recording angle (SRA)

The stereophonic recording angle (SRA) can be defined as the sector of the sound field in front of the microphone array that is localised at fully left or right between the two loudspeakers and becomes an important parameter for designing a stereophonic microphone technique [Williams 2004]. The SRA is not necessarily equal to the angle between the microphones, but is determined by the horizontal angle of sound field that produces the interchannel difference required for the full phantom image shift for a given microphone technique. This is controlled by the angle or distance between the microphones, or the combination of both depending on the type of design concept for the microphone technique. For instance, for a given sound source position, when the angle between two uni-directional microphones is increased, the

ICID that is produced for the sound source will be increased but the SRA will be decreased relatively. Similarly, when the distance between two omni-directional microphones is decreased, the ICTD will be decreased but the SRA will be increased.

However, since the value of ICTD or ICID required for a particular phantom image shift tends to vary depending on the source of data one relies upon, as shown in **Table 1.1**, the calculation of the SRA would also be dependent on which psychoacoustic values are used, although Simonsen [1984]'s values have been most practically used to date. Based on Simonsen's data, Williams [1987] calculated the relationship between SRA and specific combinations of angle and distance between various directional microphones. The results obtained for cardioid microphones are shown in **Figure 1.3** and these are the so-called 'Williams curves'. Wittek [2001] developed a tool for the design of two- or three-channel microphone technique called 'Image Assistant' (see **Figure 1.4**), which enables one to calculate the SRA as well as the localisation curve and angular signal relationship based on the microphone polar pattern and the angle and distance between microphones that are controlled by the user. The psychoacoustic principles for this model are based on the interchannel trading relationship proposed by Theile [2001] and the interchannel difference data obtained by Wittek [2000], and therefore the SRA based on this tool will differ from that based on the Williams curves. It is considered that the Image Assistant seems to provide a more flexible and precise way of calculating the SRA than the Williams curves since, with the former, combinations of virtually any microphone angles and distances together with various microphone polar patterns are possible.

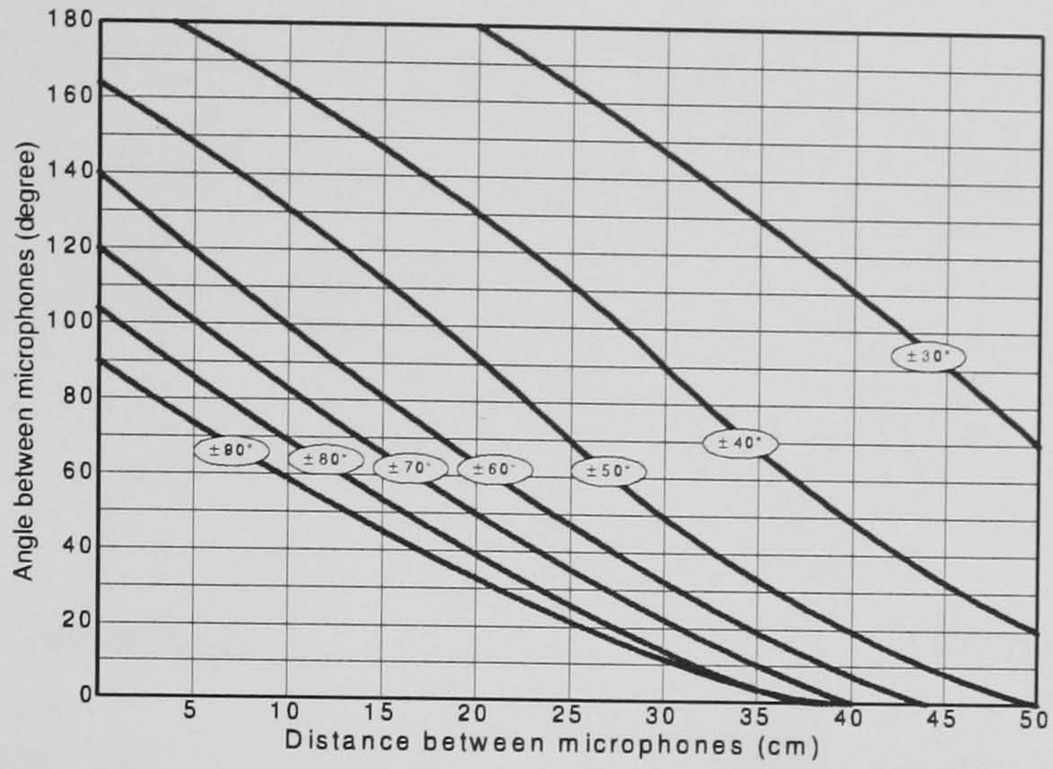


Figure 1.3 SRA diagram for cardioid microphones [after Williams 1987]

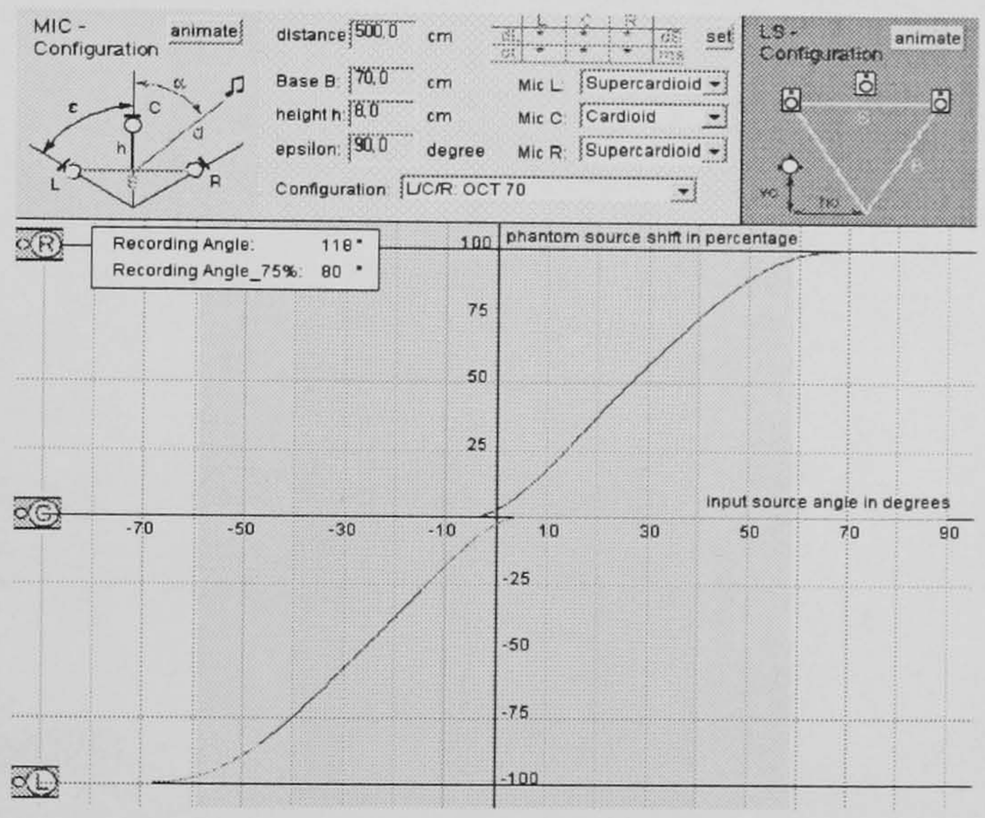


Figure 1.4 Layout of the 'Image Assistant' tool [Courtesy of Wittek 2001a]

The SRA of a microphone array will be a crucial factor for the recording engineer to control the amount of space between the loudspeakers that is occupied by the phantom images in the reproduction. For example, as can be seen in **Figure 1.5**, when the SRA is greater than the extent of the sound source ensemble, the extent of the reproduced phantom sources will be narrower than that of the loudspeakers. On the other hand, when the SRA is smaller than the extent of the ensemble (see **Figure 1.6**), the sound sources that are located at the left and right limits of the SRA and outside

will be reproduced at fully left and right respectively. In this case a linear distribution of the phantom images becomes impossible even though the phantom images are created at the full stereophonic extent.

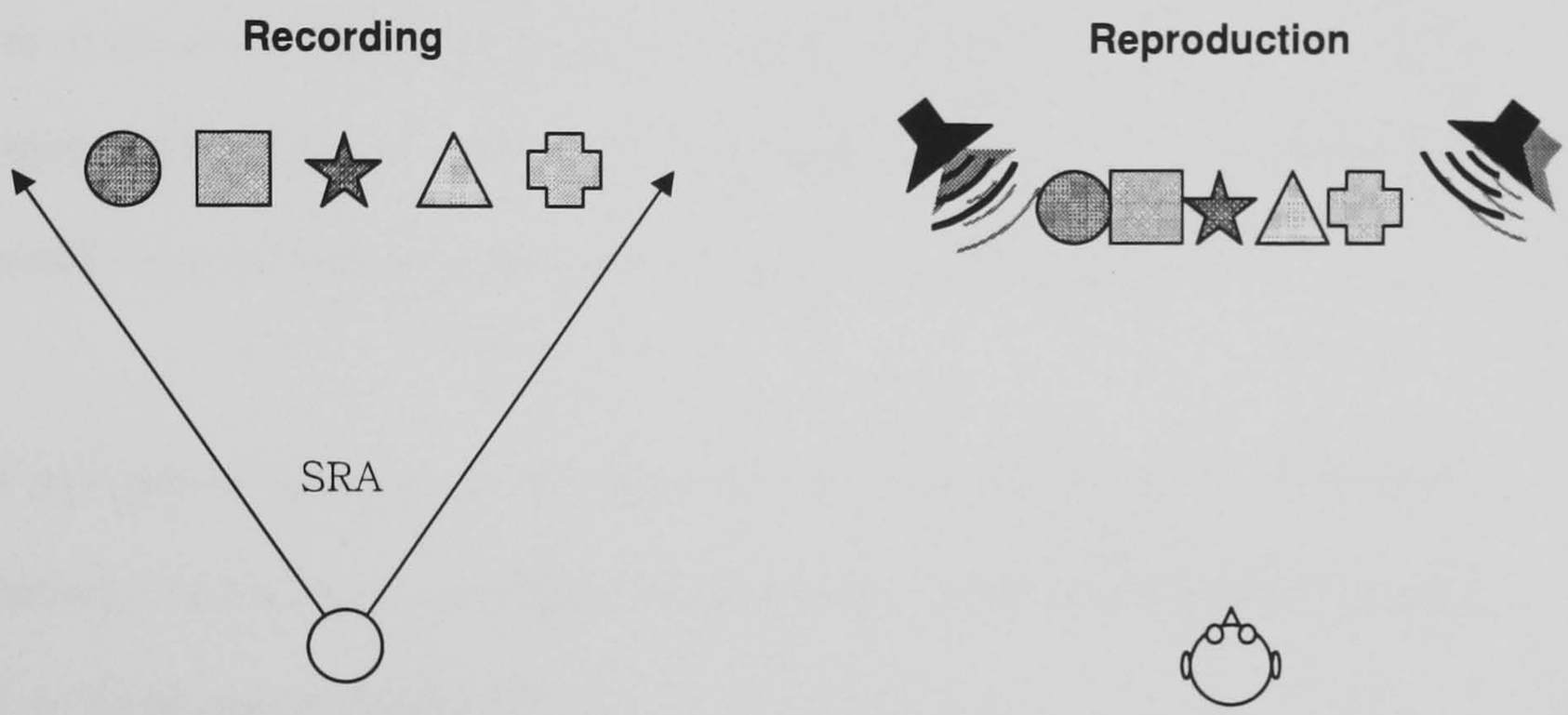


Figure 1.5 Stereophonic recording and reproduction in relation to the stereophonic recording angle; when the SRA is greater than the spread of the sound sources

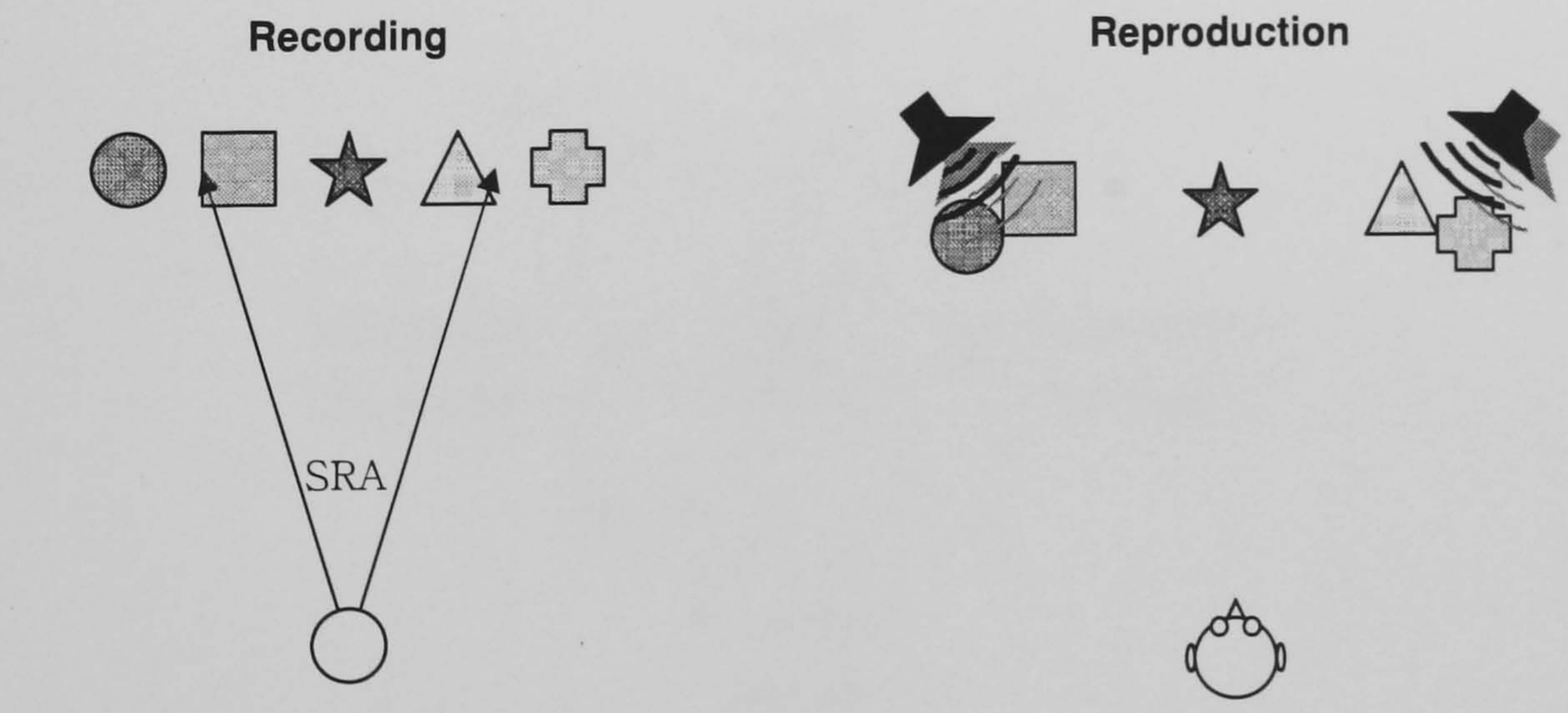


Figure 1.6 Stereophonic recording and reproduction in relation to the stereophonic recording angle; when the SRA is smaller than the spread of the sound sources

1.2.2 Coincident pair microphone technique

Coincident pair microphone techniques consist of two directional microphones that are placed together, with the angle between the microphones usually being adjustable depending on the SRA that is desired. The phantom imaging of a coincident array is based on the summing localisation principle that converts pure ICID to low frequency (<700Hz) ITD [Clark *et al* 1958]. Due to the spacing between the microphones, little time difference information is encoded between the microphone channels.

The microphone technique of this type that was first developed is the 'Blumlein' technique, which uses a pair of figure-8 microphones arrayed at a fixed lateral angle of 90° as can be seen in **Figure 1.7**.

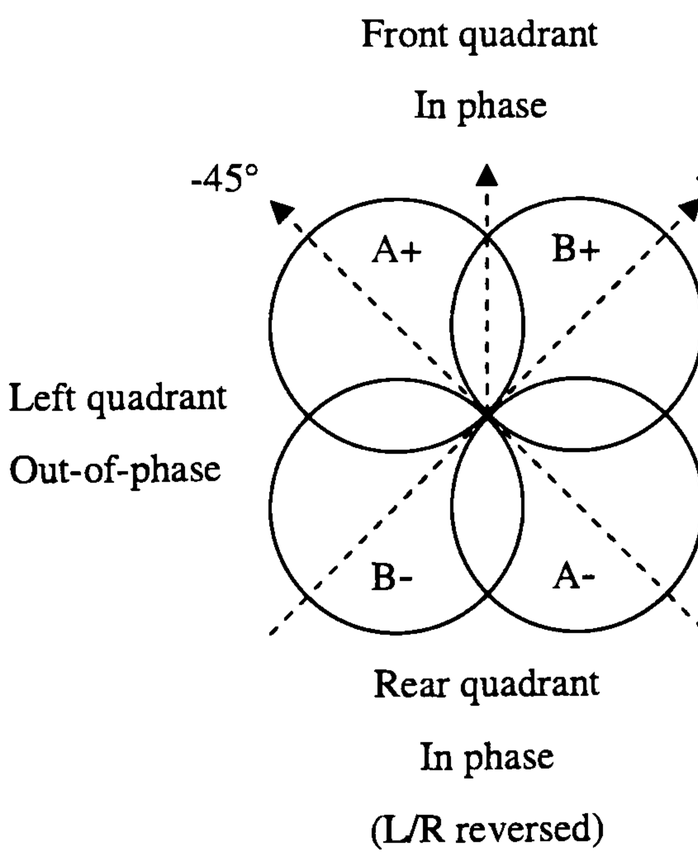


Figure 1.7 Configuration of the 'Blumlein' coincident pair technique

In this technique the maximum ICID is caused at ±45°, where the off-axis of one microphone corresponds to the on-axis of the other microphone. The intensity of the summed signals for any sound source located in the front quadrant of the array remains constant due to the identical cosine response over the whole pickup angle [Eargle 2001, Rumsey 2001]. The reversed polarity in the side quadrants results in out-of-phase information. The sound picked up in these regions, typically being reflections or reverberation, will suffer from a spatial ambiguity [Eargle 2001] and cancellation if the channels are summed to mono [Rumsey 2001]. The sound picked up by the rear quadrant is in-phase but the left - right polarity of the reproduced image will be opposite to that of the front quadrant. The localisation curve for this array, which is calculated based on the image assistant, is shown in **Figure 1.8**. The SRA for this array is 72°, which means that this array may be required to be placed far away from the performance stage in order to achieve a linear distribution of phantom sources. It has to be noted that the SRA of a coincident array does not vary with the distance of sound source from the array. The Blumlein array generally provides 'crisp' and 'accurate' phantom imaging in the reproduction [Rumsey 2001] as well as a 'good sense of acoustical space' [Eargle 2001].

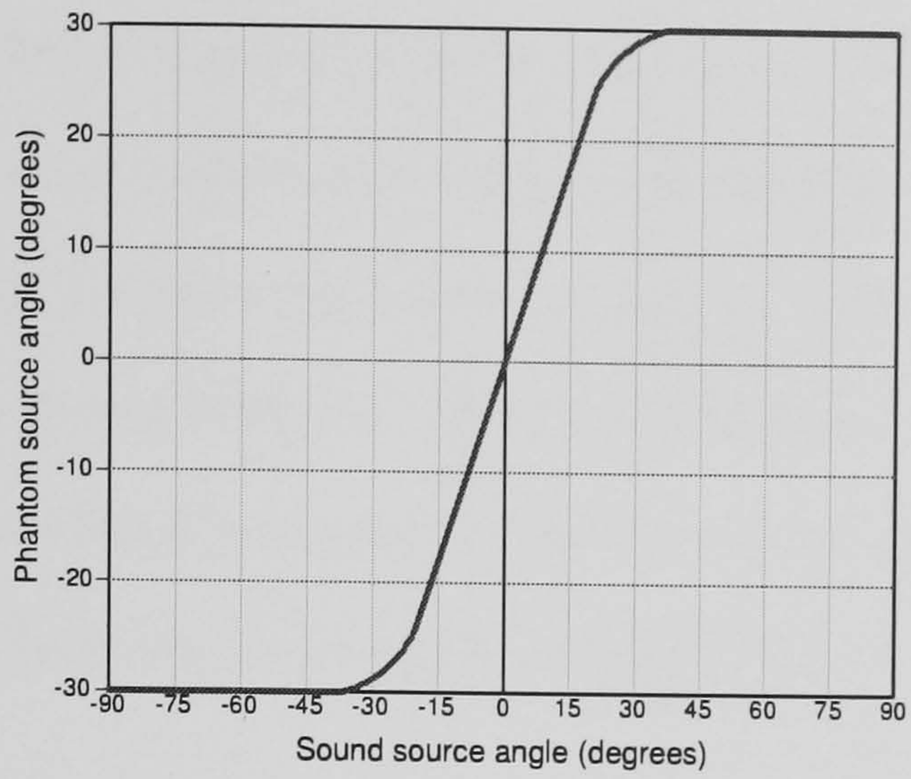


Figure 1.8 Localisation curve for the 'Blumlein' array, calculated using the Image Assistant [Wittek 2001a]; the SRA is 72°.

When a front-biased stereophonic image is desired, microphones of cardioid patterns can be used instead of figure-8 patterns. A cardioid-crossed coincident pair technique is normally called XY technique. The angle between the microphones for an XY array normally varies from 90° to 180° depending on the SRA desired for specific recording situations. For instance, when a fully wide phantom imaging of sound sources is desired, as the microphone array moves farther away from the sound sources, the angle between the microphones has to be increased to reduce the SRA. Examples of the lateral angles that are most popularly used for XY techniques are 90°, 131° and 180° (the so-called back-to-back). The SRAs calculated for the microphone arrays with these later angles using the image assistant model are 180°, 136° and 92° respectively. The polar pattern of microphone can also be changed to super-cardioid or hyper-cardioid with a corresponding lateral angle for the desired SRA. A practical example is a crossed pair of super-cardioids with a lateral angle of 120° with the SRA of 98° [Eargle 2001]. Similarly to the figure-8 pair technique, the cardioid-crossed pair techniques are advantageous for accurate phantom imaging

[Eargle 2001]. It also has a good monophonic compatibility as there is virtually no comb-filter effect that is caused by phase cancellation [Streicher and Everest 1998]. However, the cardioid-crossed pair techniques in general have a 'poor sense of acoustical space' [Dooley and Streicher 1982] due to the lack of interchannel time difference information and a poor frequency response of the central signal due to the wide angle between the microphones facing the off-axis [Rumsey 2001].

1.2.3 Spaced pair microphone technique

Spaced pair microphone techniques have been widely used since they were first introduced in the 1930s [Steinberg and Snow 1934, cited in Snow 1953]. Two omni-directional microphones are most frequently used for a spaced array since they tend to provide a wider and flatter frequency response than uni-directional ones [Rumsey 2001].

For a spaced omni array the amount of interchannel time or intensity differences resulting from a sound source at a particular lateral position largely depends on the distance of the source from the array. When the distance is very short (1–2m), both ICTD and ICID can be effective for the localisation of phantom source. However, as the microphone array is moved away from the source, the ICID will become negligible and the localisation will be mainly governed by the ICTD. This means that the SRA also can vary to some extent depending on the distance between the microphone array and sound source.

The spacing between the microphones is taken into account for the linearity of phantom source distribution. According to the Image Assistant model, when the microphones are spaced about 40cm apart and sound sources are five metres distant from the array, the SRA becomes 98°, which means that the phantom images for the sound sources located at greater than ±49° from the centre axis of the array will be localised at fully left or right (see **Figure 1.9**). However, if the microphone spacing is increased to 1m with the identical source distance, the phantom images for the sound sources located at greater than only ±18° will come to be localised at fully left or right and this causes a perception of the so-called 'hole in the middle' effect (see **Figure 1.9**). This strong microphone spacing dependency of phantom image distribution in spaced omni arrays is caused because the ICTD required for triggering the precedence effect is produced at smaller angle of sound source as the spacing between the microphones is increased. From the above, it might be important for recording engineers to consider the width of sound sources and the distance between the sound sources and microphone array when they decide the spacing between the microphones. Dooley and Streicher [1982] propose that the spacing between microphones need to be between 1/3 and 1/2 of the total width of the sound sources in order to achieve a satisfactory phantom image localisation.

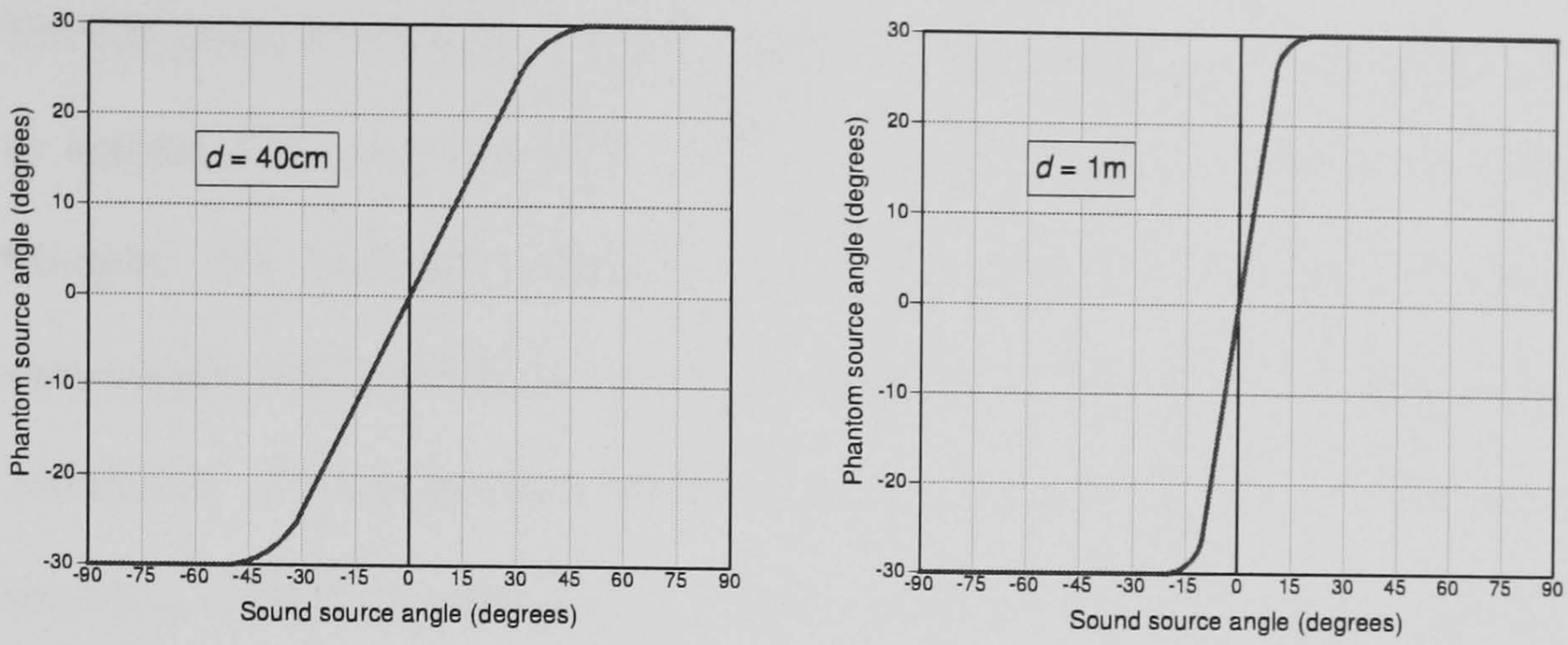


Figure 1.9 Comparison of the localisation curves for the spaced omni arrays with different distances between microphones (*d*), calculated using the Image Assistant [Wittek 2001a]

It is widely accepted that accuracy of phantom image localisation for a spaced omni array is not as good as that for a corresponding coincident array. It is pointed out by Rumsey [2001] that this is particularly true for continuous sounds as the precedence effect is triggered mainly by transient sounds (A more detailed discussion on the characteristics of the precedence effect is presented in Section 2.2.1). Furthermore, spaced omni arrays tend to suffer from low frequency comb filtering due to the lack of phase coherence at low frequencies between the sound arriving at each microphone [Dooley and Streicher 1982]. However, the highly decorrelated signals caused by these arrays are claimed to provide a good sense of spatial impression to recordings and this makes spaced pair techniques suitable for ambience pickup in multichannel stereophonic recordings [Rumsey 2001]. As the distance between the microphones for a spaced array is increased, there will be more decorrelated ambient sounds such as reflections and reverberation picked up by the array, thus increasing the spatial impression.

The hole-in-the-middle effect resulting from a large spacing between microphones can be avoided if an additional microphone is added for a central sound pickup. For example, the so-called 'Decca tree' technique uses three omni-directional microphones and is known to produce a solid centre image as well as good spatial impression. The operational principle of this technique will be discussed in s later section as it might be considered to be more useful for the purpose of multichannel stereophonic recording.

In addition, for recordings of wide sources such as large scale orchestra and choir, spaced pair techniques are often used as outriggers in addition to coincident pair technqiues as main pickup, in order to provide a sufficient spatial impression as well as a more detailed imaging of the direct sounds located at the extremes of the stage [Eargle 2001].

1.2.4 Near-coincident pair microphone technique

Near-coincident pair microphone techniques employ a pair of uni-directional microphones that are spaced closely and angled outward, thus having forms of both coincident and spaced pair techniques. Designs of these techniques rely on a combination of ICTD and ICID that can be traded-off for certain SRAs, although it would depend on which trading relationship is believed, and therefore it is possible to vary the distance and angle between microphones in various ways depending on the attributes of phantom images that are desired by recording engineers. As stated by

Rumsey [2001], near-coincident microphone techniques have their advantages in the compromise between an accurate phantom image localisation and a good sense of spatial impression since these attributes cannot be always conveyed simultaneously by a coincident or spaced pair technique alone. For instance, if the microphone distance was increased against the angle for a certain SRA, the resulting image would benefit more from a good sense of spatial impression rather than accurate imaging. On the other hand, if the angle between the microphones played a more important role than the distance in deciding the SRA, the image would be accurately localised but would not necessarily be spacious.

A number of near-coincident arrays with fixed distances and angles have been used for many years. Arguably the near-coincident pair technique that has been most widely used to date is the 'ORTF' (the Office de Radiodiffusion – Television Francaise) technique. As can be seen in **Figure 1.10**, the two cardioid microphones are spaced 17cm apart with the lateral angle of 110°. The SRA based on Wittek [2001a]'s image assistant model is about 102° whereas that based on the Williams curves [1987] is 95°. In this technique, signals from the two microphones are virtually phase coherent at low frequencies while minimal phase difference is produced only at the highest frequencies [Streicher and Everest 1998]. Therefore, the low frequency comb-filter effects, which tend to be caused from pure spaced techniques, are avoided and an 'open and airy' sound is produced [Streicher and Everest 1998]. Another popular example of near-coincident techniques is the 'NOS' (Nederlande Omroep Stichting) technique, which uses cardioid microphones with the distance of 30cm and the lateral angle of 90°. The SRA for the NOS is 82°,

according to Wittek [2001a]'s model.

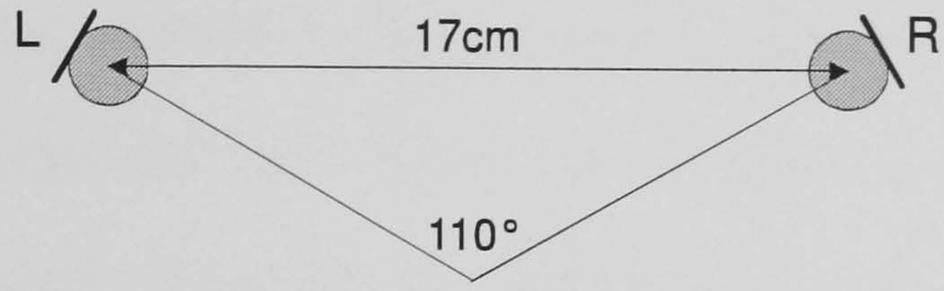


Figure 1.10 Configuration of 'ORTF' near-coincident array

1.3 Phantom Imaging Principles for 3-2 Stereophonic Reproduction

Phantom imaging principles for 3-2 stereophonic reproduction are mainly based on the summing localisation theory of 2-0 stereophonic reproduction, which was discussed earlier. However, the psychoacoustics involved in the 3-2 stereophonic reproduction are more complicated than those in the 2-0 one since the former deals with multiple sound sources generated from five different directions around the listener (see **Figure 0.1**). Therefore, it can be expected that each of the front, side and rear reproduction segments has unique imaging characteristics. The imaging characteristics of 3-2 stereophonic reproduction are an important basis for the designs and applications of 3-2 stereophonic microphone techniques.

1.3.1 Front image localisation

Due to the addition of a centre loudspeaker in 3-2 stereophonic reproduction, the frontal listening area is divided into two stereophonic segments (i.e. the left (L) – centre (C) loudspeaker pair and the right (R) – centre (C) loudspeaker pair). Similarly to the operation of summing localisation in 2-0 stereophonic reproduction, if coherent signals are fed into the two loudspeakers in one segment, the localisation of a phantom source within the segment can be controlled by the relationship between ICTD and ICID. (This perfect separation of the two segments in terms of phantom imaging based on the individual interchannel relationship is virtually impossible in the reproduction of the signals recorded using multichannel microphone techniques due to the problem of interchannel crosstalk.) When there is no ICTD or ICID, the phantom image will be localised at ±15° from the centre loudspeaker. Therefore, in 3-2 stereo the range of the maximum phantom image shift becomes 15°. Theile [2001] suggests that the degree of phantom image shift for a certain ICTD or ICID is decreased linearly with decreasing loudspeaker angle. For example, the ICTD and ICID required for the phantom image shifts of 10°, 20° and 30° in 2-0 stereo, which were shown in **Table 1.1**, will be used for the shifts of 5°, 10° and 15° respectively in 3-2 stereo. This also suggests that the same SRAs of 2-0 stereophonic microphone arrays will still be effective even when the output signals of the arrays are reproduced from the L – C or R – C pair. Theile's hypothesis seems to be confirmed to some extent by the results of Martin *et al* [1999]'s experiments that were conducted to investigate the localisation behaviour in 3-2 stereophonic reproduction in an anechoic chamber using a speech signal, although determination of exact ICTD and ICID values for

localisation were not the main interest of this experiment. The experiment was designed for the subject to point to the locations of perceived phantom images that were created with varying ICTDs or ICIDs for a randomly chosen pair of adjacent loudspeakers (L – C or R – C). The ICTD was varied between 0ms and 2ms in 0.2ms intervals and the ICID was varied between 0dB and 16dB in 2dB intervals. The resulting localisation plots are shown in **Figure 1.11**. It can be seen that in cases where the centre channel is attenuated or delayed relative to the left or right channel (i.e. the phantom images are expected to be localised between 15° and 30°), the ICIDs and ICTDs required for the phantom image shifts of 5°, 10° and 15° correspond roughly to Wittek [2000]'s or the author's data obtained for the image shifts of 10°, 20° and 30° in the conventional 2-0 stereophonic reproduction (see **Table 1.1**). However, this constant relationship does not seem to appear as obvious when the left or right channel is attenuated or delayed relative to the centre channel, especially in the case of ICTD. This might be explained by the following hypothesis. The centre loudspeaker is placed on the median plane, and therefore a signal radiated from it will cause no interaural time and intensity differences while that from the left or right loudspeaker will naturally cause certain interaural differences for the direction of the loudspeaker. Therefore, when the centre channel is not attenuated or delayed relative to the left or right channel at all, the position of centre loudspeaker itself might become a confusing factor for creating interaural differences that are suitable for the operation of the summing localisation.

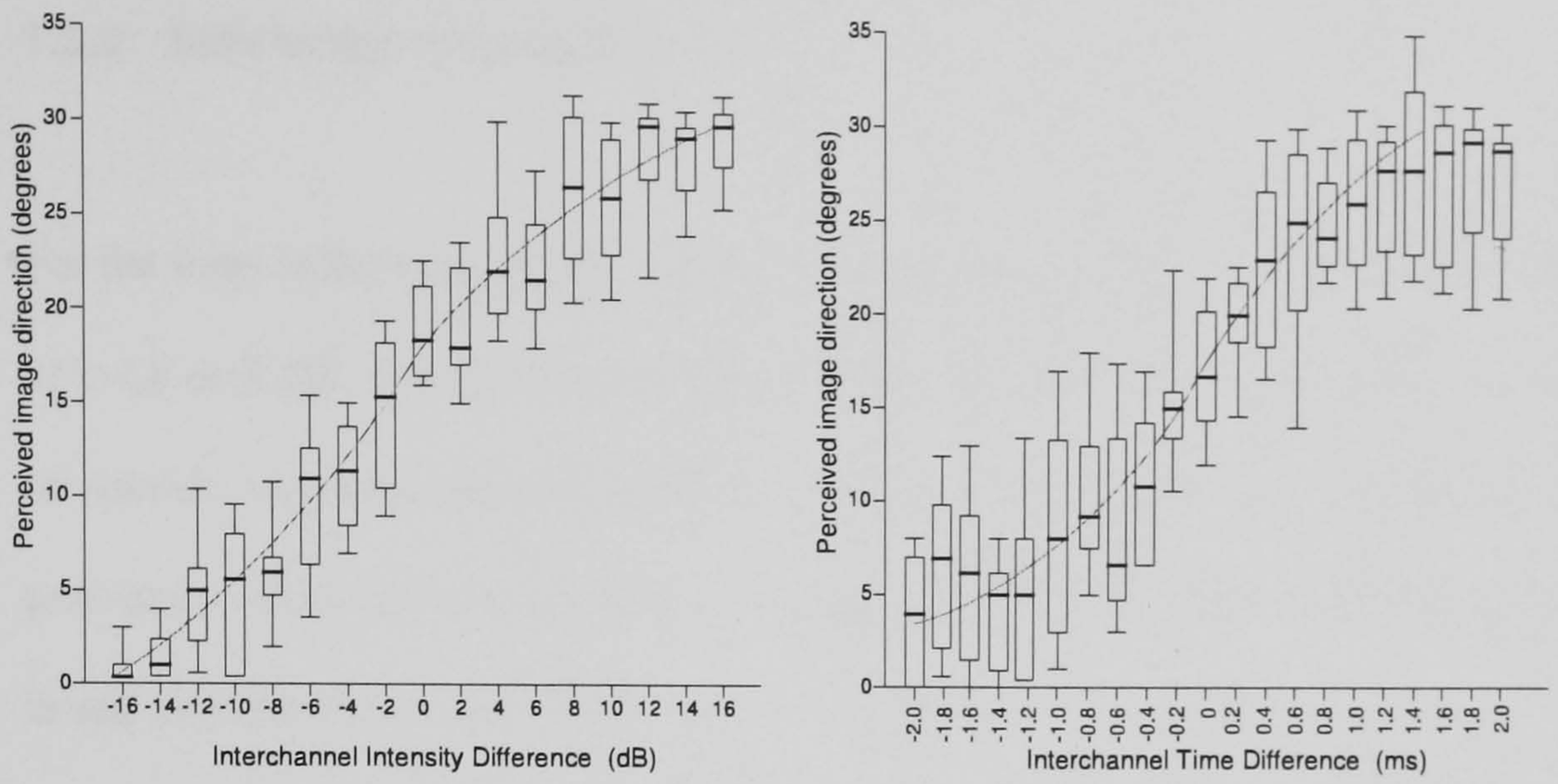


Figure 1.11 Comparison between localisation characteristics for the front images created using ICID and those created using ICTD, obtained from a subjective listening test using a speech source [after Martin *et al* 1999]

It can also be seen from the above plots that the ranges of localisation errors for the images created using ICTD were generally greater than those created using ICID, and this seems to support the dominance of coincident pair microphone techniques relying on the ICID cue over spaced pair techniques relying on the ICTD cue with respect to the accuracy or certainty of phantom image localisation. It was also pointed out by Martin *et al* [1999] that the comb-filter effect that would result from three-channel spaced pair microphone arrays would be more obvious than that resulting from a two-channel spaced pair array, since the relative lack of head shadowing of the centre channel to the ears would increase the effect of interference between the signals radiated from adjacent loudspeakers.

1.3.2 Side image localisation

For the front facing listener only one ear is toward each of the side loudspeaker pairs of L-LS or R-RS, and therefore in the localisation of side images there will be a lack of suitable interaural differences that are required for the summing localisation or precedence effect in the usual manner. In fact, the difficulty in achieving stable side image localisation has been proven in several studies.

The result of Ratliffe [1974, cited in Theile and Plenge 1977]'s experiment carried out with a quadraphonic reproduction system showed that even small intensity differences between the front-left and rear-left loudspeakers could cause large angular shifts, and that the phantom sources tended to jump randomly between the front and rear. Theile and Plenge [1977] conducted a similar experiment to Ratliffe's. They used a pair of loudspeakers splayed at a fixed angle of 60° and the centre of the pair was varied laterally anti-clockwise. It was found that as the loudspeaker pair was moved closer to the side of the listener, the localisation curve became steeper and the degree of uncertainty in localisation increased, as can be seen in **Figure 1.12**. These limitations of side image localisation were also confirmed in Martin *et al* [1999]'s experiments, which were described in the previous section. It was reported that the certainty of phantom image localisation in the side area of the standard 3-2 stereophonic loudspeaker arrangement (30°−120°) was the worst in among all the sub-listening areas. It was further found that the degree of uncertainty in the side image localisation was greater with ICTD panning than with ICID. Martin *et al* note that side phantom images created using ICTDs suffer from noticeable comb-filter effects

since there is little intensity difference between the two signals arriving at each of the listener's ears.

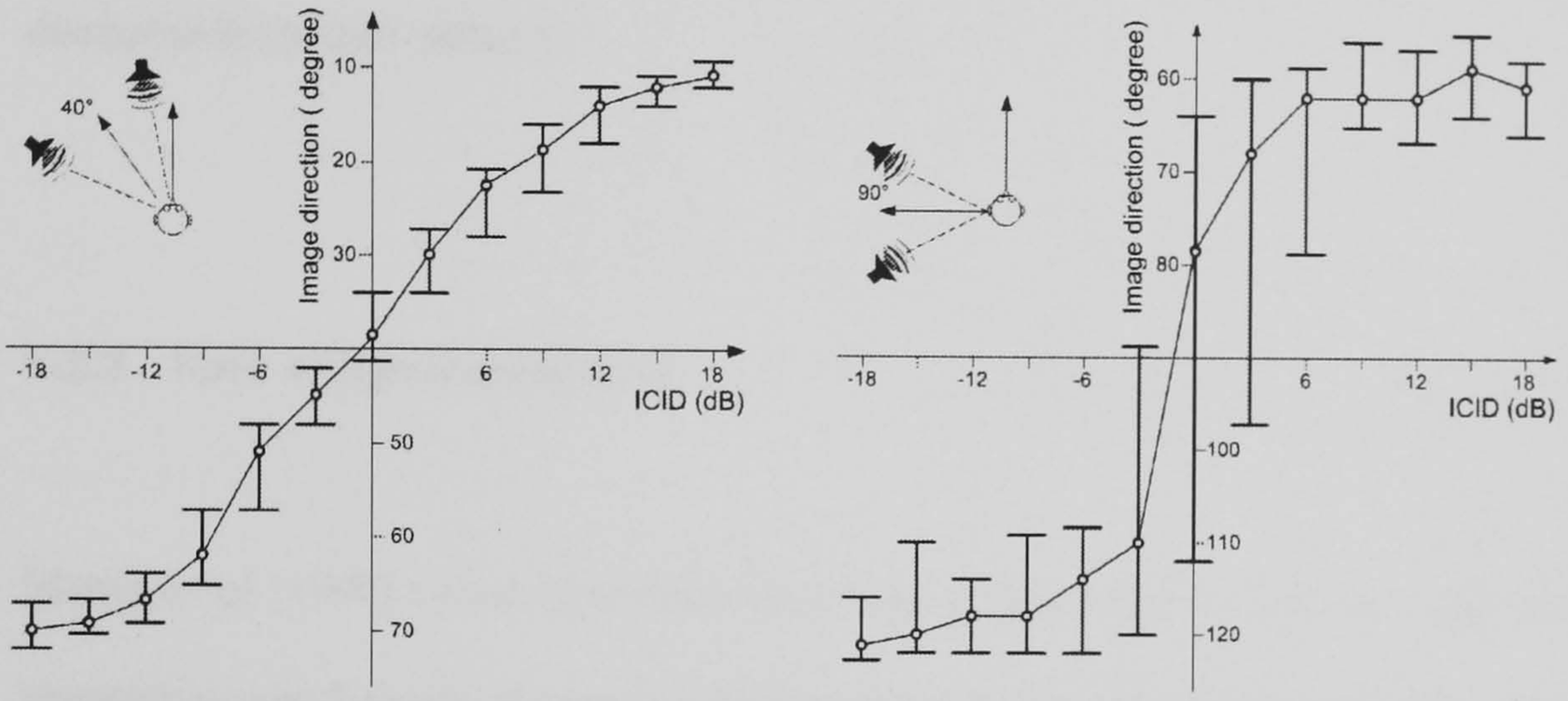


Figure 1.12 Comparison of the localisation characteristics of the phantom images created from loudspeaker pairs having different lateral displacements of stereo-base centre [after Theile and Plenge 1977]

The above findings might lead to a discussion about what kind of sound source the side listening area should be used for. It should be basically dependent on how much the localisation accuracy is required for a certain type of sound source. For example, ambient sounds created by reflections or reverberation would not be required to be accurately localised; rather they might benefit from less precise imaging, which is normally produced by decorrelated low frequency signals [Griesinger 1996]. Therefore, these kinds of sounds would be suitable to be reproduced from the side pair of loudspeakers. However, many recording engineers particularly for classical music might favour stable localisation of phantom images for direct sound sources and therefore such side imaging characteristics as the dramatic angular shift of phantom

image or low degree of localisation certainty might make side pairs of loudspeakers unsuitable for the reproduction of direct sounds. This issue is also related to the design and application of 3-2 stereophonic microphone techniques, which will be discussed in the later sections.

1.3.3 Rear image localisation

Martin *et al* [1999] found from their experiments (described in Section 1.3.1) that phantom images between the rear pair of loudspeakers were localised more stably than those for the front pairs. This is due to the fact that the rear pair of loudspeakers is symmetrical across the median plane while the other pairs of loudspeakers are not [Martin *et al* 1999]. This suggests that it would be acceptable to use the rear region for reproducing the phantom images of direct sounds, although it is usually used for reproducing ambient sound images. Another interesting result obtained from their experiments is that ICTD of only about 0.6ms was required for a phantom source to appear at fully one loudspeaker in the rear region. This value is approximately a half of the ICTD required for the same effect in the conventional 2-0 stereo. This seems to be due to the wider angle subtended by the rear loudspeakers. However, the wide angle between the rear loudspeakers might lead to the 'hole in the middle' effect, as the images tend to pull to the loudspeakers rapidly [Rumsey 2001].

1.4 3-2 Stereophonic Microphone Techniques

In recent years, a number of novel 3-2 stereophonic microphone techniques have been proposed for surround sound recording and reproduction. Their design and operational principles are based on the principles of 2-0 stereophonic microphone techniques. However, 3-2 techniques are still being evaluated and developed as the psychoacoustics involved in surround sound have not been fully investigated yet, for example the effect of interchannel crosstalk. This section reviews the design concepts and operating characteristics of various 3-2 stereophonic microphone techniques.

1.4.1 Design concepts

Rumsey [2001] suggests a way of classifying the design concepts of current microphone techniques intended for 3-2 stereophonic reproduction, based upon the purpose of the rear channels. According to his classification, there are two main groups: those that use 'five-channel main microphone techniques' and those that use 'techniques with front and rear separation'. Five-channel main microphone techniques consist of five microphones that are placed relatively close to one another, forming a single array (normally a front triplet with two microphones further back). Each microphone signal is routed to one of the loudspeakers in 3/2 stereo reproduction: Left (L), Centre (C), Right (R), Left Surround (LS) and Right Surround (RS). Such microphone techniques attempt to provide both satisfying spatial

impression and continuous phantom imaging around the 360° in the horizontal plane simultaneously with a fixed pattern of microphone placement. However, due to the limitation of balanced phantom imaging in the side listening area, which was discussed in Section 1.3.2, linear 360° imaging seems difficult to realise. Furthermore, Theile [2001] points out that the creation of natural images requires much effort because of the complicated relationship between the psychoacoustic parameters involved. For example, accurate localisation will rely on the summing localisation and precedence effect across the various two-channel stereo segments (for example, between L & C, or R & RS in the 3/2 stereo configuration) due to the short distances between the microphones. The listening position and front-rear balance will therefore affect the performance of the technique [Rumsey 2001]. Furthermore, the fixed positions and polar patterns of the front and rear microphones would result in an inevitable compromise between the representation of optimised directional images and spatial impression. For example, the front triplet should be optimised not only with respect to the recording angle of direct sound from the front but also with respect to the balance of direct and indirect sound intensity in conjunction with the rear microphones [Theile 2001]. In addition, the position and directivity of the rear microphone array should not be decided exclusively for the characteristics of the ambient sound, but also for the suppression of the direct sound due to the relatively short distance between the front and rear microphones.

'Techniques with front and rear separation', on the other hand, use a 'frontal' main microphone array that is used primarily to image the direct sound from the front, together with a separate 'rear' microphone array that is intended to pick-up

decorrelated ambient sound to supply (primarily) the rear loudspeakers. Usually the frontal microphone array is a variation of a conventional stereo technique or the front triplet of a five-channel main microphone technique. Different rear microphone arrays can be combined with different front arrays depending on desired directional and ambience characteristics [Theile 2001]. The distance between the front and the rear arrays can vary depending on different recording situations. The further the rear array is from the recorded sources, the more early reflections, the higher the reverberant-to-direct ratio and the higher the density of reflections. However, according to Theile [2001], at least 10dB suppression of the direct sound is required in the rear channels versus the front channels. It is considered that 'techniques with front and rear separation' afford recording engineers more freedom to choose 'front' and 'rear' microphone techniques depending on the desired characteristics of frontal image and spatial impression than fixed five-channel main microphone arrays. Moreover, they would enable the engineer to subjectively balance the direct and ambient sounds using artistic and technical judgment. In this respect, microphone techniques with front and rear separation appear to be more practical in a wider range of recording applications. However, both groups in common tend to prefer a narrowly or widely spaced microphone configuration to a coincident one since the coincident technique does not provide a satisfying natural spatial impression due to the lack of decorrelated low frequency phase difference [Griesinger 1997].

1.4.2 Frontal main microphone techniques

The 'Decca tree' technique shown in **Figure 1.13** has been one of the most popular two-channel main microphone techniques. However, this technique also can be adopted for three-channel purposes due to the number of microphones used. It employs three widely spaced omni- directional microphones, thus relying on the precedence effect. The spaced pair of L and R produces sufficient time difference information and therefore provides a good sense of 'openness' [Theile 2001]. The centre microphone provides 'articulation' to the phantom image [Streicher and Everest 1998] and prevents the hole in the middle, which would be likely to occur with the spaced pair itself.

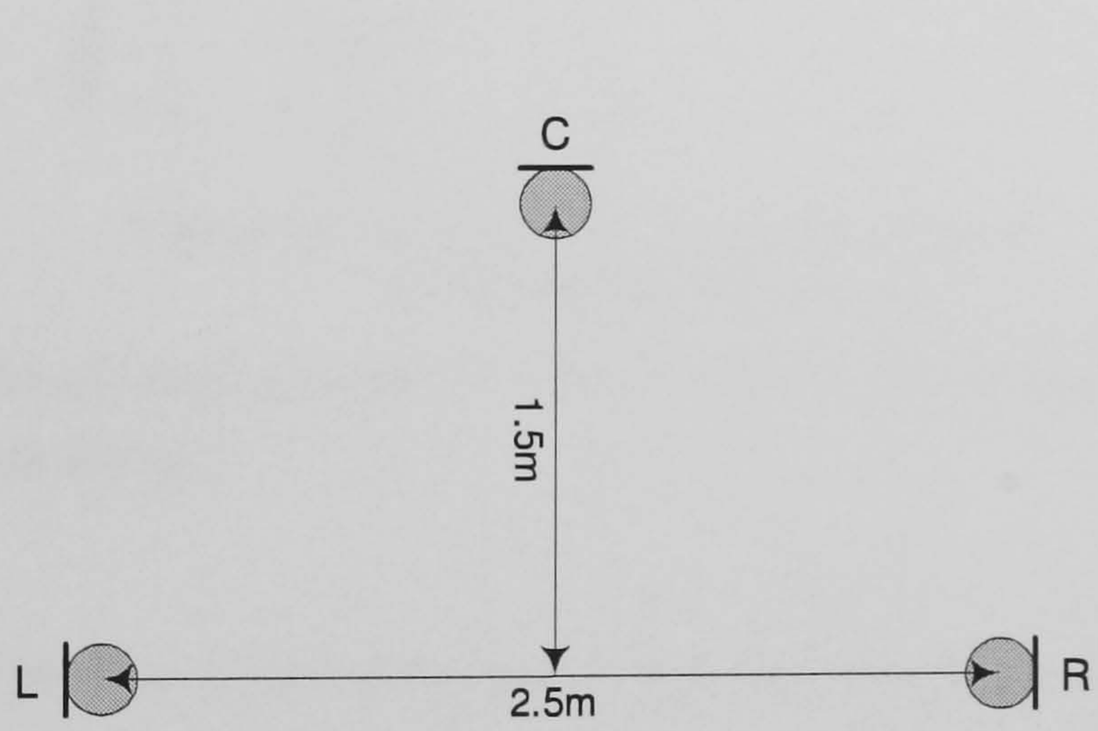


Figure 1.13 'Decca tree' configuration with three spaced omni microphones

However, the addition of the centre microphone without intensity reduction causes an imbalance in phantom image distribution. As can be seen in **Figure 1.14**, due to the large spacing between the L – R pair and C, sound sources located at up to ±45° are reproduced in the centre loudspeaker. Beyond this angle the phantom image rapidly

shifts toward the left or right loudspeaker. This means that the Decca tree essentially has three solid localisation areas owing to the strong precedence effect. Fukada *et al* [1997] suggests that when this technique is used for surround recording, cardioid microphones should be used instead of omnis because the latter could pick up too much ambient sound, thus causing exaggerated spatial impression when surround channels are added.

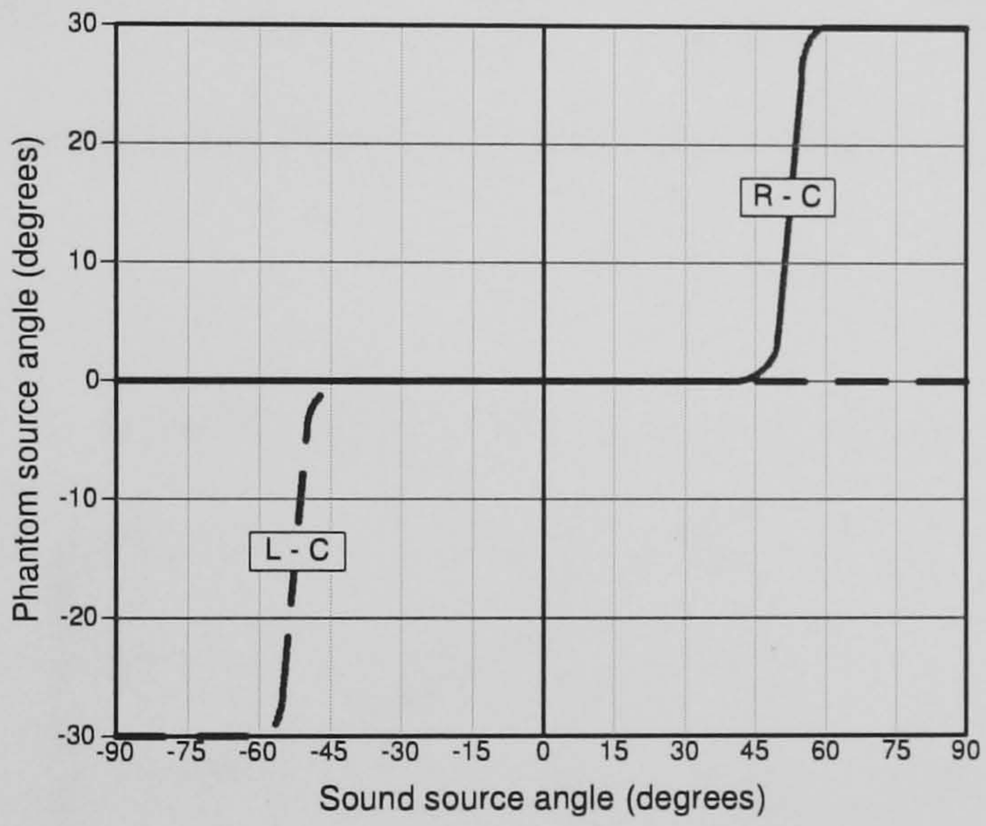


Figure 1.14 Localisation curve for the Decca Tree array, calculated using the Image Assistant [Wittek 2001a]

Klepko [1997] proposed a three-channel near-coincident technique (see **Figure 1.15**), which consists of three microphones placed in line with a distance of 17.5cm between each microphone. In order to avoid a strong centre phantom image, the outer channel employs a super-cardioid microphone, which has increased directivity, while the centre channel uses a cardioid microphone. However, despite the use of super-cardioids, this technique suffers from a high degree of interchannel crosstalk, in that the centre and left or right channels produce an intensity difference of only 1–8dB and

a time difference of less than 0.5ms [Theile 2001]. Therefore, a huge overlap between the recording area L-C and R-C is inevitable as can be seen in **Figure 1.16**. The stereophonic recording angle (SRA) of this array is very wide (180°) due to the small lateral angle and this may result in a narrow stereophonic image with a usual microphone distance from the stage.

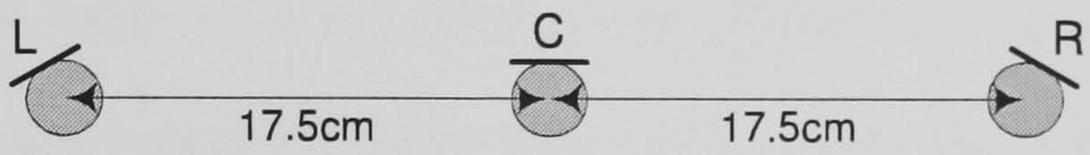


Figure 1.15 Near-coincident triplet with cardioid microphones, proposed by Klepko [1997]

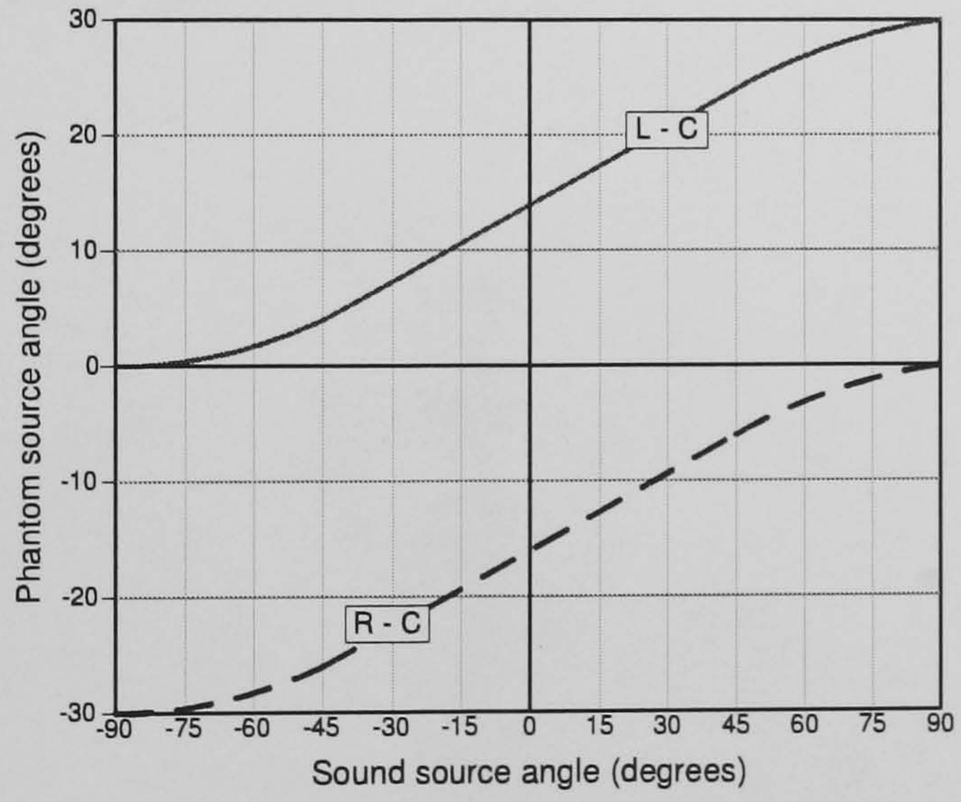


Figure 1.16 Localisation curve for Klepko [1997]'s three-channel near-coincident array, calculated using the Image Assistant [Wittek 2001a]

Williams and Le Du [1999, 2000] proposed a microphone technique aiming to achieve balanced distribution of phantom images. This technique is based on the design method they called 'critical linking' which attempts to link the SRAs for the two microphone pairs of L – C and C – R without overlap as can be seen in **Figure 1.17**,

and the combination of distance and angle between microphones for achieving certain SRAs depends on the 'Williams curves', which were introduced in Section 1.2.1. The critical linking is achieved by using either 'electronic offset' or 'microphone position offset'. The electronic offset is achieved by varying the value of ICID or ICTD while the microphone position offset is achieved by changing the physical position of the microphones with respect to the time and intensity trading function. A benefit from using the critical linking technique is that it enables recording engineers to create microphone arrays with various distances and angles sharing the same SRA depending on the characteristics of recorded sound desired. Since the SRA is based on the time and intensity trading function, a more spaced microphone array will have a smaller angle between microphones.

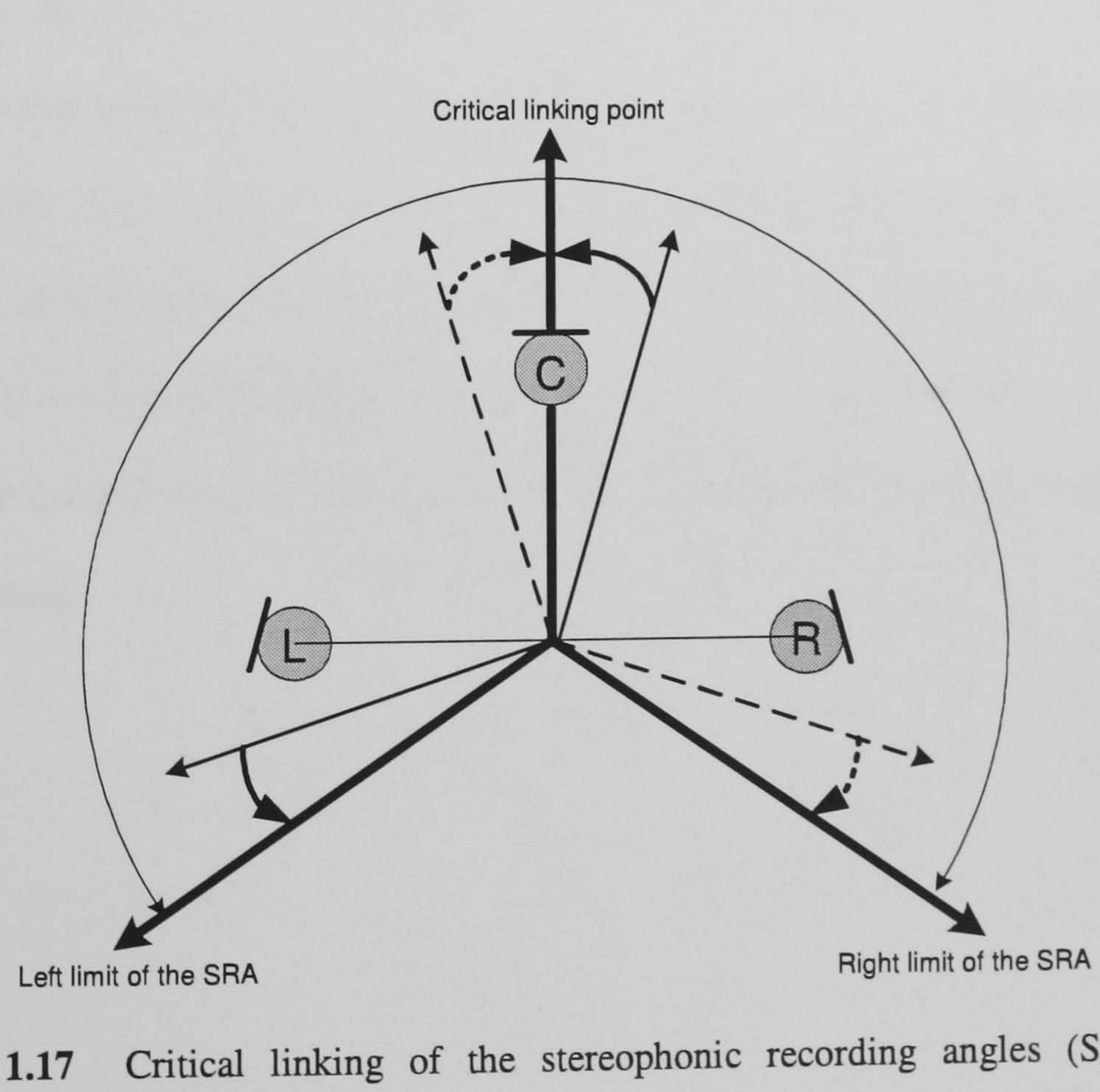


Figure 1.17 Critical linking of the stereophonic recording angles (SRAs) of microphone pair L – C and C – R [Willams and Le Du 1999, 2000]

Herrmann and Henkels [1998] developed a three-channel microphone technique they call 'ICA-3' (Ideal Cardioid Array) technique on the basis of the critical linking approach. The distance and angle between the microphones are based on the Williams curves and they can be varied depending on the SRA desired, as shown in **Table 1.2**. However, the angle between the outer microphones should always match the SRA. Therefore, in order to obtain the full spread of phantom images, the array can simply be placed so that the outer microphones face the edges of the recording stage and this feature might be convenient when recording engineers choose the SRAs suitable for particular microphone array placements. **Figure 1.18** shows the localisation curve of an ICA-3 array with the SRA of 120°. According to this curve, critical linking appears to be achieved successfully in this technique. However, Theile [2001] claims that the phantom imaging for this array is compromised by a considerable amount of interchannel crosstalk due to the lack of sufficient channel separation. For example, according to the calculation using the Image Assistant [Wittek 2001a], when the sound source is located at the front of the array with 5m distance, the intensity difference between L and C is only 3dB greater than that between L and R, which means that the impact of localisation by the L-R pair cannot be neglected.

Stereophonic Recording Angle	**Horizontal Distance between L and R**	**Vertical distance between L – R and C**
100°	126cm	29cm
120°	92cm	27cm
140°	68cm	24cm
160°	49cm	21cm
180°	35cm	17.5cm

Table 1.2 Distances and angles for the microphones of the ICA-3 array, required for certain stereophonic recording angles (SRAs); the angle between left and right microphones match the SRA [Herrmann and Henkels 1998]

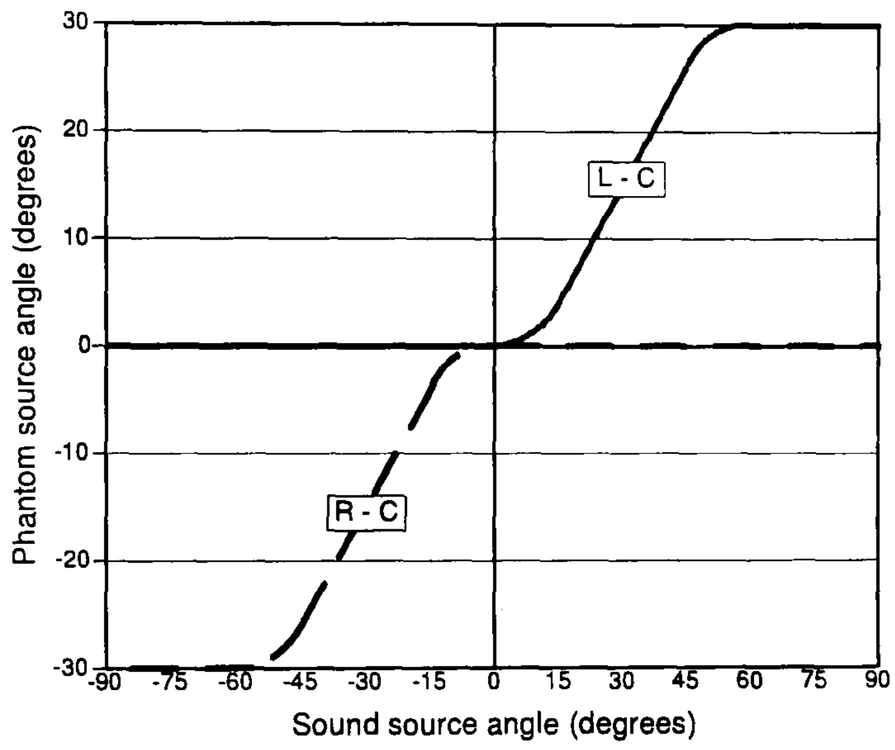


Figure 1.18 Localisation curve for Herrmann and Henkels [1998]'s ICA-3 array with the SRA of 120°, calculated using the Image Assistant [Wittek 2001]

Having firstly raised the issue of interchannel crosstalk, Theile [2001] proposed a three-channel microphone technique called 'OCT' (Optimal Cardioid Triangle). This technique attempts to reduce the amount of interchannel crosstalk as much as possible, particularly in the associated intensity of the stereophonic pair L-R, so that only the pairs of L-C and R-C become effective in localisation. In order to achieve this aim, the OCT configuration, shown in **Figure 1.19**, employs a cardioid microphone for the centre microphone and super-cardioid microphones for the outer microphones. The

outer microphones are oriented towards the sides in order to obtain maximum channel separation and owing to this feature the associated intensity of the unwanted phantom sources L-R is about 10dB lower compared to that of the wanted phantom sources L-C or R-C. This appears to be a clear improvement against the ICA-3. The spacing between L and C can be adjusted depending on the recording angle. The relationship between the recording angle and distance *d* calculated using the Image Assistant [Wittek 2001a] is shown in **Figure 1.20**.

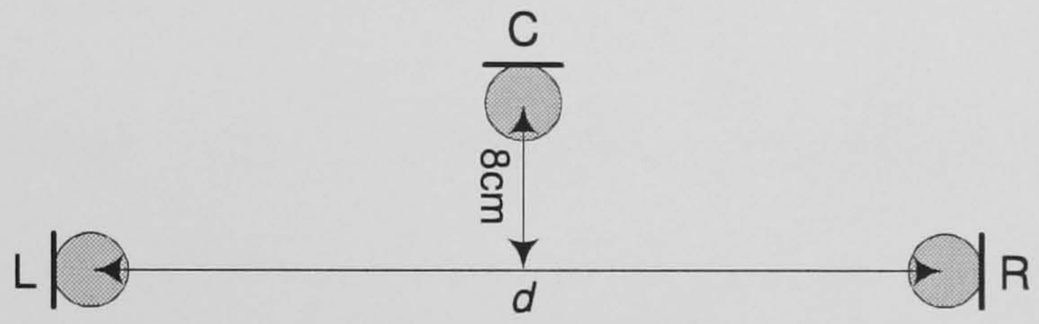


Figure 1.19 'OCT' frontal microphone array using super-cardioid microphones for L and R and cardioid microphone for C, proposed by Theile [2001]; spacing between L and R is adjustable depending on the stereophonic recording angle.

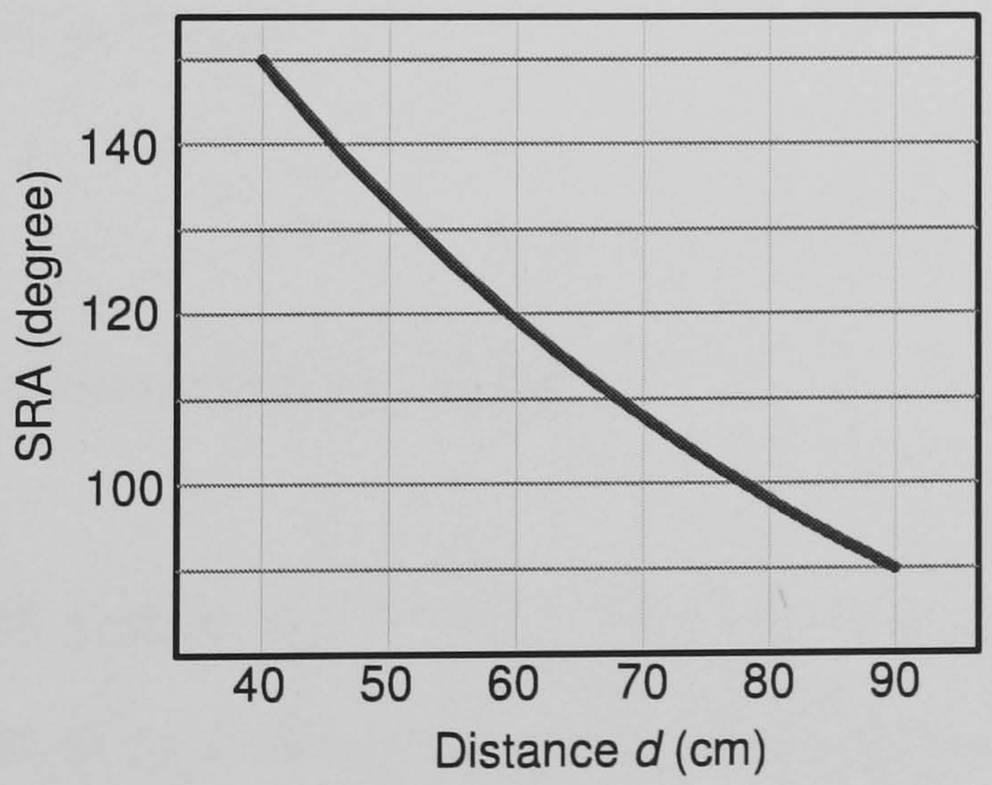


Figure 1.20 Stereophonic recording angle (SRA) of the OCT array for various distances between left and right microphones, calculated using the Image Assistant [Wittek 2001a]

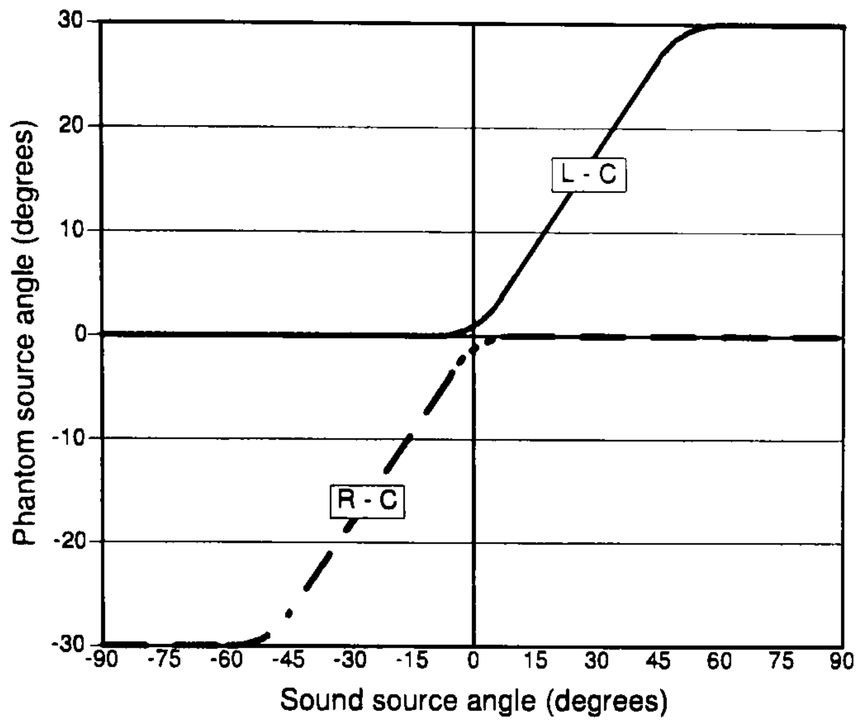


Figure 1.21 Localisation curve for Theile[2001]'s OCT array with the SRA of 118°, calculated using the Image Assistant [Wittek 2001a]

Despite the crosstalk optimisation, however, it seems that the OCT is limited in providing a linear directional transition of phantom sources across L-C-R. As can be seen in **Figure 1.21**, there is an obvious overlap between the localisation curves for L-C and R-C around the centre region and this might be comparable with the linear transition for the ICA-3 shown in **Figure 1.18**. The reason for this nonlinearity is not explained in Theile's paper.

1.4.3 Rear microphone techniques

Theile [2001] suggests that in order to create a realistic image of enveloping atmosphere in sound recording and reproduction, a rear microphone array should employ four channels, with each pair covering each side of the recording space. This can be supported by Hiyama *et al* [2002]'s finding. They compared a number of different loudspeakers arrangements using band-passed noise signals in order to

investigate the number of loudspeakers required for the reproduction of realistic diffused sound field. The reference arrangement was 24 loudspeakers placed at every 15° making a circle and the number of loudspeaker used for the reproduction was reduced from 24 to 12, 8, 6, 5, 4, 3 and 2. The spatial impression created from each arrangement was compared with the reference. It was found that at least six loudspeakers were required to obtain spatial impression similar to that created from the reference. However, it was also found that almost the same spatial impression could be perceived with only four loudspeakers when they were arranged at the positions similar to those of the left, right, left surround and right surround loudspeakers in the standard 3-2 arrangement.

Theile [2001] proposed a four-channel rear microphone technique called 'IRT-Cross'. As can be seen in **Figure 1.22**, this technique employs four cardioid microphones arranged in a square. Due to the front-side facing cardioid microphones in the array, interchannel crosstalk from direct sounds can become considerable unless this array is placed far enough away from the front array. The spacing between the microphones can be decided depending on the characteristics of spatial impression desired, although the range of 20-25cm is recommended by the author [Theile 2001]. For example, a closer spacing will provide a more balanced distribution of enveloping sources, while a wider spacing will provide a more diffused reverberation. However, extreme spacing of either too close or too wide will cause a 'loss of envelopment' [Theile 2001].

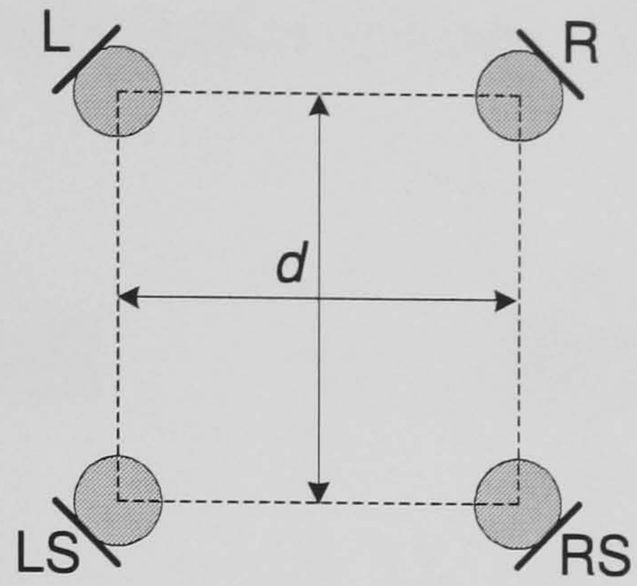


Figure 1.22 'IRT-Cross' configuration [Theile 2001]; the distance d is in the range of 20cm and 25cm.

The 'Hamasaki-Square' [Hamasaki *et al* 2000], shown in **Figure 1.23**, is another example of four-channel technique for ambience pick-up. It employs four figure-8 microphones with the side of each microphone facing the front in order to reduce the amount of interchannel crosstalk from direct sounds as much as possible. It is suggested that the microphones LS and RS are routed to loudspeakers LS and RS while the microphones L and R are routed to loudspeakers L and R or panned between L-LS and R-RS depending on the amount of desired spatial information in the front loudspeakers. The distance between each microphone that was originally suggested by the authors was 1m, but later Hamasaki and Hiyama [2003] suggested the distance of 2-3m from subjective investigations. They measured interaural cross-correlation coefficients (IACC) for the signals recorded using two omni-directional microphones with various spacings in a reverberant sound field and reported that low frequency decorrelation required for generating the most satisfying spatial impression was achieved at the distance in the range of 2-3m. This array is guided to be placed far beyond the critical distance, where the intensities of direct and reverberant sounds become the same, and at a high position in the recording space in order to obtain the

maximum R/D ratio (intensity of reverberation relative to direct sound) [Hamasaki *et al* 2000].

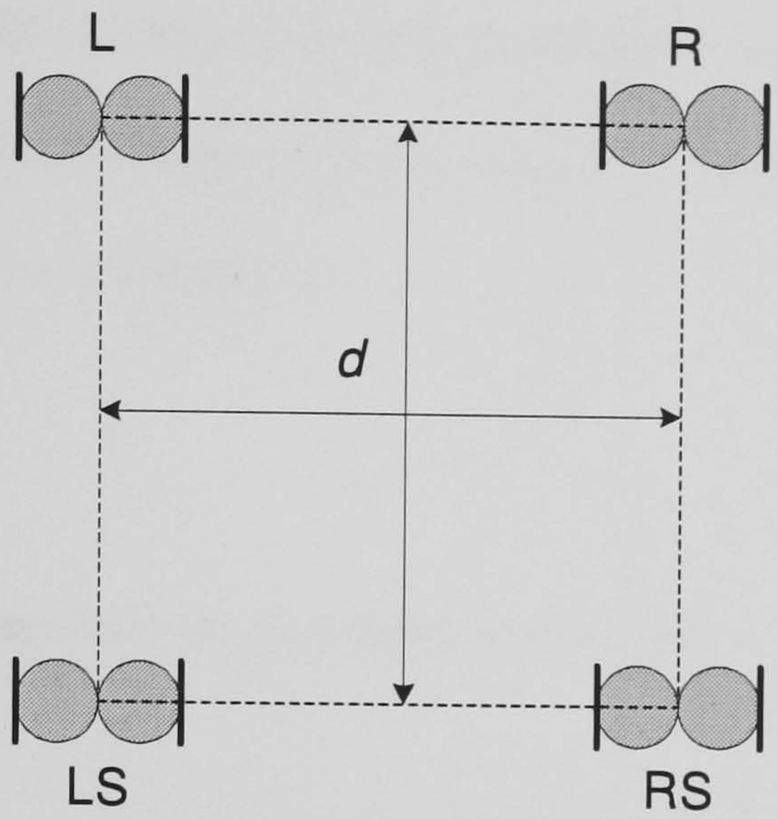


Figure 1.23 'Hamasaki-Square' configuration [Hamasaki *et al* 2000]; the distance *d* is in the range of 2-3m.

Klepko [1997] proposed using a dummy-head binaural microphone in order to provide a 'continuous' lateral spatial impression. He affirms that the limitation of the loudspeaker reproduction of binaural signals caused due to acoustic crosstalk in the conventional two-channel reproduction can be naturally overcome when the signals are reproduced through the rear loudspeakers LS and RS. This is based on the fact that the rear loudspeakers are placed almost at the sides of the listener. In such case the listener's head will act as a diffracting barrier to high frequencies above 1kHz, which carry the most effective HRTF cues. Klepko reported that 'continuous and clear' spatial images were perceived between ± 30° and ± 90° from the listening test using the dummy head microphone coupled with the near-coincident front triplet introduced in the previous section. The distance between the front triplet and the dummy head used for his experiment was 124cm. However, with this distance the

interchannel crosstalk from the direct sound will have almost the same intensity and short delay time (about 0.38ms). This might become a critical problem with regard to achieving accurate localisation of the front image and it might be more reasonable to place the dummy head microphone further back from the front array and let it face the back in order to increase the R/D ratio.

1.4.4 Five-channel main microphone techniques

The 'critical linking' technique [Williams and Le Du 1999, 2000], which was introduced in Section 1.4.2, can be applied for the design of a five-channel main microphone array. The SRA for each of the five stereophonic recording segments is linked without any overlap in order to enable phantom images around the full 360°. Similarly to the three-channel critical linking techniques, calculation of the SRA for each stereophonic segment is based on the Williams curves (see Section 1.2.1). In the design process the SRA for the front triplet is decided first depending on the distance of the microphone array from the recording stage and then the SRA for the rear pair LS-RS is determined as desired. Finally the distance between the front triplet and the rear pair is decided depending on the necessary SRA for the side segments and the 'critical linking' between the front and rear segments is achieved using suitable electronic time or intensity offset. **Figure 1.24** shows an example of the critical linking five-channel array. In this particular example, the SRA for the front triplet is 120° and that for both side and rear pair is 80°. Williams [2003] states that an advantage of this technique is the flexibility for design since the SRA of each

segment can be decided flexibly depending on the type of recording sources, e.g. large orchestra requiring wider front SRA and small ensemble requiring narrower SRA. However, as Rumsey [2001] points out, it is doubtful if the Williams curves that were derived from a front two-channel based experiment can also be applied correctly for the calculation of SRA of the side or rear pairs of microphones. In fact, as discussed in Section 1.3, the localisation of phantom image for the side and rear listening areas has different characteristics to that for the front.

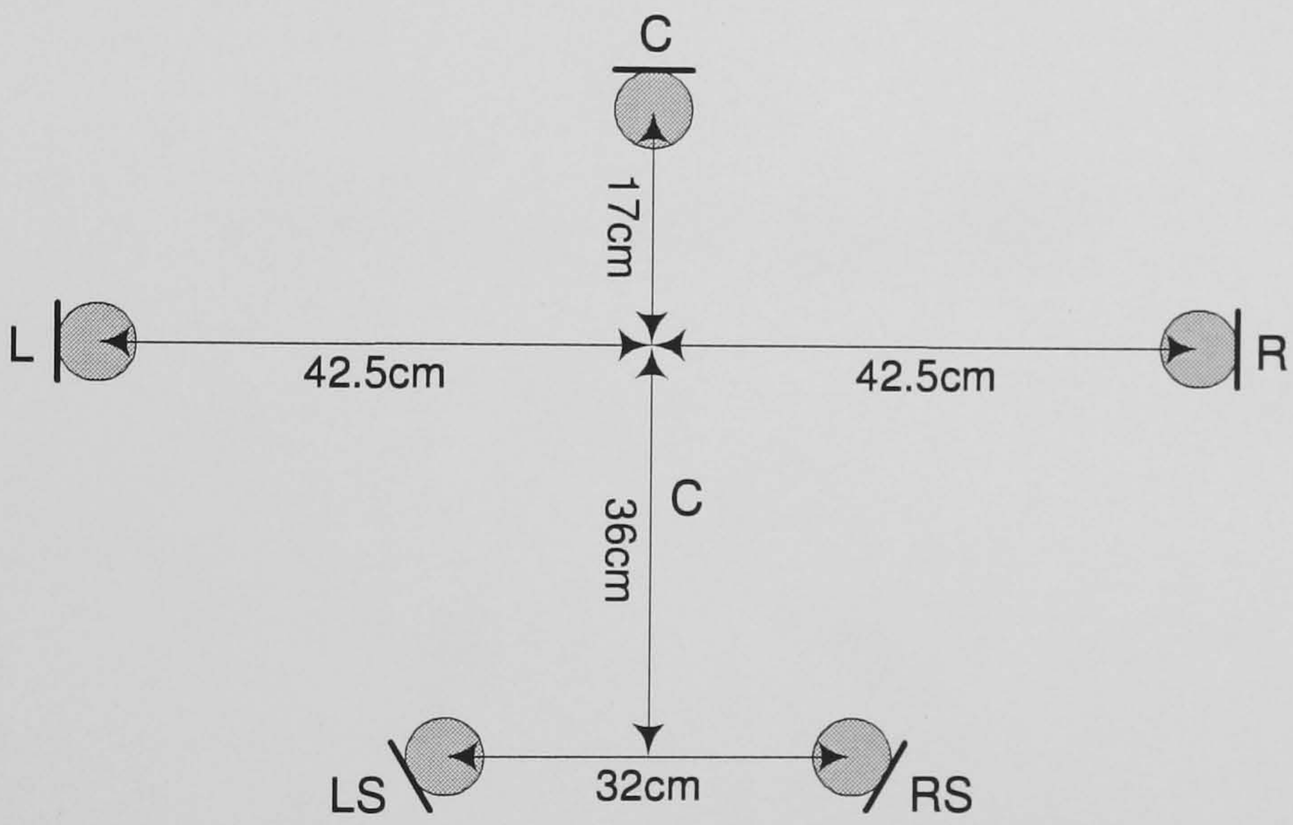


Figure 1.24 'Critical linking' five-channel microphone array [Williams 2003]

The 'ICA-5' technique developed by Herrmann and Henkels [1998] consists of the ICA-3 front triplet (see Section 1.4.2) and two rear cardioid microphones, as shown in **Figure 1.25**. This configuration also is designed to achieve the SRA of 360° using the critical linking technique. Calculation of the SRA for each stereophonic segment is again based on the Williams curves. The proposed SRA for the front triplet is 180° and that for the side or rear pair is 60°. This is based on the authors' subjective judgment on the balanced phantom image distribution in the front listening area. However, even though the wide front SRA is the correct choice in terms of the

attempted 360° imaging, it would be likely to cause the frontal stereophonic images to become too narrow if the array was placed at usual distances from the recording stage. Therefore, in order to increase the width of the frontal stereophonic image, the array could be placed very close to the stage, but in this case the rear microphones would suffer from a high degree of interchannel crosstalk from the direct sound in the front.

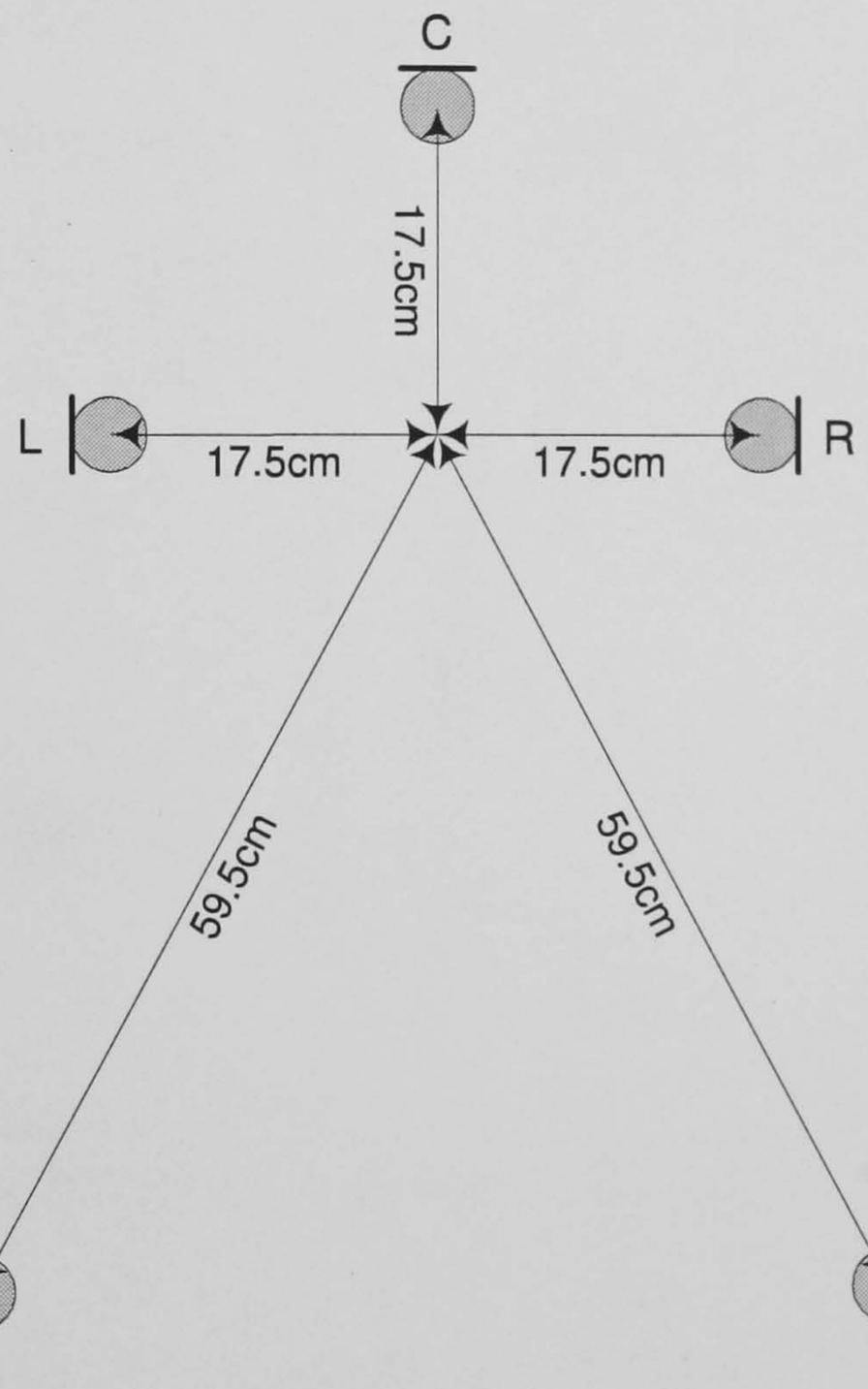


Figure 1.25 'ICA-5' five-channel microphone array [Herrmann and Henkels 1998]

The 'OCT-Surround' technique [Theile 2001], shown in **Figure 1.26**, was adapted from the 'OCT' three-channel technique that was introduced in the earlier section. As can be seen in the figure, two additional cardioid microphones are added to the OCT front triplet for rear pick-up. This technique is optimised in order to obtain a

natural intensity balance between direct and indirect sounds without affecting the frontal image localisation. The rear microphones face backward in order to obtain the maximum suppression of interchannel crosstalk from the direct sound from the front, which becomes 13-25dB for the frontal sound arriving from 0° to 45°. For this sufficient intensity reduction of the direct sound, it is not crucial to increase the delay time between the front and rear channels [Theile 2001]. It is suggested by the author that the OCT-Surround technique is most suitable for the recording of a small ensemble or a soloist. This seems to be because the short distances between the microphones would not result in sufficient low frequency decorrelation that is required for creating satisfying spatial impression for larger scale sources.

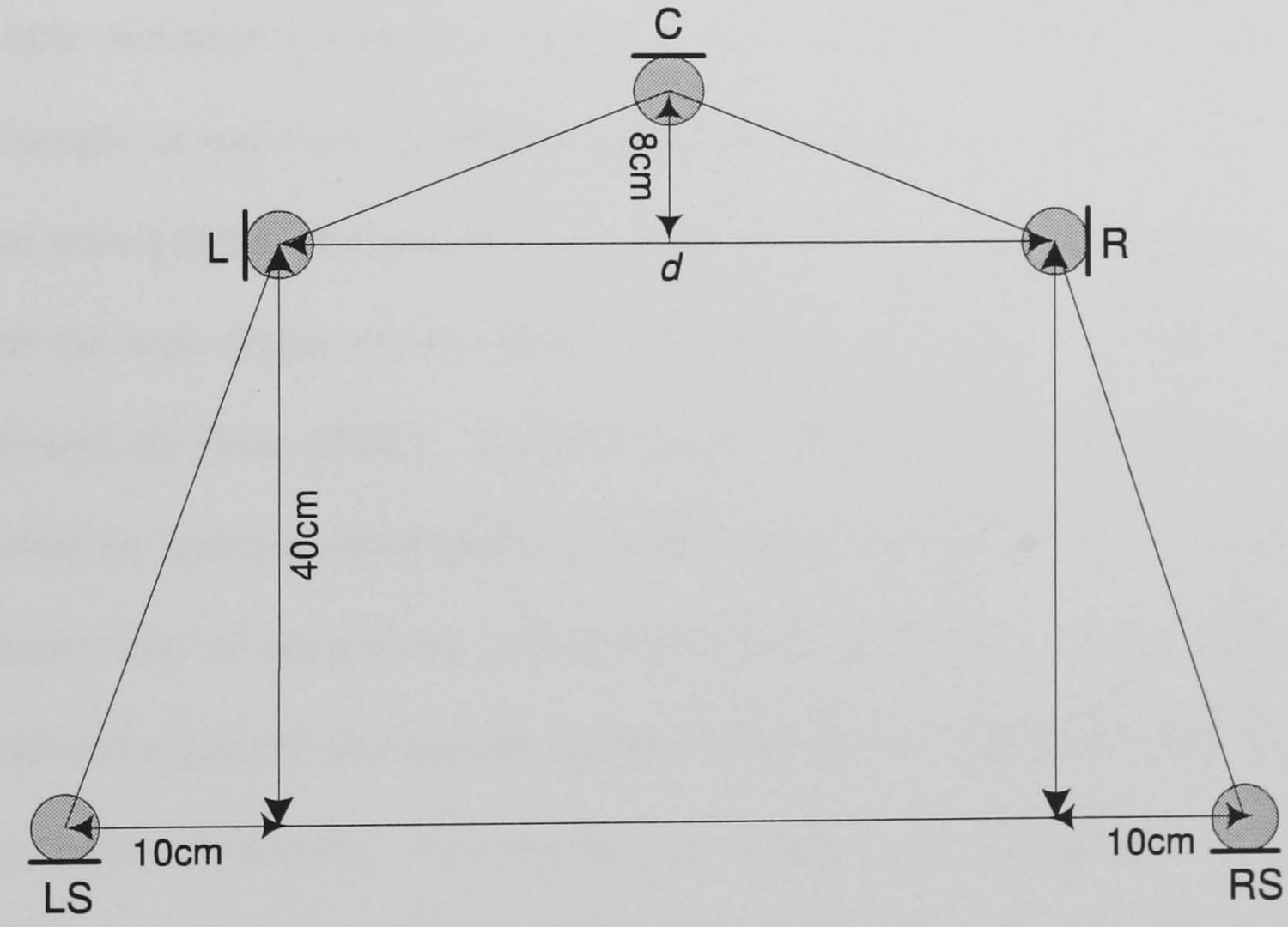


Figure 1.26 'OCT-Surround' five-channel main microphone array; the distance *d* varies according to the relationship shown in **Figure 1.20**

1.4.5 Discussions on the issue of interchannel crosstalk

There seems to be a strong disagreement between the viewpoints of Theile [2001] and Williams [2003] concerning the significance of interchannel crosstalk in multichannel microphone techniques for perceived sound quality. Theile claimed that the localisation quality would be decreased if interchannel crosstalk was not reduced enough in microphone technique design. However, Williams [2003] considered the linear attachment of SRAs for L-C and C-R (the critical linking) as a more crucial factor for improving the localisation quality than the suppression of interchannel crosstalk. Both authors attempted to achieve the aim of balanced phantom imaging with their own novel concepts but it seems that no one has achieved the aim perfectly. For example, it was shown in **Figure 1.18** that the ICA-3 array based on Williams' critical linking approach produced a continuous and balanced localisation across L-C-R, but the high degree of interchannel crosstalk in the array was claimed to be problematic by Theile [2001]. The OCT array [Theile 2001], on the other hand, is optimised for interchannel crosstalk by maximising channel separation and therefore the interference of unnecessary interchannel relationship of the two-channel based stereophonic segments other than the segment that is desired for phantom image of the source is not considerable. Nevertheless, this technique does not seem to provide a continuous transition of phantom images around the central listening area since the localisation curves for L-C and C-R slightly overlap (**Figure 1.21**). This discussion seems to suggest that the effect of interchannel crosstalk might not necessarily be problematic regarding the linearity of localisation curve, but more importantly related to the perception of various auditory attributes depending on the interchannel time and

intensity relationship involved in the crosstalk signal.

Several subjective experiments were conducted to compare the perceived auditory attribute qualities of the OCT and ICA-3, and they showed contradictory results. Wittek [2001b] compared the performances of different front microphone techniques. For each microphone technique, the phantom source image was compared with the monophonic source image of a single loudspeaker that was placed at the same position as the phantom image position. It was found that with regard to 'image focus' and 'sound colour' attributes, the phantom image created with the OCT was more similar to the monophonic image than that created by the ICA-3. From this result Wittek [2001b] suggested the superiority of the OCT technique in sound quality. This result certainly shows that the OCT provides a more precise localisation than the ICA-3. However, it might be questioned whether the preference of sound quality between two different techniques can be evaluated by the degree of similarity between the phantom image and the corresponding monophonic source image for each technique. The results of an experiment conducted by Heck and Riesebeck [2001, cited in Fukada 2001] are somewhat contradictory to Wittek's results. They evaluated the attributes of 'breadth', 'localisation', 'depth', 'transparency' and 'spatial impression' of the OCT-Surround, ICA-5, Fukada-Tree and critical linking techniques. It was found that there was no perceivable difference between the OCT-Surround and INA-5 in localisation quality. Moreover, the INA-5 was ranked as the best sounding array in overall attributes while the OCT-Surround was ranked as the worst. This author also conducted a subjective listening test to make a comparison between the OCT and critical linking techniques in the preference of perceived sound quality using various

types of sound sources comprising string quartet, percussion ensemble, solo violin and solo piano. It was found that the preference was dependent on the type of sound source used. For instance, the ICA-3 was preferred over the OCT for the solo piano recording while the reverse was true for the solo violin. The detailed method and results of this test are presented in Chapter 4.

The above results seem to suggest that interchannel crosstalk would not necessarily be an absolute parameter for decreasing the perceived sound quality. However, it is considered to be important for recording engineers to be aware of the perceptual effects of interchannel crosstalk on particular sound quality attributes and to be able to control the degree of interchannel crosstalk in microphone array design or application depending on the imaging characteristics required for the sound source to be recorded.

To date, no experimental data have been provided on the perceived attributes of interchannel crosstalk and their relative weights. From direct comparisons between different microphone techniques, it will be difficult to judge the effect of interchannel crosstalk alone since such parameters as microphone spacing and the amount of reflections or reverberation will also have their effects on the perceived sound. Therefore, further investigations that control the interchannel crosstalk as a single parameter are required and this is why the experiments described in Chapter 4 were carried out.

Additionally, for the microphone techniques with front and rear separation, it is considered that the interchannel crosstalk issue is most relevant to the front three channels only since the rear microphone arrays are usually placed at a distance that is

long enough for the direct sound to be decorrelated by reflections or reverberation. The rear channels in the five-channel main microphone techniques might suffer from considerable crosstalk due to the relatively short distances from the front channels. Furthermore, such techniques seem to be less practical in general applications than the techniques with front and rear separation due to the limitations mentioned in Section 1.4.1.

1.5 Summary

This chapter described the psychoacoustic principles involved in 2-0 and 3-2 stereophonic recording and reproduction. Firstly, the principles of phantom image localisation and the design and operational principles of microphone techniques for 2-0 stereo were reviewed. Then, the unique features of 3-2 stereophonic phantom imaging were discussed, followed by reviewing the design and operational principles of recent 3-2 stereophonic microphone techniques particularly with regard to interchannel crosstalk.

To summarise, the summing localisation theory suggests that the individual spatial cue of interchannel time difference (ICTD) or interchannel intensity difference (ICID) can cause the phantom image to shift to particular positions between two loudspeakers in 2-0 stereophonic reproduction, provided the ICTD is less than about 1.1ms. Within the range the summing localisation operates, the ICTD and ICID can be traded for desired phantom image localisations. Coincident pair 2-0 microphone techniques rely on the

ICID cue while the spaced pair techniques largely rely on the ICTD cue. The near-coincident microphone technique uses both cues depending on the relevant trading ratio for a certain stereophonic recording angle (SRA). The SRA for a microphone array is an important parameter for providing a balanced phantom image distribution and creating the width of the stereophonic images between the loudspeakers. It is calculated depending on the trading relationship between ICTD and ICID. Phantom imaging principles for 3-2 stereophonic reproduction are mainly based on those for 2-0 stereo. However, there are unique imaging characteristics in the 3-2 stereo due to the increased number of channels and the arrangement of loudspeakers. In particular, the localisation of side phantom images is typically unstable or inaccurate. This characteristic becomes relevant in the designs of 3-2 stereophonic microphone techniques, which are divided into two groups: techniques with front and rear separation and five-channel main microphone techniques. For the former techniques, the interchannel crosstalk is considered to be more relevant to the three-channel front techniques due to the small spacings between microphones rather than to the rear microphone techniques, which are normally placed further back in the recording space. The latter techniques, on the other hand, seem to suffer from interchannel crosstalk more seriously since the spacings between the five microphones are relatively small. However, such techniques are considered to be less practical than the other techniques in terms of flexibility. The representative novel front microphone techniques for 3-2 reproduction are Williams and Le Du [1999]'s 'Critical Linking' and Theile [2001]'s 'OCT'. Both techniques share the goal of accurate and balanced localisation of phantom images. The former technique attempts to link between the SRAs for the two stereo-base segments L-C and C-R without overlap whereas the latter attempts to

reduce the interchannel crosstalk as much as possible. Even though there is a much debate with regard to the significance of interchannel crosstalk between the two authors, neither microphone technique seems to provide ideal phantom imaging characteristics. Furthermore, the results of several subjective comparisons show that there is no absolute winner. However, more importantly, to date no systematic experimental data have been provided on the perceptual attributes of interchannel crosstalk and their relative weights, which would be likely to be important for recording engineers to know in order to design and operate microphone techniques more appropriately for particular recording situations.

2 PERCEPTUAL AND PHYSICAL EFFECTS OF DELAYED SECONDARY SIGNALS

The nature of interchannel crosstalk in multichannel microphone arrays, which takes the form of delayed and attenuated repetitions of a primary signal, could be compared well with the relationship of reflections to a direct sound in acoustics. For many years the effects of reflections on the perception of auditory attributes have been researched extensively in the field of room and concert hall acoustics. Studying the findings of those works could be the basis for understanding the perceptual effects of interchannel crosstalk in multichannel microphone technique. However, reflection in rooms typically has a much greater range of delay time than that of the interchannel crosstalk signals studied in this project. Furthermore, the direction of a reflection depends on the acoustic pattern of the environment, while that of interchannel crosstalk is determined by the placement of microphones, which is controllable. Therefore, it might be suggested that the context of concert hall studies does not directly correspond to that of this crosstalk study. However, most of the reflection experiments were simulated using stereophonic reproduction systems in anechoic chambers, for the purpose of controlling experimental variables. It is possible, therefore, to derive a useful hypothesis for the perceptual effects of interchannel crosstalk and also to map the relationship between those effects and physical parameters.

2.1 Perceptual Attributes of Reflection

In general, the room or concert hall research has shown that the presence of one or more reflections would be mainly related to the perception of three categories of auditory attributes, which are 'localisation', 'spatial impression' and 'tone colours' or 'timbre'. Even though there has been a large amount of research conducted on aspects of localisation and spatial impression, it seems that the properties of tone colouration have not been fully examined. In the literature it is only generally explained that a change in timbre is likely to be caused by the interference between direct sound and its reflection producing a comb filter effect, typically when the delay time of the reflection is in the range between 10ms and 50ms [Barron 1971, Haas 1972]. That is, since the reflection lags in phase relative to the direct sound, there will be cancellation at certain frequencies where the two are 180° out of phase, and augmentation at other frequencies where the direct and the reflected sounds arrive in phase. Because it is a function of wave length, the comb filter effect will create notches in portions of the frequency spectrum at regularly spaced intervals. However, reports on the subjective effects of reflections on tone colouration do not seem to provide a clear answer as to which specific timbral attributes such colouration affects. For example, Haas [1972] reported that the addition of reflections provides timbral richness to the direct sound and considered this as a desirable effect, whereas Barron [1971] regarded the tone colouration as a negative effect of reflection causing the perceived sound to be sharp or shrill. Haas's finding was only for a speech source while Barron's report was related to musical sources such as violin. It is also stated in Barron [1971]'s paper that the tone colouration effect would become particularly

dominant with broad band sources, heavy instrumentation and percussion instruments. This suggests that the tone colouration effect would be dependent on the spectral and temporal characteristics of the sound source. More research seems to be required on this issue. Since the colouration effects have not been studied widely or documented in detail in the literature, this chapter will concentrate on aspects of localisation and spatial impression.

2.2 Localisation

This section covers various aspects of the precedence effect, which is a primary influence on localisation in an acoustic environment. Firstly, the lower and upper threshold of the precedence effect is discussed. Then the physical characteristics of sound sources that are required to trigger the precedence effect are reviewed. Finally, the cognitive aspect of this effect is introduced.

2.2.1 Precedence Effect

The precedence effect can be described as a psychoacoustic phenomenon that enables one to easily localise the accurate position of direct sound in a reflected environment. The presence of reflection could be a disturbing factor for localisation. However, when the precedence effect operates, the auditory image will be consistently localised at the position of the direct sound regardless of the interference of reflection. It has already been mentioned in Section 1.1 that when the interchannel time difference

between the original (direct) and delayed (reflection) signals radiated from loudspeakers arranged in the standard two-channel stereophonic configuration is greater than approximately 1ms, the auditory image will be localised consistently at the position of the earlier loudspeaker, provided that both signals have equal intensity. This delay time is widely regarded as a lower boundary of the precedence effect [Blauert 1997]. As the delay time increases beyond the lower boundary, the auditory image will be consistently localised at the earlier loudspeaker until the delay time exceeds the upper boundary. Above the upper boundary, which is typically called the 'echo threshold', the reflection begins to be perceived as a separate sound source. The echo threshold varies widely depending on the type of sound source. Transient signals tend to have shorter echo thresholds than continuous signals. For example, the echo threshold for single clicks lies in the range between 2ms and 10ms [Rosenzweig and Rosenblith 1950, Thurlow and Parks 1961]. The echo threshold for noise pulses lies around 15ms [Damaske 1971]. On the other hand, for continuous speech signals it varies from 32ms [Meyer and Schodder 1952, cited in Blauert 1997] to 50ms [Haas 1972].

It is important to note that even though only one image is perceived at the direction of the original sound source (or the leading loudspeaker), the precedence effect is not a total masking or elimination of reflection information. The literature shows that it is possible to distinguish between auditory images with and without reflections. For example, it was reported by Freyman *et al* [1991] that the former would have greater loudness and spatial extent than the latter. Perrott *et al* [1988] found that the image created by the leading and lagging sounds tended to be extended toward the lagging

source and eventually fill the space between both sources. Blauert [1997] also states that the contribution of a delayed signal on spatial distribution becomes gradually greater as the delay time is increased.

2.2.2 Physical parameters for the precedence effect

Rakerd and Hartmann [1985] investigated the effect of reflection azimuth on the operation of the precedence effect. Their experiment was carried out in an anechoic chamber and a single reflecting panel was used to change the acoustic condition of the room. Five different room conditions were simulated by placing the reflecting panel in different positions: empty-room condition, ceiling condition, floor condition and side-wall conditions (left and right walls). It was found that all five situations produced different localisation judgments and particularly the side reflections caused the greatest difficulty in localisation. Hartmann [1983] regards this dependence as a limitation of the precedence effect. This finding might be relevant to the effect of interchannel crosstalk in five-channel main microphone techniques. It can be suggested that since crosstalk signals in the rear channels of the array are reproduced through the rear loudspeakers that are placed at the sides of the listener, the rear channel crosstalk might disturb localisation to a higher degree than the front channel crosstalk.

The temporal characteristics of a sound source play an important role for the operation of the precedence effect. The statements that are presented below explain that the

transient element of a sound is the main factor to trigger the precedence effect, rather than the steady-state element.

'The precedence effect can be demonstrated best when the sounds have some discontinuous or transient character. Steady tones or continuous and uniform noises are obviously not suitable because there is no way to define precedence. Clicks, on the other hand, work quite well, and speech or piano music are reasonably satisfactory.' [Wallach *et al* 1949]

'In a realistic acoustic environment steady-state sounds do not provide reliable information about the location of a sound source. Reflectors of the sound wave have as much influence on the waveforms present at any two points in the space as the locus of the sound source. Thus transient wavefronts, especially if later echoes can be suppressed, provide the most reliable cue to the location of the sound source.' [Yost *et al* 1971]

Evidence supporting the above statements can be found in Rakerd and Hartmann [1986]'s report. Rakerd and Hartmann conducted a subjective experiment to investigate the effect of onset duration on the operation of precedence effect in room condition. The experiment was carried out in an anechoic room using 500Hz and 2,000Hz sine tones. The stimulus was radiated from a single loudspeaker and a single reflection was produced using a single reflective panel. The delay time of the reflection was varied by changing the distance of the reflective panel. The onset duration of the stimulus was varied gradually from 0ms. It was observed that the

precedence effect was triggered maximally when the onset was instantaneous. As the duration of the onset was increased, the accuracy in localisation decreased. The experimenters proposed that this was due to a 'misdirection' effect by localisation cues in the steady-state sound field. The maximum onset duration that was effective for triggering the precedence effect was 100ms. Rakerd and Hartmann presumed that this negative effect of ongoing sound on localisation would be related to the 'plausibility' of the ongoing cue. They asserted that the ongoing (steady-state) cue of a tone would be typically unreliable for localisation in a room, based on a 'plausibility hypothesis' (this will be discussed in detail in the next section). Additionally, they found that onset rate (sound pressure level/unit time) was also a critical factor for the precedence effect; a signal with a higher peak intensity per unit onset duration gave rise to a more accurate localisation performance.

The 'Franssen effect' [1960, cited in Hartmann and Rackerd 1989] is a good example of the dominance of transient energy over steady-state energy for localisation in a room. Franssen conducted an experiment in an ordinary room with the standard stereophonic loudspeaker arrangement using low frequency sine tones (500Hz). As illustrated in **Figure 2.1**, a tone was sounded instantaneously at the left loudspeaker and decayed steadily to silence over 30ms. During the same period the signal at the right loudspeaker was increased steadily from zero to its peak and then maintained at the same intensity. It was found that the left loudspeaker was perceived to be still sounding. This is due to the illusion resulting from the persuasiveness in the instantaneous onset cue of the left signal [Hartmann 1993].

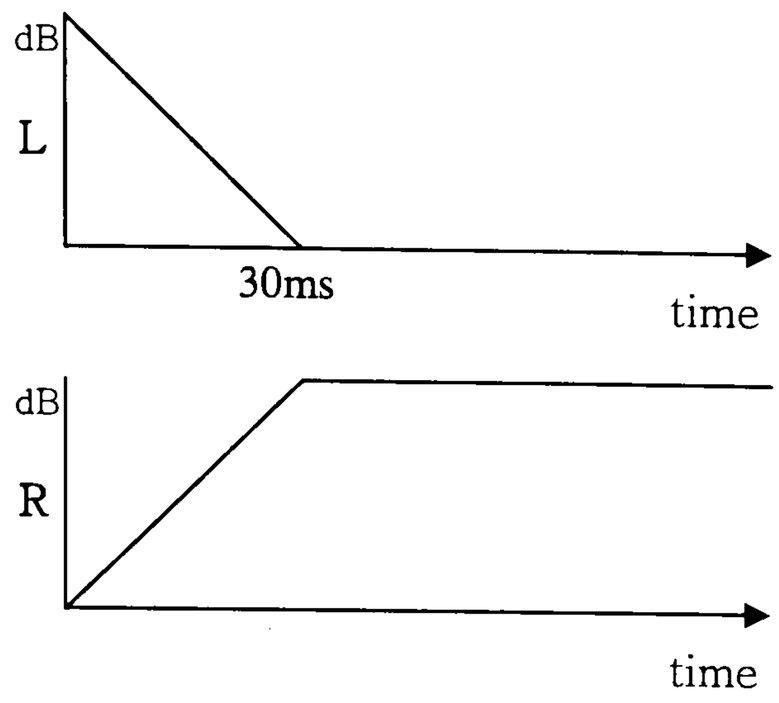


Figure 2.1 Illustration of the conditions for the Franssen effect

It seems important to note that while the above described investigations using pure tone stimuli reported the dominance of transient sound over steady-state sound in localisation, some investigations using more complex stimuli presented contradictory results. From Tobias and Zerlin [1959] and Perrott and Baars [1974]'s investigations using noise band signals, it was found that the ongoing cues became more effective for localisation than the onset transient cues as the duration of the signal increased. The onset transient cues lost their effect when the total signal duration exceeded 100ms. As Hartmann [1993] asserts, ongoing noise cannot be called a steady-state sound becuase it has too many random fluctuations. The high-frequency fluctuations can be described as a series of small transients that cause interaural time differences themselves, thus potentially triggering the precedence effect. This seems to suggest that for complex musical sound sources their temporal characteristics do not necessarily have to be discretely transient (e.g. percussion and piano) to be localised accurately. For example, when continuous stringed instruments were considered, the series of small transients caused by every bow or note change would be likely to contribute to the precedence effect.

Low frequency energy was also found to play a significant role in the precedence effect. Yost *et al* [1971] measured the effect of a single transient on the location of sound source using headphones. Two identical transient noise signals, one without a time delay and the other with some delay, were fed to the listeners' ears through headphones. The signals were low-pass filtered and high-pass filtered at various frequencies and the listeners were asked to discriminate the positions of the source as the filter cut-off frequency changed. The result showed that the listeners were better able to discriminate the position of the source when the signal contained energies below 1,500Hz than when frequencies below that were excluded [Yost *et al* 1971]. Banks and Green [1973] repeated the experiment that had been undertaken by Yost *et al*, using loudspeakers instead of headphones and obtained very similar results to Yost *et al's*. The only difference was the cut-off frequency value of the high-pass filtering, which was 2000Hz. A comparison of the results of Yost *et al* and Bank and Green gives rise to the hypothesis that 'binaural' precedence and 'stereophonic' precedence have very similar behaviour. Yost *et al* [1971] explain the reason for the low frequency significance for localisation from a physiological standpoint; low frequency transients vibrate more space in the cochlear partition than high frequency ones and excite more fibres, thus producing more substantial positional displacement.

Additionally, it was found by some researchers that reflection would not have to be an exact copy of the direct sound for triggering the precedence effect. Zurek [1980] found that uncorrelated white noise bursts showed the operation of the precedence effect. Blauert and Divenyi [1988] also reported that the precedence effect occured even when the frequency bands of direct sound and reflection did not overlap. It was

reported by Clifton *et al* [1994] that even though reflecting surfaces in a room absorbed low and high frequencies differentially, causing spectral distortion, reflections were still suppressed perceptually.

2.2.3 Cognitive processes in the precedence effect

The previous section reviewed some physical aspects of the precedence effect. However, there is experimental evidence that the precedence effect is not just a 'hard-wired' low-level process, but a high-level cognitive process.

Firstly, it was reported by Clifton [1987] that the precedence effect would require a 'build-up process'. In her experiment using two loudspeakers in an anechoic chamber, it was observed that when a single burst of pure tone was reproduced followed by a lagging version of the same signal with a delay time beyond its echo threshold, the listener initially heard both clicks separately. However, when the burst pair with the same delay time was repeated for a certain period of time, the perception of separate bursts halted triggering the precedence effect. This means that the echo threshold was raised gradually and the precedence effect built up during the ongoing stimulation supplying increasing information about the leading (direct) and lagging (reflected) sounds [Clifton *et al* 1994]. In a further investigation into the build-up process in the precedence effect that was conducted by Clifton and Freyman [1989] using the same experimental setup as Clifton [1987], it was found that the delay time between the clicks was one of the factors that would contribute to the build-up process.

When the delay time was shorter than 4ms, the perception of separate clicks disappeared in only 1–2 seconds; when the delay time was increased to 5–9ms it took 5–6 seconds. From the above findings, it may be possible to derive a hypothesis that the precedence effect is triggered depending on musical performance. For instance, for a large scale orchestra piece, the stringed instruments tend to generate transient information constantly at note changes or in tremolo passages and this might trigger a build-up process, whereas the percussion instruments such as timpani tend to be played occasionally and therefore there would be insufficient time to build up the precedence effect.

Secondly, it was also found in Clifton [1987] that the precedence effect could break down depending on changes in the acoustic condition. In the experiment click trains were presented through two loudspeakers, one leading the other by 5 ms. At the beginning of a click train (1 click/s), the listener localised a single click mainly at the leading loudspeaker. However, when the original and delayed clicks were spatially switched half way into the click train, most listeners heard two clicks from both loudspeakers for a few clicks. With repeated hearings, however, the precedence effect built up again as described above. This means that the sudden change in spatial location of the leading sound source caught the listener's selective attention and broke down the normal process of the precedence effect for a while, following an instance of a re-figuration in the set of stimuli [Blauert 1997]. This phenomenon is often called the 'Clifton effect'.

Finally, Rakerd and Hartmann [1985] proposed the 'plausibility hypothesis' and this supports the cognitive aspect of the precedence effect. According to this hypothesis, a listener evaluates the reasonability or reliability of the ITD (interaural time difference) cue perceptually and weights it accordingly [Hartmann 1993]. In other words, the human brain uses some kind of rapid decision-making process in sound localisation. In Rakerd and Hartmann [1985]'s experiments conducted in a single reflected room with a 500Hz sine tone, it was found that listeners almost ignored the ITDs that were unreasonably large. This means that implausible ITD cues are excluded subconsciously in the process of localisation judgment.

The importance of onset transient sound compared to ongoing steady-state sound in the precedence effect can now be explained by the plausibility hypothesis. It has been discussed that transient ITD cues trigger the precedence effect in a room with reflections [Rackerd and Hartmann 1985, 1986, Wallach *et al* 1949, Yost *et al* 1971, Zurek 1980]. The plausibility hypothesis suggests that transient cue is plausible as it wins in a competition with reflections and its ITD becomes apparently detectable, but the steady-state ongoing cue is implausible because it conflicts with room effect.

Hartmann and Rakerd [1989] point out that the steady-state cue can also be plausible in an anechoic room. They conducted an experiment regarding the 'Franssen effect', which has been described above, in a room with anechoic acoustics. The listeners' ability to detect transitions from one loudspeaker to another was investigated and it was found that in an anechoic room, the listeners could perfectly detect the transition from the transient source to the steady-state source, and thus the Franssen effect failed.

2.3 Spatial Impression

This section first defines various terminologies that are relevant to interpreting the concept of spatial impression (SI) and introduces the different paradigms of SI perception that have been proposed so far. Then, the objective parameters that can be used for measuring SI are discussed in detail. Finally, reports on the subjective preference for SI are reviewed.

2.3.1 Conceptual properties of SI

2.3.1.1 Classification of terminologies

In the past, spatial impression was often understood as a unidimensional attribute and the term was used to describe such spatial phenomena as 'source broadening' [Barron 1971, Barron and Marshall 1981], 'spaciousness' [Blauert 1997], and 'listener envelopment' [Beranek 1996]. However, most research on spatial impression conducted after 1995 tends to agree with defining SI as a multidimensional characteristic of an auditory event having two distinct sub-dimensions of 'apparent or auditory source width' (ASW) and 'listener envelopment' (LEV) [Bradley and Soulodre 1995, Hidaka *et al* 1995, Morimoto 2002], although there is still a lack of common definitions for these terms. Blauert and Lindemann [1986] and Beranek [1996] used the term 'spaciousness' also as a generic term comprising ASW and LEV. Morimoto and Maekawa [1988] referred to spaciousness as ASW. However, it was

claimed by Griesinger [1997] that ASW and spaciousness should be distinguished because the former should describe 'the impression of a large and enveloping space', thus being possibly equated with LEV. He asserts that spaciousness or envelopment is included in SI, and therefore ASW is a different impression from SI. Despite the variety in the use of the terms that are shown above, the current studies will follow the trend that assumes ASW and LEV are the properties of SI.

2.3.1.2 Paradigms of ASW and LEV perception

ASW and LEV are described in various ways by different authors although the main concepts of the terms are broadly similar. The descriptions for ASW include the following:

'The width of a sound image fused temporally and spatially with the direct sound image' [Morimoto and Maekawa 1988]

'The apparent auditory width of the sound field created by a performing entity as perceived by a listener in the audience area of a concert hall' [Hidaka *et al* 1995]

'The apparent width of the sound source' [Soulodre *et al* 2002]

On the other hand, LEV is described as follows:

'The fullness of sound images around a listener' [Morimoto and Maekawa 1988]

'The subjective impression by a listener that (s)he is enveloped by the sound field, a condition that is primarily related to the reverberant sound field' [Hidaka *et al* 1995]

'Listener's impression of the strength and directions from which the reverberant sound seems to arrive' [Beranek 1996]

It can be conceptualised from the above descriptions that ASW is a source-related attribute whereas LEV is more of an environment-related attribute. It is generally accepted in the literature that ASW is mainly related to early lateral reflections, while LEV is a property of late reflections or reverberation [Kuhl 1978, Barron and Marshall 1981, Bradley and Soulodre 1995, Hidaka *et al* 1995, Beranek 1996, and Okano *et al* 1998]. Barron and Marshall [1981] proposed that the upper limit of delay time for early parts of reflections should be 80ms in their experiment on spatial impression using musical sound sources of anechoic orchestral recordings. This was based on Schubert [1966, cited in Barron and Marshall 1981]'s threshold for the reflection to be perceived as a separate echo for a musical signal. Bradley and Soulodre [1995] conducted a similar type of experiment also with anechoic orchestral recordings, and used 80ms delay time as the threshold for dividing early and late parts of reflections. Okano *et al* [1998] also state that ASW is a property of reflections arriving within 80ms of the arrival of the direct sound, while LEV is generated by a reverberant sound field beginning after 80ms.

However, Griesinger [1997] proposes a different paradigm for explaining the perception of ASW and LEV, based on complex psychoacoustics of human perception of sound events. He firstly separates the spatial perception into 'foreground' and 'background' streams depending on different time divisions. Early reflections arriving within 50ms of the direct sound are interpreted as a foreground stream by the brain while reflections or reverberation arriving at the ears at least 120ms after the end of all foreground sound events are interpreted as a background stream. Griesinger [1997] asserts that the perception of the background stream is largely inhibited between 50ms and 120ms because this is perceptually the most insensitive region for spatial impression. According to his hypothesis, ASW is clearly distinguished from spatial impression (SI). Here SI describes a perception of being in an enclosed space. Separation between ASW and LEV is not determined by a simple time division of the signal, but by the relationship between the onset time of the direct sound and the delay time of early reflections. That is, ASW is increased only by reflections arriving during the onset time of the direct sound. Therefore, for a certain delay time of early reflections, a source signal with a faster onset will have a smaller ASW, while one with a slower onset will have a greater ASW. Reflections arriving after the offset or during the sound segments increase SI, which can convey an impression of LEV or spaciousness. It has to be noted that according to Griesinger, LEV and spaciousness are considered to be similar impressions. For discrete sound sources such as speech, the reflections in the foreground stream arriving during the sound segments or after the end of the direct sound contribute to the perception of 'early spatial impression' (ESI), while reflections in the background stream contribute to the perception of 'background spatial impression' (BSI). On the other hand, for continuous sound such

as a continuous part of orchestral music, neither ESI nor BSI can be produced. The SI produced in this case is called 'continuous spatial impression' (CSI). According to Griesinger [1997]'s descriptions, ESI is an acoustic impression that is associated with the direct sound in the foreground stream. It does not convey LEV information. BSI is an acoustic impression of envelopment that surrounds the listener, which is separated from the direct sound and usually created by diffused reverberation in the background stream. Finally, CSI is an impression that is related to both foreground and background streams. Therefore, it can be understood that both ASW and ESI are properties of the foreground stream that are source-related, whereas BSI is a property of the background stream that is environment-related and conveys an impression of LEV. In addition, CSI can be related to both source and environment, and therefore can convey both ASW and LEV information.

Although Griesinger's foreground-background paradigm introduced above is based on a complex psychoacoustic model of auditory perception and it seems that a number of detailed subjective investigations are still required to confirm the hypothesis, it introduces a unique and valuable concept which is the separation between source and environment perceptions depending on the temporal characteristics of the musical sound source. In particular, the introduction of the additional source-related attribute ESI to the conventional ASW suggests that auditory width perception can be multidimensional depending on the delay time of reflection and the envelope of the direct sound. From the viewpoint of the application of various musical sound sources having different characteristics in the measurement of perceived spatial effect, this seems to be more reasonable than the previously introduced methods that separate

ASW and LEV by dividing the early and late parts of reflections at a single value of 80ms. Even though the value of Griesinger's paradigm is acknowledged here, the trend of including the ASW in the SI category will be continued below since it is a more common way of classifying those attributes.

2.3.2 Objective parameters for SI measurement

Various ways in which SI could be measured objectively have been investigated by a large number of researchers in the field of concert hall acoustics. The objective parameters that could be used for SI measurement mainly include intensity and direction of reflection, frequency component of sound source, interaural cross-correlation and interaural fluctuation over time. The first two are related to the physical property of sound source, whereas the other two are related to the binaural relationship of ear input signals.

2.3.2.1 Intensity and direction of reflection

Barron and Marshall [1981] investigated the effect of a single early lateral reflection on the perceived spatial impression (as ASW). The experimental parameters included delay time, frequency spectrum, direction and intensity of the single reflection. They simulated sound fields in an anechoic chamber by reproducing a monophonic direct sound and discrete reflections derived from tape delay machines

through loudspeakers. The stimuli were anechoically made orchestral recordings. They found that the greatest spatial impression was observed for reflections arriving from the side of the listener (azimuth angles of 90°), while reflections arriving from directions in the median plane did not produce any increase in spatial impression. It is interesting to compare this finding with Rakerd and Hartmann [1985]'s finding, which showed that the side reflections were the most disturbing factor for localisation accuracy. This seems to suggest a conflicting relationship between localisation and spatial impression.

It was also reported by Barron and Marshall that perceived spatial impression increased as the ratio of reflected sound intensity to total sound intensity within the 80ms delay time became higher. They interpreted the detectable change of reflection intensity as being dependent on changes in spatial impression, and from this established the relationship between the magnitude of spatial impression and the ratio of early lateral reflection to direct sound intensity, which can be seen in **Figure 2.2**. This relationship suggests the significance of the intensity of early 'lateral' reflection on spatial impression. Based on these findings, Barron and Marshall proposed a physical measure for spatial impression named 'lateral fraction' (*Lf*), and the equation is shown below.

$$Lf = \sum_{t=5ms}^{80ms} r\cos\varphi \Big/ \sum_{t=0ms}^{80ms} r$$

where r = sound intensity

φ = azimuth angle of reflection from the lateral plane

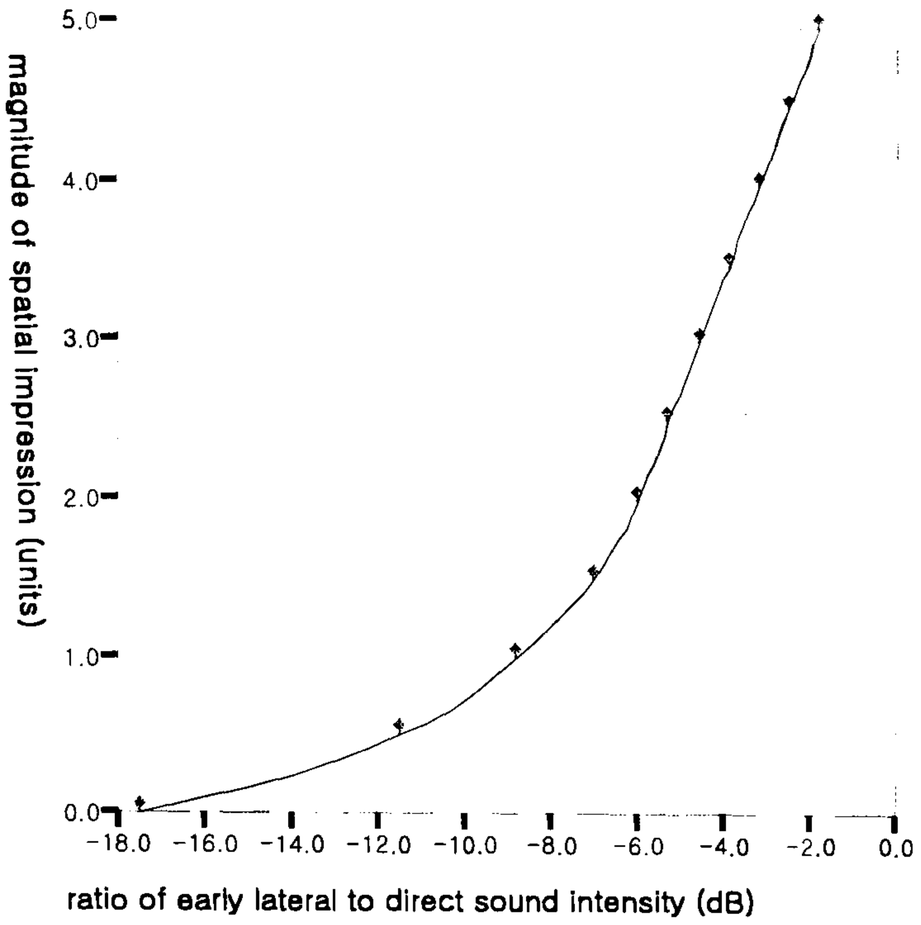


Figure 2.2 Relationship between the magnitude of spatial impression and the ratio between the early lateral to direct sound intensity [after Barron and Marshall 1981]

While Barron and Marshall [1981]'s investigation was limited to early reflections arriving at the ears within 80ms of the direct sound, Bradley and Soulodre [1995] were interested in the effect of sound arriving beyond 80ms such as reverberation. They conducted a subjective experiment using a similar method to Barron and Marshall's, with an anechoically recorded orchestral recording and simulated early and late arriving sounds generated by loudspeakers in an anechoic space. The experimental variables included intensity and direction of reverberation signals generated by loudspeakers. They found that reverberant energy arriving after 80ms produced a sense of LEV. It was further found that the effect of reverberation on the LEV perception had a very similar tendency to the effect of early reflections on ASW perception that was shown by Barron and Marshall. That is, the magnitude of the perceived LEV is proportional to the intensity of lateral sound arriving beyond 80ms after the direct sound. From these findings, Bradley and Soulodre proposed a

physical measure for LEV named 'late lateral energy fraction'. This measure is obtained using impulse response beyond 80ms after the direct sound, and defined in the equation shown below.

$$LF_{80}^{\infty} = \int_{80}^{\infty} p^2(t)\cos^2(\alpha)dt \Big/ \int_{80}^{\infty} p^2(t)dt$$

where p(t) = room impulse response
α = azimuth of angle of late reflection from the lateral plane

2.3.2.2 Frequency components of sound source and reflection

A number of investigations have been carried out with respect to the effect of frequency component on the perceived spatial impression. However, it is first necessary to distinguish the research that considered the frequency component of the sound source itself from those that dealt with the frequency component of the reflection signal. Examples of the former include the work of Morimoto and Maekawa [1988] and Hidaka *et al* [1995] and examples of the latter include the work of Barron and Marshall [1981] and Blauert and Lindemann [1986]. These are summarised and discussed in this section.

Morimoto and Maekawa [1988] investigated the subjective effects of the low frequency components of sound sources and the interaural cross-correlation coefficient (IACC) on spaciousness (in the form of ASW), using high-pass band limited white noise signals. The lower cut-off frequencies were 100, 200, 300, 400 and 510Hz while the upper cut-off frequency was constantly 5300Hz. The stimuli were

reproduced by three loudspeakers arranged at 0° and ±22.5° in an anechoic chamber, and their IACCs were varied by manipulating the ratio of lateral to frontal energy. In order to determine the independency of the effects of frequency components on ASW, the value of the IACC was kept constant while stimuli with different frequency ranges were tested by the subjects. The results showed that keeping the IACC equal, perceived ASW increased as the lower cut-off frequency decreased below 510Hz, with a particularly remarkable magnitude of increase between 100Hz and 200Hz. The relationship between the IACC and lower cut-off frequency that was found by Morimoto and Maekawa will be described in more detail in the next section.

The significance of the intensity of low frequencies for the ASW increase was also reported by Hidaka *et al* [1995]. Based on Barron and Marshall [1981]'s 'lateral fraction' theory, which was introduced above, Hidaka *et al* investigated the effect of increased sound intensities of orchestral music at frequencies above or below 355Hz on ASW. The result showed that increases of the intensities at lower frequencies caused greater increases of ASW than those at higher frequencies.

In Barron and Marshall [1981]'s subjective experiment described in the previous section, it was also investigated how the frequency components of lateral reflections affected the perception of spatial impression in concert halls. With the same experimental setup as described above, stimuli that were filtered into six octave band frequencies (125, 250, 500, 1000, 2000 and 4000Hz) were compared. It was found that 'source broadening' (as increased ASW) was caused by middle frequencies around 1000Hz, while lower frequencies contributed to an increase of 'envelopment'.

However, it is not entirely clear from their paper what they meant by the term 'envelopment' because it was described as 'the apparent area of the source is large' [Barron and Marshall 1981], which could also be interpreted as ASW increase. In fact, many writers tend to equate the envelopment perception found in this research with ASW perception.

Blauert and Lindemann [1986] conducted an experiment to determine the effective frequency components of early lateral reflections for 'spaciousness' (as spatial impression). They simulated reflective and reverberant sound fields in an anechoic chamber using three loudspeakers placed at 0° and ±45° or ±90° from the listener position. The centre loudspeaker was used for generating the original sound, and the side loudspeakers for generating the delayed sounds. The sound source was anechoically recorded orchestral music. A total of 12 test signals having different bandwidth of delayed sounds were created for comparisons between different frequency components of delayed sounds in terms of spaciousness perception. The acoustically synthesised sound fields were recorded with the dummy head placed at the listener position, and the binaural signals were presented to the subjects. According to their results, all frequency components of early lateral reflection contributed to spaciousness. Furthermore, it was reported that frequencies below 3kHz produced the 'sense of feeling enveloped by the sound' and 'expanded depth', while the higher frequencies caused the ASW to be increased. However, the perception of listener envelopment with only a few early reflections seems to be somewhat unreasonable based on the notion that the perception of listener envelopment is produced by late reflections or reverberation rather than early

reflections [Hidaka *et al* 1995, Bradley and Soulodre 1995 and Okano *et al* 1998].

For the effect of the frequency component of sound source, it was commonly found from Morimoto and Maekawa [1988] and Hidaka *et al* [1995]'s research that the low frequencies caused greater increases of ASW than the higher frequencies. For the effect of the frequency component of reflection, Barron and Marshall's and Blauert and Lindemann's reports seem to suggest that different frequencies of reflection signal might produce different width increasing effects related to the source. For example, Barron and Marshall observed two different perceptual attributes of envelopment and source broadening for low and middle frequencies respectively, although both attributes were described to be related to the source. This suggests that at least two separate 'source-related' width attributes could be perceived for different frequencies. Blauert and Lindemann also reported that all frequencies were taken into account in the perception of spatial impression. These findings might suggest that different frequency components of the reflection signal would produce different source-related width attributes. In fact, there is no standard way of describing the perceptual effects of ASW, and therefore it is possible that researchers use the term ASW commonly for different source-related effects that are frequency dependent. Therefore, detailed subjective elicitation experiments are required in order to examine the effects of frequency components of early reflections on the perceptions of various source-related width attributes, and accordingly new terminologies need to be developed together with clear definitions.

2.3.2.3 Interaural cross-correlation

The above sections discussed the effects of the physical properties of a sound source such as intensity, direction and frequency components on the perceived spatial impression. This section covers the effect of the relationship between the signals reaching the ears containing combinations of direct and reflected sounds on the perceived spatial impression. Over the years the 'interaural cross-correlation', which means the similarity between sound signals arriving at each ear, has been confirmed by researchers as one of the important binaural parameters related to the magnitude of perceived spatial impression in a concert hall. In a concert hall the degree of interaural cross-correlation will largely depend on the temporal and spectral patterns of reflections. The relationship between each ear signal is calculated using the 'interaural cross-correlation function' (IACF), which is defined in the following equation.

$$IACF_t(\tau) = \frac{\left[\int_{t1}^{t2} P_L(t)P_R(t+\tau)dt\right]}{\left[\int_{t1}^{t2} P_L^2(t)dt \int_{t1}^{t2} P_R^2(t)dt\right]^{1/2}}$$

where P = binaural impulse response (sound pressure)

L = left ear signal

R = right ear signal

t1 and t2 = period of time under measurement

τ = time offset between the two ear signals

The value of IACF depends on the value of τ, which varies in the same range of maximum interaural time difference, –1ms to +1ms [Hidaka *et al* 1995]. The maximum absolute value of IACF over all frequencies obtained within this range of τ is called the 'interaural cross-correlation coefficient' (IACC) and this is widely used as a standard measure for the calculation of interaural cross-correlation. The equation

for the IACC is shown below.

$$IACC_t = \left| IACF_t(\tau) \right|_{MAX} \text{, for } -1\text{ms} < \tau < +1\text{ms}$$

The background for the relationship between IACC and spatial impression can be found from psychophysical experiments conducted into the subjective effect of the magnitude of interaural cross-correlation on spatial attributes. For example, Chernyak and Dubrovsky [1968] investigated the subjective effects of different magnitudes of IACC on the perceived 'position' and horizontal 'extent' of the auditory event, with two wideband noise signals reproduced over headphones. The results indicated that a single fused auditory event with relatively smaller extent was perceived when the signals were perfectly correlated (IACC = 1). However, as the degree of cross-correlation decreased, the extent of the auditory event appeared to be greater even though the position of the auditory event kept unchanged. Although it is not clear whether this finding of consistency in the position of auditory event can also be understood as highly accurate localisation, this result seems to suggest that localisation accuracy is not necessarily decreased by increasing ASW.

Keet [1968] was the first to investigate the effect of the magnitude of interaural cross-correlation on the perception of spatial impression in a concert hall. He conducted a subjective experiment to judge the perceived ASW of the recordings made at various locations in a concert hall, using a near-coincident two-channel microphone technique. The sound source was a dry orchestral recording reproduced by a single loudspeaker. While the headphone and loudspeaker experiments mentioned earlier used

manipulation techniques for the variation of the IACC, the difference in the IACC values of the recordings in Keet's experiment was likely to be caused by the different reflection patterns encountered at each location in the hall where the microphones were placed. The results of the subjective experiment were compared to the IACC measured from the recordings of impulse response over the time period of 50ms ($IACC_{50}$), which were made in the same manner as those of the music signals. It was reported as a result that the values of IACC had a consistent and linear relationship with the subjective results. That is, the magnitude of perceived ASW increased as the IACC value was lowered.

This linear relationship between a low value of IACC and great magnitude of ASW was observed by other researchers in the field of concert hall acoustics. In Morimoto and Maekawa [1988]'s experiment described in section 2.2.2.2, it was also reported that keeping the lower cut-off frequency constant, the perceived magnitude of ASW increased linearly as IACC decreased, as can be seen in **Figure 2.3**. It can also be seen from the figure that the magnitude of ASW change due to the IACC change is constantly maintained regardless of the changes in lower cut-off frequency. From this research Morimoto and Maekawa concluded that IACC and low frequency contents of sound source and reflections affected ASW independently.

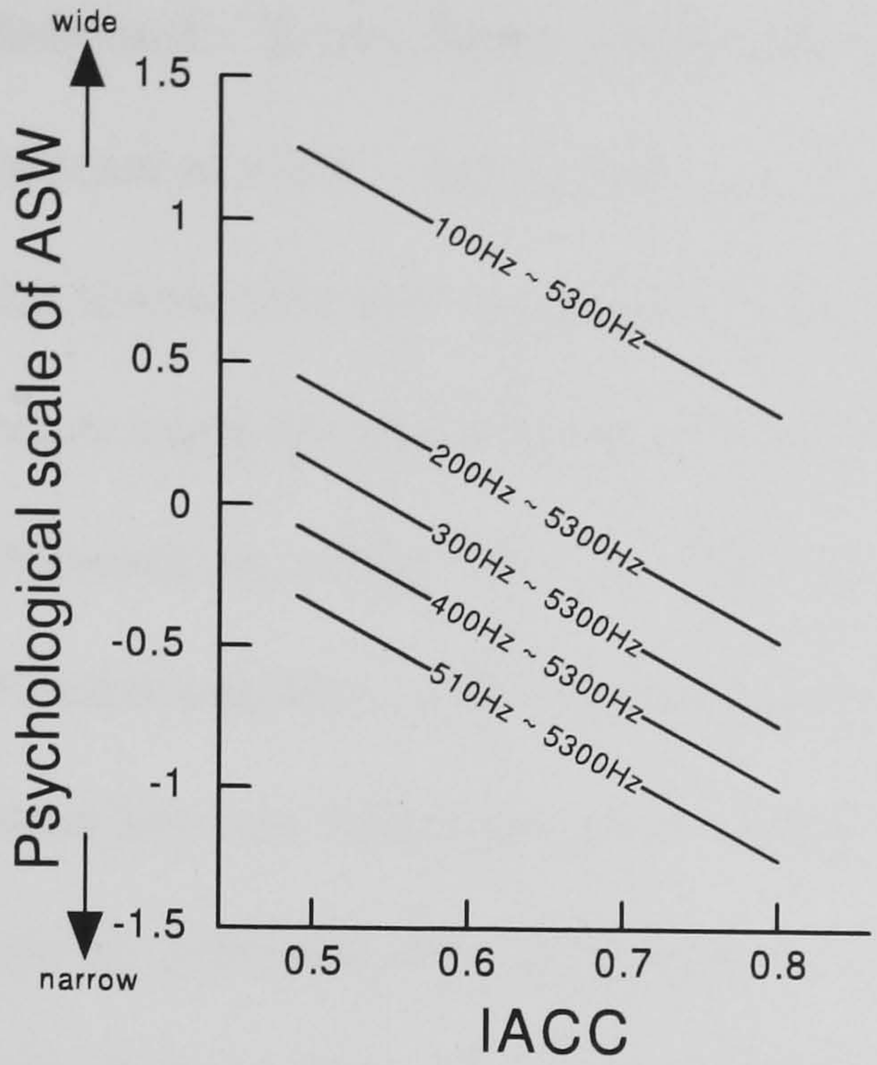


Figure 2.3 Effects of IACC and lower cut-off frequency of sound on the perceived ASW [after Morimoto and Maekawa 1988]

While Morimoto and Maekawa's results showed that there was no interaction between the IACC and low frequency contents of the stimulus, it was shown by Morimoto and Iida [1995] that the effect of IACC on the perceived ASW depended on the sound pressure level (SPL) of stimulus. In Morimoto and Iida's subjective experiment that was conducted in a simulated sound field created with three loudspeakers placed at 0° and ±45°, an anechoically recorded orchestral sound source was taken as a direct sound and it was reproduced from the centre loudspeaker. A pair of simulated reflection signals was fed into the side loudspeakers. The value of IACC was altered from 0.4 to 0.9 by varying the ratio of the direct sound and reflections, and the SPL of each stimulus with different IACC that was presented to the subjects was changed from 50dBA to 80dBA. The subjects were asked to grade the perceived ASW of the stimuli. The results generally indicated a similar pattern of IACC effect on ASW to those found in the above studies, in which the perceived ASW increased as the IACC

decreased. It was further shown that the increase in SPL also contributed to the increase of ASW. Interestingly, the change in the ASW due to the IACC change at the lowest SPL was very small while that at the highest SPL was dramatic. This result might be related to the effect of low frequency contents on ASW, which was discussed in section 2.3.2.2. The equal loudness contours that were devised by Fletcher and Munson [1933] show that with a higher SPL the ears perceive relatively more low and high frequencies compared to middle frequencies. Based on this, it can be presumed that the increase of the SPL of stimuli in Morimoto and Iida's experiment might have led to the perception of more low frequencies than the other frequency components, which might have been the main reason for the overall increase of ASW due to the SPL increase. This finding also suggests that the loudness equalisations of stimuli will be an important issue in the design of a subjective experiment investigating the perceived magnitudes of spatial impression.

As discussed so far, it appears that there is a general agreement about the effect of decreasing IACC on the increase of the perceived spatial impression, although most of the experiments mentioned above were related to the aspect of ASW rather than LEV. However, there are reports showing that this relationship is determined by the frequency range of the source signals reaching the ears. Hidaka *et al* [1995] reported a study on the frequency bands that make IACC effective for quality evaluation of concert halls. Based on Okano *et al* [1994]'s 'equal ASW' contours that shows the relationship between six octave-band frequencies and the corresponding IACC to make the source perceived equally wide (as can be seen in **Figure 2.4)**, Hidaka *et al* considered the three octave-band frequencies of 500, 1000, and 2000Hz to be the most

effective for measuring IACC. The lower frequency bands were excluded because the relative importance of IACC for ASW was small. The 4000Hz band was also excluded because its intensity for a typical orchestral music was considered to be 15dB lower than those of the 1000, and 2000Hz bands, therefore having little effect on ASW [Hidaka *et al* 1995]. From this choice of the most sensitive frequency bands for IACC measurement, Hidaka *et al* proposed two objective measurements for spatial impression, $IACC_{E3}$ and $IACC_{L3}$, with each being the average of the IACCs for the three bands. The former is measured based on the impulse response from 0ms to 80ms, thus being related to ASW perception, whereas the latter is from 80ms to 750ms, thus being associated with LEV perception.

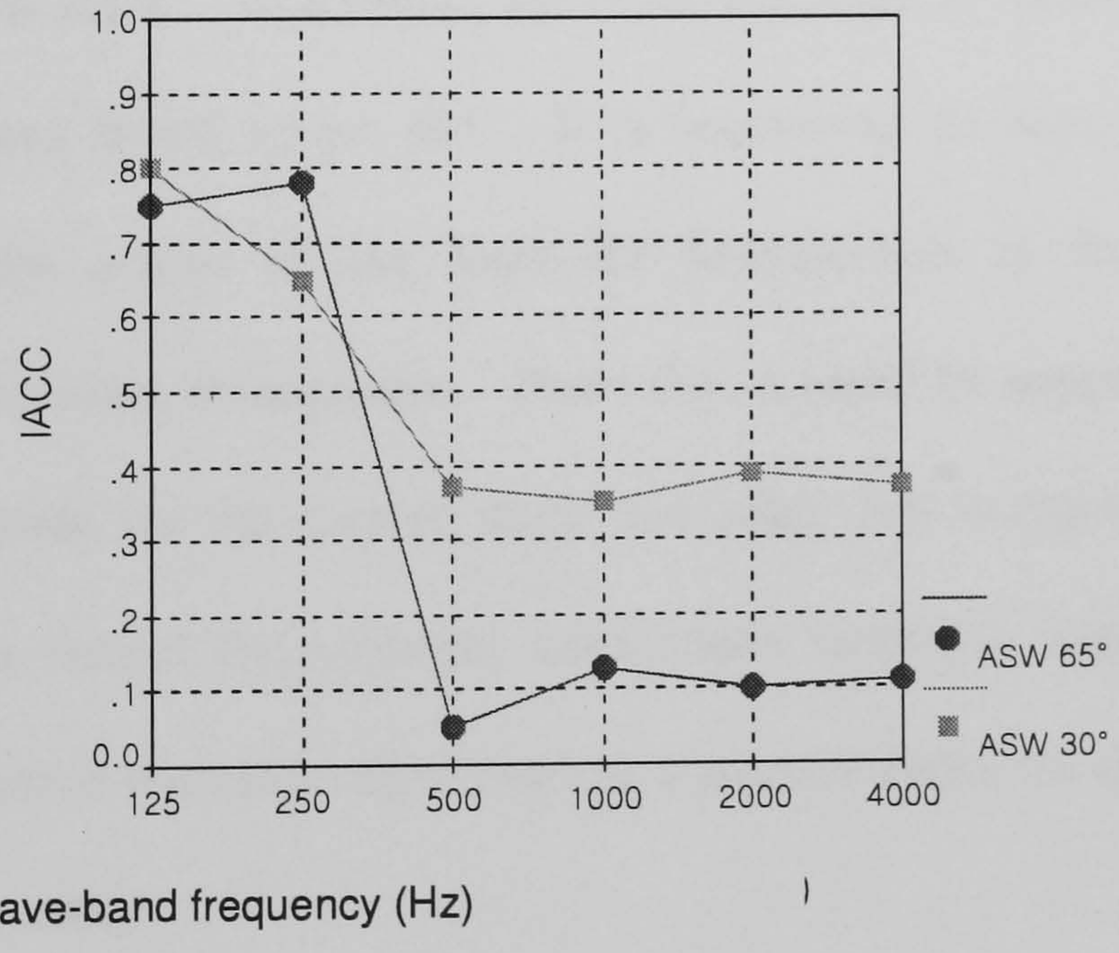


Figure 2.4 Equal ASW contours for octave-band frequencies [after Okano *et al* 1994]

In addition, IACC measurement is also found to be closely related to the prediction of the subjective preference of sound quality. Ando and Kageyama [1977] investigated

the relationship between the subjective preference and the magnitude of IACC of sound. In an anechoic chamber, a speech signal was reproduced by a loudspeaker placed at a central position in front of the listener. A simulated reflection was reproduced by a loudspeaker, and its direction was varied from the angle of 15° to 180° with intervals of 15°. The listener judged preferences for the direct sound only and the sound with the reflection of each direction, and the magnitude of IACC for each sound was measured. It was first found that the sound with a reflection was always preferred to the sound without a reflection. It was further found that a sound with a smaller magnitude of IACC tended to have a higher degree of preference. According to the relationship between the IACC and the magnitude of ASW discussed earlier, this finding means that a sound having greater ASW is likely to be preferred to that having smaller ASW. In addition, the reflection angle at which the sound was most preferred was found to be 30°. It is interesting to note that this angle corresponds to the angles of the front-side loudspeakers in the standard 3-2 stereophonic loudspeaker arrangement. From this, it could be suggested that, under the assumption made for the current study that either left or right channel signal generated from a frontal three-channel microphone technique becomes unwanted crosstalk, interchannel crosstalk might even be a positive factor for the preference of perceived sound quality.

2.3.2.4 Limitation of the current IACC measurement technique

Currently the most widely acknowledged IACC measure as a predictor of perceived

ASW is Hidaka *et al* [1995]'s $IACC_{E3}$ that was introduced earlier. This technique uses a binaurally recorded impulse response as a source signal for measurement since in this way it is possible to analyse the temporal characteristics of sound easily. However, the use of an impulse response is claimed to have a serious limitation in predicting the perceived effects accurately because a transient impulse signal has different spectral and temporal characteristics to the more complex musical signal that is actually heard in the listening space [Griesinger 1997, Mason 2002]. Mason *et al* [2004] state that it is the spectral characteristics of the sound source as well as the pattern of reflections that determine the interaction between the direct and reflected sound, which affect IACC measurement. Therefore, the interaction between direct and reflected sound resulting from a transient impulse will be small compared to a more complex musical signal. In fact, Griesinger [1997] reported that the IACC measured with a musical signal had a lower value than that with the corresponding impulse response. Mason *et al* [2004] also reported that there was a great difference between the IACC measured with an impulse response and that with a complex and continuous tonal signal. They suggested that the more accurate objective judgment of spatial impression in a concert hall using the method of IACC measurement could be achieved with representative source signals having spectral and temporal characteristics similar to musical signals.

Another arguable aspect of Hidaka *et al*'s approach is the use of the specific time value of 80ms for dividing the attributes of ASW and LEV. This value seems to be simply based on Barron and Marshall [1981]'s value, which was originally taken from Schubert [1966]'s echo threshold for musical signals. However, it was discussed

earlier that the value for defining the boundary between the early and late parts of spatial impression could be assumed to vary depending on the echo threshold, which also depends on the type of sound source. For example, Haas [1972]'s echo threshold obtained for speech signal was 50ms, and many sources show that transient clicks generally have much shorter echo thresholds compared to continuous and complex signals [e.g. Rosenzweig and Rosenblith 1950, Thurlow and Parks 1961]. Therefore, it is debatable whether the division of a source signal at 80ms would accurately separate attributes of ASW and LEV in every case.

Mason *et al* [2004] developed a new IACC measurement model to overcome the limitations of the conventional IACC measurement technique described above. Basically, this model is designed to measures the time-varying IACC of musical sound source instead of the transient impulse response, making it suitable for the applications of both concert hall and sound reproduction. For this reason, this model is considered to be useful for predicting the perceived effect of interchannel crosstalk in multichannel microphone technique in an objective way. The detailed working principles of this model are described in Chapter 5.

2.3.2.5 Fluctuations in interaural time and intensity differences

Another important objective parameter for the measurement of spatial impression is interaural fluctuation, which is based on the measurement of the magnitude of variations in interaural time difference (ITD) or interaural intensity difference (IID)

over time. Unlike the conventional IACC measurement, the measurement of interaural fluctuations over time is applied for continuous musical signals, thus being more suitable to be applied to the evaluation of sound quality in sound recording and reproduction [Mason 2002]. This suggests that interaural fluctuations over time could be directly related to understanding the causes for the resulting effects of interchannel crosstalk in multichannel microphone technique. The research that has been conducted to investigate into the effect of ITD and IID fluctuations is based on Blauert's finding of the phenomenon called 'localisation lag'. Blauert [1972] investigated the pattern of lateralisation affected by different rates of interaural fluctuation. A continuous train of pulse signals was presented to both ears using headphones and the interaural time and intensity difference between each channel were altered with various rates. It was found that the created sound images were perceived to be moving at low rates of the fluctuation and this phenomenon disappeared as the fluctuation rate increased. Grantham and Wightman [1978] further investigated the threshold of this effect using frequency modulated noise signals and found that the perception of movement was changed to that of increased width beyond the fluctuation rate of 20 Hz. Griesinger [1997] also investigated the same effect with a continuous band-limited pink noise and indicated that the threshold of the localisation lag was 3 Hz. He also reported that the source was perceived to be 'stationary' in the presence of a 'surround', and this seems to suggest the effect of interaural fluctuation on the increase of spatial impression or ASW.

Blauert and Lindemann [1986] investigated the effect of fluctuation in time or intensity difference on the perceived spaciousness (as ASW) individually, using a

band-limited impulsive signal. In their experiment, two simulated sound fields were considered. They were both created from the direct impulsive sound and simulated reflections in an anechoic chamber, but one was created with fluctuation in ITD only and the other with IID only. Two impulse responses were produced for each sound field and they were recorded with a dummy head. One of the binaural impulse responses for the sound field with either the ITD or IID fluctuation was then modified so that the fluctuation was removed, thus having identical signals at both channels. The original and manipulated signals were finally convolved with an anechoic orchestral recording. The subjects were asked to compare the difference in terms of the perceived spaciousness between the original and manipulated signals in headphone reproduction, and the results indicated that in the cases of both sound fields the original signals, which contained the fluctuations, were perceived to be more spacious (wider).

The individual effect of ITD or IID fluctuation was further evaluated by Griesinger [1992] although the detailed experimental method was not indicated in his paper. From a subjective listening test conducted with 1/3 octave band noise signal modulated with 5 Hz fluctuation in either ITD only or IID only, it was reported that both the ITD and the IID fluctuations contributed to the creation of spatial impression with each having different localisation characteristics. That is, the latter provided a well localised sound image while the former produced a poorly localised image.

The above results appear to suggest that the interaural fluctuations created by the interaction of a direct sound and reflections influence the increase of perceived

magnitude of ASW, and this was recently confirmed by the results of experiments conducted by Mason [2002] although his work was focused on the aspect of ITD fluctuation only. In his series of subjective elicitation experiments the effect of different frequencies and magnitudes of ITD fluctuations were investigated with both headphones and loudspeakers using frequency modulated noise stimuli. The perceived attributes of ITD fluctuations were analysed from the results of graphical elicitation tests. In terms of the effect of fluctuation frequency, it was found that in experiments with both headphones and loudspeakers the 'localisation lag' effect, which was introduced above, was observed as the frequency of the fluctuation rose. In terms of the fluctuation magnitude it was reported that with headphone listening mainly the perceived 'width', 'depth' and 'height' of the source were increased as the magnitude increased. On the other hand, with loudspeaker listening, the increase of the fluctuation magnitude was found to cause increases of perceived 'width' and 'envelopment'. These results suggest that the measurement of ITD fluctuation can be successfully related to the measurement of spatial impression.

In an acoustical environment, these fluctuations are naturally produced by the interaction between a direct sound and reflections [Griesinger 1992]. In order to explain the creation of interaural fluctuations simply, Mason [2002] simulated the interaction of a direct sound and a single reflection in an acoustical environment by modelling a sound source 15 metres directly in front of a dummy head and a single side wall placed 5 metres away from the lateral plane of the dummy head. It is stated by Mason that if the direct sound in this model is a complex signal, the interaction that might result between the numerous frequency components of the direct and reflected

signals will produce changes in the interaural time and intensity differences of ear input signals over time. **Figure 2.5** shows the examples of ITD and IID fluctuations that are measured in this particular model using a complex source signal consisting of three continuous sine tones of 480, 500, and 520 Hz [Mason 2002]. The fluctuation patterns shown in these figures are repeated in the same manner over time.

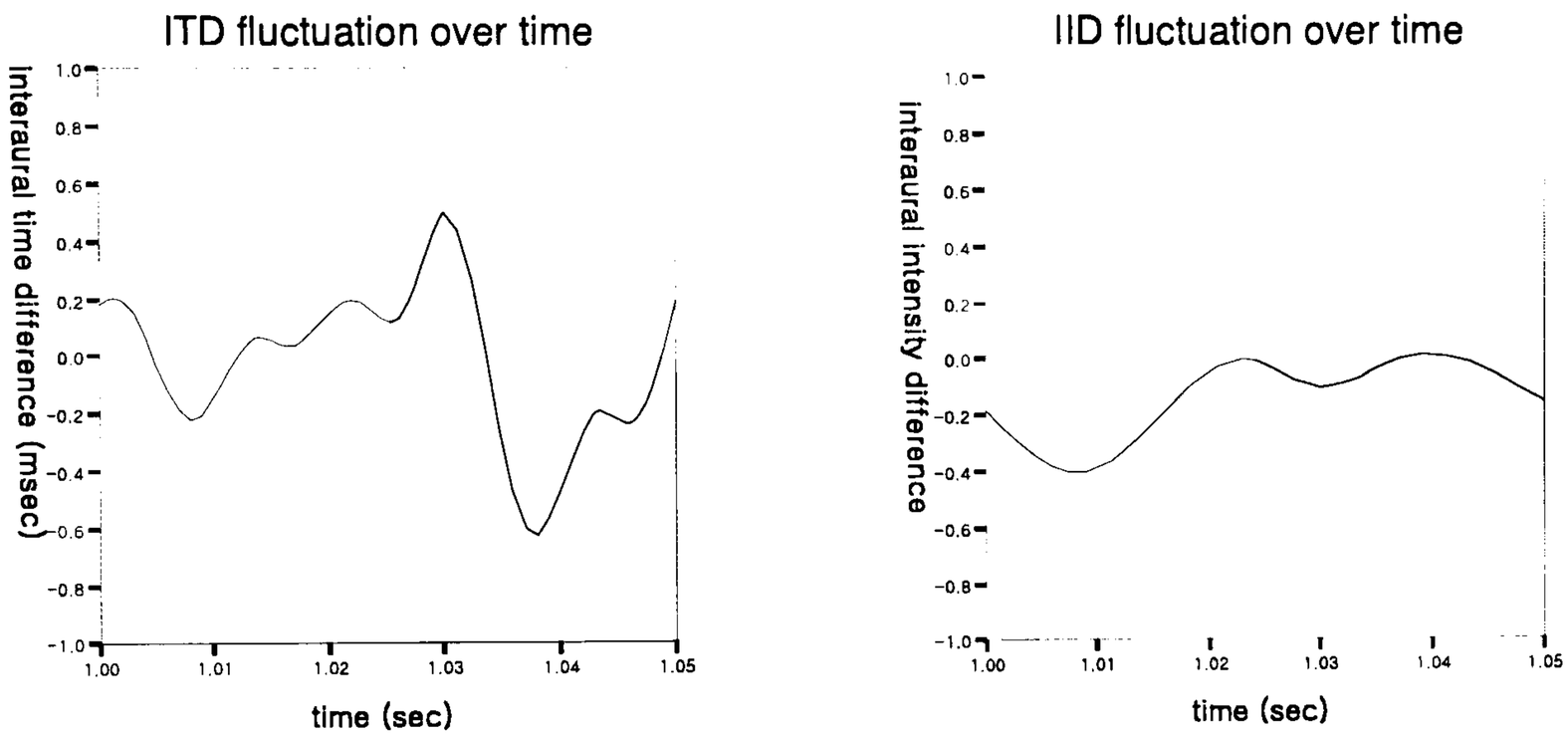


Figure 2.5 Plots of the ITD and IID fluctuations over time measured for Mason's simulation model of an acoustical environment producing a single reflection, with a source signal consisting of three continuous sine tones of 480, 500, and 520 Hz [after Mason 2002]

Mason [2002] indicated that the creation of these fluctuations was influenced by the properties of the source signal (e.g. frequency response and intensity) as well as the reflection pattern (e.g. delay time and direction). In the acoustical model introduced above, he found that the frequency response of the ITD fluctuation changed as they altered the delay time and the direction of the reflection by changing the distance between the side wall and the dummy head. However, it was further indicated that despite the changes in the reflection pattern, the peak of the frequency response of the

fluctuation was maintained at 20 Hz, which was the frequency of spacing of the three sine tones (480, 500, and 520 Hz). Furthermore, it was also found that there was no linear dependency of the magnitude of the fluctuation on the reflection pattern. These findings suggest that the properties of the sound source have a more significant effect on the characteristics of interaural fluctuation than the reflection pattern. From this, Mason [2002] moved on to the measurement of interaural fluctuation using musical source signals, which were anechoic recordings of a continuous cello note and a transient acoustic guitar chord. For simulating the interaction between the source and reflections in a room, the stimuli were convolved with the binaural impulse response of a room simulation. The duration of the source from the onset to the offset was 0.5 second for the cello note, and 2 seconds for the acoustic guitar. **Figure 2.6** and **2.7** show the plots of the ITD and IID fluctuations over time for these stimuli. It can be firstly seen that the ITD fluctuations are more obvious and erratic than the IID fluctuations for both sources. It can be also observed that the erratic ITD fluctuations are not generated at the onset, but during the note and after the offset due to the reflections. This means that the spatial impression is not generated at the onset of a sound, but at the arrivals of reflections during the note and at the offset of a sound where the reflections and reverberation have maximum energies. This is supported by Griesinger [1996], asserting that a great magnitude of spatial impression is produced in the space between the notes of a musical sound source, where the energies of reflections are maximal and therefore the fluctuations in ITD and IID are particularly large. For the cello note, the erratic ITD fluctuations are generated continuously during the note and the reverberation. This is likely to be because the ongoing variations in frequency and intensity during the length of the note interacted

with reflections continuously. On the other hand, the acoustic guitar chord has a distinctive difference in fluctuating pattern between the source and the reverberation. The ITD fluctuations are relatively constant during the length of the chord, even thought there are some erratic fluctuations after the onset. This is likely to be because the ringing after the transient plucking of the acoustic guitar chord is relatively constant over the length of the chord. The fluctuations become much more erratic in the reverberation part. These findings suggest that the characteristics of interarual fluctuations are dependent on the temporal characteristics of a source.

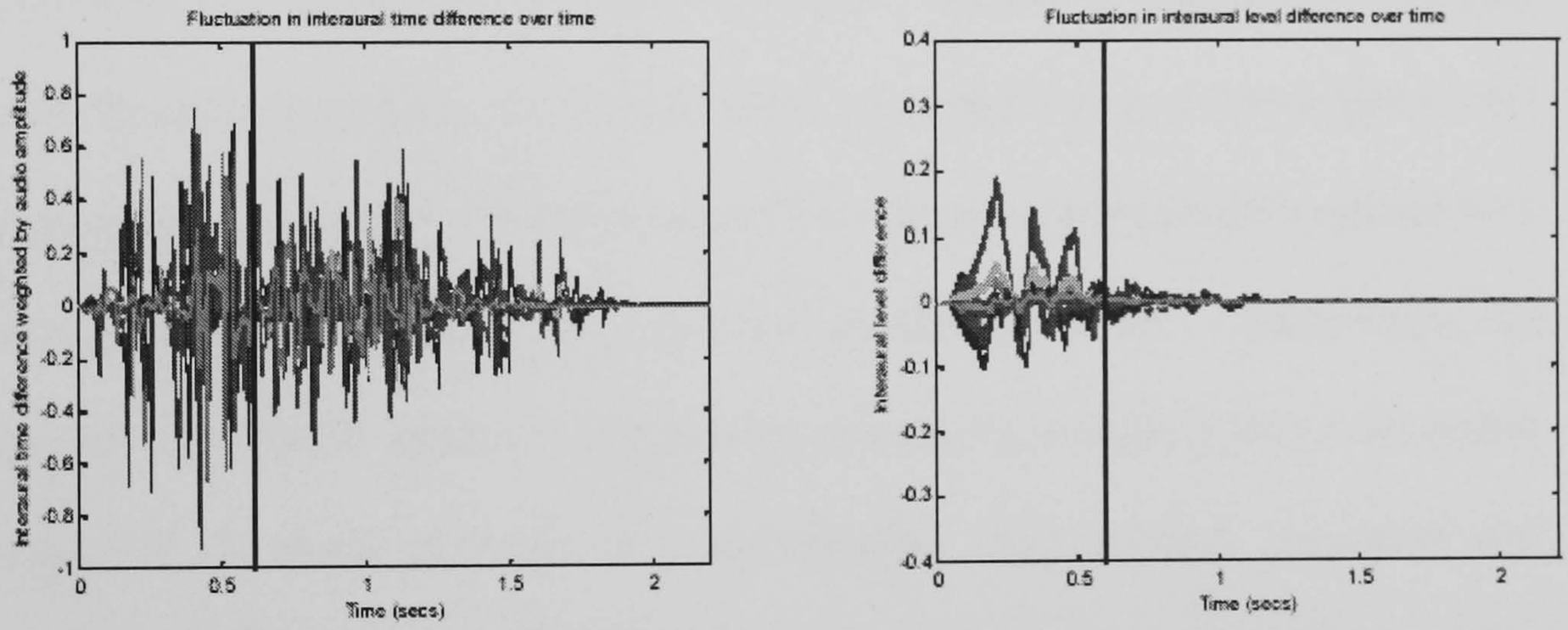


Figure 2.6 ITD and IID fluctuations for cello note [courtesy of Mason 2002]

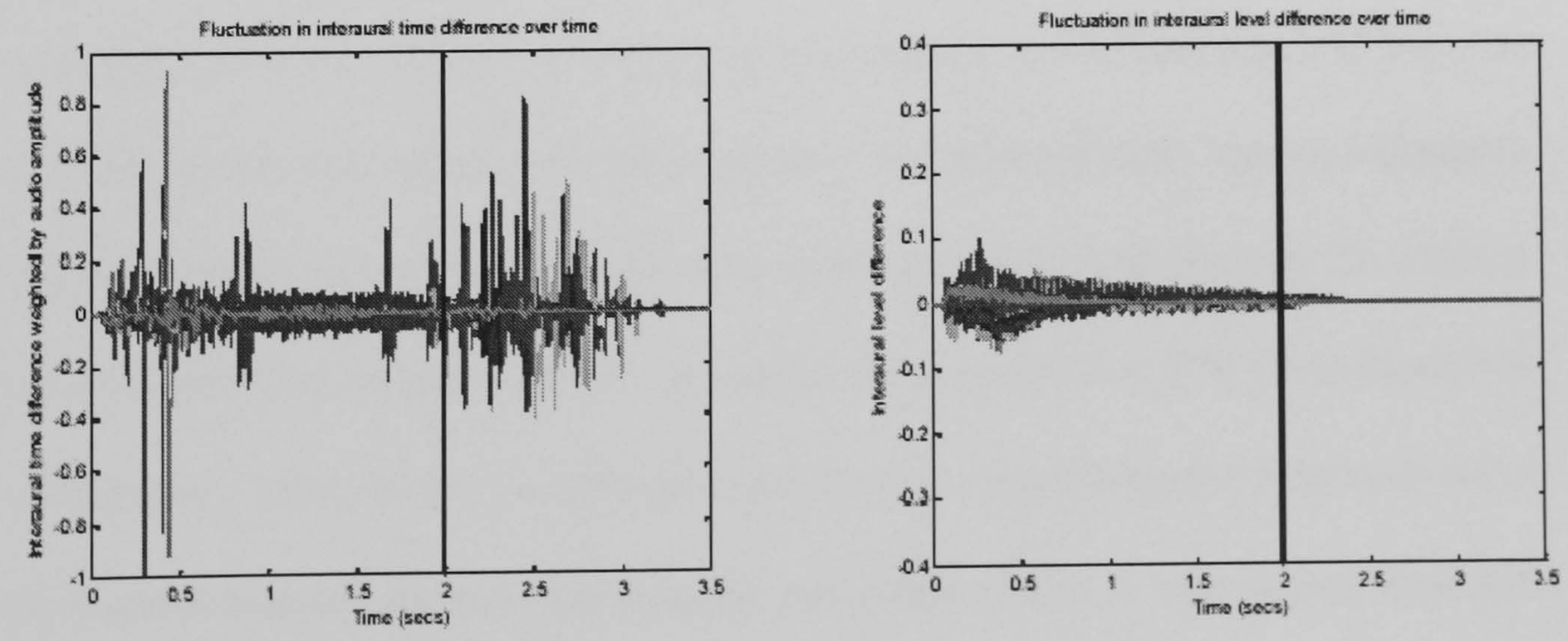


Figure 2.7 ITD and IID fluctuations for acoustic guitar chord [courtesy of Mason 2002]

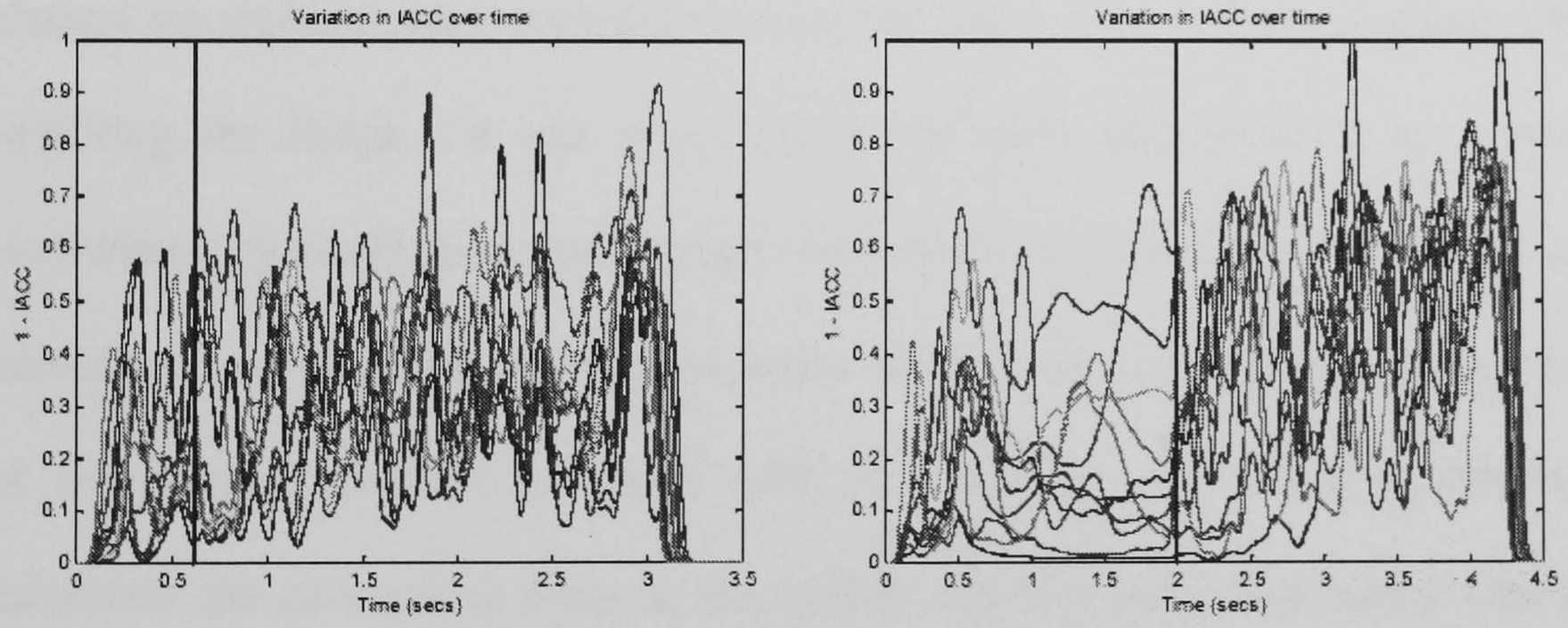
From the above findings, the interaural fluctuation can be considered as a physical cue that is directly related to the properties of a complex musical source, such as frequency response, intensity and temporal characteristics of the musical performance, as well as the pattern of resulting reflections in an acoustical environment. Therefore, the measurement of interaural fluctuation of musical sound sources over time would be more effective for investigating the effect of reflection in a concert hall than the measurement of IACC using the response of a rather unrealistic impulsive sound source within a fixed time window. In fact, having claimed that the IACC has limitations as an objective measure for spatial impression created with complex musical source in a concert hall as mentioned in the above section, Griesinger [1997] proposes that the measurement of interaural fluctuation is a more suitable method for a more accurate prediction of the perceived spatial impression. Furthermore, the interaural fluctuation measure could also be suitable for evaluating perceived spatial impression in sound recording and reproduction. For example, coincident and spaced pair microphone techniques will differ in perceived source width due to the different magnitudes of interaural fluctuations created by each technique. Coincident techniques produce signals that are largely correlated at low frequencies and therefore the created ITD fluctuations will be minimal. On the other hand, spaced techniques will minimise the correlation between the resulting signals depending on the distance between the microphones, therefore increasing the magnitude of ITD fluctuations that are created. From this, it is considered that the interaural fluctuation over time could be a useful measure for understanding the perceptual effects of interchannel crosstalk that might be dependent on the interchannel relationship in multichannel microphone technique.

2.3.2.6 Relationship between interaural fluctuation measurement and IACC measurement

Griesinger [1992] noted that the IACC would be closely related to fluctuations in both intensity and phase of a signal in a certain pattern. Mason [2002] conducted an investigation into the relationship between interaural cross-correlation and fluctuations in interaural time and intensity difference over time using various stimuli. He firstly analysed the effect of varying magnitude of ITD and IID fluctuation on the maximum IACC across the range of +/- 1ms. The stimulus signal consisted of a pair of 500Hz sine tones that were modulated either in frequency or amplitude at 5Hz. The magnitude of ITD or IID fluctuation was varied by creating different magnitudes of frequency or amplitude modulation and the IACC was measured for each variation. The results showed that as the fluctuation in both ITD and IID increased, the maximum IACC value decreased. However, it was also found that a change in the IACC caused by a change in IID fluctuation was less than that caused by a similar change in ITD fluctuation.

From this, Mason moved on to a further investigation using musical stimuli. He compared the characteristics between the IACC variation over time and the fluctuation in ITD or IID over time for the stimuli of a single cello note and a single acoustic guitar chord, which were shown in **Figure 2.6** and **2.7**. The measurement plots of the IACC over time for these sources are shown in **Figure 2.8**. It can be seen in general that for both sources the measurement of the ITD fluctuation is more similar to the measurement of the IACC than the measurement of the IID fluctuation. Even though

the cello source has erratic variations in both ITD and IACC, it is difficult to observe an obvious similarity between the two in terms of the pattern of variation. On the other hand, the acoustic guitar source appears to have more similarities between the ITD fluctuation and the IACC variation in that there are several peaks in the region of the early reflections and there are more erratic fluctuations in the region of the late reflections or reverberation. From these findings, Mason [2002] concluded that the frequency and envelope dependent interaural fluctuations over time are the main factors that affect the interaural cross-correlation of a signal. This also suggests that frequency and envelope of a sound would be directly related to the width perception for the sounds with secondary delayed signals.



(a) Cello (b) Acoustic guitar

Figure 2.8 Plots of the reversed IACC (1-IACC) for single cello note and acoustic guitar chord, measured for different frequency bands of the signal [Courtesy of Mason 2002]

2.4 Discussions

Since early reflection in acoustic space and interchannel crosstalk in microphone technique are commonly in the form of secondary delayed signals, the findings from the reflection studies that have been reviewed in this chapter are considered to be useful for hypothesising the kinds of attributes interchannel crosstalk would affect and which physical cues would give rise to the perception of those attributes.

It is firstly proposed that interchannel crosstalk would be a disturbing factor for phantom image localisation. Similarly to early reflections, however, if the precedence effect was triggered between the crosstalk and wanted signals in a three channel microphone array, interchannel crosstalk would be perceptually masked when localising the image. It has to be noted that early reflections in an acoustic environment typically have much longer delay times than interchannel crosstalk in a microphone array. Also, for near-coincident microphone techniques, the delay time of interchannel crosstalk is traded with its intensity. For example, the more coincident the microphone array is, the smaller intensity and shorter delay time the crosstalk signal will have. Therefore, even if the delay time of the crosstalk signal fell under the threshold for the precedence effect in highly coincident arrays, its small intensity might still lead to the localisation of an image at the desired position. However, the accuracy or easiness of localisation would depend on the temporal and spectral characteristics of sound source. Based on findings related to the precedence effect in acoustical environments, transient nature and low frequency components in a sound source are necessary for triggering the precedence effect. Such continuous

sound as a pure tone would therefore be difficult to localise. From this, it could be predicted that interchannel crosstalk would also affect the accuracy of phantom image localisation depending on the temporal and spectral characteristics of the sound source. However, since musical signals, which are most likely to be the sound sources in recordings using microphone techniques, have complex and unique characteristics in their spectra and envelopes, findings relating to the precedence effect obtained using pure tones or noise signals might be applied differently in the context of interchannel crosstalk.

Secondly, it was reported by many authors that the addition of a reflection arriving within about 80ms after the direct sound would contribute to the perception of apparent source width (ASW). Griesinger [1997] proposed from a different viewpoint that only the reflection arriving within the onset time of the direct sound contributes to the increase of ASW. No matter which paradigm is believed, it can be predicted from the above that interchannel crosstalk might contribute to the increase of the perceived width of a phantom source image since it has a relatively small range of delay time, which is normally less than a few milliseconds. If this is the case, similarly to the case of reflection effects, the perception of ASW due to interchannel crosstalk would be affected by such physical parameters as the intensity of crosstalk signal and the frequency components of sound source. IACC and ITD fluctuations might also become useful parameters for measuring the source width increase caused due to interchannel crosstalk objectively. However, because the ratio between interchannel time and intensity differences in a crosstalk signal will vary as the microphone array configuration changes, crosstalk intensity should be taken into account together with

crosstalk delay time when the magnitude of crosstalk effect is considered.

Finally, it is known that certain changes in timbral characteristics are caused by the addition of reflection due to the comb-filter effect. However, it is not entirely clear from the literature which specific timbral attributes are affected by reflections in which conditions. Therefore, it is difficult to make a precise prediction about the effect of interchannel crosstalk on timbral attributes using the information provided on reflection effects. However, based on the finding that the timbre changing effect of reflection becomes most obvious when the range of delay time is 10-50ms [Barron 1971, Haas 1972], it might be hypothesised that an interchannel crosstalk, which would typically have a delay time in the range of a few milliseconds, would cause a smaller degree of timbral change than a reflection.

2.5 Summary

This chapter reviewed the studies relating to the effects of delayed secondary signals that have been conducted in the context of concert hall or room acoustics for the purpose of obtaining a useful basis for understanding the effect of interchannel crosstalk in multichannel stereophonic recording and reproduction. To summarise, the perceptual effects of such delayed secondary signals are related to attributes in three main categories comprising localisation, spatial impression and tone colour. While the tone colouration effect is known to be caused by the comb filtering, there seems to be no conclusive experimental data available about which timbral attributes

are directly related to this effect. It is suggested that the tone colouration depends on the spectral and temporal characteristics of a sound source.

Localisation of a sound in a reflective environment owes much to the precedence effect. When this effect operates, a reflection arriving at the listener's ears about 1ms after the direct sound is perceptually suppressed and the auditory image is localised constantly at the position of the direct sound. However, this is effective only up to the delay time of the upper (echo) threshold, which varies depending on the type of sound source and the direction of reflection. Beyond the echo threshold the reflection is perceived as a separate source, which is likely to disturb accurate localisation. It is widely found that low frequency transient energy is essential for triggering the precedence effect. The precedence effect is also found to involve a cognitive process of human perception and the evidence for this includes the experimental findings on the build-up process, the Clifton effect and the plausibility hypothesis.

The addition of reflections is also found to increase the perceived spatial impression. These days spatial impression (SI) is generally accepted to include at least two sub-attributes of apparent source width (ASW) and listener envelopment (LEV). A great deal of research has been carried out especially to develop parameters for objective measurement of SI. The objective parameters that have been mainly investigated include intensity and direction of reflection, frequency component of sound source, interarual cross-correlation coefficient (IACC) and interaural fluctuation over time. The first two are related to the physical properties of source signals, whereas the last

two are related to the binaural relationship of ear input signals. The lateral fraction theory suggests that perceived ASW increases as the intensity of lateral reflection increases. With regard to the effect of frequency component on perceived ASW, reports from different researchers do not totally match and this seems to be due to the use of different sound sources or the lack of standard definitions for terminologies. It is generally found that the perceived ASW increases when the IACC decreases or the magnitude of interarual fluctuation increases. It is also found that the IACC and the interaural fluctuation are related to each other in that the latter is the main factor affecting the former.

3 PERCEPTUAL ATTRIBUTES OF PHANTOM IMAGES IN 2-0 STEREOPHONIC SOUND REPRODUCTION

This chapter summarises subjective experiments carried out to examine the perceptual attributes of phantom images in 2-0 stereophonic sound reproduction. These experiments were designed from the following backgrounds. As Rumsey [2001] suggested, the perceived auditory attributes of phantom images created from the interference of interchannel crosstalk signals between the adjacent microphones in multichannel microphone arrays would be likely to depend on the combination of relevant time and intensity differences between the signals. It was found from the studies related to the localisation of phantom images in stereophonic reproduction, which were discussed in Chapter 1, that images created with pure interchannel time difference (ICTD) would be less easily or accurately localised than those with pure interchannel intensity difference (ICID) in general (i.e. spaced pair microphone techniques vs. coincident pair microphone techniques). To date, however, there seem to be no conclusive experimental results, of which the author is aware, which describe the specific kinds of attributes that can be perceived from stereophonic phantom images created with certain ICTD and ICID relationships. Neither is it clear how such attributes might be weighted perceptually. Therefore, it would be first necessary to understand the effect of ICTD and ICID on the perception of relevant attributes in two-channel format prior to the investigation of interchannel crosstalk in multichannel format. Furthermore, research conducted to investigate the perceptual effects of reflections, which was covered in Chapter 2, has suggested that the spectral and temporal characteristics of sound sources would be significant for accurate

localisation and perceived source width. From this, it became also of interest to see how different types of sound sources affect the perception of phantom images in stereophonic reproduction. In this study, the perceptual effects of the ICTD - ICID relationship and the type of sound source were investigated by comparing stereophonic phantom source images with referential monophonic sources, which were intended to be localised at the same position as the stereophonic images.

From the above backgrounds, the following research questions were formulated for investigation.

- What are the perceptual attributes of 2-0 stereophonic phantom images?
- Are the perceptions of these attributes significantly influenced by the type of panning method and the type of sound source?
- Do any of these attributes have correlations?

3.1 Experimental Hypotheses

In Section 2.2.1, it was mentioned that the auditory image created by the precedence effect, which operates in the perception of an original (direct) sound and its delayed (reflected) sound, would be perceived to be more spacious compared to that created by the original sound alone [Freyman *et al* 1991, Perrott *et al* 1988]. It was also stated by Blauert [1997] that the degree of such spatial distribution would become greater as the delay time increased. It was discussed in Chapter 2 that the magnitude of

perceived spatial impression could be determined by such objective measures as IACC and ITD fluctuations over time. As Mason [2002] states, in the context of stereophonic sound recording and reproduction, the magnitude of interaural fluctuations over time for the reproduced signals arriving at the ears could be determined by the combination ratio of ICTD and ICID. Furthermore, the comb-filter effect for the stereophonic signals, which would be likely to cause certain timbral differences between the stereophonic and monophonic images, might be dependent on the time and intensity relationship between the signals since it is a function of phase between two signals. From these, it was hypothesised that the perceived differences between stereophonic phantom source image and monophonic source would be perceived in both spatial and timbral attributes, and the magnitudes of those differences would significantly vary depending on the combination ratio of ICTD and ICID in the panning method used.

In addition, the temporal characteristics of the sound source have been found to be significant for accurate localisation in the literature reviewed in Section 2.2.2. It was discussed in Section 2.3.2.2 that in the presence of reflection the spectral characteristics of sound source would be important for the perception of source width. The spectral characteristics of the sound source would also be closely related to the timbre of the source. Therefore, it was predicted that the perceptions of phantom image attributes would be significantly affected by the temporal and spectral characteristics of the sound sources used in the current experiments.

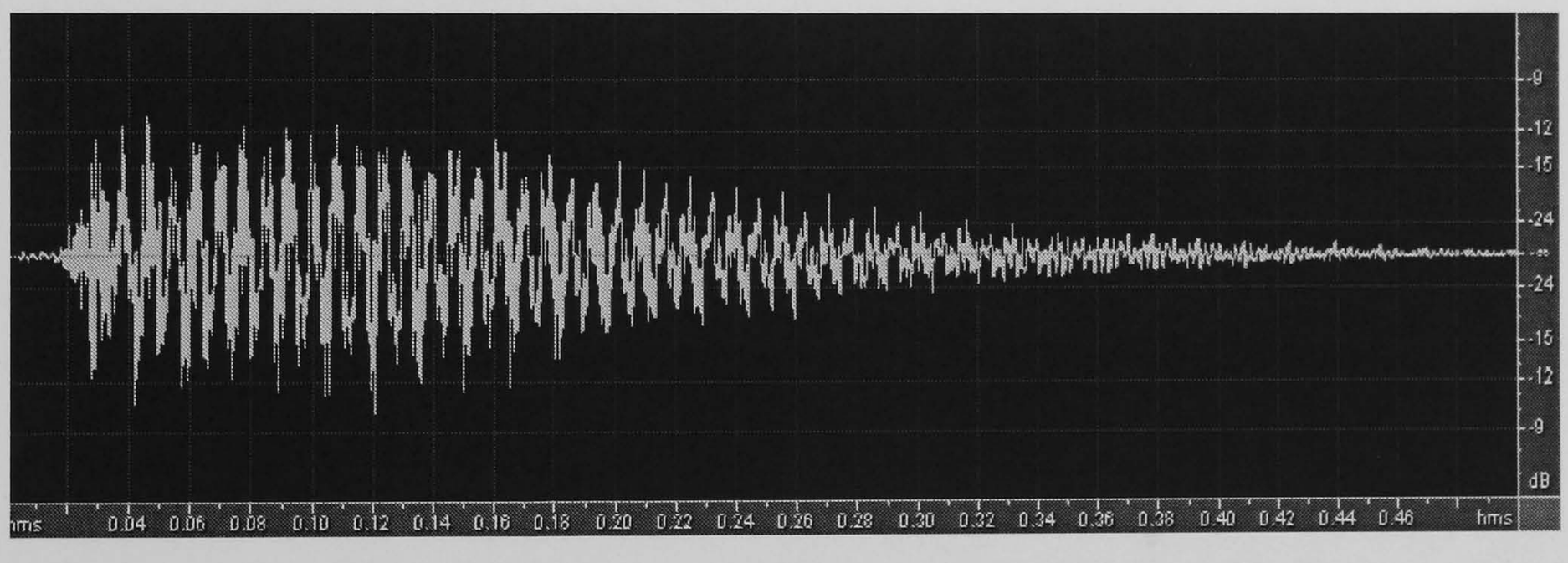
3.2 Experimental Design

3.2.1 General methodology

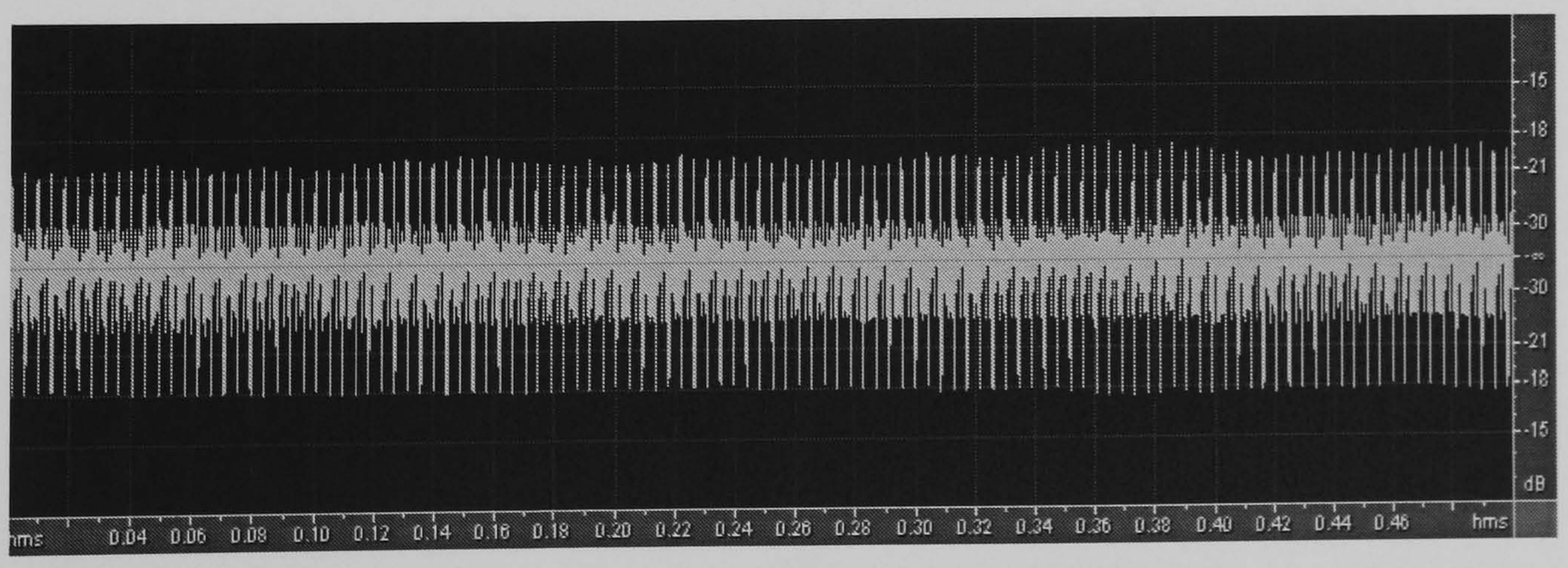
This investigation was inspired by the Quantitative Descriptive Analysis (QDA) method, which was originally developed for the evaluation of sensory attributes of products. The original QDA basically consists of three stages: elicitation, grouping analysis and grading [Bech 1999]. Firstly, a group of qualified subjects are presented with stimuli and generate descriptive terms for the attributes of the product through discussion. Secondly, the elicited terms are grouped into a limited number of attributes through discussion based on the similarity of meaning. Finally, the stimuli are graded using the obtained scales. This method is particularly suitable for investigating undeveloped areas in that the subjects are actively involved in choosing the relevant attributes to be graded. As Kjeldsen [1998] and Berg and Rumsey [1999] point out, the use of 'provided' attribute scales has a significant limitation in this kind of sound quality evaluation in that the subjects would be restricted to respond only in the experimenter's own terms even if they found other relevant attributes for evaluation. To make this investigation more effective in terms of time, the original QDA was modified. Instead of undertaking the grouping analysis with all the subjects involved in the discussion, the elicited terms were interpreted and grouped by the experimenter through informal discussions with individual subjects on the meanings of the terms they used. Therefore, the whole investigation consisted of two subjective experiments, namely elicitation and grading phases.

3.2.2 Creation of stimuli

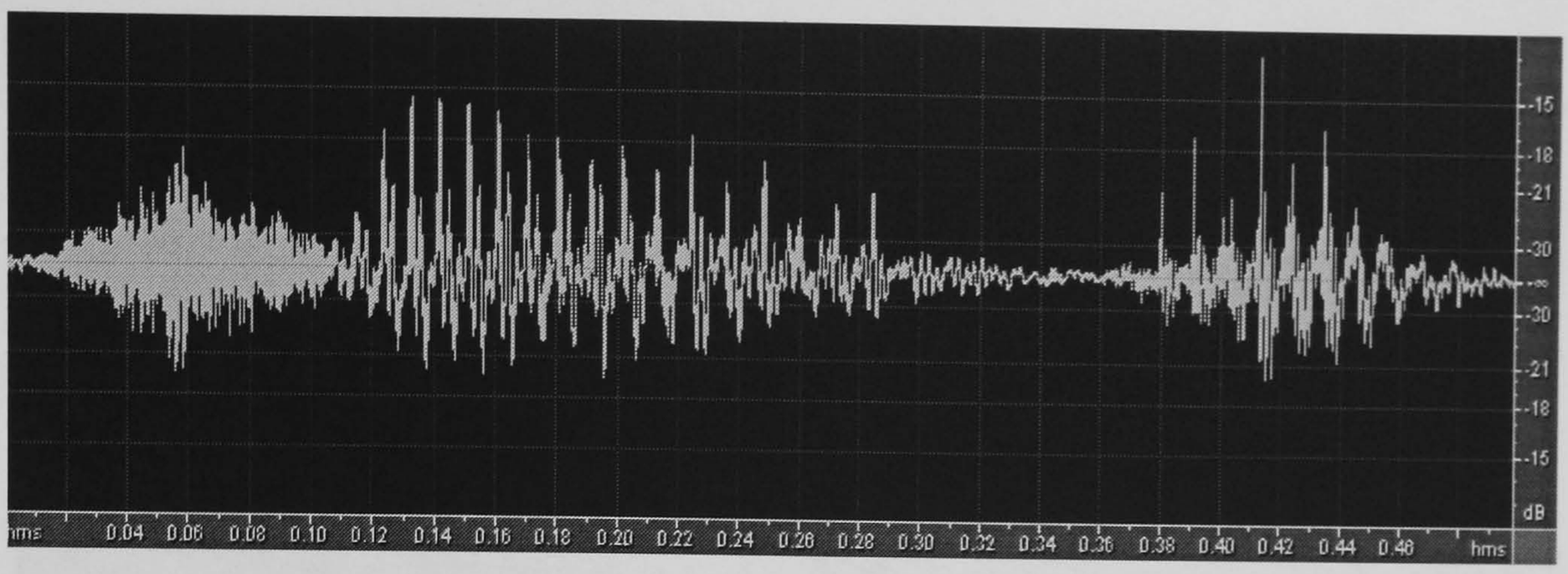
For the experiments three types of sound source were chosen, comprising piano 'staccato' note of C3 (f_o = 130 Hz), trumpet 'sustain' note of B flat3 (f_o = 228 Hz) and male speech dialogue. The piano and trumpet sources were chosen in order to examine the perceived effects that might change depending on the different temporal characteristics of musical instruments, i.e. transient and continuous characteristics (staccato vs. sustain). The short term extracted waveform for each sound source is shown in **Figure 3.1**.



(a) Transient piano note



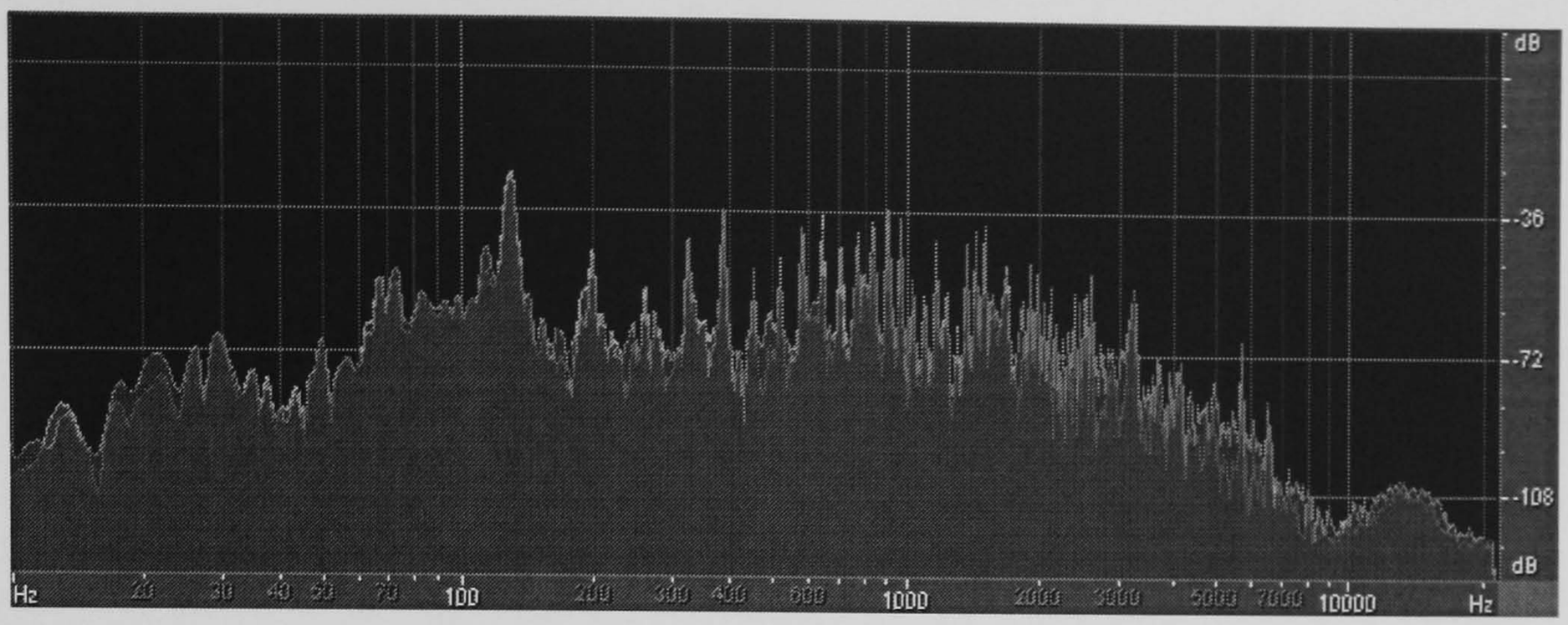
(b) Continuous trumpet note



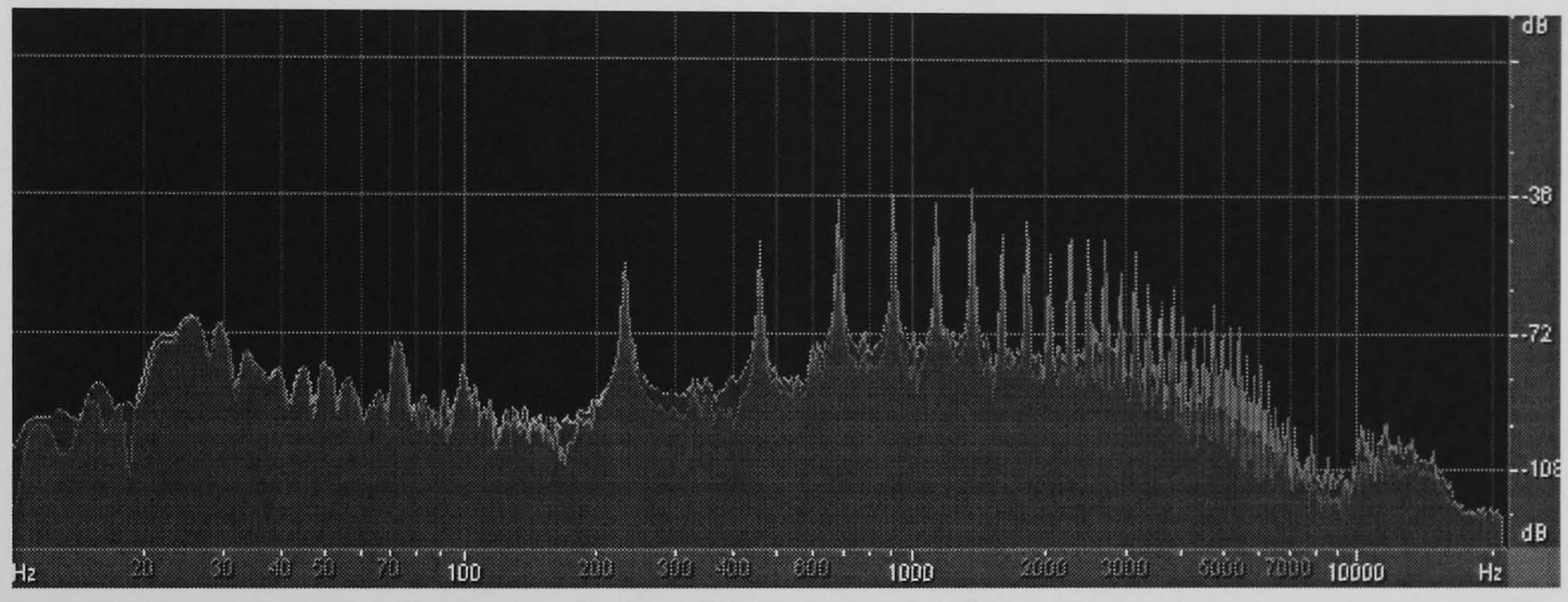
(c) Continuous male speech dialogue

Figure 3.1 Short term extracts of waveforms for each sound source

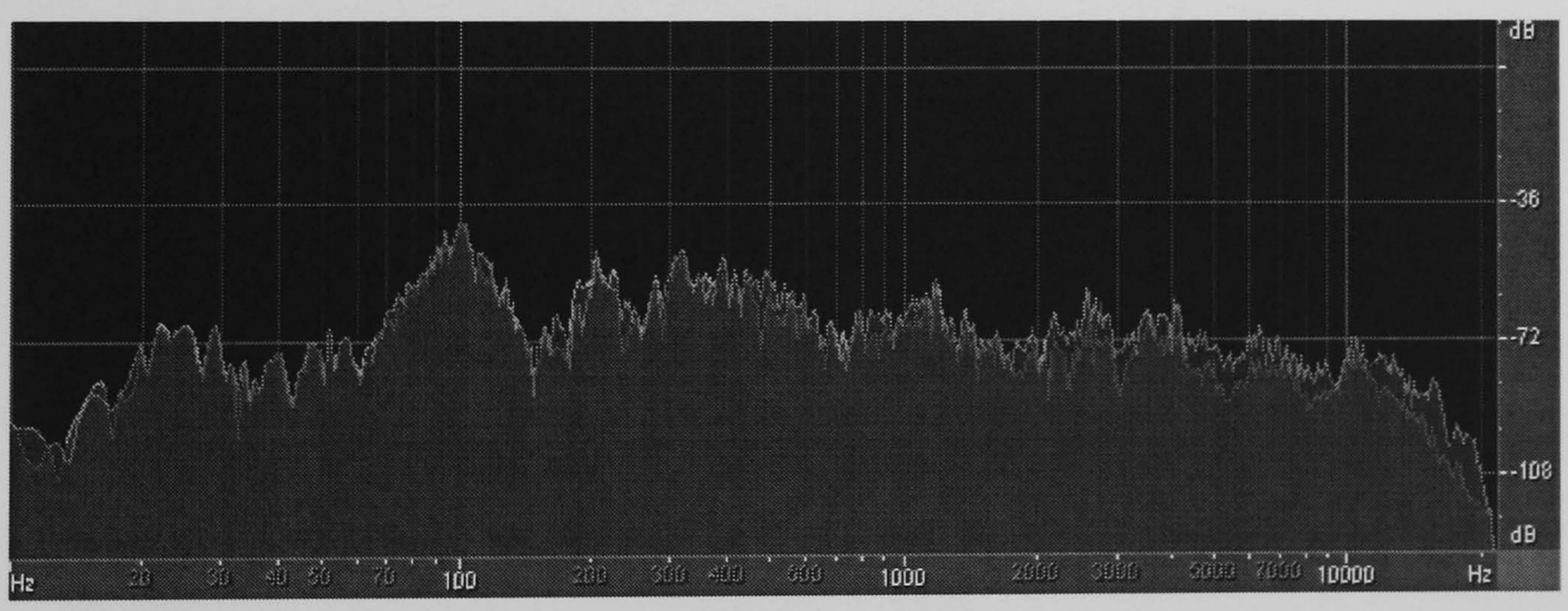
Single notes of those sources were used instead of musical extracts in order to limit the variables strictly within the experimental scope. The piano source was recorded using a single cardioid microphone placed about 30cm over the hammers for the desired note. The piano was completely covered with thick cloth in order to reduce unwanted acoustic effects as much as possible. The trumpet sources were recorded in a small overdub booth of Studio 3 of the University of Surrey, using a single cardioid microphone placed about 1m away from the instrument. The recording space was acoustically isolated, had a very low reverberation time and was almost anechoic. In order to investigate the continuous nature of the trumpet strictly, the onset and offset transients of the trumpet source were removed by fading in and out the beginning and ending for one second each, making the total duration of the stimulus four seconds. The speech signal was chosen because it has a combination of both transient and continuous characteristics as well as a wide range of frequencies. The speech recording used was Danish male speech that was anechoically recorded for Bang and Olufsen's Archimedes project [Hansen and Munch 1991]. Additionally, each sound source differs in spectral characteristics, as can be seen in **Figure 3.2**.



(a) Piano note C3



(b) Trumpet note B3



(c) Male speech dialogue

Figure 3.2 Long-term averaged frequency spectrum of each sound source

For each sound source, one monophonic stimulus and three stereophonic stimuli were created using three different panning methods of time, intensity and a combination of the two. The loudness of the stereophonic stimuli was naturally greater than that of the monophonic ones simply due to the number of loudspeakers used. Therefore, in order to enable the subject to judge differences other than loudness, the peak sound pressure levels of all stimuli were calibrated at 75dBA. From an informal test that had been conducted before the main experiments, it was recognised that the choice of panning angle had a very small effect on the perceived attributes. However, the test angle was fixed at 20° since this angle was considered to provide a reasonably balanced combination of ICTD and ICID. The interchannel time and intensity differences required for localising the sound image at 20° were calculated based on a combination function developed by the author using the psychoacoustic values that were obtained from a localisation experiment conducted using the same types of sound sources. The details of the localisation experiment and the development of the combination function are described in Appendix A. The composition of the test stimuli is shown in **Table 3.1**.

	Time Panning	Combination panning	Intensity panning
Speech	0.5ms	0.25ms + 4dB	8dB
Piano			
Trumpet			

Table 3.1 Composition of the test stimuli, showing interchannel time and intensity differences: a total of nine stimuli were produced using these different panning methods

3.2.3 Physical setup

The experiment was conducted in an ITU-R [1994] BS.1116-compliant listening room at the University of Surrey. The physical setup of the listening room is shown in

Figure 3.3. Two Genelec 1032A loudspeakers L and R were set up at 60° from the listening position and 3m apart. The reference loudspeaker was placed at the 20° position so that its auditory image would appear at the same (or as similar as possible) direction as that of the phantom image created by L and R. An acoustically transparent curtain was used in order to hide the nature of the experiment from the listener.

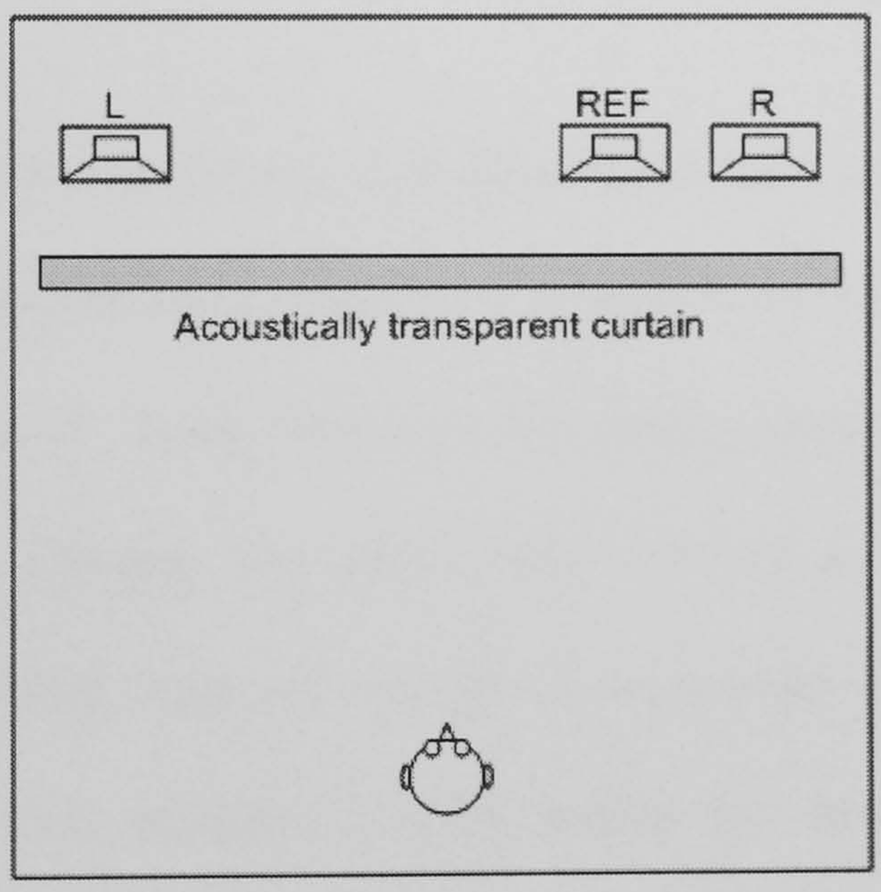


Figure 3.3 Physical setup of the listening room

3.2.4 Subjects

A total of eight subjects participated in the test. All were experienced in spatial listening, being selected from staff members, doctoral students and final year undergraduate students on the University of Surrey's Tonmeister course. For a

subjective experiment such as a preference test, a large number of naïve subjects are often used. However, this was considered to be unsuitable for the current experiments since the nature of the listening test required subjects' critical listening skills to discriminate subtle differences between stimuli and therefore naïve subjects would be likely to provide inconsistent data.

3.3 Experiment Part 1: Elicitation of Perceptual Attributes

3.3.1 Listening test method

This experiment was designed such that the subjects were provided with two sound stimuli 'A' and 'B' and asked to complete a statement written as 'Stimulus B is ___ compared to stimulus A', using their own descriptive terms. The control interface was designed using Cycling 74's MAX-MSP software as shown in **Figure 3.4**. There were a total of nine trials and their presentation order was randomised for each subject. In every trial, stimulus 'B' represented the stereophonic stimulus and stimulus 'A' was the corresponding monophonic stimulus. The stereophonic signals of stimulus A were fed into the loudspeakers L and R while the monophonic signal of stimulus B was fed into the reference loudspeaker. The stimulus pair of A and B was synchronised and looped so that the subjects were able to switch between them freely and to listen repeatedly. The subjects were allowed to spend as much time as they wanted in order to find all the audible differences. The natures of the stimuli were veiled to the listener.

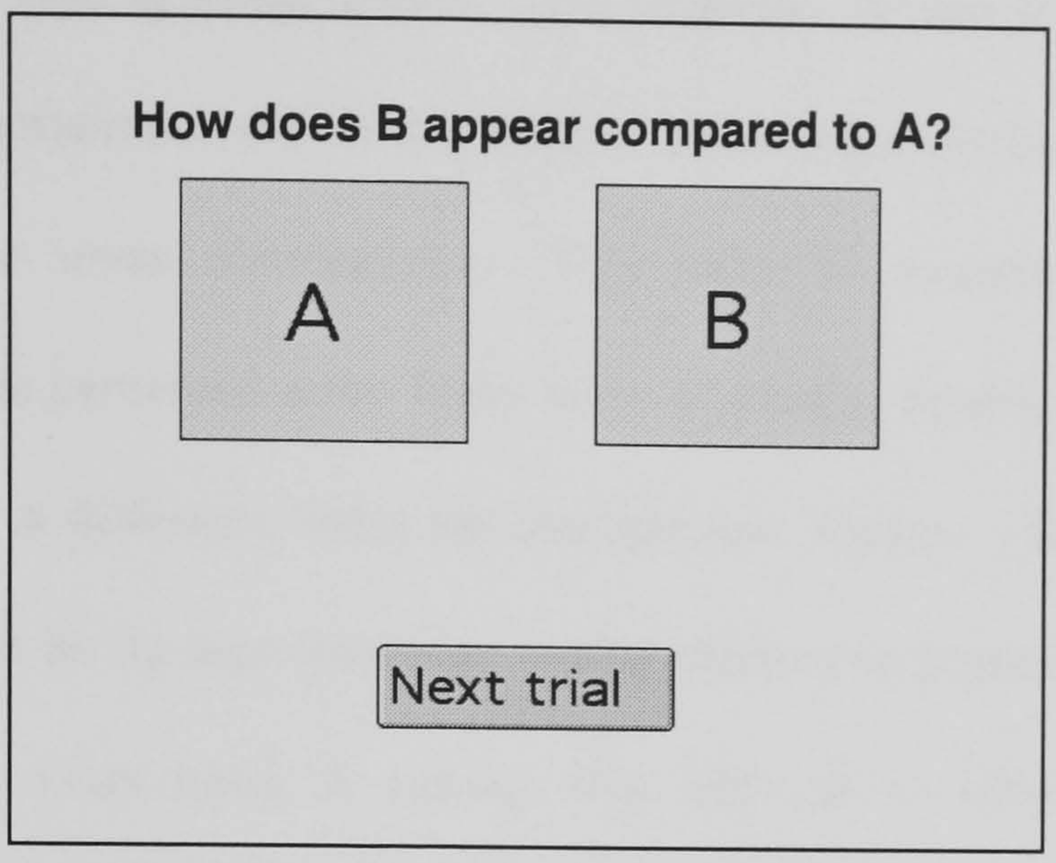


Figure 3.4 Layout of the control interface used for comparing mono and phantom images

3.3.2 Results and discussions

A number of descriptive terms were elicited from the subjects and the interpretation and grouping analysis of the terms were carried out by the author, with informal discussions with the individual subjects on the meanings of some unclear terms. The results are summarised in **Tables 3.2** and **3.3**. Firstly it was possible to separate the terms into two broad groups: spatial and timbral attributes. The individual terms were then separated into six sub-groups based on the similarity in meaning: source focus, source width, source distance, brightness, hardness and fullness. The definitions given for these attributes are listed in **Table 3.4**. The number in brackets that can be seen in **Tables 3.2** and **3.3** represents the number of occurrences for each specific term. It is interesting to see from the tables that every sound source had the same types of perceived spatial and timbral attributes. However, for the spatial attributes, it appears that the total number of occurrences for each attribute varies

depending on the type of sound source. For example, for the trumpet source there was just one observation of the source width attribute whereas for speech and piano there were six or seven observations. This might be because the monophonic trumpet image was perceived to be fairly wide originally, making it difficult for the subjects to detect a difference from the stereophonic images. The source distance attribute appears to be the least dominant spatial attribute in general. For the timbral attributes, on the other hand, it appears that there is no obvious sound source dependency for any of attributes. It is also seen that hardness is the least salient attribute.

The spatial differences are likely to be due to the difference in the degree of interaural cross-correlation or fluctuation in interaural time and intensity differences. In Chapter 2 it was explained that the addition of one or more reflections would decrease the degree of interaural cross-correlation or increase that of fluctuations in ITD and IID, leading to increased spatial impression. Similarly, the degree of interarual cross-correlation for sounds radiated from two loudspeakers with a certain difference in time and intensity would be likely to be higher than that for a sound from a single loudspeaker. The explanation of timbral differences also seems to be found in the reflection studies. It was mentioned in Chapter 2 that the interference between a direct sound and its delayed reflection produces a comb-filter effect. Similarly, the summation of leading and lagging sounds in stereophonic reproduction is likely to cause comb-filtering when the ICTD and ICID are transmitted to the ears with acoustic crosstalk.

Sound source	Spatial Attributes	
	Group	Descriptive terms
SPEECH	*Source focus*	Less localised (4) Less focused (2) Less present (1) Less stable (1) Less Coherent (1)
	Source width	Wider (7)
	Source distance	More distant (1) Further away (1)
PIANO	*Source focus*	Harder to locate (2) Less defined (2) Less focused (1)
	Source width	Wider (6)
	Source distance	More distant (1) Closer (1) More reverberant (2)
TRUMPET	*Source focus*	Harder to locate (2) Less focused (2) Less solid (1) More diffused (1)
	Source width	Wider (1)
	Source distance	More distant (1) Further away (1) Closer (1)

Table 3.2 Summary of spatial attributes drawn from the elicited descriptive terms

Sound source	Timbral Attributes	
	Group	Descriptive terms
SPEECH	*Brightness*	Less bright (2) More cloudy (1) Duller (1) Muddier (1) Less breathy (1)
	Hardness	Softer (1)
	Fullness	Fuller (1) Bassier (1) Less bassy (1) Less body (1)
PIANO	*Brightness*	Brighter (1) Duller (2) Less dark (1) Less bright (1) Less toppy (1) Less harsh (1)
	Hardness	Softer (1) Less attack (2)
	Fullness	Less bassy (1) Less punch (1) Bassier (1) Fuller (1)
TRUMPET	*Brightness*	Brighter (3) Duller (1) More present (1) More nasal (1)
	Hardness	Stronger (1) Harsher (1)
	Fullness	Fuller (1) Less bassy (2)

Table 3.3 Summary of timbral attributes drawn from the elicited descriptive terms

Source focus	The easiness of localisation of a sound source i.e. How easy is it to pinpoint the apparent location of a source?
Source width	The perceived width of a sound source itself i.e. Is one source perceived to be wider than the other?
Source distance	The perceived distance from the listener to a sound source i.e. Can the sources be discriminated in terms of their distances?
Brightness	The timbral characteristics of a sound depending on the level of high frequencies i.e. bright / dull
Hardness	The timbral characteristics of a sound depending on the level of mid-high frequencies (typically in the range of 2 – 4kHz) i.e. hard / soft
Fullness	The timbral characteristics of a sound depending on the level of low frequencies i.e. full / thin

Table 3.4 Definitions of the attributes that were grouped from the elicited subjective terms

3.4 Experiment Part 2: Grading of the Magnitude of Perceptual Effect

3.4.1 Listening test method

Based on the attributes that were derived from the previous experiment, the magnitudes of the perceived differences between the stereophonic and the monophonic stimuli were graded. The listening test was designed so that for each sound source type the subjects compared each of the three stereophonic stimuli created using three different panning methods with the reference monophonic stimulus. The control interface used for this test is shown in **Figure 3.5**. In order to obtain sufficient data for statistical analysis, the trial for each type of sound source was

repeated twice with the order of stimulus presentation randomised. Therefore, there were six trials to be tested in total. For each trial the subjects were asked to grade the magnitudes of the perceived differences between the monophonic stimulus REF and each stereophonic stimulus A, B and C on an 11-point continuous grading scale for each attribute, labelled from –5 to 5.

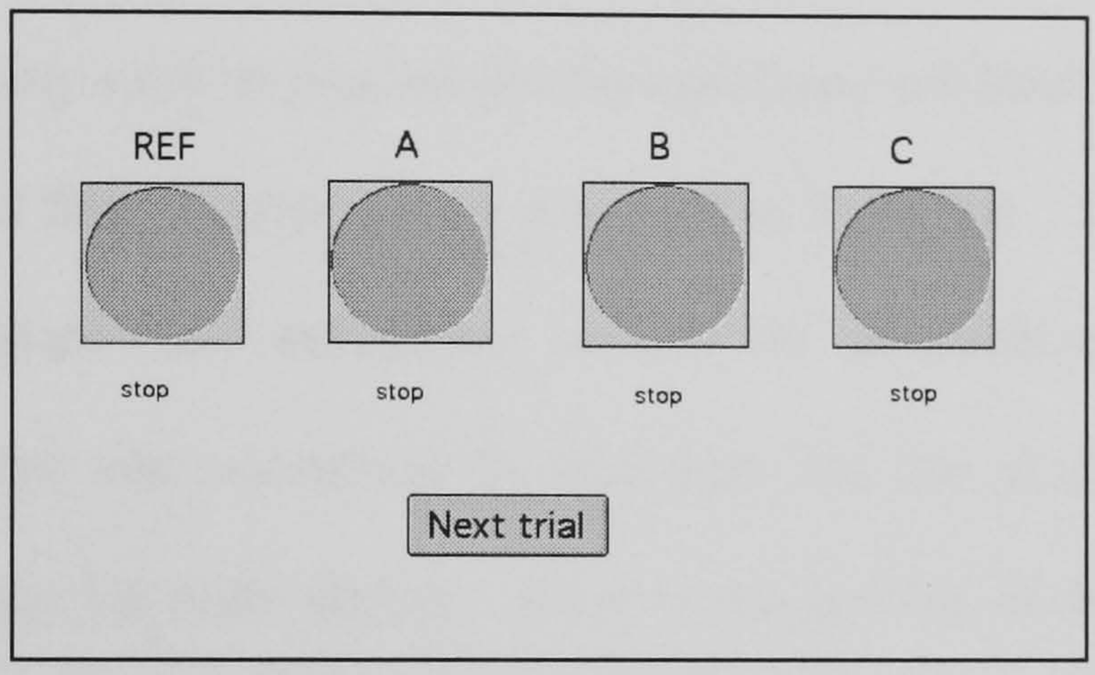


Figure 3.5 Layout of the control interface used in the grading test

The choice of scale type was influenced by the following considerations. It was thought that using a semantic differential scale with word labels would not be appropriate for this experiment for the following two reasons. Firstly, the potentially nonlinear nature of the scale would not be ideal for parametric statistical analysis. Secondly, the meanings of the labels might be differently interpreted by different subjects. This is likely to be particularly true for an attribute such as source width because it would be difficult for subjects to define the meanings of such labels as 'much wider' and 'slightly wider' in the same way. With this in mind, using a continuous grading scale was considered to be a more appropriate method since the data would be potentially more reliable for parametric statistical analysis due to the linearity of the scale, although the data would need to be normalised before statistical

analysis. However, using a pure continuous grading scale without any labels, the subjects might have difficulties in maintaining consistency in testing through many trials individually. Therefore, numerical labels were added to a classical continuous rating scale as guidelines for helping subject consistency.

This type of subjective experiment, for investigating fine perceptual differences, would typically carry a risk of psychological errors [Stone and Sidel 1993]. The list of such errors and their descriptions are presented in **Table 3.5**. In order to avoid contrast, convergence and anticipation errors, the presentation order of the stereophonic stimuli was randomised for each trial, and that of the trial was also arranged differently for each subject. Prior to the grading of the magnitude of perceived difference against the reference stimulus, the subjects were instructed to familiarise themselves with the differences between the stereophonic stimuli first. This was in order to avoid central tendency and time-order errors.

Psychological error	**Description**
Central tendency error	Subjects tend to use the midrange of a scale, avoiding the extremes, especially when they are unfamiliar with the stimuli or a test method.
Time-order error	Subjects tend to give the first product a higher score than expected.
Contrast error	The difference between two stimuli is exaggerated, occurring when a 'smaller' stimulus is followed by a 'larger' stimulus, and vice versa.
Convergence error	The difference between stimuli is underestimated, occurring when a few relatively small stimuli are compared with a distinctively larger stimulus.
Anticipation error	Occurs when the subjects can anticipate the pattern of systematic changes in a series of stimuli.
Logical error	Occurs when the subjects are not precisely instructed. The subject follows a logical but self-determined process in evaluating stimuli.
Proximity error	Adjacent characteristics tend to be rated more similar than those that are farther apart. Thus the correlations between adjacent pairs may be higher.

Table 3.5 Potential psychological errors to be considered in subjective listening test and their descriptions, based on Stone and Sidel [1993]

3.4.2 Statistical analysis

The grading experiment was designed so that all conditions were tested within the same group of subjects. Therefore, a repeated measure ANOVA (RM ANOVA) test was performed for statistical analysis of the data obtained from the grading experiment. The independent variables were the panning method and the sound source, and the dependent variable was the grading data. Because of the nature of the scale used, it

was predicted that each subject would use a different range of the scale. This problem of subject variability in use of the scale might cause inaccurate results from statistical analysis. Therefore, the original data were normalised based on the ITU-R BS.1116 Recommendation [1994] and the equation used for this is shown below.

$Zi = Xi - Xsi + Xs$ where Zi = normalised results

Xi = score of subject i

Xsi = mean score of subject i in session s

Xs = mean score of all subjects in session s

There were a total of 144 observations, consisting of 16 observations for each of the 9 'sound source type–panning method' combinations obtained from 8 subjects. The result of the RM ANOVA test for each attribute is presented in the following sections.

In the presentation of the results, each independent variable is termed 'source' and 'panning' for convenience. In order to interpret the results of the RM ANOVA test correctly, it was necessary to examine the 'assumption of sphericity' (equal variances of the differences between conditions) by using Mauchly's test of sphericity. An insignificant statistic of Mauchly's test ($p>0.05$) means that the variances of the data for each condition compared are not significantly different, and thus the assumption of sphericity is met. In this case, the 'sphericity assumed' significance value should be used as a result of the RM ANOVA. However, if Mauchly's test statistic is significant ($p<0.05$), the assumption of sphericity is violated and one of the corrected significance values should be used instead of the sphericity assumed one. The result of Mauchly's test for each attribute is presented in the following section.

3.4.3 Results

3.4.3.1 Source focus

Table 3.6 shows the results of the RM ANOVA test for the grading data obtained for the 'source focus' attribute. The significance value p for each condition 'sound source' and 'panning method' was determined according to the results of Mauchly's test of sphericity presented in **Table 3.7**, as explained in the above section. The results indicate that both sound source (p = 0.013) and panning method (p = 0.000) had highly significant effects on source focus difference between the stereophonic and monophonic images. The experimental effect size, which can be estimated from the Partial Eta Squared value, was greater for panning method (0.750) than for sound source (0.129). With respect to the interaction between each factor, it is shown that the effect was also significant (p = 0.016).

Measure: MEASURE_1

Source		F	Sig.	Partial Eta Squared
SOURCE	Sphericity Assumed	4.684	.013	.129
	Greenhouse-Geisser	.4.684	**.013**	.129
	Huynh-Feldt	4.684	.013	.129
	Lower-bound	4.684	.013	.129
PANNING	Sphericity Assumed	20.975	**.000**	.750
	Greenhouse-Geisser	20.975	.000	.750
	Huynh-Feldt	20.975	.000	.750
	Lower-bound	20.975	.003	.750
SOURCE * PANNING	Sphericity Assumed	3.675	**.016**	.344
	Greenhouse-Geisser	3.675	.047	.344
	Huynh-Feldt	3.675	.026	.344
	Lower-bound	3.675	.097	.344

Table 3.6 Result table of repeated measure ANOVA test for the data obtained for 'source focus' difference between stereophonic and monophonic stimuli

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.
SOURCE	.299	7.240	2	.027
PANNING	.640	2.682	2	.262
SOURCE * PANNING	.061	15.151	9	.100

Table 3.7 Result table of Mauchly's test of sphericity for the data obtained for 'source focus' difference between stereophonic and monophonic stimuli

Figure 3.6 shows the mean values and 95% confidence intervals for each sound source and each panning method. It initially shows that the stereophonic images for every sound source and panning method were perceived to be 'less focused' than the monophonic image. From the plots of sound source, it can be seen that the speech source had the greatest effect, followed by piano and trumpet sources in order. It can also be seen from the plots of panning method that pure time panning caused the greatest difference and pure intensity panning the smallest difference. In addition, the magnitude of difference appears to decrease almost linearly as the panning method moves from time to intensity. **Table 3.8** presents the results of pairwise comparisons between each sound source and between each panning method. From these results, it can be confirmed that the significance of the sound source effect was caused by the significant difference between speech and trumpet ($p = 0.010$).

Measure: MEASURE_1

(I) SOURCE	(J) SOURCE	Mean Difference (I-J)	Std. Error	Sig.
speech	piano	-.419	.232	.228
	trumpet	-.707	.232	.010
piano	speech	.419	.232	.228
	trumpet	-.288	.232	.660
trumpet	speech	.707	.232	.010
	piano	.288	.232	.660

Measure: MEASURE_1

(I) PANNING	(J) PANNING	Mean Difference (I-J)	Std. Error	Sig.
Time	Combi	-1.104	.290	.020
	Intensity	-2.396	.467	.004
Combi	Time	1.104	.290	.020
	Intensity	-1.292	.330	.017
Intensity	Time	2.396	.467	.004
	Combi	1.292	.330	.017

Table 3.8 Result tables of pairwise comparisons between each sound source and between each panning method for 'source focus' attribute

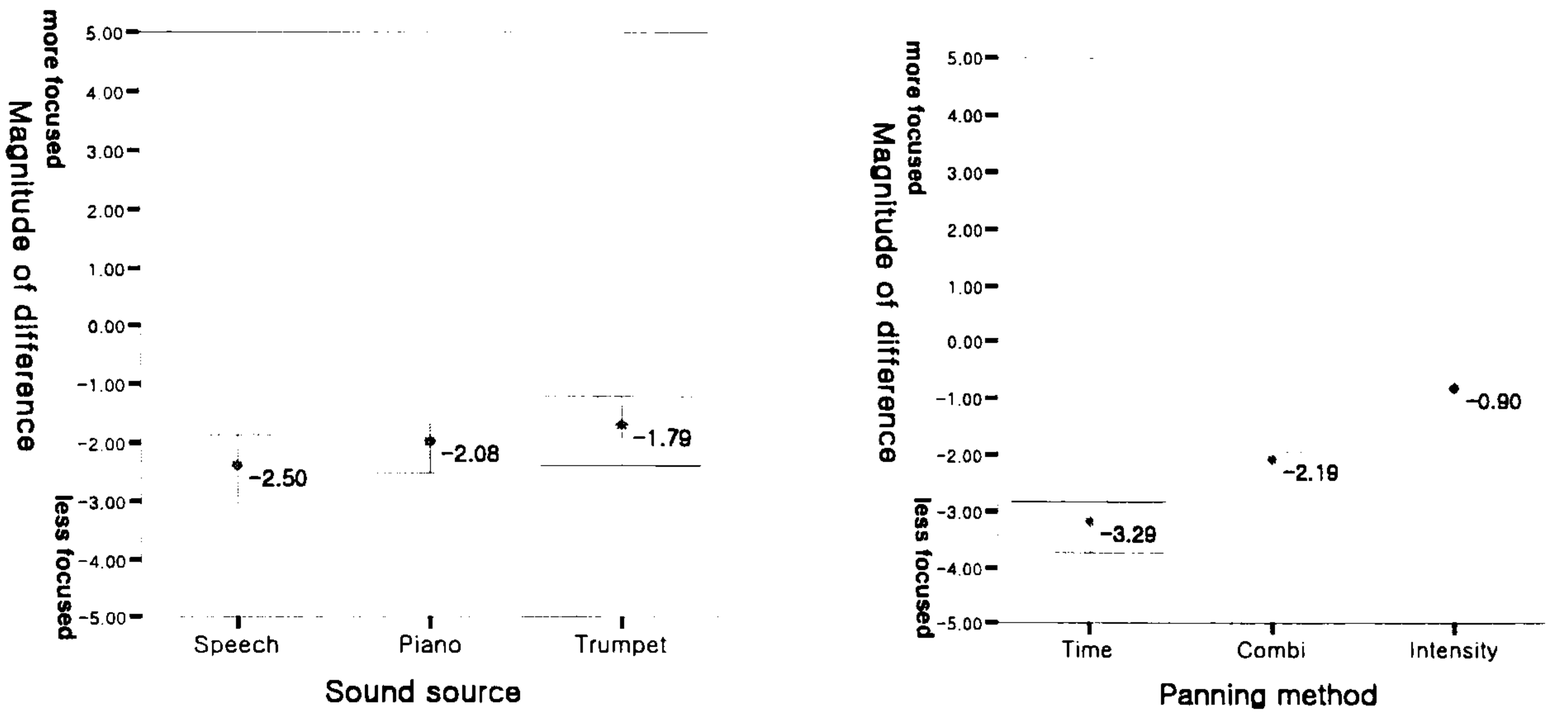


Figure 3.6 Mean values and the associated 95% confidence intervals of the grading data of 'source focus' difference between stereophonic and monophonic stimuli by sound source and panning method

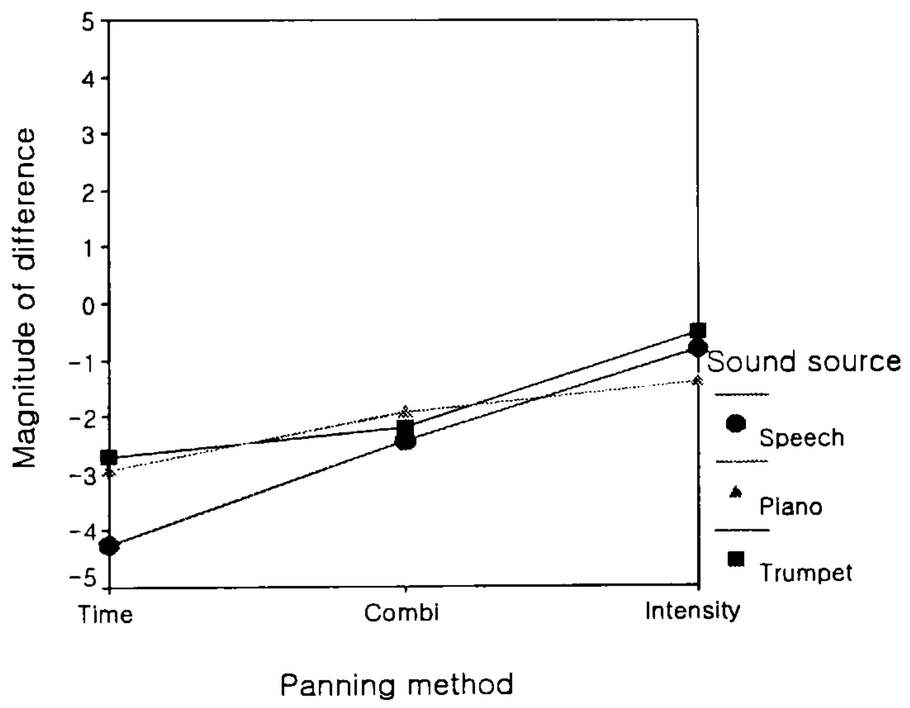


Figure 3.7 Interaction between panning method and sound source for source focus attribute

Plots of the source*panning interaction are shown in **Figure 3.7**. In order to examine the significance of the sound source effect for each panning method shown in the plots, a paired samples T-test was carried out. The results shown in **Table 3.9** indicate that for time panning the effect of the speech source was significantly greater compared to that of the piano or trumpet source. The effects of the piano and trumpet sources did

not have any significant difference for the time panning. For combination panning, there was no significant difference between any sound sources. Intensity panning gave rise to a significant difference between the piano source and the speech or trumpet source, while the difference between the speech and trumpet sources was insignificant.

		t	Sig. (2-tailed)
Time	Speech*Piano	-3.815	.007
Time	Speech*Trumpet	-4.074	.005
Time	Trumpet*Piano	-1.005	.348
Combi	Speech*Piano	-1.664	.140
Combi	Speech*Trumpet	-1.615	.150
Combi	Trumpet*Piano	.818	.440
Intensity	Speech*Piano	1.521	.172
Intensity	Speech*Trumpet	-.667	.526
Intensity	Trumpet*Piano	-3.487	.010

Table 3.9 Result table of paired samples T-test carried out for the interaction effect of sound source and panning method for source focus attribute

3.4.3.2 Source width

The results of the RM ANOVA test for the grading data obtained for the 'source width' attribute are shown in **Table 3.10**, and the results of Mauchly's test of sphericity are shown in **Table 3.11**. The results indicate that the effects of sound source (p = 0.009) and panning method (p = 0.000) on the source width difference between stereophonic and monophonic images were highly significant, although panning method had a greater experimental effect (Partial Eta Squared value = 0.742) than sound source (0.138). The source*panning interaction is shown to be insignificant.

Measure: MEASURE_1

Source		F	Sig.	Partial Eta Squared
SOURCE	Sphericity Assumed	5.037	.009	.138
	Greenhouse-Geisser	5.037	.009	.138
	Huynh-Feldt	5.037	.009	.138
	Lower-bound	5.037	.009	.138
PANNING	Sphericity Assumed	20.100	.000	.742
	Greenhouse-Geisser	20.100	.000	.742
	Huynh-Feldt	20.100	.000	.742
	Lower-bound	20.100	.003	.742
SOURCE * PANNING	Sphericity Assumed	1.993	.123	.222
	Greenhouse-Geisser	1.993	.185	.222
	Huynh-Feldt	1.993	.174	.222
	Lower-bound	1.993	.201	.222

Table 3.10 Result table of repeated measure ANOVA test for the data obtained for 'source width' difference between stereophonic and monophonic stimuli

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.
SOURCE	.007	29.866	2	.000
PANNING	.666	2.443	2	.295
SOURCE * PANNING	.015	22.880	9	.009

Table 3.11 Result table of Mauchly's test of sphericity for the data obtained for 'source width' difference between stereophonic and monophonic stimuli

Figure 3.8 presents the mean values and 95% confidence intervals for each sound source and each panning method. It can be firstly seen that the stereophonic images were perceived to be 'wider' than the monophonic image for every sound source and panning method. Similarly to the results for the 'source focus' attribute shown above, the speech source appears to have the greatest effect and trumpet the smallest effect. Also the magnitude of effect appears to increase linearly as the panning method moves from intensity to time. From the results of pairwise comparisons between each sound source shown in **Table 3.12**, it can be observed that the difference in the speech and piano pair was insignificant while that in the other pairs was significant. It can be also observed that every pair of panning methods had a significant difference.

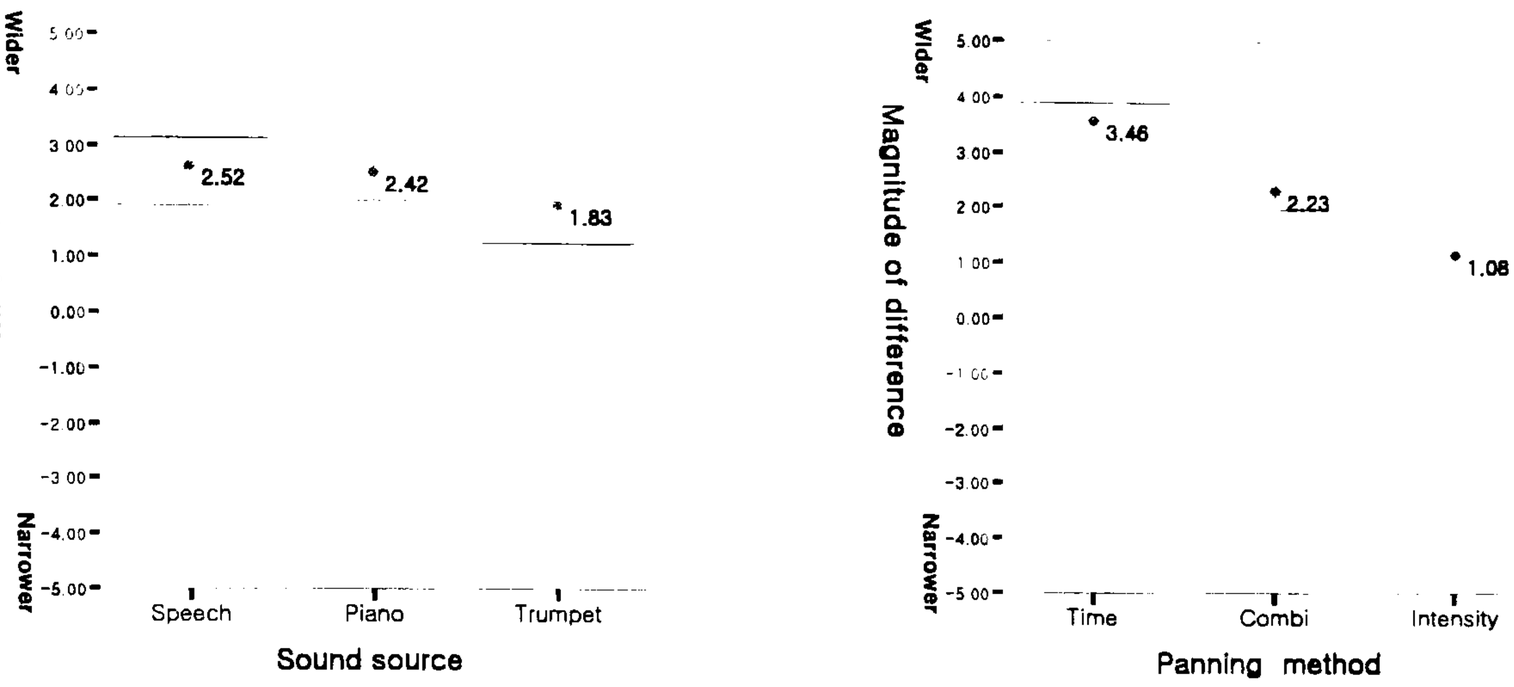


Figure 3.8 Mean values and the associated 95% confidence intervals of the grading data of 'source width' difference between stereophonic and monophonic stimuli by sound source and panning method

Measure: MEASURE_1

(I) SOURCE	(J) SOURCE	Mean Difference (I-J)	Std. Error	Sig.
Speech	Piano	.103	.234	1.000
	Trumpet	.688	.234	.014
Piano	Speech	-.103	.234	1.000
	Trumpet	.585	.234	.045
Trumpet	Speech	-.688	.234	.014
	Piano	-.585	.234	.045

Measure: MEASURE_1

(I) PANNING	(J) PANNING	Mean Difference (I-J)	Std. Error	Sig.
Time	Combi	1.201	.281	.011
	Intensity	2.347	.461	.004
Combi	Time	-1.201	.281	.011
	Intensity	1.146	.346	.039
Intensity	Time	-2.347	.461	.004
	Combi	-1.146	.346	.039

Table 3.12 Result tables of pairwise comparisons between each sound source and between each panning method for 'source width' attribute

3.4.3.3 Source distance

Table 3.13 shows the results of the RM ANOVA test for the grading data obtained for the 'source distance' attribute, and **Table 3.14** shows the results of Mauchly's test of sphericity. There was no significant effect for either sound source (p = 0.510) or panning method (p = 0.417) and the source*panning interaction effect is also shown to be insignificant (p = 0.532).

Measure: MEASURE_1

Source		F	Sig.	Partial Eta Squared
SOURCE	Sphericity Assumed	.681	.510	.021
	Greenhouse-Geisser	.681	.510	.021
	Huynh-Feldt	.681	.510	.021
	Lower-bound	.681	.510	.021
PANNING	Sphericity Assumed	.930	.417	.117
	Greenhouse-Geisser	.930	.395	.117
	Huynh-Feldt	.930	.408	.117
	Lower-bound	.930	.367	.117
SOURCE * PANNING	Sphericity Assumed	.624	.649	.082
	Greenhouse-Geisser	.624	.532	.082
	Huynh-Feldt	.624	.570	.082
	Lower-bound	.624	.455	.082

Table 3.13 Result table of repeated measure ANOVA test for the data obtained for 'source distance' difference between stereophonic and monophonic stimuli

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.
SOURCE	.978	.134	2	.935
PANNING	.625	2.820	2	.244
SOURCE * PANNING	.026	19.790	9	.024

Table 3.14 Result table of Mauchly's test of sphericity for the data obtained for 'source distance' difference between stereophonic and monophonic stimuli

Figure 3.9 shows the mean values and 95% confidence intervals for each sound source and each panning method. From the plots the magnitudes of the effects of both panning method and sound source do not appear to be considerable, although the stereophonic images appear to be 'more distant' than the monophonic image in all conditions. It is also indicated that none of the panning methods had significant differences between each other.

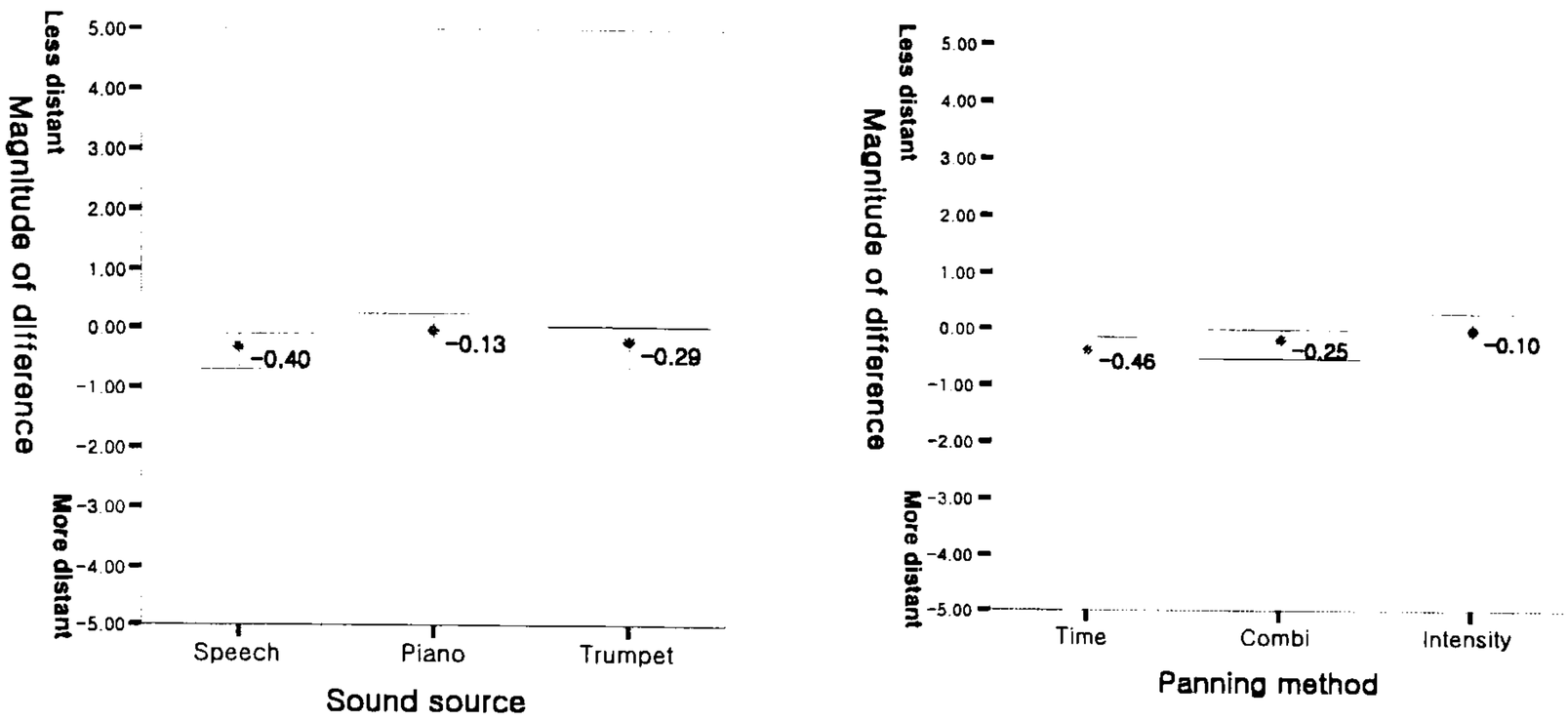


Figure 3.9 Mean values and the associated 95% confidence intervals of the grading data of 'source distance' difference between stereophonic and monophonic stimuli by sound source and panning method

3.4.3.4 Brightness

The results of the RM ANOVA test for the grading data obtained for the 'brightness' attribute are shown in **Table 3.15**, and the results of Mauchly's test of sphericity are shown in **Table 3.16**. The results indicate that sound source (p = 0.007) had a significant effect on the difference in brightness attribute while panning method did not (p = 0.419). However, the estimated effect size of sound source appears to be small (Partial Eta Squared value = 0.289). The source*panning interaction effect is shown to be insignificant (p = 0.667).

Measure: MEASURE_1

Source		F	Sig.	Partial Eta Squared
SOURCE	Sphericity Assumed	14.596	.000	.289
	Greenhouse-Geisser	14.596	.007	.289
	Huynh-Feldt	14.596	.007	.289
	Lower-bound	14.596	.007	.289
PANNING	Sphericity Assumed	.785	.475	.101
	Greenhouse-Geisser	.785	.419	.101
	Huynh-Feldt	.785	.425	.101
	Lower-bound	.785	.405	.101
SOURCE * PANNING	Sphericity Assumed	.598	.667	.079
	Greenhouse-Geisser	.598	.579	.079
	Huynh-Feldt	.598	.639	.079
	Lower-bound	.598	.465	.079

Table 3.15 Result table of repeated measure ANOVA test for the data obtained for 'brightness' difference between stereophonic and monophonic stimuli

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.
SOURCE	.000	52.273	2	.000
PANNING	.252	8.276	2	.016
SOURCE * PANNING	.077	13.888	9	.143

Table 3.16 Result table of Mauchly's test of sphericity for the data obtained for 'brightness' difference between stereophonic and monophonic stimuli

The mean values and 95% confidence intervals for each sound source and each panning method are shown in **Figure 3.10**. It can be seen that even though the stereophonic images were graded to be 'duller' than the monophonic image in general, the magnitudes of the grading differences appear to be negligible. The results of pairwise comparisons between each sound source shown in **Table 3.17** indicate that the significant differences occurred between speech and piano and between trumpet and piano, which means that the significance of the sound source effect was caused by the piano source.

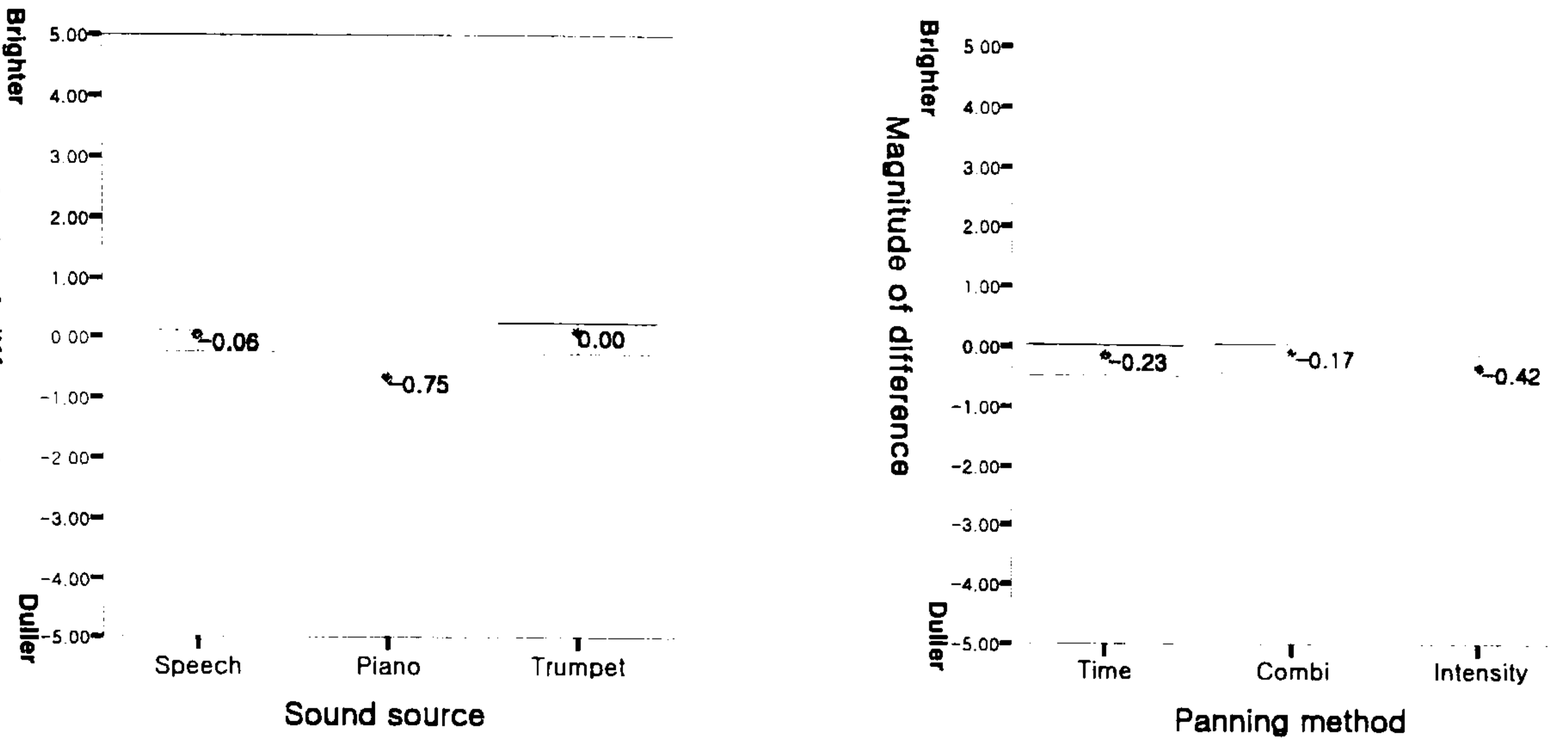


Figure 3.10 Mean values and the associated 95% confidence intervals of the grading data of 'brightness' difference between stereophonic and monophonic stimuli by sound source and panning method

Measure: MEASURE_1

(I) SOURCE	(J) SOURCE	Mean Difference (I-J)	Std. Error	Sig.
Speech	Piano	.689	.001	.000
	Trumpet	-.062	.189	1.000
Piano	Speech	-.689	.001	.000
	Trumpet	-.751	.189	.016
Trumpet	Speech	.062	.189	1.000
	Piano	.751	.189	.016

Measure: MEASURE_1

(I) PANNING	(J) PANNING	Mean Difference (I-J)	Std. Error	Sig.
Time	Combi	-.063	.083	1.000
	Intensity	.187	.262	1.000
Combi	Speech	.063	.083	1.000
	Intensity	.250	.231	.948
Intensity	Speech	-.187	.262	1.000
	Combi	-.250	.231	.948

Table 3.17 Result tables of pairwise comparisons between each sound source and between each panning method for 'brightness' attribute

3.4.3.5 Hardness

The results of the RM ANOVA test for the grading data obtained for 'hardness' ashown in **Table 3.19**. Similarly to the results for the 'brightness' attribute, the difference between sound sources is found to be significant (p = 0.000) while that between panning methods is not (p = 0.210). However, the estimated effect size of sound source appears to be fairly small (Partial Eta Squared value = 0.240). The source*panning interaction effect is shown to be insignificant (p = 0.257).

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		F	Sig.	Partial Eta Squared
SOURCE	Sphericity Assumed	9.949	.000	.240
	Greenhouse-Geisser	9.949	.000	.240
	Huynh-Feldt	9.949	.000	.240
	Lower-bound	9.949	.000	.240
PANNING	Sphericity Assumed	1.750	.210	.200
	Greenhouse-Geisser	1.750	.220	.200
	Huynh-Feldt	1.750	.214	.200
	Lower-bound	1.750	.227	.200
SOURCE * PANNING	Sphericity Assumed	1.512	.226	.178
	Greenhouse-Geisser	1.512	.257	.178
	Huynh-Feldt	1.512	.251	.178
	Lower-bound	1.512	.259	.178

Table 3.18 Result table of repeated measure ANOVA test for the data obtained for 'hardness' difference between stereophonic and monophonic stimuli

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.
SOURCE	.986	.082	2	.960
PANNING	.670	2.402	2	.301
SOURCE * PANNING	.003	32.334	9	.000

Table 3.19 Result table of Mauchly's test of sphericity for the data obtained for 'hardness' difference between stereophonic and monophonic stimuli

Figure 3.11 shows the mean values and 95% confidence intervals for each sound source and each panning method. It appears that the stereophonic images were graded to be 'softer' than the monophonic image, but the magnitudes of the differences appear to be very small. Similarly to the brightness attribute, significant differences appear to have occurred between speech and piano and between trumpet and piano as shown in **Table 3.20**, suggesting the dominant effect of piano source.

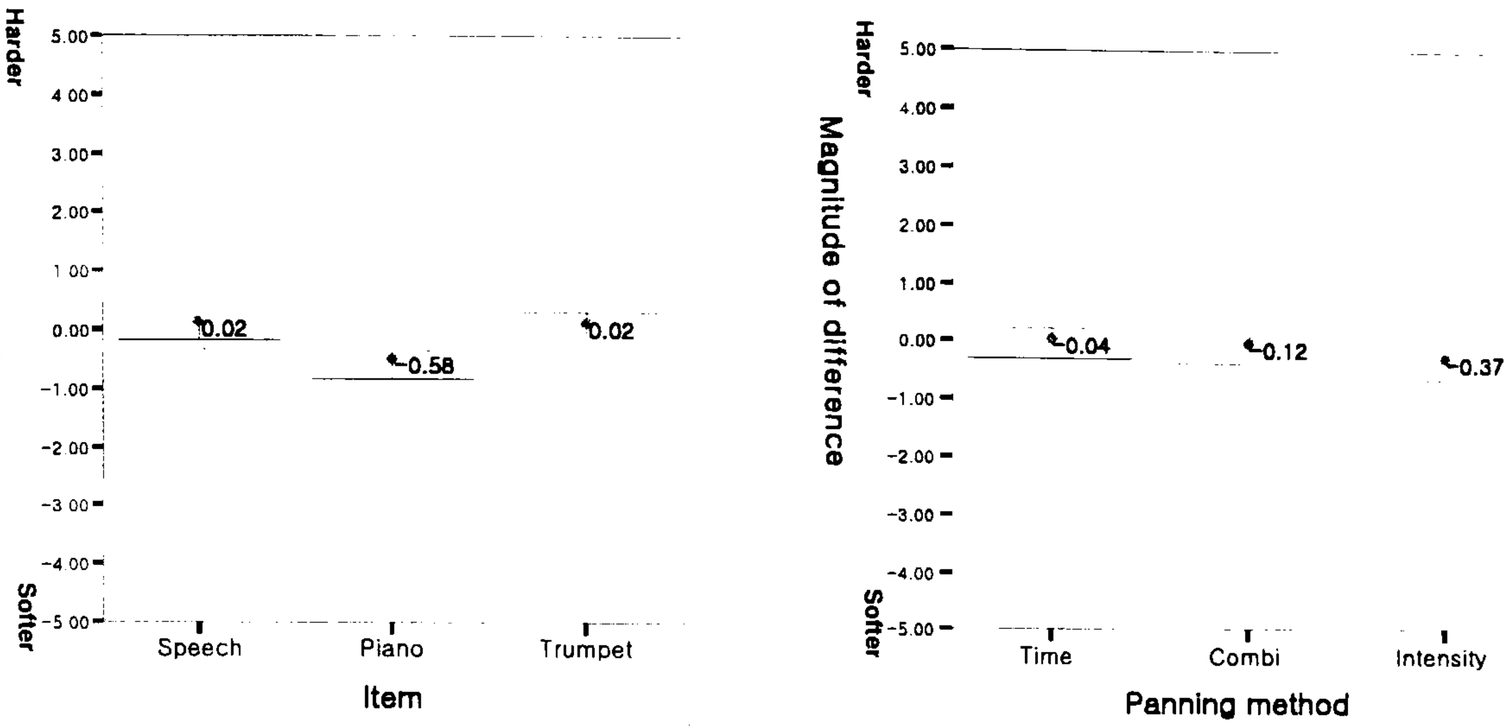


Figure 3.11 Mean values and the associated 95% confidence intervals of the grading data of 'hardness' difference between stereophonic and monophonic stimuli by sound source and panning method

Measure: MEASURE_1

(I) SOURCE	(J) SOURCE	Mean Difference (I-J)	Std. Error	Sig.
Speech	Piano	.604	.001	.000
	Trumpet	-3.469E-18	.001	1.000
Piano	Speech	-.604	.001	.000
	Trumpet	-.604	.002	.000
Trumpet	Speech	3.469E-18	.001	1.000
	Piano	.604	.002	.000

Measure: MEASURE_1

(I) PANNING	(J) PANNING	Mean Difference (I-J)	Std. Error	Sig.
Time	Combi	.083	.126	1.000
	Intensity	.333	.223	.534
Combi	Time	-.083	.126	1.000
	Intensity	.250	.194	.716
Intensity	Time	-.333	.223	.534
	Combi	-.250	.194	.716

Table 3.20 Result tables of pairwise comparisons between each sound source and between each panning method for 'hardness' attribute

3.4.3.6 Fullness

Table 3.21 shows the results of the RM ANOVA test for the grading data obtained for 'fullness' attribute and **Table 3.22** shows the results of Mauchly's test of sphericity. The results indicate that the effect of sound source was significant (p = 0.039) while that of panning method was not (p = 0.156). Nevertheless, similarly to the other timbral attributes described above, the estimated effect size of sound source appears to be negligible (Partial Eta Squared value = 0.089). It is also found that the interaction effect between sound source and panning method was insignificant (p = 0.203).

Measure: MEASURE_1

Source		F	Sig.	Partial Eta Squared
SOURCE	Sphericity Assumed	3.427	.039	.098
	Greenhouse-Geisser	3.427	.039	.098
	Huynh-Feldt	3.427	.039	.098
	Lower-bound	3.427	.039	.098
PANNING	Sphericity Assumed	1.915	.156	.057
	Greenhouse-Geisser	1.915	.156	.057
	Huynh-Feldt	1.915	.156	.057
	Lower-bound	1.915	.156	.057
SOURCE * PANNING	Sphericity Assumed	1.535	.203	.089
	Greenhouse-Geisser	1.535	.203	.089
	Huynh-Feldt	1.535	.203	.089
	Lower-bound	1.535	.203	.089

Table 3.21 Result table of repeated measure ANOVA test for the data obtained for 'fullness' difference between stereophonic and monophonic stimuli

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.
SOURCE	.974	.157	2	.924
PANNING	.196	9.789	2	.007
SOURCE * PANNING	.094	12.821	9	.189

Table 3.22 Result table of Mauchly's test of sphericity for the data obtained for 'fullness' difference between stereophonic and monophonic stimuli

Figure 3.12 shows the mean values and 95% confidence intervals for each sound source and each panning method. It can be seen that the stereophonic images were perceived to be 'fuller' than the monophonic image in all conditions. However, like the other timbral attributes, the magnitude of the effect does not appear to be considerable. **Table 3.23** shows the results of pairwise comparisons between each sound source and it is indicated that the piano and trumpet pair was the only pair that had a significant difference.

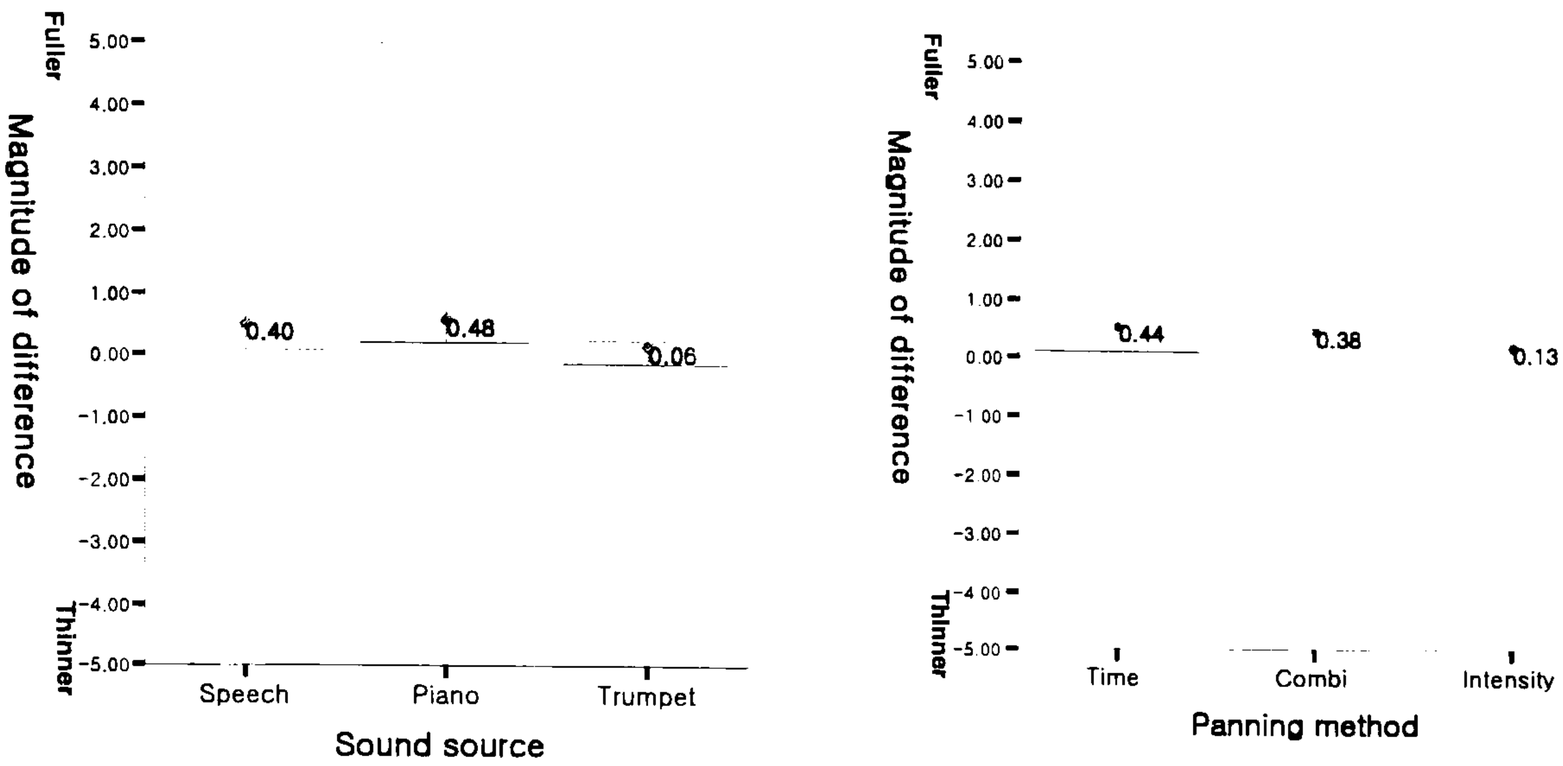


Figure 3.12 Mean values and the associated 95% confidence intervals of the grading data of 'fullness' difference between stereophonic and monophonic stimuli by sound source and panning method

Measure: MEASURE_1

(I) SOURCE	(J) SOURCE	Mean Difference (I-J)	Std. Error	Sig.
Speech	Piano	-.083	.169	1.000
	Trumpet	.335	.169	.156
Piano	Speech	.083	.169	1.000
	Trumpet	.418	.169	.048
Trumpet	Speech	-.335	.169	.156
	Piano	-.418	.169	.048

Measure: MEASURE_1

(I) PANNING	(J) PANNING	Mean Difference (I-J)	Std. Error	Sig.
Time	Combi	.063	.204	1.000
	Intensity	.312	.257	.789
Combi	Time	-.063	.204	1.000
	Intensity	.250	.083	.060
Intensity	Time	-.312	.257	.789
	Combi	-.250	.083	.060

Table 3.23 Result tables of pairwise comparisons between each sound source and between each panning method for 'fullness' attribute

3.4.4 Discussions

3.4.4.1 Discussion of the results for the individual attributes

From the results presented in the previous section, it was found that the effect of sound source was statistically significant for all the attributes except source distance, while that of panning method was significant only for the source focus and the source width

attributes. The significance found in the sound source effect for the timbral attributes seems to be a natural result to some extent because each sound source has different spectral characteristics. However, the estimated size of the sound source effect was shown to be very small for every attribute, whereas that of the panning method effect for the source focus and the source width attribute was great. This means that the most dominant differences between the stimuli, for the source focus and source width attributes, were caused by using different panning methods. Considering that noticeable comb-filtering effects due to reflections in an acoustical space are usually caused when the delay time is in the range between 10 and 50ms as mentioned in Section 2.1, it is suggested that the small effect sizes for the timbral attributes seem to be due to the small range of ICTD (<< 1ms) involved in the signals. However, this result cannot be generalised since only a limited range of spectral characteristics in sound source was considered in this experiment; only low note piano and trumpet sources were used, for example.

The results showed that the source focus and the source width attributes had similar patterns in the effects of both sound source and panning method, although the polarity of the scale used was opposite. For instance, the magnitude of panning method effect increased in the order of intensity, combination, and time panning (Intensity < Combination < Time). This result suggests that when there is a greater ratio of time difference to intensity difference information involved in a two-channel stereophonic microphone technique, the perceived phantom image will be less focused and wider. This confirms the widely known, but mostly anecdotally reported, spatial characteristics of coincident, near-coincident, and spaced-omni techniques.

The result for the source width attribute might be explained by the effect of interaural fluctuations over time on the perceived width of a source. As described in Section 2.3.2.4, Mason and Rumsey [2001] undertook research into interaural time difference (ITD) fluctuations as an objective measure related to auditory spatial perception in sound reproduction and they reported that the perceived source width increases as the magnitude of ITD fluctuations becomes greater. In the reproduction of conventional stereophonic recordings, the amount of interchannel time difference (ICTD) between each signal can determine the magnitude of ITD fluctuations. A larger ICTD will cause a higher degree of decorrelation between the interaural signals, therefore a greater magnitude of ITD fluctuations, which would also mean a smaller degree of interaural cross correlation (IACC) according to Mason [2002]. This explains why a spaced microphone technique would produce a wider phantom image than a coincident technique. Although fluctuation in IID would also be taken into account in the perception of source width to some extent, as mentioned in Section 2.3.2.4, ITD fluctuation tends to have a more dominant effect on the increase of perceived width.

The results show that the effect of sound source type on the source focus attribute was significant. From the interaction between sound source and panning method it was further found that the significant difference between sound sources was mainly caused by the difference between the speech source and the piano or trumpet sources for time panning. Piano and trumpet sources did not give rise to a significant difference. This might initially look rather contradictory to the findings of classical literature relating to the precedence effect discussed in Chapter 2. That literature suggested that a more continuous sound would be more difficult to localise than a more transient

sound. However, in the context of the current experiment the task was not to compare the three different sound sources directly with each other, but to compare the stereophonic phantom images for those sources with the reference monophonic images for each. Therefore, assuming that the trumpet source was originally difficult to localise due to its continuous nature, it might have been that the difference between the monophonic and stereophonic sounds was hardly detected in terms of the source focus attribute. On the other hand, assuming that the speech source was originally easily localised due to its ongoing transients, the difference between the monophonic and stereophonic sounds in respect of the source focus attribute would have been likely to be more distinctive. This might also be related to the 'plausibility hypothesis' proposed by Rakerd and Hartmann [1985], which was introduced in Section 2.2.3. That is, the continuous nature of the trumpet sound might have been recognised to be implausible for detecting necessary interaural time differences required for localisation of both monophonic and stereophonic images since it might have caused a strong interaction with room reflections,

For the source width attribute, it was found that the perceived differences for the speech and piano sources were significantly greater than that for the trumpet source. This could be initially explained by the fact that the speech and piano signals have more dominant low frequency energies than the trumpet signal (see **Figure 3.2**), since some literature suggests that the low frequency components of sound sources are significant for the perceived source width as reviewed in Section 2.3.2.2. However, it can be seen from the results that the perception of the source width difference has a similar tendency to that of the source focus difference and this might suggest that

these two attributes are correlated. If this is the case, it could be considered that the plausibility hypothesis might also have been applied for the perception of source width, although this is an issue that requires further investigation.

It is interesting to observe that the brightness and hardness attributes had similar sound source effects. It can be found that for both attributes the significance of the sound source effect was caused by the piano source regardless of the type of panning method (see **Figures 3.10** and **3.11**). This means that the stereophonic image became significantly duller or softer than the monophonic image when the piano source was used. This might be due to a comb-filtering effect occurred in the region of the upper harmonics. However, this result cannot be generalised because the piano source used in this experiment was only a single C3 note having spectral characteristics generated from a relatively low fundamental frequency. The result might have differed if a piano note with a higher fundamental frequency had been used.

3.4.4.2 Discussion of the relationships between the attributes

It was observed in the results that some attributes had similar patterns in the effects of sound source and panning method, e.g. source focus – source width, and brightness – hardness. In order to identify the perceptual dimensions of the six attributes, a principal component analysis was carried out. **Figure 3.13** displays the 'eigenvalue' for each component that was initially extracted. An eigenvalue conceptually represents the proportion of the total variance accounted for by a particular component,

and determines which components are retained in the analysis: only the components having an eigenvalue of greater than 1 are extracted. From the current analysis, therefore, only three effective components (component number 1, 2, and 3) are finally extracted.

Table 3.24 presents the rotated component matrix containing the partial correlation values of the six attribute tests on the three components extracted. It is shown that source focus and source width attributes essentially constitute the same perceptual dimension of Component 1; brightness and hardness attributes of Component 2; fullness and source distance attributes of Component 3. The interactions between each component based on the matrix are also shown in **Figure 3.14**.

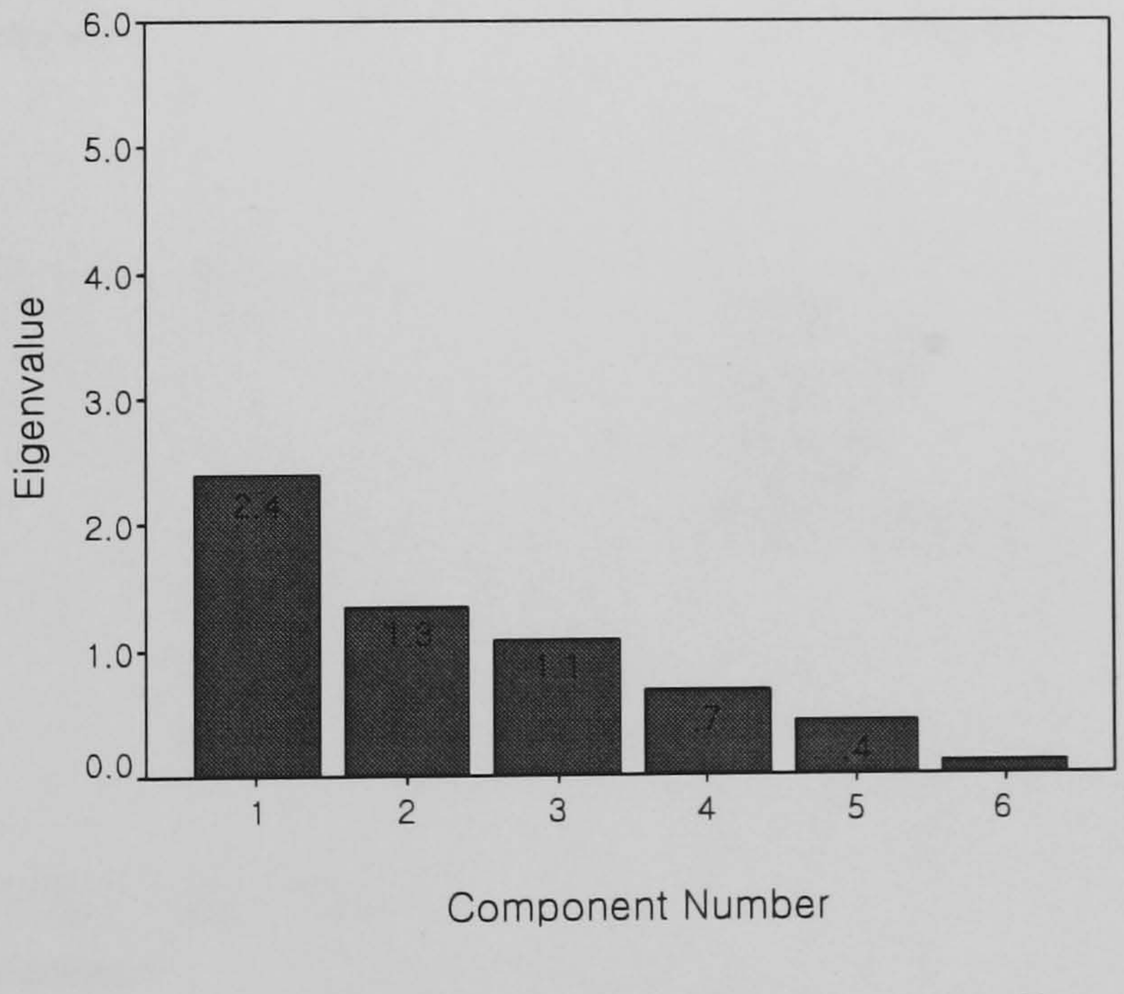


Figure 3.13 Display of the eigenvalues for the components initially extracted from principal component analysis

	Component		
	1	2	3
source focus	**−.966**	−.110	−.089
source width	**.949**	.048	.212
brightness	.051	**.902**	−.122
hardness	.102	**.803**	.319
fullness	.177	−.023	**.811**
source distance	−.083	−.136	**−.780**

Table 3.24 Table of the rotated component matrix obtained by principal component analysis

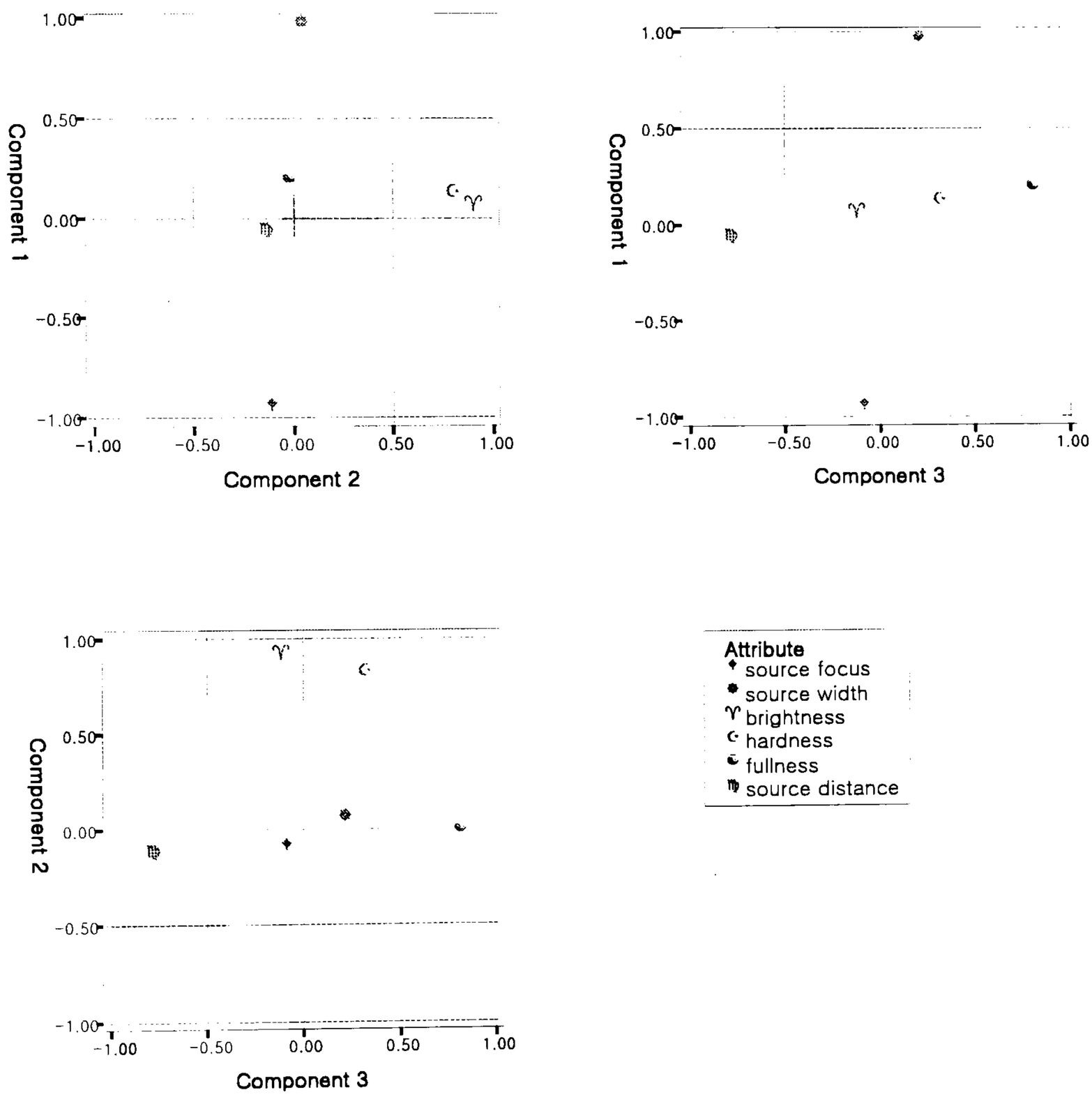


Figure 3.14 Component plots based on the rotated component matrix obtained by principal component analysis

This result is compared with the result of a bivariate correlation test, presented in **Table 3.25**. This directly indicates the relative strength of correlation between each attribute and each component. It can be found in the result that the source width and source focus attributes (Component 1) had a very strong negative correlation (r = -0.893); whereas brightness and hardness (Component 2) had a moderate correlation (r = 0.494). Correlation between fullness and source distance (Component 3) is shown to be weak (r = - 0.341). From this it might be suggested that even though three hidden perceptual dimensions were discovered from principal component analysis, Components 2 3 might not be particularly relevant or directly interpretable. However, it can be strongly suggested that the subjects perceived source focus and source width attributes in the same dimension in the listening test. For example, a less easily localised source might have been perceived as a wider source, whereas a more easily localised source might have been perceived as a narrower source.

		source focus	source width	source distance	brightness	hardness	fullness
source focus	R	1	**-.893**	.180	-.143	-.204	-.239
	Sig. (2-tailed)	.	.000	.130	.232	.085	.043
source width	R	**-.893**	1	-.271	.058	.211	.306
	Sig. (2-tailed)	.000	.	.021	.628	.075	.009
source distance	R	.180	-.271	1	-.078	-.240	-.341
	Sig. (2-tailed)	.130	.021	.	.516	.042	.003
brightness	R	-.143	.058	-.078	1	.494	-.051
	Sig. (2-tailed)	.232	.628	.516	.	.000	.670
hardness	R	-.204	.211	-.240	.494	1	.253
	Sig. (2-tailed)	.085	.075	.042	.000	.	.032
fullness	R	-.239	.306	-.341	-.051	.253	1
	Sig. (2-tailed)	.043	.009	.003	.670	.032	.

Table 3.25 Result table of bivariate correlation test

3.4.4.3 Limitations

The investigation described in this chapter was designed and conducted systematically but there are also a number of limitations that must be considered.

The fundamental frequencies of the musical sound sources used for this investigation were limited to low frequencies and this limited the scope of the elicitation and grading experiments, especially for the timbral attributes.

Single notes of musical sound sources were used and this certainly enabled the author to strictly control the variables of the temporal and spectral characteristics of sound. However, the musical stimuli were generally said by the subjects to be somewhat uncomfortable to listen to. Especially the continuous trumpet stimuli were found to be tiring when listened to repeatedly for a long period and this might have affected the subject's ability for consistent judgment. The piano stimuli were also found to be difficult to compare simultaneously because they were single transient hits. For these reasons, it was recognised that it could be more appropriate to use performance extracts of single instruments for the next investigation.

In the course of instructing the subjects for the grading experiment, it was found that some subjects were not fully familiar with the definitions of some attributes because those attributes were not directly developed from the terms that those subjects had described individually. This is likely to be due to the lack of group discussion in the process of developing pooled subjective terms, which would have given each subject

the opportunity to familiarise themselves with the meanings of the terms that were elicited by other subjects. Therefore, extra verbal explanations on the definitions of the provided attributes were required in the instruction to avoid a logical error, which was described in **Table 3.5**.

It was reported in Section 3.4.4.2 that the source focus and source width attributes were negatively correlated at a high level and this might be a natural result. However, it might also be that this strong correlation was caused by a proximity error (see **Table 3.5**). That is, since the two attributes are conceptually adjacent, biases on the relationship between the attributes might have been involved in the subjects' gradings when they were graded in the same test. The results might have been different if the two attributes had been tested separately.

3.5 Summary

A series of subjective experiments were conducted in order to investigate the perceptual attributes of 2-0 stereophonic phantom images. There were three different sound sources: speech, transient piano hit and continuous trumpet note. The stereophonic stimuli were created by using three different panning methods: pure time panning, pure intensity panning and a combination of the two. Firstly, the subjects described the perceived differences between the stereophonic and the reference monophonic sounds using their own terms. The subjective terms were then separated into six attribute groups. Finally, the subjects graded the magnitudes of the perceived

differences between the stimuli on the attribute scales developed. The data obtained from the grading experiment were analysed using the RM ANOVA statistical model. The findings of this investigation are summarised below:

- Six common attributes were developed from the elicited terms for all sound source types. There were three spatial attributes, comprising source focus, source width and source distance, and three timbral attributes comprising brightness, hardness and fullness.
- Source focus and source width were perceptually the most dominant attributes of 2-0 stereophonic images.
- The type of sound source had a significant effect on the difference between stereophonic and monophonic images for all attributes except source distance.
- The type of panning method had a significant effect only for the spatial attributes of source focus and source width.
- Source focus and source width were correlated at a high level.

4 PERCEPTUAL EFFECTS OF INTERCHANNEL CROSSTALK IN 3-2 STEREOPHONIC MICROPHONE TECHNIQUES

This chapter describes a series of subjective experiments conducted to investigate the perceptual effects of interchannel crosstalk in 3-2 stereophonic microphone techniques. As introduced in chapter 0, interchannel crosstalk in the context of the current studies is defined as an extra signal to the primary signals that are responsible for the localisation of phantom image in the desired two-channel based stereophonic segment. For instance, if a three-channel microphone technique is to be used for recording a sound source located in the right stereophonic segment, the signals of the centre (C) and right (R) microphones are regarded as the signals primarily responsible for image localisation while the signal of the left microphone (L) is regarded as crosstalk. If it is assumed that the image localisation resulting from the interchannel relationship between the signals of C and R is not affected significantly by the crosstalk signal L, the perceptual effect of the crosstalk signal can be investigated by comparing the image created by C and R (crosstalk-off) with that created by L, C and R (crosstalk-on). The investigation described in this chapter is based on the above assumption.

The primary research questions formulated for this investigation were as follows:

- What kind of auditory attributes are perceived when interchannel crosstalk is present in multichannel microphone techniques?

- How audible are these attributes?
- Does the subjective grading for these attributes depend on the configuration of microphone array (combination ratio of interchannel time and intensity differences), the type of sound source, or acoustic condition?
- Does interchannel crosstalk have a significant effect on the subjective preference for perceived sound quality?

In order to answer these questions, a series of listening experiments were designed and undertaken. The first two experiments were inspired by the QDA method, which was described in Section 3.2. The first experiment was conducted to elicit the perceptual attributes of interchannel crosstalk and examined the relative perceptual weights of those attributes and the second experiment investigated the significance of the effects of microphone array configuration, sound source type and acoustic condition. The results of these two experiments were of the main interests in the current research. However, as mentioned above, it was of additional interest to see the effects of interchannel crosstalk on subjective preference on perceived sound quality. Therefore, the third experiment examined the preference between the crosstalk-off and crosstalk-on stimuli, which were selected from the stimuli that were used for the elicitation and grading experiments. Finally, an additional experiment was carried out to further investigate the preference for interchannel crosstalk using practical recordings made with microphone techniques having different crosstalk characteristics.

4.1 Experimental Hypotheses

The literature reviewed in Chapter 2 generally suggested that in the context of concert hall or room acoustics, the addition of delayed secondary signals to the original signal would influence the perception of localisation accuracy, spatial impression and tone colour of the auditory image. From the experiments described in the previous chapter, this was confirmed to be the case in the context of 2-0 stereophonic sound reproduction. Those experiments investigated the perceptual difference between monophonic source images and 2-0 stereophonic phantom images created with various ratios of interchannel time and intensity differences, using different types of sound source. It was shown that the differences were perceived in both spatial and timbral attributes comprising source focus, source width, source distance, brightness, hardness and fullness. It was predicted that similar differences would be perceived between two-channel phantom images with crosstalk off (CR) and three-channel images with crosstalk on (LCR) in three-channel microphone techniques, based on the similarity between the contexts of the two stereophonic experiments (i.e. comparison between one-channel and two- channel images vs. comparison between two-channel and three-channel images). The results from the previous experiment also showed that the panning method or sound source had a significant effect on the perceptual difference between stereophonic phantom image and monophonic source image depending on the type of perceptual attribute. From this, it was logical to hypothesise that the combination ratio of interchannel time and intensity differences involved in the crosstalk signal, and the type of sound source, would also affect the perceptual difference between crosstalk-off (CR) and crosstalk-on (LCR) images.

The acoustical characteristic of the recording environment was also predicted to be an important factor since such acoustic parameters as reflections and reverberation would be likely to affect the pattern of perception of the sound images as discussed in Chapter 2. Additionally, it was predicted that the subjective preference for sound images created with interchannel crosstalk would be dependent on the type of sound source since the specific attributes of sound images desired by recording engineers would be likely to vary depending on the temporal or spectral characteristics of sound sources.

4.2 Designs of Elicitation and Grading Experiments

This section describes the experimental design involved in the elicitation and grading experiments. This will include discussions on the choices of independent variables and the process of experimental stimuli creation.

4.2.1 Choice of microphone technique

4.2.1.1 Basic philosophy

As discussed in Section 1.4.1, current 3-2 stereophonic microphone techniques can be divided into two main groups according to Rumsey [2001]'s classification: those that use five-channel main microphone arrays and those that use separate front and rear arrays. To recap briefly, the former consists of five microphones that are placed relatively close to each other and form a single array, pursuing the recreation of a

natural sound field of the recording space. With these techniques, interchannel crosstalk is likely be an issue not only between the front channels but also between the front and surround channels due to the relatively short distance between the front and rear microphones. The techniques in the other group use frontal main microphone arrays that are used specifically for accurate pickup of direct sound so that sources can be easily localised on reproduction, together with separate rear microphone arrays that are designed to pick up decorrelated ambient sound to feed the surround loudspeakers. Different rear microphone arrays can be combined with different frontal arrays depending on the desired directional and ambience characteristics. For the techniques in this group interchannel crosstalk between the front and rear microphones would not be significant because of the sufficiently long distance between them. In this regard, it seems that techniques in this group give recording engineers more freedom to control the spatial impression and enables them to use their artistic and technical creativity more than the five-channel main microphone technique. For this reason, a technique with separate treatment of front and rear was chosen as the basis for the elicitation, grading and controlled preference experiments.

4.2.1.2 Simulation of microphone technique

If a microphone technique were operated in a practical recording venue, such uncontrolled acoustic artefacts as reflections and reverberation might lead to difficulty when analysing the factors that caused the resulting perceptual effects. In order to obtain data about the effects of interchannel crosstalk on phantom images in the

absence of room reflections the experiment included a simulation of recordings made in an anechoic condition, rather than using recordings made in a practical venue. For the anechoic experiment, only a three-channel frontal microphone technique was needed. Even though the primary aim of this research was to understand the effect of interchannel crosstalk in anechoic recording conditions, which enable one to obtain the controlled results, it was also of interest to see how the perception of this effect would differ in the context of different reverberant recording conditions. As discussed in the previous section, the purpose of the rear microphone array in the context of this experiment is to provide a diffuse ambience rather than a localisable image of the direct sound. The ambient sound picked up by a rear microphone array was simulated by using an artificial reverberator.

4.2.1.3 Frontal microphone technique

The frontal microphone technique chosen for these experiments was the so-called 'critical linking' three-channel microphone technique, proposed by Williams and Le Du [1999] (detailed descriptions of this technique were presented in Section 1.4.2). The basic design concept of this technique aims to achieve a continuous distribution of phantom images across channels L, C and R by linking the stereophonic recording angles (SRAs) of each stereophonic segment C-L and C-R without overlap. Within one segment, the psychoacoustic laws for localisation in conventional two-channel stereophonic reproduction such as summing localisation or the precedence effect are applied independently without considering the influence of the other segment. For

example, when a sound source is located at 45° to the right of the centre line, localisation of the phantom image should be governed by the summing localisation effect between C and R only, and in this case L can be regarded as crosstalk to the channels C and R. Ideally, L should not be taken into account in the localisation process since it is to be suppressed by the same effect or the precedence effect operating between C and L. It was shown in **Figure 1.18** in Section 1.4.2 that the linear attachment of two separate recording segments could be successful for microphone techniques of critical linking type.

However, from the reports on the perceptual effects of reflections that were reviewed in Chapter 2, it could be hypothesised that even though the position of the phantom image can be solely determined by C and R without the aid of L, the presence of L will influence the spatial or timbral quality of the image to some extent. This could also be supported by the results of the previous experiment indicating that the stereophonic phantom image created with certain time and intensity differences between two channels was perceived to have differences to the corresponding monophonic image in both spatial and timbral attributes. In this regard, it is logical to examine the effect of interchannel crosstalk by comparing the image that is created with the crosstalk channel turned on (image formed by contributions from LCR) and that with the crosstalk channel turned off (CR only). The critical linking technique supposedly enables one to create various array styles having different distances and angles between microphones while keeping the SRA across L, C and R constant. Therefore, the effect of the ratio of time to intensity differences between the crosstalk

signal and the other channels can be investigated by comparing different microphone arrays sharing the same SRA.

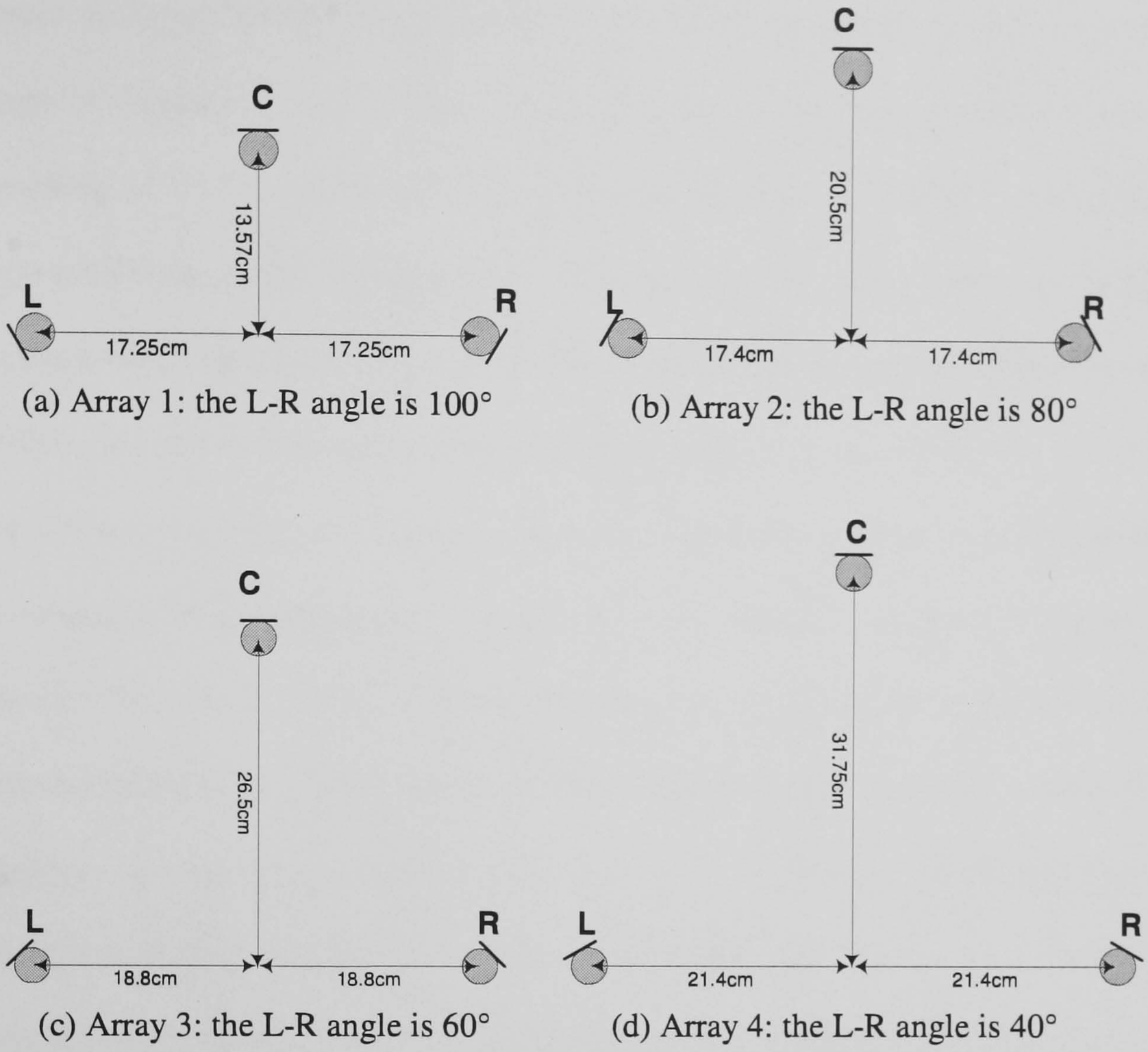
Williams and Le Du provided various examples of critically linked microphone arrays. For the current experiment, four sample arrays were selected from the examples as shown in **Figure 4.1**. These particular arrays were chosen because the difference between each array in the distance and angle between microphones was considered to be large enough to provide four distinctive interchannel relationships for the crosstalk signals. The common SRA for these arrays was 180°, the simulated direction of the sound source was 45° from the centre line of the array and the distance from the centre point of the array was five metres. The particular source direction was chosen because the interchannel relationship caused by a source located at that direction was considered to be a good compromise between the extreme interchannel relationships required for the hard-centre and fully-right images that can be created within the SRA of 90° for the C-R segment.

The interchannel time and intensity differences between L and C and between R and C calculated for each array are shown in **Table 4.1**. As found by the authors mentioned in **Table 1.1**, e.g. Simonsen [1984], Wittek [2000], Lee [2004] (see Appendix A) , in a conventional 2-0 stereophonic reproduction the minimum interchannel time difference (ICTD) required for localising a phantom image at a fully one loudspeaker is 1.0-1.1ms, provided that there is no interchannel intensity difference (ICID). On the other hand, the minimum ICID required for the same effect is in the range of 15-18dB, provided that there is no ICTD. Certain

combinations of relevant ICID and ICTD can also cause the same effect and they can be calculated based on the time-intensity trade-off curves of Williams [1987] (see **Figure 1.1**) or those of this author [2004] (see **Figure A.6** in Appendix A) depending on whose psychoacoustic values are believed. It was suggested by Theile [2001] that this trading relationship could be applied constantly in three-channel application. That means that the ICTD and ICID relationship required for localising the phantom image at fully one side between L and R in a two-channel stereo would cause the phantom image to be localised at fully one side between C and L or C and R in a three-channel stereo. It appears that whatever trade-off curve is used, the combined ICTD and ICID values for C - L segment shown in **Table 4.1** are more than enough to cause the full phantom image to be localised fully at C. This suggests that the crosstalk signal L would theoretically have no effect on determining the position of the phantom image.

	C to L delay	**C to L intensity**	**C to R delay**	**C to R intensity**
Array 1	0.64ms	- 20.5dB	- 0.08ms	- 0.7dB
Array 2	0.79ms	- 12.8dB	0.06ms	0.6dB
Array 3	0.94ms	- 8.0dB	0.16ms	1.2dB
Array 4	1.09ms	- 4.6dB	0.21ms	1.4dB

Table 4.1 Time and intensity differences between the centre channel and the left or right channel for each array: the simulated direction of sound source is 45° and the simulated distance of the sound source from the arrays is 5m.



(a) Array 1: the L-R angle is 100°

(b) Array 2: the L-R angle is 80°

(c) Array 3: the L-R angle is 60°

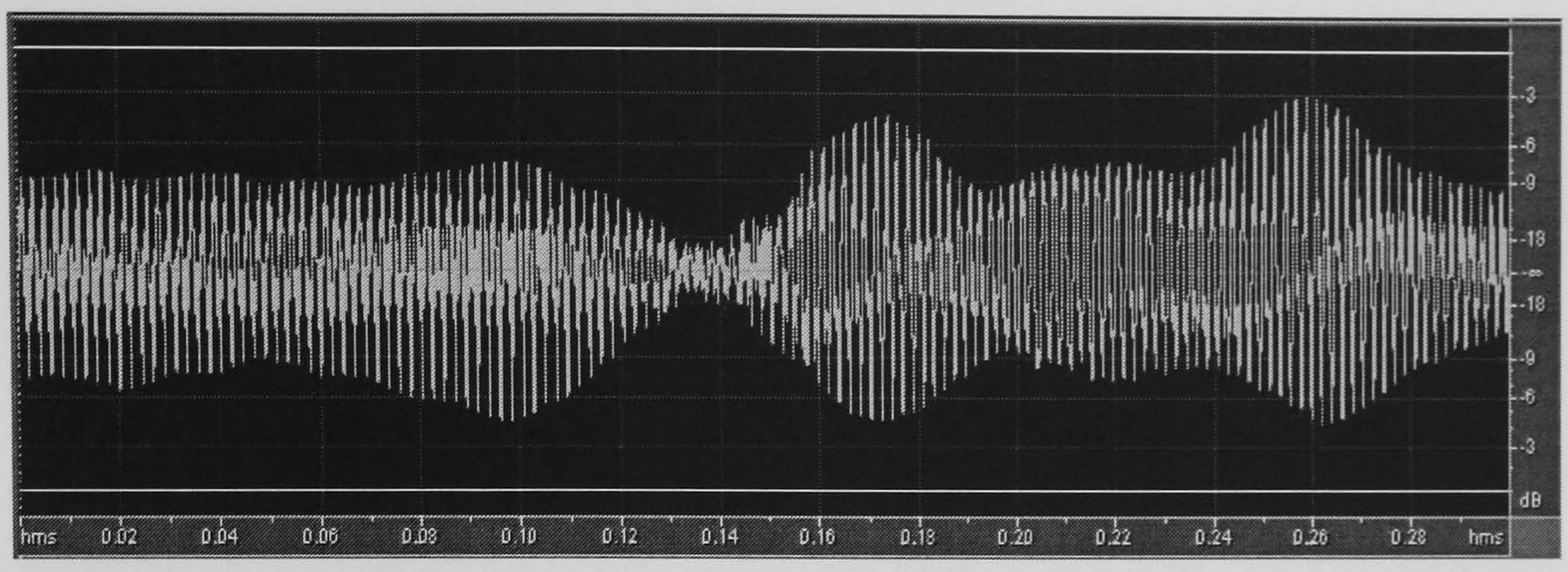
(d) Array 4: the L-R angle is 40°

Figure 4.1 Configuration of 'Critical linking' microphone arrays simulated for the elicitation and grading experiments

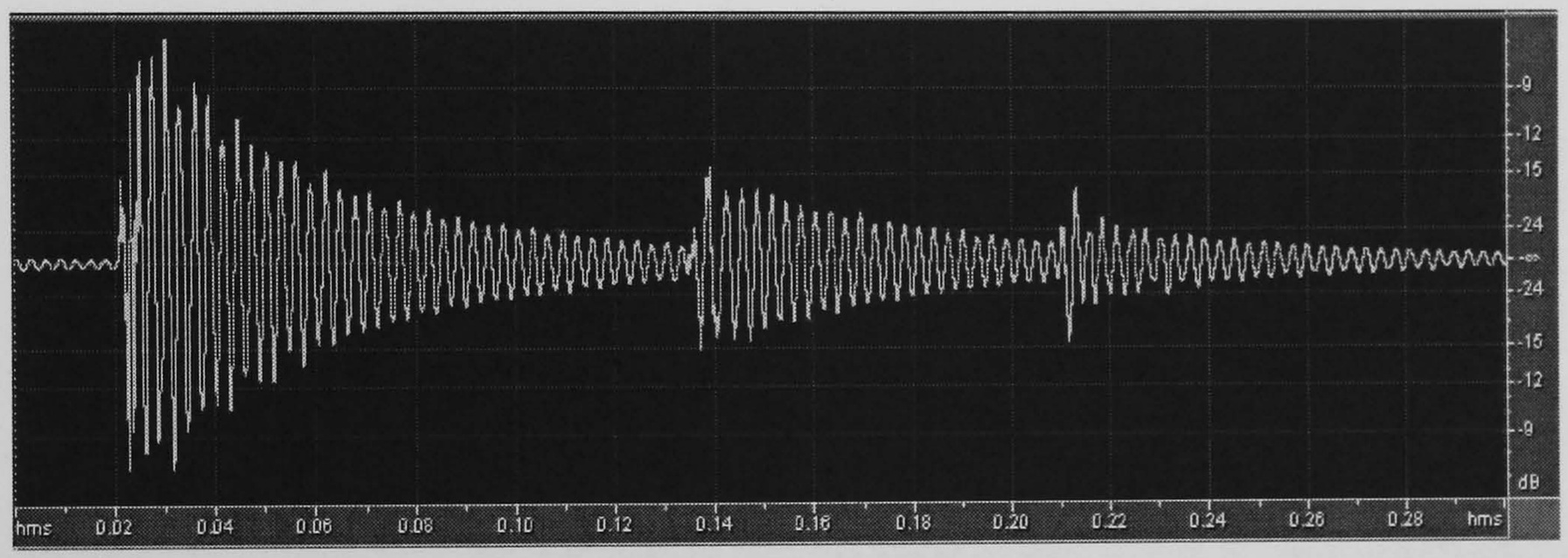
4.2.2 Choice of sound source

It was of interest to examine whether the effect of interchannel crosstalk depends on the type of sound source. Three types of natural sound source comprising cello, bongo and speech were chosen for this experiment due to their distinctive temporal and spectral characteristics, with the cello being relatively continuous and having a complex harmonic structure, the bongo having a strong transient nature, and the

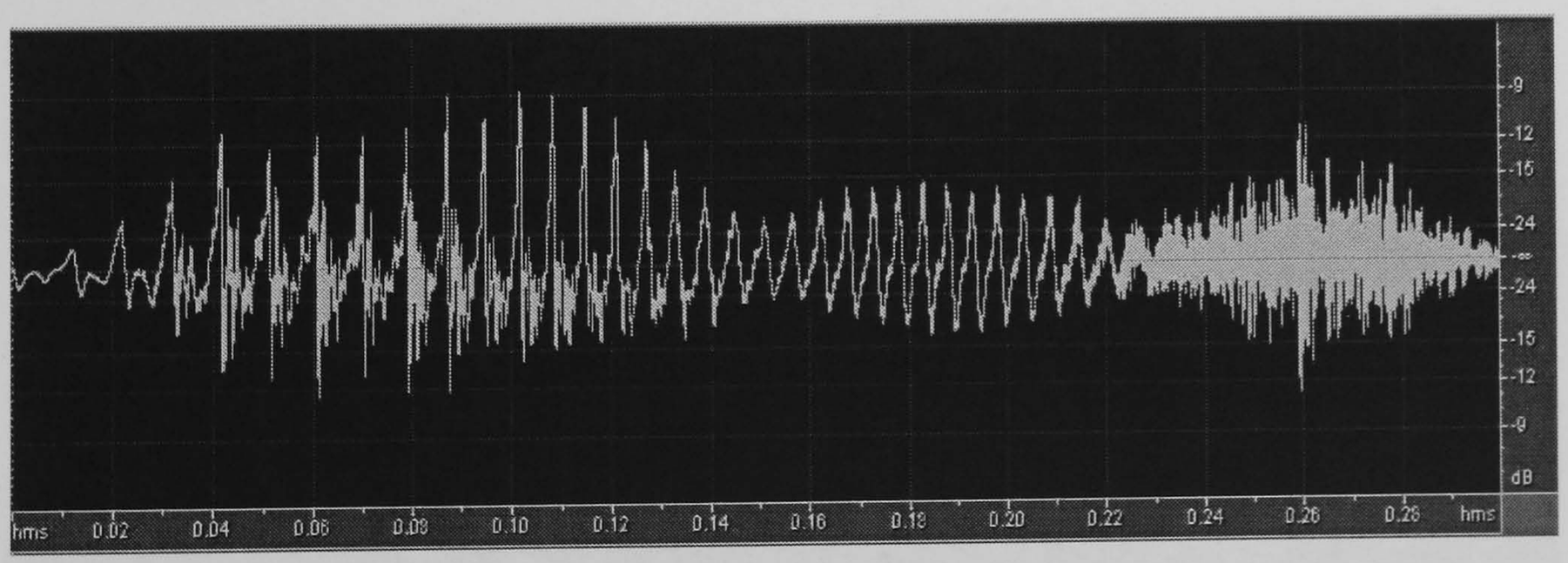
speech having a fine mixture of transient and continuous sounds as well as a wide range of frequencies. The signal for each sound source was an anechoic mono recording of a performance excerpt taken from the Bang & Olufsen Archimedes project CD [Hansen and Munch 1991]. From a psychophysical viewpoint, it might be claimed that the characteristics of natural sound sources are too complex to strictly analyse the effect of spectral or temporal characteristics of the sound. In fact, the use of pure sine tones or bandpass noise signals might allow a more controlled investigation of various aspects. However, results obtained with strictly controlled stimuli often lack ecological validity and might not be applicable to natural sound sources because the characteristics of the latter are more complex and invoke cognitive associations as well as basic perceptual responses. Therefore it was deemed to be more appropriate to use sound sources likely to be encountered in practical recording situations. The waveform and frequency analysis plots for each sound source are shown in **Figures 4.2** and **4.3**. The waveform shows temporal variations during specific 0.3 second extracts taken from the performance, which show representative temporal characteristics, and the frequency analysis is a plot of the average intensity by frequency over the whole performance.



(a) Cello source

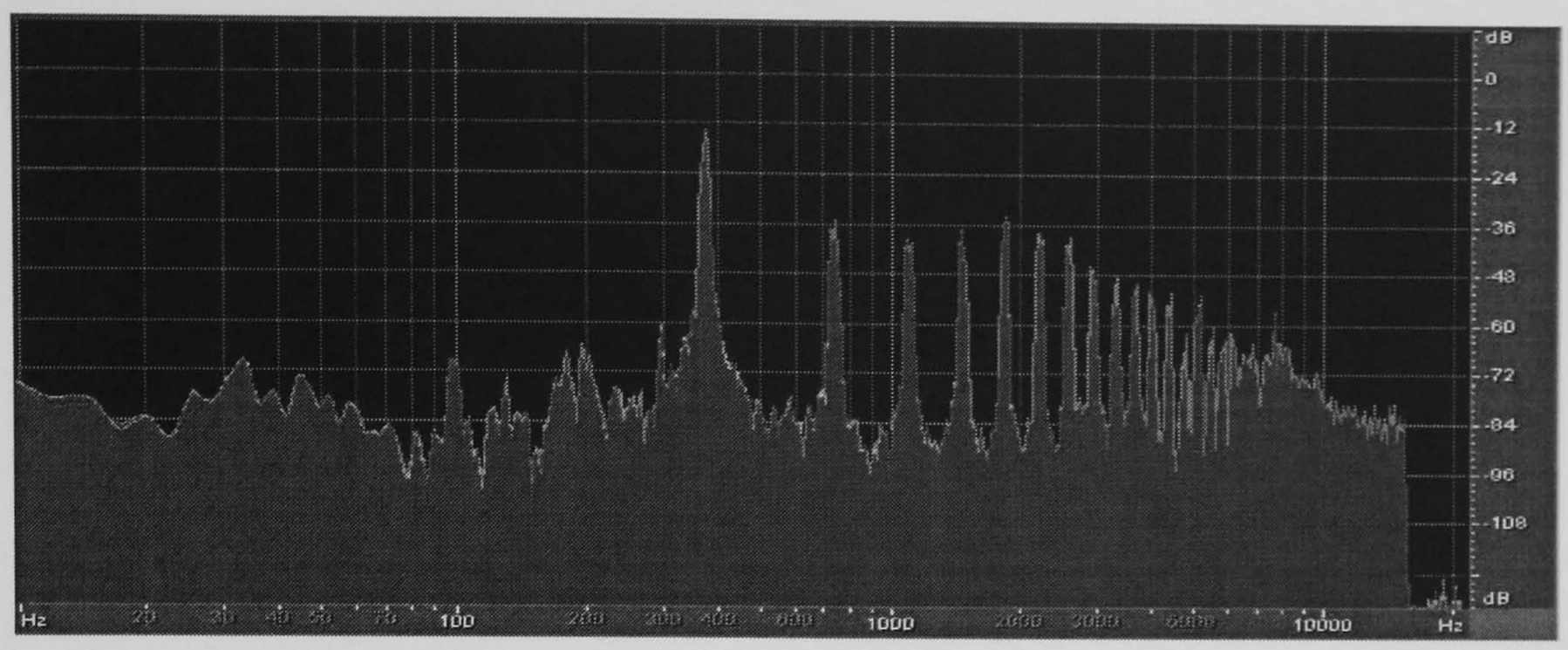


(b) Bongo source

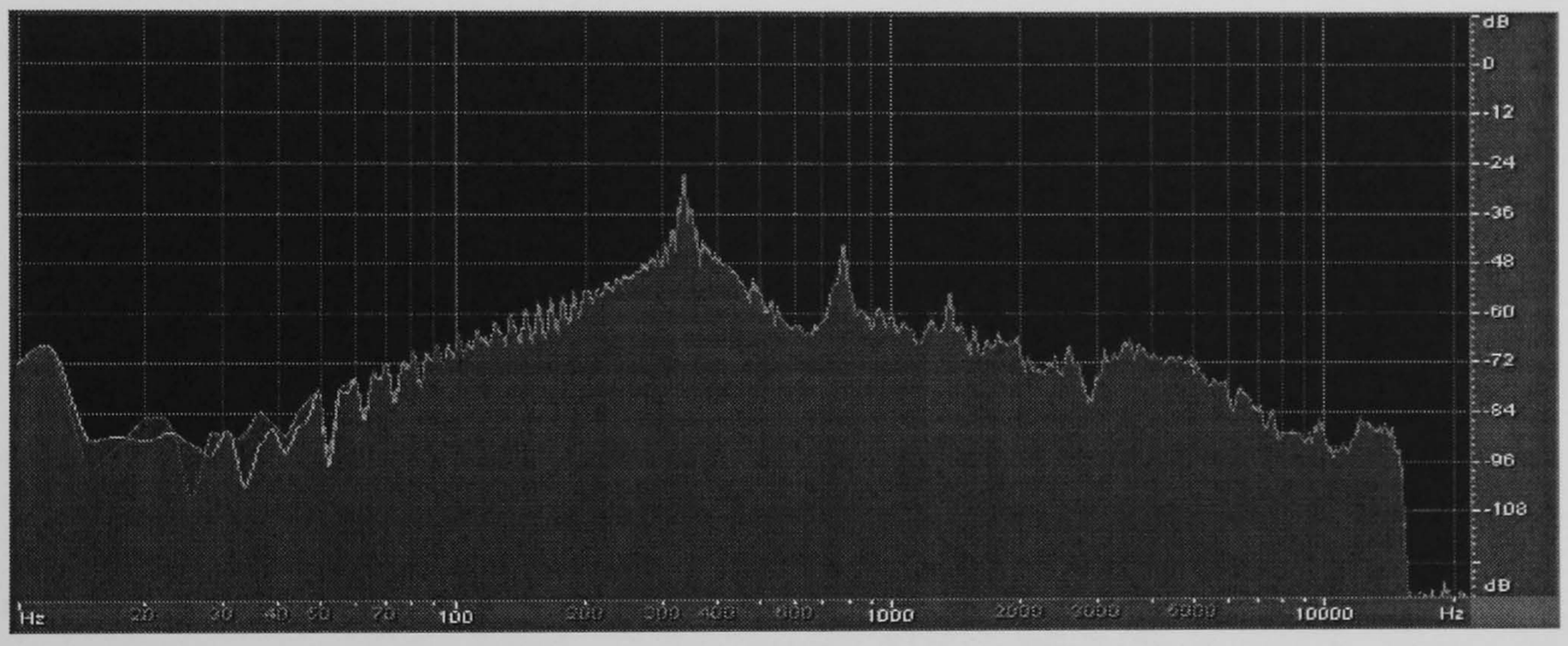


(c) Speech source

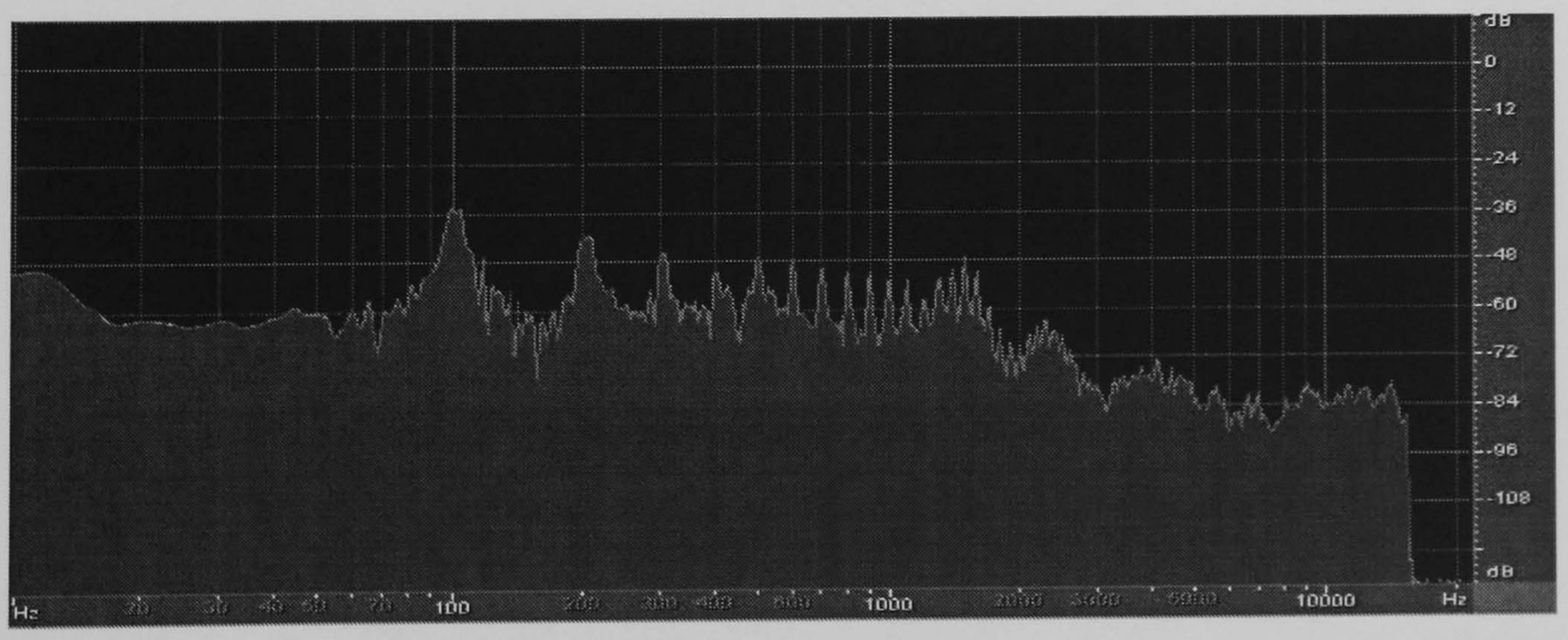
Figure 4.2 Short term extracts of waveforms for each sound source



(a) Cello source



(b) Bongo source



(c) Speech source

Figure 4.3 Long-term averaged frequency spectrum of each sound source

4.2.3 Acoustic conditions

The acoustic conditions considered in this experiment comprised anechoic, 'room' and 'hall'. As mentioned above, the anechoic condition was of primary interest since it enabled the strict control of variables, and it was created naturally by using anechoically recorded sound sources. Simulations of recordings made in different acoustic conditions were also used in order to predict the behaviour of interchannel crosstalk in practical recording venues such as room and hall. In this simulation, the interchannel crosstalk signals and artificial reflections or reverberation could be controlled separately, which means that the effect of interchannel crosstalk was only imposed on the direct sound component. However, this is not possible in practical situations because the front microphone array would normally pick up reflections or reverberation at the same time. Therefore, what was intended in this simulation was to observe in a controlled manner how the reflections or reverberation with a reasonable mixing level would influence the effect of interchannel crosstalk perceptually. The detailed characteristics of the simulated room and hall conditions are described in the next section.

4.2.4 Stimuli creation process

A set of multichannel stimuli, involving 36 combinations of four microphone arrays, three sound sources and three acoustic conditions, was processed for the experiment. The process was carried out in Studio 3, a multichannel sound control room of the

University of Surrey's Department of Music and Sound Recording. The diagram for the stimuli creation process is shown in **Figure 4.4**. For the creation of the anechoic stimuli, monophonic signals of each anechoic sound source were first fed into three separate channels on a Sony Oxford-R3 digital console and they were processed in accordance with the time and intensity relationship of each microphone array shown in **Table 4.1**. The processed signal of each channel was then routed to each group output of L, C and R for the reproduction of three front channels. On the other hand, the room and hall stimuli were mixed for the reproduction of all five channels. The monophonic signal of the anechoic sound was sent to a Lexicon 480L reverberator through an auxiliary output of the mixer. The four purely ambient output signals generated from the reverberator were then routed to two group outputs for reproduction of the front channels L and R as well as those for the surround channels LS and RS, with the intensities of each signal kept the same, thus being mixed with the original anechoic sound signals in L and R. The basis for using the four outer channels for reproduction of the reverberation signals is as follows. As mentioned in Section 1.4.3, Hiyama *et al* [2002] investigated the number of loudspeakers required for the reproduction of the optimum spatial impression of a diffuse sound field. To recap, a reference loudspeaker arrangement consisting of 24 loudspeakers placed at every 15° making a circle was compared with various arrangements having a different number of loudspeakers (12, 8, 6, 5, 4, 3 and 2) with regard to spatial impression. They found that at least four loudspeakers, which were arranged in similar positions to the BS.775-1 recommendation, were required for listeners to perceive a similar spatial impression to the reference sound. For creating ambient sounds of room and hall, the presets of 'large room' and 'large hall' setup existing in the reverberator

were used. The details of the reverberator setup used for creating the room and hall ambient sounds are shown in **Table 4.2**. In general, the 'large room' set can be described as producing coloured and comb-filtered ambient sounds with slapping echoes. The 'large hall' creates an ambient sound that has a longer reverberation time and is more diffused without colouring the direct sound.

	Size	**RT Mid**	**RT Low**	**HF Cut-off**	**Pre-delay**
Large Room	19m²	0.70s	0.70s	6.593kHz	0ms
Large Hall	37m²	2.19s	2.63s	2.862kHz	24ms

Table 4.2 Parameters of the 'Lexicon 480L' reverberation setup used for simulations of room and hall (RT Mid = middle frequency reverb. time, RT Low = low frequency reverb. time)

The mixing ratio of the direct sound and reverberation was up to the author's aesthetic judgment as an experienced balance engineer, aiming to compromise between maintaining the clarity of the direct sound and achieving sufficient listener envelopment. The signals from each group output were individually recorded to computer hard disk using a Protools hard disk recording interface and were eventually transformed as monophonic audio files.

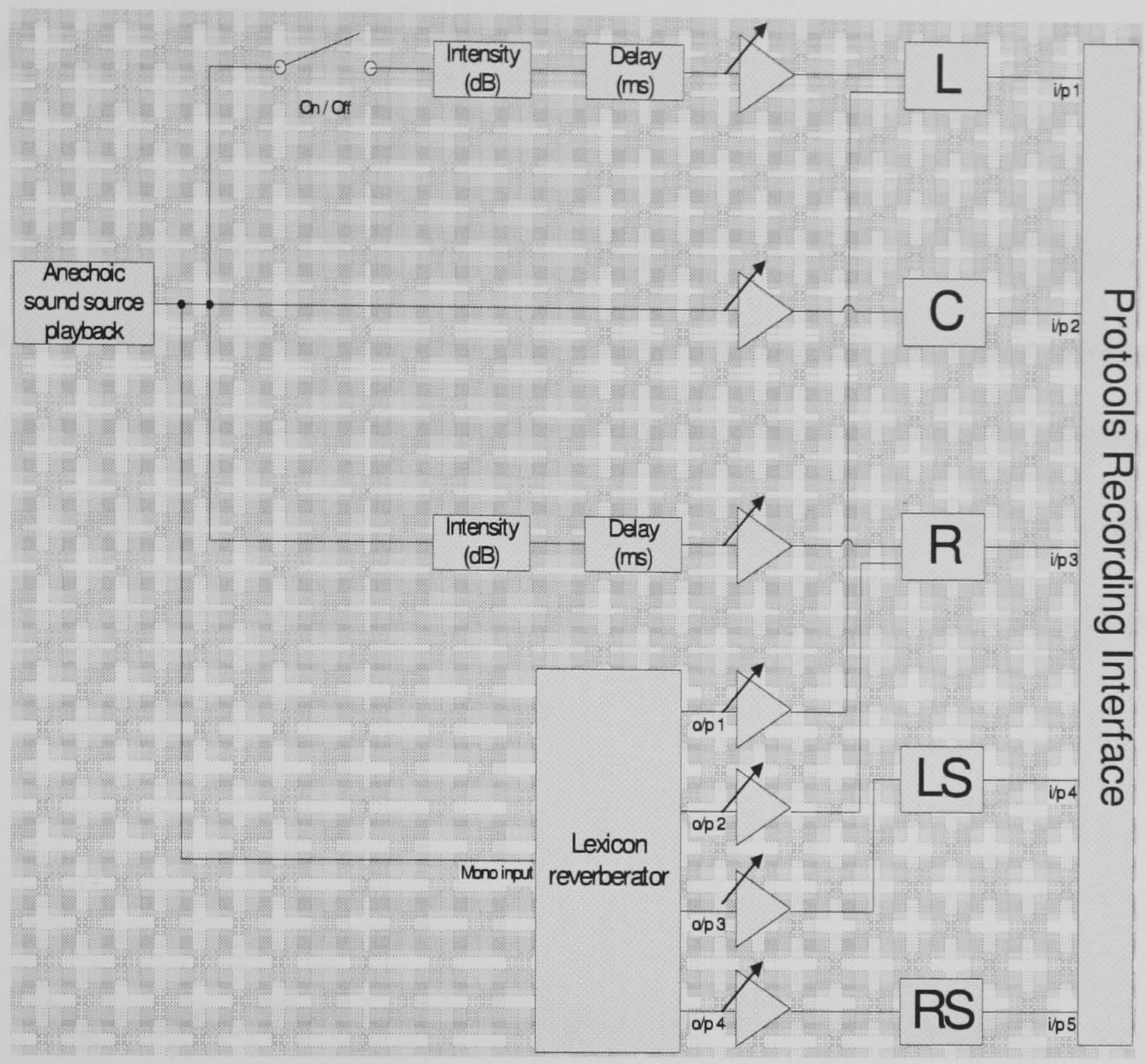


Figure 4.4 Diagram of signal processing for stimuli creation

4.2.5 Physical setup

The experiments were conducted in an ITU-R BS.1116-compliant [1994] listening room at the University of Surrey. In accordance with the ITU-R BS.775-1 recommendation [1993], five Genelec 1032A loudspeakers were set up at 0°, 30° and 110°, with a distance of 2m from the subject's seat. In order to avoid the effects of loudness difference on the subjects' judgments, the peak sound pressure levels of all stimuli were calibrated at 75dBA. The stimuli were played back through a Yamaha O2R mixing console and controlled by a computer-based control interface placed in front of the listener's seat.

4.2.6 Test subjects

Similarly to the case of the experiments described in the previous chapter, it was deemed to be more reasonable to employ experienced listeners for tests requiring fine perceptual distinctions, as suggested in ITU-R BS.1116 rec. [1994]. Therefore, a total of eight experienced subjects took part in the experiment. They were selected from staff members, research students and final year undergraduate students on the University of Surrey's Tonmeister course.

4.3 Experiment Part 1: Elicitation of Perceptual Attributes

4.3.1 Listening test method

This process used only six representative stimuli from the whole set of stimuli created. They were each anechoic sound source combined with microphone arrays 1 and 4, which were considered to have the most distinctive difference in perception of the resulting images. The reason for using only the anechoic stimuli was that they enabled the most focused listening to the effect of interchannel crosstalk without any artefacts of recording room acoustics. This test was designed to give the subject the freedom to control the playback of the stimuli. **Figure 4.5** shows the control interface used for this test, which was written using MAX-MSP software. There were a total of six trial pages and the buttons A and B in each page presented the images of CR (crosstalk-off) and LCR (crosstalk-on) in random orders. The stimuli

pair A and B was synchronised and looped so that the subjects could switch between them freely and listen repeatedly.

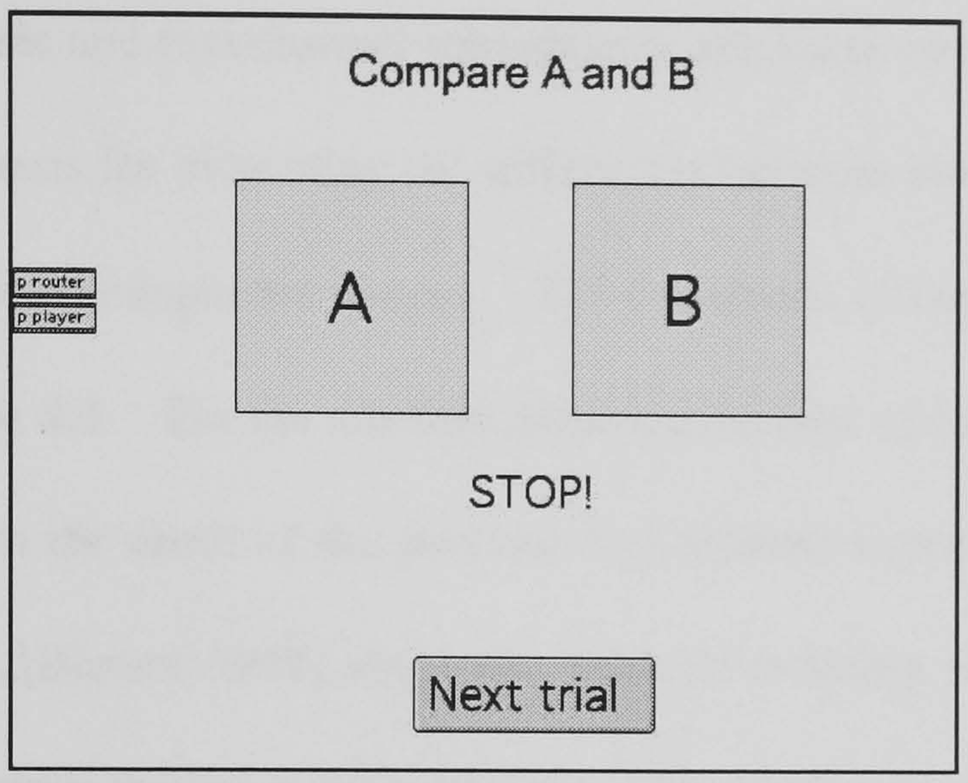


Figure 4.5 Layout of the control interface used for the pair-wise comparison and elicitation of auditory attributes

There were two tasks for the subjects to complete in this test, comprising:

- To define the global set of auditory attributes for the perceived differences between the images of CR and LCR.
- To grade the overall intensities of audibility for those attributes.

The first task was given in order to understand the basic auditory percepts arising from interchannel crosstalk. As mentioned earlier, the subjects were provided with a list of potential attributes and asked to select the ones relevant to the perceived differences. Any additional differences perceived were also to be described using the subjects' own terms and they were unified into common terms by informal discussions between the subjects. The choice of the provided attributes was based on the results of the previous experiment. A number of other spatial or timbral

attributes were also available to choose from various elicitation experiments [Berg and Rumsey 1999, Zacharov and Koivuniemi 2001, Gabrielsson and Sjogren 1979]. However, due to the similarity of the experimental contexts, the attributes perceived between monophonic and two-channel stereophonic attributes were found to form the most appropriate basis for evaluating the differences between two-channel (CR) and three-channel (LCR) stereophonic images. The definitions of the provided attributes are shown in **Table 4.3**. For the attribute meaning the ease of localisation, the term 'source focus' from the result of the previous two-channel experiment was replaced with 'locatedness' [Blauert 1997] since the semantic meaning of the former could well be confused with that of 'source width'. The 'source location' attribute was additionally included because a small degree of source location shift was noticed between the images of CR and LCR in the author's own informal listening test.

The purpose of the second task was to limit the number of attributes to be graded in the next test. Grading all the elicited attributes was considered to be ineffective since minor attributes are likely to have small experimental effects. The 10-point scale shown in **Figure 4.6** was used for the subjects to grade the audibility of the elicited attributes. The degree of audibility might vary for different stimuli, but the grading was to be related to the most audible one.

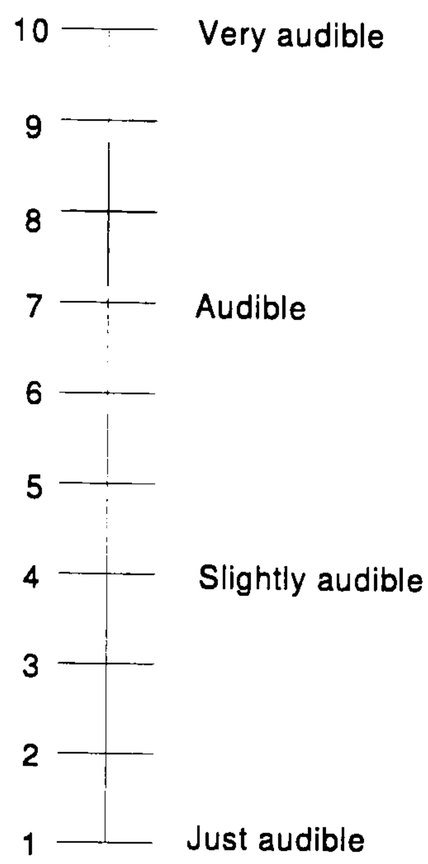


Figure 4.6 Scale used for grading the audibility of each attribute elicited

4.3.2 Results and discussions

As a result of the elicitation test, a total of eleven attributes were elicited from the subjects comprising all seven of the provided attributes and four additional attributes. **Table 4.4** shows the attributes that were elicited, the number of their occurrences, and their audibility indexes. The audibility index represents the average degree of audibility for each attribute, and it was obtained by dividing the sum of the audibility grading values obtained for each attribute by the number of subjects.

According to the results shown in the table, 'source width' is the most audible attribute, having an audibility index of 6.5. The second most audible attribute is shown to be 'locatedness'. The audibility index is 4.7 and this value indicates that the attribute was more than 'slightly audible' according to the semantic labels on the

scale. The audibility indexes of all other attributes are shown to be lower than 4.0. This means that the differences for those attributes were in the range between just audible and slightly audible, which are considered to be minor effects. Therefore, the 'source width' and 'locatedness' attributes, which were graded above the 'slightly audible' level, were finally selected to be used for the next grading test.

Source width	The perceived width of a sound source itself i.e. is one source perceived to be wider than the other?
Source distance	The perceived distance from the listener to a sound source i.e. can the sources be discriminated in terms of their distances?
Source location	The perceived location of a sound source i.e. does the apparent location of the source appear to change?
Locatedness	The easiness of localisation of a sound source i.e. how easy is it to pinpoint the apparent location of a source?
Brightness	The timbral characteristics of a sound depending on the level of high frequencies i.e. bright / dull
Hardness	The timbral characteristics of a sound depending on the level of mid-high frequencies (especially in the range of 2 – 4kHz) i.e. hard / soft
Fullness	The timbral characteristics of a sound depending on the level of low frequencies i.e. full / thin

Table 4.3 Definitions of the auditory attributes provided for selection

Attribute	Occurrences	Audibility index
Source width	7	6.5
Locatedness	6	4.7
Source location	6	3.6
Fullness	5	3.5
Source distance	7	3.1
Hardness	3	2.3
Brightness	5	1.4
Diffuseness	1	1.3
Naturalness	1	1.3
Envelopment	1	0.7
Phasiness	1	0.5

Table 4.4 Attribute group, number of occurrences and audibility index obtained for the differences perceived between the images of CR and LCR with cello, bongo and speech sources

4.4 Experiment Part 2: Grading of Perceptual Effect

4.4.1 Listening test method

The grading experiment was designed based on the result of the elicitation experiment and required subjects to grade the perceived difference between the images of CR and LCR. It was considered that the locatedness and source width attributes might have adjacent characteristics, and therefore a proximity error might be caused if they were graded simultaneously in the same session. In other words, they might be graded as unnecessarily correlated due to a possible biasing effect between each other. In fact, this might have been the case for the strong correlation between source width and source focus attributes that was found in the previous two-channel experiment.

Therefore, it was decided to test each attribute individually in order to avoid a psychological bias. To this end, the whole experiment was divided into two sub-tests: locatedness change test and source width change test.

A total of 36 stimulus-pairs were created for comparison. In each attribute test, each subject was asked to compare the 36 stimulus pairs twice, and therefore a total of 72 trial sets were produced. Grading all the 72 trials in one session might have caused experimental errors due to subject fatigue, so the 72 trials were distributed evenly into three separate sessions by the type of acoustic condition, each session thus containing 24 trials. In order to avoid such psychological errors as contrast, convergence and anticipation errors, which were introduced in Section 3.3.1, the order of presentation for the trials was randomised for each session and for each subject. The orders of sessions and attribute tests were also arranged differently for each subject.

This experiment used a 7-point continuous grading scale labelled from -30 to 30. The reason for using a continuous grading scale rather than a semantic differential scale was explained in detail in Section 3.3.1. The ends of the scale for the locatedness attribute were labelled as 'more located – less located', and those for the source width attribute were labelled as 'wider – narrower'.

An example of the control interface used for the experiment is shown in **Figure 4.7**. As can be seen, a vertical slider was used for grading, without showing the value to the subjects. The graded value was saved automatically by clicking the 'next trial' button. The question presented to the subjects was as shown in the figure, but the

order of the crosstalk-off (CR) and crosstalk-on (LCR) images presented by the buttons 'A' and 'B' was randomised for each trial. Prior to the main grading tests a few familiarisation trials were provided to the subjects in order to encourage them to use consistent scale ranges and also avoid central tendency errors [Stone and Sidel 1993]. Six representative stimuli comprising the extreme arrays of 1 and 4 combined with three sound sources were selected for the familiarisation trials.

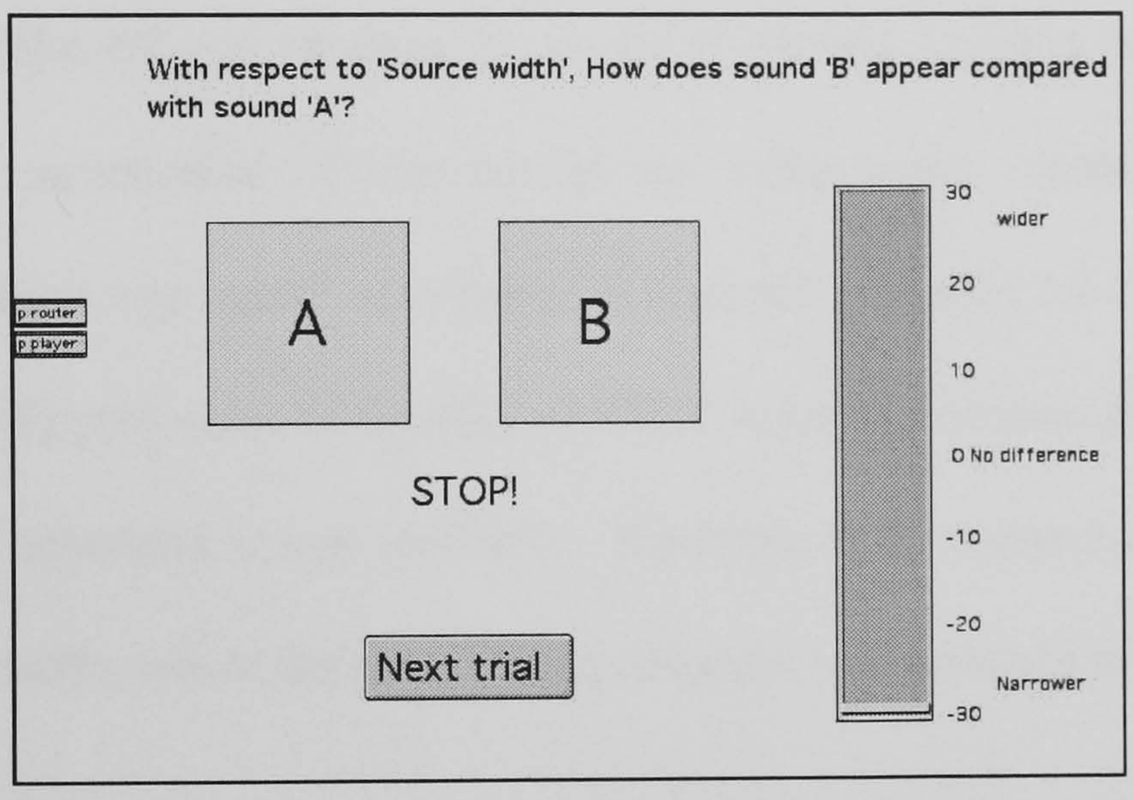


Figure 4.7 Layout of the control interface used for the pairwise comparison and grading for source width attribute

4.4.2 Statistical analysis

A repeated measure ANOVA (RM ANOVA) was carried out for statistical analysis of the data obtained from the grading experiment, since all conditions were tested within the same group of subjects. The independent variables were the type of acoustic condition, the type of sound source and the type of microphone array. The dependent variable was the grading of the perceived magnitude of difference between

crosstalk-off and crosstalk-on sounds on a scale of -30 to 30. There were a total of 576 observations, consisting of 16 observations for each of the 36 'acoustic condition – sound source type – microphone array type' combinations obtained from eight subjects. Prior to the RM ANOVA test, the original grading data were normalised based on the ITU-R BS.1116 recommendation [1994] for the reason described in Section 3.3.2. Mauchly's test of sphericity was carried out for each attribute test in order to examine the assumption of sphericity. The results are shown in **Tables 4.5** and **4.6**. **Tables 4.7** and **4.8** show the results of the RM ANOVA for each attribute test. In the presentation of the results the independent variables are termed 'acoustic', 'source' and 'array'. As explained in detail in Section 3.3.2, the 'sphericity assumed' significance value in the RM ANOVA result can be used provided that the assumption of sphericity is met (p>0.05). However, if the assumption of sphericity is violated (p<0.05), one of the corrected significance values should be used instead.

Measure: MEASURE_1

Within Subjects Effec	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
ACOUSTIC	.621	6.675	2	.036	.725	.782	.500
SOURCE	.802	3.084	2	.214	.835	.927	.500
ARRAY	.541	8.418	5	.136	.751	.891	.333
ACOUSTIC * SOURCE	.365	13.505	9	.144	.666	.823	.250
ACOUSTIC * ARRAY	.095	30.143	20	.075	.600	.813	.167
SOURCE * ARRAY	.019	50.972	20	.000	.444	.549	.167

Table 4.5 Mauchly's test of sphericity for source width change

Measure: MEASURE_1

Within Subjects Effec	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
ACOUSTIC	.893	1.582	2	.453	.903	1.000	.500
SOURCE	.827	2.651	2	.266	.853	.951	.500
ARRAY	.071	36.232	5	.000	.442	.468	.333
ACOUSTIC * SOURCE	.449	10.736	9	.298	.723	.915	.250
ACOUSTIC * ARRAY	.022	48.742	20	.000	.372	.440	.167
SOURCE * ARRAY	.152	24.096	20	.252	.596	.805	.167

Table 4.6 Mauchly's test of sphericity for locatedness change

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
ACOUSTIC	Sphericity Assumed	13.014	2	6.507	.344	.711	.022
	Greenhouse-Geisse	13.014	1.450	8.974	.344	.644	.022
	Huynh-Feldt	13.014	1.565	8.317	.344	.660	.022
	Lower-bound	13.014	1.000	13.014	.344	.566	.022
Error(ACOUSTIC)	Sphericity Assumed	566.708	30	18.890			
	Greenhouse-Geisse	566.708	21.751	26.054			
	Huynh-Feldt	566.708	23.471	24.146			
	Lower-bound	566.708	15.000	37.781			
SOURCE	Sphericity Assumed	495.003	2	247.502	6.733	.004	.310
	Greenhouse-Geisse	495.003	1.670	296.437	6.733	.007	.310
	Huynh-Feldt	495.003	1.854	266.956	6.733	.005	.310
	Lower-bound	495.003	1.000	495.003	6.733	.020	.310
Error(SOURCE)	Sphericity Assumed	1102.719	30	36.757			
	Greenhouse-Geisse	1102.719	25.048	44.025			
	Huynh-Feldt	1102.719	27.814	39.646			
	Lower-bound	1102.719	15.000	73.515			
ARRAY	Sphericity Assumed	17141.102	3	5713.701	156.563	.000	.913
	Greenhouse-Geisse	17141.102	2.254	7603.759	156.563	.000	.913
	Huynh-Feldt	17141.102	2.673	6412.793	156.563	.000	.913
	Lower-bound	17141.102	1.000	17141.102	156.563	.000	.913
Error(ARRAY)	Sphericity Assumed	1642.259	45	36.495			
	Greenhouse-Geisse	1642.259	33.814	48.567			
	Huynh-Feldt	1642.259	40.094	40.960			
	Lower-bound	1642.259	15.000	109.484			
ACOUSTIC * SOURCE	Sphericity Assumed	40.007	4	10.002	.531	.714	.034
	Greenhouse-Geisse	40.007	2.663	15.022	.531	.643	.034
	Huynh-Feldt	40.007	3.292	12.153	.531	.680	.034
	Lower-bound	40.007	1.000	40.007	.531	.478	.034
Error(ACOUSTIC*SOURCE)	Sphericity Assumed	1130.771	60	18.846			
	Greenhouse-Geisse	1130.771	39.948	28.306			
	Huynh-Feldt	1130.771	49.378	22.900			
	Lower-bound	1130.771	15.000	75.385			
ACOUSTIC * ARRAY	Sphericity Assumed	167.944	6	27.991	2.337	.038	.135
	Greenhouse-Geisse	167.944	3.602	46.626	2.337	.073	.135
	Huynh-Feldt	167.944	4.881	34.409	2.337	.052	.135
	Lower-bound	167.944	1.000	167.944	2.337	.147	.135
Error(ACOUSTIC*ARRAY)	Sphericity Assumed	1078.111	90	11.979			
	Greenhouse-Geisse	1078.111	54.030	19.954			
	Huynh-Feldt	1078.111	73.212	14.726			
	Lower-bound	1078.111	15.000	71.874			
SOURCE * ARRAY	Sphericity Assumed	630.788	6	105.131	8.097	.000	.351
	Greenhouse-Geisse	630.788	2.663	236.855	8.097	.000	.351
	Huynh-Feldt	630.788	3.292	191.621	8.097	.000	.351
	Lower-bound	630.788	1.000	630.788	8.097	.012	.351
Error(SOURCE*ARRAY	Sphericity Assumed	1168.601	90	12.984			
	Greenhouse-Geisse	1168.601	39.948	29.253			
	Huynh-Feldt	1168.601	49.378	23.667			
	Lower-bound	1168.601	15.000	77.907			

Table 4.7 Results of repeated measure ANOVA test for source width change

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
ACOUSTIC	Sphericity Assumed	634.292	2	317.146	7.063	.003	.320
	Greenhouse-Geisse	634.292	1.807	351.025	7.063	.004	.320
	Huynh-Feldt	634.292	2.000	317.146	7.063	.003	.320
	Lower-bound	634.292	1.000	634.292	7.063	.018	.320
Error(ACOUSTIC)	Sphericity Assumed	1347.042	30	44.901			
	Greenhouse-Geisse	1347.042	27.105	49.698			
	Huynh-Feldt	1347.042	30.000	44.901			
	Lower-bound	1347.042	15.000	89.803			
SOURCE	Sphericity Assumed	24.385	2	12.193	.457	.637	.030
	Greenhouse-Geisse	24.385	1.706	14.296	.457	.608	.030
	Huynh-Feldt	24.385	1.902	12.818	.457	.628	.030
	Lower-bound	24.385	1.000	24.385	.457	.509	.030
Error(SOURCE)	Sphericity Assumed	799.948	30	26.665			
	Greenhouse-Geisse	799.948	25.586	31.265			
	Huynh-Feldt	799.948	28.537	28.032			
	Lower-bound	799.948	15.000	53.330			
ARRAY	Sphericity Assumed	14067.505	3	4689.168	87.488	.000	.854
	Greenhouse-Geisse	14067.505	1.325	10619.696	87.488	.000	.854
	Huynh-Feldt	14067.505	1.404	10022.507	87.488	.000	.854
	Lower-bound	14067.505	1.000	14067.505	87.488	.000	.854
Error(ARRAY)	Sphericity Assumed	2411.911	45	53.598			
	Greenhouse-Geisse	2411.911	19.870	121.385			
	Huynh-Feldt	2411.911	21.054	114.559			
	Lower-bound	2411.911	15.000	160.794			
ACOUSTIC * SOURCE	Sphericity Assumed	163.104	4	40.776	2.901	.029	.162
	Greenhouse-Geisse	163.104	2.894	56.367	2.901	.047	.162
	Huynh-Feldt	163.104	3.659	44.576	2.901	.034	.162
	Lower-bound	163.104	1.000	163.104	2.901	.109	.162
Error(ACOUSTIC*SOUF CE)	Sphericity Assumed	843.229	60	14.054			
	Greenhouse-Geisse	843.229	43.404	19.428			
	Huynh-Feldt	843.229	54.885	15.364			
	Lower-bound	843.229	15.000	56.215			
ACOUSTIC * ARRAY	Sphericity Assumed	297.583	6	49.597	3.113	.008	.172
	Greenhouse-Geisse	297.583	2.230	133.420	3.113	.052	.172
	Huynh-Feldt	297.583	2.638	112.805	3.113	.043	.172
	Lower-bound	297.583	1.000	297.583	3.113	.098	.172
Error(ACOUSTIC*ARRA Y)	Sphericity Assumed	1433.750	90	15.931			
	Greenhouse-Geisse	1433.750	33.456	42.854			
	Huynh-Feldt	1433.750	39.571	36.233			
	Lower-bound	1433.750	15.000	95.583			
SOURCE * ARRAY	Sphericity Assumed	146.698	6	24.450	2.123	.058	.124
	Greenhouse-Geisse	146.698	3.574	41.042	2.123	.098	.124
	Huynh-Feldt	146.698	4.830	30.371	2.123	.074	.124
	Lower-bound	146.698	1.000	146.698	2.123	.166	.124
Error(SOURCE*ARRAY	Sphericity Assumed	1036.302	90	11.514			
	Greenhouse-Geisse	1036.302	53.615	19.329			
	Huynh-Feldt	1036.302	72.454	14.303			
	Lower-bound	1036.302	15.000	69.087			

Table 4.8 Results of repeated measure ANOVA test for locatedness change

4.4.3 Results

4.4.3.1 Source width change

The results of the RM ANOVA test shown in **Table 4.7** indicate that microphone array is the most significant factor in source width change (p = 0.000). The main effect of sound source is also highly significant (p = 0.004), but the effect size is small (0.310) compared to that of microphone array (0.913). On the other hand, acoustic condition does not have a significant effect (p = 0.644). With respect to the interactions between each factor, the largest effect is observed between source and microphone array (p = 0.000), followed by between acoustic condition and microphone array (p = 0.038). The acoustic*source interaction is shown to be insignificant (p = 0.714).

Figure 4.8 shows the mean values and 95% confidence intervals for each microphone array. It can be seen that array 4 has the largest increase of source width when affected by the crosstalk signal, followed by array 3, 2 and 1 in order. Also, there is no overlap of 95% confidence intervals between any pair of arrays, thus causing highly significant differences between all the arrays (see **Table 4.9**).

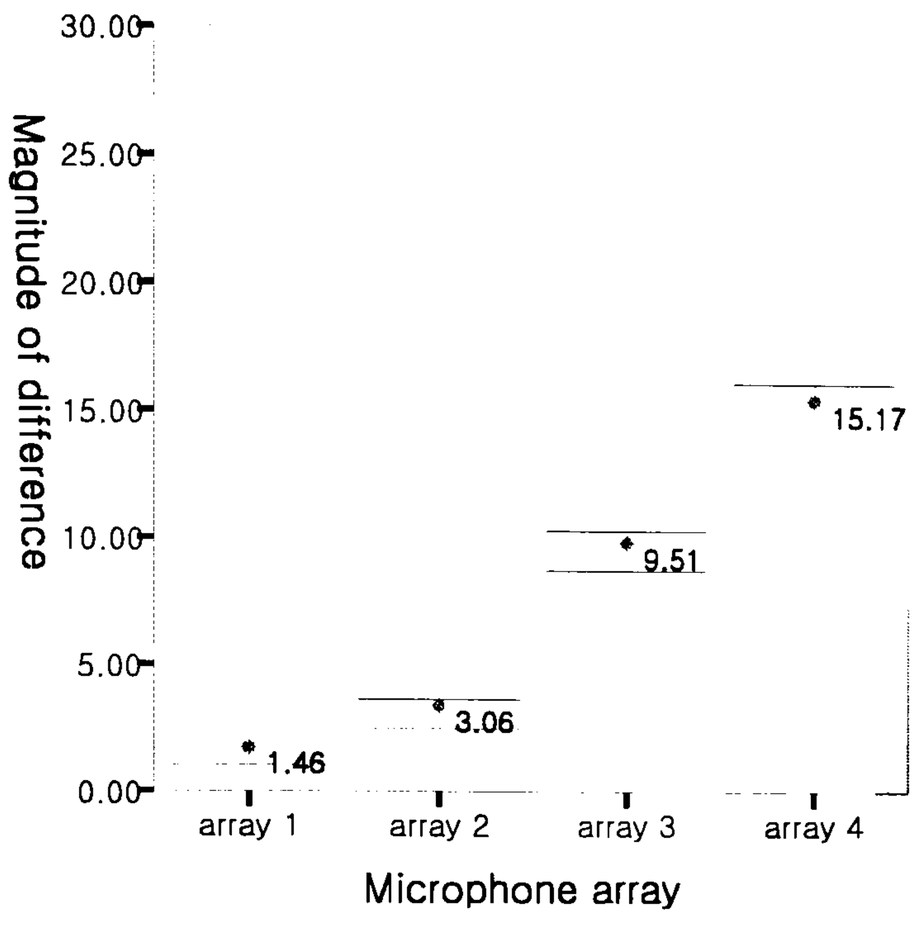


Figure 4.8 Mean value and associated 95% confidence intervals of the grade of source width difference between the crosstalk-off (CR) and crosstalk-on (LCR) images for each microphone array

Measure: MEASURE_1

(I) ARRAY	(J) ARRAY	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
1	2	-1.597	.461	.021	-2.996	-.199
	3	-8.056	.693	.000	-10.159	-5.952
	4	-13.715	.881	.000	-16.391	-11.039
2	1	1.597	.461	.021	.199	2.996
	3	-6.458	.714	.000	-8.625	-4.292
	4	-12.118	.761	.000	-14.429	-9.807
3	1	8.056	.693	.000	5.952	10.159
	2	6.458	.714	.000	4.292	8.625
	4	-5.660	.695	.000	-7.771	-3.549
4	1	13.715	.881	.000	11.039	16.391
	2	12.118	.761	.000	9.807	14.429
	3	5.660	.695	.000	3.549	7.771

Table 4.9 Result of multiple pairwise comparison between each microphone array for source width change

Figure 4.9 shows the mean values and 95% confidence intervals for each sound source. It appears that the speech source is outstanding compared to the cello and bongo sources. The multiple pairwise comparisons between each sound source

indicated in **Table 4.10** confirm the significant difference between the speech and the other sources. The cello and bongo are shown to have the same effect (p = 1.000).

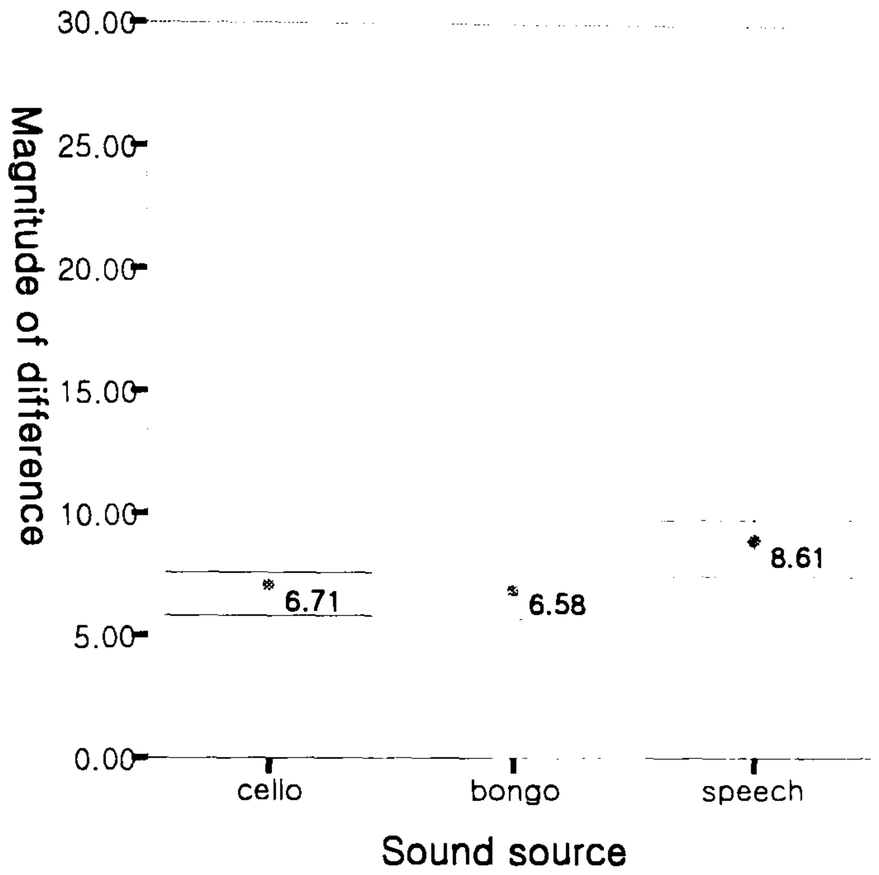


Figure 4.9 Mean value and associated 95% confidence intervals of the grade of source width difference between the crosstalk-off (CR) and crosstalk-on images (LCR) for each sound source

Measure: MEASURE_1

(I) SOURCE	(J) SOURCE	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Cello	Bongo	.125	.463	1.000	-1.122	1.372
	Speech	-1.901	.699	.047	-3.784	-.018
Bongo	Cello	-.125	.463	1.000	-1.372	1.122
	Speech	-2.026	.667	.025	-3.824	-.228
Speech	Cello	1.901	.699	.047	.018	3.784
	Bongo	2.026	.667	.025	.228	3.824

Table 4.10 Result of multiple pairwise comparison between each sound source for source width change

The main effect of the acoustic condition on source width change is shown in **Figure 4.10**. Adding multiple reflections and reverberation to an anechoic sound might have increased the source widths for both images of CR and LCR. The insignificant

main effect means that the magnitude of the individual increase was similar. This result suggests that the source widening effect of interchannel crosstalk is independent of the acoustic condition of recording space.

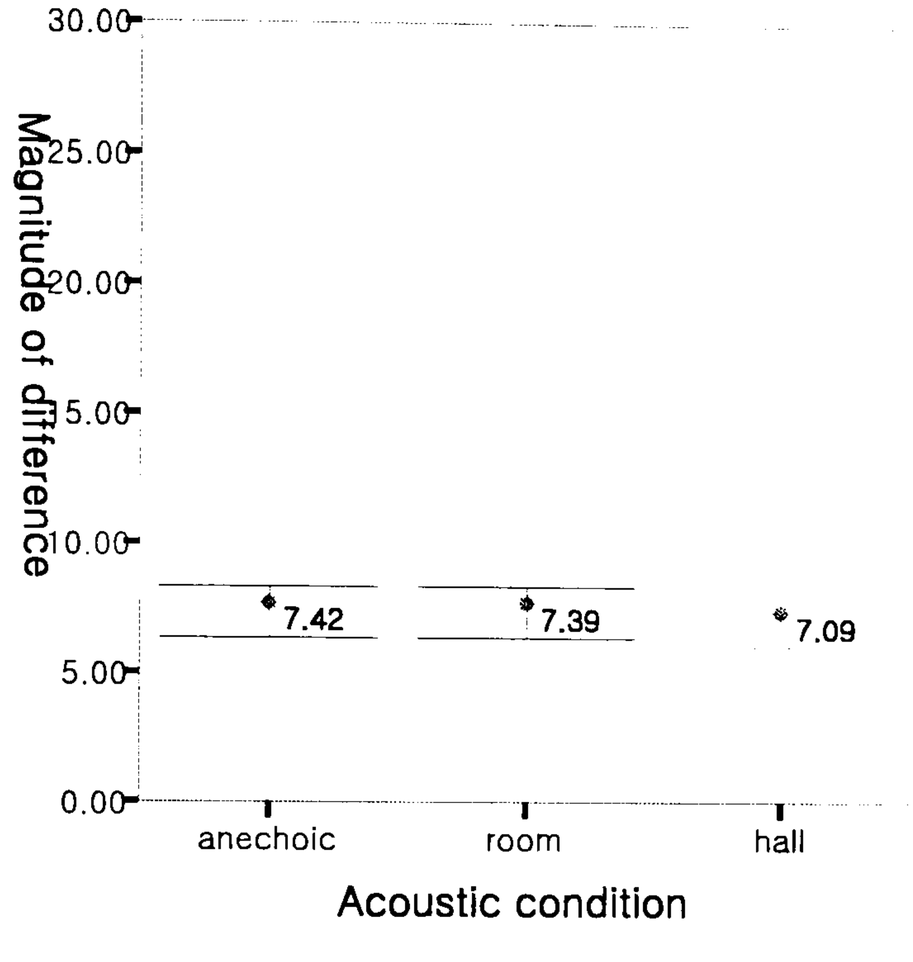


Figure 4.10 Mean value and associated 95% confidence intervals of the grade of source width difference between the crosstalk-off (CR) and crosstalk-on images (LCR) for each acoustic condition

The source*array interaction is shown in **Figure 4.11**. Even though this interaction effect was found to be significant, the order of microphone array in the magnitude of change was the same for all sound sources. Also, since the estimated effect size is only 0.351 (Partial Eta Squared), this interaction could possibly be ignored. The acoustic*array interaction was also found to be significant, but again the estimated effect size is shown to be very small (0.135), and the order of microphone array stays the same regardless of the acoustic condition (see **Figure 4.12**). Therefore, this interaction could be also ignored. The acoustic*source interaction was found to be insignificant.

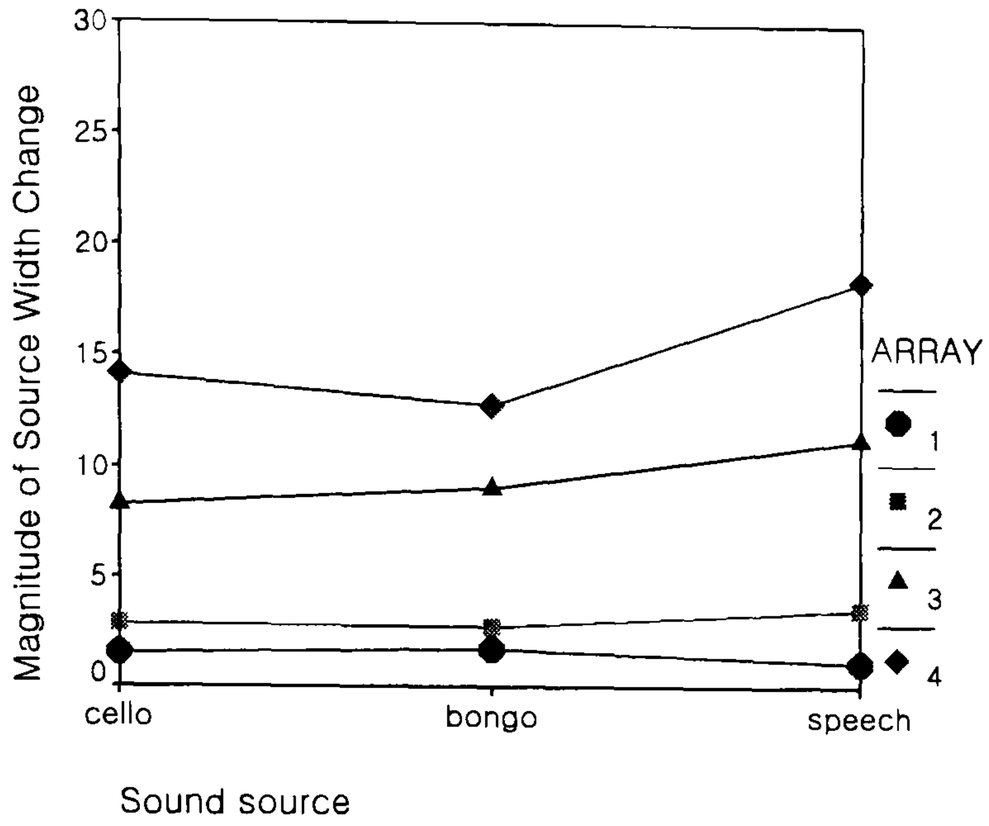


Figure 4.11 Interaction between microphone array and sound source

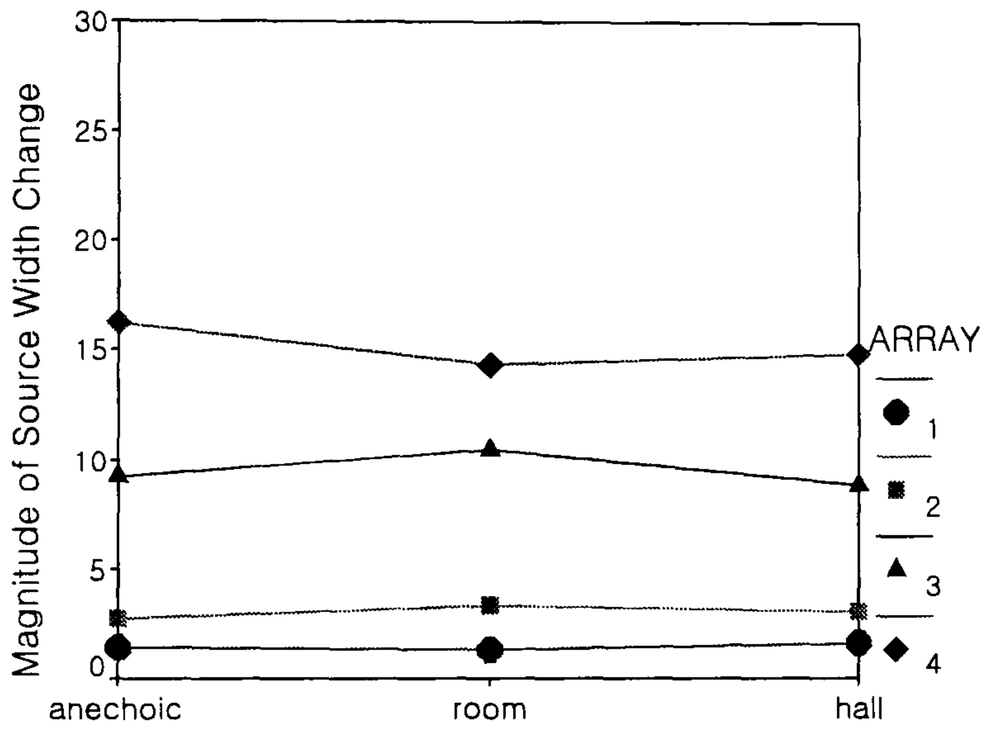


Figure 4.12 Interaction between microphone array and acoustic condition

4.4.3.2 Locatedness change

Taking an overview of the results of the RM ANOVA test indicated in **Table 4.8**, 'microphone array' has the most significant effect on locatedness change (the significance value p is 0.000, and the estimated size of effect is 0.854). The main effect of 'acoustic condition' is shown to be significant (p = 0.003), but its

experimental effect (0.320) is much smaller than that of microphone array. 'Sound source' does not have a significant main effect (p = 0.637), which means that the magnitude of locatedness change was similar for all sound sources. The largest interaction effect is observed between acoustic and source (p = 0.029). The interaction effect between acoustic and array can be judged differently depending on which corrected significance value is used because sphericity is violated. That is, the Hyunh-Feldt value (0.043) indicates significance while the Greenhouse-Geisser value (0.052) does not. However, the small partial eta-squared values for acoustic*source (0.162) and acoustic*array (0.172) suggest that the experimental effects of those interactions are relatively minor regardless of the significance value. The source*array interaction is shown to be insignificant (p = 0.058).

Figure 4.13 shows the mean value and associated 95% confidence intervals of the grade given for each microphone array. It can firstly be seen that the magnitude of locatedness change between CR and LCR increases as the array number increases from 1 to 4. This basically means that the most 'time-difference' based array gave rise to the greatest effect, whereas the most 'intensity-difference' based array gave rise to the smallest effect. It is interesting to note that the magnitude of locatedness change tends to increase almost linearly from array 2 to array 4. It can also be observed that there is no overlap between any pair of arrays in 95% confidence interval, which means that the differences between those four microphone arrays were clearly distinguished by the subjects. The significant difference between each array is confirmed by the result of the multiple pairwise comparison test shown in **Table**

4.11 (all *p* values are 0.000). This result suggests that the intensity and delay time of the crosstalk signal is a crucial factor governing the perception of locatedness.

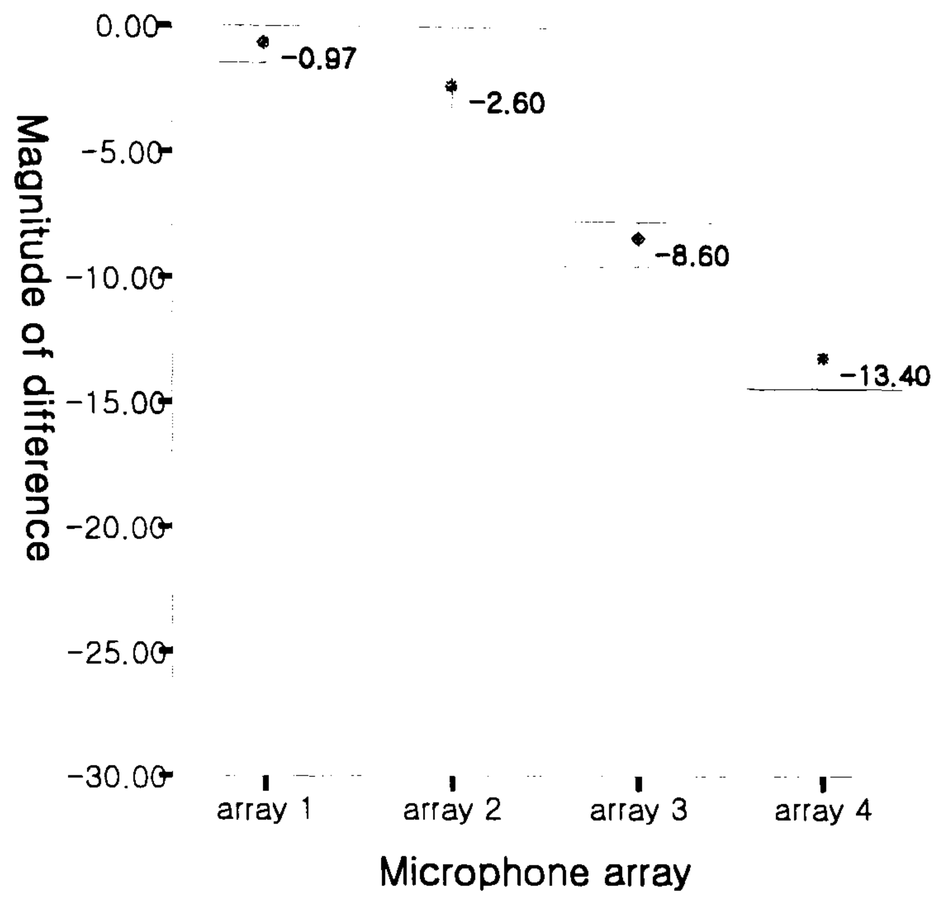


Figure 4.13 Mean value and associated 95% confidence intervals of the grade of locatedness difference between the crosstalk-off (CR) and crosstalk-on images (LCR) for each microphone array

Measure: MEASURE_1

(I) ARRAY	(J) ARRAY	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
1	2	1.625	.287	.000	.755	2.495
	3	7.625	.766	.000	5.300	9.950
	4	12.424	1.251	.000	8.626	16.221
2	1	−1.625	.287	.000	−2.495	−.755
	3	6.000	.677	.000	3.944	8.056
	4	10.799	1.117	.000	7.409	14.189
3	1	−7.625	.766	.000	−9.950	−5.300
	2	−6.000	.677	.000	−8.056	−3.944
	4	4.799	.727	.000	2.590	7.007
4	1	−12.424	1.251	.000	−16.221	−8.626
	2	−10.799	1.117	.000	−14.189	−7.409
	3	−4.799	.727	.000	−7.007	−2.590

Table 4.11 Result of multiple pairwise comparisons between each microphone array for locatedness change

The plot for the effect of each acoustic condition is shown in **Figure 4.14**. Even though the graph shows a noticeable decreasing pattern in the magnitude of difference

as the microphone array changes from 1 to 4, there is a large overlap between each nearby condition in 95% confidence intervals, which might have led to the relatively small effect size (0.320). The result of a pairwise comparison test shown in **Table 4.12** indicates that the only significant difference is between the anechoic and hall conditions ($p = 0.003$).

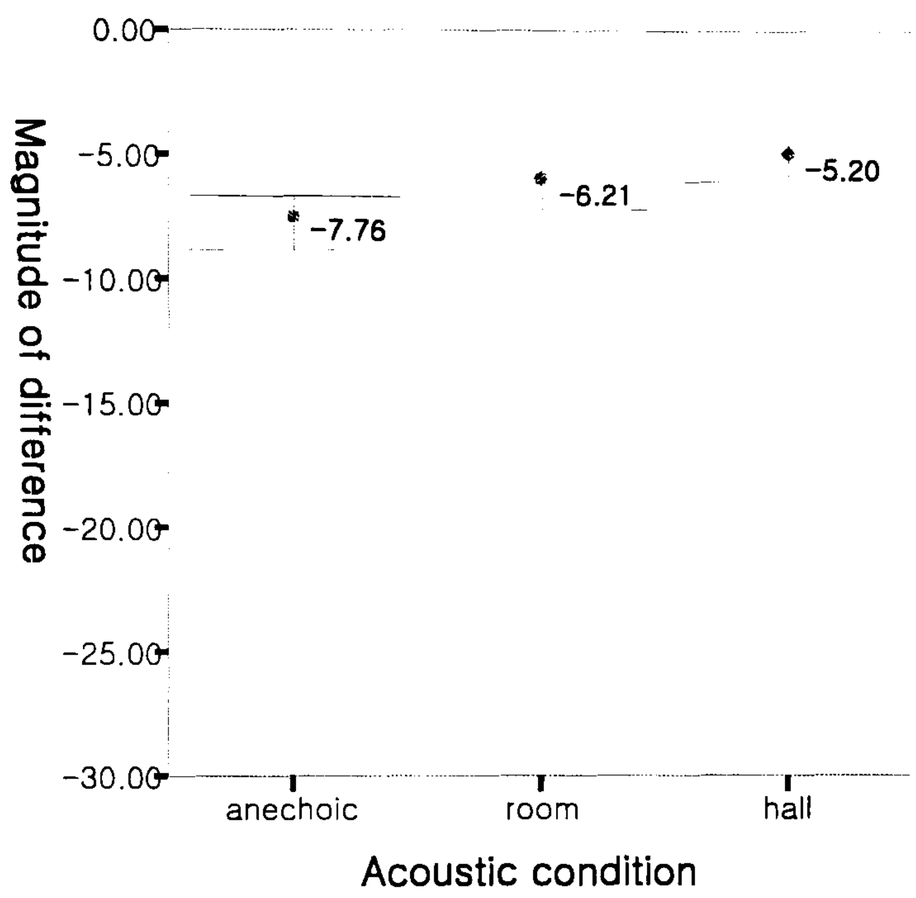


Figure 4.14 Mean value and associated 95% confidence intervals of the grade of locatedness difference between the crosstalk-off (CR) and crosstalk-on images (LCR) for each acoustic condition

Measure: MEASURE_1

(I) ACOUSTIC	(J) ACOUSTIC	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Anechoic	Room	-1.542	.787	.207	-3.662	.579
	Hall	-2.552	.614	.003	-4.206	-.898
Room	Anechoic	1.542	.787	.207	-.579	3.662
	Hall	-1.010	.638	.402	-2.728	.707
Hall	Anechoic	2.552	.614	.003	.898	4.206
	Room	1.010	.638	.402	-.707	2.728

Table 4.12 Result of multiple pairwise comparison between each acoustic condition for locatedness change

The mean values and associated 95% confidence intervals of the normalised data for each sound source are shown in **Figure 4.15**. As can be seen, all sound sources have small differences in mean values and large overlaps in 95% confidence intervals.

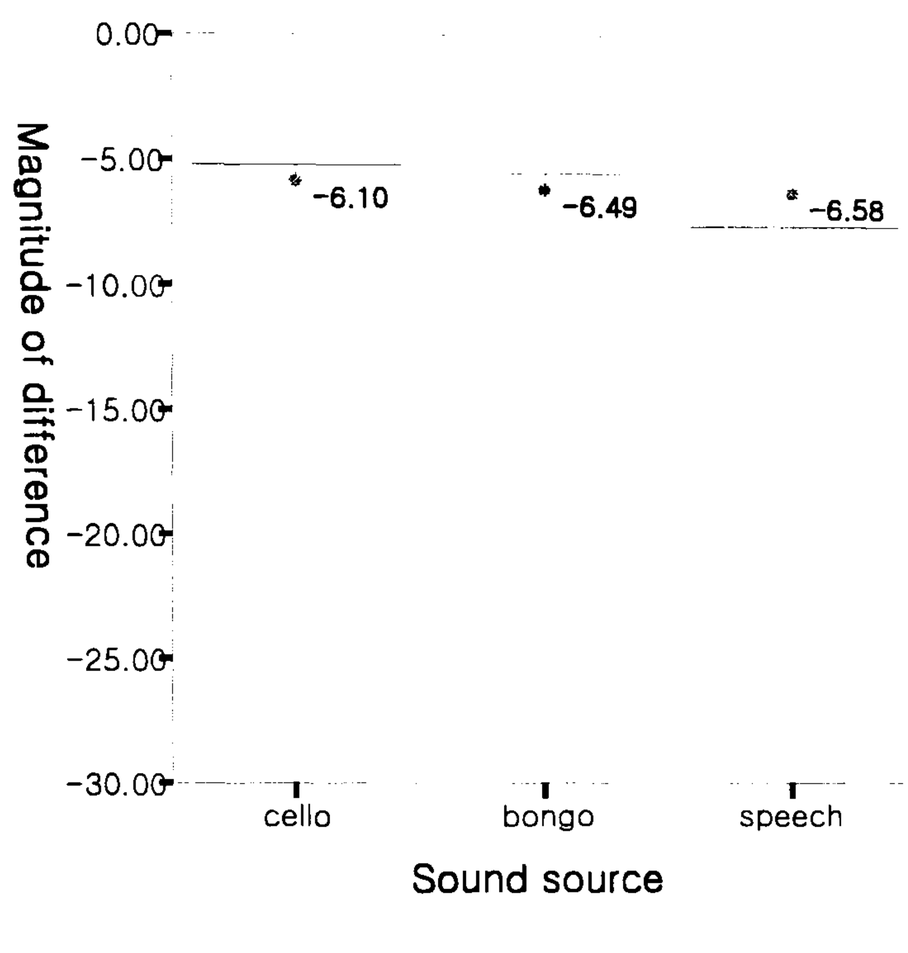


Figure 4.15 Mean value and associated 95% confidence intervals of the grade of locatedness difference between the crosstalk-off (CR) and crosstalk-on images (LCR) for each sound source

Figure 4.16 shows the interaction graph between acoustic condition and sound source. There are significant contrasts observed between the anechoic and hall conditions when cello is compared to bongo ($p = 0.011$), and when cello is compared to speech ($p = 0.028$). These contrasts mean that the difference between the cello and the bongo (or speech) in the anechoic condition is significantly bigger than the difference between them in the hall condition. A more detailed interaction can be found in the relationship between each sound source for each acoustic condition. For this investigation, a 'Paired-Samples T-test' was performed, and the result summary is shown in **Table 4.13**. Firstly, in the comparison between sound sources for the

anechoic condition, it can be seen that there are significant differences between cello and bongo (p = 0.007), and between cello and speech (p = 0.048), although the main effect of sound source is not significant (when acoustic and array are ignored). Bongo and speech do not have a significant difference. **Figure 4.17** shows the acoustic*array interaction graph. Arrays 3 and 4 have a significant difference when the room and hall conditions are compared. Also, arrays 2 and 3 are significantly different when the anechoic and hall conditions are compared. Nevertheless, this effect might be ignored since the order of microphone arrays is the same for all acoustic conditions, and the size of experimental effect is small. This result seems to suggest that the significance of the intensity of the crosstalk signal does not change regardless of the acoustic condition of recording space. The source*array interaction was found to be insignificant.

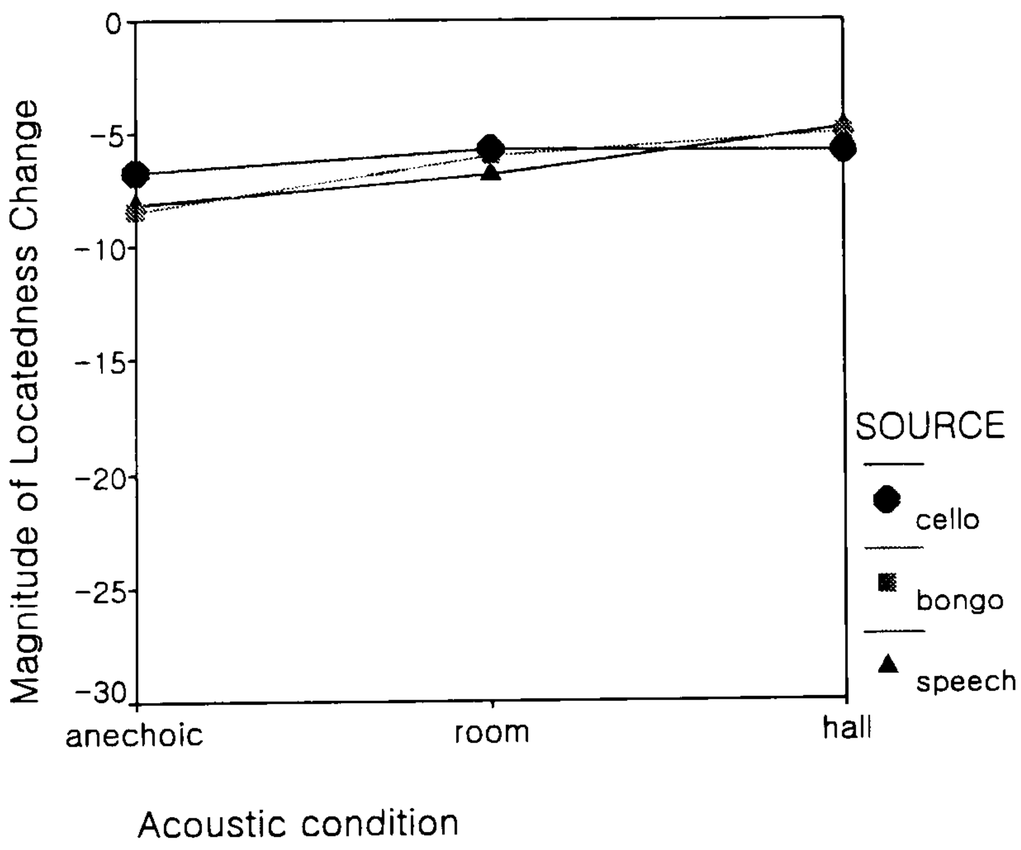


Figure 4.16 Interaction between acoustic condition and sound source

		t	Sig. (2-tailed)
Anechoic	Cello - Bongo	2.750	.007
Anechoic	Cello - Speech	-.631	.529
Anechoic	Bongo - Speech	-2.827	.005
Room	Cello - Bongo	.662	.509
Room	Cello - Speech	-.283	.778
Room	Bongo - Speech	-.863	.390
Hall	Cello - Bongo	-1.376	.171
Hall	Cello - Speech	-3.103	.002
Hall	Bongo - Speech	-1.849	.067

Table 4.13 Result table of paired samples T-test for acoustic condition and sound source

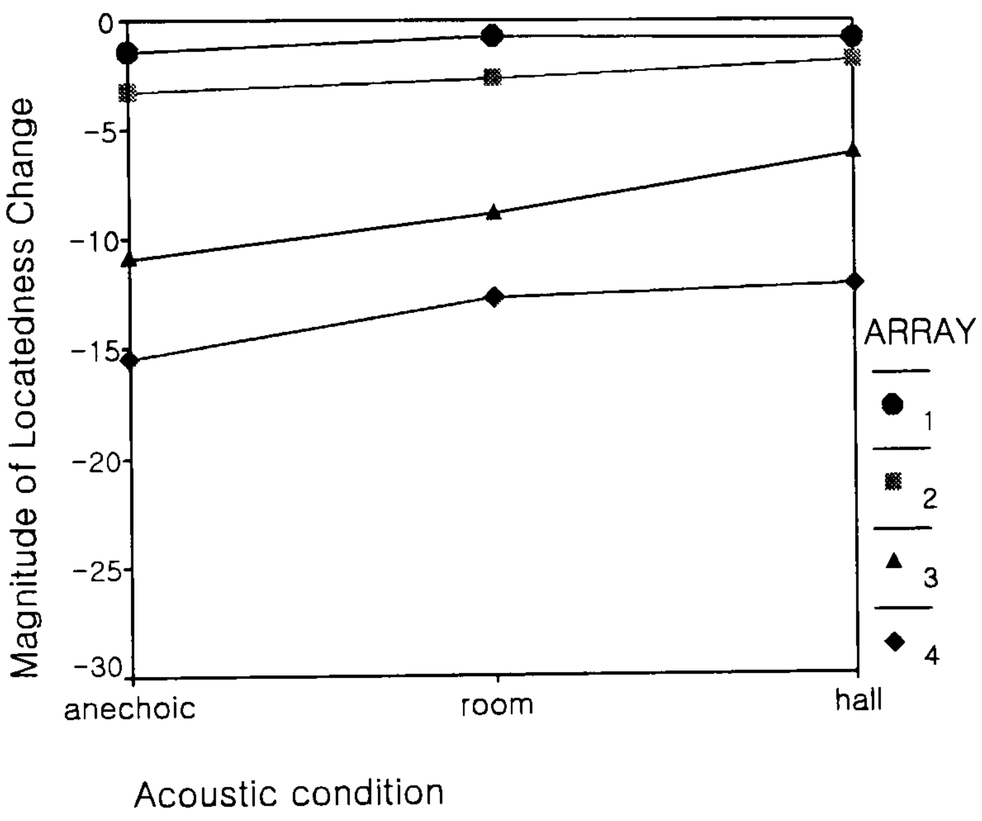


Figure 4.17 Interaction between microphone array and acoustic condition

4.4.4 Discussions

4.4.4.1 Discussion of the results for the individual attributes

The result showing that the type of microphone array had a significant effect suggests that the effect of interchannel crosstalk on source widening and locatedness

decreasing becomes greater as a more spaced microphone technique is used, in other words as the ratio of time difference to intensity difference increases. It also suggests that this effect can be almost ignored when a more coincident type of microphone technique is used. Therefore, this leads to a discussion on the influence of interchannel time and intensity differences between L and C. The basis for this discussion might be found in the result of the previous two-channel investigation, showing that two-channel stereophonic images were perceived to be wider and more focused compared to the corresponding monophonic image and the magnitude of this effect became greater as the ratio of time difference to intensity difference was increased (see Section 3.3.4.1). This seems to hold true in the case of the current experimental conditions. Since the microphone technique used in this experiment was a near-coincident type, the ratio between the interchannel time difference (ICTD) and interchannel intensity difference (ICID) changes for each array style. That is, the ICTD between L and C in the arrays 1 to 4 increases from 0.5ms to 1.1ms, while the corresponding interchannel intensity difference (ICID) decreases from 20.5dB to 4.6dB (see **Table 4.1**). The decrease in the ICID between L and C means an increase in the intensity of the crosstalk signal. This suggests that the crosstalk signal L in the array 4 might not only have been most audible due to its greatest intensity but might also have caused the largest change in the magnitude of ITD fluctuations between the ear input signals of CR and LCR, thus leading to the largest source width and locatedness change. For the array 1, however, the ICID between L and C is 20.5dB and this is greater than the usual psychoacoustic values required for full phantom image shifts in two-channel stereophonic reproduction as indicated in **Figure 1.1**. This means that the effects of crosstalk L would have been barely

detectable regardless of the amount of ICTD. In fact, from the visual indications in **Figure 4.8**, the crosstalk effects for arrays 1 and 2 appear to be very small compared to those for arrays 3 and 4. The above discussion might suggest that a more widely spaced three-channel microphone array will tend to give rise to a greater effect of interchannel crosstalk on source widening as it will always have greater ITD fluctuations and greater intensity of the crosstalk signal due to the nature of the near-coincident microphone technique design requiring a trade-off between interchannel time and intensity differences. Conversely, it might also suggest that in order to minimise the effects of interchannel crosstalk in the design of three-channel microphone techniques, one should pursue a more coincident style of microphone technique by shortening the delay time and increasing the intensity difference between channels.

It was shown that the effect of sound source type was significant for source width increase due to interchannel crosstalk. In particular, the source width increase for the speech signal was found to be significantly greater than that for the cello or bongo source and this might be related to the frequency components of these sound sources. In section 2.3.2.2, it was discussed that the findings of reflection studies relating to the effects of frequency components on the perceived source width increase were contradictory. For example, Hidaka *et al* [1995] reported that for orchestral music sources, frequencies below 355Hz would become most significant for increasing perceived source width in a concert hall. Morimoto and Maekawa [1988] found that for a noise signal, frequencies of sound source above 510Hz became less effective for source width increase than the lower frequencies and frequencies around 100 – 200Hz

resulted in an especially marked increase. These findings suggest the significance of low frequency energy on source width perception. However, Barron and Marshall [1981] found that for orchestral music sources, 'source broadening' was mainly governed by middle frequencies of the reflection around 1000 – 2000Hz while 'envelopment' was related to the lower frequencies, although it seems that both attributes were related to the source itself from the authors' definitions of the terms. Additionally, Blauert and Lindemann [1986] reported that all frequency components of reflection would contribute to the perception of source width. Despite this lack of definite results, it can be at least suggested from the above literature that the greater source width increasing effect of the speech source might have resulted from the broad frequency range and the rich frequency components of the source. The frequency spectrum of the cello and speech sources shown in **Figure 4.3** clearly shows the dominance of speech over cello in terms of spectral richness. The speech source also has greater low frequency energy around 100Hz, which might support Morimoto and Maekawa's findings.

In addition to this, the onset time of the speech source might also have been taken into account for the perception of source width. **Figure 4.3** shows that the frequency spectrum of the bongo source is also reasonably rich although its low frequency energies around 100Hz appear to be weaker than those of the speech. However, due to its strong transient nature, the crosstalk signal for the bongo source would have produced a smaller degree of interaural fluctuations than that for the speech source during the onset and this might have caused the significant difference between the two sources in perceived source width. In other words, the interaction between the

wanted and crosstalk signals in the onset region was shorter for the bongo source than for the speech. In order to show this aspect visually, **Figure 4.18** illustrates two cases of simulated interactions between leading and lagging signals for sound sources having different onset times with the same ongoing and decay times. As can be seen, when the delay time is constant, while the size of time region for the ongoing and offset interaural fluctuation is the same for both sources, that for the onset fluctuation differs depending on the onset time of the source. Therefore, source B, which is more transient, has a smaller magnitude of interaural fluctuation than source A. According to Griesinger [1996]'s hypothesis about the cognitive perception of spatial impression, which was introduced in Section 2.3.1.2, source width perception in a concert hall is related to the interaction between the direct and reflected signals in the onset region. Based on this, the difference between the bongo and speech sources in perceived source width is considered to be caused by the different magnitudes of interaural fluctuations in the onset regions. If this hypothesis is valid, it could be suggested that a certain trade-off relationship between the spectral characteristics and temporal characteristics of sound source exists for the perception of source width.

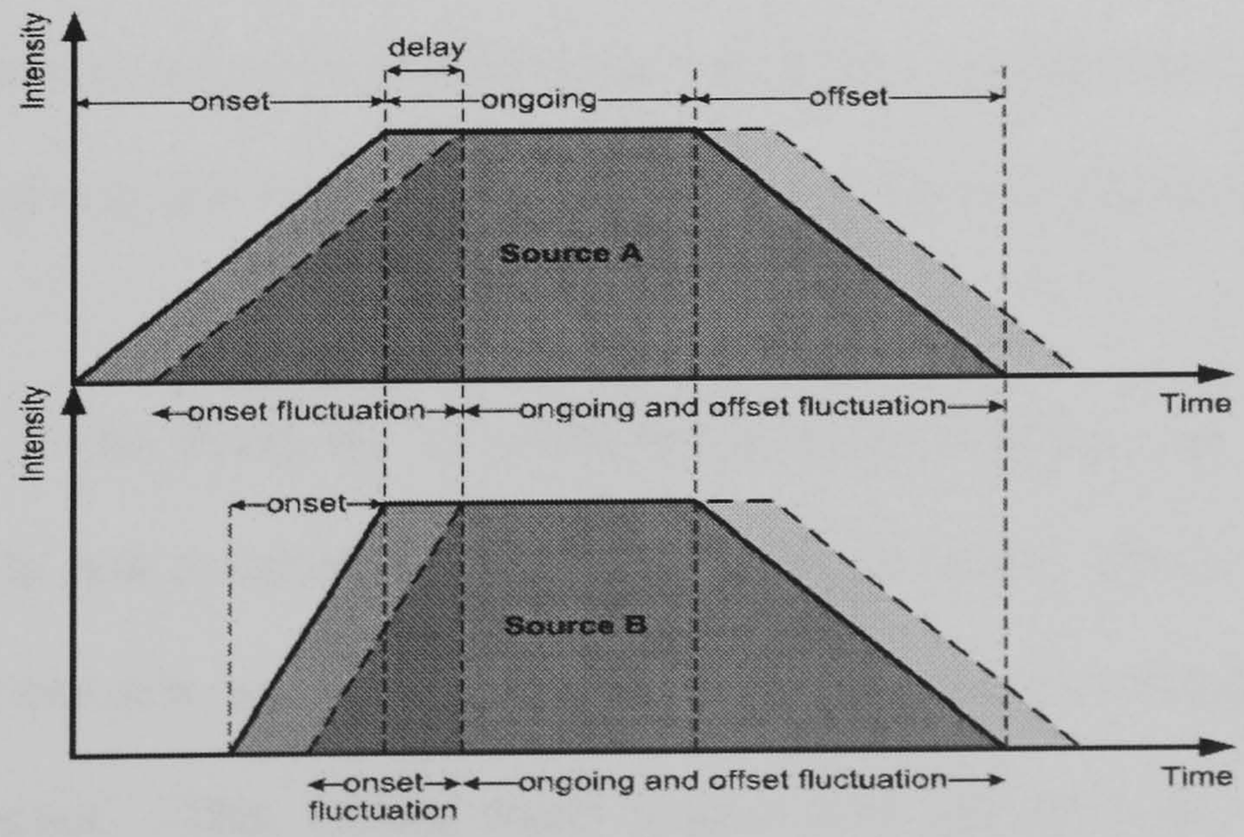


Figure 4.18 Interaction between leading and lagging signals for sound sources having different onset times with the same ongoing and offset times

It was shown that the type of sound source was not significant for the locatedness decreasing effect. Locatedness perception seems to be related to the precedence effect, which was discussed in Section 2.2.2. It has been reported by many authors that the precedence effect would mainly be triggered by transient sounds rather than continuous sounds. From this, one may presume that the continuous nature of the cello source would cause a greater locatedness decrease in the crosstalk-on image than the transient nature of the bongo source would. However, it should be noted that the above finding strictly only relates to pure tone signals. Rakerd and Hartmann [1986] pointed out that in the case of a complex signal such as noise, the precedence effect could also be operated by continuous sounds. Furthermore, Tobias and Zerlin [1959] found that for noise signals, the continuous part became more influential on localisation than the onset transient as the duration of the signal increased. The sound sources used in the experiments reported here have complex spectral and temporal characteristics as shown in **Figures 4.2** and **4.3** and it is likely that all of them have sufficient transient information to retrigger the precedence effect. For example, the speech source has a fine structure of transients at every syllable, the cello source also has a continuous musical phrase containing ongoing fluctuations at every note or bow change and every hit in the bongo source contains a rapid onset transient.

In addition, it was found that the perceived locatedness change was significantly smaller in the hall condition than in the anechoic condition. This is likely to be because the crosstalk was perceptually masked by the long reverberation from the previous sounds. This finding might suggest that the effect of crosstalk on locatedness change would become less audible in a more diffused recording space.

4.4.4.2 Discussion of the relationships among the attributes

Table 4.14 shows the summary of significance values for each attribute. The main effect of microphone array was significant for both locatedness and source width changes. However, the significances of the sound source and acoustic condition effects were found to be opposite for each attribute. That is, the effect of sound source was significant for the source width change, but not for the locatedness change. In contrast, the effect of acoustic condition was significant for the locatedness change, but not for the source width change. For interaction effects also, only the acoustic*source interaction was significant for the locatedness change while it was the only insignificant interaction for the source width change. There is a tendency in the literature for source width and locatedness attributes to be regarded as negatively correlated. For example, in Berg and Rumsey's research [2002], 'source width' and 'localisation', although a different term was used for the definition of 'locatedness', were found to be negatively correlated at a moderate level. The result of the previous investigation, as was discussed in Section 3.4.4.2, also showed that 'source width' and 'source focus' attributes in 2-0 stereophonic images had a strong negative correlation.

	Main Effect			**Interaction Effect**		
	Array	Source	Acoustic	Array *Source	Array *Acoustic	Source *Acoustic
Locatedness	0.000	0.637	0.003	0.058	0.052	0.029
Source width	0.000	0.004	0.711	0.000	0.038	0.714

Table 4.14 Summary of significance values of the main effects and interaction effects for locatedness and source width changes caused by interchannel crosstalk

However, the differences found in the significance levels between locatedness and source width for each factor shown in the above table led to a hypothesis that the correlation between those attributes depends on sound source and acoustic condition. Therefore, a set of bivariate correlation tests was carried out. Since the microphone array effects in both attributes have similar tendencies, the level of correlation was expected to be considerable when all the independent variables were included in the test. The result was in fact a moderate negative correlation (-0.670). This means that the ratio of interchannel time and intensity differences affects the changes in both attributes similarly. However, it was also predicted that if only one microphone array was considered, the correlation would be at a low level due to the different main effects of the sound source and acoustic condition. Therefore, individual correlation tests were also performed with each microphone array and the results confirmed the prediction as can be seen in **Table 4.15**. In general this result suggests that with respect to the effect of interchannel crosstalk in a microphone technique, a large source width increase resulting from interchannel crosstalk does not necessarily mean a large locatedness decrease nor vice versa. This finding might also lead to a discussion on the relationship between source width and locatedness perceptions in general. As mentioned above, it seems to be a widely accepted concept that a wider source is more difficult to localise. However, based on the above result, it might be suggested that the correlation between those two attributes is dependent on the type of sound source (This issue is further discussed in Chapter 5.).

	Array 1	**Array 2**	**Array 3**	**Array 4**
Correlation	-0.280	-0.323	-0.169	-0.201

Table 4.15 Correlation value between locatedness change and source width change by microphone array

4.5 Experiment Part 3: Preference for Interchannel Crosstalk

4.5.1 Background

From the previous experiment it was found that interchannel crosstalk had a significant effect on the increase in perceived source width and the decrease in perceived locatedness of the sound. It is asserted by Theile [2001] that interchannel crosstalk should be suppressed as much as possible in the design of multichannel microphone techniques as it is considered to be a negative factor for achieving balanced localisation of sound sources in the reproduction. If the aim of sound recording was only to capture a precisely localised sound image, then Theile's claim might be fully supported. However, in practice localisation is not the only criterion determining the perceived sound quality. Rather it is often found that the creation of sufficient spatial impression is more desirable than the achievement of accurate localisation [Mckinnie 2004]. The popularity of spaced microphone techniques such as the Decca tree over the pure coincident technique such as the XY could be a good example of this. Furthermore, in the context of concert hall acoustics, the increase in source width caused by early reflections is found to be a positive factor for perceived sound quality as outlined in Section 2.3.2.3 [Schroeder *et al* 1974, Barron 1971, Ando and Kageyama 1977, Barron and Marshall 1981, Blauert and Lindemann 1986]. However, studies in concert hall acoustics typically consider much longer delay times of reflections (10ms <...< 80ms) than those of crosstalk signals that might be encountered in general microphone techniques and therefore investigation in the context of sound recording and reproduction is required to confirm whether

interchannel crosstalk would also be a positive factor for the perceived sound quality. From this background, a pairwise comparison experiment was conducted to examine the subjective preference between crosstalk-on (LCR) and crosstalk-off images (CR).

4.5.2 Stimuli selection

This experiment used only 12 pairs of representative stimuli from the whole 36 pairs of stimuli used in the grading experiment. The results of the grading experiment showed that the differences between crosstalk-on and crosstalk-off images in arrays 1 and 2 were not as obvious as those in arrays 3 and 4. From this, it was considered that it would be hard to distinguish the difference in preference for the stimuli of arrays 1 and 2. In addition, it was considered that the preference testing of the anechoic stimuli would be inappropriate from a practical point of view. Therefore, only the stimuli of arrays 3 and 4 with room and hall simulations were used for this experiment.

4.5.3 Test subjects

For the evaluation of the subjective preference of sound quality, normally a large number of naïve listeners are used. However, for this particular preference test, the listeners' critical listening skills were crucial for distinguishing the fine perceptual differences resulted from interchannel crosstalk. Also, this whole study was focused on the viewpoints of classical music recording engineers about interchannel crosstalk.

Therefore, again the same subjects as in the previous elicitation and grading experiments, who are trained and experienced sound engineers, took part in this experiment.

4.5.4 Listening test method

The subjects were asked to test a total of 12 trials using the same control interface described in the grading experiment. Each trial presented two sounds A and B, which presented CR and LCR in random orders. The order of the presentation for the trials was randomised for each subject in order to avoid potential psychological errors. There were two tasks for the subjects to complete in this test. The first task was to judge which sound they preferred and to grade the preference on the scale shown on the provided answer sheet. The scale chosen for this experiment was a nine-point bipolar semantic scale, which was adopted from the hedonic acceptance scale [Stone and Sidel 1993]. The subjects were requested to circle the term that best reflected their attitude about the sound and the results were later transformed to numerical values for statistical analysis. It was considered in the design of the grading experiment that some semantic scales might not be ideal for parametric statistical analysis because of the psychological nonlinearity of the scale. However, this would not be the case for the hedonic scale as the psychological distances between each semantic label are equal [Stone and Sidel 1993], and therefore the numerically transformed data could be directly used for parametric statistical analysis. The semantic labels that were used in the scale and their corresponding numerical values for statistical analysis are shown in **Table 4.16**.

The purpose of the second task was to understand the attributes that influenced preference and the priority among them. **Table 4.17** shows the questionnaire used for this task. As can be seen, the subjects were given a list of the crosstalk attributes that were elicited in the experiment part 1. They were firstly asked to select the attributes that contributed to their choice of sound for each trial and then to rank them according to the degrees of the contributions. They were then asked to complete a statement written as 'The preferred sound is _____ than the other' for each of the selected attributes by circling the relevant comparative words provided. If there were additional reasons for their choice of sound, the subjects were encouraged to describe them using their own words and also rank them. The data obtained from the second task were analysed so as to understand the relative perceptual weight of each of the preference attributes.

Semantic labels	Numerical values
Prefer sound A *Extremely*	4
Prefer sound A *Very Much*	3
Prefer sound A *Moderately*	2
Prefer sound A *Slightly*	1
Prefer *Neither* sound A *nor* B	0
Prefer sound B *Slightly*	-1
Prefer sound B *Moderately*	-2
Prefer sound B *Very Much*	-3
Prefer sound B *Extremely*	-4

Table 4.16 Nine-point bipolar semantic scale used for the preference grading and the numerical values given for each label

Rank	Attributes for preference	The preferred sound is _____than the other.	
	Source width	Wider	Narrower
	Locatedness	More located	Less located
	Source distance	More distant	Less distant
	Brightness	Brighter	Darker
	Hardness	Harder	Softer
	Fullness	Fuller	Thinner
	Describe additional attributes		

Table 4.17 Questionnaire used for preference test

4.5.5 Results

For the analysis of the grading data obtained, the number of the preference for each sound was investigated first in order to look at the general polarity of the preference. There were a total of 96 observations, consisting of 12 observations obtained from 8 subjects. For this analysis, the semantic data were modified by giving a value of +1 where the crosstalk-on image was preferred, -1 for the crosstalk-off image and 0 for no preference. The percentages for the frequencies of the numerical values were then analysed as shown in **Figure 4.19**. It can be seen that the crosstalk-off sounds (50) were preferred to the crosstalk-on sounds (36) more frequently, while in 10 cases there was no preference. This shows that the crosstalk-off sounds were not exclusively preferred to the crosstalk-on sounds.

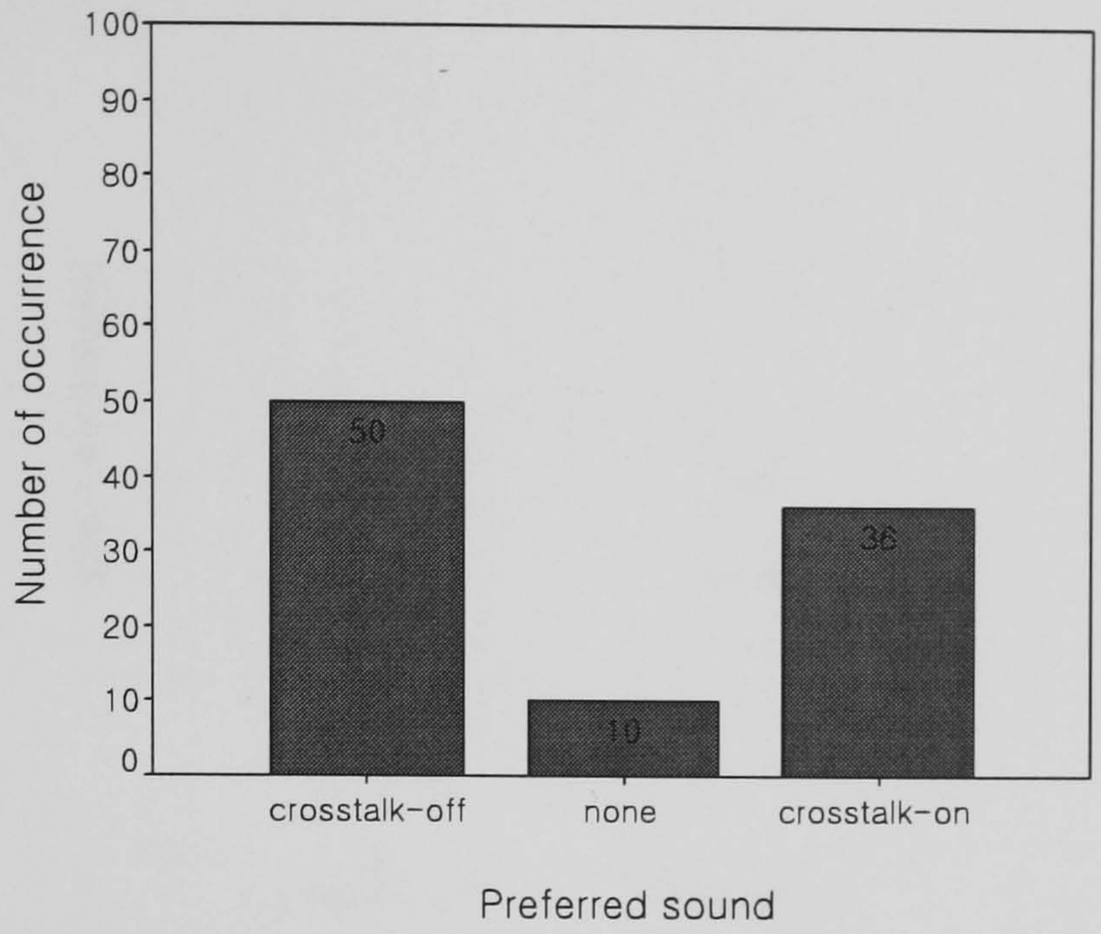


Figure 4.19 Frequency percentages of preference choices for sounds with and without crosstalk

In order to see the overall degrees of preference gradings for each crosstalk condition, the mean values and 95% confidence intervals of the grading data were obtained as shown in **Figure 4.20**. It can be seen that the degree of preference is very similar for both types of stimuli, being around the *moderately prefer* range. In order to examine the statistical significance of the difference between crosstalk-on and crosstalk-off sounds in the degree of preference, a nonparametric method was used since the number of observations for each case was different. Therefore, the Mann-Whitney U test was performed and the result shown in **Table 4.18** confirmed that the difference was insignificant ($p = 0.977$).

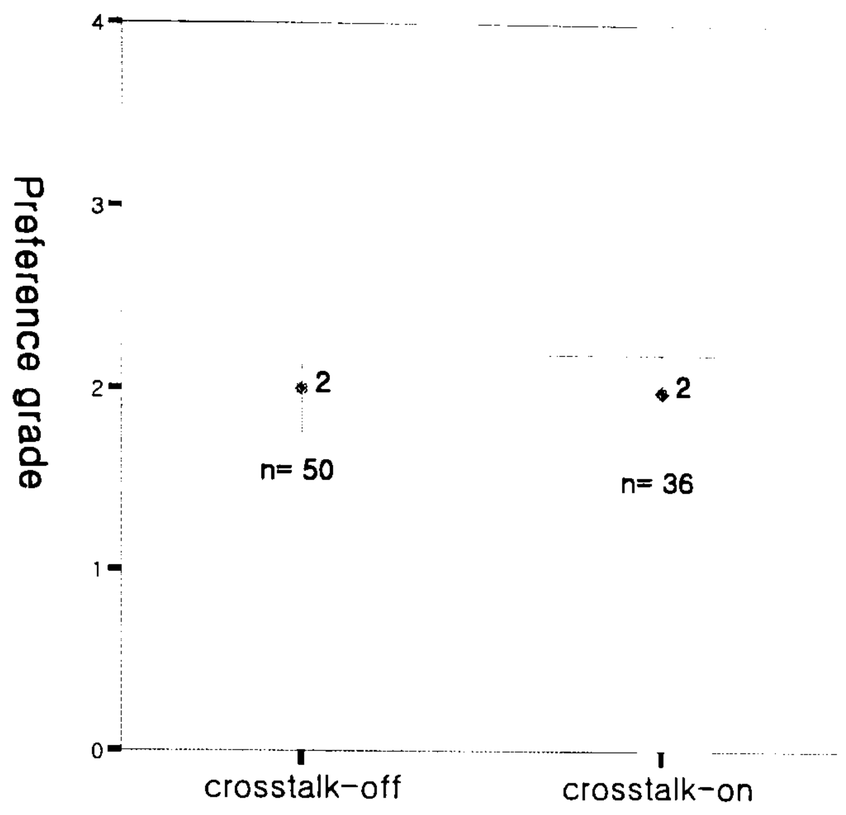


Figure 4.20 Mean values and 95% confidence intervals for the grading values of crosstalk-off and crosstalk-on stimuli

Test Statistics

	GRADE
Mann-Whitney U	897.000
Wilcoxon W	1563.000
Z	-.028
Asymp. Sig. (2-tailed)	.977

Table 4.18 Result of the Mann-Whitney U test result for the preference grading data of crosstalk-off and crosstalk-on stimuli

The effects of the independent variables on the preference were also analysed using a repeated measure ANOVA method. The summary of the results of the conducted RM ANOVA is shown in **Table 4.19**. It can be seen from the results that none of the effects of independent variables were significant. Also the interactions between each independent variable are shown to be insignificant.

Model	F	Sig.
Acoustic	0.000	1.000
Source	0.862	0.444
Array	1.109	0.327
Acoustic*Source	0.963	0.406
Acoustic*Array	0.562	0.478
Source*Array	3.245	0.079

Table 4.19 Summary of the Repeated Measure ANOVA performed for the analysis of the preference gradings

From the data obtained from the second task of the listening test, it was analysed what kinds of attributes of interchannel crosstalk contributed to the subjects' preference choice and how they were relatively weighted. **Table 4.20** shows the result of the analysis. Firstly, it can be seen that all the provided attributes and two additional attributes were related to the subjects' preference. In order to examine the relative perceptual importance of these attributes, a weighting factor was calculated for each attribute using the equation below, as used in Neher [2004]. As the attributes had been ranked by the subjects in order of priority, a specific index was given to each rank number. The rank number 1 was assigned the rank index 1, the rank number 2 the index 1/2, the rank number 3 the index 1/3, and so on. The calculation of the weighting factor was equated so that the maximum value became 1.

$$\frac{\text{(Sum of the number of occurrences) + (Sum of rank index)}}{\text{(Number of trials) x (Number of subjects) x 2}}$$

Preference Attribute	Occurrences	Preference Polarity (occurrences)		Weighting Factor
Locatedness	60	More located (47)	Less located (13)	0.57
Source width	67	Wider (26)	Narrower (41)	0.54
Source distance	34	More distant (16)	Less distant (18)	0.27
Brightness	26	Brighter (18)	Darker (8)	0.21
Fullness	19	Fuller (15)	Thinner (4)	0.18
Hardness	18	Harder (14)	Softer (4)	0.13
Naturalness	9	More natural (9)	Less natural (0)	0.06
Phasiness	7	More Phasey (0)	Less Phasey (7)	0.05

Table 4.20 Group of attributes that contributed to the choice of sound and their relative weights

4.5.6 Discussions

4.5.6.1 Discussions on the results of the controlled experiment

From **Table 4.20**, it can be seen that the weighting factors for the locatedness and source width attributes are 0.57 and 0.54 respectively, which are noticeably higher compared to the values for the other attributes. This means that the locatedness and source width attributes were the most important contributors to the choice of sound. It is worth pointing out that the relative weightings for the preference attributes have a similar tendency to those for the attribute audibility that were shown in Section 4.3; the source width and locatedness attributes have the most dominant effects while timbral attributes have weak weights. This might suggest that subjective preference for a sound is likely to be determined by the most dominant perceptual attribute of the

sound. However, from a different viewpoint, this could also mean that there might have been strong psychological biases in the subjects' judgments due to the perceptual dominance of certain attributes. In other words, the subjects might have initially paid more attention to the locatedness or source width attribute for its strong audibility and judged the preferences directly by their prejudices about how the attribute should be perceived, without considering the aspects of other relevant attributes. As mentioned earlier, it tends to be taken for granted that the locatedness and source width attributes always have a strong correlation, although the results of the correlation test of the previous experiment suggested that they would not necessarily do so. Some classical recording engineers tend to prefer a more easily localised and narrow sound while others prefer a less easily localised and wider sound. On the other hand, in the context of concert hall acoustics, a wider or more diffused sound is usually regarded to be preferable to a narrower sound by normal audiences. The results shown in **Table 4.20** indicate that for the locatedness attribute, sounds were preferred largely because they were more located. It can also be seen that for the source width attribute, narrower sounds were more preferred to wider sounds. It is not so clear from these results which perceptual polarity was given to crosstalk-on or crosstalk-off sounds. However, based on the results of the previous experiments, it could be seen that crosstalk-on sounds were always perceived to be less located and wider compared to crosstalk-off sounds. Given that the subjects used in the current experiment were all experienced sound recording engineers, the possibility of bias that might have occurred during the listening test has to be acknowledged. It also has to be admitted that the result cannot be generalised since only a small number of subjects from a particular group was used for the experiment.

In general, the results obtained from this experiment seem to indicate that interchannel crosstalk would not significantly decrease the subjective preference of sound quality and this seems to present a challenge to Theile [2001]'s negative viewpoint on the influence of interchannel crosstalk. However, it has to be admitted that this experiment considered only a limited range of single sound sources and the perspectives of only a small number of experienced listeners in a controlled manner. Therefore, these results would not provide a conclusive answer about the acceptability of interchannel crosstalk.

4.5.6.2 Discussions on the limitations of the controlled experiment

It was found from this experiment that the sound source type did not have a considerable effect on the subjective preference of interchannel crosstalk. Certainly, this experiment enabled the subjects to focus solely on the changes due to interchannel crosstalk and in this regard the obtained results could be validated. However, it seems that these results stand alone from a practical point of view since they were obtained from using controlled stimuli manipulated with the simulated interchannel relationships and acoustic conditions. That is, in this experiment such acoustical factors as reflections and reverberation were mixed in the stimuli with the same patterns regardless of the interchannel relationships of the simulated microphone arrays. However, this kind of control is not possible in practical recordings with microphone techniques due to the fact that the distance and angle between the front microphones would directly affect the interchannel decorrelation of

the reflected and reverberant sounds. In these respects, it is not clear whether the results of the controlled experiment would be able to represent what would actually happen in practical recording situations. It is deemed that in practical microphone techniques the subjective preference for interchannel crosstalk would strongly depend on the type of sound source and the acoustic condition of the recording space due to the interaction between these two factors. For instance, such instruments as trumpet and clarinet would have relatively poor locatedness when they were performed in a reflective space, due to their continuous characteristics interacting with reflections (see Section 2.2.3). In this case, the interchannel crosstalk, which would be likely to decrease perceived locatedness, might become a negative factor for preference. However, for such instruments as piano, which would not particularly require good locatedness of every single note, the interchannel crosstalk might not become a problematic factor. In addition, such percussive instruments as conga and bongo would be easily localised regardless of the existence of reflection due to their strong transient characteristics and therefore the source width increasing effect of interchannel crosstalk might even provide a balanced locatedness and spatial impression to the images of recorded sounds.

4.6 Experiment Part 4: Comparisons of Practical 3-2 Stereophonic Microphone Techniques

4.6.1 Background

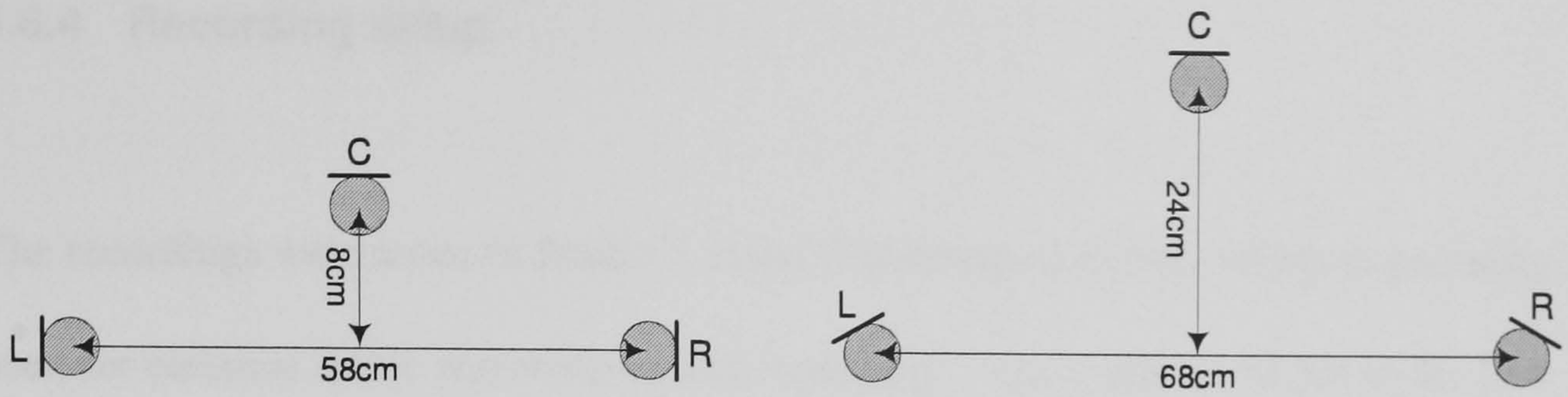
Based on the above discussions, an additional subjective experiment was carried out in order to provide supplementary findings about the practical implications of interchannel crosstalk. This kind of experiment using 'real world' recordings is typically limited when it comes to the question of controlling experimental variables other than interchannel crosstalk. That is, as mentioned above, the configuration of a microphone array would also be likely to contribute to the interchannel decorrelation of the reflected and reverberant sounds, which will have an effect on perceived spatial impression, and therefore it becomes difficult to distinguish between the effect of microphone configuration on front phantom imaging and that on spatial impression. Therefore, the interpretations of the causes for preference data obtained from a comparison between two different microphone techniques are likely to be somewhat arbitrary and indirect. If a microphone technique giving rise to stronger crosstalk was preferred to one giving rise to weaker crosstalk, then the crosstalk effect might be considered as either a positive or negligible factor for the perceived sound quality, depending on whether or not the crosstalk was the main contributor to the perceived sound quality. However, if a microphone technique with a weaker crosstalk was preferred to that with a stronger crosstalk, then it would be difficult to judge whether the preference was directly due to the crosstalk or not. Therefore, in the latter case, in order to find the main contributor for the preference, it would first be necessary to know if the subjective attributes resulting in the preference choice

matched any of those resulting from the crosstalk. If there were to be no match, the crosstalk effect could be disregarded. However, in the opposite case, it would be possible to regard the crosstalk only as a 'potential' negative factor since it would still not be clear if other types of variables produced similar attributes to those of the crosstalk and if they were actually the main contributor. For the above reasons, this experiment was designed for the subjects to describe the attributes that were most relevant to their preference choices as well as to grade the degrees of preferences.

4.6.2 Choice of microphone technique

The listening test was designed such that the subjects could compare recordings made with two different front microphone techniques, which differed in their crosstalk characteristics, in the presence of ambient sounds recorded with a common rear microphone technique. As mentioned earlier, interchannel crosstalk between front and rear arrays that are placed far apart will be not large. Nevertheless, the rear technique was used in this experiment to create a listening environment of a type likely to be encountered in a practical 3-2 stereophonic classical music reproduction. The front microphone techniques chosen for this comparison were the 'OCT' [Theile 2001] and 'ICA-3' [Herrmann and Henkels 1998], and the rear technique was the 'Hamasaki-square' [Hamasaki *et al* 2000]. The detailed descriptions of these microphone techniques were presented in Sections 1.4.2 and 1.4.3. Briefly summarising, the OCT technique using a cardioid centre microphone and two super-cardioid side microphones has a better interchannel crosstalk rejection than the ICA-3

using three cardioid microphones, although they are based on their own unique design concepts for balanced phantom imaging. The Hamasaki-square technique attempts to produce natural spatial impression across the front and rear channels using four figure-8 microphones arranged in a square. As mentioned in Section 1.4.2, there are various configurations available for both techniques depending on the desired SRA or microphone spacing and angle. Using the Image Assistant model [Wittek 2001a], which was introduced in Section 1.2.1, it was attempted to match the stereophonic recording angles (SRAs) of the front microphone arrays for localising the phantom sources created with both techniques at similar positions between the loudspeakers. Also, the microphone spacings were matched as closely as possible during the process of the SRA matching in order to minimise the effect of microphone spacing in the comparison of the two techniques. This made it possible to separate the perception of the crosstalk effect from that of interchannel decorrelation of reflections due to microphone spacing to some extent. The resulting SRA was 132° and the configurations of the microphone arrays are shown in **Figure 4.21**. Basically, the ICA-3 had larger microphone spacings than the OCT array, whereas the latter had a wider lateral microphone angle (90°) than the former (70°). The interchannel relationship between channels L (crosstalk) and C for each array, calculated under the assumption that the sound source is located at 45° of the centre line of each array with 5m distance from the centre base of the array, is as shown in **Table 4.21**. As can be seen, the OCT technique is superior to the ICA-3 technique in terms of the reduction of interchannel crosstalk.



(a) OCT (b) ICA-3

Figure 4.21 Configurations of the OCT and ICA-3 arrays having the same stereophonic recording angle (SRA) of 132°

	OCT	**ICA-3**
C to L Time difference	0.78ms	1.21ms
C to L Intensity difference	-16.6dB	-11.5dB

Table 4.21 Interchannel relationship of crosstalk channel L against channel C for the OCT and ICA-3 microphone arrays used for the preference experiment; the simulated direction of sound source is 45° and the distance of the sound source from the arrays is 5m.

4.6.3 Choice of sound source

This experiment used a range of musical sound sources comprising performance excerpts of string quartet, solo percussion pair (conga and bongo), solo violin and solo piano. They were chosen for the variety of both musical contexts and physical characteristics (i.e. ensemble vs. solo, continuous vs. transient, and syllabic vs. wide source). The musicians were students of Music Department of the University of Surrey.

4.6.4 Recording setup

The recordings were made in Studio 1 at the University of Surrey, which is primarily used for classical music performance and recording. The studio is 14.5m wide, 17m long and is 6.5m high. The reverberation time RT_{60} is approximately 1.5 seconds. **Figure 4.22** shows the dimensions of the studio and the positions of the front and rear microphone arrays.

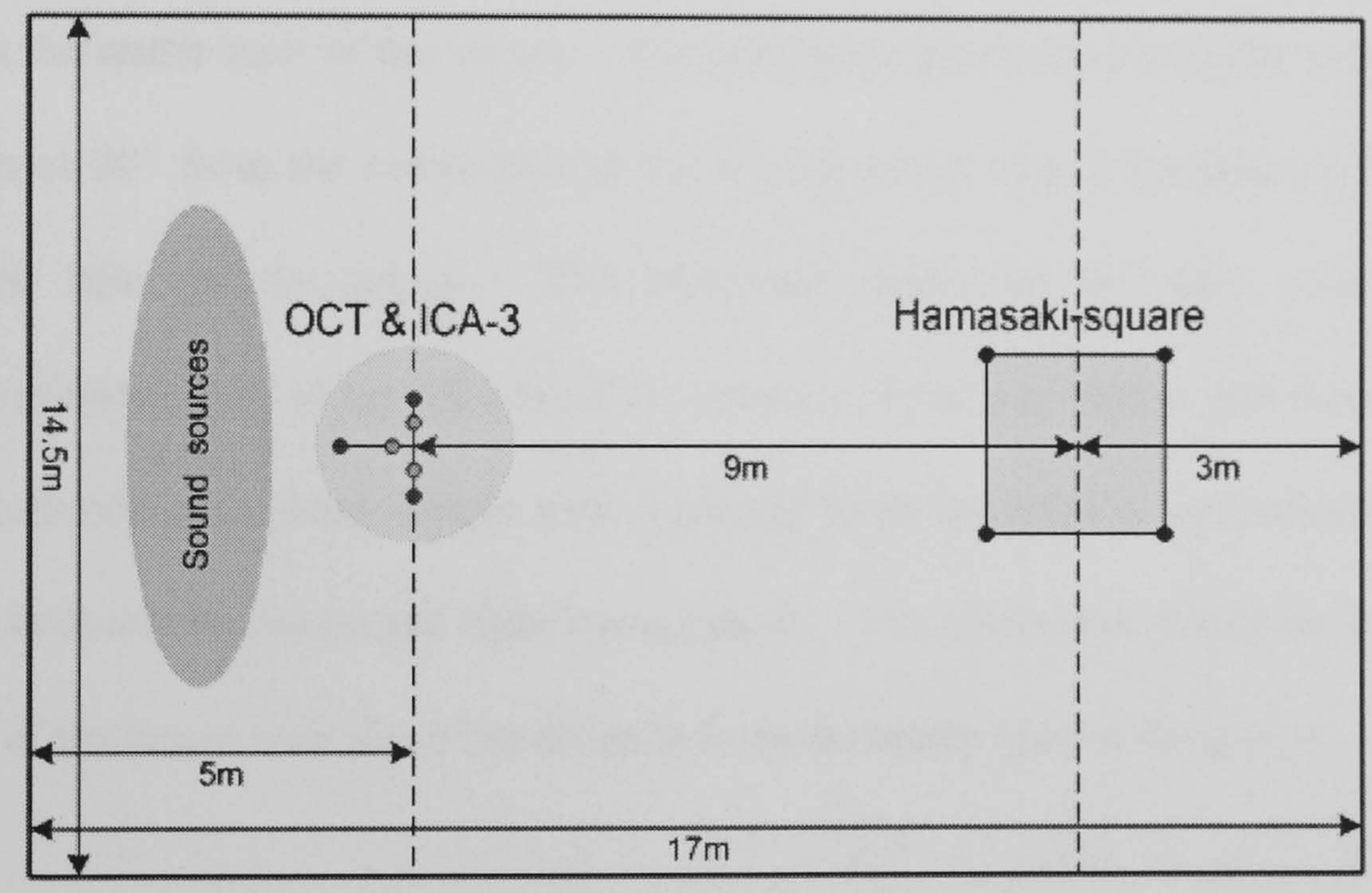


Figure 4.22 Recording studio setup

As can be seen in the above figure, the front and rear arrays were placed in the centre line of the studio. The front arrays, which were manipulated to have the same SRA, were placed 5m from the front wall and centred at the same location in order to create similar stereophonic sound stages. The distance between the front and rear arrays was 9m. The height of the front arrays was 2.2m while that of the rear array was 4m. It was recognised that using microphones of different models or manufacturers for each front array would be likely to cause differences in the timbral qualities of each

array, which should be distinguished from the timbral differences caused due to the interchannel relationship of each array. Therefore, Schoeps CCM models of various polar patterns were exclusively used for this experiment, which are closely matched in terms of timbral quality.

The placements of the sound sources varied. The string quartet was arranged in a normal concert configuration in front of the front arrays and placed about 2m away from the centre base of the arrays. The percussion pair and solo violin were placed at about 30° from the centre axis of the frontal arrays with 2.5m distance from the centre base of the arrays. The off-centre angle for the solo sources was approximately half of the SRA for C-R segments of the front arrays and therefore the corresponding phantom images were expected to be localised at approximately half way between the centre and right loudspeakers. The piano was placed on the centre axis of the arrays with about 3m distance from the centre base of the arrays.

The microphone output signals were fed through a Sony Oxford R-3 digital console and recorded as ten discrete channels on Sony PCM-800 recorders at 16bit/48kHz, which were eventually mixed as five channels for each combination of front and rear arrays to be reproduced in 3-2 stereophonic system. The mixing ratio of the front and rear array signals was decided by the author's artistic and technical judgment as an experienced balance engineer, aiming to achieve a reasonable combination of the clarity of the direct sound and sufficient listener envelopment.

4.6.5 Test subjects

This experiment was conducted using the same eight subjects selected from the previous experiments, who are all trained sound engineers. This number of subjects was probably too small to draw a general conclusion about the perceived sound quality, but was potentially representative of a population of trained recording engineers. Furthermore, the main purpose of this experiment was to evaluate the results of the previous controlled experiment in a practical manner and indirectly or informally map this subject group's preference patterns for interchannel crosstalk depending on the types and contexts of sound sources.

4.6.6 Listening test method

The listening test was conducted in the same listening condition as the previous experiments. The peak sound pressure levels of the recordings made with the two techniques were calibrated at 75dBA. There were a total of four trials to be tested. Subjects were asked to compare between the sounds recorded with the OCT and ICA-3 techniques for each sound source, which were arranged in random orders for each trial, using a control interface. The order of the trial was also randomised. The subjects' tasks were to grade the degree of preference on a nine-point hedonic scale, which was described in the previous experiment, and to describe the reasons for their preference choices using their own terms.

4.6.7 Results and discussions

For statistical analysis of the data obtained from the listening test, the semantic labels of the grading scale were first converted into numerical values in the same manner that was described in the previous controlled experiment (e.g. prefer *Extremely* = 4, *Very much* = 3, *Moderately* = 2, *Slightly* = 1, prefer *Neither* = 0).

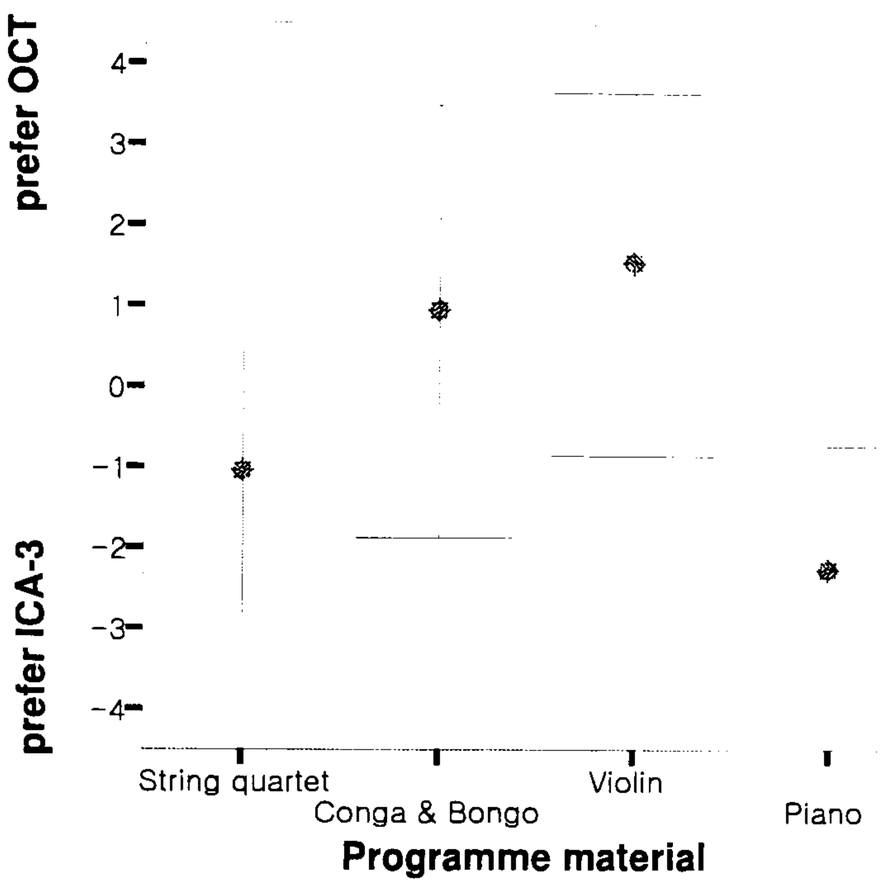


Figure 4.23 Mean value and associated 95% confidence intervals of the preference grading for each programme material

Figure 4.23 shows the plot of mean value and associated 95% confidence intervals for the preference gradings made for each programme material. The positive values in the grading scale represent the preference for the recording made with the OCT array and the negative values represent the preference for that with the ICA-3 array. It can be initially seen that for the string quartet and solo piano recordings, the ICA-3 was preferred to the OCT while for the solo conga & bongo and solo violin recordings the OCT was preferred to the ICA-3. It appears that the ICA-3 was most

preferred for the piano recording while the OCT was most preferred for the violin recording. In order to examine the significance of the difference between the results for each programme item, a paired samples T-test was performed and the results are shown in **Table 4.22**. As can be seen, every pair of sound sources except the pair of 'conga & bongo – violin' had a significant difference.

		t	df	Sig. (2-tailed)
Pair 1	string quartet - conga&bongo	-4.472	4	.011
Pair 2	string quartet - solo violin	-3.474	4	.025
Pair 3	string quartet - solo piano	3.207	4	.033
Pair 4	conga&bongo - solo violin	-.612	4	.573
Pair 5	conga&bongo- solo piano	4.824	4	.008
Pair 6	solo violin - solo piano	10.156	4	.001

Table 4.22 Result table of paired samples T-test for each sound source

As mentioned earlier, the ICA-3 array produces a stronger crosstalk than the OCT. Therefore, the fact that the ICA-3 was more preferred to the OCT for the recordings of the string quartet and piano sources suggests that the presence of interchannel crosstalk would have been either a positive or negligible factor for the perceived sound quality for those sound sources.

Table 4.23 presents the list of the terms that were used by the subjects to describe the reasons for their preference choices. The number in brackets represents the number of occurrences for each specific description. As mentioned earlier, from this list of

descriptive terms, it is possible to examine whether or not the preference for the ICA-3 was due to the strong interchannel crosstalk.

	OCT	ICA-3
String quartet		More comfortable (1) More pleasant tonal balance (1) Closer (1) More central (1) Less hard (1)
Solo conga & bongo	Better locatedness (2) Narrower (2) More focused (1) More clarity (1) More natural (1)	Brighter (1) Fuller (1)
Solo violin	Better locatedness (2) Narrower (2) More stable (1) Slightly brighter (1) Less phasey (1) Softer (1)	More pleasant stereo image (1) More pleasant tonal balance (1)
Solo piano		Wider (2) Less localisable (1) Closer (1) Better tonal balance (1) More natural (1) Softer (1) Fuller (1)

Table 4.23 Summary of subjective terms that describe the reasons for preference choice of the OCT and ICA-3 microphone techniques

It can be seen from the above table that for the string quartet there is no term that describes or alludes to any of the crosstalk attributes that were elicited previously.

This means that the preference choice was probably not directly due to the presence of interchannel crosstalk but to another physical factor such as microphone spacing and angle. In this case, interchannel crosstalk can be regarded as a negligible factor. For the piano source, however, a number of crosstalk-related attributes are included in the list of preference reasons (e.g. source width (wider), locatedness (less localisable), hardness (softer) and fullness (fuller)). Although it is still not entirely clear whether these attributes were perceived directly from the presence of crosstalk or from the larger microphone spacing that might have caused a greater signal decorrelation, at least it can be suggested that interchannel crosstalk could potentially be a positive factor for perceived sound quality for such a sound source as solo piano. A possible explanation for this result based on the descriptions shown in **Table 4.23** is as follows. For the piano recording, easy localisation of the sound of each individual note would not have been a main factor for the subjects to determine the perceived sound quality. Rather, the subjects might have focused on an overall stereophonic image, which was perceived to be spatially wide and tonally full.

The results also show that the OCT was preferred for the recordings of the solo percussion and solo violin. It is indicated in **Table 4.23** that for both sources the reasons for choosing the OCT were directly related to the crosstalk attributes. That is, the OCT was preferred mainly because the resulting phantom source images were easier to localise and narrower than those created with the ICA-3. This suggests that it is highly possible that the stronger crosstalk in the ICA-3 array was a negative factor for the subjective preference of sound quality for those sound sources, although it is also possible that the microphone spacing was the main factor.

It is difficult to draw any general conclusion about the preference for interchannel crosstalk from the results presented here, for several reasons. Firstly, only a small number of subjects participated in the experiment. Secondly, preference for sound quality will depend partly on the type of subject. That is, the subjects used for this experiment were all trained sound engineers, but the results might have differed if naïve subjects had been used since they would have different perspectives on the judgment of sound quality. Nevertheless, the results obtained from this experiment at least suggest that for sound engineers the effects of interchannel crosstalk on perceived sound quality are not always regarded negatively but can be regarded positively, depending on the characteristics desired for recordings of different types of sound source. For example, for recordings of such ensembles as string quartet and orchestra, sufficient spatial impression and natural blending of instruments in an overall stereophonic image might be more desired than precise localisation of each individual note or instrument. For the recording of such a wide solo instrument as a piano, localisation of each individual note might not be as important as a broad spatial impression in an overall stereophonic image. In these cases, the perceived quality of recorded sound could benefit from the presence of interchannel crosstalk. However, for recordings of narrow solo instruments, any decrease in locatedness caused due to interchannel crosstalk would become relatively more noticeable compared with a similar degree of decrease in locatedness for individual instruments in an ensemble and this might become a disturbing factor for listening.

4.7 Summary

A series of subjective experiments were conducted in order to investigate the perceptual effect of interchannel crosstalk in multichannel microphone techniques, using trained sound engineers. Firstly, elicitation and grading experiments were conducted in order to investigate the types of audible attributes and their relative weights depending on various physical variables. The independent variables were microphone array type, sound source type and acoustic condition. The experimental stimuli were created by simulations of multichannel recordings made with the above variables. Subjects were asked to compare the perceptual differences between crosstalk-on and crosstalk-off sounds. The audible attributes of interchannel crosstalk were first elicited from the subjects and only the most dominant ones were selected. Then the magnitudes of the selected attributes were graded. The obtained grading data were statistically analysed using the repeated measure ANOVA method. Finally, the effect of interchannel crosstalk on subjective preference of perceived sound quality was investigated in both controlled and practical manners. The controlled preference experiment was conducted so that the controlled stimuli of crosstalk-off and crosstalk-on, which were used in the previous experiments, were compared for preference choice. The practical preference experiment involved various recordings of musical performances made with two different three-channel microphone techniques of OCT and ICA-3, which differ in the crosstalk characteristics.

The main findings obtained from the experiments are as follows.

- The audible attributes of interchannel crosstalk images elicited from the subjects were source width, locatedness, source direction, fullness, source distance, hardness, brightness, diffuseness, naturalness, envelopment and phasiness.
- Source width and locatedness were found to be the only attributes that were more than 'slightly audible'.
- In general, the interchannel crosstalk caused an increase in perceived source width and a decrease in locatedness.
- Statistically, the magnitudes of both source width increase and locatedness decrease significantly depended on the ratio of interchannel time and intensity differences in three-channel frontal microphone technique. For both attributes, an array employing a greater interchannel time difference (conversely, a greater intensity of crosstalk signal) caused a greater effect.
- Sound source type was a significant factor for the source width effect but not for the locatedness effect. In general, the sound source having a broader frequency range caused greater source width increase.
- Acoustic condition had a significant effect on the locatedness decrease, but not on the source width increase. The locatedness decreasing effect became less perceptible as the reverberation became more diffused.
- Interactions between microphone array type and sound source type, and between microphone array and acoustic condition were significant for the source width effect, but not for the locatedness effect. The experimental effects for these interactions were very small, thus can probably be ignored.

- Interaction between sound source type and acoustic condition was significant for the locatedness changing effect, but not for the source width changing effect. The experimental effect for this interaction was very small, thus can probably be ignored.
- For each microphone array type, the source width and locatedness changing effects of interchannel crosstalk had a low correlation.
- There was no noticeable difference between the crosstalk-on and crosstalk-off sounds in the preferences graded by a group of trained sound engineers using the controlled stimuli.
- Microphone array type, sound source type and acoustic condition had no significant effects on the preference grading of the controlled stimuli.
- Locatedness and source width attributes were the most salient preference cues for the controlled stimuli.
- In the comparison between the OCT and ICA-3 microphone techniques, the ICA-3 was preferred to the OCT technique for the string quartet and solo piano recordings while the OCT was preferred to the ICA-3 for the solo violin and percussion recordings (i.e. the OCT had a greater reduction of interchannel crosstalk than the ICA-3).

5 OBJECTIVE MEASUREMENTS OF THE EFFECTS OF INTERCHANNEL CROSSTALK

The series of experiments described in the previous chapter investigated the perceptual effect of interchannel crosstalk. The results showed that the most dominant perceptual effects of interchannel crosstalk were the increase in source width and the decrease in locatedness. The statistical analysis of the data obtained from the grading experiment indicated that the type of microphone array had a significant effect for both the source width and locatedness attributes. The effect of sound source type was significant only for the source width while that of acoustic condition was significant only for the locatedness attribute.

This chapter discusses the objective measurements that were made to investigate the effect of interchannel crosstalk in a perceptual model, and to map the relationship between the perceived results and their physical causes. Firstly, the measurement model used in this investigation is introduced and the procedure of stimuli creation is described. Secondly, the results of the measurements are compared with those of the subjective experiment. Finally, the frequency and envelope dependencies of the measurements and their relationship with the perceived effect are discussed.

5.1 Measurement Model

The measurement model that was chosen for the current studies was an IACC-based width and location prediction model that was developed by Mason *et al* [2005c]. This model was designed to overcome the limitations of the conventional IACC-based width measurement technique such as Hidaka *et al* [1995]'s using impulse response, which was discussed in Section 2.3.2.4. In this model, Mason *et al* attempted to develop the IACC into a more complete and practical source width prediction model by including a simulation of the binaural hearing system and taking into account the effect of physical properties of musical source signals on perceived width.

This model is particularly suitable for the purpose of the current studies for the following reasons. Firstly, it divides the source signal into 22 frequency bands and measures them separately, so that the influence of different frequency components of the interchannel crosstalk signals on the measurement can be investigated. Secondly, the time-variant IACC measurement and the indication of loudness envelope for each frequency band enable one to examine the relationship between the temporal characteristics of the sound and the measurement. Finally, this model can provide a prediction of time variant source location as well as source width. The measurement of location change over time might well be related to locatedness perception. A block diagram of the main processing stages of this model is shown in **Figure 5.1** and the basic aspect of each stage is summarised in the following sections. More detailed descriptions of the principles of this model can be found in Mason *et al* [2005c].

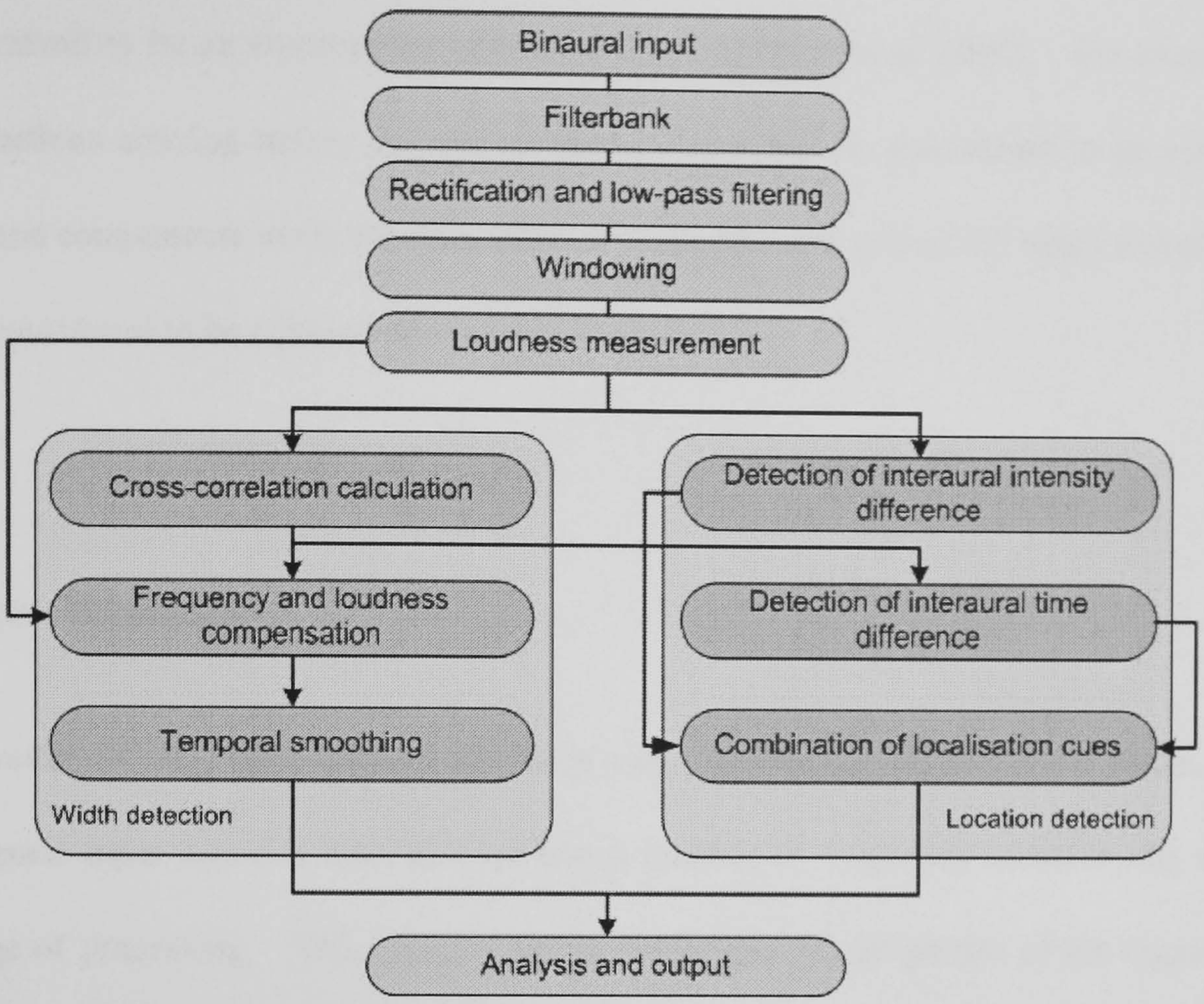


Figure 5.1 Block diagram of the processing stages of the IACC-based width and location prediction model that was developed by Mason et al [2005c]

5.1.1 Binaural input

The limitations of using impulse response and the fixed time division value of 80ms for separating source-related and environment-related segments were discussed earlier. This model is designed to measure musical source signals and the source and environment related segments are separated based on perceptual grouping, which is a concept inspired by Griesinger [1996, 1997]'s 'foreground-background' paradigm that was introduced in Section 2.3.1.2. Perceptual grouping is a simple division of a total input signal into segments containing physical parameters that are perceived to be a source-related attribute, and segments containing physical parameters that are

perceived to be an environment-related attribute [Mason *et al* 2004]. For example, reflections arriving before the end of direct sound will be considered to be source-related components while the reflections arriving after the end of the direct sound will be considered to be environment-related components.

5.1.2 Filterbank

In order to take into account the frequency dependency of perceived width, the binaural input signal is first divided into a number of frequency bands in the early stage of processing. This enables one to investigate the properties of the signal for different frequency ranges.

5.1.3 Half-wave rectification and low-pass filtering

This stage is included to simulate a physiological phenomenon in the binaural hearing system that is related to the perception of width. Mason *et al* [2004] found that in order to accurately predict the perceived width of high frequency stimuli, it is necessary to simulate the breakdown of phase-locking in the ear. The breakdown of phase-locking causes the fine temporal detail at higher frequencies to be lost and therefore the perceived width of high frequency stimuli to be dependent on the IACC of the signal envelope. This effect is simulated by passing the input signal through half-wave rectification and a 6th order Butterworth low-pass filter with a cut-off frequency of 1 kHz prior to the IACC measurement of the signal.

5.1.4 Windowing

It is known that the IACC of musical signals varies over time and this can be perceived. In order to predict the perceived width of a musical signal with a time-variant IACC, the signal is divided into a number of time windows and the IACC is measured in each window. The length of each window used for the current model is 50ms, although it may vary between 35ms and 80ms depending on whether the prediction should be made for the most critical listener or an average listener.

5.1.5 Loudness measurement

The effect of loudness on perceived width was investigated by Mason *et al* [2004 *et al*] with a number of narrow-band stimuli and it was found that the perceived width of a source signal had a loudness dependency. In order to take this into account in the prediction model, the sound pressure level (SPL) of the input signal in each time window for each frequency band is measured. The result is converted into the value of phons and sent to the processing stage of loudness and frequency compensation.

5.1.6 Cross-correlation calculation

The IACC is calculated using the IACF (interaural cross-correlation function) described in section 2.3.2.3. The common IACC is the maximum 'absolute' value of

IACF. In this model, however, the maximum value of IACF is taken because the positive and negative polarity of the value is considered to be related to different perceived effects.

5.1.7 Loudness and frequency compensation

It was mentioned above that the relationship between the perceived width and the measured IACC of a sound depends on the frequency and loudness of the sound. The dependencies on these physical factors of a sound should be taken into account in the measurement model; otherwise direct comparisons of the predicted widths will be possible only for the sound sources having identical characteristics of frequency and loudness [Mason 2004 *et al*]. Therefore, the differences in frequency and loudness are compensated in order to increase the accuracy and practicality of the prediction.

5.1.8 Temporal smoothing

Although it is known that the IACC of a sound varies over time, it is not clear how the variations are perceived temporally. Therefore, in order to model the temporal response of human hearing system to the varying width, Mason *et al* [2004] investigated subjective effects of the variations of IACC over time on width perception. It was found that decreases in the IACC appear to be perceived more rapidly than increases. This is simulated in this model by selecting the optimum

measurement window length, as described previously and using a complex state-dependent filter with a relatively fast onset and a slow offset.

5.1.9 Detection of interaural intensity difference

For each measurement window for each frequency band, the interaural intensity difference (IID) is detected by the measurement of the difference between the mean sound pressure levels in each channel.

5.1.10 Detection of interaural time difference

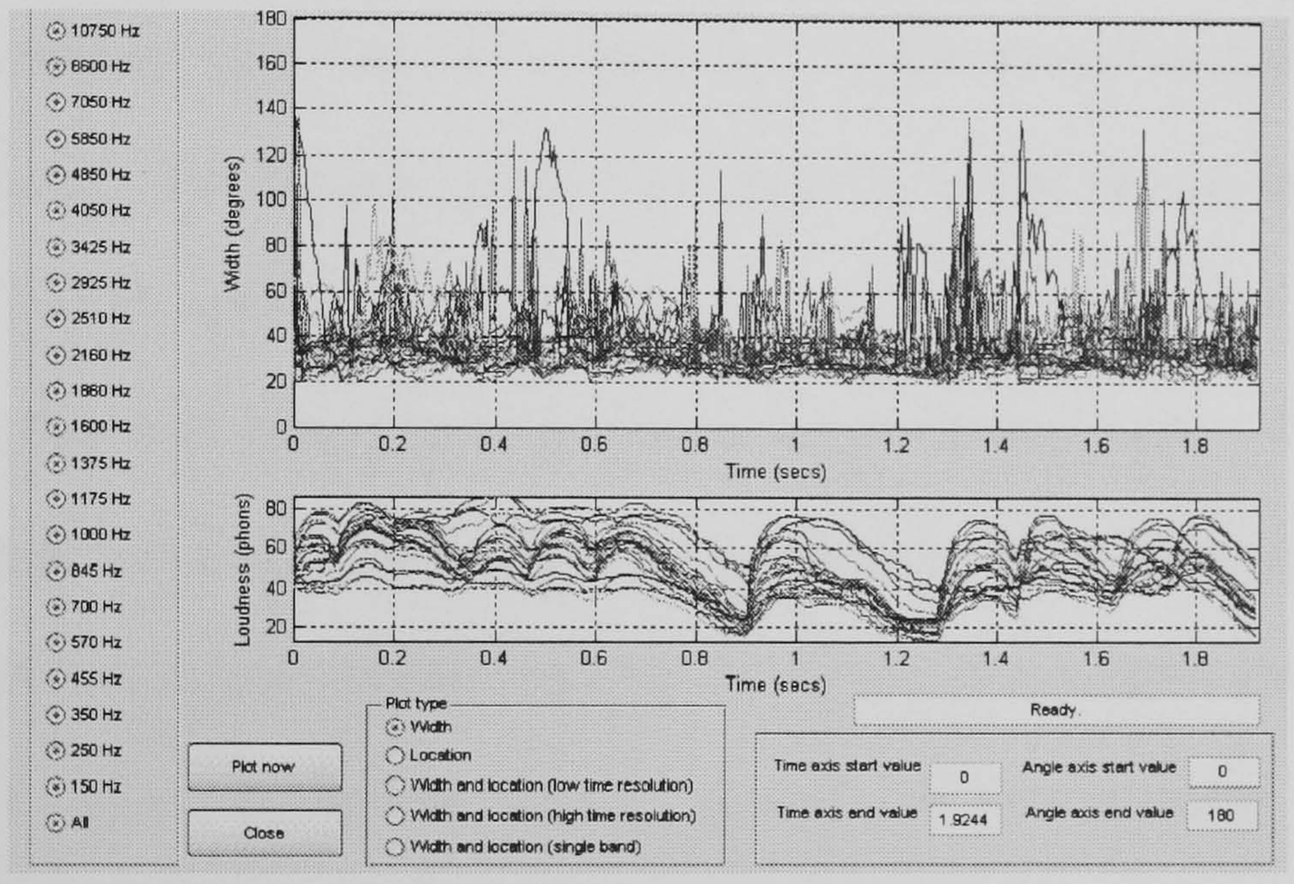
The value of time offset τ that relates to the peak in the measured cross-correlation is the prediction of the interaural time difference (ITD).

5.1.11 Combination of localisation cues

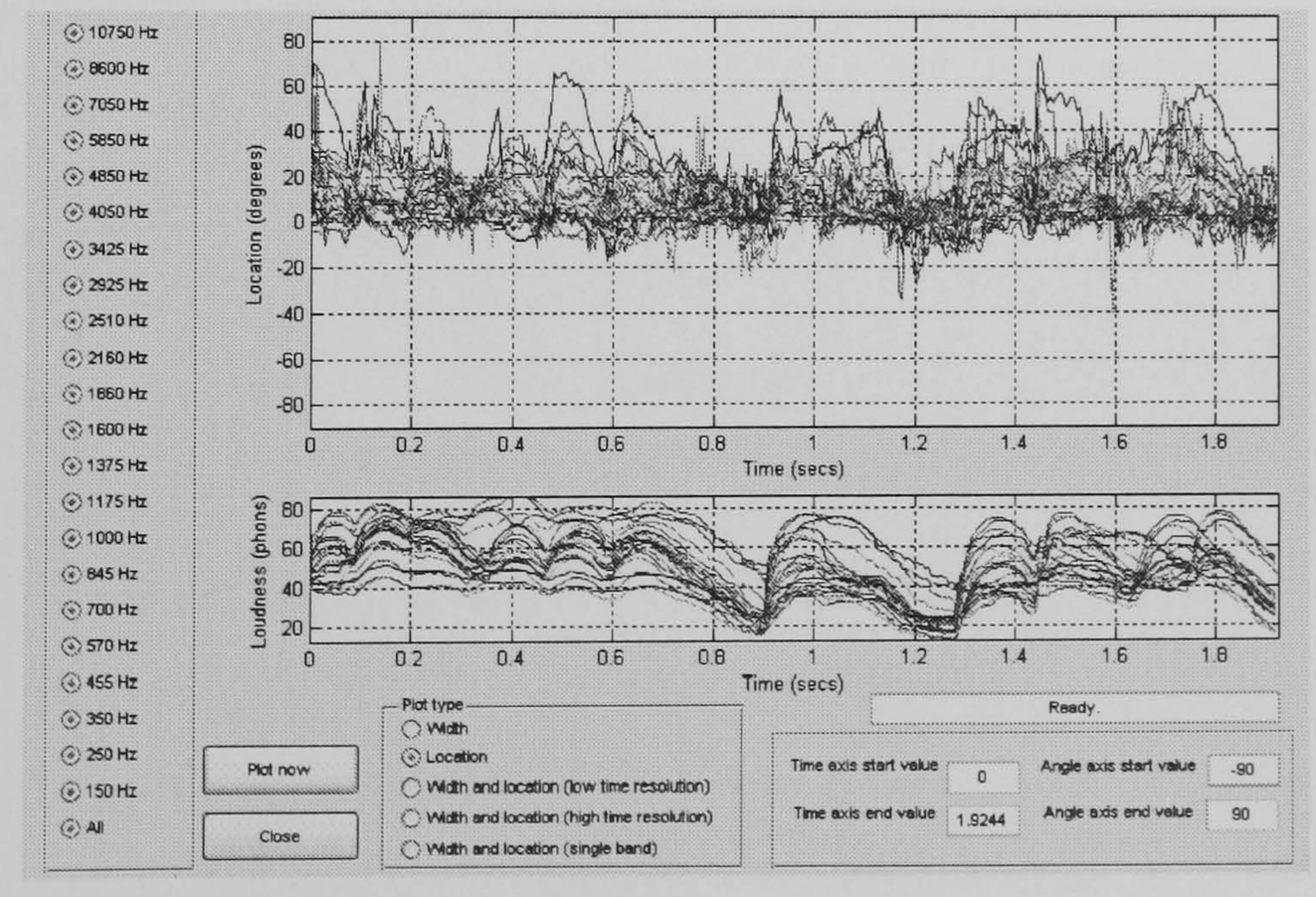
The resulting data of the IID and ITD are combined using a trade-off equation based on subjective data from Damaschke *et al* [2000].

5.1.12 Analysis and output

The data for the width and location detections described above are integrated and the results are converted to angles based on data of from Kuhn [1977]. Displays of the final outputs of width and location measurements are shown in **Figure 5.2**.



(a) Display of the result of width measurement

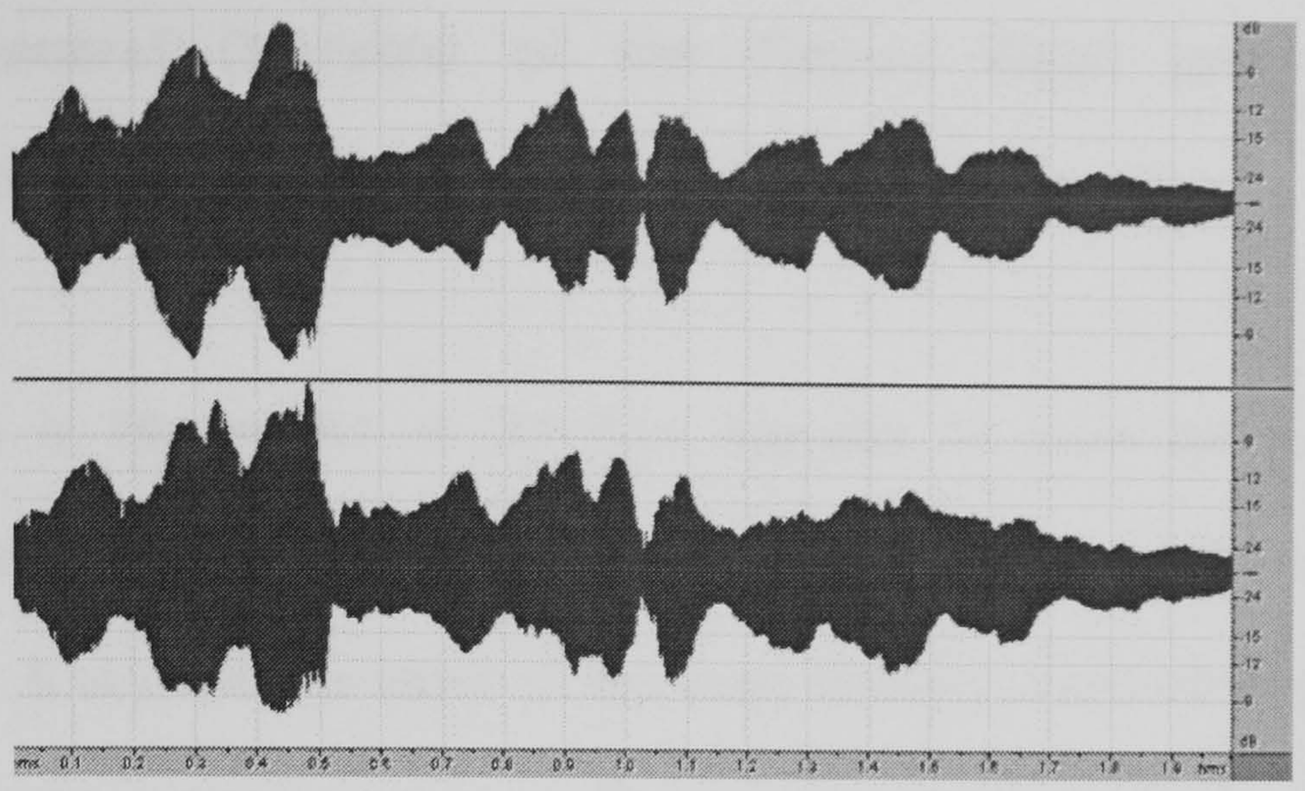


(b) Display of the result of location measurement

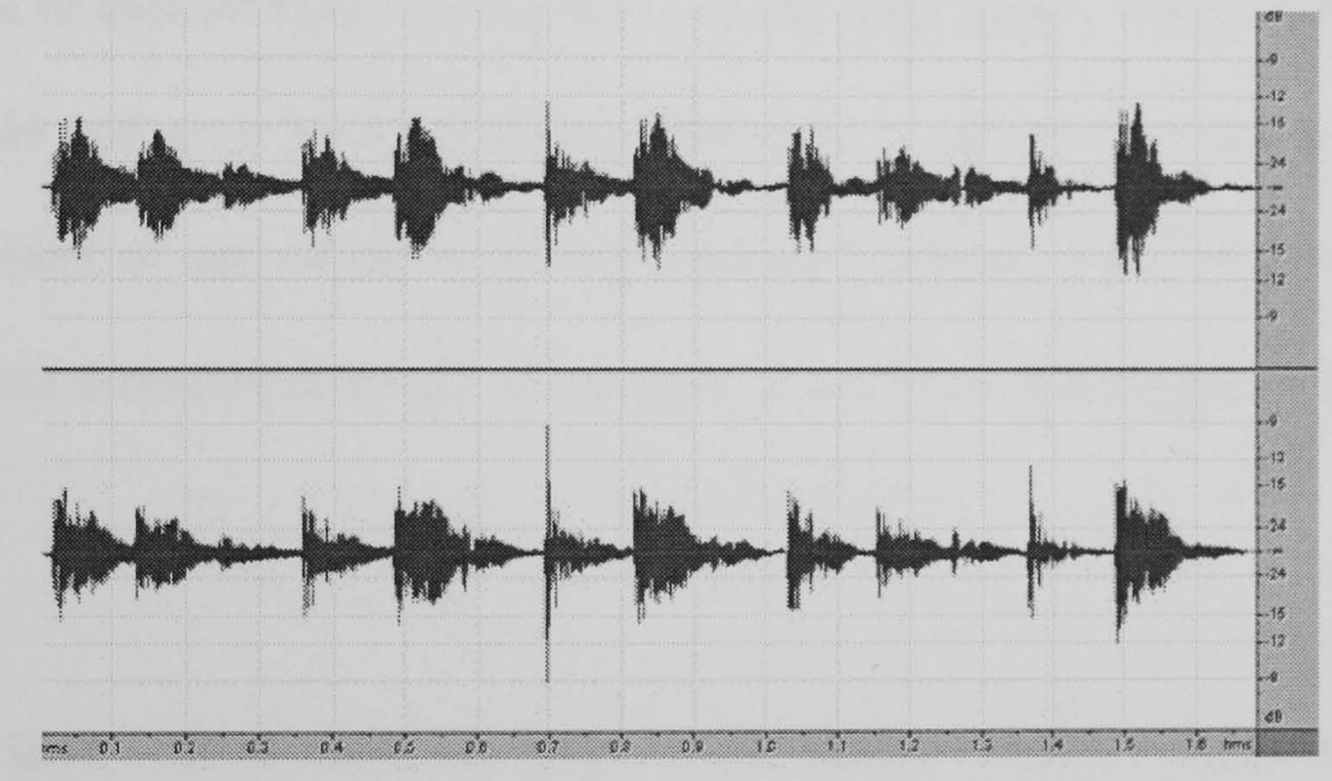
Figure 5.2 Displays of the width and location measurements made using the model developed by Mason *et al* [2005c]

5.2 Stimuli Creation

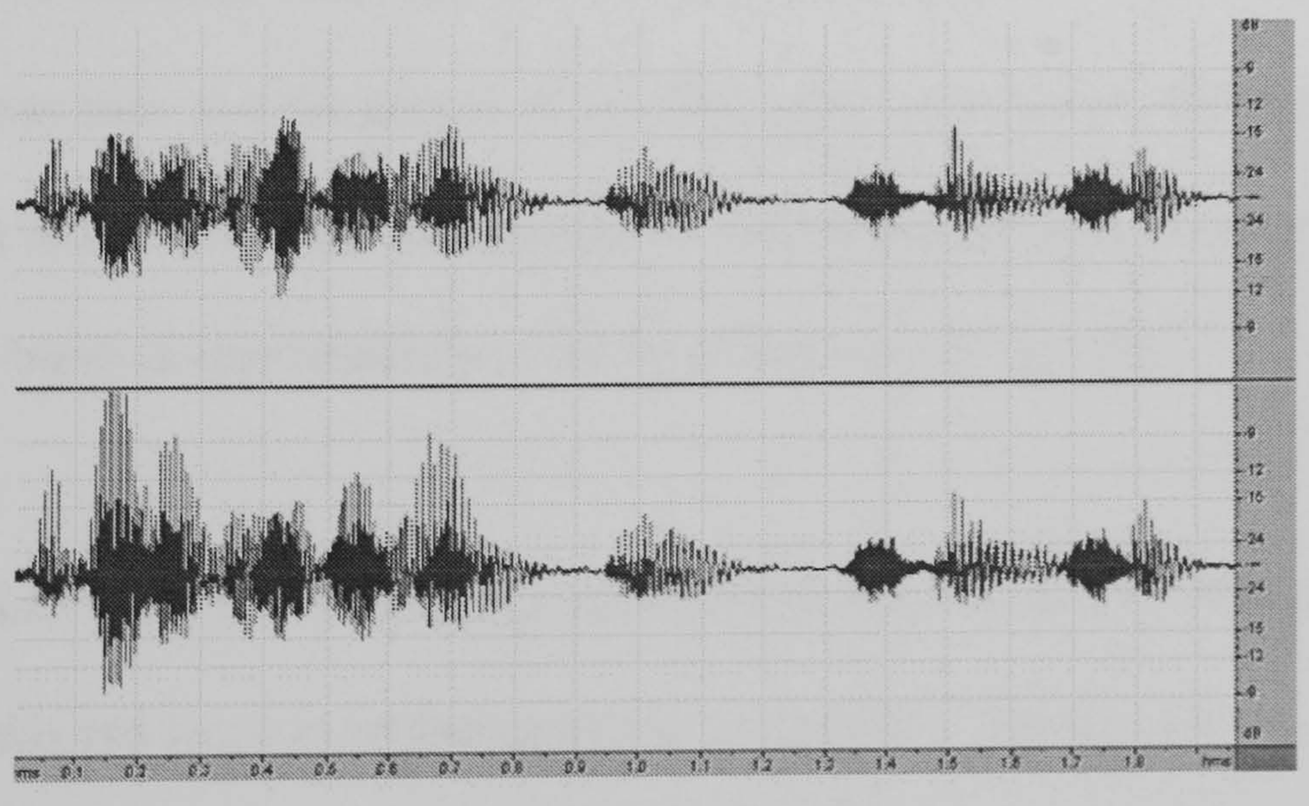
For the measurements, a set of binaural stimuli needed to be created first. For this, the stimuli used for the subjective experiments were reproduced in the same listening room that was used for the listening test and the created sound field was recorded using a dummy head placed in the listener position. The recorded binaural signals were converted into wave sound files so that they could be processed by the computer based software. The original stimuli were found to be too long to be processed. Therefore, selective excerpts of around 1.6 - 2.0 seconds were taken from the original stimuli and they were created as the new sound files for measurement. The selection of the excerpts was made so that they included representative temporal characteristics of the sound sources (e.g. note and bow changes for the cello, syllable changes for the speech, and ongoing hits for the bongo). Examples of the waveforms of the short anechoic stimuli created for each sound source are shown in **Figures 5.3** to **5.5**.



(a) Cello



(b) Bongo



(c) Speech

Figure 5.3 Waveforms of the binaural stimuli used for the objective measurement

5.3 General Overview of the Source Width and Location Measurements

The plots in **Figures B.1** to **B.9** (see Appendix B) show the source width measurements made in 22 different octave frequency bands for each experimental stimulus. It can firstly be observed that every frequency band exhibits a different pattern in the measured source width. Certain frequency bands have erratic variations in measurement, occasionally creating large peaks, which mean that there were rapid and great IACC changes. However, it can be seen that the predicted source widths of most of the frequency bands are crowded in the lower region of the plots, being measured relatively consistently in the range approximately between 20° and 40°.

Figures B.10 – B.18 show the plots of the source location measurements. The variations of locations over time are measured depending on the fluctuations of ITD and IID over time and the pattern of the fluctuations varies with different frequency bands. It can be seen from the figures that the average location of the source for all frequency bands is approximately in the range between 15° and 20°.

Figures B.19 – B.27 show the high resolution plots produced for representing both source width and location measurements in the previous figures at once. Therefore, these plots can provide a more integrated visual comparison. The darker and lighter parts in the plots represent the relative loudness levels of the sound signal depending on the number of frequency bands crowded in a certain measurement region as well as

the sound pressure level in each band. For example, the source widths measured for single frequency bands are represented as the lightest parts, while those for the largest number of frequency bands are in the darkest parts.

5.4 Comparisons between Measured Data and Perceived Data

The results of the statistical analysis for the perceived data were presented in Section 4.4. To recapitulate, the type of microphone array had a significant crosstalk effect on both source width and locatedness changes. On the other hand, the type of sound source was significant only for the source width change, while the type of acoustic condition was significant only for the locatedness change. In order to validate the usability of the current prediction model as a tool for analysing the physical factors causing the perceived effects, it is first necessary to compare the measured results with the perceived results and to discover whether the former can be used to predict the latter. If they do not match reasonably, it might be because the model does not implement all the aspect of complex cognitive aspects of spatial perception. Due to the nature of this model using time-varying measurement of IACC, there is no method of converting the magnitudes of measured differences between crosstalk-on stimuli and crosstalk-off stimuli into single numerical values for statistical analysis. Nonetheless, it is possible to measure the magnitudes of the visual changes indicated in the measured plots approximately, and compare the general trends between the statistical results and the measured results.

5.4.1 Difference between microphone arrays

When comparing crosstalk-on stimuli (LCR) and crosstalk-off stimuli (CR) with respect to the magnitude of change in measured source width, it was expected that the LCR would appear to have a greater source width than the CR since the addition of the crosstalk signal would decrease the IACC. From the plots shown in **Figures B.1 – B.9**, it appears in every source type and acoustic condition that the source width measurements of CR and LCR have similar trends for microphone array 1, although there are some minor differences in the variation pattern of certain frequency bands. Microphone array 4, on the other hand, shows more obvious changes between CR and LCR in general. There are more frequency bands that produce large peaks in the LCR, and therefore the plot of LCR shows more erratic variations in source width measurement over time. This means that when the crosstalk signal has a higher ratio of time difference to intensity difference, it causes a higher degree of decorrelation, leading to the perception of a greater source width. According to Mason [2002] asserting the close relationship between the interaural fluctuations over time and the IACC, this can be also explained as a more time-difference-based crosstalk signal producing a larger magnitude of fluctuations in ITD over time. These measurement results agree well with the results of the statistical analysis showing a significant difference between microphone arrays.

With regard to the source location changes between CR and LCR over time, microphone array 4 appears to have a greater magnitude of change than microphone array 1 for every source type and acoustic condition (see **Figures B.10 – B.18**).

Based on the principle of the model, this suggests that there was a greater magnitude of variation in interaural time and intensity differences over time with a stronger crosstalk signal. It can be also seen that the general magnitude of change between CR and LCR in location is similar to that in width. The variation in location over time might also be related to the decrease in locatedness, and if this is the case, the results also agree with the statistical results showing that the magnitude of decrease in locatedness due to crosstalk became greater with a more spaced microphone array.

The above observations suggest that the effect of the ratio between interchannel time and intensity differences on the perceived crosstalk effect could be well predicted by the current measurement model.

5.4.2 Difference between acoustic conditions

It was shown above that array 4 has more obvious changes between CR and LCR than array 1 overall. However, from **Figure B.1 – B.9** the magnitude of source width change between CR and LCR in array 4 appears to differ slightly with different acoustic conditions. However, the magnitude of difference between each acoustic condition appears to be minor compared to that between each microphone array, and the measured results do not seem to differ greatly from the statistical results which showed that the difference between each acoustic condition was statistically insignificant.

In respect of location predictions, it generally appears that the changes between CR and LCR for the anechoic stimuli are more obvious than those for the room or hall stimuli. This might be due to complex reflections and reverberation causing the interaural relationship of both crosstalk-off and crosstalk-on stimuli to become similar. However, the difference between the room and hall stimuli in the change between CR and LCR is relatively difficult to determine. The statistical results in Chapter 4 indicated that the magnitude of the locatedness change between CR and LCR significantly decreases from the anechoic to the room conditions, and the room to the hall conditions. It seems apparent that these statistical results agree with the above mentioned visual indication of the measured results.

However, the objective measurements designed to predict source width might not explain all aspects of the perceived effects. It is suggested that the relationship between the source width change and the acoustic condition is more related to certain psychoacoustic factors than to physical factors. It has been discussed earlier that the perception of reverberation is mainly related to the perception of LEV (Listener Envelopment) [Hidaka et al 1995, Bradley and Soulodre 1995]. Also, based on Griesinger [1996, 1997]'s hypothesis suggesting the separate perceptions of foreground and background streams, reverberation is related to the perception of BSI (Background Spatial Impression) rather than that of ASW (Apparent Source Width) and ESI (Early Spatial Impression). Therefore, the effect of source width increase that was perceived in the anechoic condition might have been more or less independent of the effect of reflections and reverberation in the room and hall conditions.

5.4.3 Difference between sound sources

For comparison between the magnitudes of source width change for each sound source, two cases can be considered separately depending on the behaviour of predicted values in different frequency bands. The first is for source widths determined by the frequency bands in which predicted values vary erratically, and the second is for those determined by the frequency bands crowded in the lower region of the plots in which predicted values remain relatively consistent. With respect to the former case, the cello appears to have more obvious changes between CR and LCR than the bongo and speech. However, with respect to the latter case, it appears that the speech source gives rise to the most obvious change between CR and LCR. The cello does not seem to give rise to much difference between CR and LCR in this case. For example, it appears that most of the peaks for the cello source arise erratically in a single or a small number of frequency bands (**Figures B.1 - B.3**), while those for the speech source arise relatively regularly in a larger number of frequency bands (**Figures B.7 - B.9**). The differences between the cello and speech sources can be clearly observed also in the high time resolution plots. For the cello source (**Figures B.19 - B.21**), it can be seen that the predicted source width of the lighter parts in the LCR is much greater than those in the CR, while there is no noticeable change in the source width of the darker parts. However, for the speech source (**Figures B.25 - B.27**), although there are not such rapid and erratic variations as with the cello source, the colour of the parts where there are major source width changes is darker compared to the cello source. This means that the speech source has a larger number of frequency bands giving rise to source width changes compared to the cello source, thus having greater

loudness. This might suggest that the source width changes for the speech source would have been more audible than those for the cello due to the richness of frequency bands that gave rise to the changes. In this regard, the measurement results seem to agree with the perceptual results showing that the speech source had the greatest crosstalk effect on the change in source width.

From a slightly different point of view, it might be further proposed that some kind of cognitive aspect was involved in the perception of source width change. That is, plausibility of the change in each frequency band might have been taken into account in the detection of audible changes. The 'plausibility hypothesis' of Rakerd and Hartmann [1986], which was introduced in Section 2.2.3, suggests that unreasonably large ITD cues produced by the interaction between direct sound and room reflections are ignored by the brain in the process of localisation and only plausible ITD cues are used. Similarly, the rapid and large variations in IACC (or rapid and large ITD fluctuations) shown in the measurement results would have been recognised as implausible cues by brain, and therefore disregarded in the process of source width perception.

Based on the reflection studies that reported the frequency dependency of source width perception, it might also be questioned whether the actual frequency itself had an effect on the difference between the sound sources. For example, based on the reports of Morimoto and Maekawa [1988], Hidaka *et al* [1995] and Mason *et al* [2005b], the low frequency dominance of the speech source over the bongo or cello source could be claimed to be the main factor for the greater source width change.

However, it can be observed from **Figures B.28 - B.60** that each sound source has similar patterns of source width changes for similar frequencies, and this suggests that the differences between sound sources were not simply dependent on the frequency components of the sources. This issue will be further discussed in Section 5.5.

For the location measurements, it appears that the physical changes caused by the crosstalk signal are most obvious for the cello source as the patterns of the occasional large peaks for a number of frequency bands become more erratic in the LCR compared to the CR. Compared to the cello, the speech and bongo sources appear to have smaller changes between the CR and LCR in the patterns of location variation over time. That is, the major location changes are observed for a smaller number of frequency bands. This means that the cello produced a greater magnitude of interaural fluctuations over time compared to the speech and bongo, and this can be explained as follows. In terms of the temporal characteristics of the sound sources, the cello has a more continuous form than the speech and bongo, even though there are ongoing fluctuations in the envelope caused by note and bow changes. In addition, the cello has the slowest onset time for a new sound event, and the longest duration between each onset. Therefore, the cello has a higher potential for continuous interactions between the wanted signals (C and R) and the crosstalk signal (L) compared to the relatively more transient speech and bongo, thus having greater interaural fluctuations over time. The dominance of continuous sound over transient sound in the magnitude of interaural fluctuations over time was confirmed by Mason *et al* [2005a]. However, it is difficult to directly determine the correspondence between the measured results and the perceived results. This is because the physical

predictors of the locatedness attribute, unlike the source width attribute, have not been much researched yet, and therefore there is a lack of information about how the temporal variations of source location for certain frequency band signals would affect locatedness perception. That is, even if there were some measured differences between CR and LCR in the magnitude of source location change for certain frequency bands, perceptually this might not be important for locatedness change.

5.4.4 Discussions

From the above investigations, it generally appears that the current measurement model provides reasonable predictions about the perceptual effects of interchannel crosstalk on source width changes. It is apparent that the measured results strongly agree with the perceived results showing that the ratio of interchannel time and intensity differences in microphone arrays has a dominant effect on the magnitude of source width increase due to interchannel crosstalk. The effect of acoustic condition on the perceived source width increase does not seem to be explained fully by the visual indications of the measured plots, and it is proposed to be more related to the psychoacoustic effect of multiple reflections and reverberation with a wide range of delay times on the perceptions of different spatial impressions. The perceived effect of sound source type is also not directly predicted by the visual indications of the measured results. However, based on the hypothesis made in the above section, the measured results seem to match the perceived results reasonably well.

The location measurement results show that the ratio of time and intensity differences exhibited by the microphone array also has an obvious effect on the difference between crosstalk-off and crosstalk-on stimuli, which seems to match the perceived results. For the effects of acoustic condition and sound source type, on the other hand, it is not as easy to define the relationship between the measured data and the perceived data. However, as mentioned above, the objective predictors of the locatedness attribute have not been established yet, and therefore simple temporal variations in source location that are dependent on the fluctuations in ITD and IID could not be directly applied for the prediction of the easiness of localisation. Therefore, further investigation needs to be conducted to understand the mechanism of locatedness perception.

A possible hypothesis for the effect whereby interchannel crosstalk decreases the perception of locatedness could be proposed based on the combined effect of the temporal characteristics of source and the rate of ITD fluctuation. Firstly, the onset transient of each sound event, such as each note and syllable for complex musical source signals, would be responsible for operating the precedence effect, leading to an instantaneous localisation of the source. As shown in Mason *et al* [2005a], the magnitude of ITD fluctuation would be very small at the onset, and therefore there would be little variation in location. This could be considered to be when the localisation is the most accurate, leading to the locatedness being at the highest degree. During the ongoing part of the sound event, however, the role of precedence effect in accurate sound localisation would become less dominant as suggested in the literature, whereas the pattern of ITD fluctuation would become more erratic since the crosstalk

signal interacts with the wanted signals more, depending on the frequency bands involved. Here the degree of locatedness might be determined by the rate of ITD fluctuation. As discussed in Section 2.6, if the rate of the ITD fluctuation is low, the sound image will be perceived to be moving. However, this effect will disappear at higher rates and the image will be perceived to have an increased source width. This is based on the 'localisation lag' effect [Blauert 1972]. Since the spectral characteristics of the musical signals are complex, the fluctuation rate would be likely to vary randomly from low to high over time. From this, it can be considered that both the locatedness and source width attributes would be randomly perceived during the length of the ongoing part. The rate of variation between the locatedness and source width would be likely to be very high and this might be the reason for the high correlation between the locatedness and source width perceptions, which was shown in Chapter 4. It is further considered that the rate of ITD fluctuation would be likely to vary for different frequency bands, and if this is the case, the locatedness perception would be frequency-dependent. For example, the frequency bands fluctuating in ITD at low rates over the duration of the ongoing part would contribute to the perception of locatedness, while those at high rates would contribute to the perception of source width. However, the lengths of onset and ongoing parts of a sound are also considered to be important for locatedness. For example, it seems apparent that a series of transient hits of bongo sound will be highly located. A continuous speech sound having frequent and rapid syllable changes or a continuous cello sound having fast and strong note changes might have reasonably good locatedness as the precedence effect will be operated constantly. On the other hand, a cello or trumpet sound having a note that rises slowly and is sustained for a long duration is likely to

have poor locatedness.

From the above discussion, it is considered to be hard to further attempt to establish the relationship between the perceived locatedness-changing effects of interchannel crosstalk and their physical causes by using the results of the location measurements obtained from the current model. The investigation into the perceptual mechanism of the locatedness attribute is considered to be a challenging research topic that is beyond the scope of this project and worthy of further research. Therefore, the following sections will discuss the aspects of source width measurements only.

5.5 Influence of Frequency Components on Increase in Source Width

It was shown from the visual indications of the measured plots that the pattern of source width change between CR and LCR varies depending on the frequency band concerned. In order to investigate the frequency dependency of the source-width-increasing effect of interchannel crosstalk, the pattern of each frequency band was analysed in detail for each sound source. Only the stimuli for microphone array 4 were considered in this investigation because array 1 appeared to indicate no obvious changes between CR and LCR overall.

5.5.1 Cello

From the observation of the measured differences between CR and LCR for each frequency band of the anechoic cello stimuli, it was possible to separate the frequency bands into four groups based on the variation patterns. The centre frequencies included in each group are shown below.

- Group 1: 150, 250, 455, and 570Hz
- Group 2: 700, 845, 1000, 1175, and 1375Hz
- Group 3: 1600, 1860, 2160, 2510, 2925, 3425, 4050, 4850, and 5850Hz
- Group 4: 7000, 8600, and 10750Hz

Figure B.28 shows the measurements made for the frequency bands of group 1. It can be firstly observed here that there is no great difference between CR and LCR. The source width measured over time is relatively constant in both CR and LCR. However, for the frequency bands of group 2, the changes between CR and LCR become obvious in that the LCR has a greater magnitude of source width and a more erratic pattern of variations than the CR (**Figure B.29**). In comparison with the loudness plot of the source signal shown in the figure, it can also be observed that the peaks of the measured source width over time for each frequency band in the LCR appear to correspond to the peaks of the loudness envelope of the frequency band signal. The overall plots of the frequency bands also appear to be largely related to the envelope of the overall waveform shown in **Figure 5.3**. **Figure B.30** shows the measurements of the frequency bands of group 3. Even though there are obvious changes observed between CR and LCR for these frequency bands also, the patterns of

source width increase for them appear to be different from those for the lower frequency bands shown above. The changes are mainly due to the random and sharp peaks, and therefore no envelope dependency is found. On the other hand, the measurements of the highest frequency bands of group 4 shown in **Figure B.31** have similar trends to those of the lowest frequency bands. That is, the source width is mostly constant over time, and no obvious change between CR and LCR is observed. This is likely to be due to the simulation of the breakdown of phase locking that is included in the process of the current model. It was mentioned earlier that the human hearing system fails to detect fine temporal details at high frequencies due to the breakdown of phase locking, and the perceived width becomes dependent on the IACC of the signal envelope rather than the signal itself [Mason *et al* 2004]. From this, it is considered that the source widths of those higher frequency bands were determined by the envelope of the signal having relatively low frequencies.

The measurements of group 1 for the room- and hall-reverberant stimuli reflect the effect of the reverberation signals on the increase in perceived width, as shown in **Figures B.32** and **B.36**. It can be seen that the large peaks occur more continuously for the hall-reverberant stimuli compared to the room-reverberant stimuli, and this might be due to the longer decay tail of the hall reverberation causing large interaural fluctuations more continuously. However, it can be commonly seen in both cases that the measurements vary in regular patterns depending on the temporal characteristics of the source. In comparison between the measurements and the signal waveform, it can be seen that the large peaks occur at the dips of the signal envelope in general. This is likely to be because the decorrelated reverberation

signal caused large fluctuations in ITD and IID in the space between the notes or between the vibratos, and therefore decreased the IACC greatly. However, in terms of the magnitude of the overall predicted source width, it appears that there is no obvious difference between CR and LCR. This means that the crosstalk signal did not have much effect at the low frequencies where the reverberation had much effect on the width increase. However, for the measurements of group 2 frequency bands, the effect of reverberation decreases as shown in **Figure B.33** and **B.37**. There are more obvious changes between CR and LCR for these frequencies and they appear to occur at the peaks of the loudness envelope for each frequency band. The measurements of group 3 frequency bands for the room-reverberant stimuli show a similar trend to those for the anechoic stimuli, having a large number of random and sharp peaks in the LCR causing the differences from the CR (**Figure B.34** and **B.38**). Similarly to group 1, group 4 shows some effects of reverberation at the dips of the envelope, but no obvious change in the magnitude of source width is observed between CR and LCR (**Figure B.35** and **B.39**).

5.5.2 Bongo

For the measurements of the bongo stimuli, the frequency bands can be separated into four groups in the same manner as shown for the cello stimuli. Firstly, it can be seen from **Figure B.40** that the measurements of the frequency bands of group 1 for the anechoic bongo stimuli change regularly over time depending on the envelope of the signal in both CR and LCR. The peaks appear to occur at the offset of each transient

hit, and this is likely to be due to the influence of the decorrelated resonant sound during the decay. However, the differences between CR and LCR do not appear to be dominant for these frequencies. For the frequency bands of group 2, there are more noticeable differences between CR and LCR as can be observed in **Figure B.41**. The LCR appears to have slightly greater widths in general and more large peaks than the CR. However, these differences seem to be relatively small compared to the differences observed at the same frequency bands of the cello stimuli. However, it is interesting to note that the measurement of each frequency band in the LCR appears to be related to the signal envelope. Whereas the peaks of the measurements for the lower frequency bands occur at the dips of the signal envelope, the peaks for these frequency bands occur at the peaks of the signal envelope. **Figure B.42** indicates that the measurements change more randomly at the frequency bands of group 3, but the differences between CR and LCR are relatively small. The measurements for the highest frequency bands shown in **Figure B.43** (group 4) appear to be similar to those for the lowest bands in that the noticeable variations in the measurement occur at the dips of the signal envelope although their magnitudes are smaller. It can also be seen that the differences between CR and LCR are negligible.

It can be generally observed from **Figures B.44 – B.51** that the measurements of both the room- and hall-reverberant bongo stimuli have similar trends to those of the anechoic stimuli in terms of the position of temporal variation. For example, large variations for the group 1 frequencies occur at the dips of the signal envelope while those for the group 2 frequencies occur at the peaks. It is interesting that the hall and room reverberations do not show any noticeable difference in terms of the duration of

the large variations. This is contradictory to the case of the cello stimuli for the same groups of frequency bands, which show that the hall reverberation causes large peaks more continuously after the offset of a note than the room reverberation. A possible explanation for this is as follows. At low frequencies the onset transient energy of a bongo hit might have perceptually masked the energy of the long and diffused reverberation generated at the offset of the previous hit, thus resulting in a high interaural cross-correlation at every new hit regardless of the length of the reverberation signal. It can also be seen that the peaks for the reverberant stimuli at the low frequencies are greater than those for the anechoic stimuli, and this is likely to be due to the maximised effect of reverberation on decorrelating the ear signals after the end of the sound as pointed out by Griesinger [1996]. For the frequency bands of group 2, there appear to be fewer large and sharp peaks for the reverberant stimuli compared to the anechoic ones. However, the peaks are observed only for a couple of frequency bands and therefore their effects on the increase of perceived width do not seem to be great. The overall magnitudes of differences between CR and LCR do not appear to vary much between the anechoic and the reverberant conditions.

5.5.3 Speech

It was shown above that the frequency bands of the cello and bongo stimuli can be separated into four groups depending on the relationship between the signal envelope and the pattern of temporal variation in measurement as well as the magnitude of difference between CR and LCR. Even though the frequency bands of the speech

stimuli can also be separated into groups depending on the same principles, the number of groups and the range of frequencies belonging to each group differ from the cello and bongo. There are a total of three groups of frequency bands that show different trends in the measurements, as indicated below:

- Group 1: 150, 250, 455, and 570Hz
- Group 2: 700, 845, and 1000Hz
- Group 3: 1175, 1375, 1600, 1860, 2160, 2510, 2925, 3425, 4050, 4850, 5850, 7000, 8600, and 10750Hz

The measurements made for the group 1 frequency bands of the anechoic speech stimuli are shown in **Figure B.52.** It can be seen that for both CR and LCR, the measurements increase at the dips of the signal envelope, in other words in the space between each syllable. However, the difference between CR and LCR appears to be very small. For the frequency bands of group 2, as shown in **Figure B.53**, there are more obvious differences in the measurements between CR and LCR for these frequency bands. The temporal variations in the measurements for the LCR appear to occur at the peaks of the signal envelopes of the corresponding frequency bands. Finally, **Figure B.54** indicates that the measurements made for the frequency bands of group 3 have more erratic temporal variations compared to those for the lower frequency bands. Due to the sharp and random peaks, none of the frequency bands appear to exhibit envelope-dependency. However, it is interesting to note that with all the frequency bands of this group considered together, there are certain temporal regions where the large peaks become crowded (e.g. in the region around 1.3 - 1.4

seconds), and the envelope of the crowded peaks appear to be related to the signal envelope.

In general the measurements of the reverberant stimuli shown in **Figures B.55 – B.60** have similar trends to those of the anechoic stimuli described above. It is noticeable that the magnitudes of the large variations observed for the lowest frequency bands increase slightly as the reverberation becomes more decorrelated. However, there is no obvious change between CR and LCR in the magnitude of the measured width, which suggests that the crosstalk does not have much effect at these frequencies.

5.5.4 Discussions

The influence of frequency on the effect of interchannel crosstalk was investigated with respect to the source width attribute using the current measurement model. The main findings from this investigation are summarised and discussed below.

Firstly, for the frequency bands up to 570Hz, the comparisons between the crosstalk-on and crosstalk-off stimuli of all source types in the measurements of source width showed no obvious differences. This suggests that the source width increasing effect of interchannel crosstalk is small at low frequencies. Even though the addition of room or hall reverberation caused large temporal variation in the measurements regularly at the offset of sound, there was no obvious change in the magnitude of difference between the crosstalk-on and crosstalk-off stimuli.

Secondly, at the middle frequencies, the crosstalk appeared to have the most obvious effect on the increase in the measured width. It was possible to separate the middle frequencies into two groups depending on the pattern of temporal variations in the measurements and the magnitude of the variations. For the anechoic cello and bongo sources, the lower-middle frequency bands of the crosstalk signal from 700Hz to 1375Hz appeared to cause regular temporal variations in the measurements in that the width increase generally occurred at the onset or ongoing part of the signal envelope. A similar effect was observed for the anechoic speech source, but the upper threshold of the frequency bands was lower (1000Hz). For the upper-middle frequency bands of the anechoic cello and bongo sources (1600Hz – 5000Hz), the measurements had rapid and random temporal variations with large peaks. For the anechoic speech source, the range of the frequency bands having a similar effect was greater compared to the cello and bongo sources, covering high frequencies up to 10750Hz.

Finally, for the cello and bongo sources, the source width increasing effect of crosstalk was greatly diminished for the high frequency bands from 7000Hz to 10750Hz. Also the patterns of the temporal variations in the measurements became very similar to those at the low frequencies. This is likely to be because the IACC was measured by the envelope of the signal rather the signal itself based on the process of the current model simulating the loss of fine temporal details in the measurement of IACC [Mason *et al* 2004].

From the measurement results, it is interesting to compare the frequency-dependent patterns of the crosstalk signal and the acoustic reflections or reverberation. At the

low frequency bands, the addition of room or hall reverberation certainly increased the general predicted width of the stimuli compared to the anechoic condition while keeping the magnitude of the small difference between crosstalk-on and crosstalk-off measurements. This occurred at the offsets of the sound rather than the onsets; these are where the maximum interaural fluctuations are produced from the reflections and reverberation [Griesinger 1996]. However, at middle frequencies, the effect of reflections and reverberation disappeared dramatically whereas the effect of crosstalk became dominant. These findings seem to validate the dominant role of low frequency components of reflection and reverberation on width perception which was discussed in Chapter 2, but also give rise to a question of why the interchannel crosstalk did not have a similar pattern of frequency dependency. The crosstalk signal and the reverberation basically have different natures. Firstly, crosstalk is a single signal that is derived from the source itself while reverberation consists of multiple reflections that are indirect and decorrelated from the source. Secondly, the delay times of acoustic reflections are normally much longer than that of the crosstalk signal. This might suggest that the different characteristics of the crosstalk and the reflection or reverberation produce different perceptual attributes that cannot be simply defined as source width. According to Griesinger [1996], as discussed in detail in Section 2.3.1.2, the width perception caused by the low frequency energy of reflection or reverberation after the offset is related to ESI or BSI rather than source width. Griesinger [1997] hypothesises that source width is perceived only when the reflection arrives within the onset duration of the direct sound. From this point of view, the range of the delay time of the crosstalk signal used in the current studies (0.5 – 1.1ms) is small enough to contribute to the perception of source width.

It was also found that the source width increase caused by crosstalk at the lower-middle frequencies mainly occurred around the onsets of the signal envelope rather than the offsets, which is contradictory to the case of the effect of reflections and reverberation. This might also be due to the fact that the crosstalk signal arrives within the onset duration of the wanted signals. In other words, the dominance of the middle frequency effect seems to be due to the relationship between the pattern of interaural fluctuation and the delay time of the lagging signal. The range of delay times involved in the crosstalk signal corresponds to approximately the wavelength of signals around 1,000 Hz, and therefore strong interaural time and intensity fluctuation (thus low IACC) might have occurred at the middle frequencies. However, with reflections having longer delay times, the effect of interaural fluctuation on source width increase might have occurred at the lower frequencies. Based on this assumption, it might be generally suggested that the measured effect of the frequency component of the secondary signal on source width perception might be dependent on the range of delay time of the secondary signal. However, it is dubious that the low frequencies of the crosstalk signal, which did not cause any obvious measurement changes, would have had virtually no effects on the increase of perceived source width since a number of researchers including Morimoto and Maekawa [1988], Hidaka *et al* [1995] and Mason *et al* [2005b] have suggested that low frequencies would be important for the 'perception' of source width independent of the 'measurement' of IACC. It might be that different perceptual source-related attributes could be perceived dependent on the frequencies of secondary signal but independent of the IACC measured for those frequencies. However, the audibilities of those frequency-dependent attributes might be related to the magnitude of the measured IACCs. This

requires further investigation.

It was observed from the results that the upper-middle frequencies generally had the greatest variations in width measurements among all frequency bands. However, as suggested earlier, it might be that these variations were implausible for perception as they are too rapid and large. In other words, the brain might use some kind of cognitive process (or rapid decision making process) for interpreting the width of a sound depending on the frequency band of the sound. If this is the case, the upper-middle frequency bands might have been cognitively excluded in the subject's width judgments in the perceptual experiments described in Chapter 4.

The frequency dependency shown in the above measurement results might suggest that the perception of source width increase due to interchannel crosstalk is mainly a middle frequency phenomenon. However, a further subjective investigation would be required to confirm this. It might also be hypothesised from the measured results that different kinds of width attributes could be perceived for different frequencies. In other words, the source width attribute of the interchannel crosstalk effect might include perceptual sub-attributes depending on the frequency components of the crosstalk signal. If this is the case, new terminologies describing the detailed perceptual attributes will be required. For investigating these, a subjective elicitation experiment needs to be conducted, comparing crosstalk-on and crosstalk-off stimuli with the crosstalk signal band-pass filtered for various centre frequencies. This investigation can also be extended to a further investigation into the effect of frequency components of early reflections on source width perception. Although

there have already been a number of investigations conducted on this topic as introduced in Chapter 2, there seems to be no definite answer yet. This seems to be mainly because different researchers tend to use different terms for the same effect, or the same term for different effects, due to the lack of standard way of defining subtle perceptual attributes.

5.6 Summary

This chapter described the investigation into the objectively measured effects of interchannel crosstalk. The measurement model chosen for this investigation was an IACC-based width and location prediction model that was developed by Mason *et al* [2004]. This model was particularly useful for the current studies since it employs various frequency bands and loudness of complex and continuous musical signals. Firstly, the correspondences between the measured data and the perceived data were examined. For this, the measurements of the crosstalk-on stimuli and crosstalk-off stimuli were compared with respect to the independent variables, which were the type of microphone array, the type of sound source, and the acoustic condition. It was found that the measurements for the source width attribute matched the perceived data reasonably well. The temporal change of source location was also measured. The measured result showed a very similar trend to the statistical result of the locatedness attribute with respect to the type of microphone array, although it was difficult to judge the similarity between them for the other independent variables. However, the psychoacoustic and physical mechanism of the locatedness attribute has not been

established yet, and therefore the locatedness change might not be defined as the temporal change of measured location based on the simple interaural time and intensity relationship. A potential hypothesis on locatedness perception was proposed in Section 5.4.4. The influence of the frequency component and its relationship with the signal envelope were investigated with regard to the source width increasing effect of interchannel crosstalk. It was found that at low frequencies up to the centre frequency of 570Hz there was no obvious crosstalk effect although the addition of reflections and reverberation caused the general width of the stimuli to be increased at the offsets of the signal envelope regardless of the existence of crosstalk. At the middle frequencies up to around 1000Hz, the source width increasing effect of crosstalk was most dominant, having a positive correlation with the onsets of the signal envelope. At the higher frequencies, the measurements became largely erratic and the envelope dependency disappeared. These results might suggest that the significance of low frequency energy on the increase of source width, which has been largely accepted in concert hall acoustics research, is dependent on the delay time of the reflected signal and its relationship with the onset duration of the direct signal. They also seem to suggest that the perception of increased source width due to interchannel crosstalk is mainly a middle frequency phenomenon. However, the frequency and envelope dependencies that were observed from this investigation seem to lead to a hypothesis that different frequencies cause different types of source width perception. To confirm the above findings, further subjective and objective investigations are required.

6 SUMMARY AND CONCLUSIONS

This final chapter summarises the research and experimentation documented in this thesis and outlines the main conclusions resulting from the associated work. Further works that may be extended from the research described in this thesis are also discussed.

6.1 Summary and Conclusions

6.1.1 Chapter 0

This opening chapter introduced the background and aims of the research described in this thesis. Interchannel crosstalk is an inevitable artefact in the design of multichannel microphone technique and its effects on perceived sound quality have been an issue of debate recently. However, to date no experimental data has been available on the perceptual effects of interchannel crosstalk and therefore there is consequently no experimental data to which sound engineers can refer when attempting to control interchannel crosstalk in the design and application of multichannel microphone technique. The current research was therefore undertaken in order to obtain a clearer understanding of the perceptual properties of interchannel crosstalk. The specific aims were as follows:

- To elicit perceptible auditory attributes of interchannel crosstalk and weight their relative audibilities.

- To analyse significances of the effects of such physical variables as microphone array configuration, sound source type and acoustic condition of recording space on the perception of interchannel crosstalk.
- To map the relationship between the perceptual effects of interchannel crosstalk and relevant physical cues.
- To examine the subjective preferences for sound images created with interchannel crosstalk.

6.1.2 Chapter 1

In chapter 1, the summing localisation theory, which forms the basis for 2-0 and 3-2 stereophonic phantom imaging, was reviewed. Particularly, the individual influence of the interchannel time difference (ICTD) cue or interchannel intensity difference (ICID) cue on specific phantom image locations and the trade-off relationship between the two spatial cues, which becomes the basis for designing near-coincident multichannel microphone technique, were discussed. The concept of stereophonic recording angle (SRA) in stereophonic microphone technique was covered and the basic design and operational principles of conventional 2-0 stereophonic microphone techniques were briefly reviewed. The unique localisation characteristics of 3-2 stereophonic reproduction were discussed, including the limitations in respect of side image localisation. Current 3-2 stereophonic microphone techniques were divided into those with front and rear separation and those

with five-channel main microphones. The design and operational principles of those microphone techniques were reviewed in detail and their characteristics with respect to interchannel crosstalk were discussed. From this, it was suggested that:

- For the microphone techniques with front and rear separation, which would be likely to prove more practical and flexible than the five-channel main microphone techniques in terms of controlling direct sound localisation and spatial impression separately, interchannel crosstalk would be a matter of importance only for the front arrays due to the typically large distance between the front and rear arrays.
- None of the current three-channel front microphone techniques seems to be perfectly optimised with regard to interchannel crosstalk.
- Interchannel crosstalk might not necessarily decrease the perceived sound quality.
- Interchannel crosstalk in three-channel microphone technique might not necessarily be problematic with regard to balanced phantom image distribution across L-C-R, but would primarily influence the perception of certain auditory attributes depending on the interchannel time and intensity relationship involved in the signal.

6.1.3 Chapter 2

Chapter 2 reviewed the literature related to the perceptual effects of reflections in concert halls and rooms since it was considered likely to act as a useful basis for understanding the perceptual effects of interchannel crosstalk in multichannel

recording and reproduction. The auditory attributes influenced by the addition of reflection were divided into three main categories: localisation, spatial impression and timbre, but this review was solely focused on the categories of localisation and spatial impression since the timbre-changing effects had not been studied widely or documented in detail in the literature. Various aspects of the precedence effect, which becomes the main psychoacoustic principle for accurate auditory localisation in reflective environment, were discussed. The effects of such physical cues as the onset transient and low frequency content on triggering the precedence effect were also considered. Additionally, it was noted that the precedence effect was not a simple low-level physiological phenomenon but involved a high-level cognitive process of human perception. The conceptual properties of spatial impression (SI) and different perceptual paradigms of apparent source width (ASW) and listener envelopment (LEV) were reviewed. Various physical parameters that had been considered for measuring SI in concert hall acoustics were discussed in detail, including intensity and direction of reflection, frequency component of sound source, interaural cross-correlation coefficient (IACC) and fluctuations in ITD and IID. Additionally, the relationship between IACC and fluctuations in ITD and IID was considered. From this review, it was hypothesised that:

- Since both reflection and interchannel crosstalk have the similar form of a delayed secondary signal and most of the reflection studies were conducted by means of simulation using a stereophonic reproduction system, the results of these studies might become the basis for hypothesising the perceptual effects of interchannel crosstalk in multichannel microphone technique
- Accuracy of phantom image localisation in multichannel microphone technique

would be affected by the presence of interchannel crosstalk depending on the temporal and spectral characteristics of the source signal.

- Interchannel crosstalk would cause the perceived source width to be increased depending on such factors as intensity of crosstalk signal, frequency components of source signal, IACC and fluctuations in ITD and IID over time.
- Interchannel crosstalk might affect various timbral attributes depending on the temporal and spectral characteristics of source signal. However, the perceived magnitudes of the tone colouration effects of interchannel crosstalk might be minor compared to those of reflections due to the relatively short delay time of crosstalk signal.

6.1.4 Chapter 3

Chapter 3 described subjective experiments that were undertaken to investigate the effects of interchannel time and intensity relationship and sound source type on the perception of phantom image attributes in 2-0 stereophonic reproduction using trained sound engineers. In the first experiment, the perceptual attributes of stereophonic phantom images were elicited through subjective comparisons between monophonic source images and the corresponding stereophonic phantom images. From this experiment it was concluded that:

- The perceptual attributes of 2-0 stereophonic phantom images elicited for piano, trumpet and speech sources comprised three spatial attributes (source focus, source width and source distance) and three timbral attributes (brightness,

hardness and fullness).

In the second experiment, the perceived magnitudes of those attributes were graded for various interchannel time and intensity relationships and sound source types. The resulting data were statistically analysed and the correlation between each attribute was discussed. From this experiment, it was concluded that:

- The effect of sound source type was significant for all attributes except source distance.
- The effect of interchannel time and intensity relationship was significant only for source focus and source width attributes.
- Source focus and source width were correlated at a high level.

6.1.5 Chapter 4

Chapter 4 described a series of subjective experiments that were conducted to investigate the perceptual effects of interchannel crosstalk in 3-2 microphone technique using trained sound engineers. Firstly, the perceptual attributes of interchannel crosstalk for the cello, bongo and speech sources were elicited through subjective comparisons between crosstalk-off (CR) and crosstalk-on (LCR) stimuli that were created using the interchannel relationships involved in various types of critical linking three-channel microphone arrays. Then, the relative perceptual weightings of those attributes were graded and the attributes that were perceptually most dominant were selected. From this

experiment, it was concluded that:

- The perceptual attributes of interchannel crosstalk that were elicited for cello, bongo and speech sources comprised source width, locatedness, source direction, fullness, source distance, hardness, brightness, diffuseness, naturalness, envelopment and phasiness.
- Source width and locatedness were the most dominant of these.

The perceived magnitudes of the effects of microphone array type, sound source type and acoustic condition on the perceived magnitudes on the selected crosstalk attributes, which were source width and source locatedness, were graded and the resulting data were statistically analysed. Additionally, the correlation between the two attributes were analysed for each microphone array type. From this experiment, it was concluded that:

- Changes in microphone array type from a more coincident array to a more spaced array resulted in significant increases of perceived source width and significant decrease of perceived locatedness of phantom images.
- Sound source type was a significant factor for the source width increasing effect of interchannel crosstalk but not for the locatedness decreasing effect. The speech source, which had the broadest frequency range, caused the greatest increase in source width.
- Acoustic condition was a significant factor for the locatedness decreasing effect of interchannel crosstalk but not for the source width increasing effect. As the reverberation became more diffused, the locatedness decreasing effect became less perceptible.

- For each microphone array type, the source width- and locatedness-changing effects of interchannel crosstalk had a low correlation.

The effect of interchannel crosstalk on sound quality preference was also investigated in both controlled and practical manners using trained sound engineers. Firstly, the crosstalk-on and -off stimuli that had been used for the previous experiments were compared for preference. Then, 'real world' recordings made with three-channel microphone techniques of OCT and ICA-3 were compared for the musical sound sources of string quartet ensemble, solo percussions, solo violin and solo piano. From these experiments, it was concluded that:

- In the controlled experiment, there was no strong preference found on either crosstalk-on or crosstalk-off stimuli.
- Microphone array type, sound source type and acoustic condition had no significant effects on the preference grading of the controlled stimuli.
- Locatedness and source width attributes were the most salient preference cues for the controlled stimuli.
- From the results of the comparison between the recordings made with the OCT and ICA-3 microphone techniques, it was suggested that the preference for interchannel crosstalk would be likely to be dependent on the characteristics of the sound source.

6.1.6 Chapter 5

The objective measurements of the effects of interchannel crosstalk, which were made for the controlled stimuli from the perceptual experiments described in Chapter 4 using the IACC-based source width and location prediction model developed by Mason *et al* [2004], were described in chapter 5. Firstly, the correspondences between the perceived data and the measured data were investigated by comparing the statistical results presented in Chapter 4 and the visual indications in the measured plots. The influences of the independent variables of microphone array type, sound source type and acoustic condition on the perceived results were also discussed, based on the measured results. From this investigation, it was suggested that:

- Interchannel crosstalk of a more spaced microphone array would cause a higher degree of interaural decorrelation than that of a more coincident microphone array, thus leading to the perception of a greater source width.
- The source width increasing effect of interchannel crosstalk would be perceptually independent from the influence of diffused reverberation in the room and hall conditions since reverberation would be likely to contribute to a perception of listener envelopment (LEV) or background spatial impression (BSI).
- The source width increasing effect of interchannel crosstalk would be more perceivable for the sound sources that have more plausible variations in IACC over time.

It was also investigated how the frequency components of interchannel crosstalk influenced on the measured source width. From this, it was concluded that:

- At low frequencies up to around 570Hz, there was no obvious width change due to interchannel crosstalk. On the other hand, room and hall reverberation caused the measured widths of both crosstalk-on and crosstalk-off stimuli to be increased at the offsets of signal envelopes. However, this effect of reverberation disappeared at the higher frequencies.
- At low-middle frequencies up to around 1000Hz, interchannel crosstalk caused obvious and regular width increases at the onsets of signal envelopes.
- At high-middle frequencies above around 1000Hz, interchannel crosstalk caused obvious but erratic width increases, which might have been implausible for perception. No envelope dependency was observed.
- At high frequencies above around 7000Hz, the effect of interchannel crosstalk on width increase was minor.
- The frequency dependency of source width increasing effect of secondary signal might be related to the delay time of the crosstalk signal.

6.2 Further Work

It was shown in Chapter 5 that the increase of width due to interchannel crosstalk, measured using an auditory model based on time-variant IACC, was produced by the middle frequencies of the crosstalk signal around 1000Hz rather than the low frequencies. However, it needs to be determined whether the measured results would correspond to the perceptual results with regard to the source width increasing effects of other frequency components of crosstalk signal, since a number of authors reported

that the low frequencies caused a significant source width increase independently of IACC value [Morimoto and Maekawa 1988, Okano *et al* 1994 and Mason *et al* 2005].

As mentioned in Chapter 2, Barron and Marshall [1988] reported that different frequency components of the delayed secondary signal produced different auditory width attributes. For example, the low frequencies contributed to the perception of 'envelopment' while the middle frequencies broadened the 'source width'. However, in their paper the term envelopment was said to be related to the source. This might suggest that the different frequency components of the crosstalk signal could give rise to a multidimensional perception of various kinds of 'source-related' attributes, requiring a further subjective investigation. In fact, the term source width has been used for describing a single dimensional concept by most authors. However, there has been a lack of standard descriptions for this attribute and therefore it is possible that the generic term 'source width' has been used for describing different perceptual concepts. For this reason, it might be necessary to conduct a systematic elicitation experiment for creating standard descriptions for perceived source-related width attributes depending on the frequency components of secondary signal. This work will require a group of trained and critical listeners that are able to distinguish subtle spatial differences.

Throughout the current research, locatedness has been recognised as one of the most salient attributes arising from interchannel crosstalk. However, the psychoacoustic and physical mechanism of locatedness perception has not been clearly known to date. Since locatedness is likely to be an important criterion that determines perceived

sound quality in sound recording, this issue needs further attention. In Section 5.4.4 a novel hypothesis was proposed on the mechanism of locatedness perception. It was hypothesised that locatedness for complex and continuous musical sources would not only depend on the simple precedence effect but would also be affected by the rate of ITD fluctuations during the ongoing part of sound. In any validation experiment for this hypothesis, the use of musical sound sources might be unsuitable since it would be difficult to control the fluctuation rate correctly due to their complex temporal and spectral characteristics. Therefore, it might be more appropriate to employ such controlled stimuli as amplitude-modulated sinusoidal or frequency-modulated noise signals, similarly to Blauert [1972]'s or Grantham and Wightman [1978]'s approaches respectively.

Appendix A LOCALISATION OF NATURAL SOUND SOURCES IN 2-0 STEREOPHONIC SOUND REPRODUCTION

This appendix describes a subjective experiment carried out to investigate the localisation characteristics of natural sound sources in 2-0 stereophonic reproduction and to develop a novel interchannel time and intensity trade-off function that can be used for the design of stereophonic microphone techniques. Since 1940, a number of stereophonic localisation experiments have been conducted to investigate the independent influence of interchannel time difference (ICTD) or interchannel intensity difference (ICID) on the position of a phantom image perceived between two loudspeakers. The data obtained from these kinds of experiments could become the basis for the design of stereophonic microphone techniques since the localisation of phantom images and the relevant stereophonic recording angle (SRA) rely on the interchannel relationship between the recorded signals. However, the results of those experiments are divergent depending on the type of sound source used (e.g. noise [Mertens 1965], wide-band speech [de Boer 1940, Leakey 1959, Wittek 2000], speech and maracas [Simonsen 1984]). Arguably, the localisation data that have been most widely quoted for microphone technique design are Simonsen [1984]'s, which were obtained using speech and maracas as sources. Williams [1987] used Simonsen's data for developing a ICTD-ICID trade-off relationship for the phantom image locations of 10°, 20° and 30° and this relationship was used for the analysis of SRAs for existing stereophonic microphone techniques and the design of his own multichannel microphone technique. However, Simonsen's data differ largely in

ICID values from the data obtained by Wittek [2000] using a speech source. This seems to suggest that data obtained in a specific experimental condition might not be directly applied to localisation in a different condition. Even though sound recordings made with microphone techniques deal with musical sources in most cases, to date experimental data related to the localisation characteristics of musical sources have not been presented apart from those of Simonsen's using maracas. This seems to be due to the complex nature of musical sources making it difficult to control experimental variables. However, it seems more valid to use the data obtained with musical sources for the design of microphone techniques since they are most likely to be encountered in practical situations. From this background, it was decided to conduct the current localisation experiment using various musical sound sources as well as speech, which have different temporal and spectral characteristics.

The current experiment investigated the independent influences of ICTD and ICID on the localisation of phantom images at the locations of 10°, 20° and 30°. Then, significances of the differences between the results obtained with different sound sources were statistically analysed. Finally, it was attempted to develop a trade-off function of ICTD and ICID.

A.1 Experimental Design

A.1.1 Test method

In most localisation tests, the listener is presented with sound stimuli created with various interchannel differences in regular intervals and asked to judge the locations of the perceived phantom images. This type of method is useful if it is desired to obtain a continuous localisation curve and error bars for perceived angles. However, it was not deemed appropriate for the current listening test because the purpose of this experiment was to obtain useful values of ICTD and ICID required for specific phantom source locations of 10°, 20° and 30°, rather than perceived angles for certain interchannel values. Therefore, this test was designed so that the listener adjusted ICTD or ICID using a slider provided in a control interface to match the positions of the phantom images to those of the markers indicated at +10°, +20° and +30° between the loudspeakers. Time delay or intensity attenuation was applied only to the left channel so that the phantom image appeared only in the centre-right region. In this way it was expected to obtain more accurate values of ICTD and ICID that worked specifically for the desired angles. The position of the 30° marker was the centre axis of the loudspeaker. The listeners were allowed to listen to the stimuli repeatedly until they were completely sure about their decisions. The listeners were asked to face the front consistently while listening to the sounds.

The control interface was developed using Cycling 74's 'MSP' software shown in **Figure A.1**. The range of ICTD that could be applied on the left channel was from 0

to 5ms with the interval of 0.1ms. However, the scale shown in the slider was presented with the representative numbers of 0 to 50 in order to prevent the listener from being biased by their experience and knowledge about the influence of ICTD. The range of ICID scale was from 0 to –100, where 0 represents zero difference and –100 represents -∞dB, and the resulting values were later transformed into the corresponding decibel values. When the listener adjusted ICTD, ICID was maintained at 0, and vice versa. The order of the angles to be judged was randomised for each stimulus in order to avoid a psychological order effect.

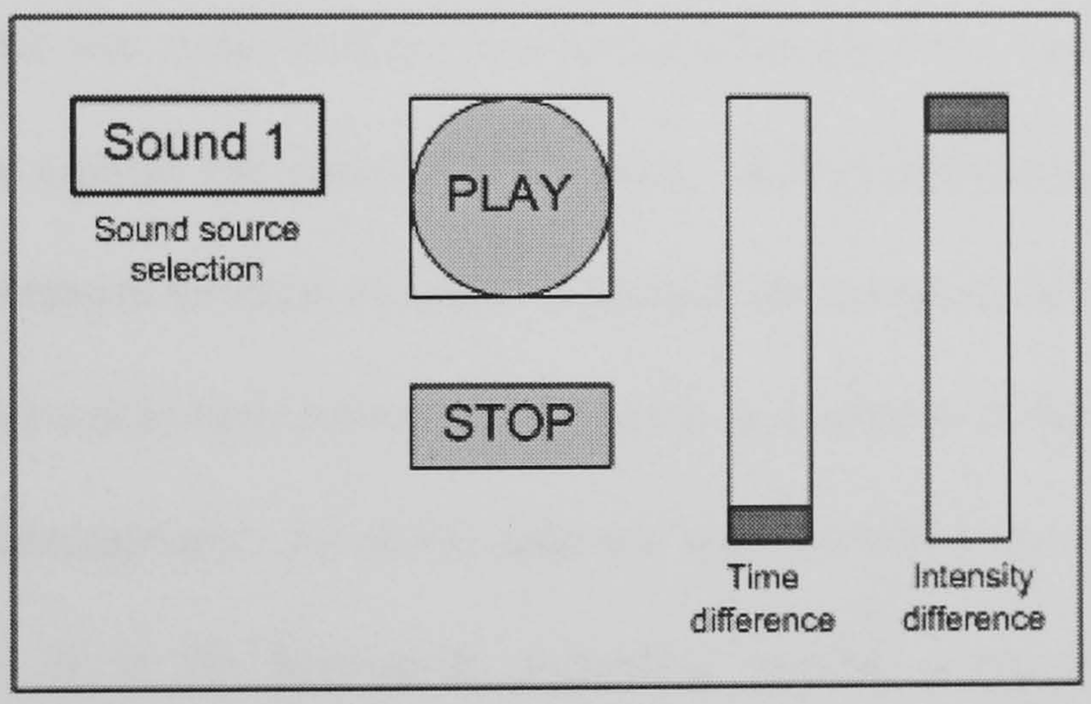


Figure A.1 Control interface for the localisation test developed using Cycling 74's MSP software

A.1.2 Sound stimuli

Five sound stimuli were chosen for this experiment, comprising:

- Piano 'staccato' note of C3 (fo = 130Hz)
- Piano 'staccato' note of C6 (fo = 1046Hz)
- Trumpet 'sustain' note of Bflat3 (fo = 228Hz)

- Trumpet 'sustain' note of Bflat5 (fo = 922Hz)
- Continuous speech

The piano and trumpet were chosen in order to examine the effect of temporal characteristics of different musical instruments (i.e. transient vs. continuous). For each musical source, low and high notes were chosen and this was for investigating the effect of spectral characteristics. The speech source was included for its broadband frequency spectrum as well as complex temporal characteristics. Also, since a number of earlier localisation tests used speech sources, the use of a speech source in this test was considered to be a useful reference for a comparison between the results of the current test and the earlier tests. It was decided to use single notes instead of performance extracts in order to control the variables strictly. Ideally all the sound sources would have been recorded under an anechoic condition, but this was unavailable. Alternatively, the piano sources were recorded in a small recording booth of studio B at the Metropolis recording studios, using a single cardioid microphone (Schoeps CMC 5-U) placed about 30cm over the hammers for the desired notes. The piano was completely covered with thick cloth in order to reduce unwanted acoustic effects as much as possible. The trumpet sources were recorded in a small overdub booth of Studio 3 of the University of Surrey, using a single cardioid microphone (AKG 414 B-ULS) placed about 1m away from the instrument. The recording space was acoustically isolated and had no audible reverberation. In order to investigate the continuous nature of the trumpet strictly, the onset and offset transients of the trumpet sources were removed by fading in and out the beginning and ending for one second each, and the total duration of the stimulus was four seconds.

The speech signal was chosen because it is a mixture of both transient and continuous natures with the wide range of frequencies. The speech recording used was Danish male speech that was anechoically recorded for the Bang and Olufsen's Archimedes project. An English speech recording was also available in the CD, but it was decided to use a foreign language rather than English in order to prevent the listener from paying attention to the language itself.

A.1.3 Physical setup

The listening test was conducted in the ITU-R BS.1116 listening room at the University of Surrey. Two Genelec 1032A loudspeakers were arranged in the standard configuration, with a distance of 2.4m between them.

A.1.4 Test subjects

A total of five listeners took part in the test. All were critical and experienced listeners, including research staff and doctoral students at the Institute of Sound Recording of the University of Surrey. Because of the nature of the test requiring highly critical listening skill, it was decided to employ a relatively small number of experienced listeners rather than a large number of inexperienced listeners, and repeat the test three times for each listener in order to ensure a sufficient amount of data for analysis.

A.2 Results and Discussions

A.2.1 Basic localisation characteristics

Figure A.2 shows the results of the localisation test using pure ICTD cue. The plots represent the median values and associated 25th and 75th percentile bars for the subjective data obtained. Firstly, all the subjects found that it was almost impossible to localise the high note trumpet. The low note trumpet, on the other hand, was reasonably localisable but the subjects still found it difficult to localise easily because the positions of phantom images randomly changed even with a very small head movement. For the transient piano sources, both low and high notes are relatively well localised. The localisation difficulty for the continuous trumpet notes with pure ICTD seems to confirm the literature reporting the importance of transient component in localisation relying on the time difference between two sounds [Rakerd and Hartmann 1985, 1986, Wallach *et al* 1949, Zurek 1980]. This result might also be explained by Rakerd and Hartmann [1986]'s 'plausibility hypothesis', suggesting that the ongoing cue of a pure tone is unreliable (or implausible) for localisation.

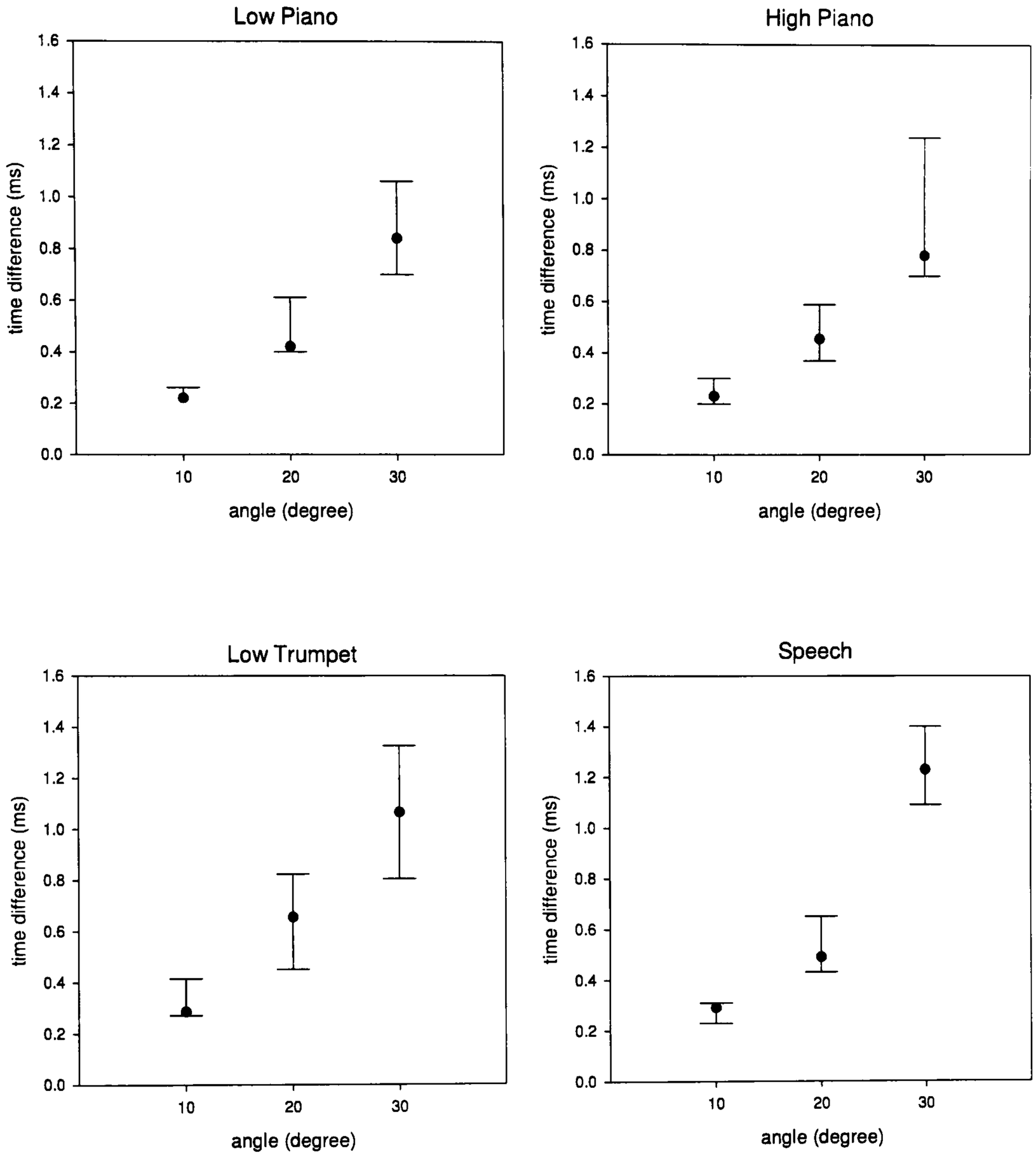


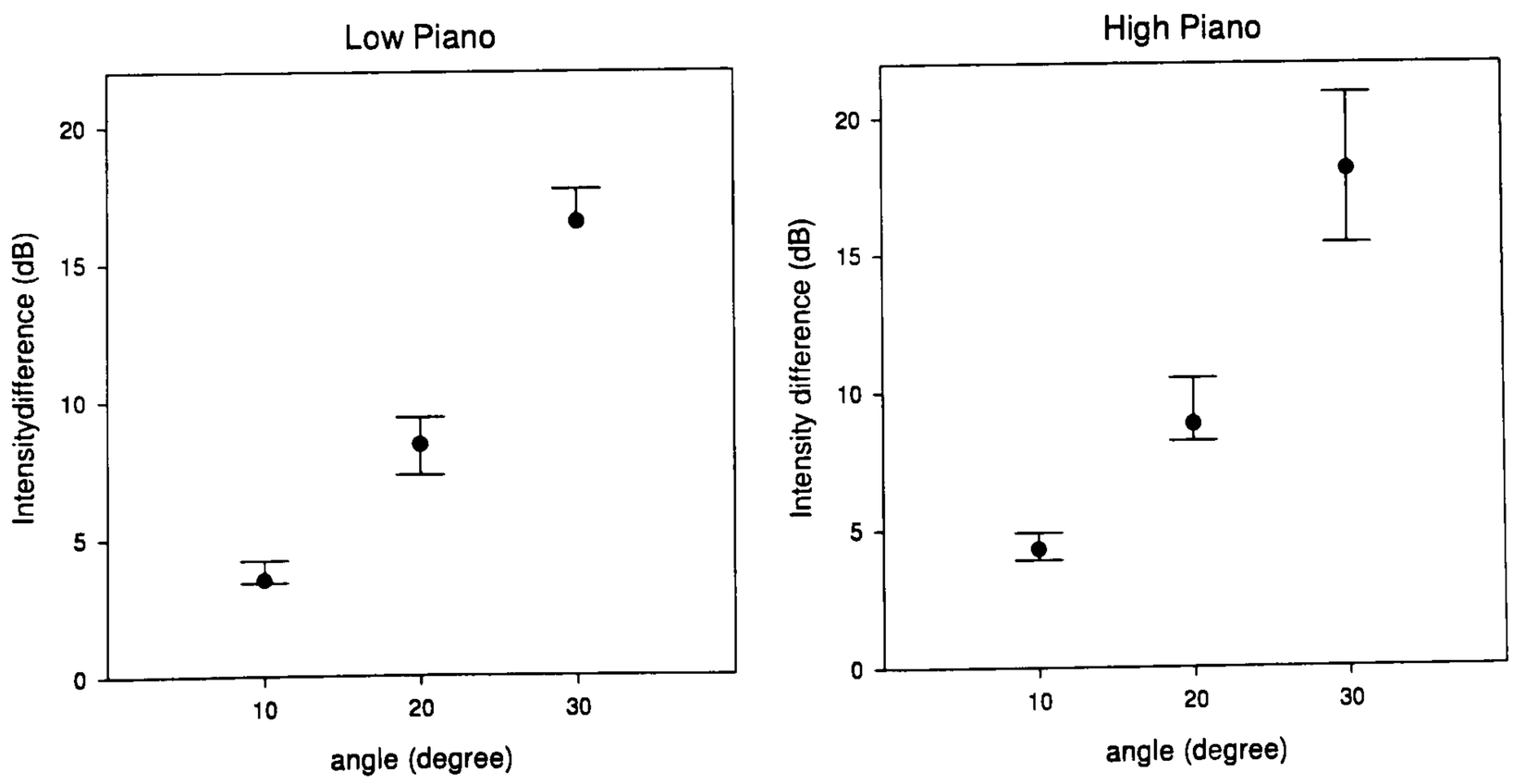
Figure A.2 Localisation by pure time difference: Median values and associated 25th to 75th percentile

It appears that the low piano note was localised slightly more certainly than the high piano note. Bank and Green [1973] found that for transient noise signals, low frequency components below about 2000Hz were essential for accurate localisation in stereophonic reproduction. Based on Yost *et al* [1971], this is because low frequency transients excite more space in the cochlear partition than high frequency ones and excite more fibres, thus producing more substantial positional displacement. The

high note piano used in this experiment has a complex tonal nature containing lower harmonics. However, the low note piano has richer low frequency components by its nature and this would have led to a better localisation certainty.

The speech source appears to have the best localisation certainty in general. This seems to be due to the fact that the continuous speech source has consecutive transients at every syllable change as well as wide frequency range with the fundamental frequency of about 100Hz.

Figure A.3 shows the results of the localisation test using a pure ICID cue. The effect of transient characteristics appears to be less dominant in the case of ICID in that the continuous trumpets of both low and high notes were localised reasonably well. This seems to suggest that the continuous nature of a sound is plausible when ICID cue is used for localisation. In general, however, the results of the ICID localisation have a similar tendency to the results of the ICTD localisation. That is, the speech and piano sources were more certainly localised than the trumpet sources.



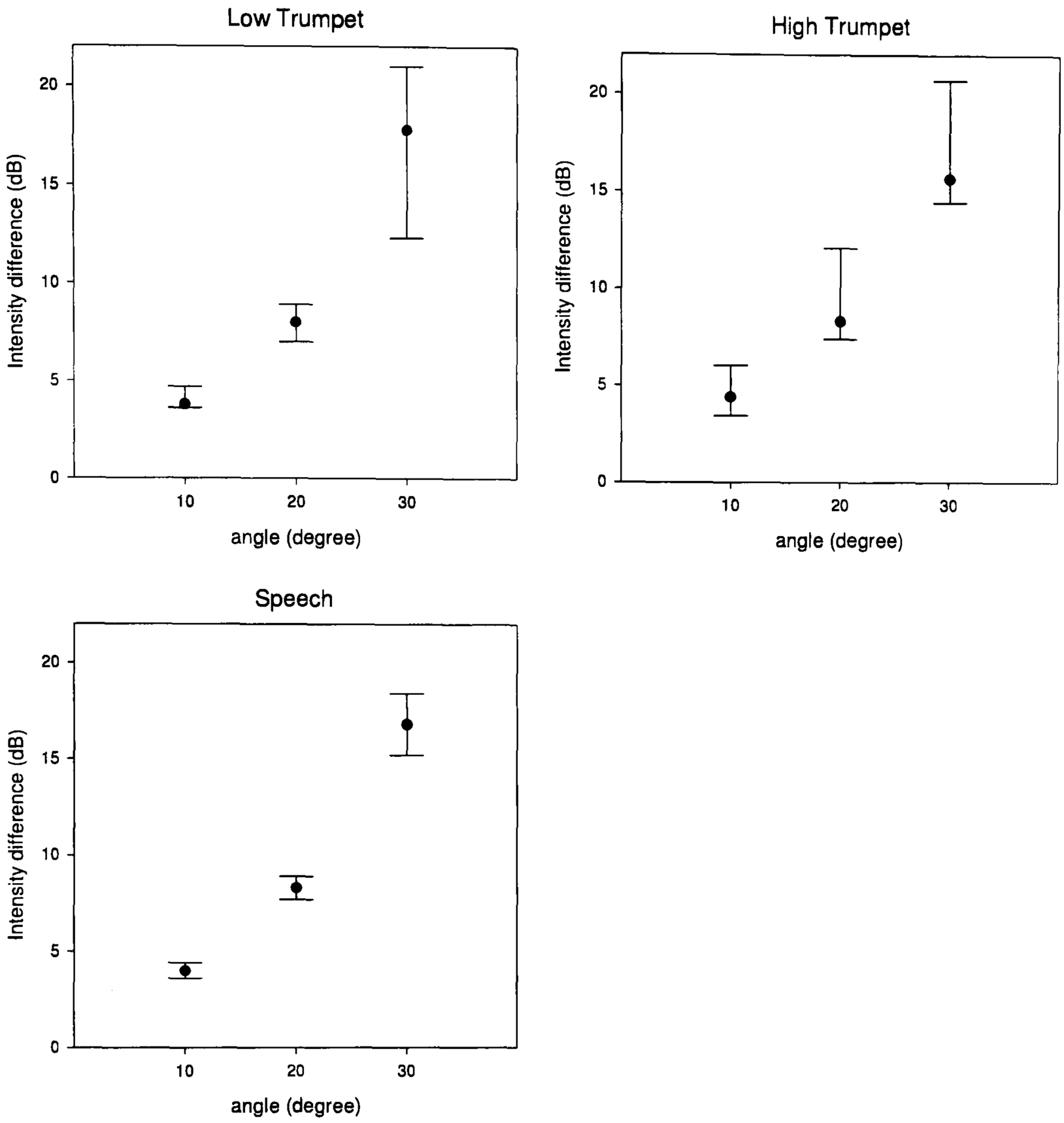


Figure A.3 Localisation by pure intensity difference: Median values and associated 25th to 75th percentile

From the above results, it can be generally seen that the localisation using pure ICID was more stable than that using pure ICTD, which supports the literature. It is interesting to observe that for both ICTD and ICID pannings the size of the error bar becomes greater as the localisation angle moves from 10° to 30°. This seems to be related to the findings of the minimum audible angle (MAA) of Mills [1958]. Mills carried out a subjective experiment to measure the smallest angular change of sound source that the listener could just detect, which is the so-called 'minimum audible

angle (MAA)', using pure tones and it was found that the MAA became larger as the loudspeaker pair moved away to the side of the listener. In addition, it can also be observed from the results that the shift factors of ICTD or ICID required for the phantom source positions of 10° and 20° have almost constant relationships. However, the shift factor for 30° appears to be much greater that those for 10° and 20°. This phenomenon can also be observed in the classic localisation curves, being almost linear up to 75% of the shift region and becoming exponential as the angle increases further up to 100% [Wittek and Theile 2002]. The results of both Mills and the author seem to suggest that in stereophonic reproduction the listener's sensitivity for localising a phantom source decreases as the direction of the source moves from the front to the side.

A.2.2 Statistical analysis

In order to examine the significance of the differences observed between the sound sources, the 'Friedman' test, which is a non-parametric statistical test, was carried out. The 'ANOVA' test, which is a parametric test, was not appropriate for this experiment since the panning angle scale (10°, 20° and 30°) had an ordinal nature and the homogeneity of variance required for the ANOVA test was not met in this case. The results of this test shown in **Table A.1** indicate that the differences between sound sources were not significant for both ICTD and ICID localisations ($p > 0.05$).

	ICTD localisation	**ICID localisation**
N	36	45
Chi-Square	3.765	5.721
Df	3	4
Asymp. Sig.	0.288	0.221

Table A.1 Effects of sound stimuli in time and intensity pannings, analysed using the Friedman test.

Since there is no significant difference between sound sources, it is possible to combine the data for all sound sources. **Table A.2** shows the overall median values and associated 25^{th} to 75^{th} percentiles, and these data are plotted in **Figure A.4** and **A.5**. It can be noted again in the unified plots that the localisation certainty tends to become worse as the angle increases. The increase of median value is almost constant up to 20° but becomes steep from 20° to 30°.

Panning method	**Angle**	**25^{th} percentile**	**median**	**75^{th} percentile**
Intensity (dB)	10	3.5	**4.0**	4.4
	20	7.6	**8.4**	9.25
	30	15.4	**17.1**	19.6
Time (ms)	10	0.22	**0.27**	0.32
	20	0.41	**0.50**	0.72
	30	0.75	**1.1**	1.36

Table A.2 Overall median values and 25^{th} to 75^{th} percentiles

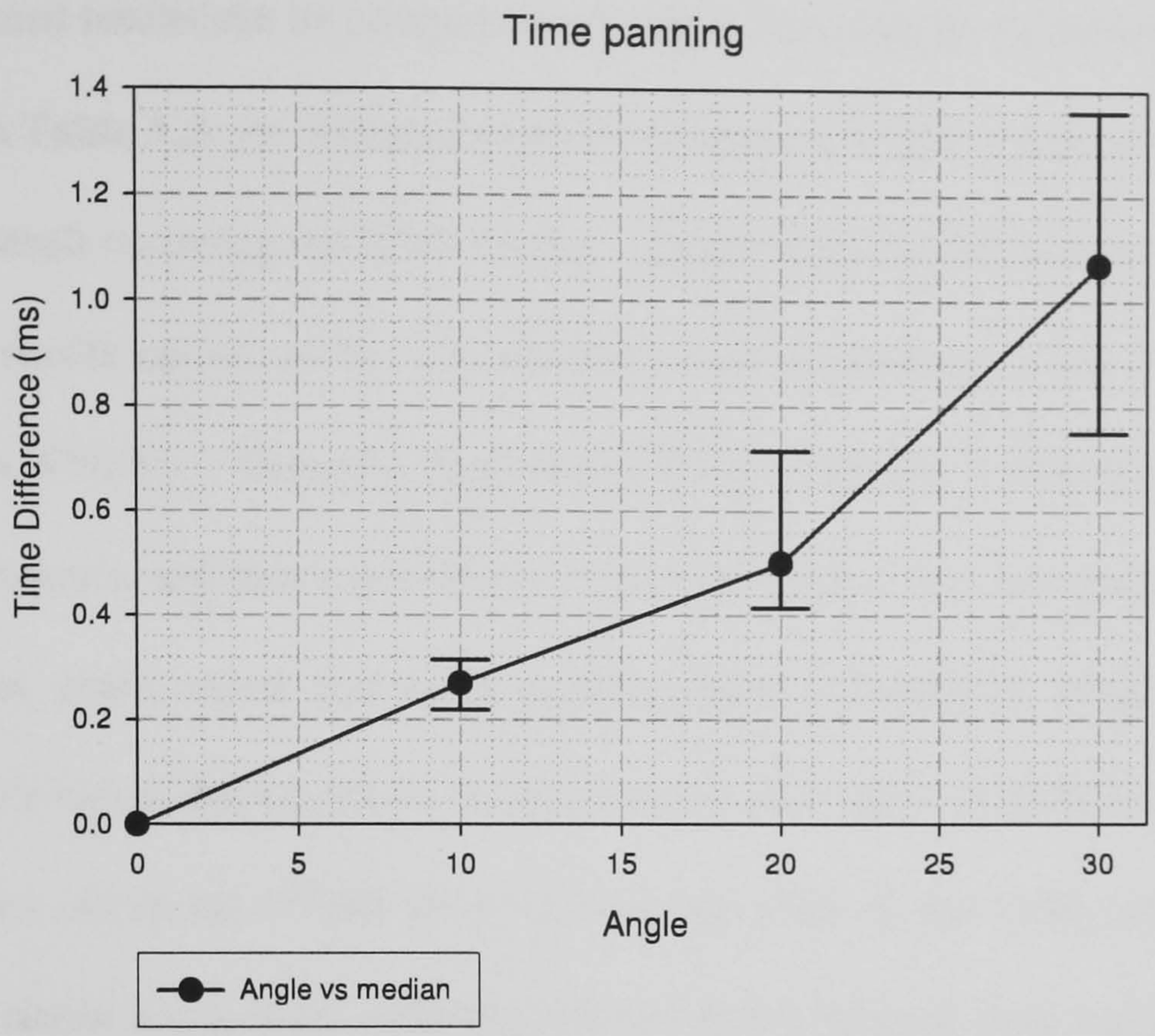


Figure A.4 Plots of overall median values and 25th to 75th percentiles for the ICTD localisation

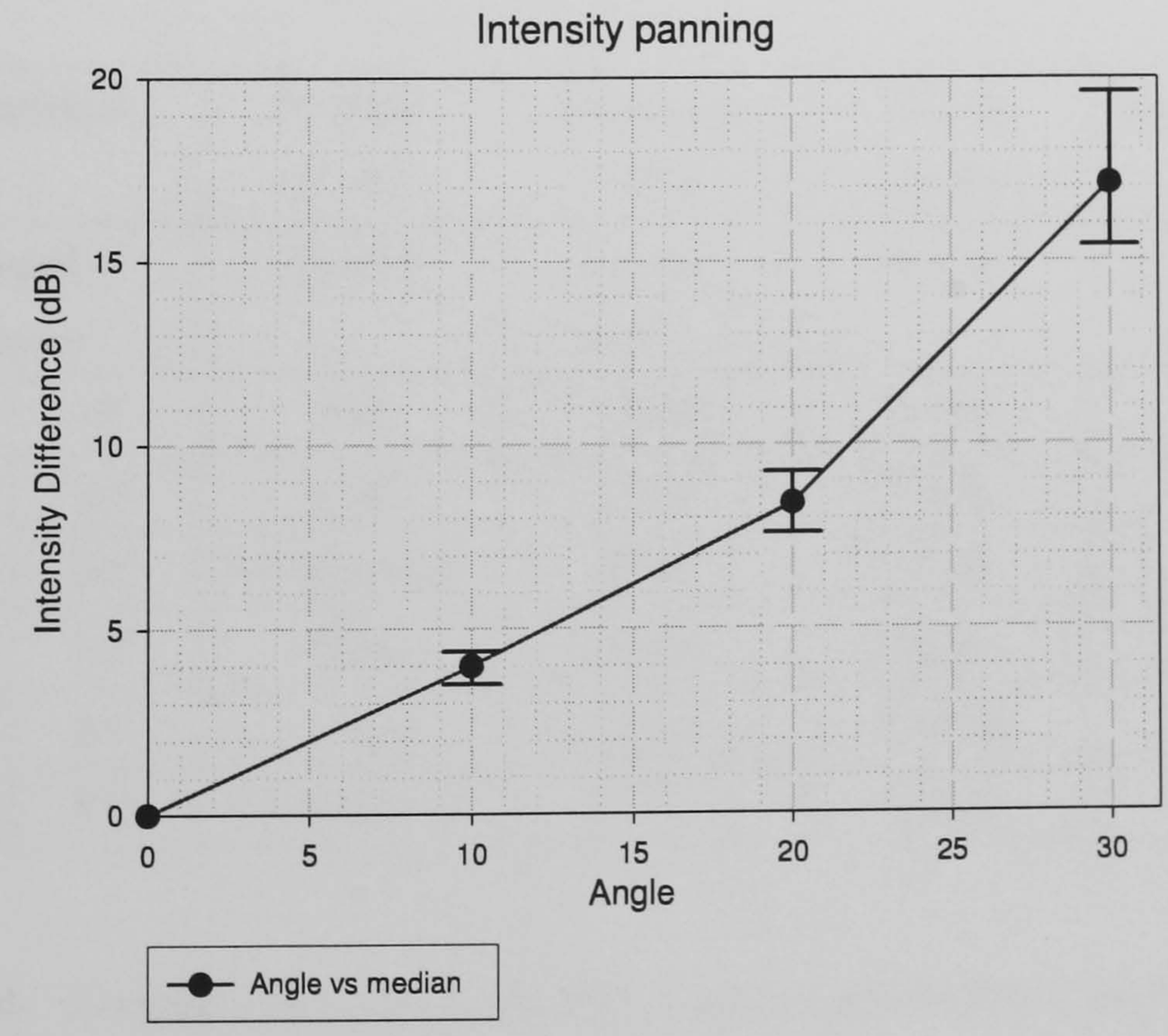


Figure A.5 Plots of overall median values and 25th to 75th percentiles for the ICID localisation

The obtained results can be compared with others from similar experiments. As can be seen in **Table A.3**, the differences among Simonsen, Wittek and the author's results are very small regarding the ICTD values. Regarding the ICID values, however, the author's results appear to be very different from Simonsen's while they are very similar to Wittek's. Generally Simonsen's ICID values are 2-3dB less than Wittek and the author's, and this is considered to be significant in that this range of intensity differences could cause noticeable angular shifts of phantom images. In fact, Simonsen's values did not satisfy the supposed angular shifts in the informal listening test carried out in the ITU-R BS.1116 listening room at the University of Surrey. The differences between the different authors' results seem to have resulted from the different experimental conditions, such as the acoustic condition of the listening room, the type of sound source used and the number of subjects.

Researcher		**De Boer** [1940]	**Simonsen** [1984]	**Wittek** [2000]	**Lee (author)** [2004]
Sound source		Speech	Speech / maracas	speech	Speech / various
ICID	10°	5dB	2.5dB	4.4dB	4.0dB
	20°	11dB	5.5dB	8.8dB	8.4dB
	30°	not indicated	15dB	18dB	17.1dB
ICTD	10°	0.7ms	0.20ms	0.23ms	0.27ms
	20°	1.7ms	0.44ms	0.45ms	0.50ms
	30°	not indicated	1.12ms	1.0ms	1.1ms

Table A.3 Comparisons of psychoacoustic values required for the localisation of 10°, 20° and 30° angles

A.3 Development of a Time-Intensity Trade-off Function

A.3.1 Method

Using the unified localisation data obtained from the current experiment, it was attempted to develop ICTD-ICID trade-off functions for the phantom image locations of 10°, 20° and 30°. The basic combination method used was based on Theile's hypothesis, which suggests that the degree of total angular shift of phantom image can be calculated simply by summing the angular shifts by individual time and intensity differences, provided the individual shift is linear. The simple equation for this hypothesis is shown below.

$$\Psi(\Delta t, \Delta I) = \Psi(\Delta t) + \Psi(\Delta I)$$

The unified data plots in **Figure A.4** and **A.5** show that the psychoacoustic values required for 10° and 20° shifts are almost linearly increased in both time and intensity pannings. In other words, the increasing factors of the 0° - 10° and 10° - 20° shift regions are almost constant and therefore it was possible to apply the above combination function in this case. In an informal listening test conducted by the author and two colleagues who are critical listeners, this combination function was found to be valid for the interchannel data of up to the 20° shift region. However, this function could not be directly applied for the 30° shift because the shift factor of the 20° - 30° region is much greater than those of the lower regions. For example, a simple combination of individual shifts by time and intensity such as

$\Delta t(10°) + \Delta I(20°)$ will not complete the desired 30° shift if $\Delta I(20°)$ is based on the 0°-20° region. Even if it is based on the region of 10°-30°, there will be two different shift factors to be considered. Therefore, for the 30° shift, it was decided to divide the whole shift region into three effective regions and consider each separately as shown below.

$$\vartheta(30°) = \vartheta\,(0° - 10°) + \vartheta(10° - 20°) + \vartheta(20° - 30°)$$

Shift factors of ICTD and ICID required for each phantom image shift region were obtained by simplifying the results of the localisation experiments shown in **Figures A.4** and **A.5** within the error ranges of 25th to 75th percentiles in such a way that the shift regions up to 20° have constant shift factors, as shown in **Table A.4**.

Shift Region	ICTD	ICID
0° - 10°	0.25ms	4dB
10° - 20°	0.25ms	4dB
20° - 30°	0.60ms	9dB

Table A.4 Phantom image shift factors of ICTD and ICID for the shift region regions of 0° - 10°, 10° - 20° and 20° - 30°

A.3.2 Result

Using the proposed shift factors shown in **Table A.4**, various combinations of ICTD and ICID were calculated. **Figure A.6** shows the obtained combination curves for

each localisation angle. It can be seen that the curves for the 10° and 20° shifts are completely linear and the calculated curve for the 30° shift is almost linear. It is not clear how the small non-linearity in the middle region is caused, but the difference between the manipulated linear curve and the calculated curve does not seem to be significant. As a result, three linear ICTD-ICID trade-off functions were developed for 10°, 20° and 30° shifts. The proposed linear trade-off curves could be advantageous to Williams [1987]'s trade-off curves shown in **Figure A.7** in that it would be much easier to calculate the required ICTD and ICID for trade-off with the linear curves.

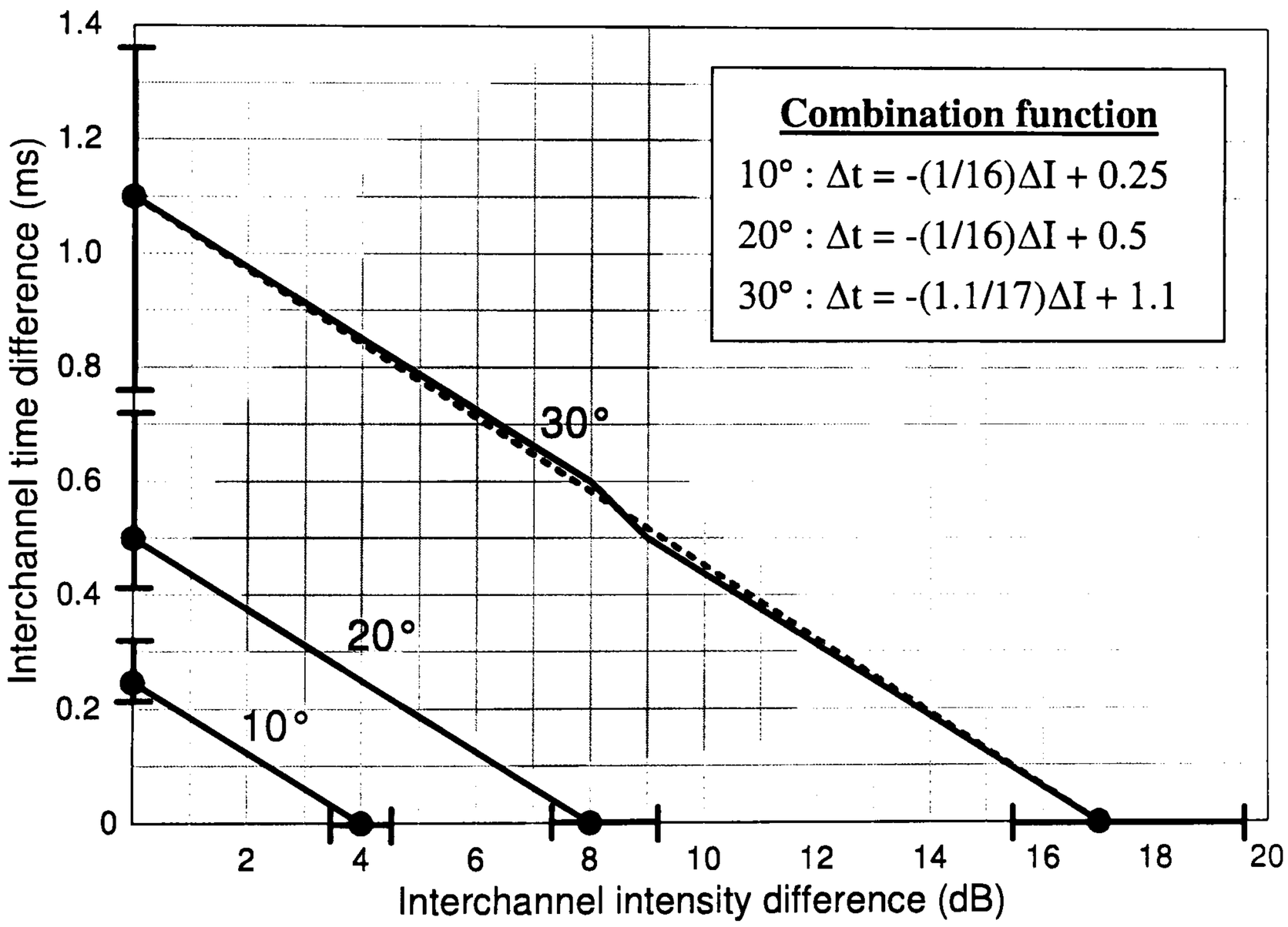


Figure A.6 Proposed ICTD and ICID trade-off curves for 10°, 20° and 30° images, based on the psychoacoustic values obtained from the localisation test (see **Table A.2**); Plots show the simplified median values and 25th to 75th percentiles.

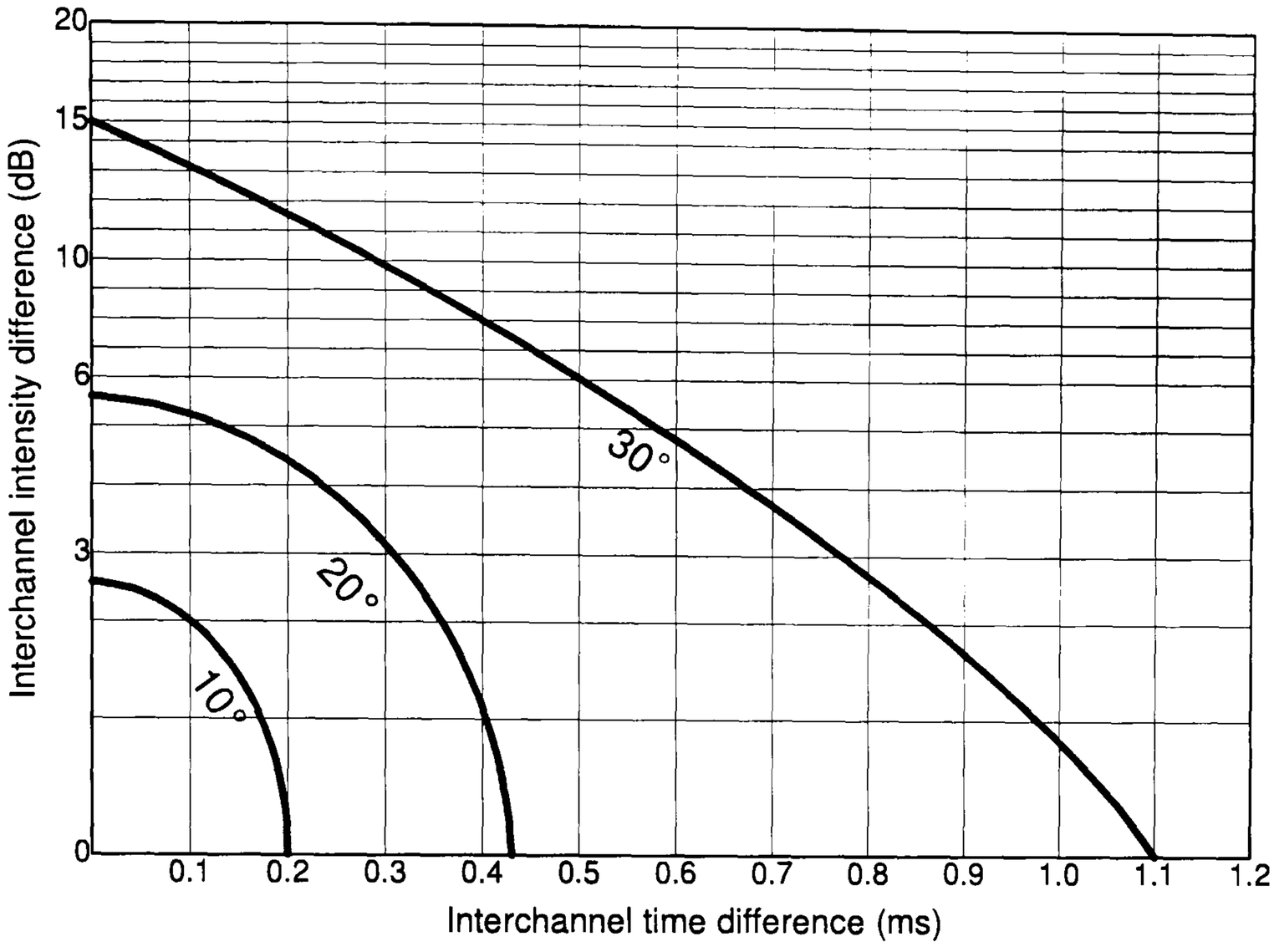


Figure A.7 Williams' ICTD and ICID trade-off curves for 10°, 20° and 30° images, based on the psychoacoustic values obtained by Simonsen [1984] [after Williams 1987]

A.3.3 Verification of the proposed combination function

In order to verify the feasibility of the proposed combination functions, an additional subjective listening test was carried out with the identical subjects using the speech source in the same listening condition. A total of 17 test stimuli were created with various combinations of ICTD and ICID based on the proposed trade-off function, as listed in **Table A.5**, and the subjects were asked to indicate the perceived locations of phantom images using reference markers placed with 5° intervals between the loudspeakers. Stimuli A to C were created aiming for 10° imaging, D to H for 20° and I to Q for 30°. The stimuli were recorded onto computer hard disk and played back to the subjects in a random order. Each stimulus was 30 seconds long, which

gave the subjects enough time for judgment. Two subjects repeated the test three times and three repeated twice. The result of the test is shown in **Figures A.8 to A.10**. As can be seen, the phantom images did not always appear at the desired locations and this suggests that the proposed combination function is not perfect. However, it appears that the deviation between the median angles and the desired angles is normally within the range of 2°-3°, and this is considered to be acceptable.

Stimuli	Combination	Stimuli	Combination
A	0dB+0.25ms	J	2dB+0.97ms
B	2dB+0.13ms	K	4dB+0.84ms
C	4dB+0ms	L	6dB+0.71ms
D	0dB+0.5ms	M	8dB+0.58ms
E	2dB+0.38ms	N	10dB+0.45ms
F	4dB+0.25ms	O	12dB+0.32ms
G	6dB+0.13ms	P	14dB+0.19ms
H	8dB+0ms	Q	17dB+0ms
I	0dB+1.1ms		

Table A.5 Sound stimuli of various time and intensity combinations, based on the linear combination functions : A – C for 10°, D – H for 20° and I – Q for 30°

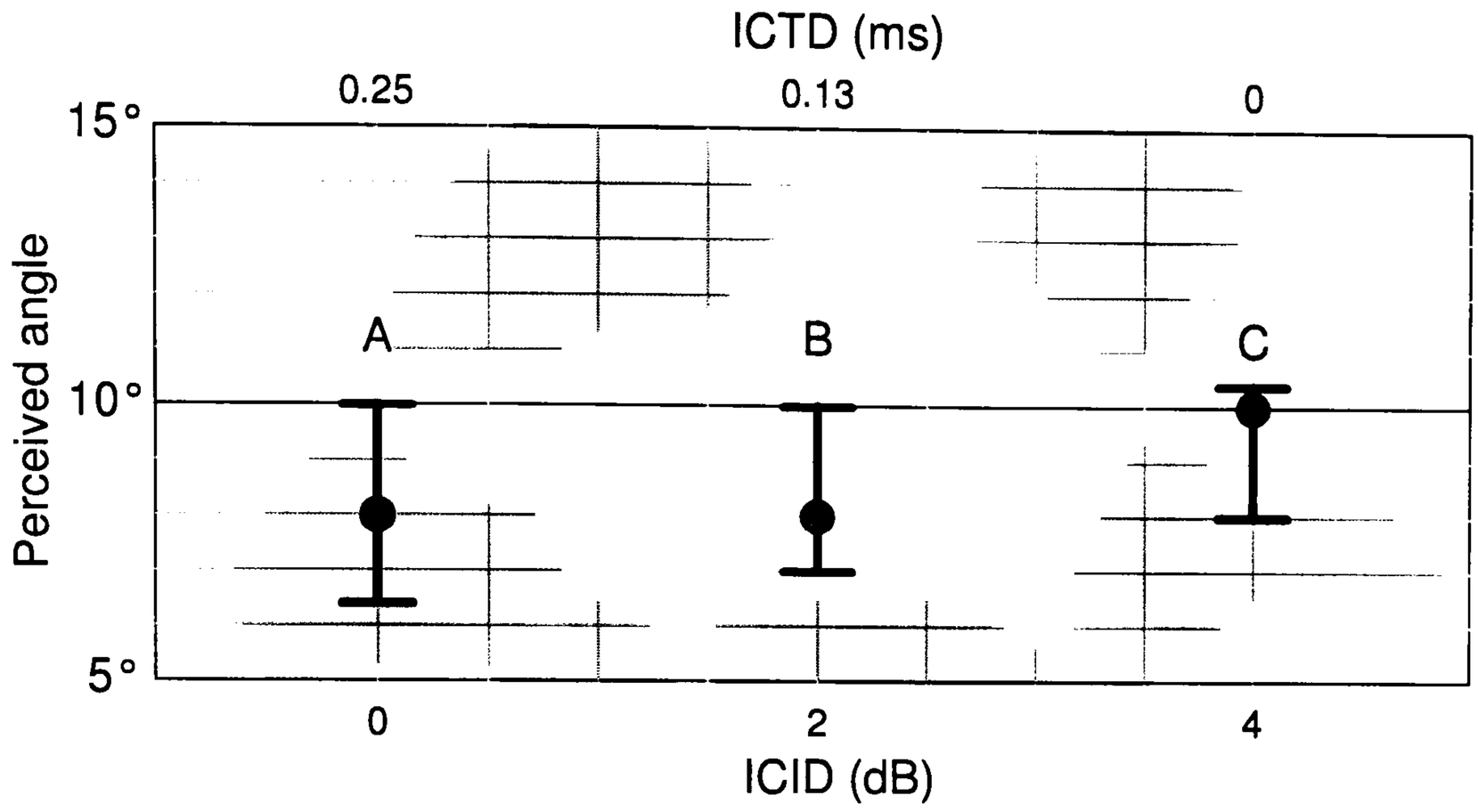


Figure A.8 Data plots of perceived phantom image angles for the stimuli A, B and C indicated in **Table A.5**: Median values and associated 25th to 75th percentiles.

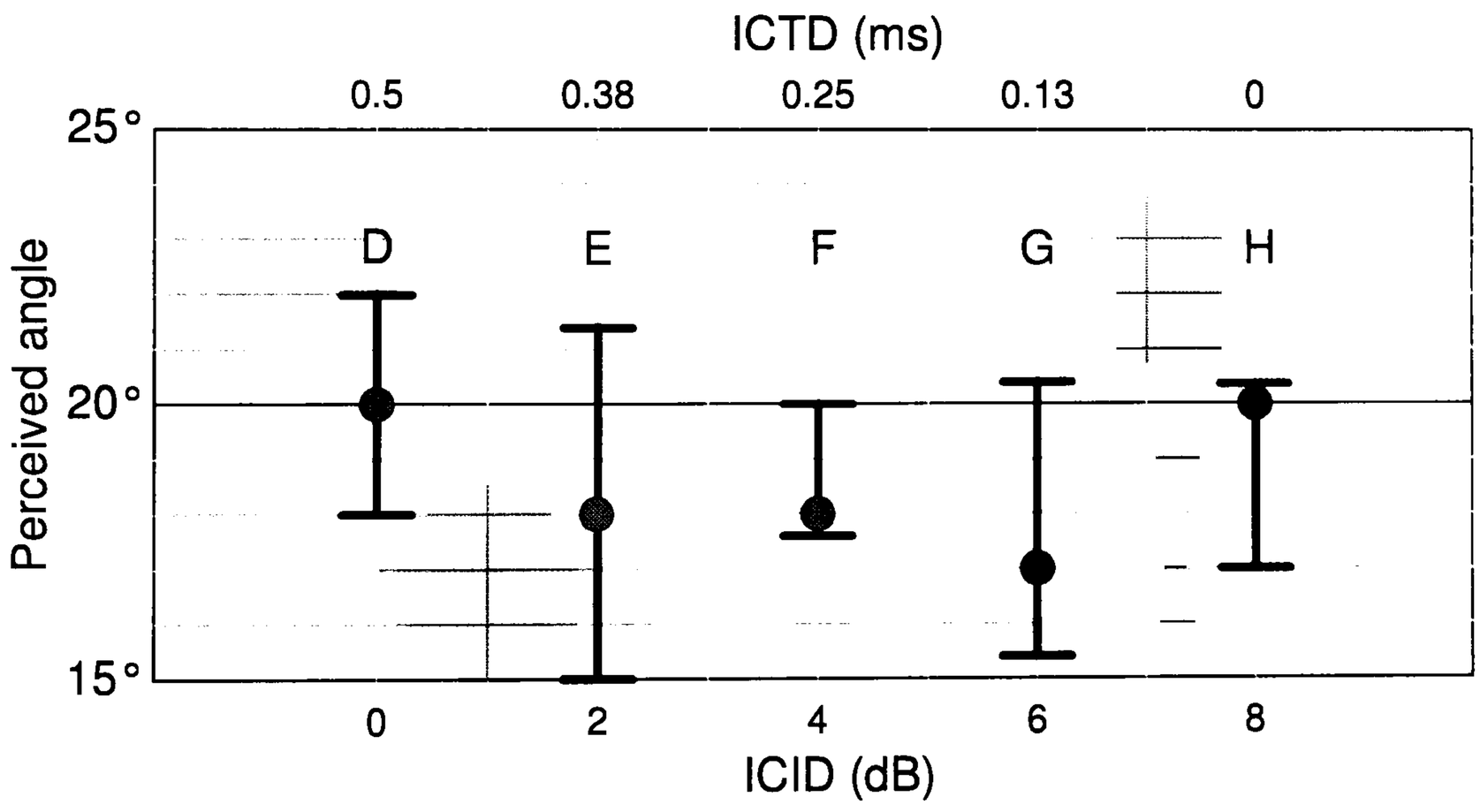


Figure A.9 Data plots of perceived phantom image angles for the stimuli D, E, F, G and H indicated in **Table A.5**: Median values and associated 25th to 75th percentiles.

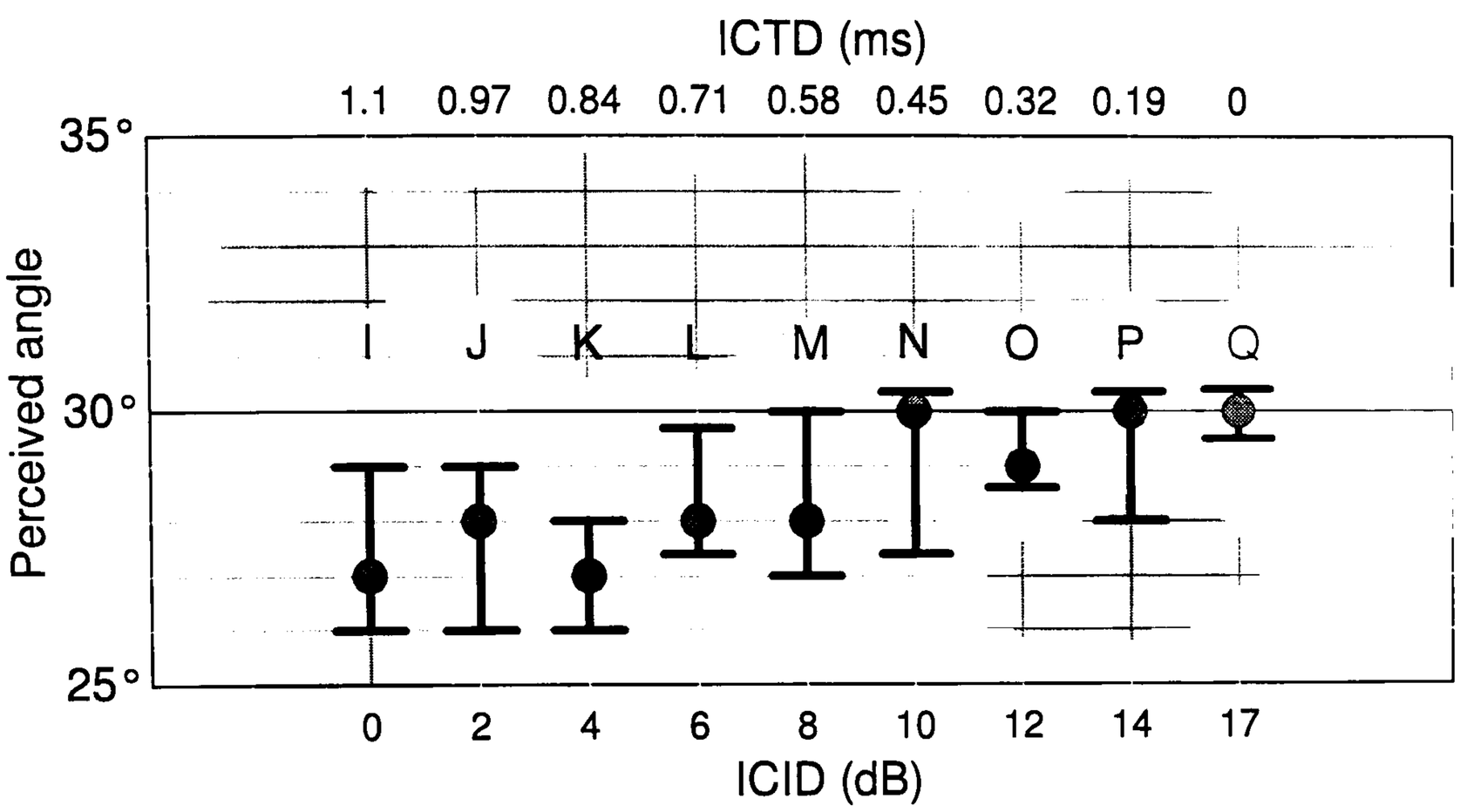
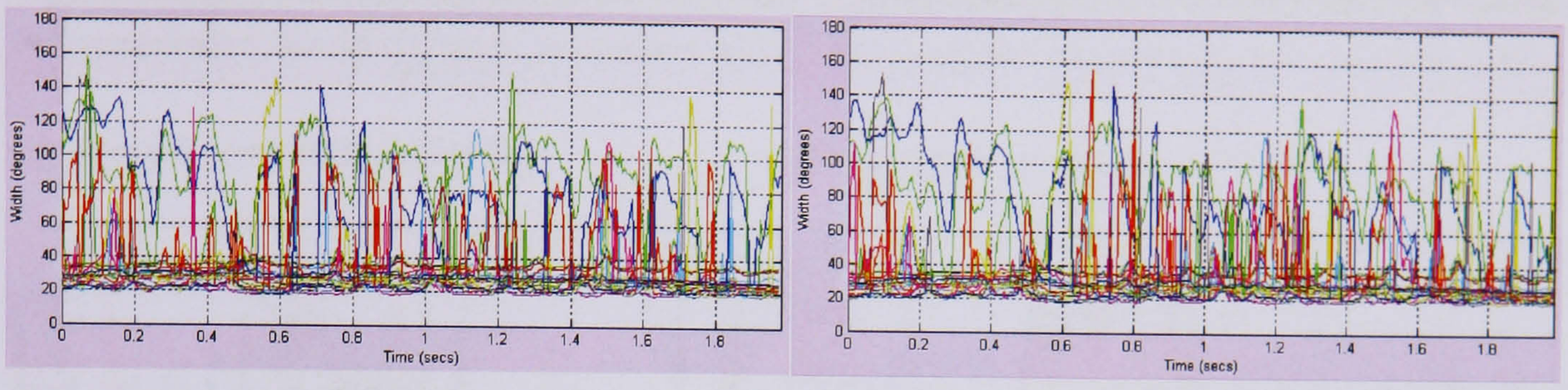


Figure A.10 Data plots of perceived phantom image angles for the stimuli I, J, K, L, M, N, O, P and Q indicated in **Table A.5**: Median values and associated 25th to 75th percentiles.

Appendix B PLOTS FROM OBJECTIVE MEASUREMENTS OF THE EFFECTS OF INTERCHANNEL CROSSTALK



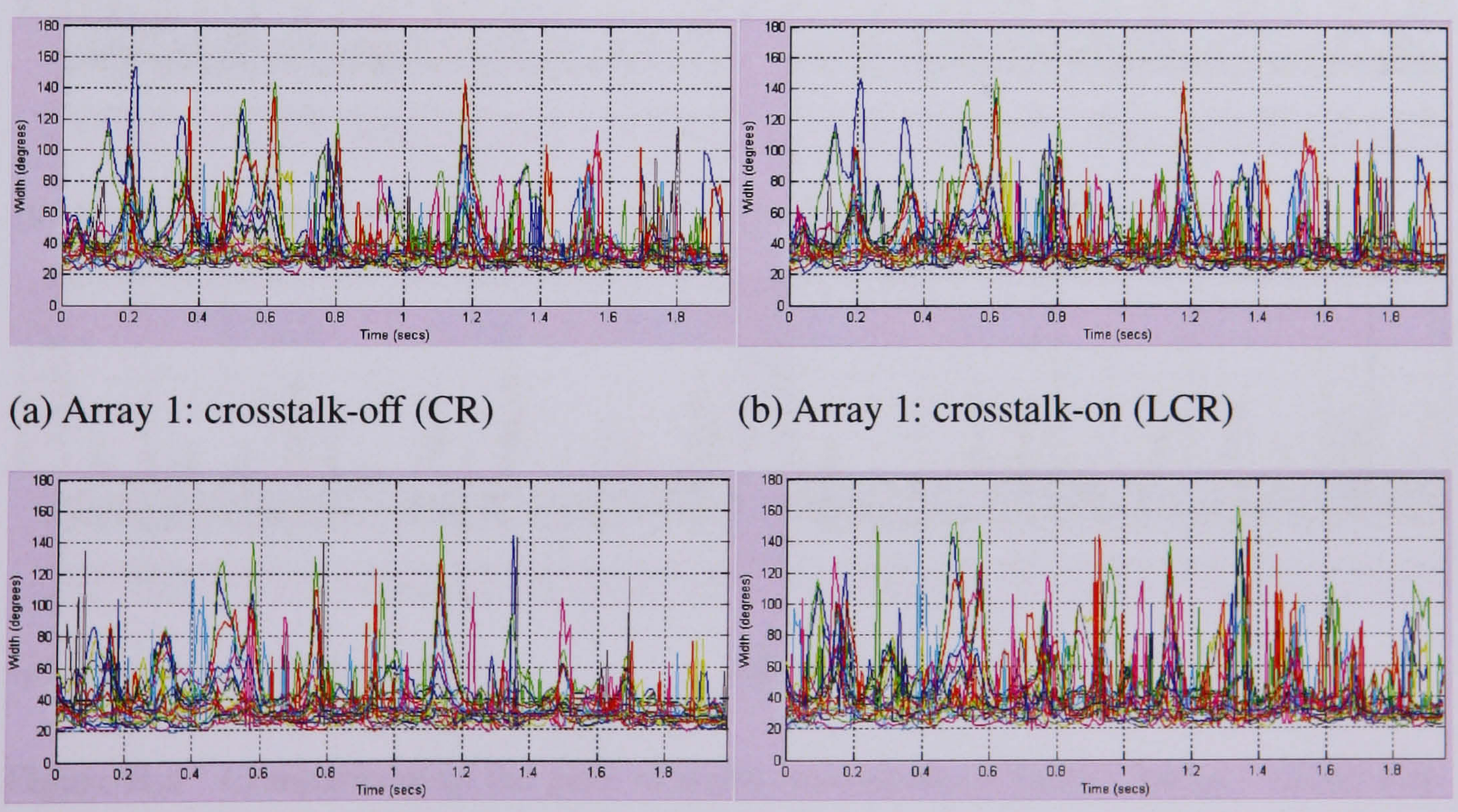
(a) Array 1: crosstalk-off (CR)

(b) Array 1: crosstalk-on (LCR)

(c) Array 4: crosstalk-off (CR)

(d) Array 4: crosstalk-on (LCR)

Figure B.1 Comparisons of the plots of width measurements for the 'cello' stimuli that were created in 'anechoic' condition



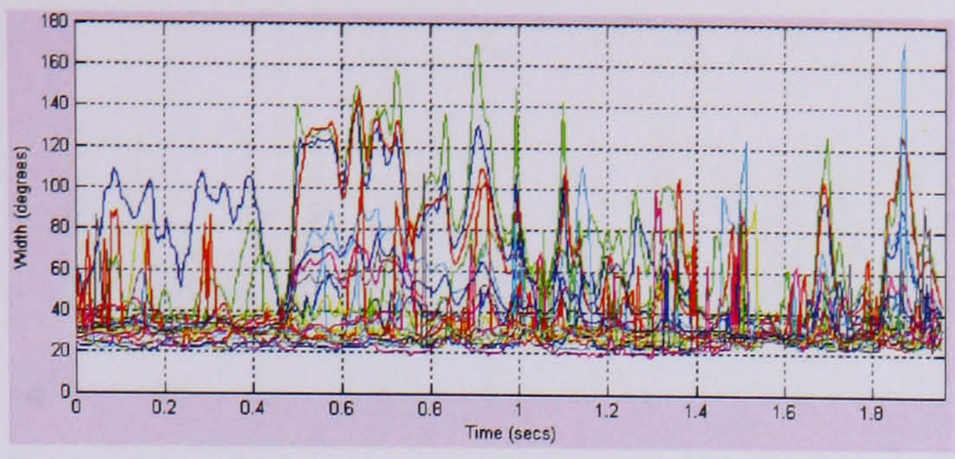
(a) Array 1: crosstalk-off (CR)

(b) Array 1: crosstalk-on (LCR)

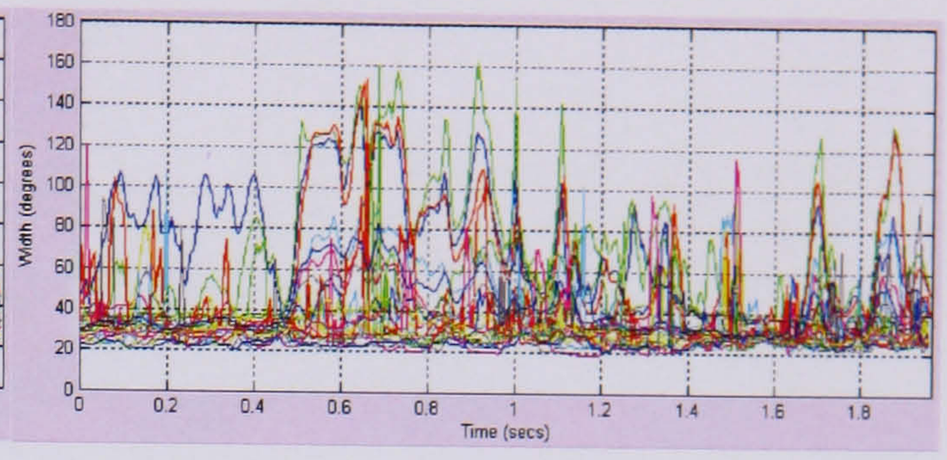
(c) Array 4: crosstalk-off (CR)

(d) Array 4: crosstalk-on (LCR)

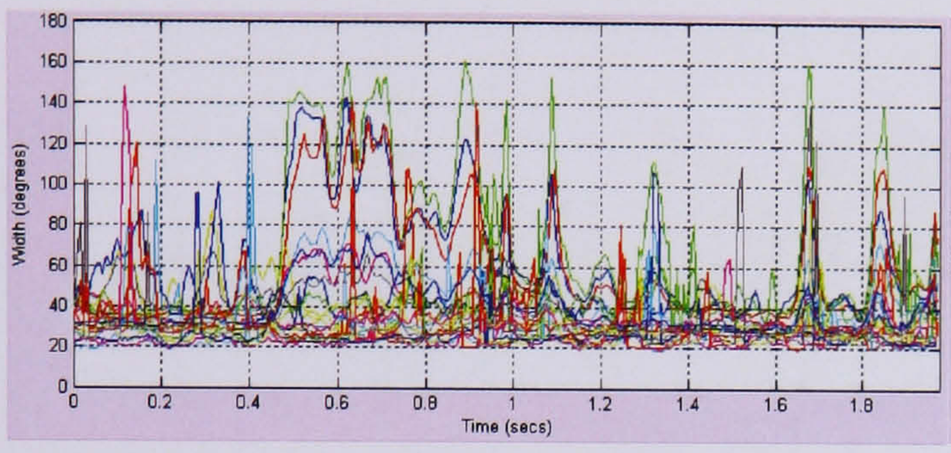
Figure B.2 Comparisons of the plots of width measurements for the 'cello' stimuli that were created in 'room' condition



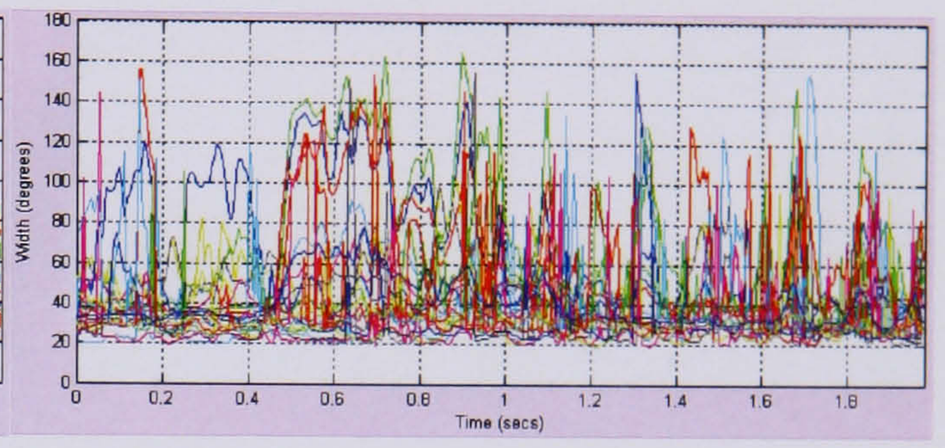
(a) Array 1: crosstalk-off (CR)



(b) Array 1: crosstalk-on (LCR)

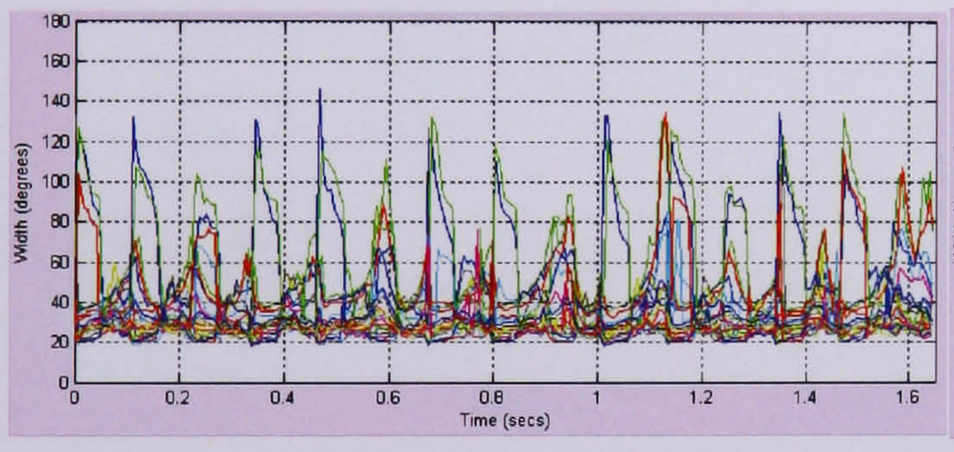


(c) Array 4: crosstalk-off (CR)

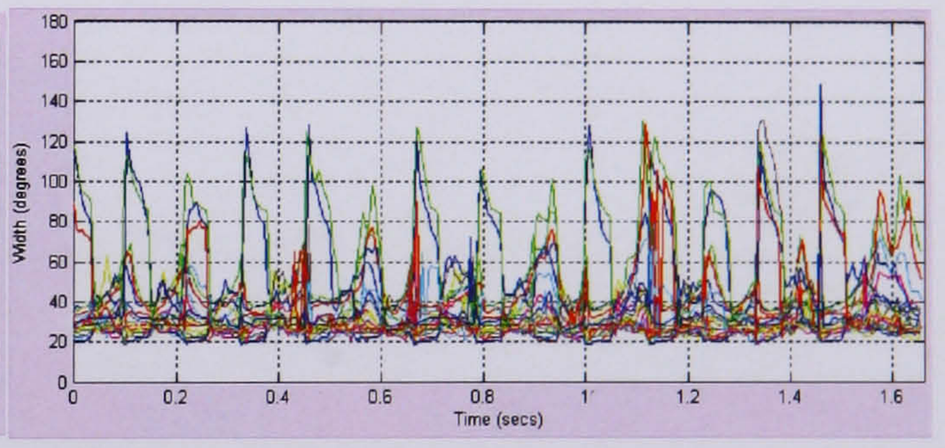


(d) Array 4: crosstalk-on (LCR)

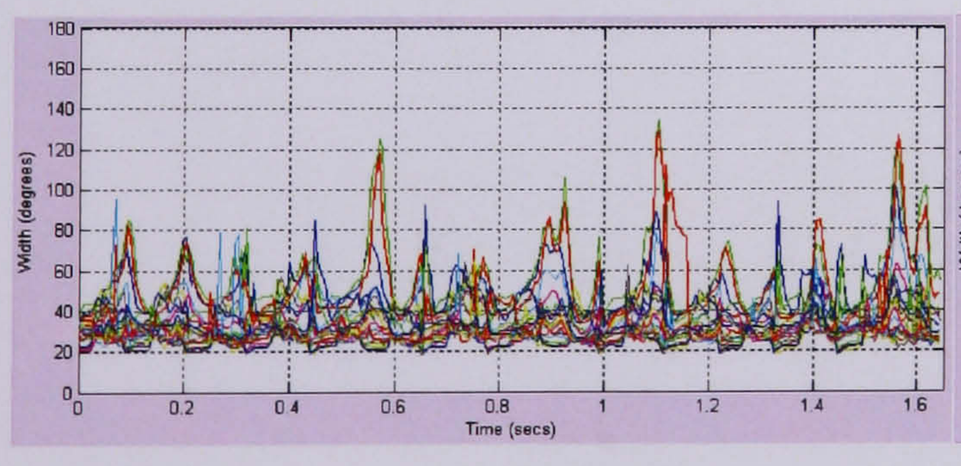
Figure B.3 Comparisons of the plots of width measurements for the 'cello' stimuli that were created in 'hall' condition



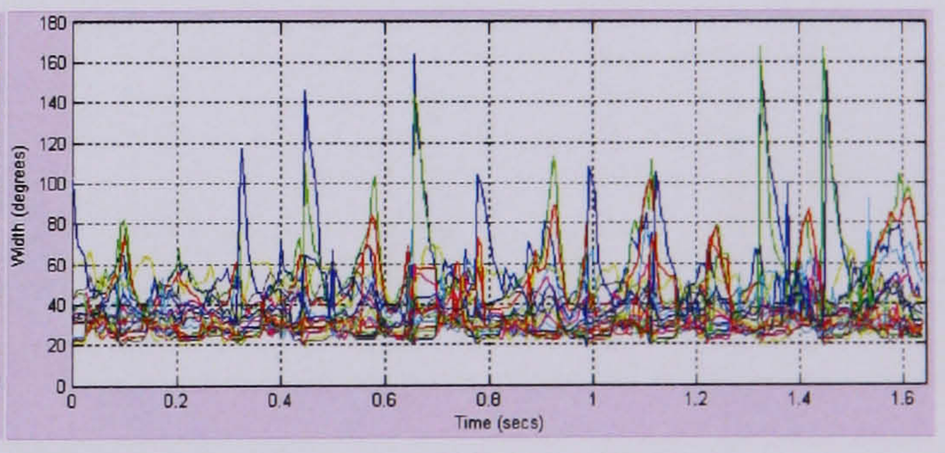
(a) Array 1: crosstalk-off (CR)



(b) Array 1: crosstalk-on (LCR)



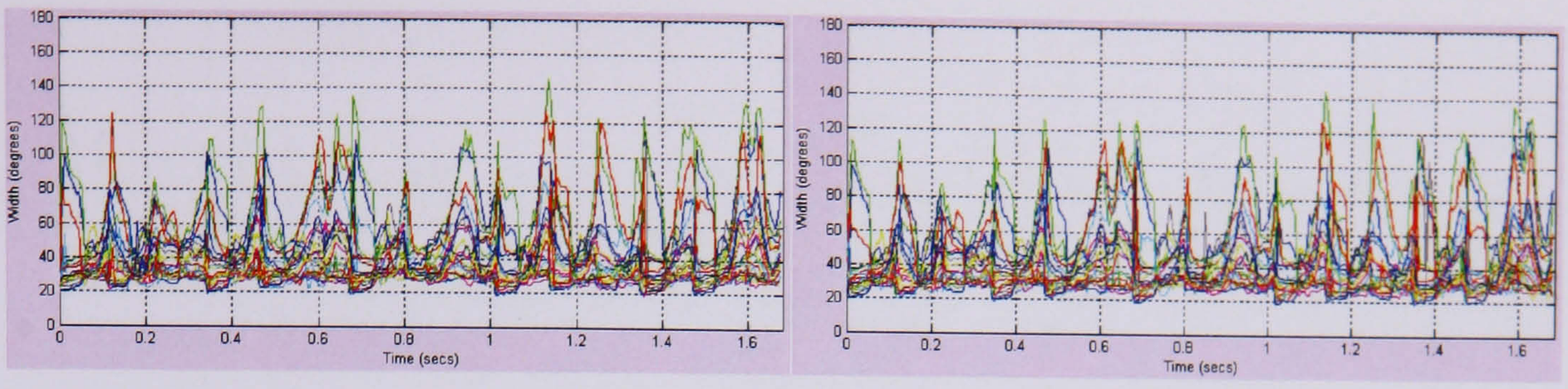
(c) Array 4: crosstalk-off (CR)



(d) Array 4: crosstalk-on (LCR)

Figure B.4 Comparisons of the plots of width measurements for the 'bongo' stimuli that were created in 'anechoic' condition

Appendix B Plots from objective measurements of the effects of interchannel crosstalk



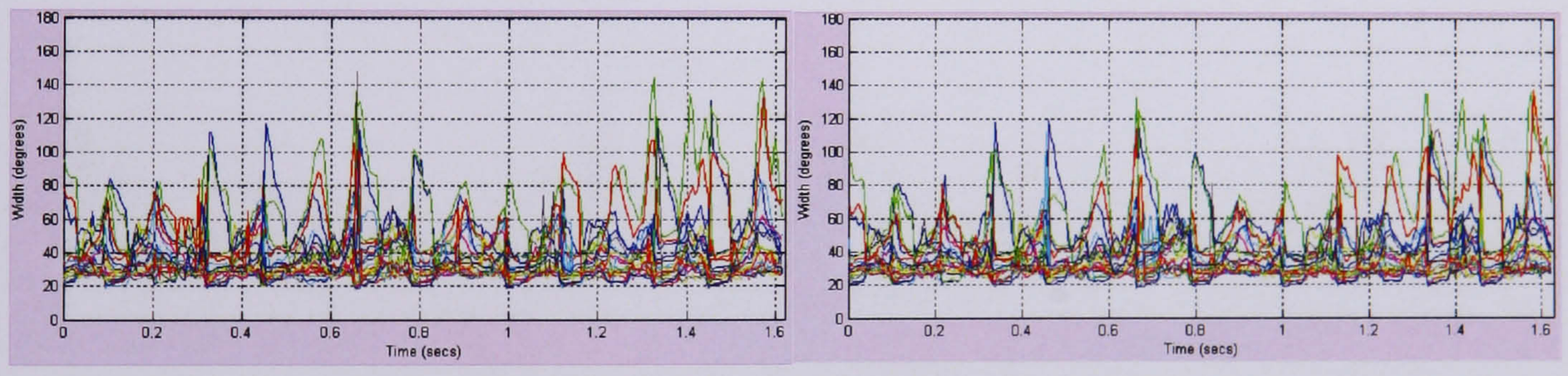
(a) Array 1: crosstalk-off (CR)

(b) Array 1: crosstalk-on (LCR)

(c) Array 4: crosstalk-off (CR)

(d) Array 4: crosstalk-on (LCR)

Figure B.5 Comparisons of the plots of width measurements for the 'bongo' stimuli that were created in 'room' condition



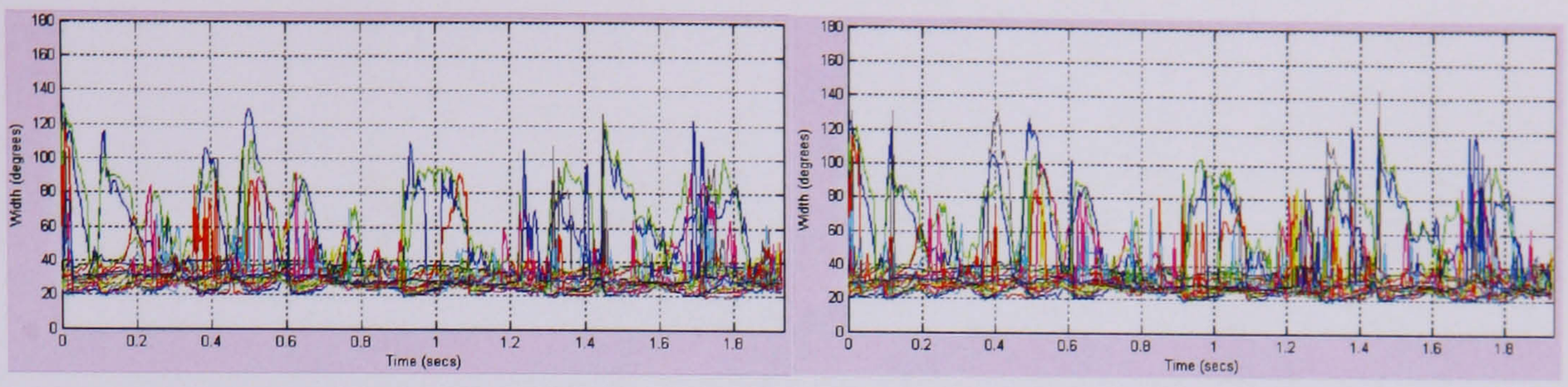
(a) Array 1: crosstalk-off (CR)

(b) Array 1: crosstalk-on (LCR)

(c) Array 4: crosstalk-off (CR)

(d) Array 4: crosstalk-on (LCR)

Figure B.6 Comparisons of the plots of width measurements for the 'bongo' stimuli that were created in 'hall' condition



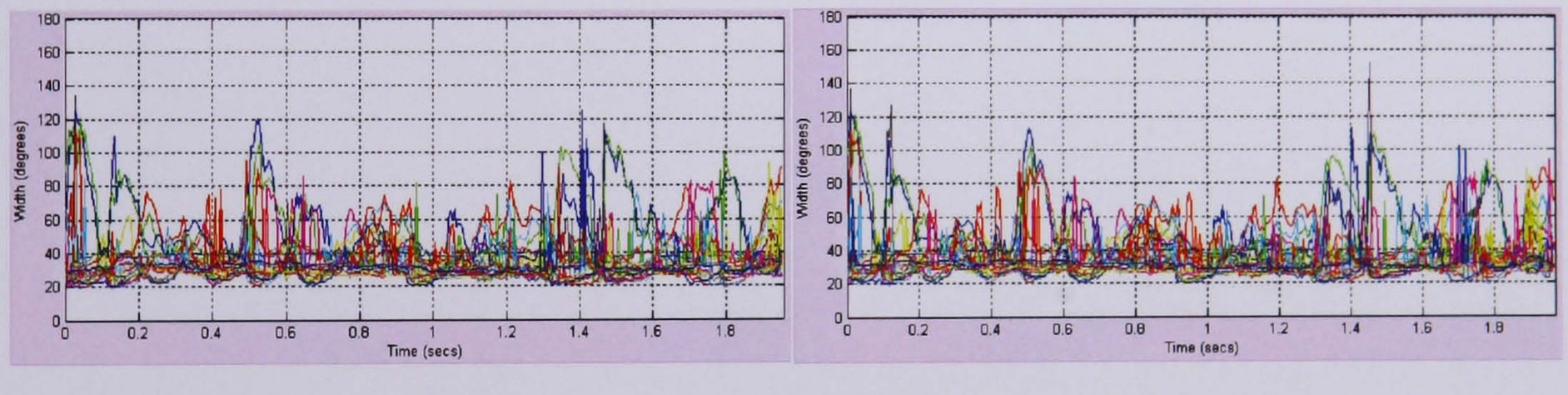
(a) Array 1: crosstalk-off (CR)

(b) Array 1: crosstalk-on (LCR)

(c) Array 4: crosstalk-off (CR)

(d) Array 4: crosstalk-on (LCR)

Figure B.7 Comparisons of the plots of width measurements for the 'speech' stimuli that were created in 'anechoic' condition



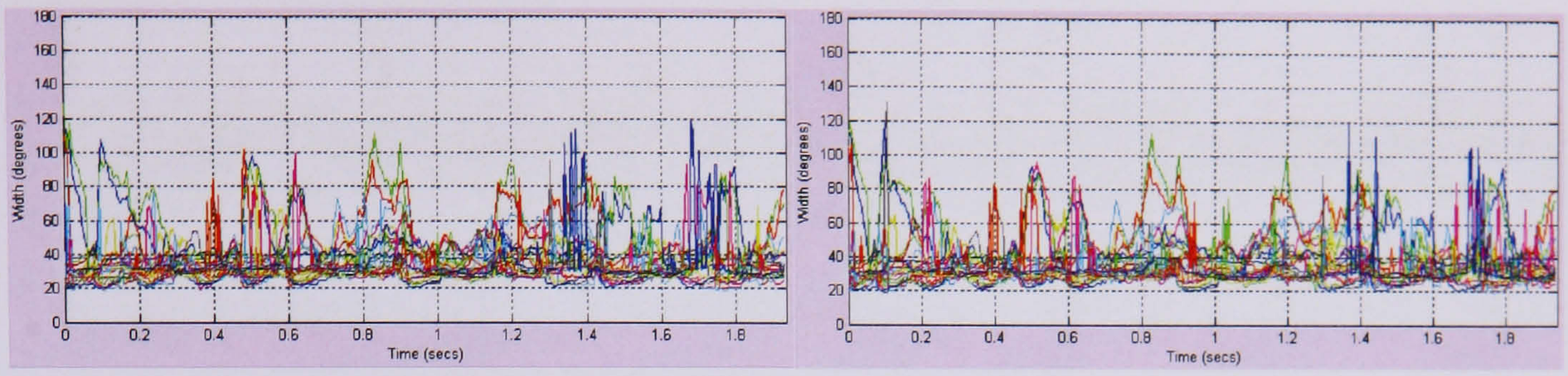
(a) Array 1: crosstalk-off (CR)

(b) Array 1: crosstalk-on (LCR)

(c) Array 4: crosstalk-off (CR)

(d) Array 4: crosstalk-on (LCR)

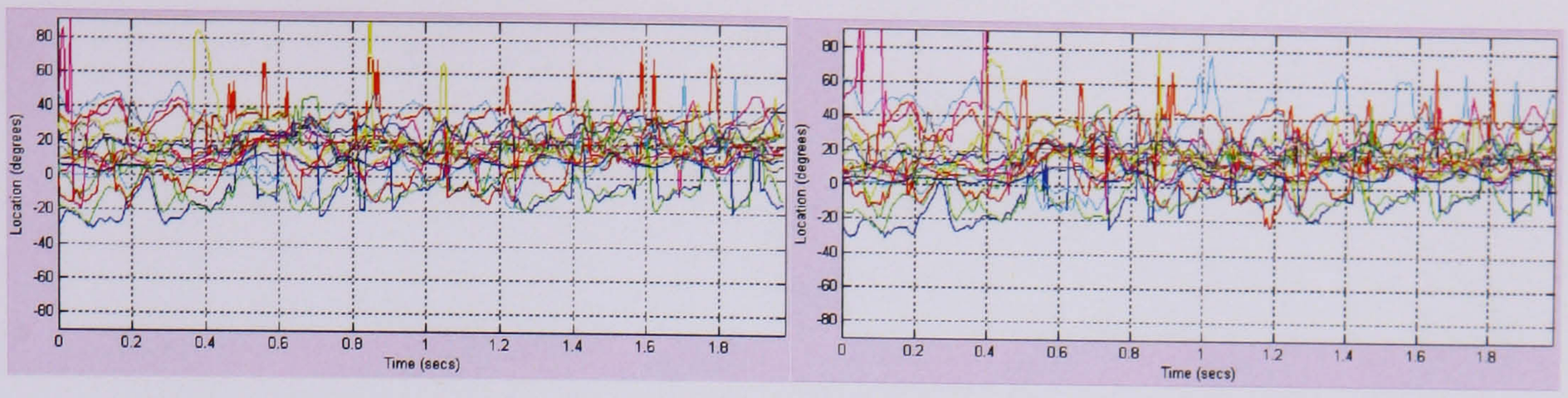
Figure B.8 Comparisons of the plots of width measurements for the 'speech' stimuli that were created in 'room' condition



(a) Array 1: crosstalk-off (CR) (b) Array 1: crosstalk-on (LCR)

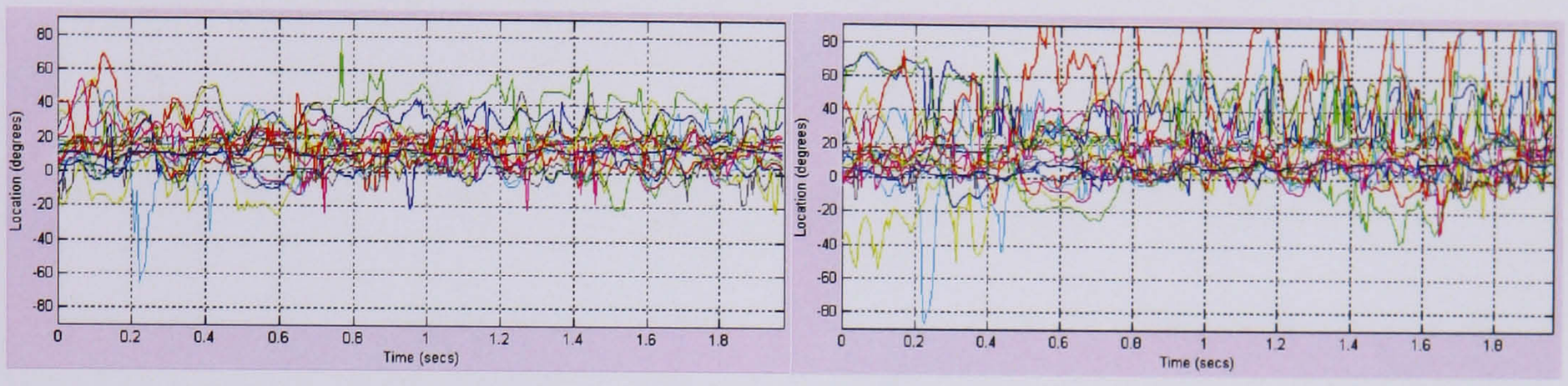
(c) Array 4: crosstalk-off (CR) (d) Array 4: crosstalk-on (LCR)

Figure B.9 Comparisons of the plots of width measurements for the 'speech' stimuli that were created in 'hall' condition



(a) Array 1: crosstalk-off (CR)

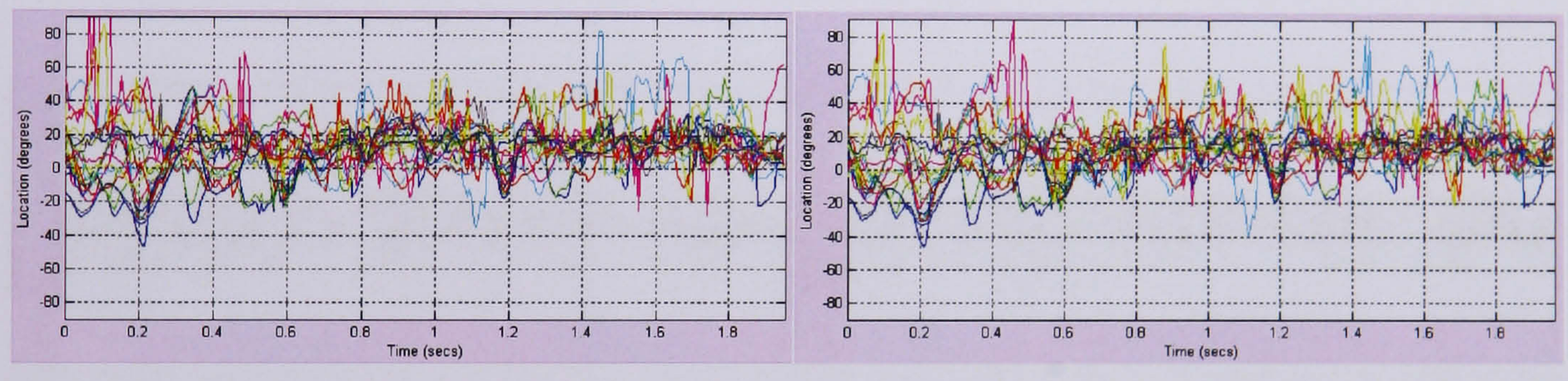
(b) Array 1: crosstalk-off (LCR)



(c) Array4: crosstalk-off (CR)

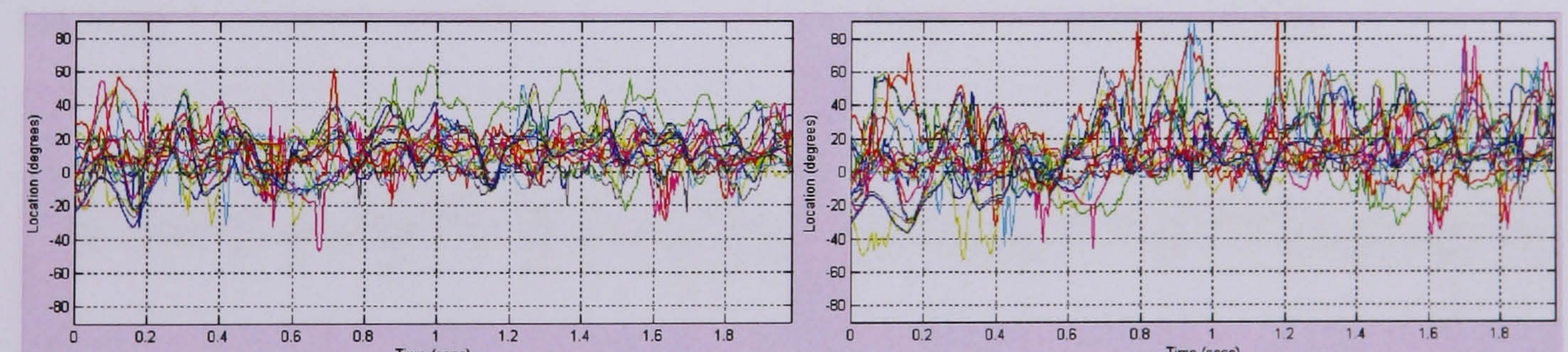
(d) Array 4: crosstalk-on (LCR)

Figure B.10 Comparisons of the plots of location measurements for the 'cello' stimuli that were created in 'anechoic' condition



(a) Array 1: crosstalk-off (CR)

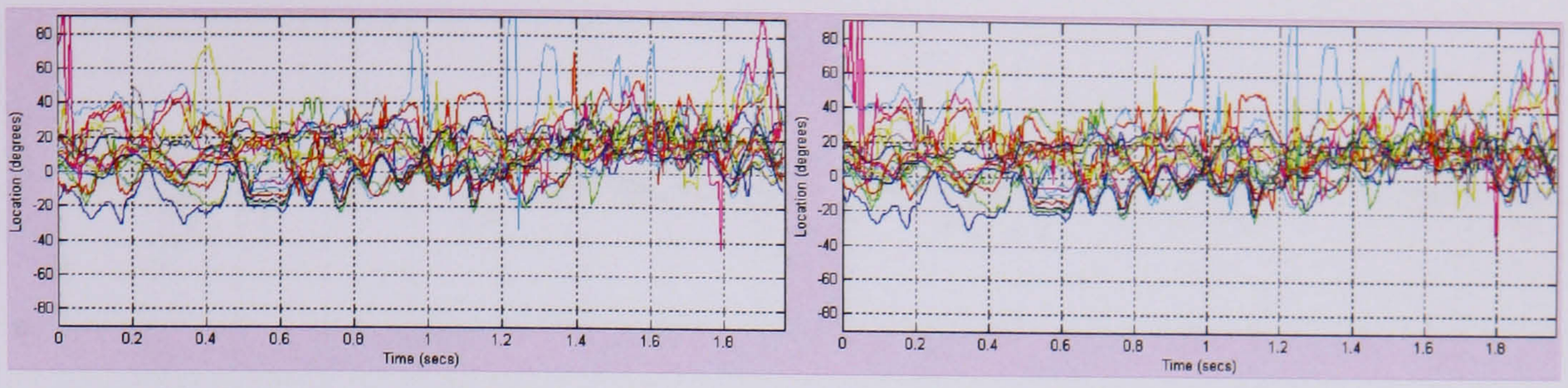
(b) Array 1: crosstalk-off (LCR)



(c) Array4: crosstalk-off (CR)

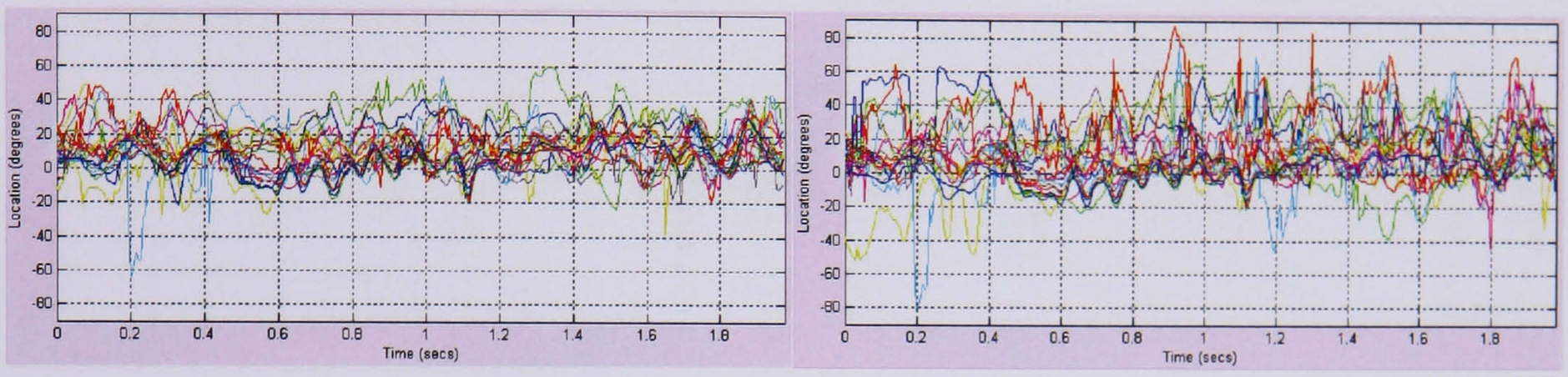
(d) Array 4: crosstalk-on (LCR)

Figure B.11 Comparisons of the plots of location measurements for the 'cello' stimuli that were created in 'room' condition



(a) Array 1: crosstalk-off (CR)

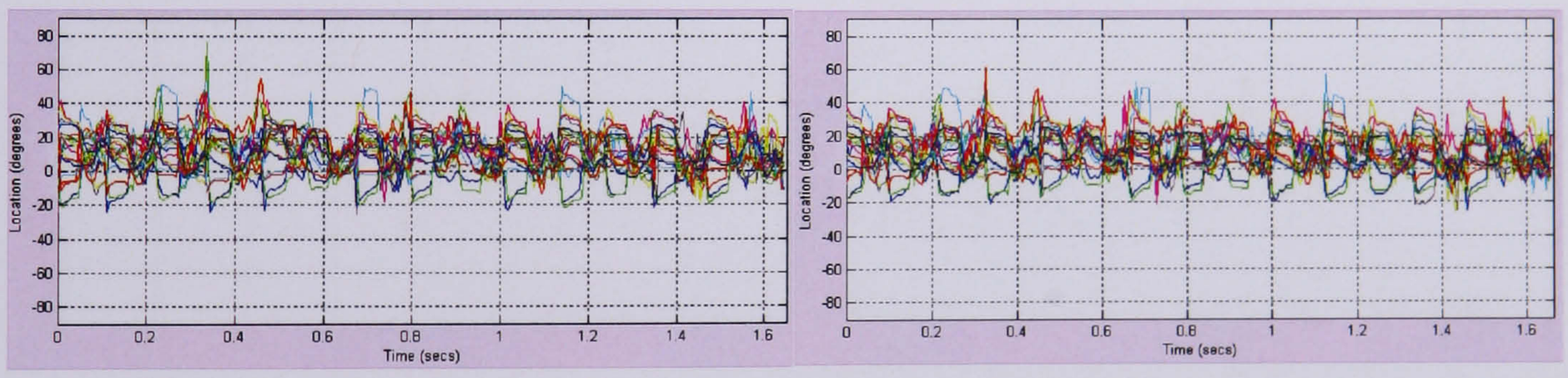
(b) Array a: crosstalk-on (LCR)



(c) Array 4: crosstalk-off (CR)

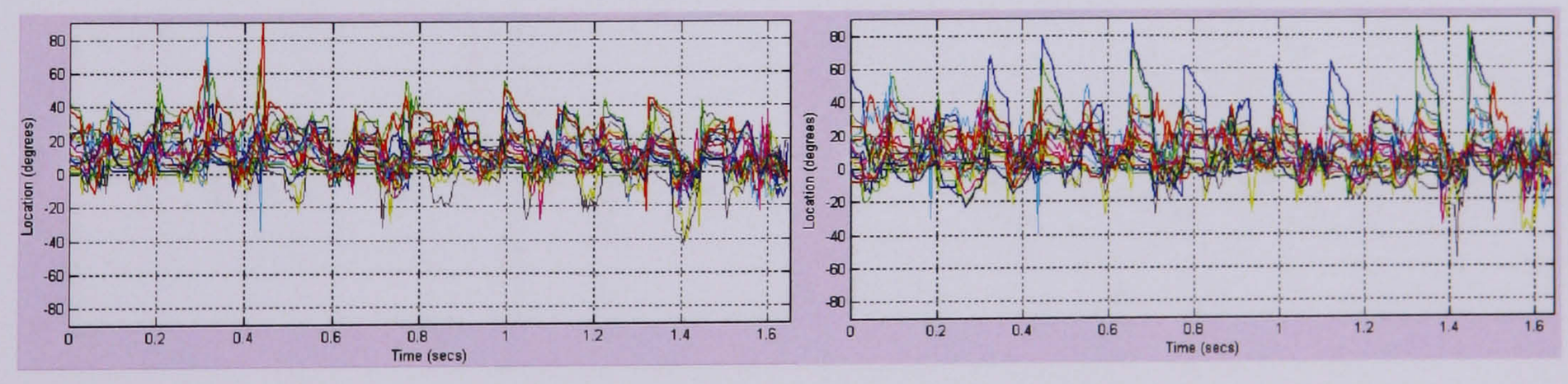
(d) Array 4: crosstalk-on (LCR)

Figure B.12 Comparisons of the plots of location measurements for the 'cello' stimuli that were created in 'hall' condition



(a) Array 1: crosstalk-off (CR)

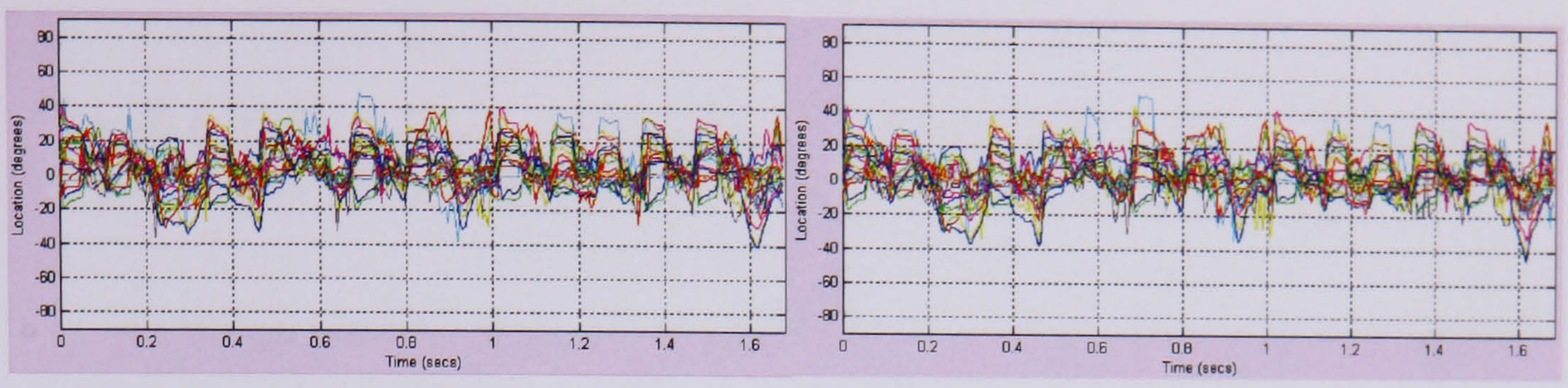
(b) Array 1: crosstalk-on (LCR)



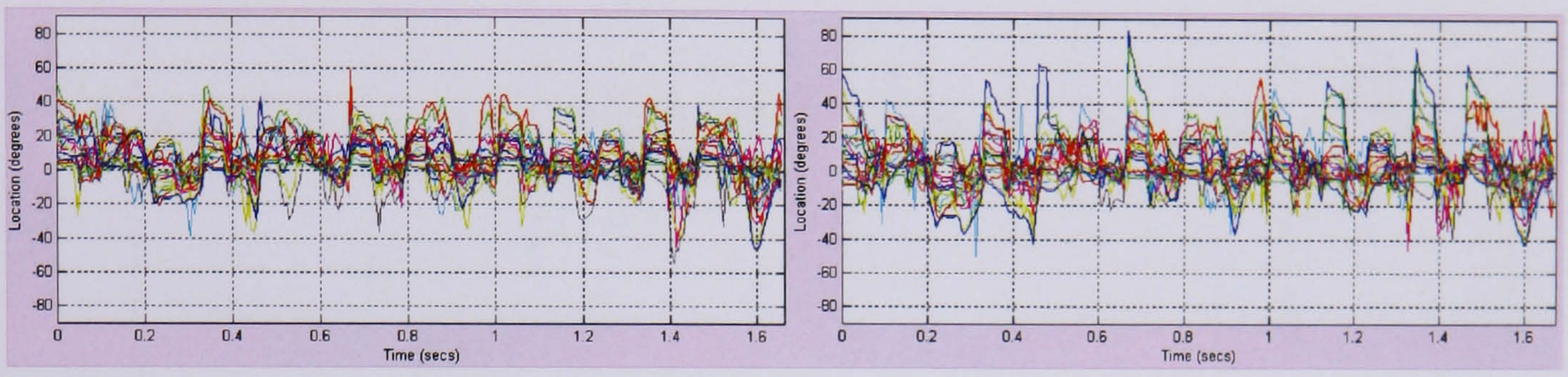
(c) Array 4: crosstalk-off (CR)

(d) Array 4: crosstalk-on (LCR)

Figure B.13 Comparisons of the plots of location measurements for the 'bongo' stimuli that were created in 'anechoic' condition

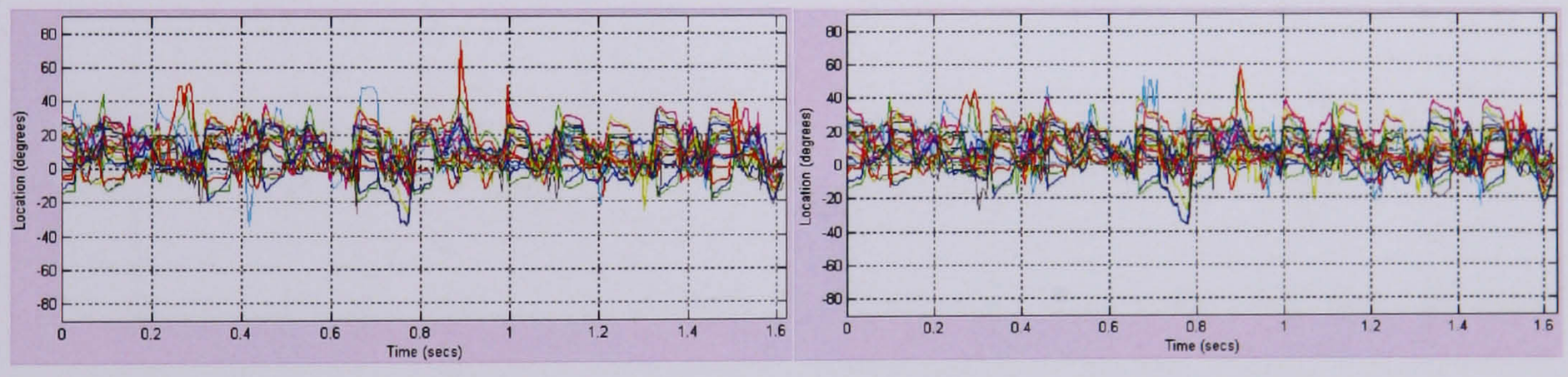


(a) Array 1: crosstalk-off (CR) (b) Array 1: crosstalk-on (LCR)

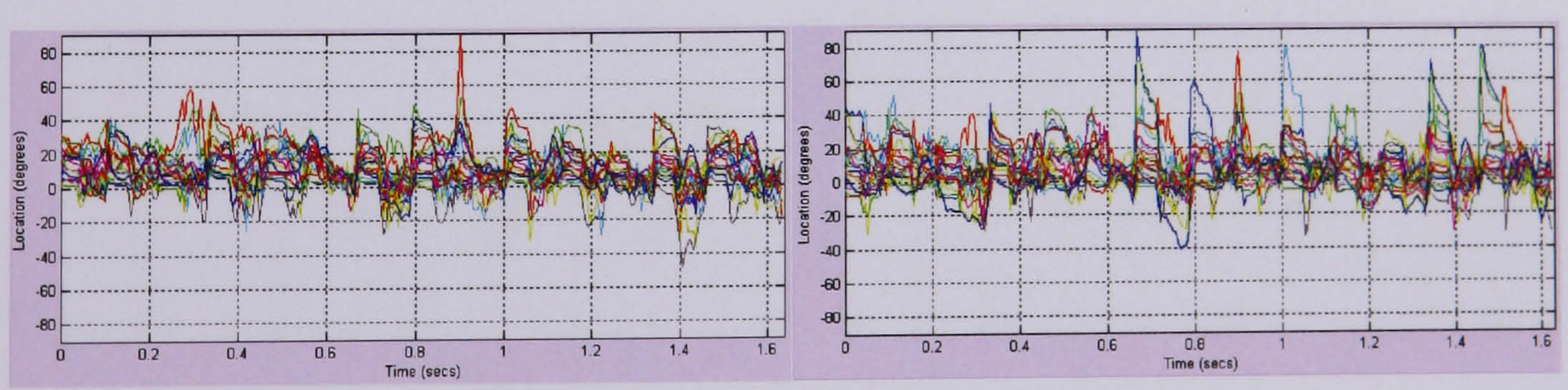


(c) Array 4: crosstalk-off (CR) (d) Array 4: crosstalk-on (LCR)

Figure B.14 Comparisons of the plots of location measurements for the 'bongo' stimuli that were created in 'room' condition

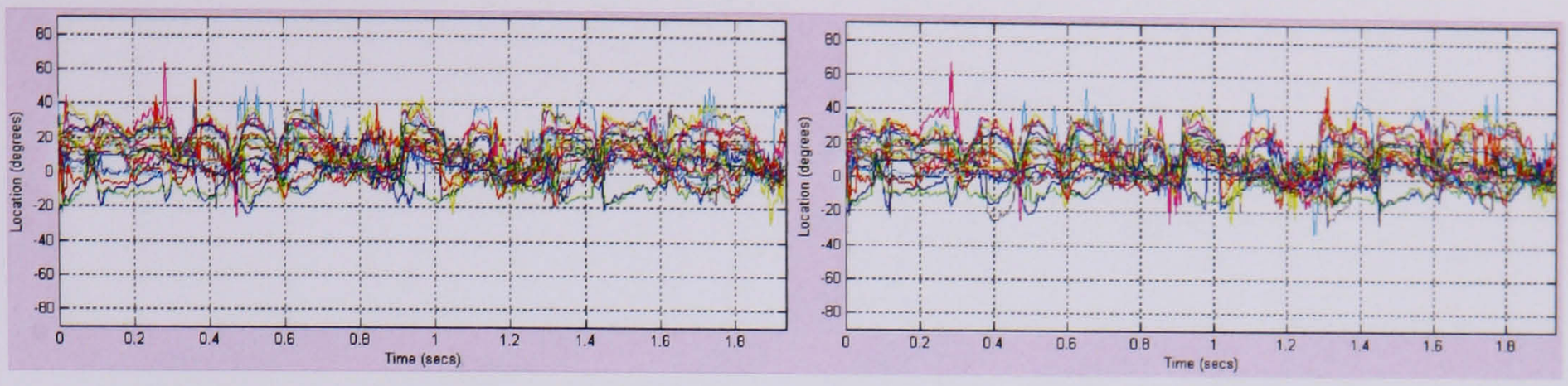


(a) Array 1: crosstalk-off (CR) (b) Array 1: crosstalk-on (LCR)

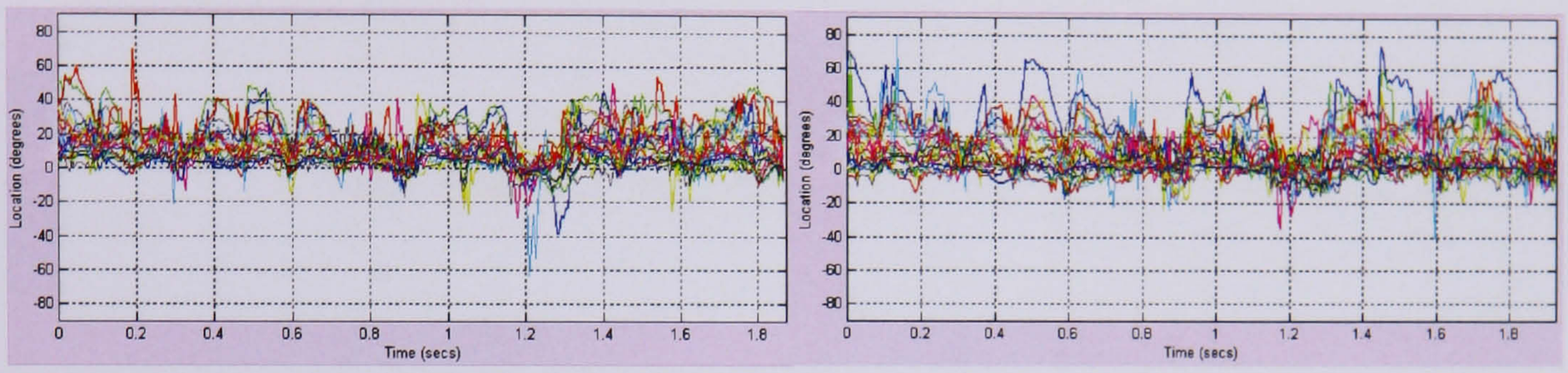


(c) Array 4: crosstalk-off (CR) (d) Array 4: crosstalk-on (LCR)

Figure B.15 Comparisons of the plots of location measurements for the 'bongo' stimuli that were created in 'hall' condition

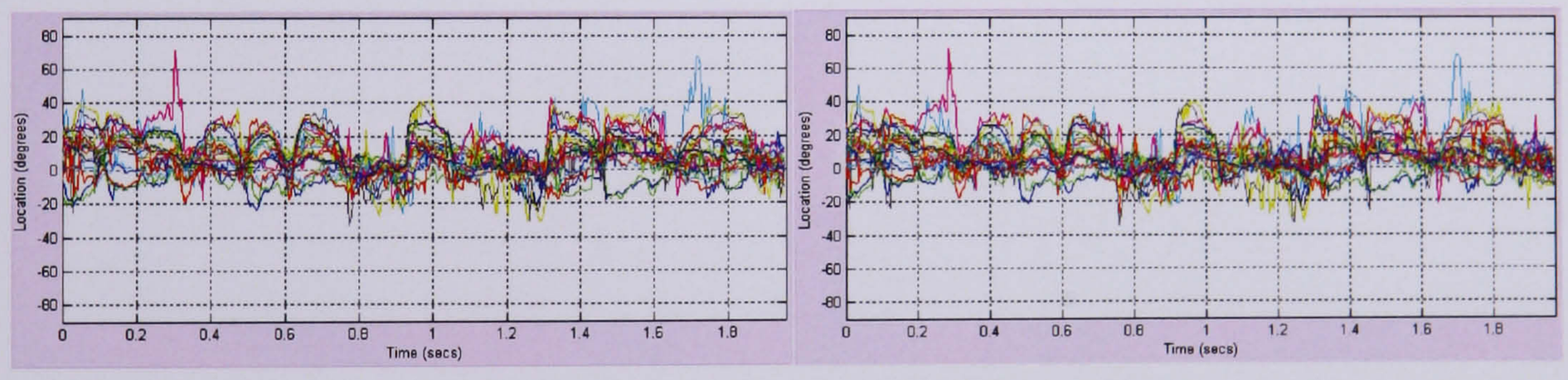


(a) Array 1: crosstalk-off (CR) (b) Array 1: crosstalk-on (LCR)

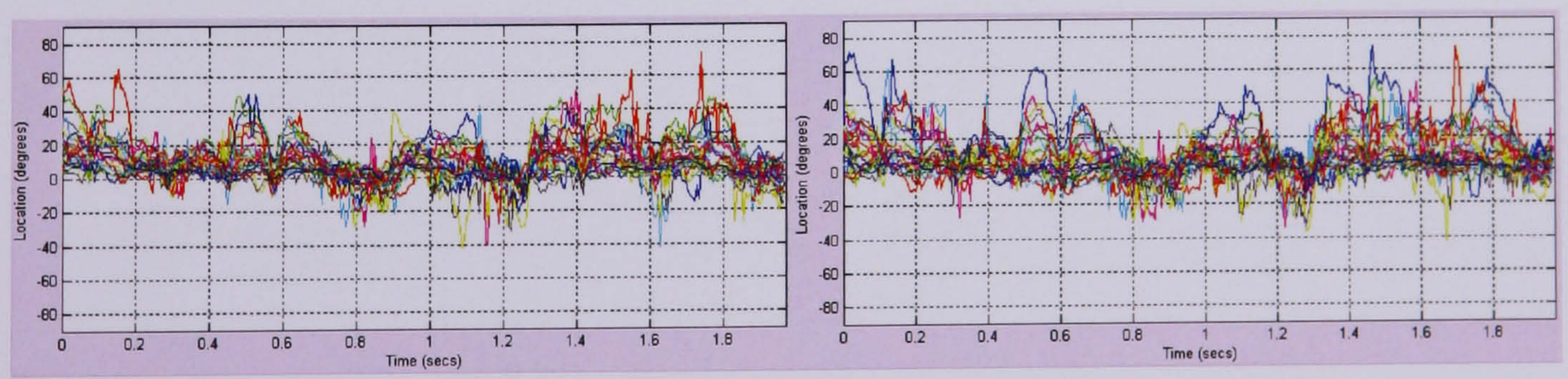


(c) Array 1: crosstalk-off (CR) (d) Array 4: crosstalk-on (LCR)

Figure B.16 Comparisons of the plots of location measurements for the 'speech' stimuli that were created in 'anechoic' condition

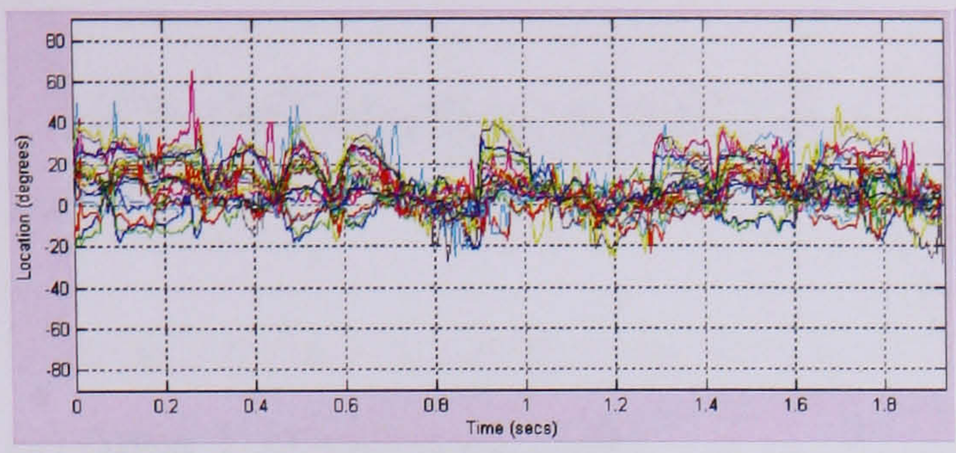


(a) Array 1: crosstalk-off (CR) (b) Array 1: crosstalk-on (LCR)

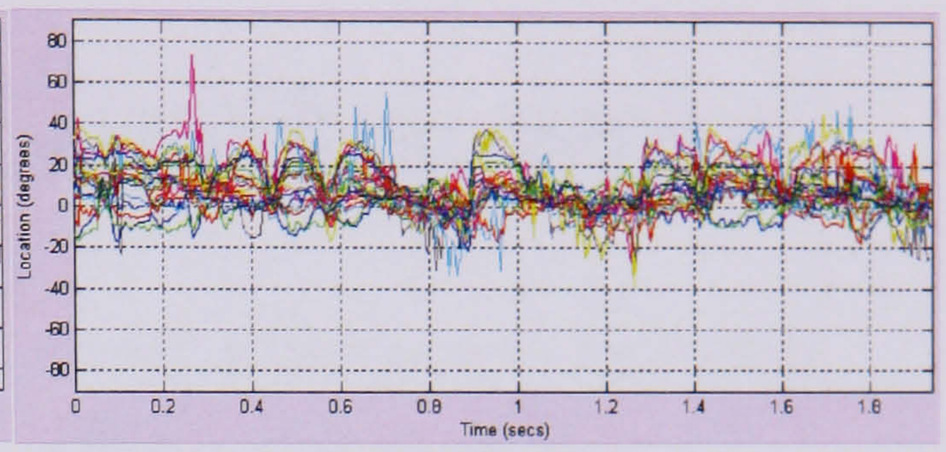


(c) Array 4: crosstalk-off (CR) (d) Array 4: crosstalk-on (LCR)

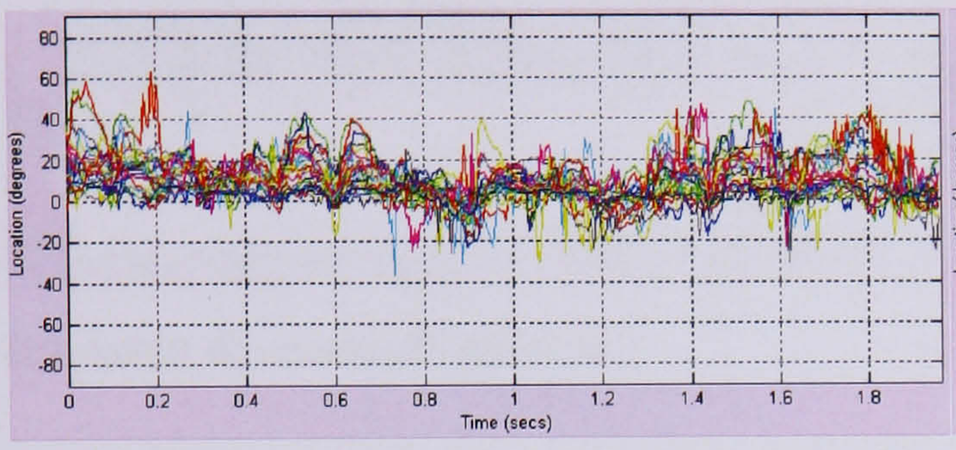
Figure B.17 Comparisons of the plots of location measurements for the 'speech' stimuli that were created in 'room' condition



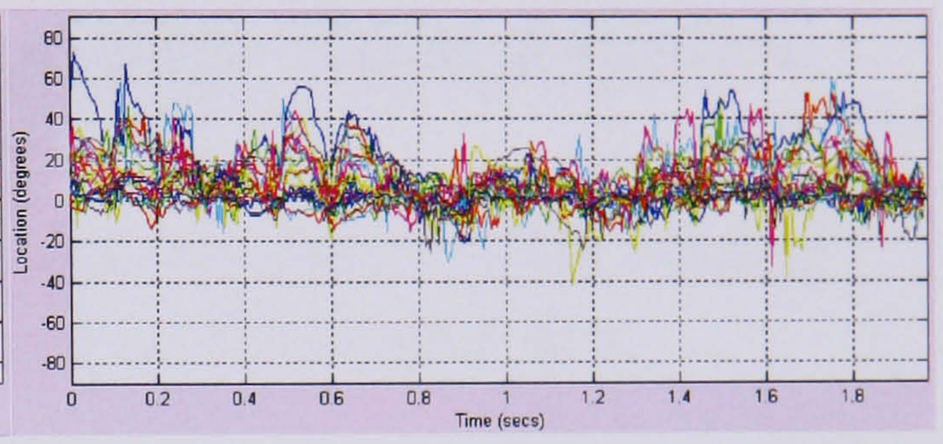
(a) Array 1: crosstalk-off (CR)



(b) Array 1: crosstalk-on (LCR)

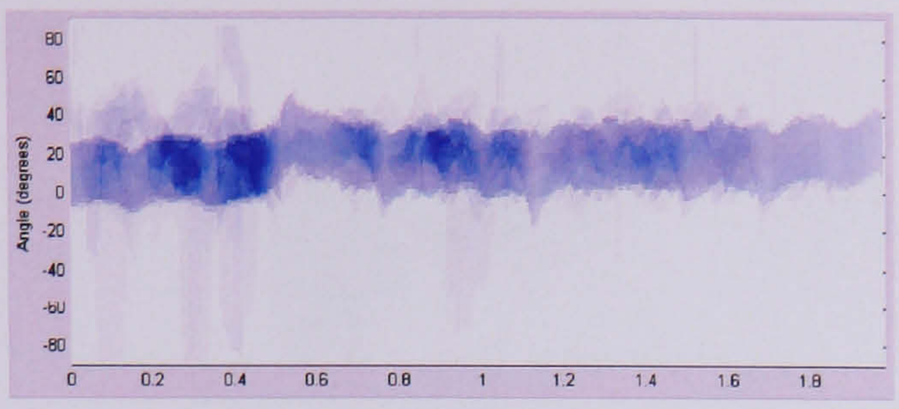


(c) Array 1: crosstalk-off (CR)

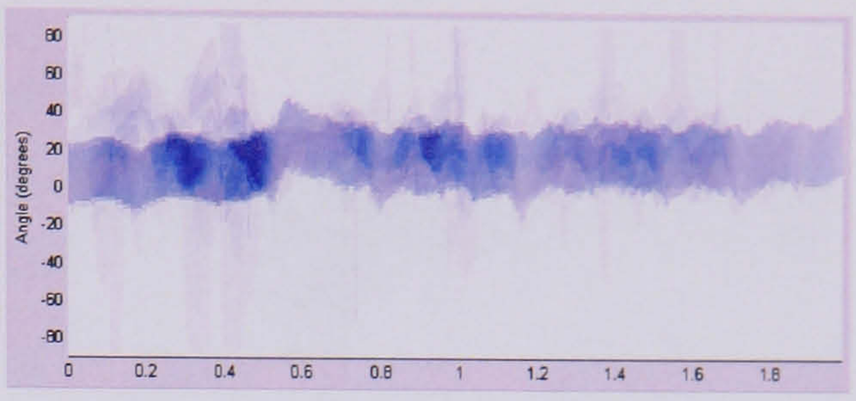


(d) Array 1: crosstalk-on (LCR)

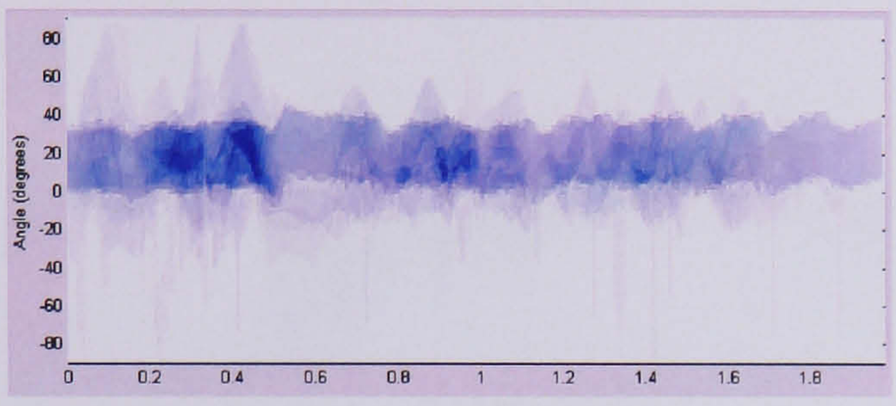
Figure B.18 Comparisons of the plots of location measurements for the 'speech' stimuli that were created in 'hall' condition



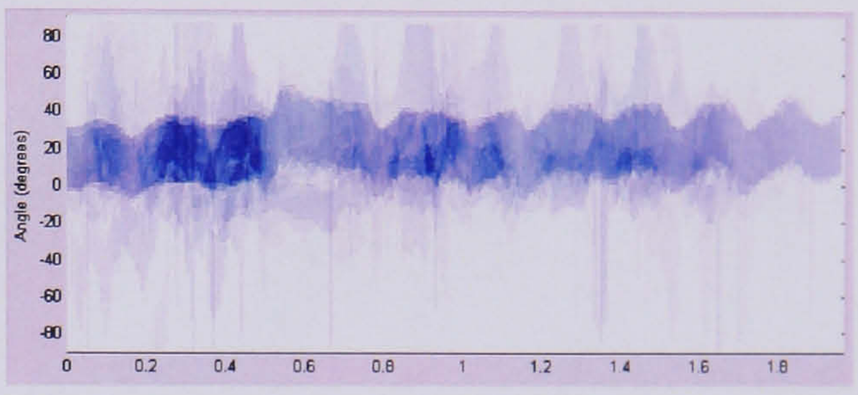
(a) Array 1: crosstalk-off (CR)



(b) Array 1: crosstalk-on (LCR)

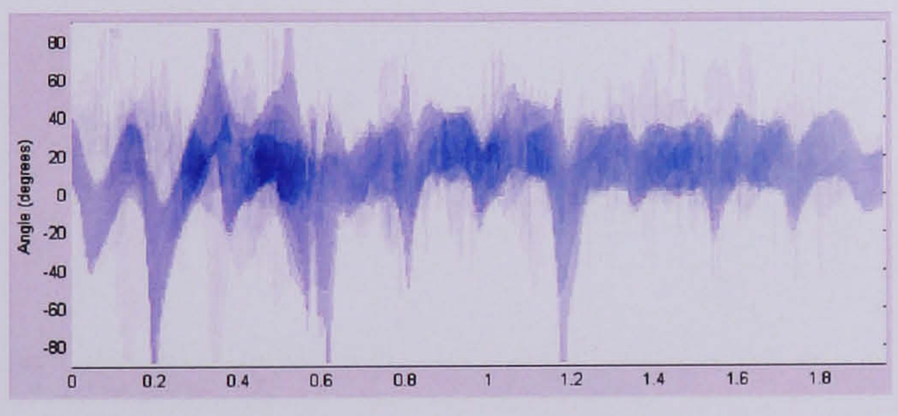


(c) Array 4: crosstalk-off (CR)

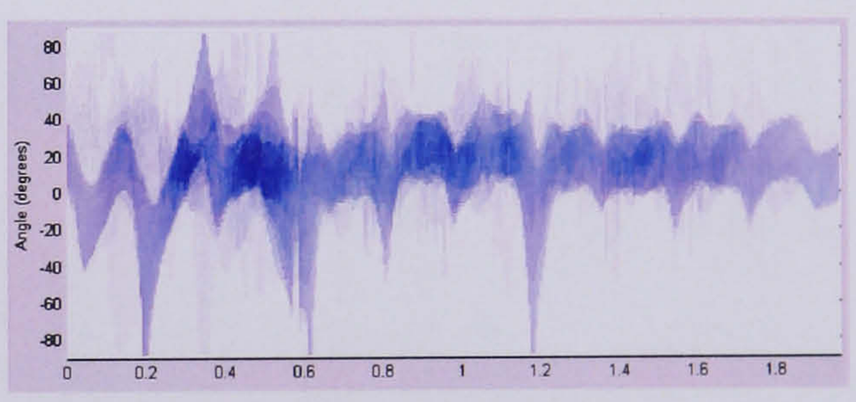


(d) Array 4: crosstalk-on (LCR)

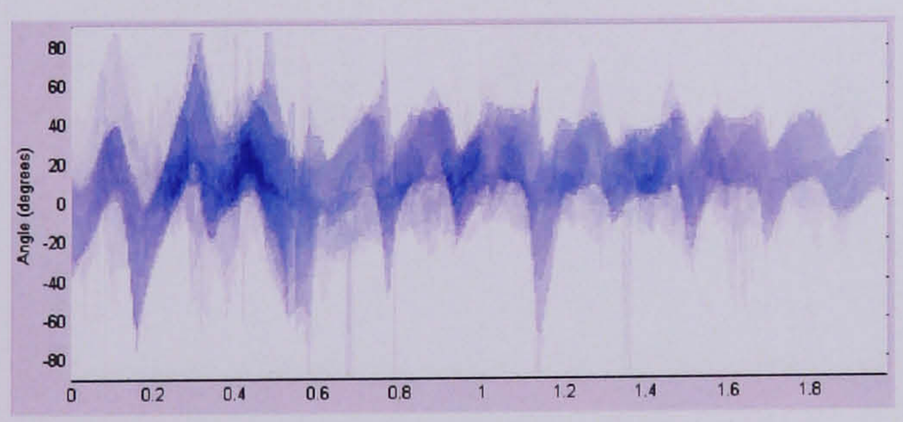
Figure B.19 Comparisons of the plots of width and location measurements for the cello stimuli that were created in 'anechoic' condition



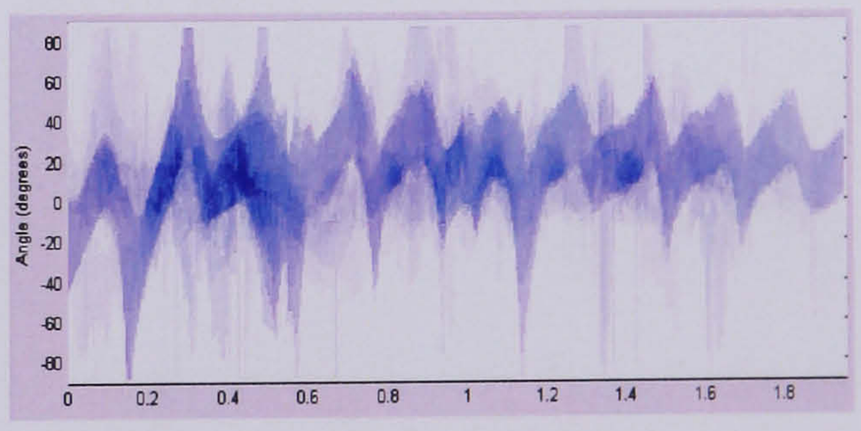
(a) Array 1: crosstalk-off (CR)



(b) Array 1: crosstalk-on (LCR)

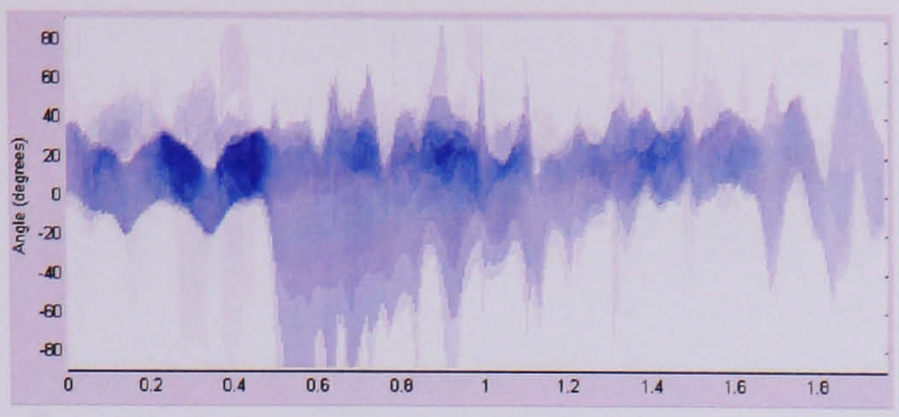


(c) Array 4: crosstalk-off (CR)

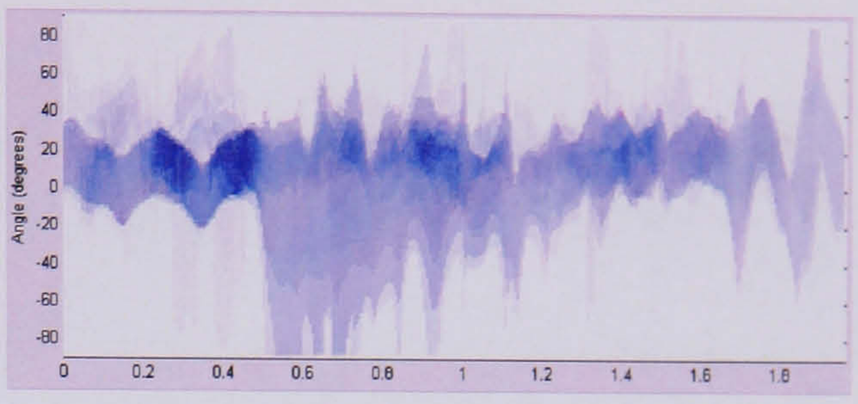


(d) Array 4: crosstalk-on (LCR)

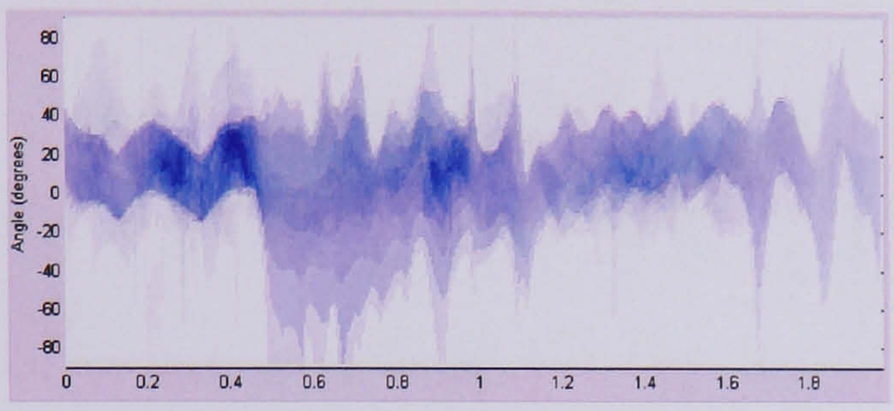
Figure B.20 Comparisons of the plots of width and location measurements for the cello stimuli that were created in 'room' condition



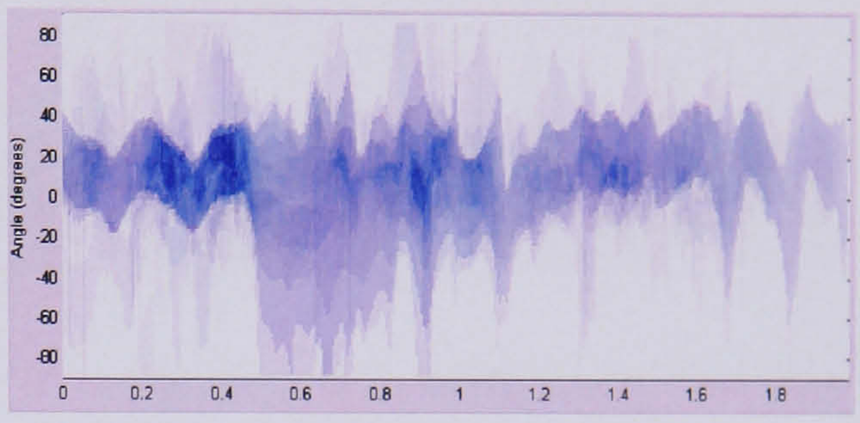
(a) Array 1: crosstalk-off (CR)



(b) Array 1: crosstalk-on (LCR)

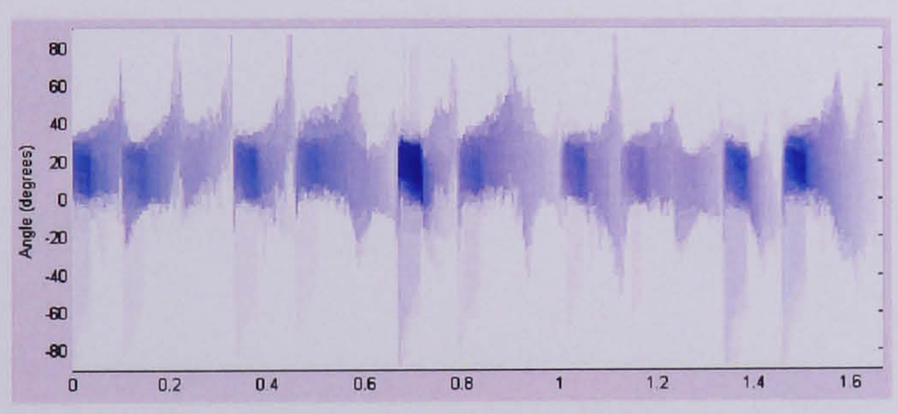


(c) Array 4: crosstalk-off (CR)

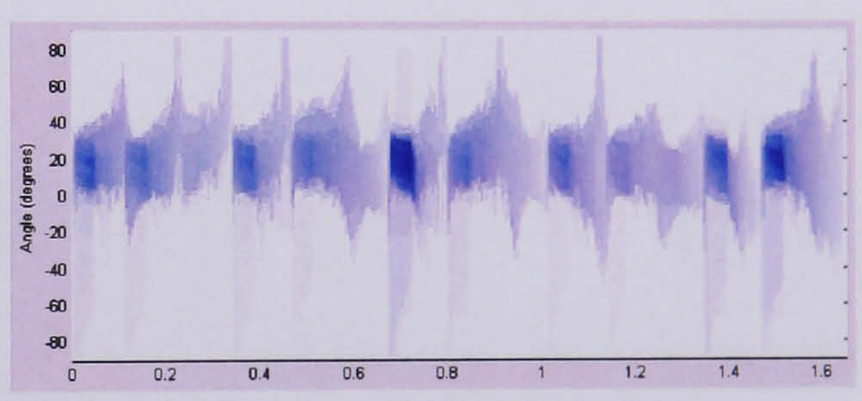


(d) Array 4: crosstalk-on (LCR)

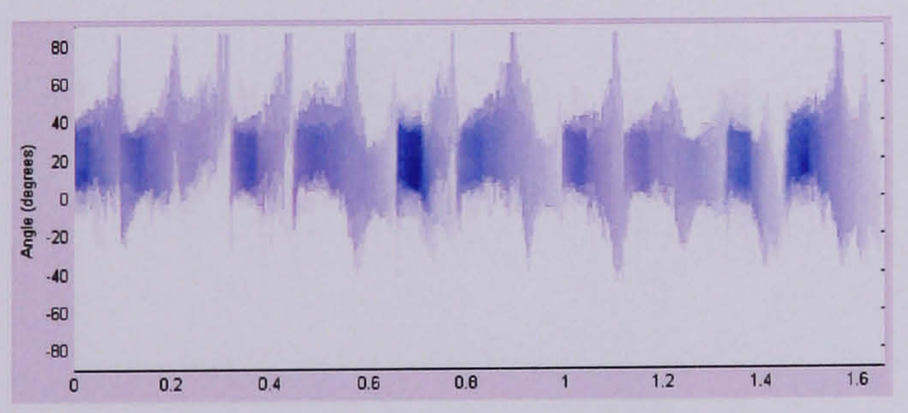
Figure B.21 Comparisons of the plots of width and location measurements for the cello stimuli that were created in 'hall' condition



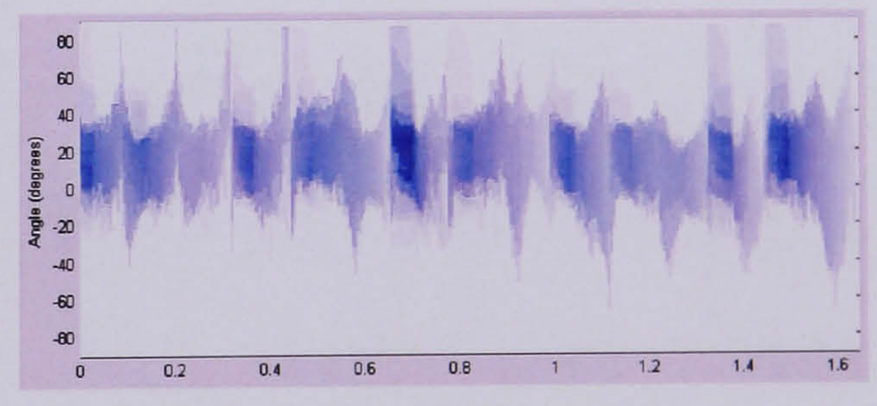
(a) Array 1: crosstalk-off (CR)



(b) Array 1: crosstalk-on (LCR)

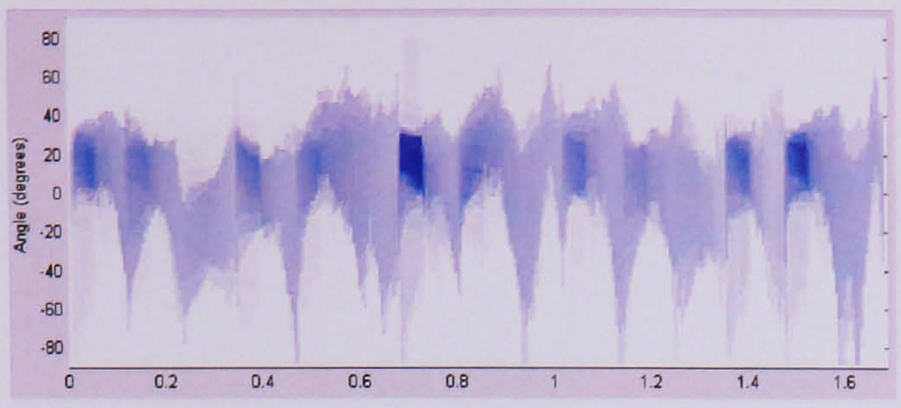


(c) Array 4: crosstalk-off (CR)

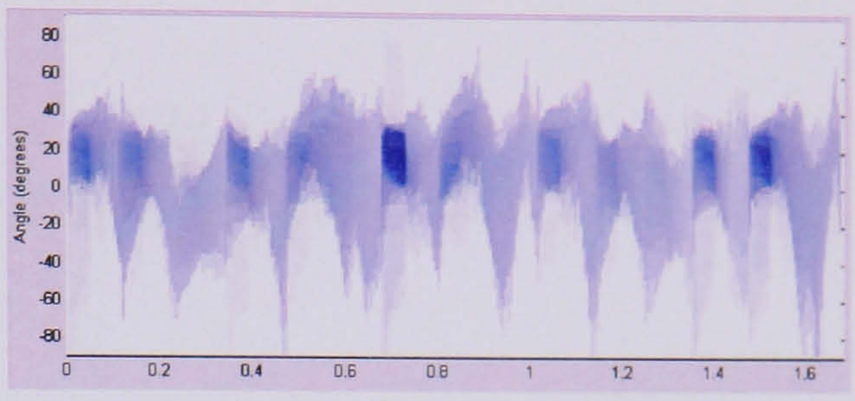


(d) Array 4: crosstalk-on (LCR)

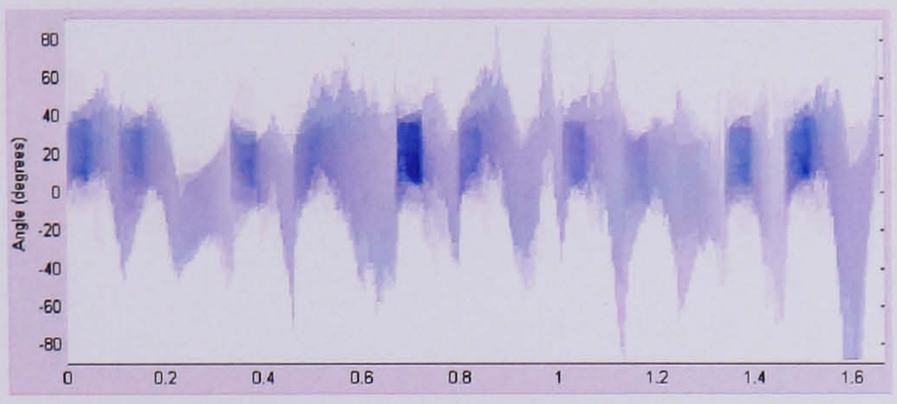
Figure B.22 Comparisons of the plots of width and location measurements for the bongo stimuli that were created in 'anechoic' condition



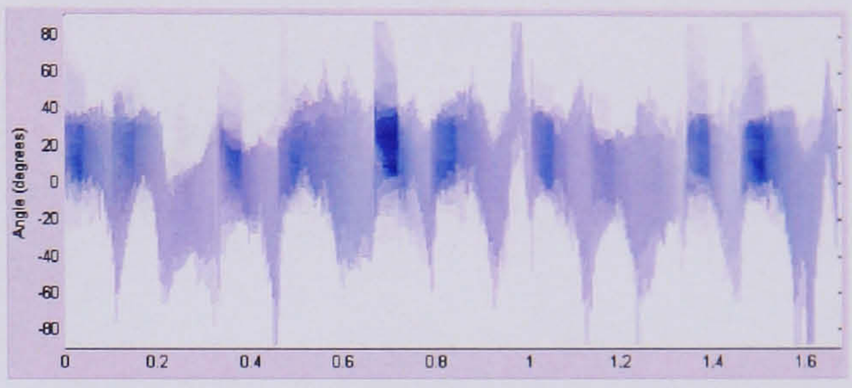
(a) Array 1: crosstalk-off (CR)



(b) Array 1: crosstalk-on (LCR)

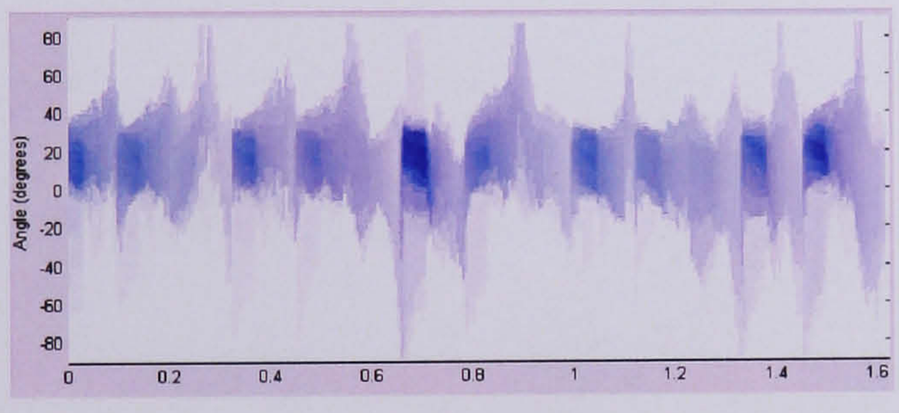


(c) Array 4: crosstalk-off (CR)

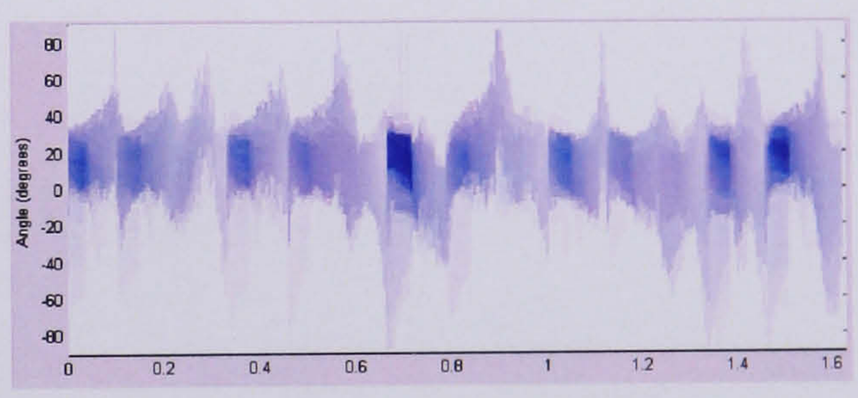


(d) Array 4: crosstalk-on (LCR)

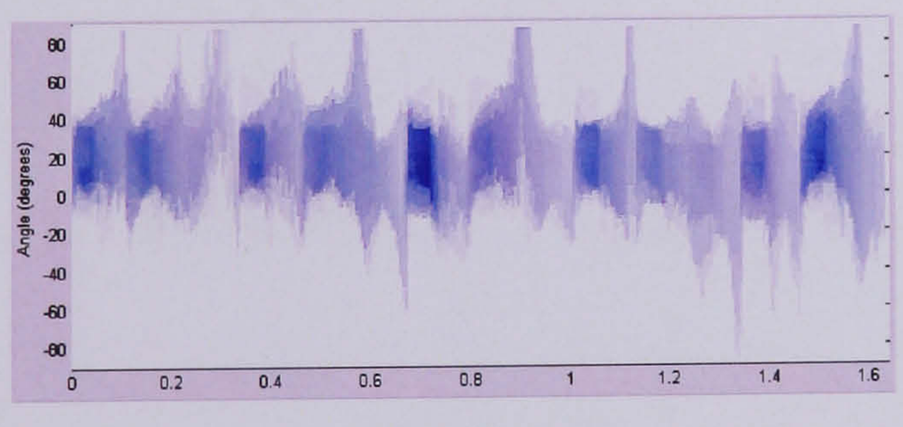
Figure B.23 Comparisons of the plots of width and location measurements for the bongo stimuli that were created in 'room' condition



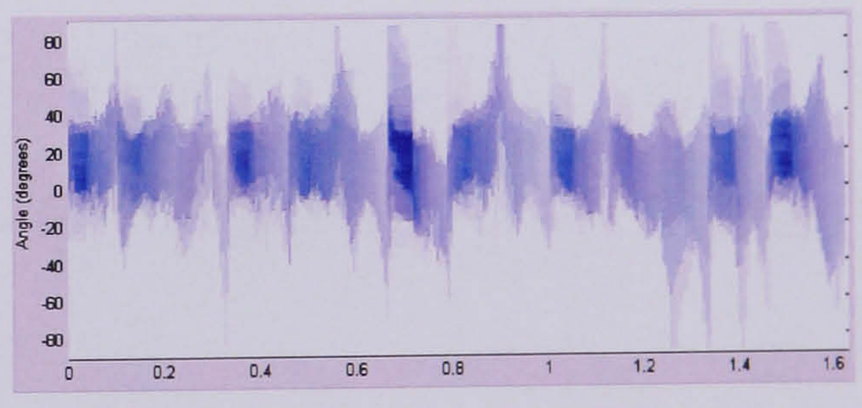
(a) Array 1: crosstalk-off (CR)



(b) Array 1: crosstalk-on (LCR)

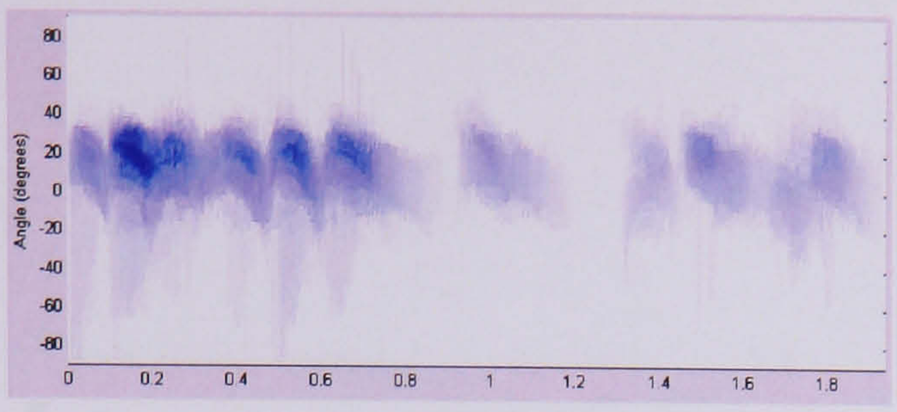


(c) Array 4: crosstalk-off (CR)

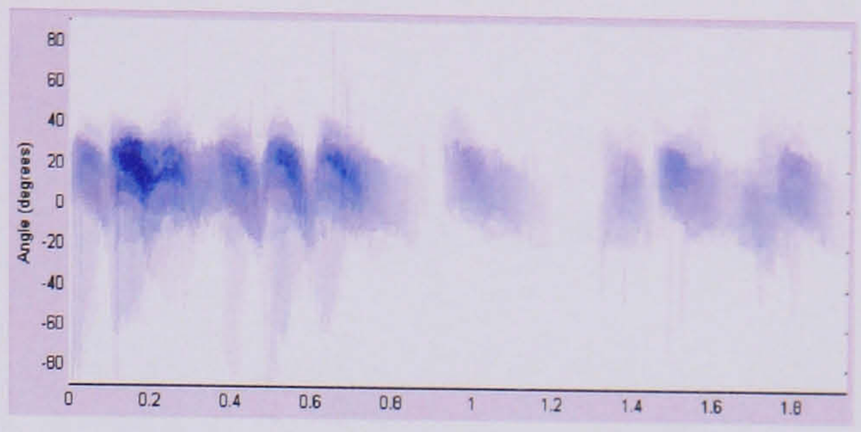


(d) Array 4: crosstalk-on (LCR)

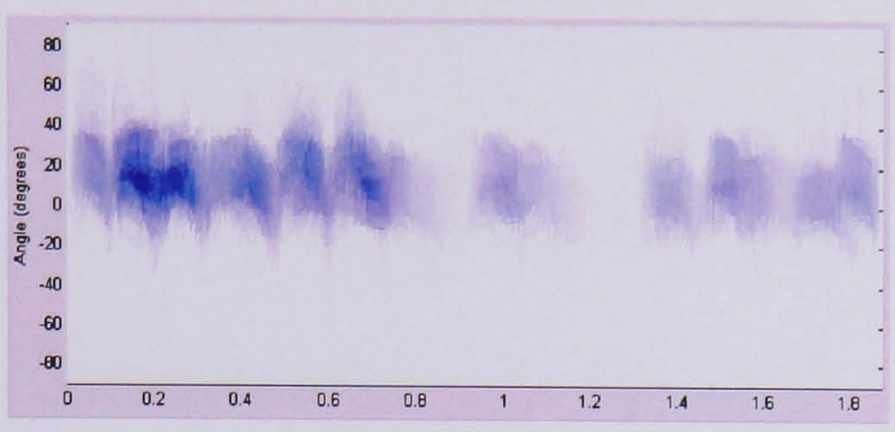
Figure B.24 Comparisons of the plots of width and location measurements for the bongo stimuli that were created in 'hall' condition



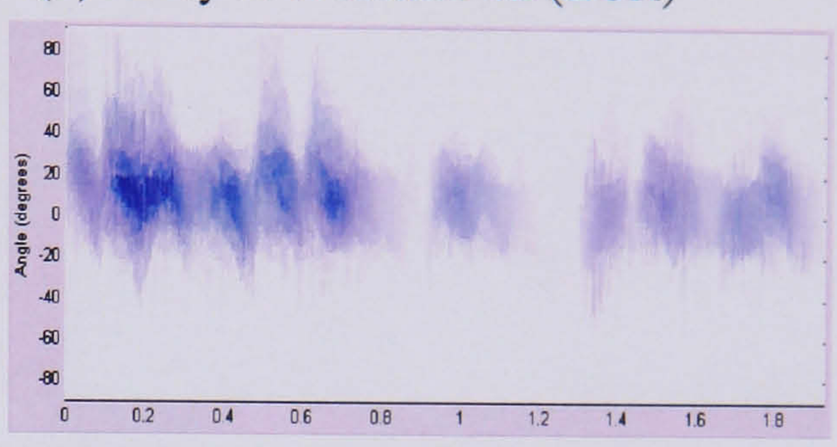
(a) Array 1: crosstalk-off (CR)



(b) Array 1: crosstalk-on (LCR)

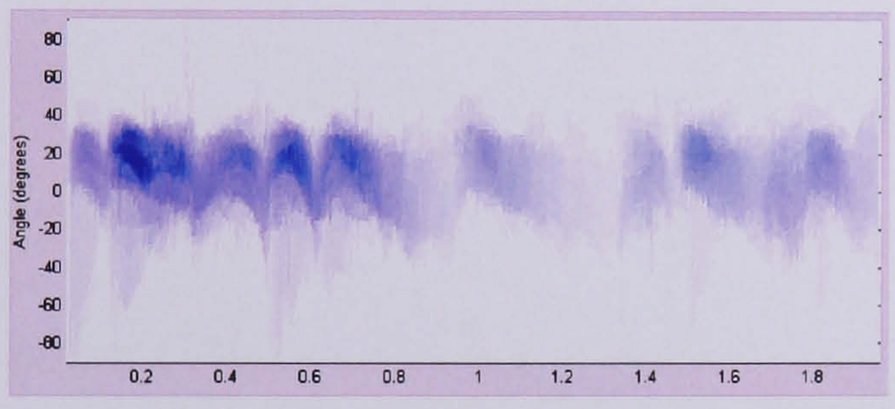


(c) Array 4: crosstalk-off (CR)

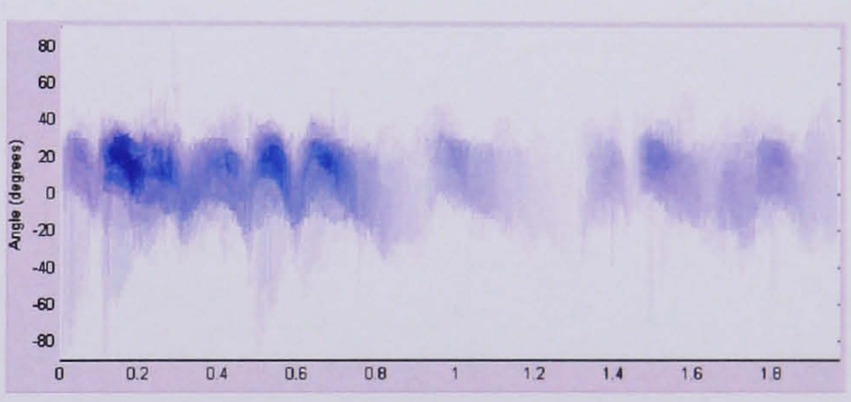


(d) Array 4: crosstalk-on (LCR)

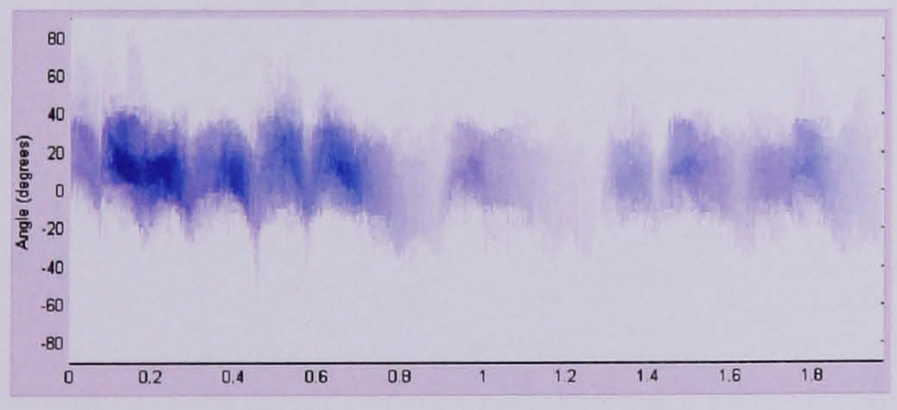
Figure B.25 Comparisons of the plots of width and location measurements for the speech stimuli that were created in 'anechoic' condition



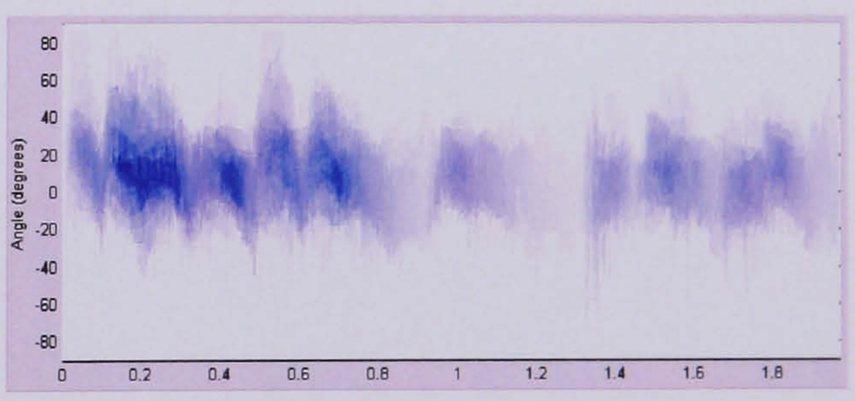
(a) Array 1: crosstalk-off (CR)



(b) Array 1: crosstalk-on (LCR)

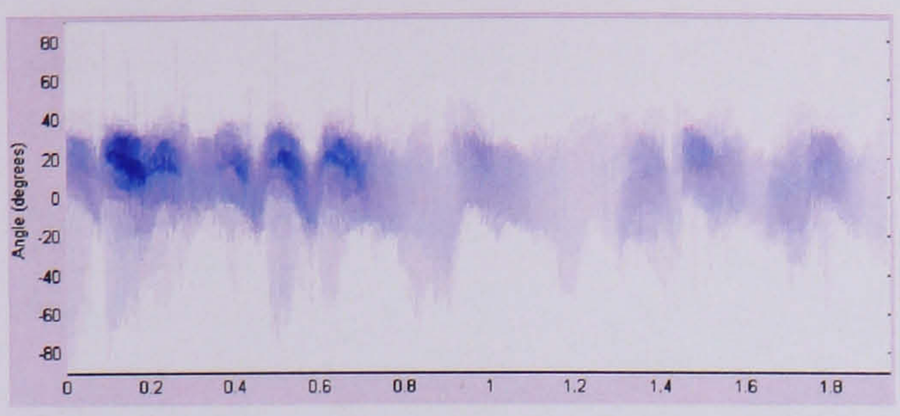


(c) Array 4: crosstalk-off (CR)

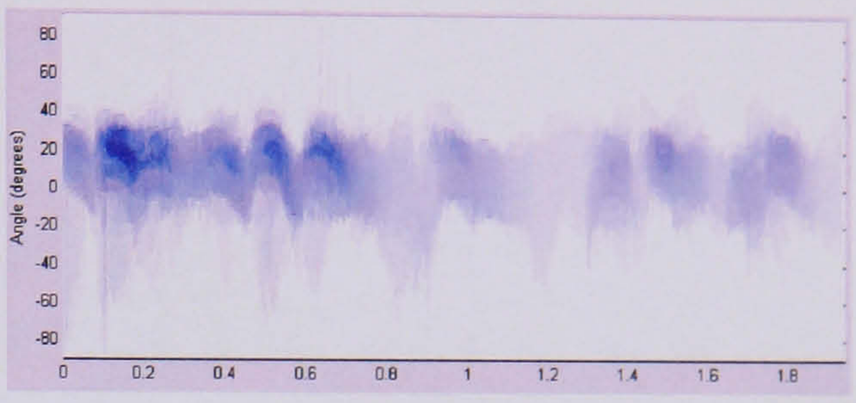


(d) Array 4: crosstalk-on (LCR)

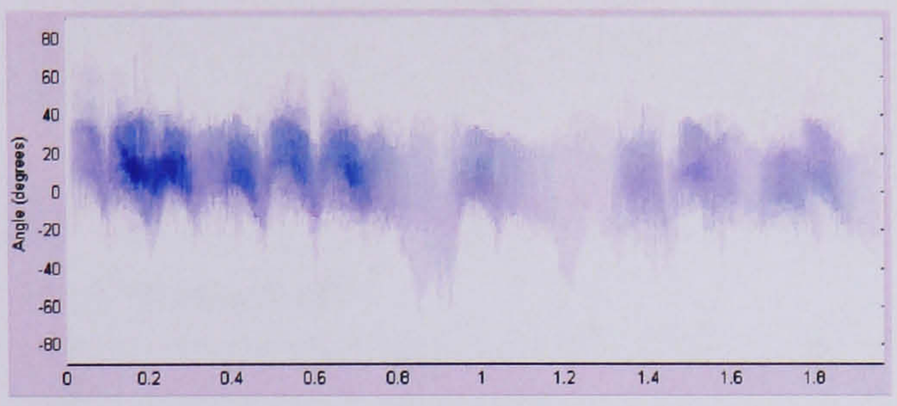
Figure B.26 Comparisons of the plots of width and location measurements for the speech stimuli that were created in 'room' condition



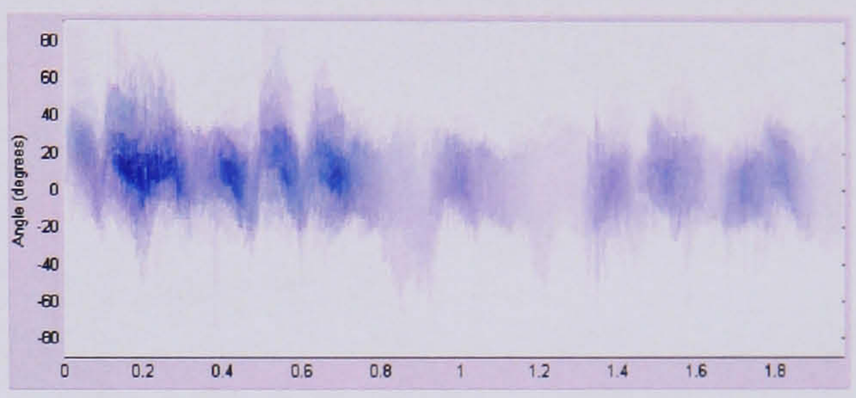
(a) Array 1: crosstalk-off (CR)



(b) Array 1: crosstalk-on (LCR)



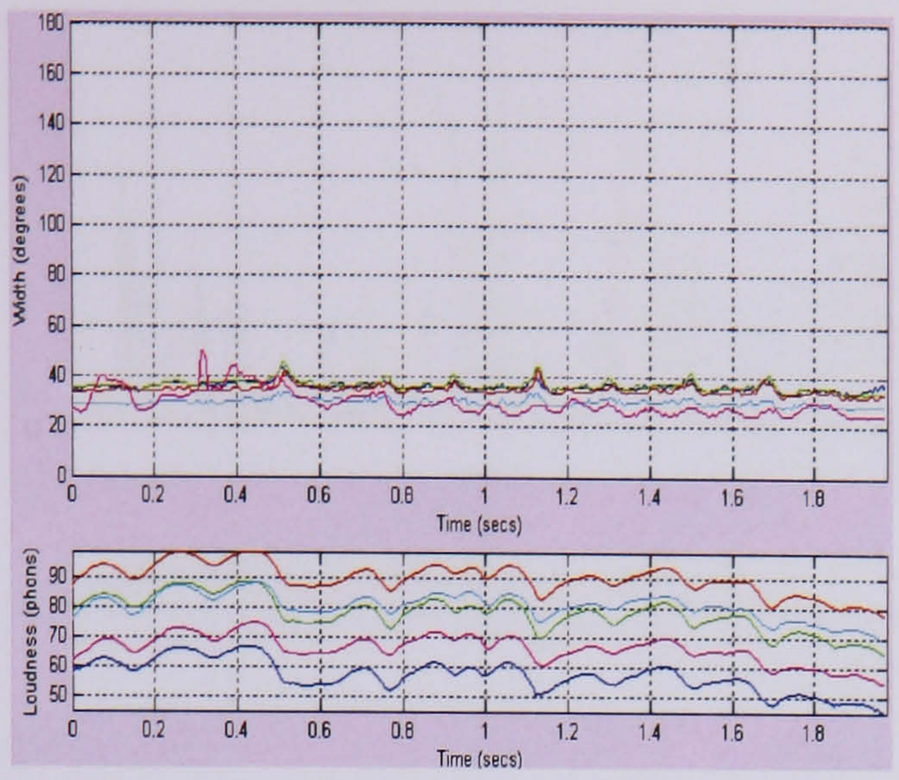
(c) Array 4: crosstalk-off (CR)



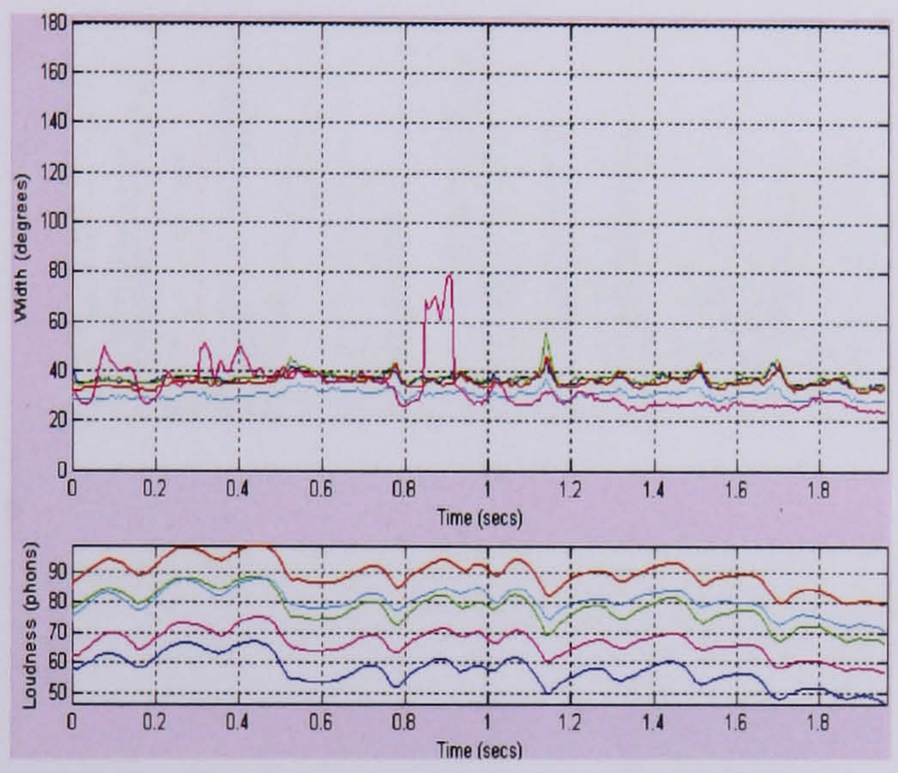
(d) Array 4: crosstalk-on (LCR)

Figure B.27 Comparisons of the plots of width and location measurements for the speech stimuli that were created in 'hall' condition

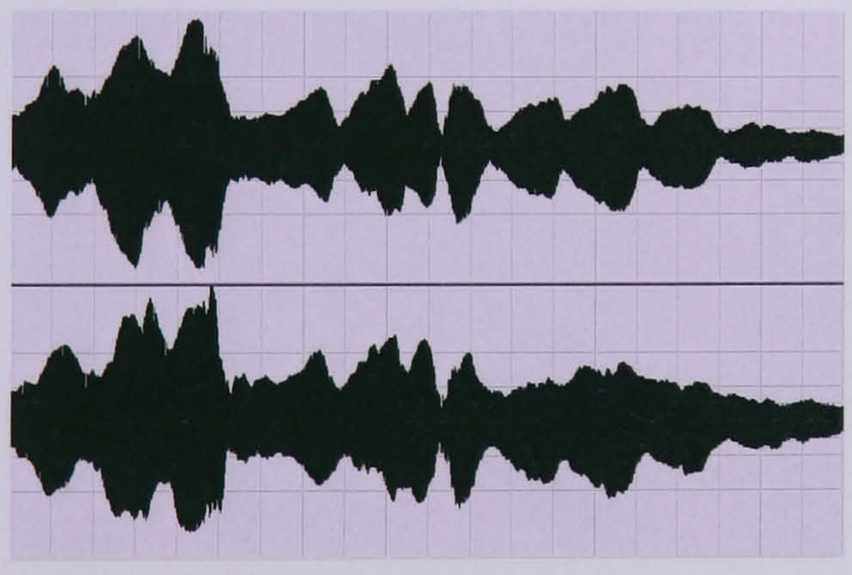
Appendix B Plots from objective measurements of the effects of interchannel crosstalk



(a) Crosstalk-off

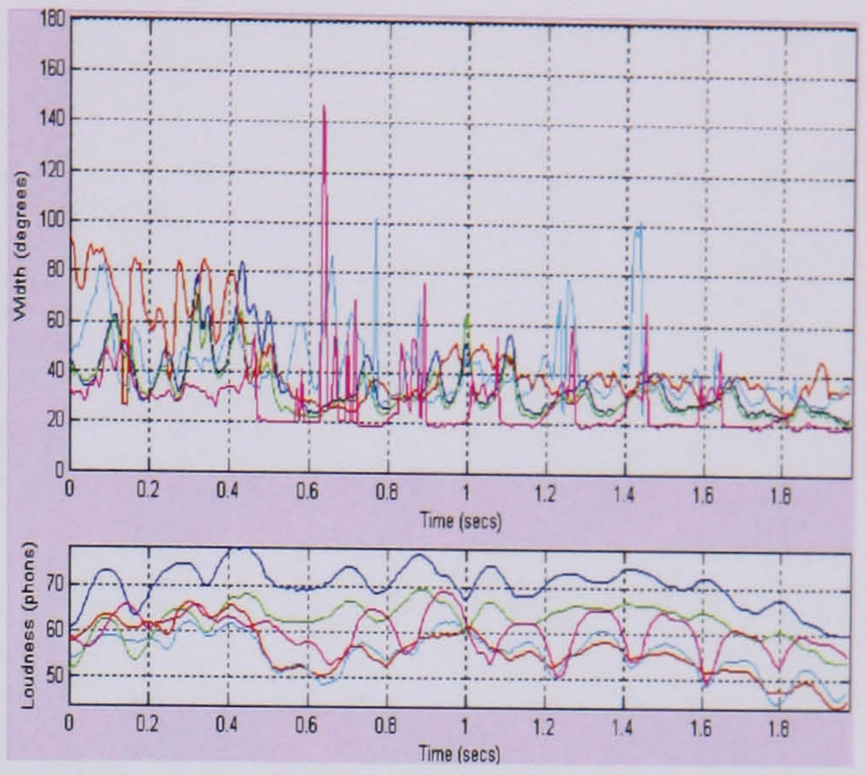


(b) Crosstalk-off

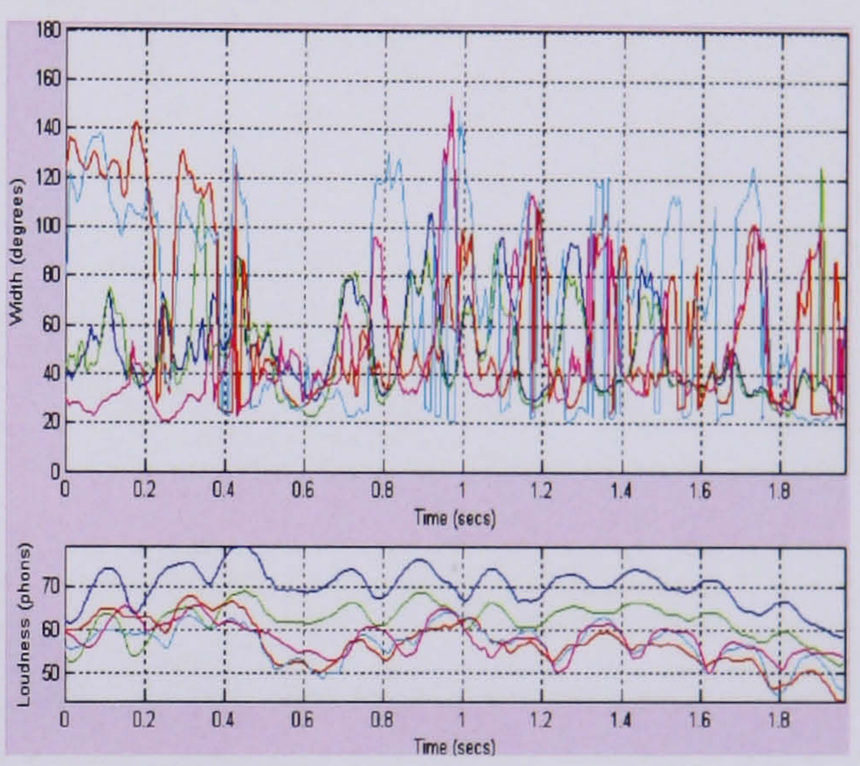


(c) Waveform

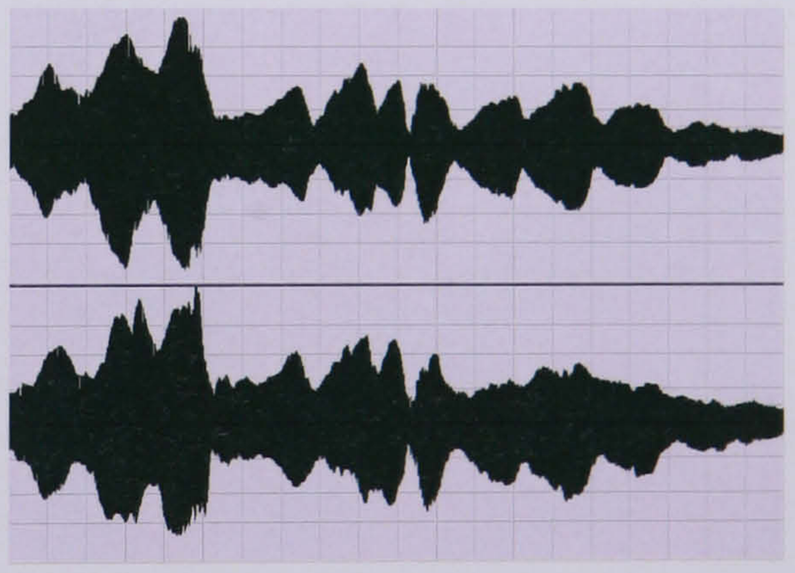
Figure B.28 Plots of the width measurement made for the anechoic cello stimuli of microphone array 4, with *fc* = 150, 250, 350, 455, 570 Hz, and waveform of the binaural signal



(a) Crosstalk-off

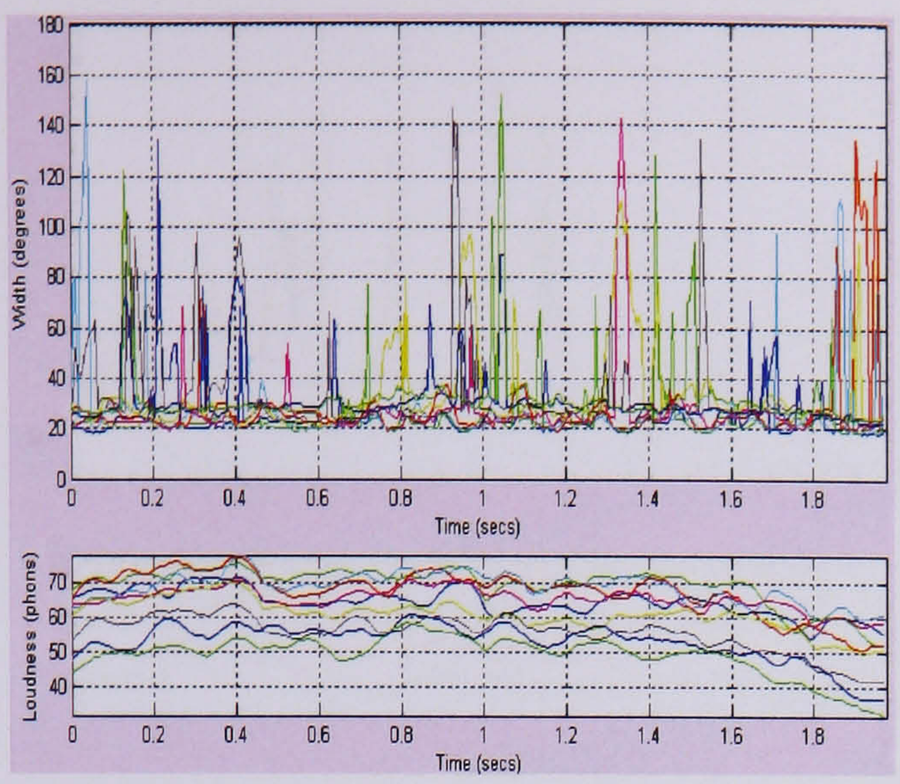


(b) Crosstalk-on

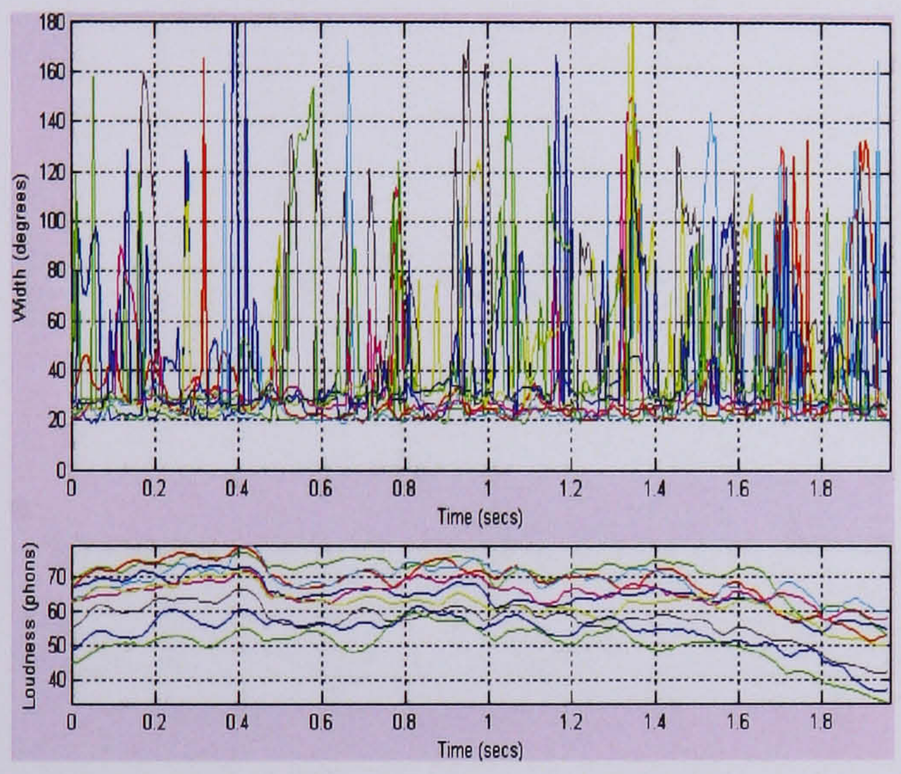


(c) Waveform

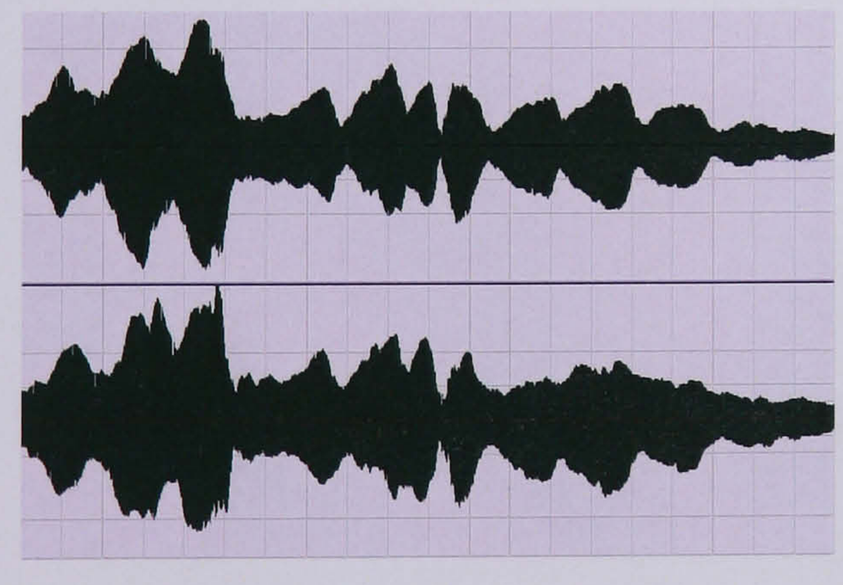
Figure B.29 Plots of the width measurement made for the anechoic cello stimuli of microphone array 4, with *fc* = 700, 845, 1000, 1175, 1375 Hz, and waveform of the binaural signal



(a) Crosstalk-off

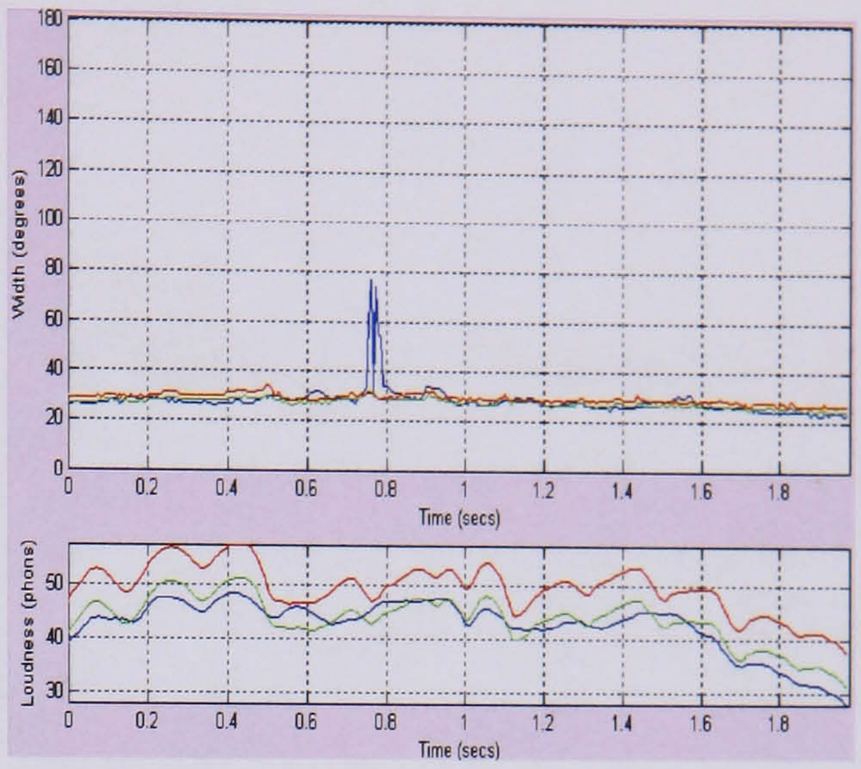


(b) Crosstalk-on

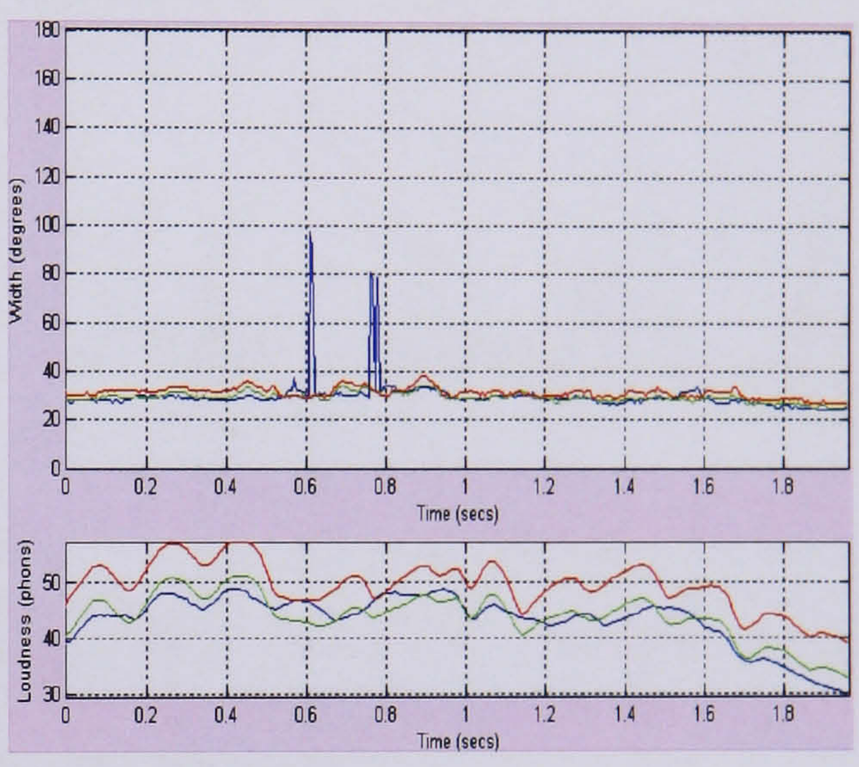


(c) Waveform

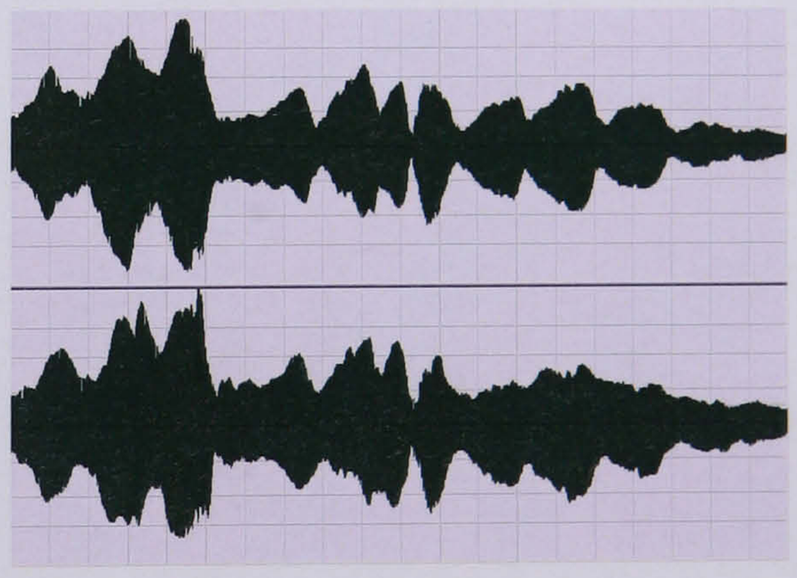
Figure B.30 Plots of the width measurement made for the anechoic cello stimuli of microphone array 4, with *fc* = 1600, 1860, 2160, 2570, 2925, 3425, 4050, 4850, 5850 Hz, and waveform of the binaural signal



(a) Crosstalk-off

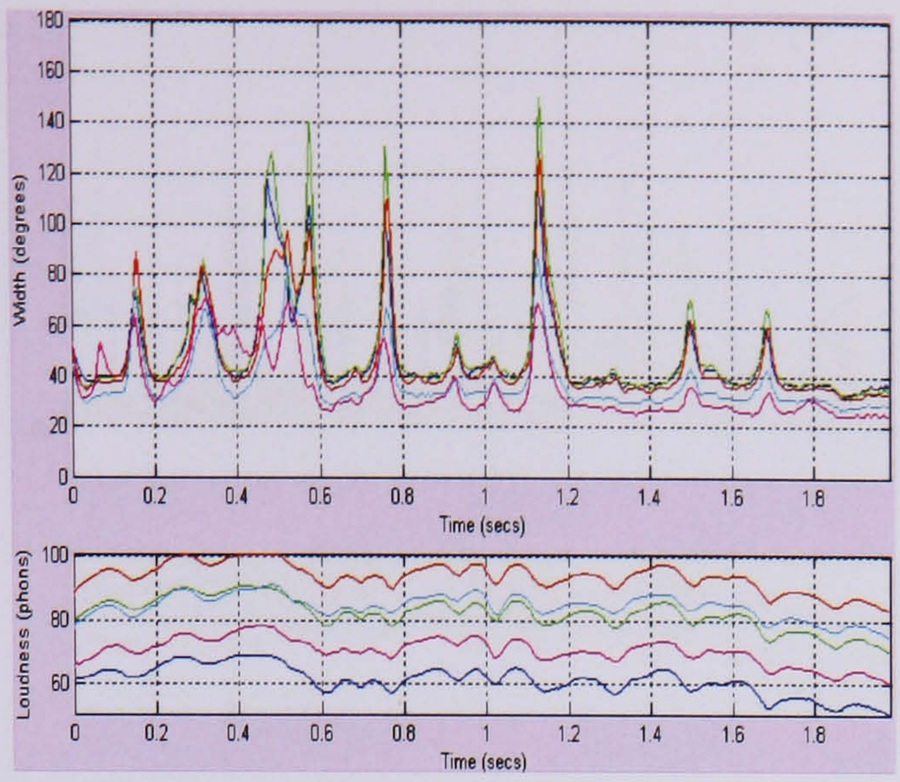


(b) Crosstalk-on

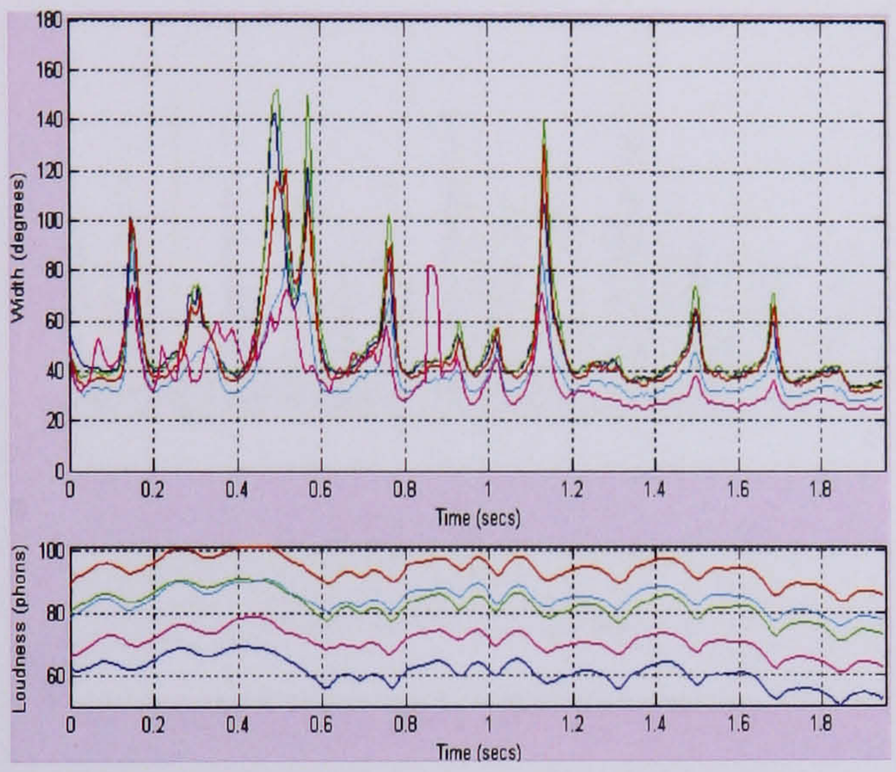


(c) Waveform

Figure B.31 Plots of the width measurement made for the anechoic cello stimuli of microphone array 4, with *fc* = 7050, 8600, 10750 Hz, and waveform of the binaural signal



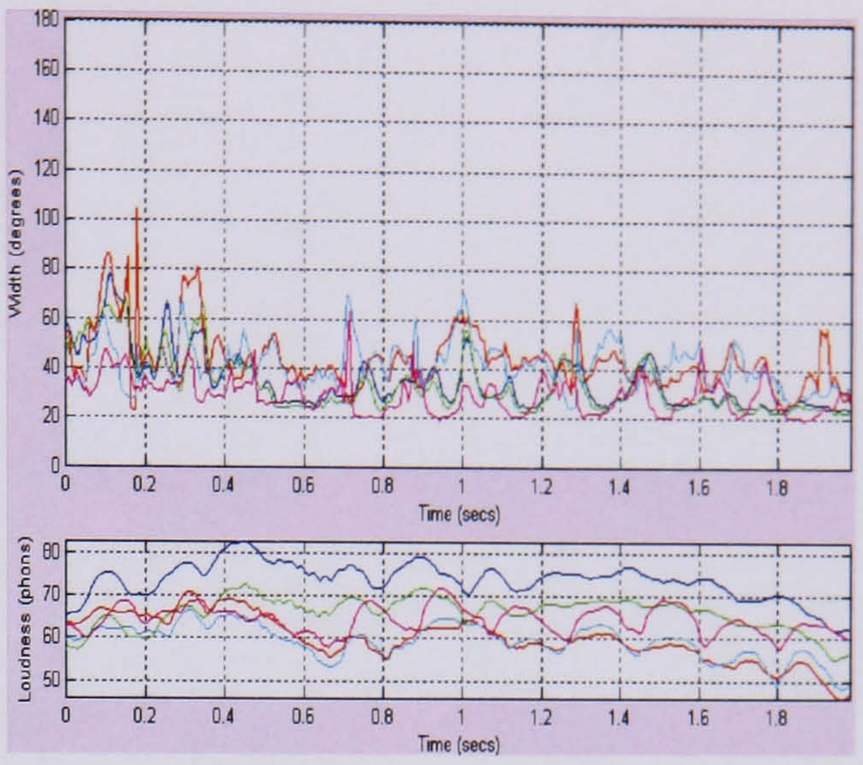
(a) Crosstalk-off



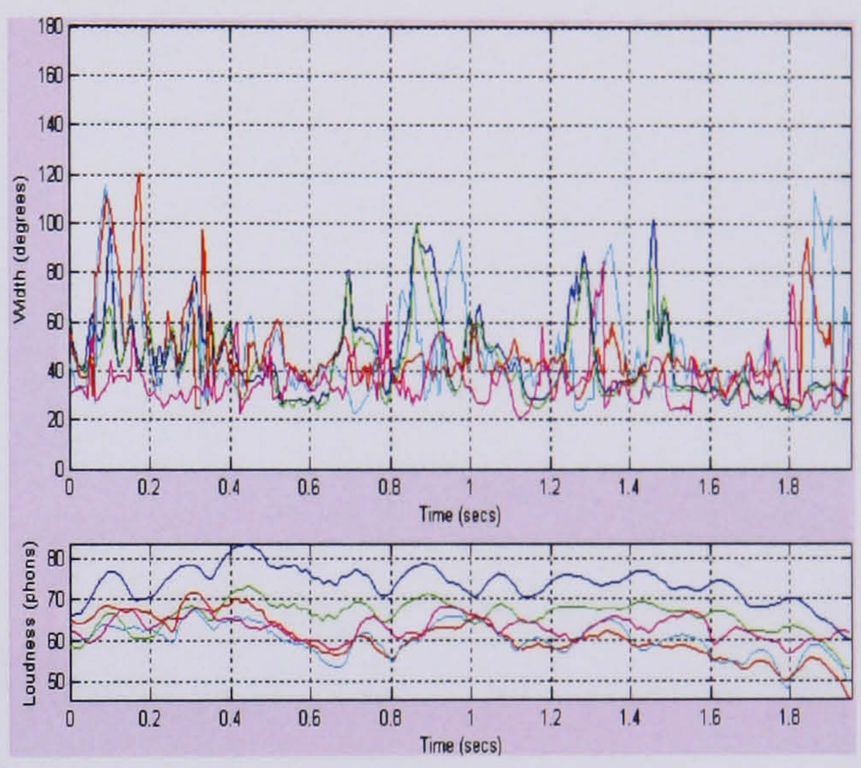
(b) Crosstalk-on

(c) Waveform

Figure B.32 Plots of the width measurement made for the room-reverberant cello stimuli of microphone array 4, with *fc* = 150, 250, 350, 455, 570 Hz, and waveform of the binaural signal



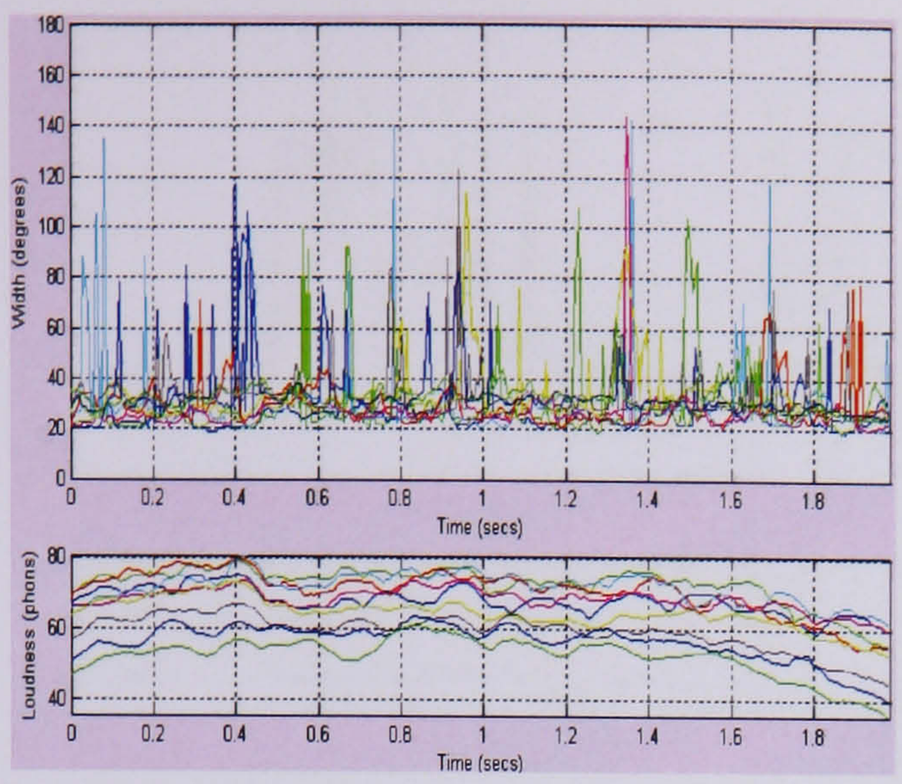
(a) Crosstalk-off



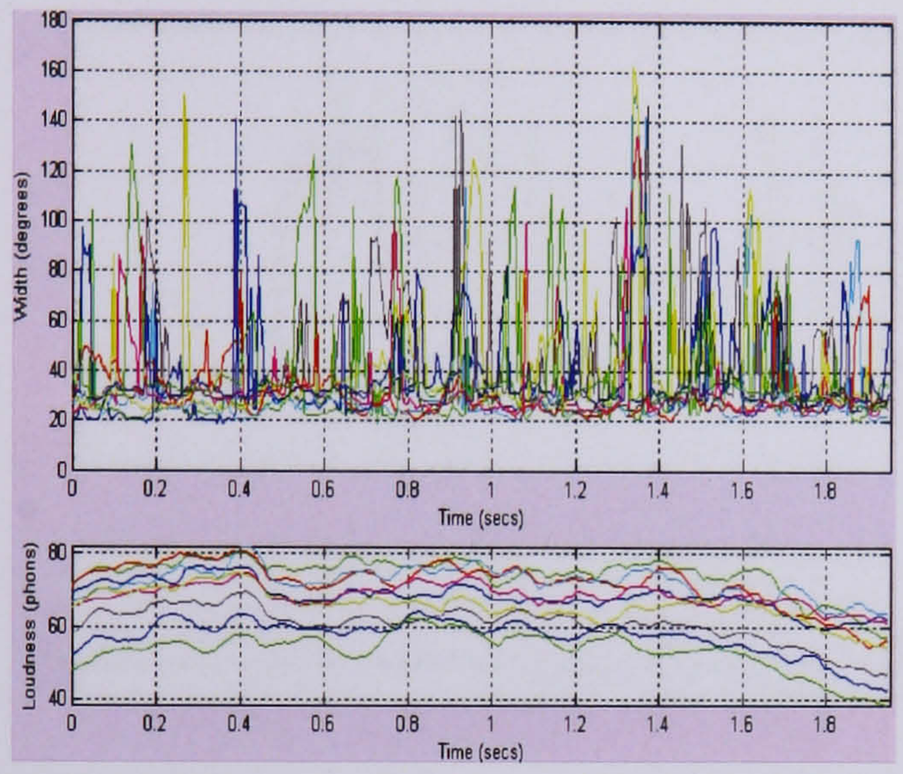
(b) Crosstalk-on

(c) Waveform

Figure B.33 Plots of the width measurement made for the room-reverberant cello stimuli of microphone array 4, with *fc* = 700, 845, 1000, 1175, 1375 Hz, and waveform of the binaural signal



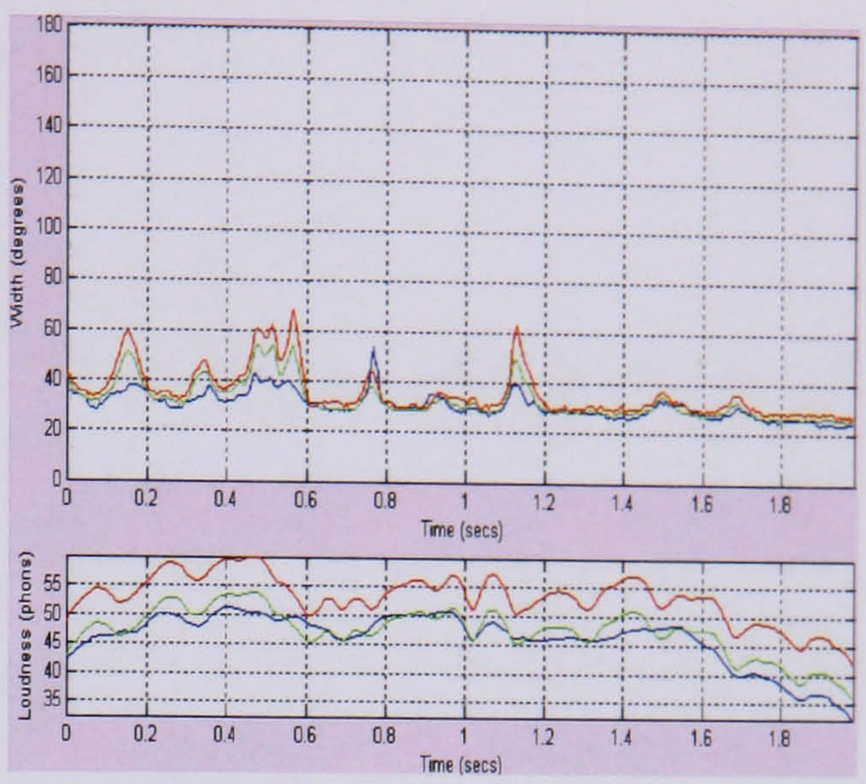
(a) Crosstalk-off



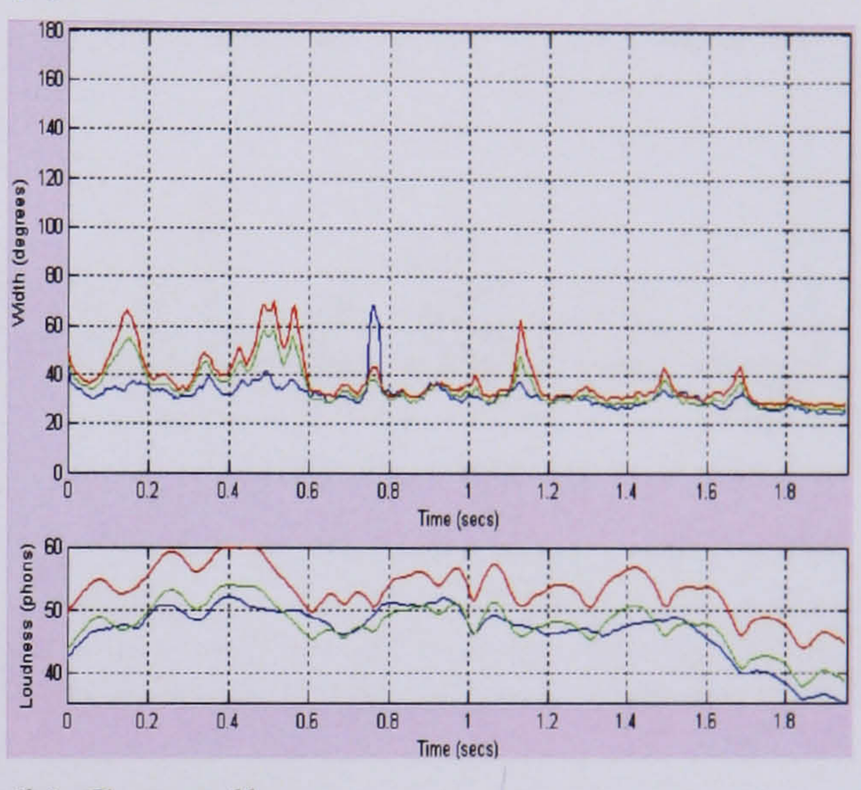
(b) Crosstalk-on

(c) Waveform

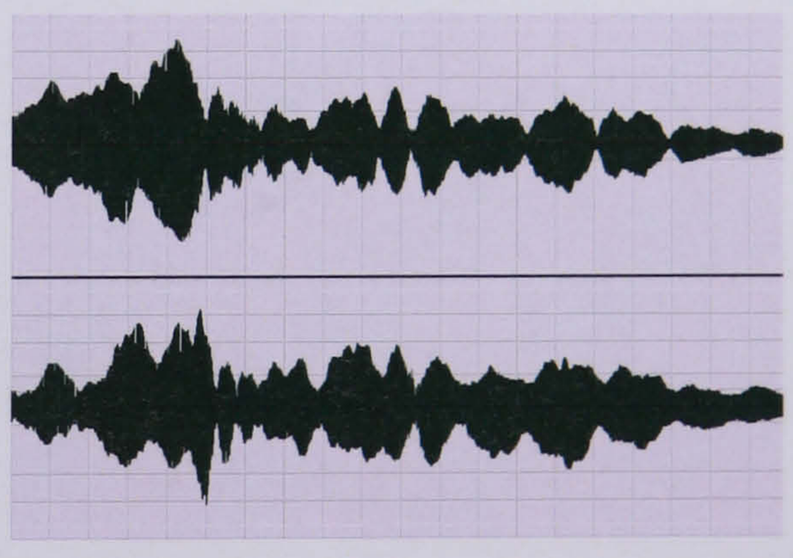
Figure B.34 Plots of the width measurement made for the room-reverberant cello stimuli of microphone array 4, with *fc* = 1600, 1860, 2160, 2570, 2925, 3425, 4050, 4850, 5850 Hz, and waveform of the binaural signal



(a) Crosstalk-off

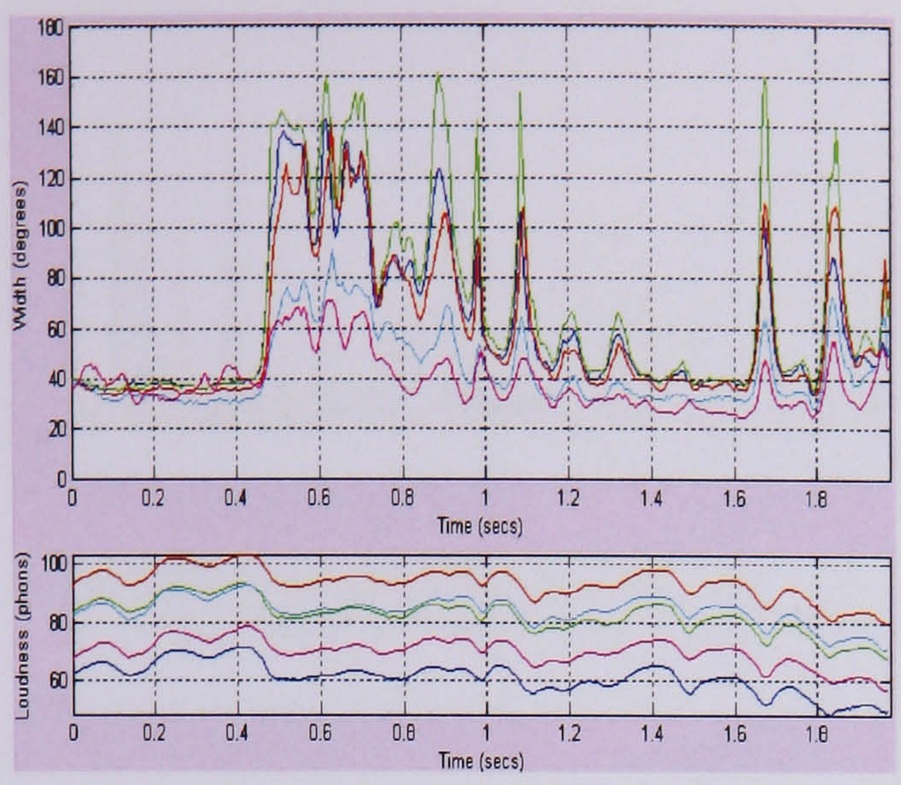


(b) Crosstalk-on

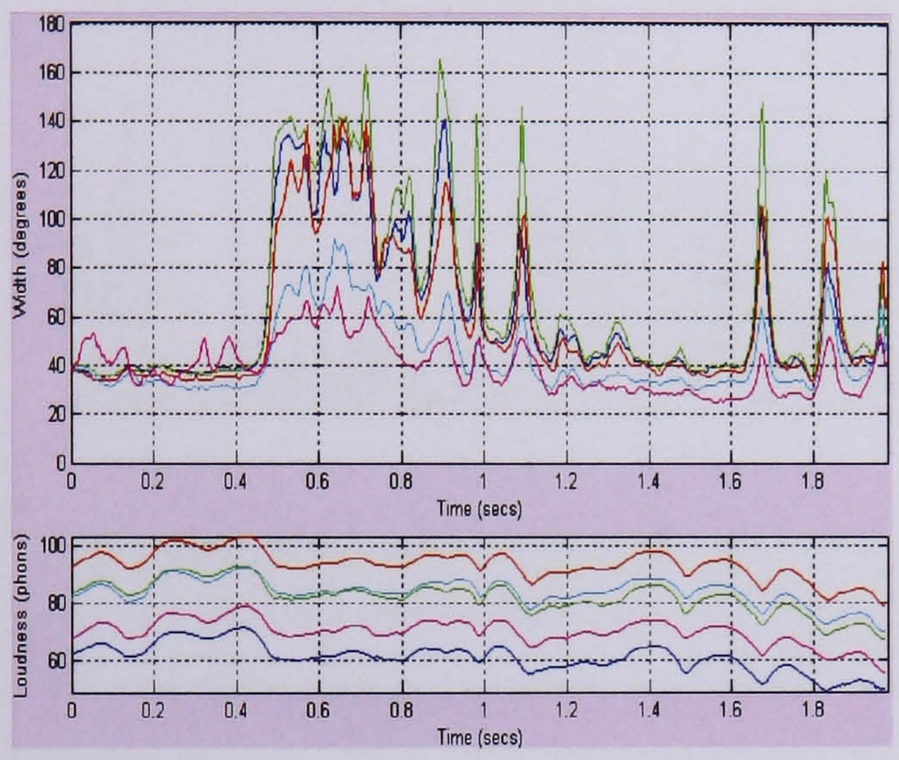


(c) Waveform

Figure B.35 Plots of the width measurement made for the room-reverberant cello stimuli of microphone array 4, with *fc* = 7050, 8600, 10750 Hz, and waveform of the binaural signal



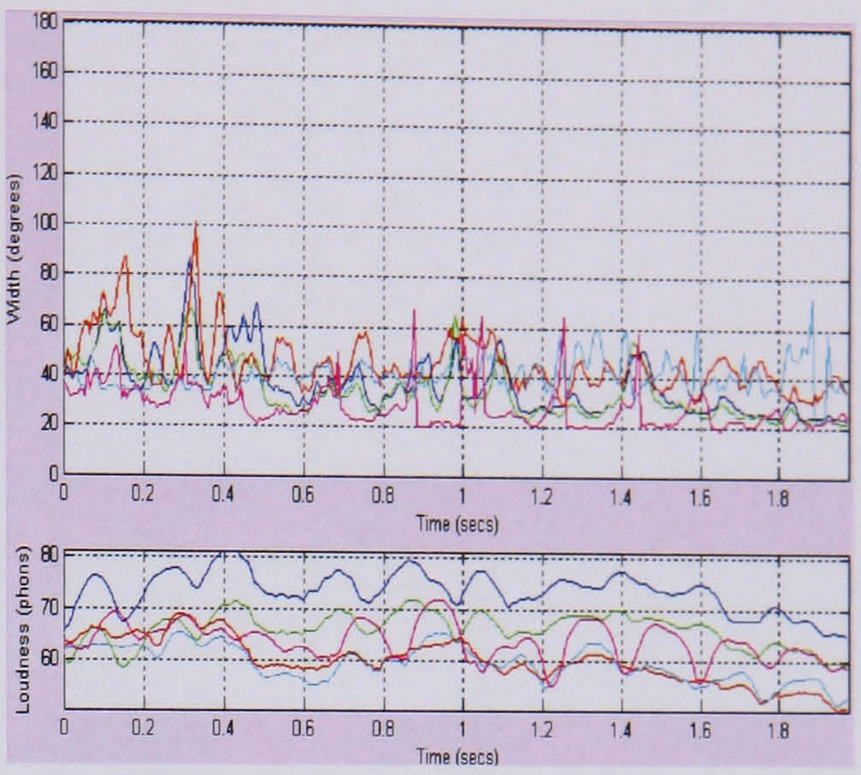
(a) Crosstalk-off



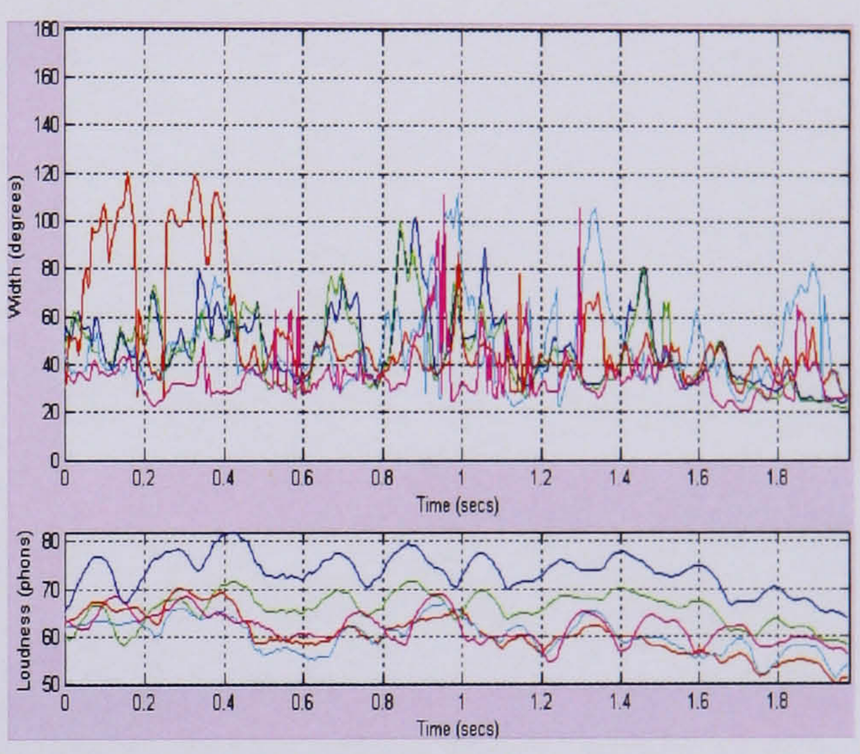
(b) Crosstalk-on

(c) Waveform

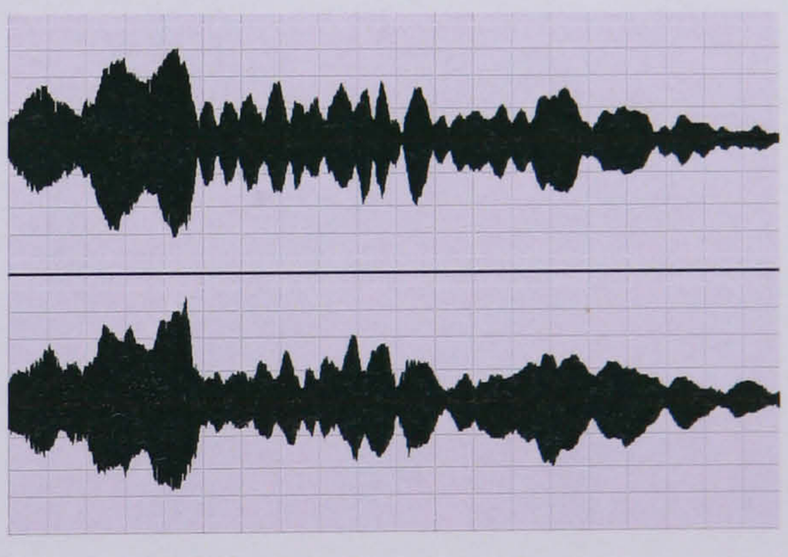
Figure B.36 Plots of the width measurement made for the hall-reverberant cello stimuli of microphone array 4, with *fc* = 150, 250, 350, 455, 570 Hz, and waveform of the binaural signal



(a) Crosstalk-off

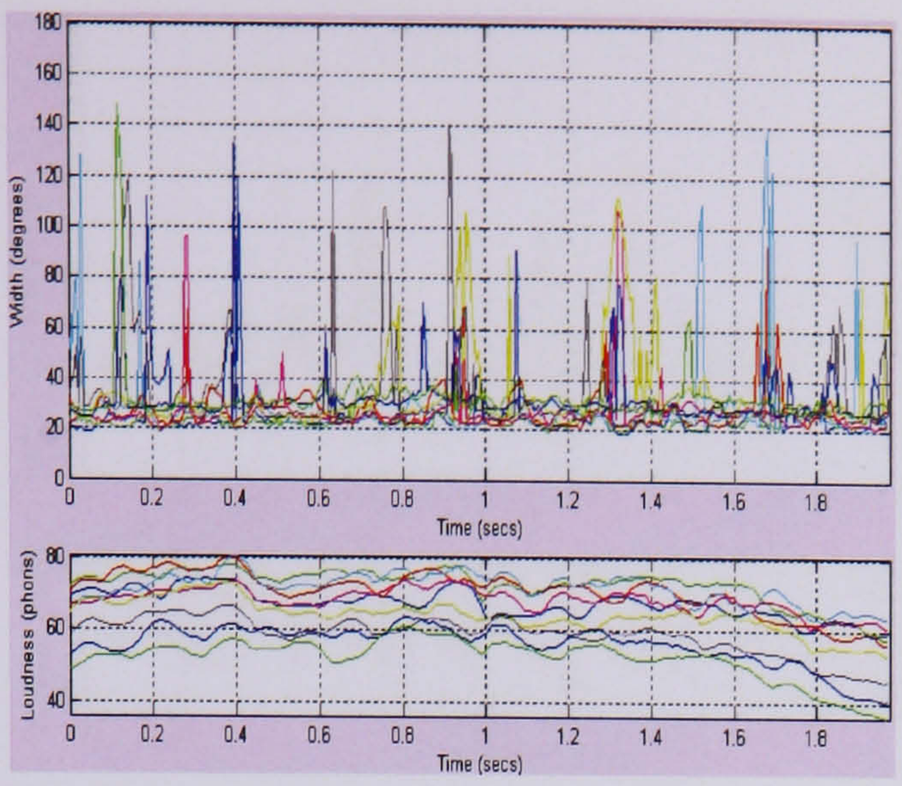


(b) Crosstalk-on

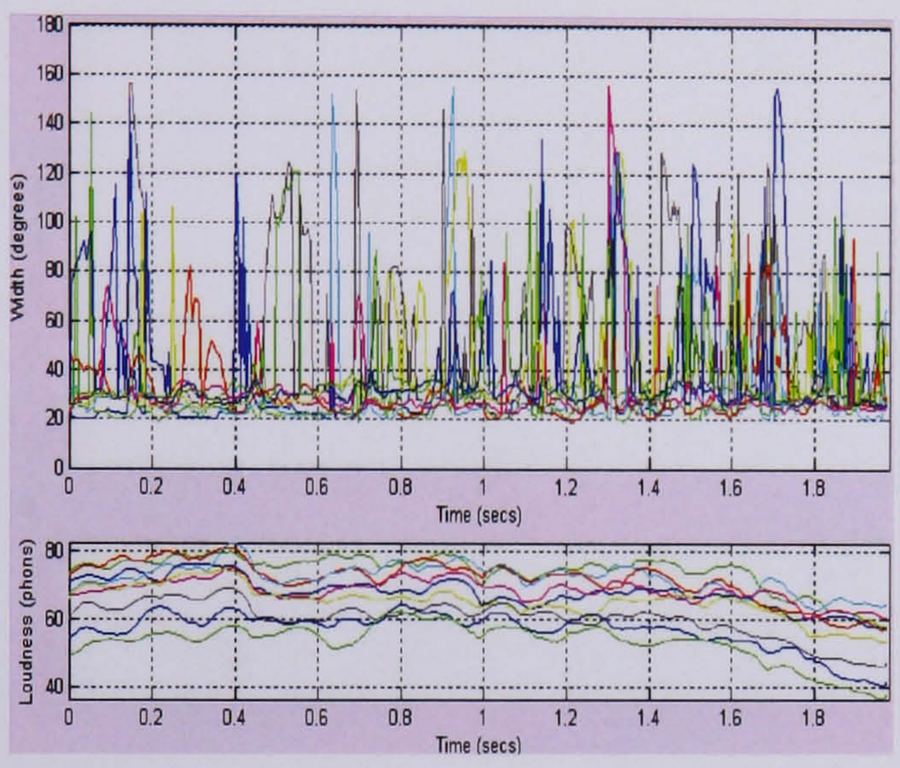


(c) Waveform

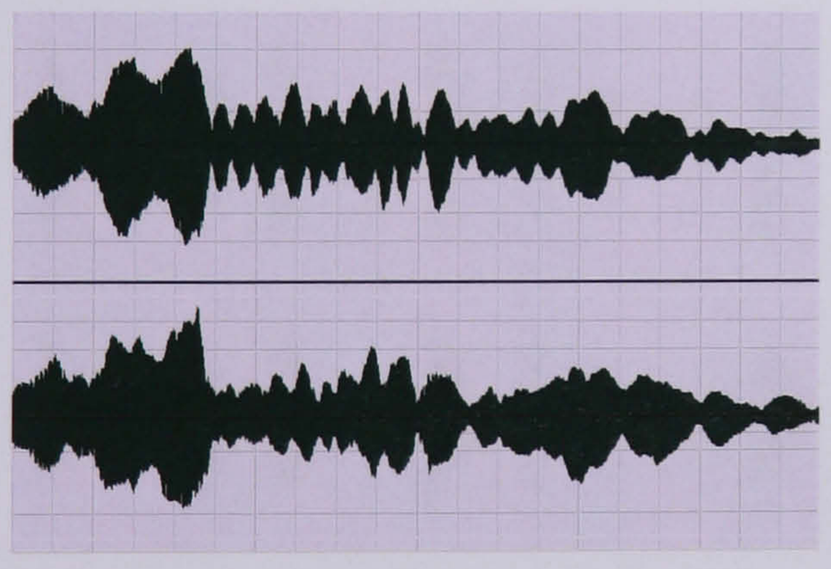
Figure B.37 Plots of the width measurement made for the hall-reverberant cello stimuli of microphone array 4, with *fc* = 700, 845, 1000, 1175, 1375 Hz, and waveform of the binaural signal



(a) Crosstalk-off

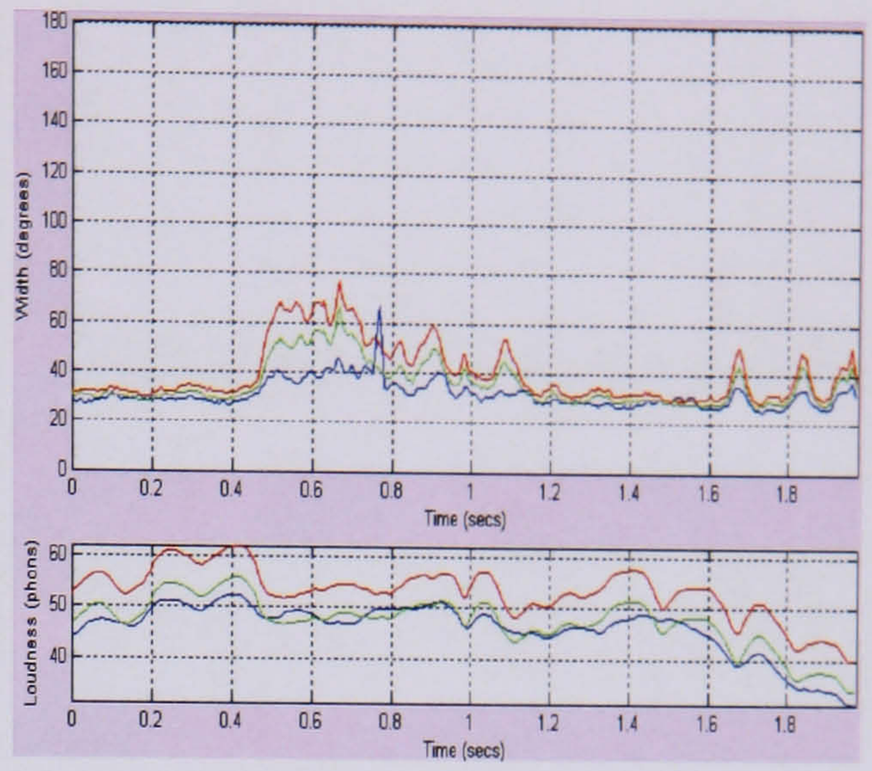


(b) Crosstalk-on

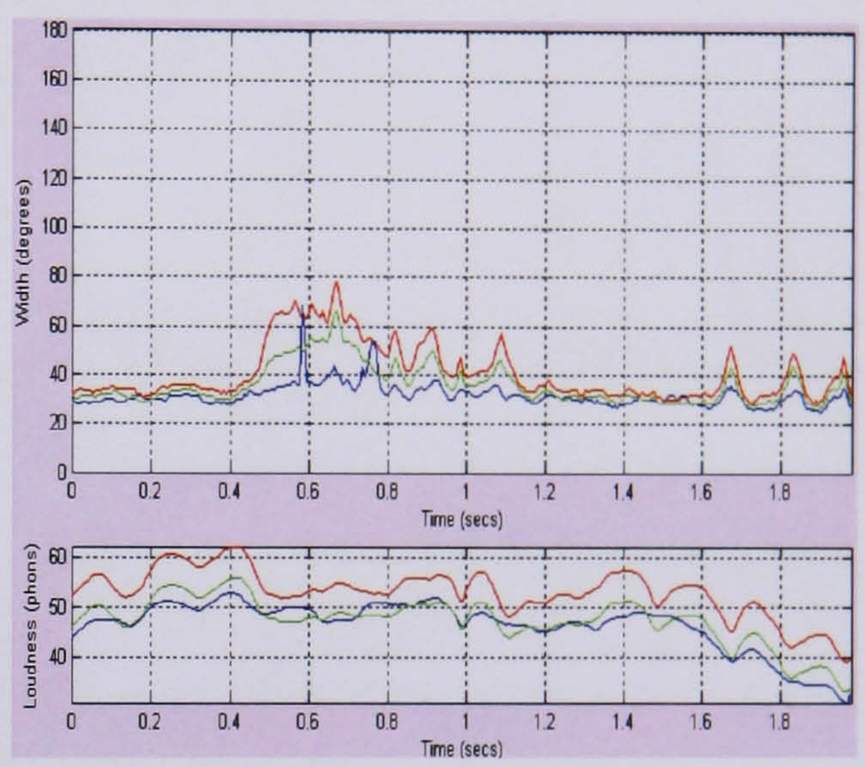


(c) Waveform

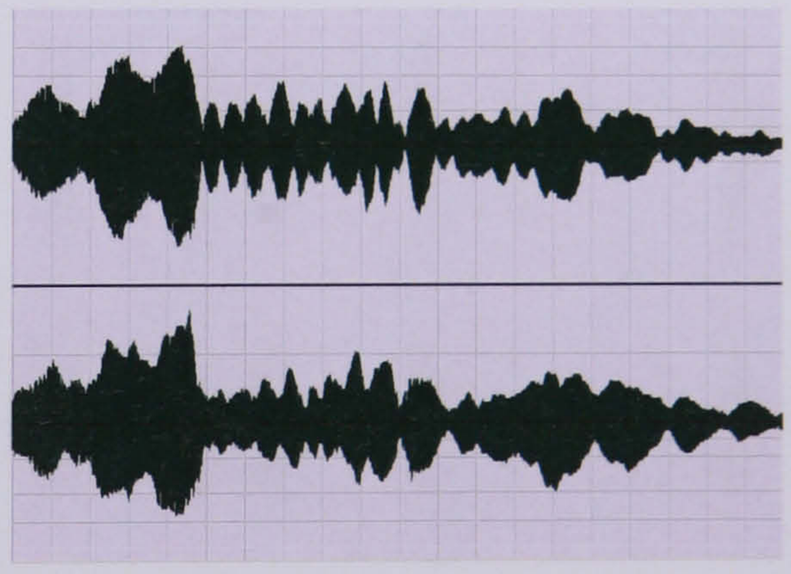
Figure B.38 Plots of the width measurement made for the hall-reverberant cello stimuli of microphone array 4, with *fc* = 1600, 1860, 2160, 2570, 2925, 3425, 4050, 4850, 5850 Hz, and waveform of the binaural signal



(a) Crosstalk-off

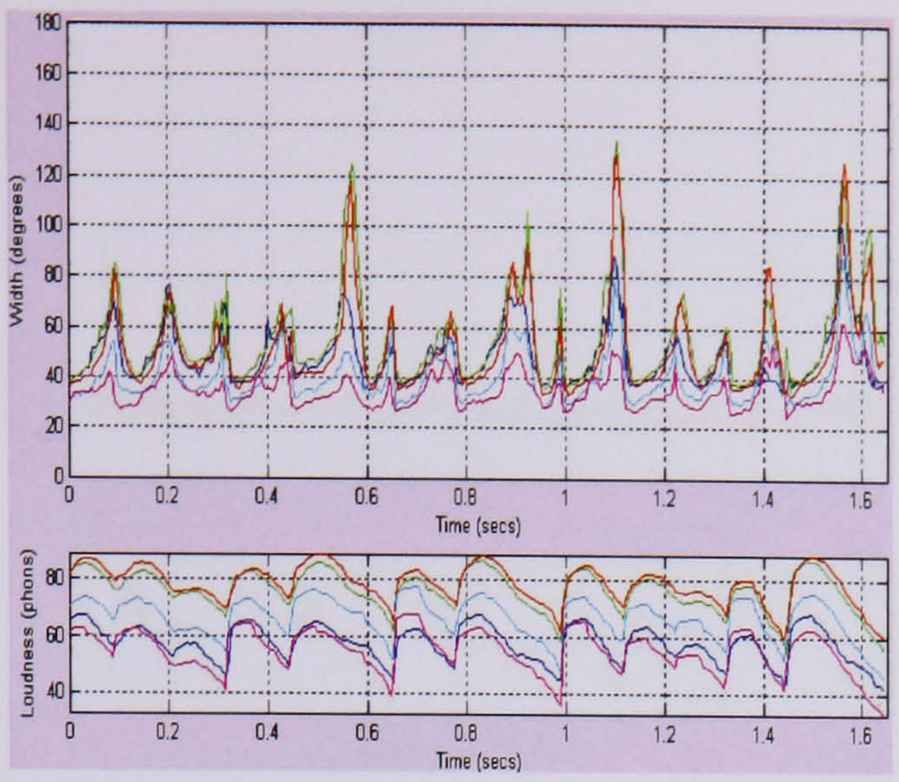


(b) Crosstalk-on

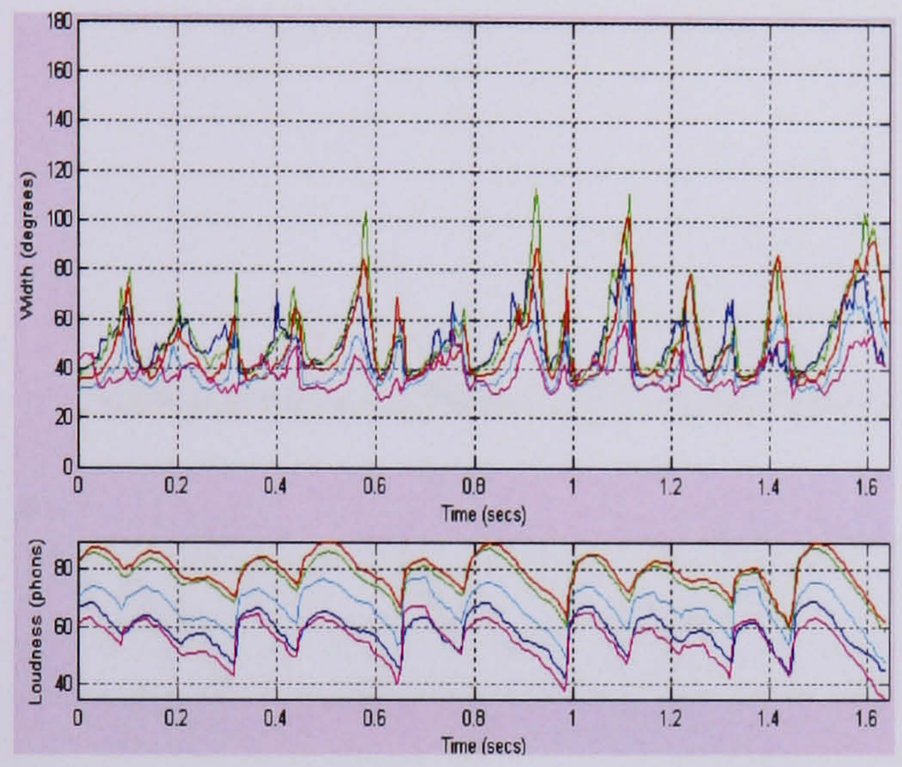


(c) Waveform

Figure B.39 Plots of the width measurement made for the hall-reverberant cello stimuli of microphone array 4, with *fc* = 7050, 8600, 10750 Hz, and waveform of the binaural signal



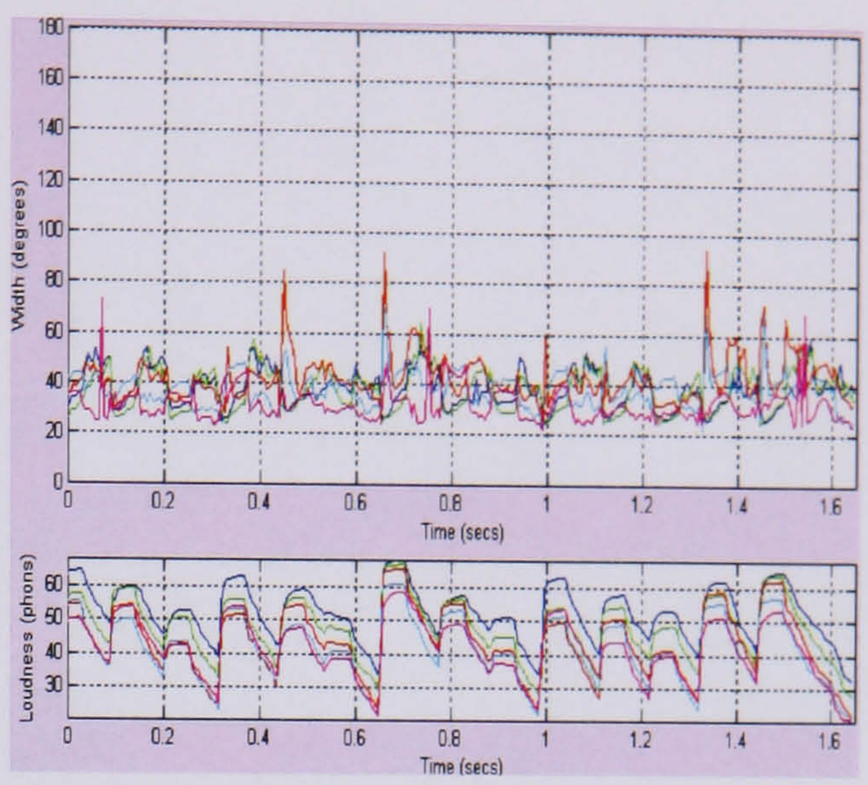
(a) Crosstalk-off



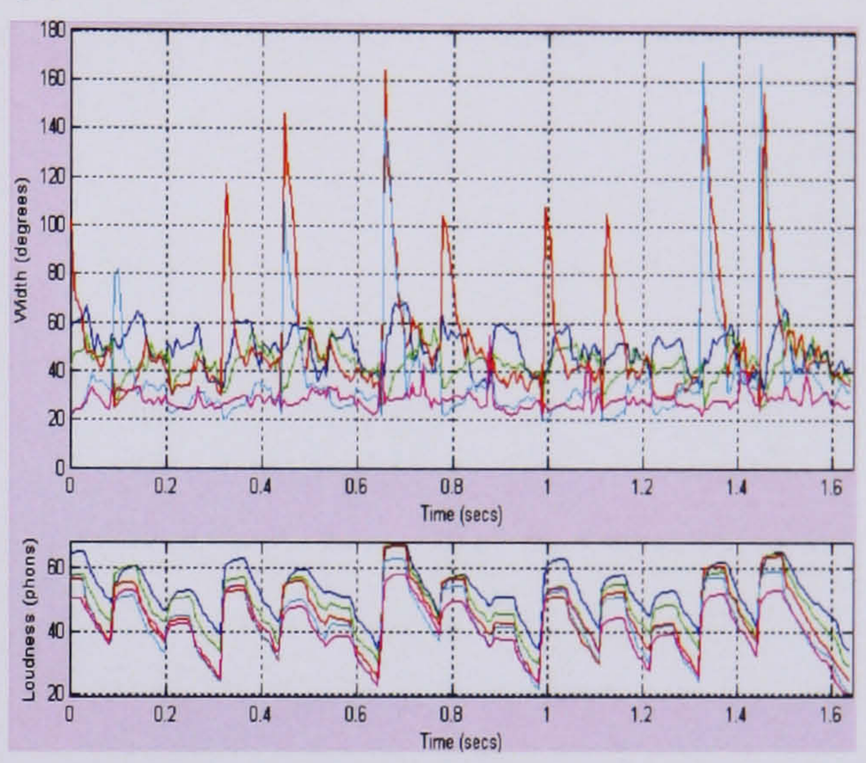
(b) Crosstalk-on

(c) Waveform

Figure B.40 Plots of the width measurement made for the anechoic bongo stimuli of microphone array 4, with *fc* = 150, 250, 350, 455, 570 Hz, and waveform of the binaural signal



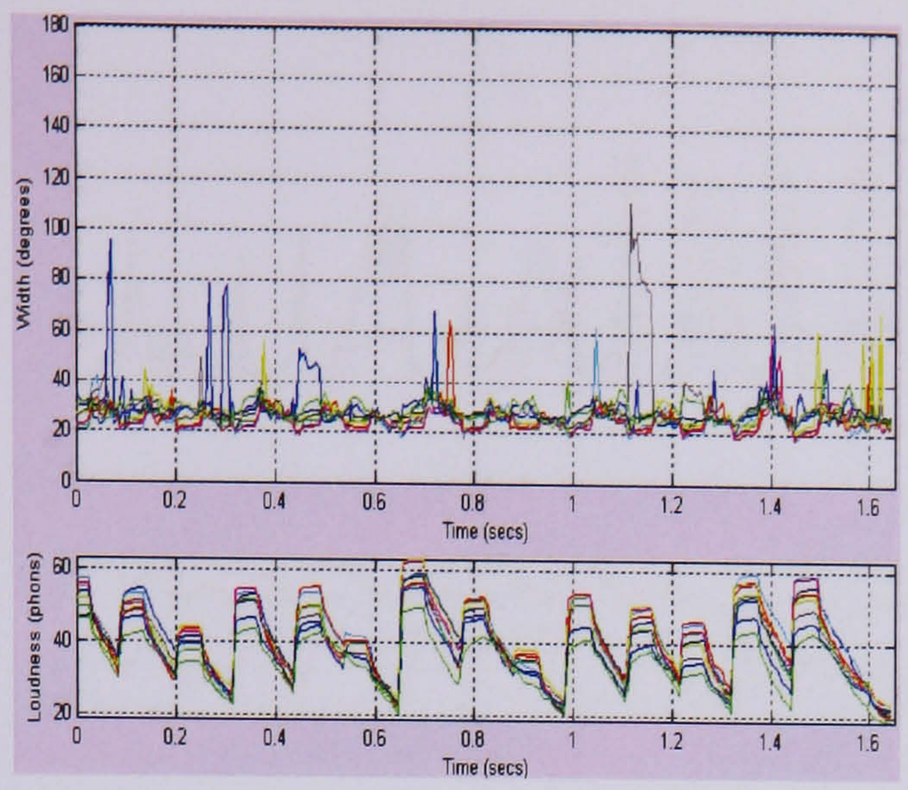
(a) Crosstalk-off



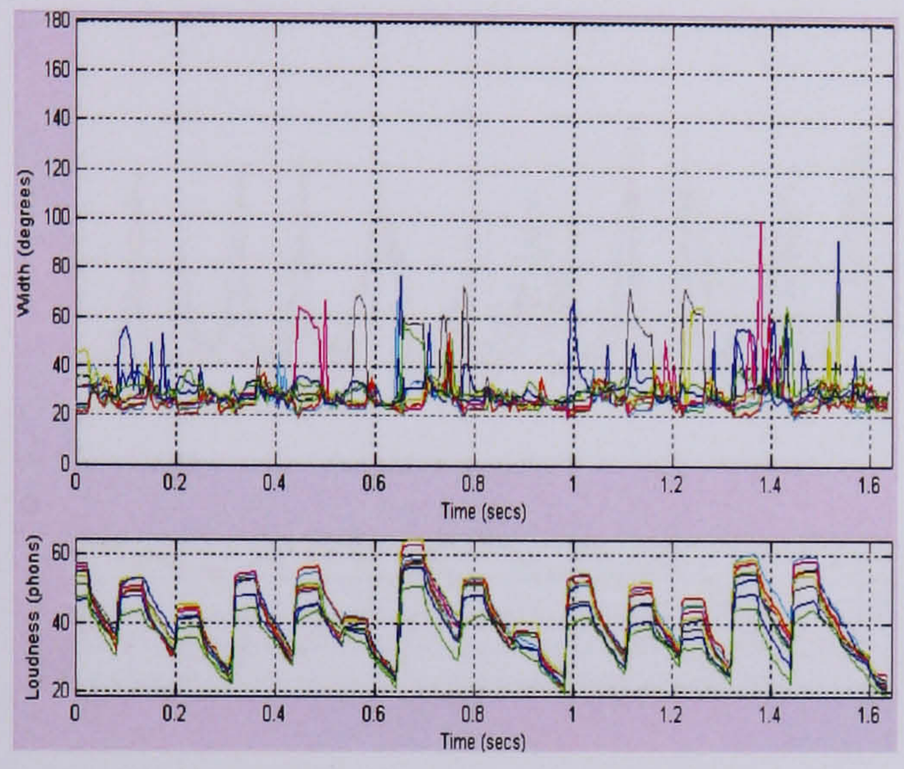
(b) Crosstalk-on

(c) Waveform

Figure B.41 Plots of the width measurement made for the anechoic bongo stimuli of microphone array 4, with *fc* = 700, 845, 1000, 1175, 1375 Hz, and waveform of the binaural signal



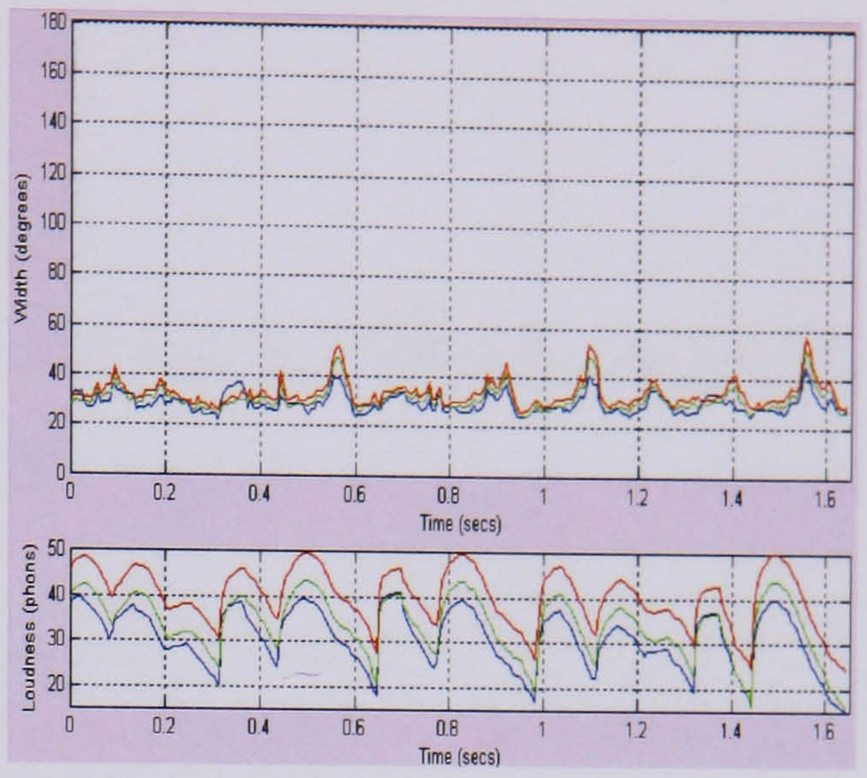
(a) Crosstalk-off



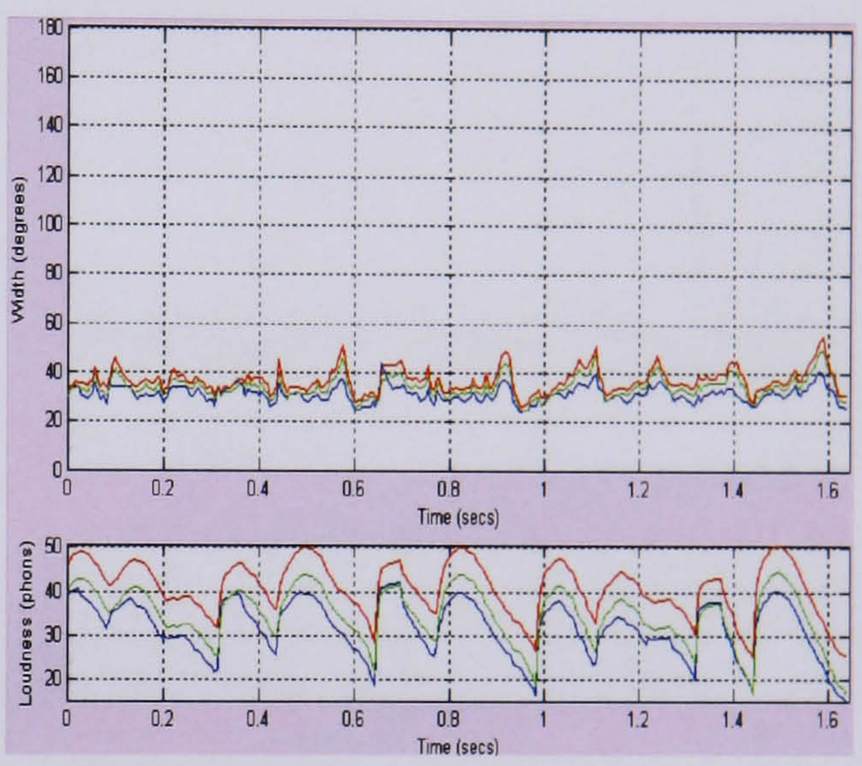
(b) Crosstalk-on

(c) Waveform

Figure B.42 Plots of the width measurement made for the anechoic bongo stimuli of microphone array 4, with *fc* = 1600, 1860, 2160, 2570, 2925, 3425, 4050, 4850, 5850 Hz, and waveform of the binaural signal



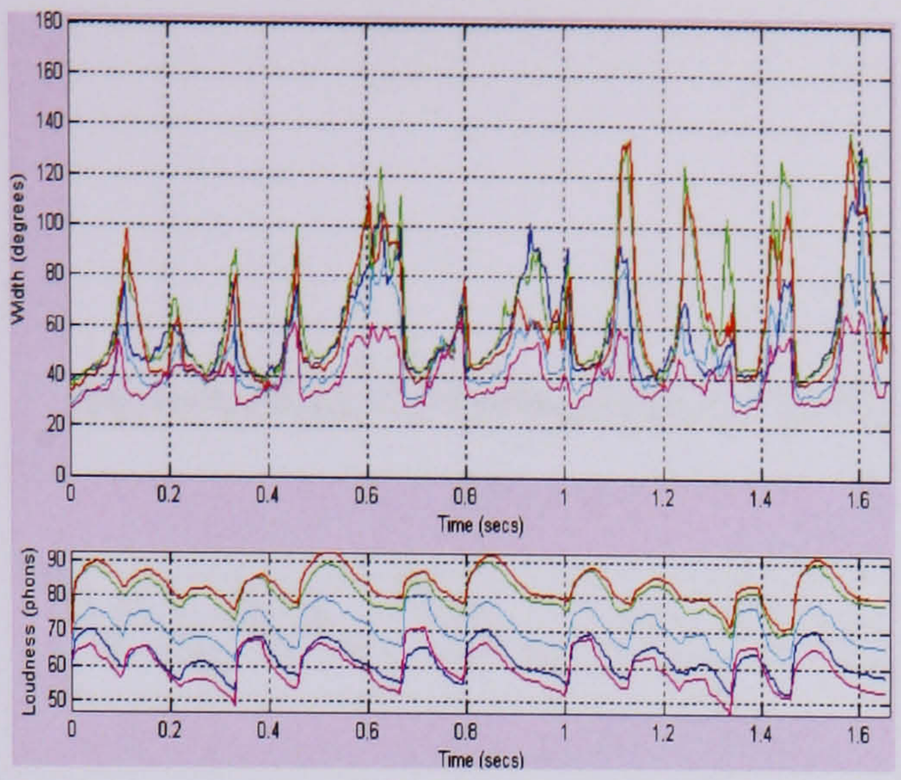
(a) Crosstalk-off



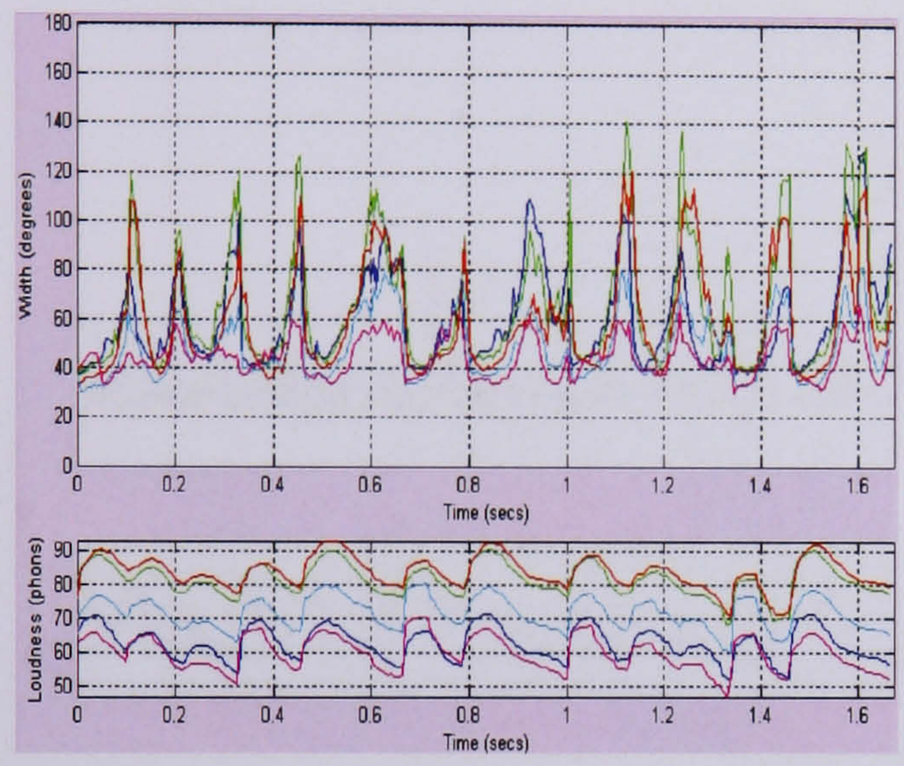
(b) Crosstalk-on

(c) Waveform

Figure B.43 Plots of the width measurement made for the anechoic bongo stimuli of microphone array 4, with *fc* = 7050, 8600, 10750 Hz, and waveform of the binaural signal



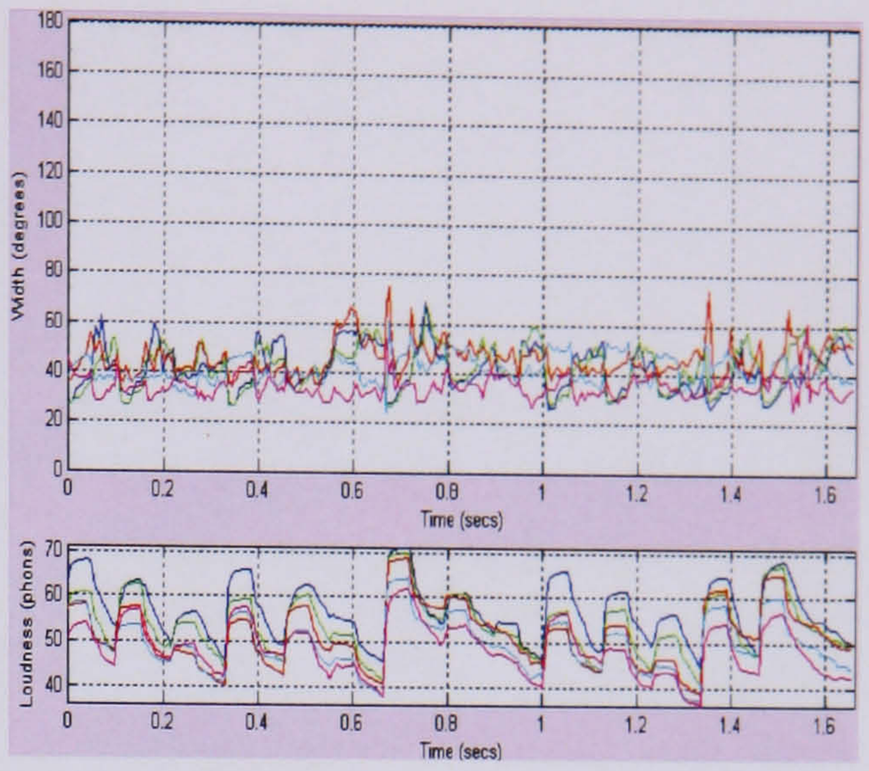
(a) Crosstalk-off



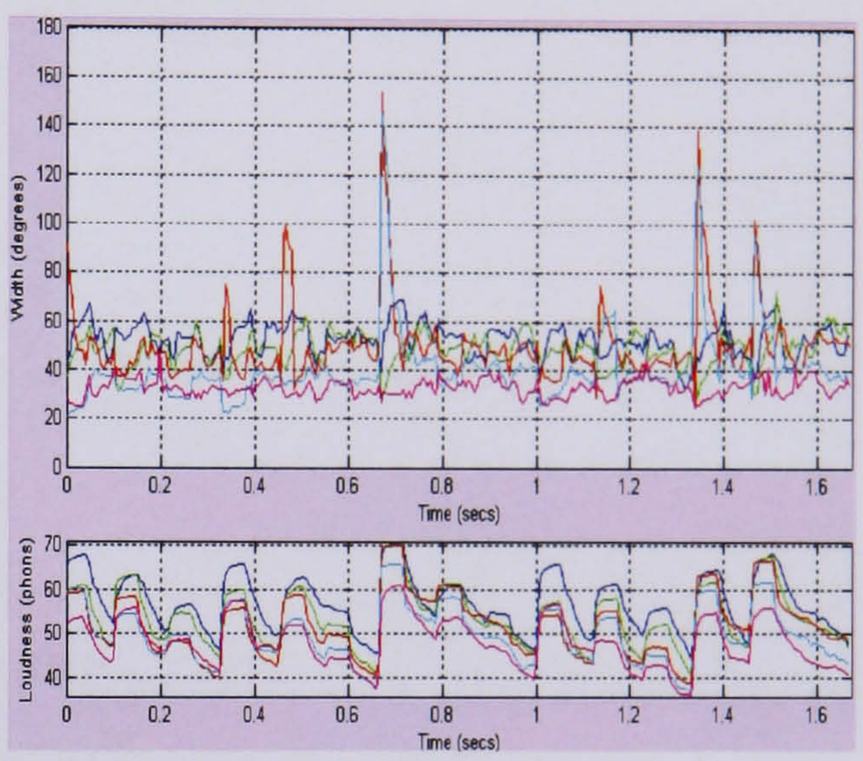
(b) Crosstalk-on

(c) Waveform

Figure B.44 Plots of the width measurement made for the room-reverberant bongo stimuli of microphone array 4, with *fc* = 150, 250, 350, 455, 570 Hz, and waveform of the binaural signal



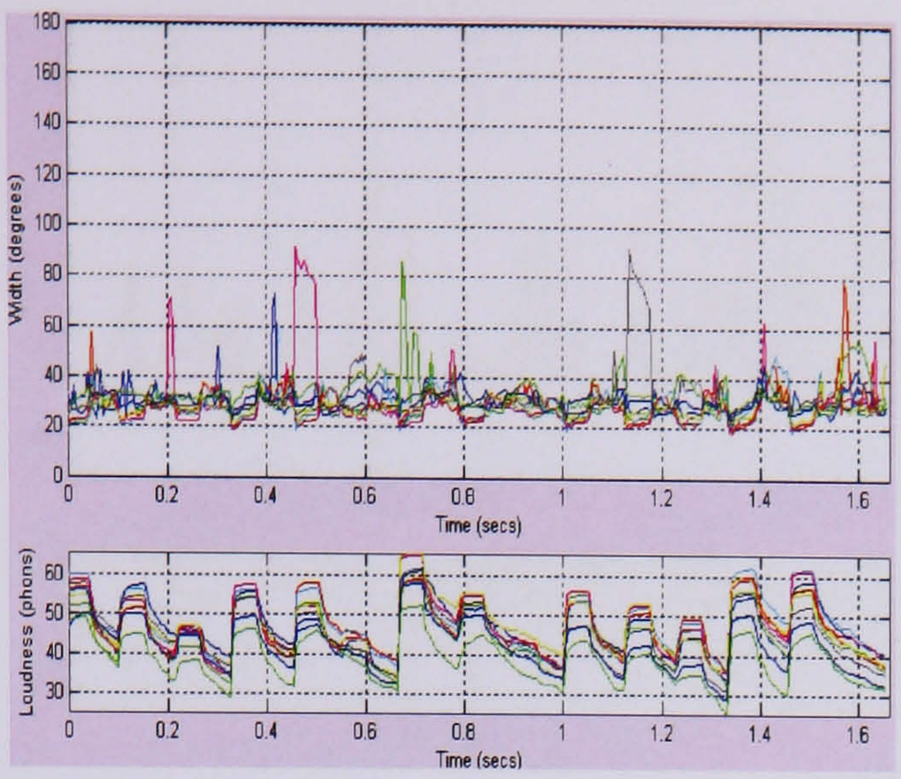
(a) Crosstalk-off



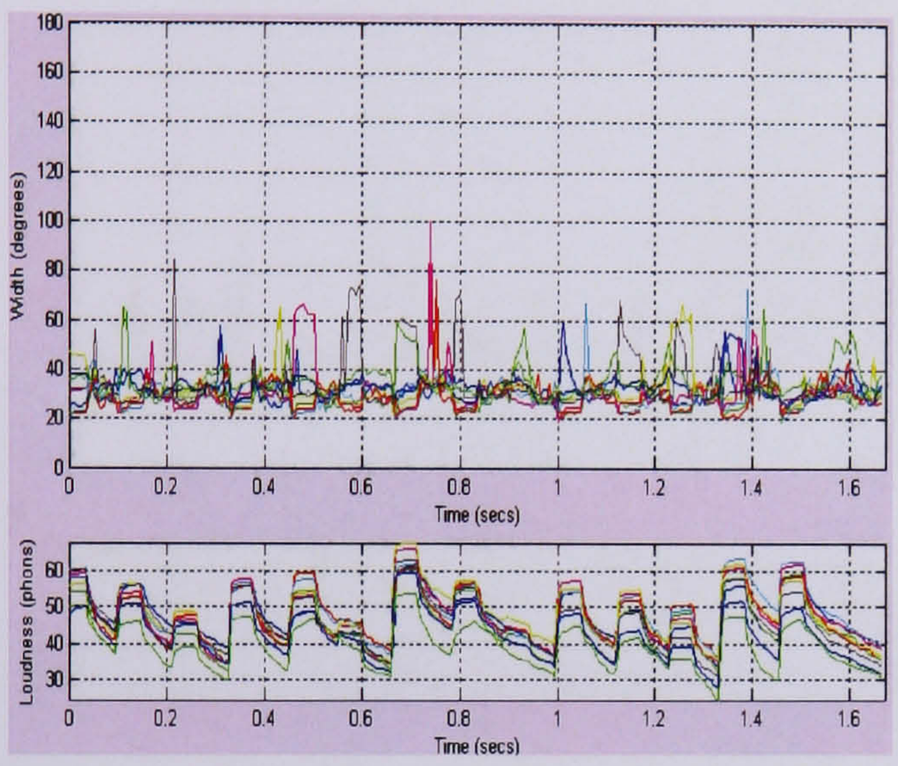
(b) Crosstalk-on

(c) Waveform

Figure B.45 Plots of the width measurement made for the room-reverberant bongo stimuli of microphone array 4, with *fc* = 700, 845, 1000, 1175, 1375 Hz, and waveform of the binaural signal



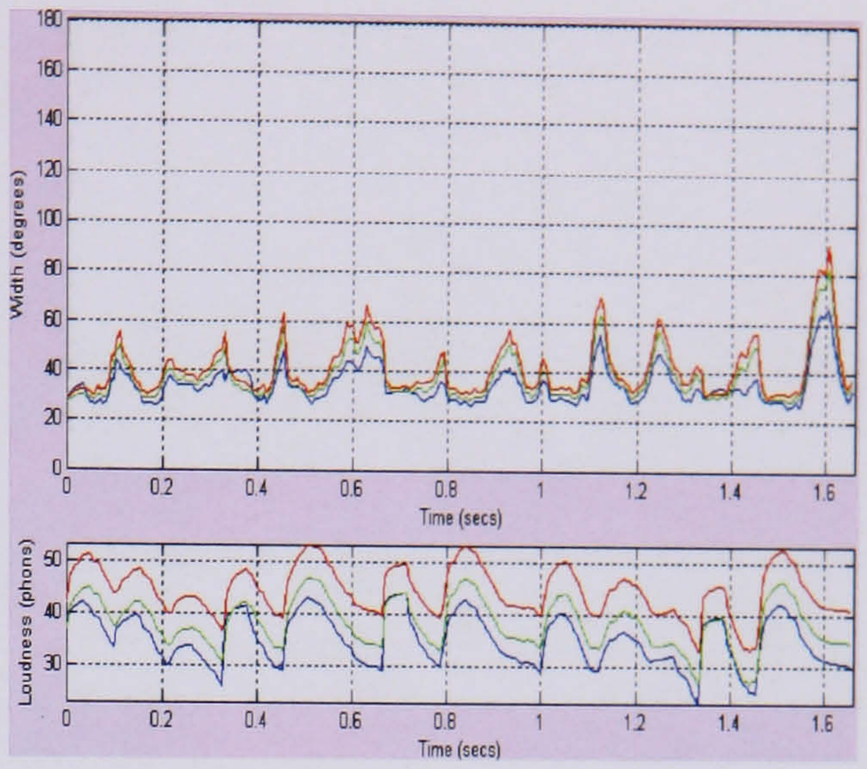
(a) Crosstalk-off



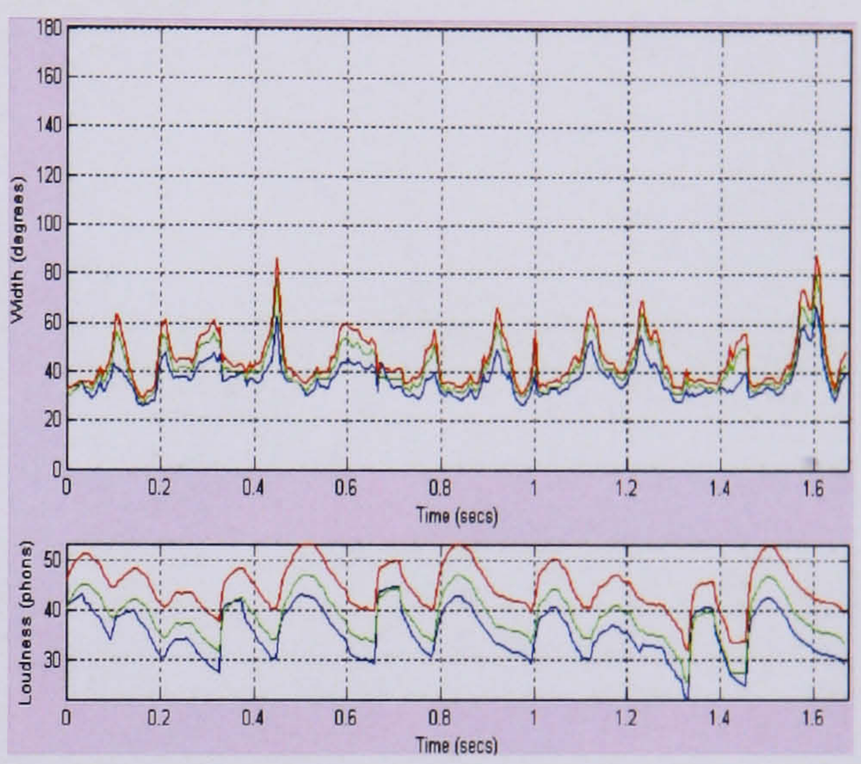
(b) Crosstalk-on

(c) Waveform

Figure B.46 Plots of the width measurement made for the room-reverberant bongo stimuli of microphone array 4, with *fc* = 1600, 1860, 2160, 2570, 2925, 3425, 4050, 4850, 5850 Hz, and waveform of the binaural signal



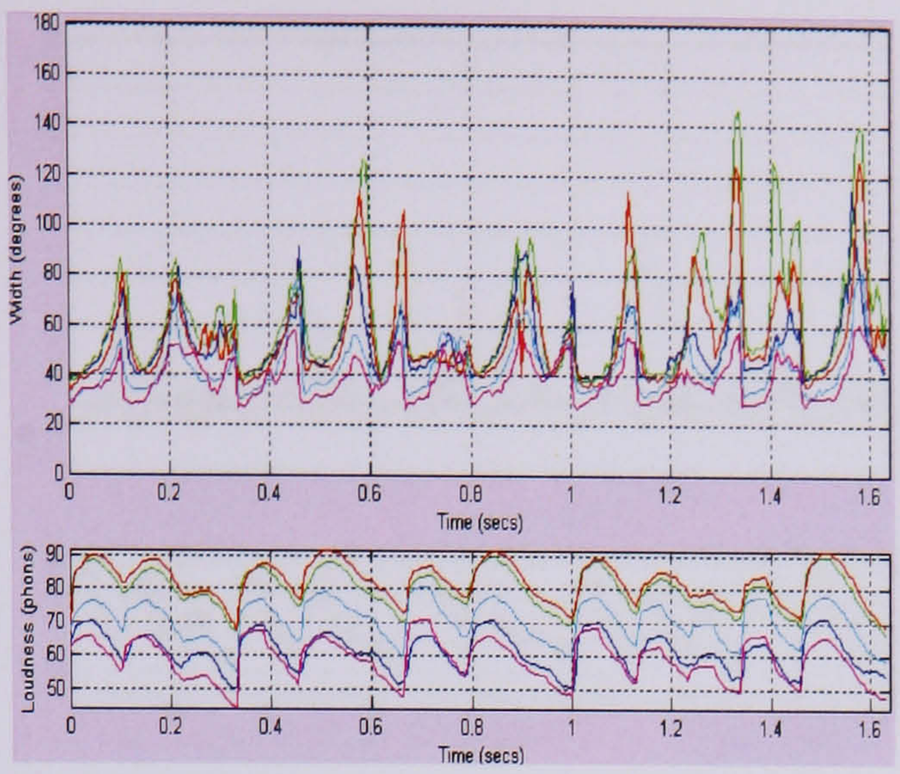
(a) Crosstalk-off



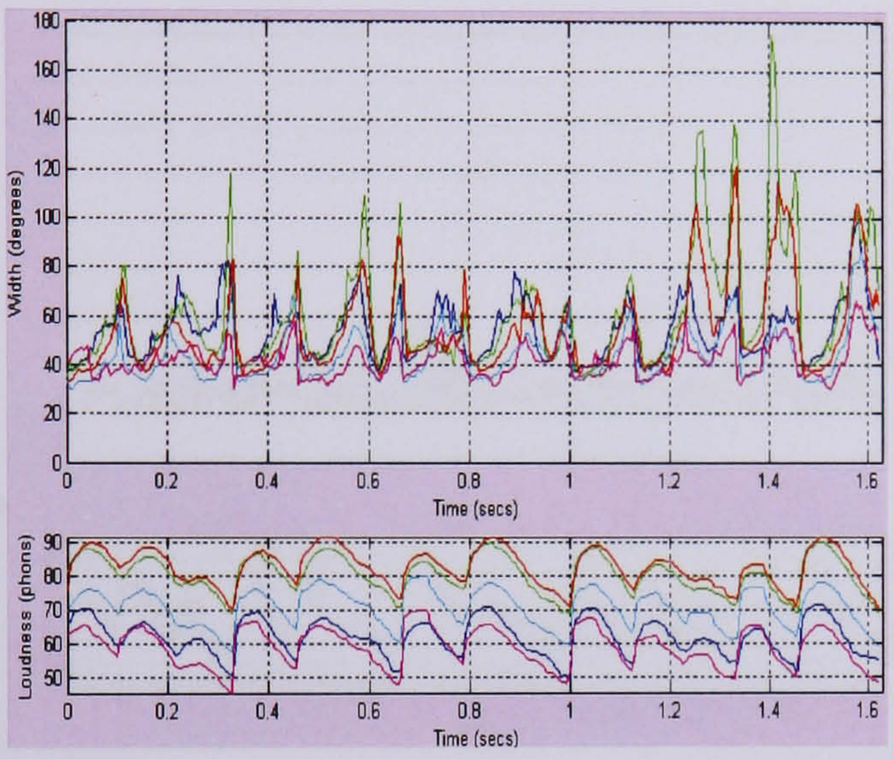
(b) Crosstalk-on

(c) Waveform

Figure B.47 Plots of the width measurement made for the room-reverberant bongo stimuli of microphone array 4, with *fc* = 7050, 8600, 10750 Hz, and waveform of the binaural signal



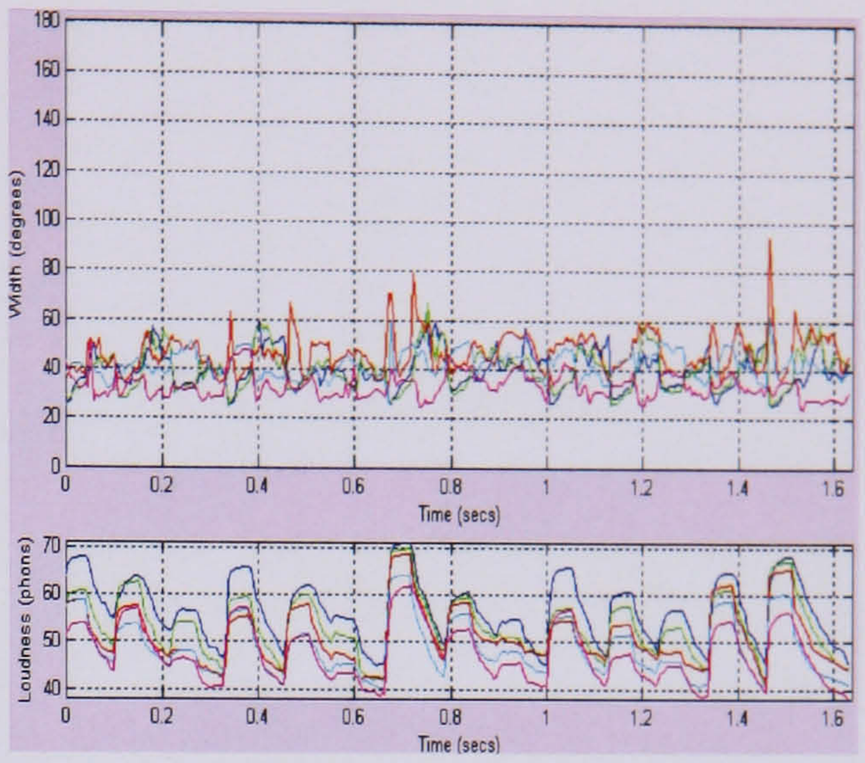
(a) Crosstalk-off



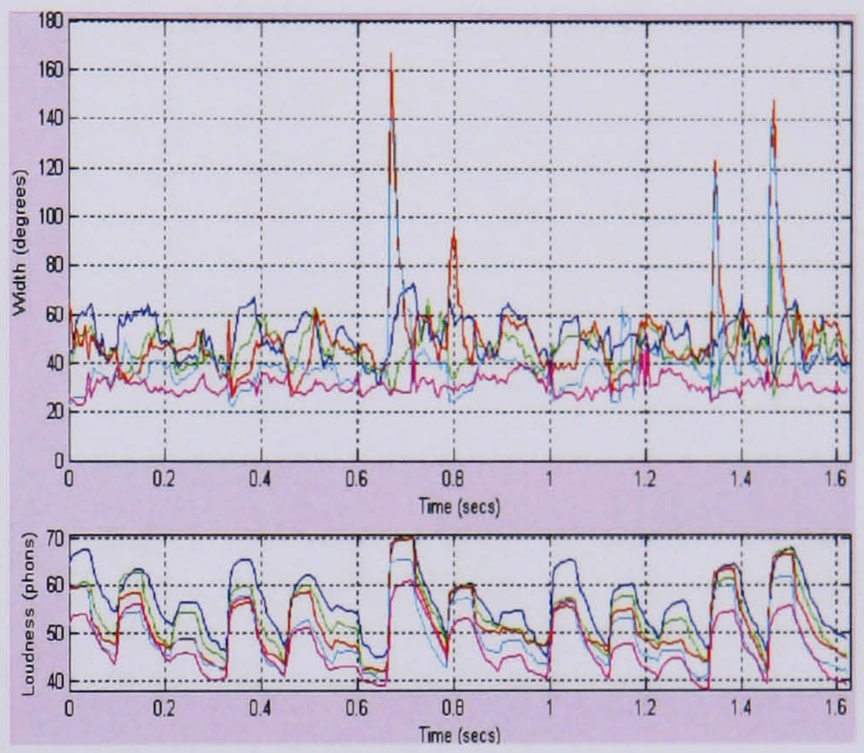
(b) Crosstalk-on

(c) Waveform

Figure B.48 Plots of the width measurement made for the hall-reverberant bongo stimuli of microphone array 4, with *fc* = 150, 250, 350, 455, 570 Hz, and waveform of the binaural signal



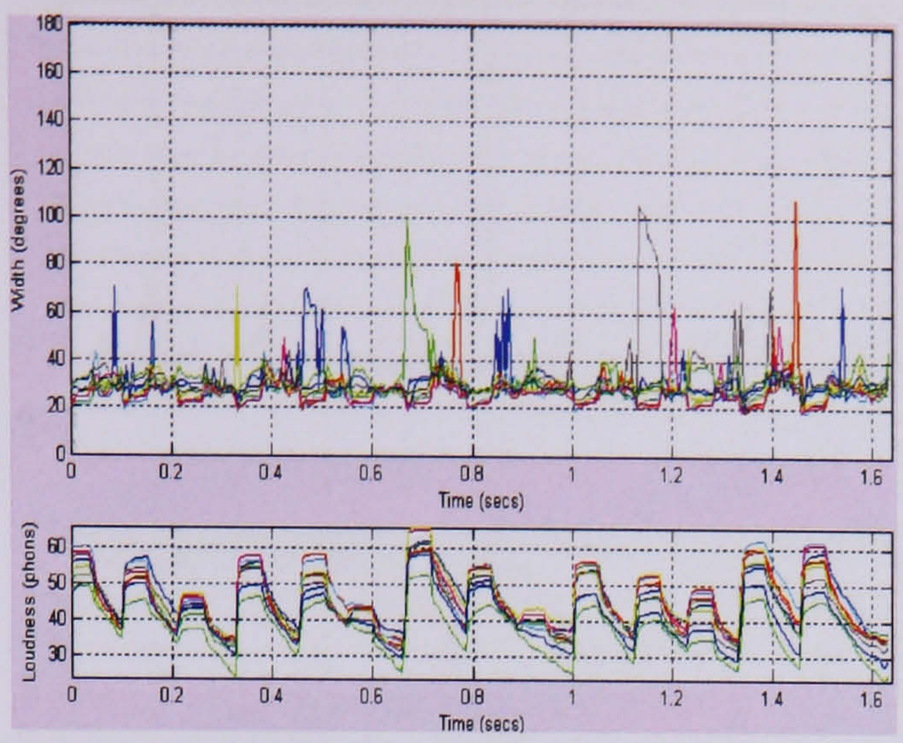
(a) Crosstalk-off



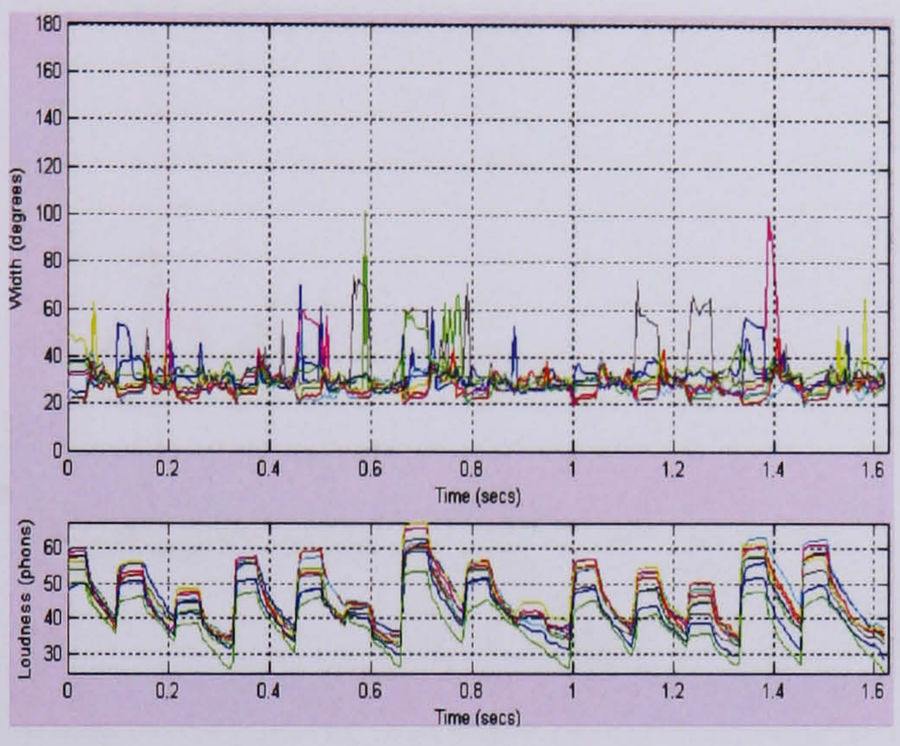
(b) Crosstalk-on

(c) Waveform

Figure B.49 Plots of the width measurement made for the hall-reverberant bongo stimuli of microphone array 4, with *fc* = 700, 845, 1000, 1175, 1375 Hz, and waveform of the binaural signal



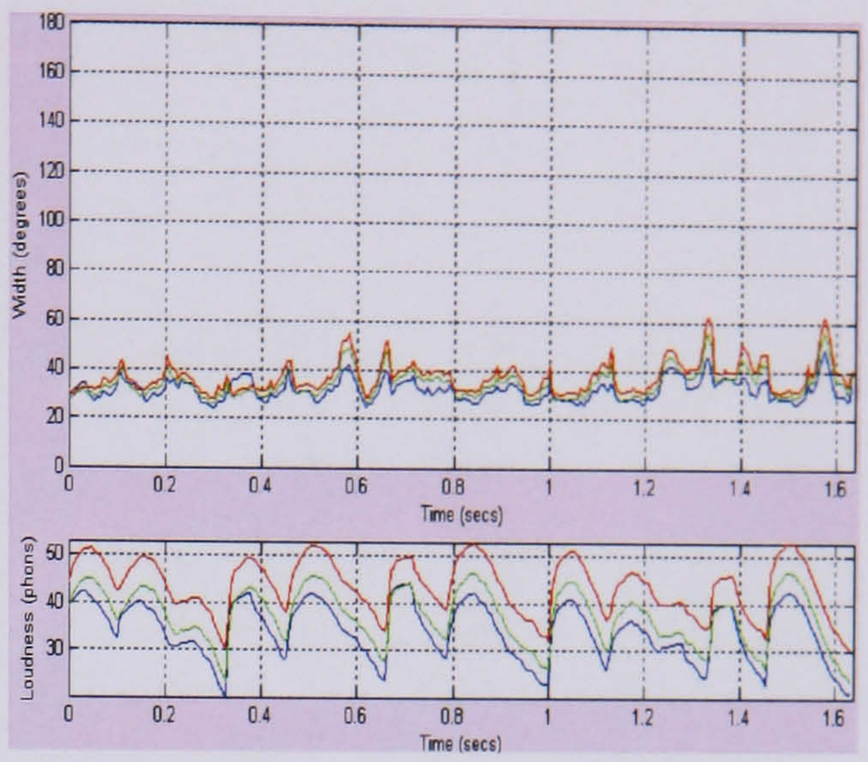
(a) Crosstalk-off



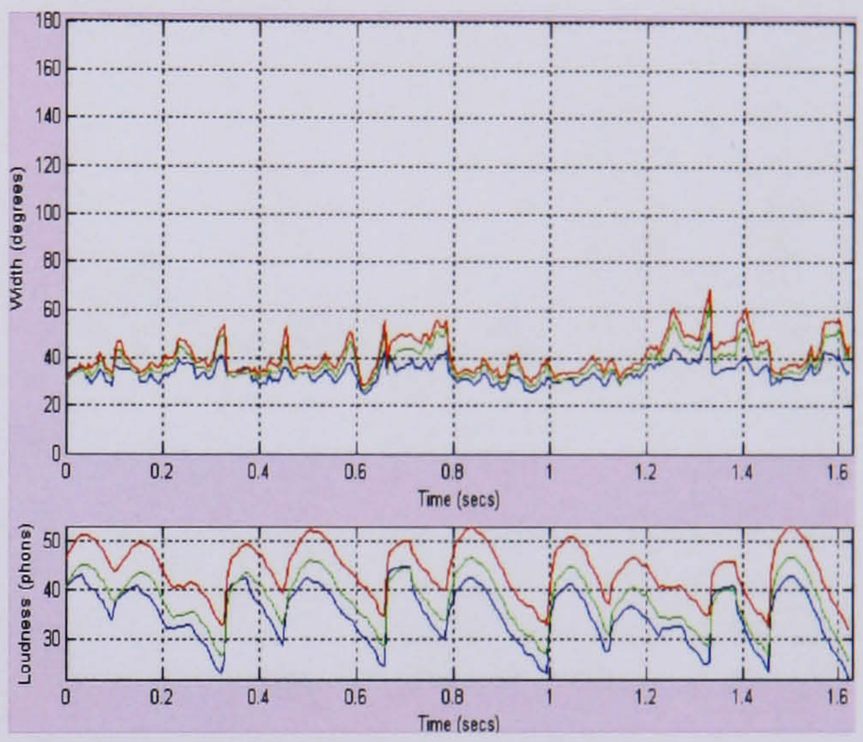
(b) Crosstalk-on

(c) Waveform

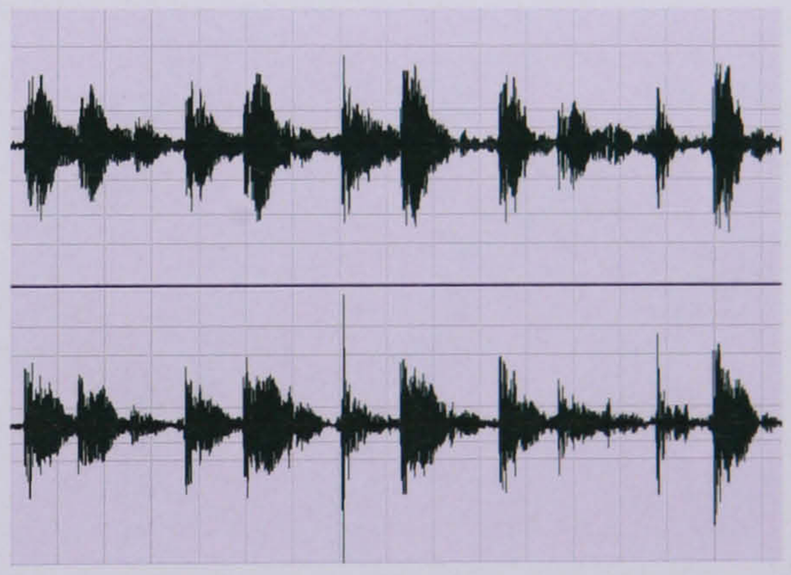
Figure B.50 Plots of the width measurement made for the hall-reverberant bongo stimuli of microphone array 4, with *fc* = 1600, 1860, 2160, 2570, 2925, 3425, 4050, 4850, 5850 Hz, and waveform of the binaural signal



(a) Crosstalk-off

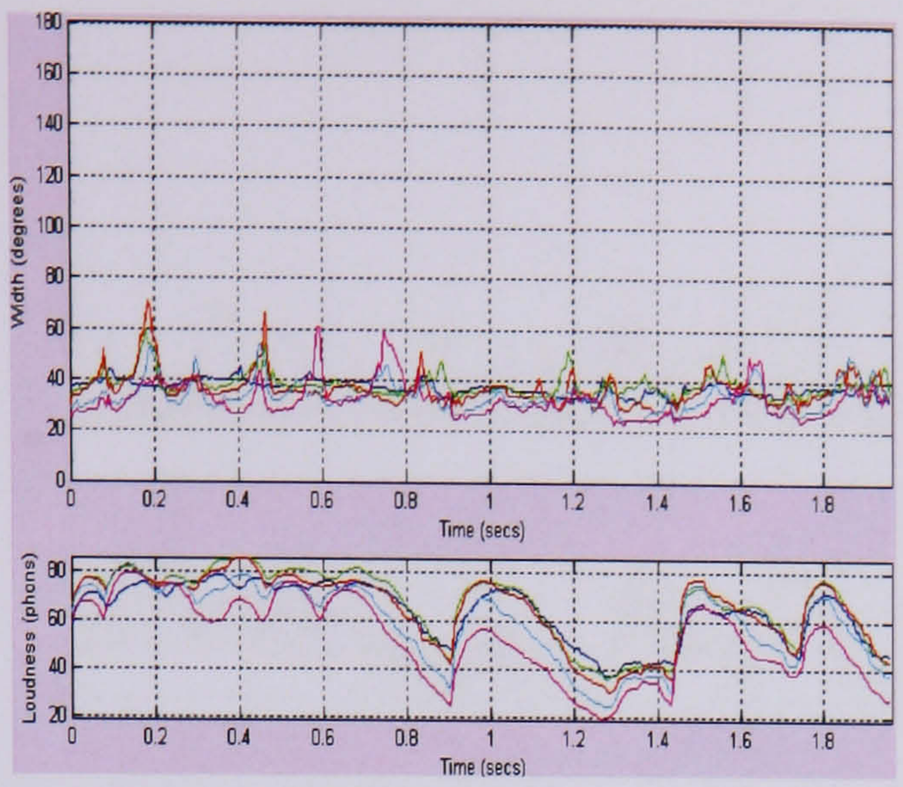


(b) Crosstalk-on

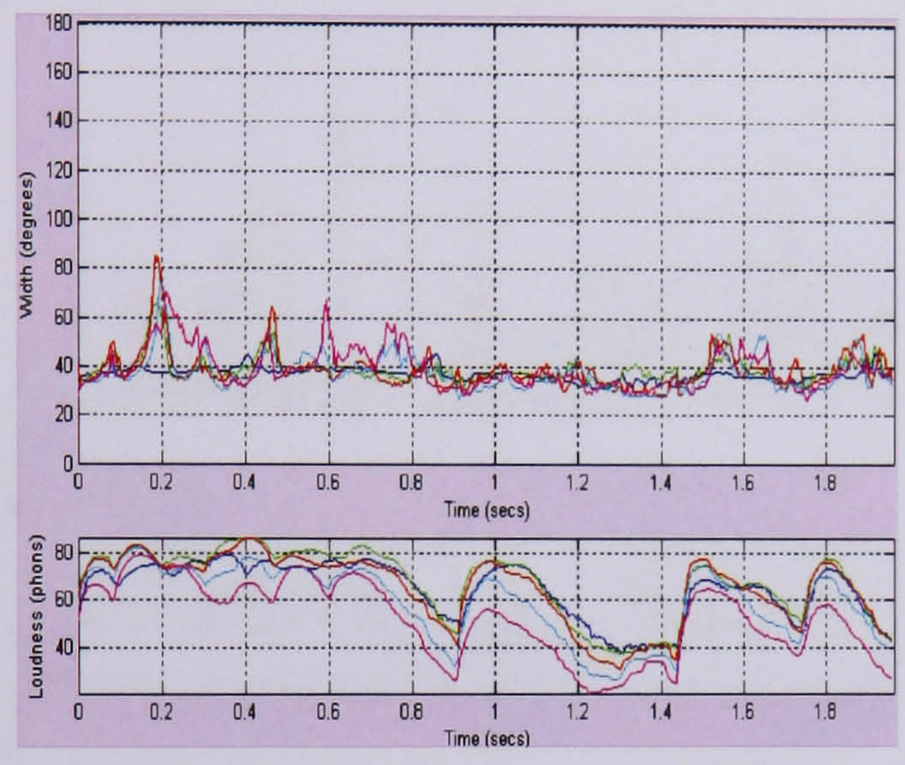


(c) Waveform

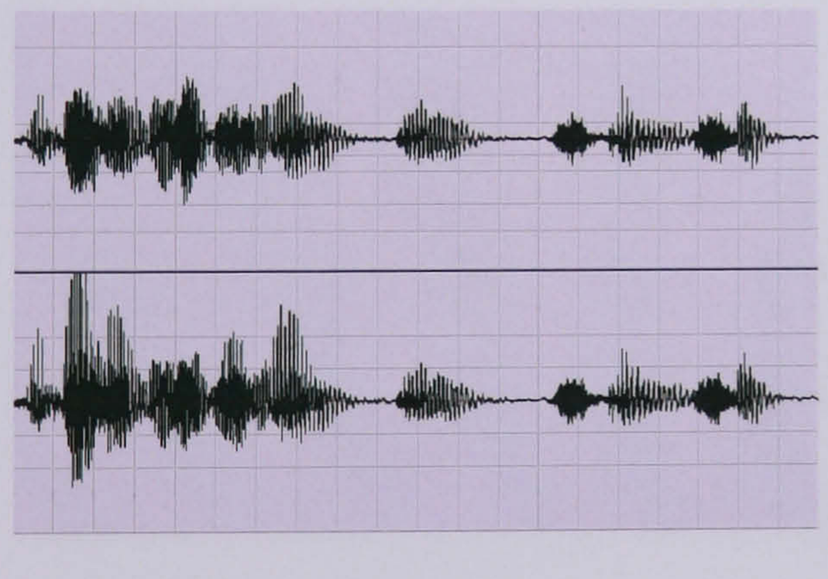
Figure B.51 Plots of the width measurement made for the hall-reverberant bongo stimuli of microphone array 4, with *fc* = 7050, 8600, 10750 Hz, and waveform of the binaural signal



(a) Crosstalk-off

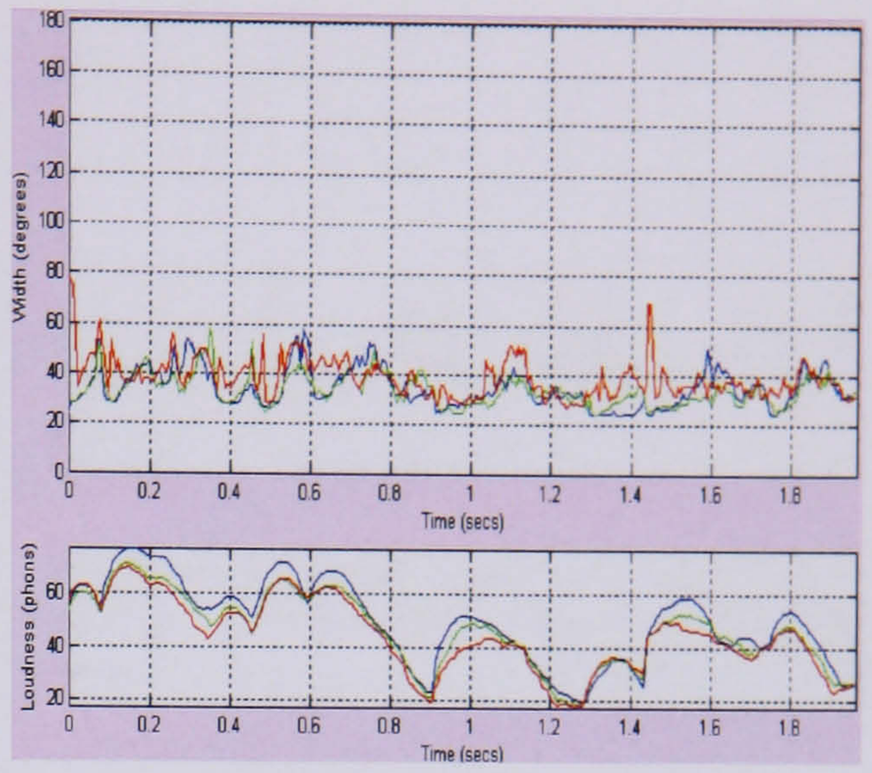


(b) Crosstalk-on

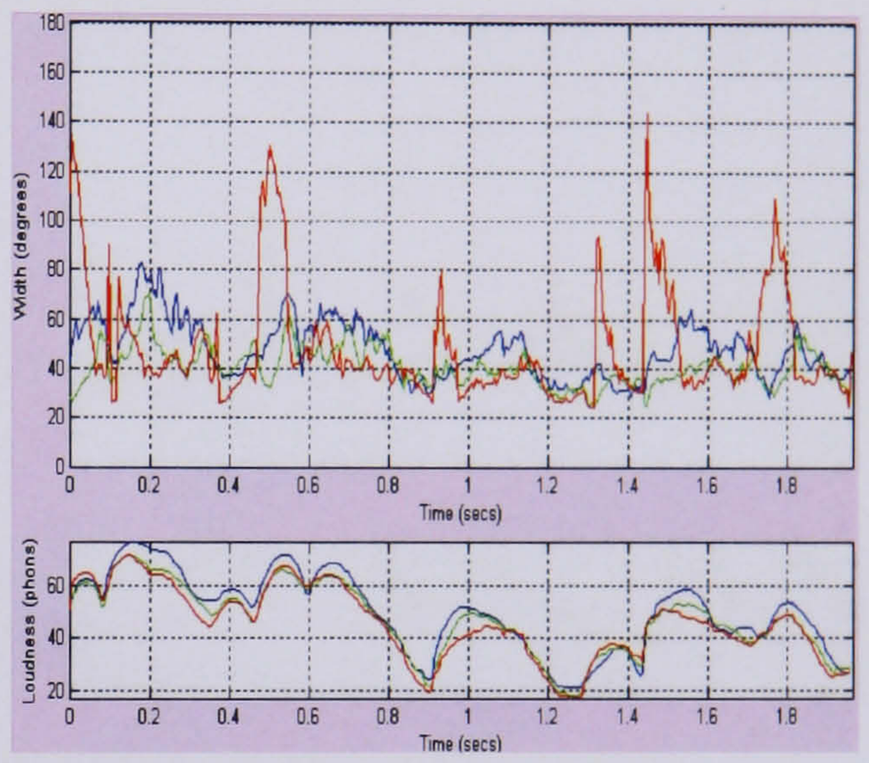


(c) Waveform

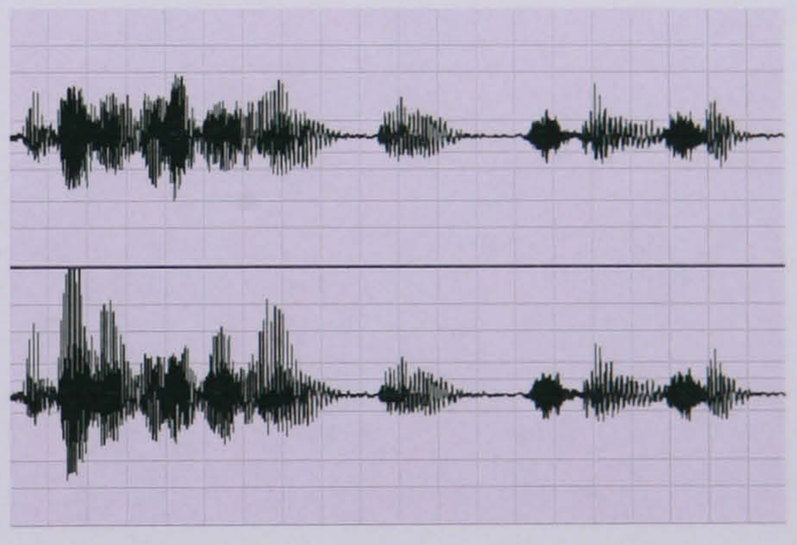
Figure B.52 Plots of the width measurement made for the anechoic speech stimuli of microphone array 4, with *fc* = 150, 250, 350, 455, 570 Hz, and waveform of the binaural signal



(a) Crosstalk-off

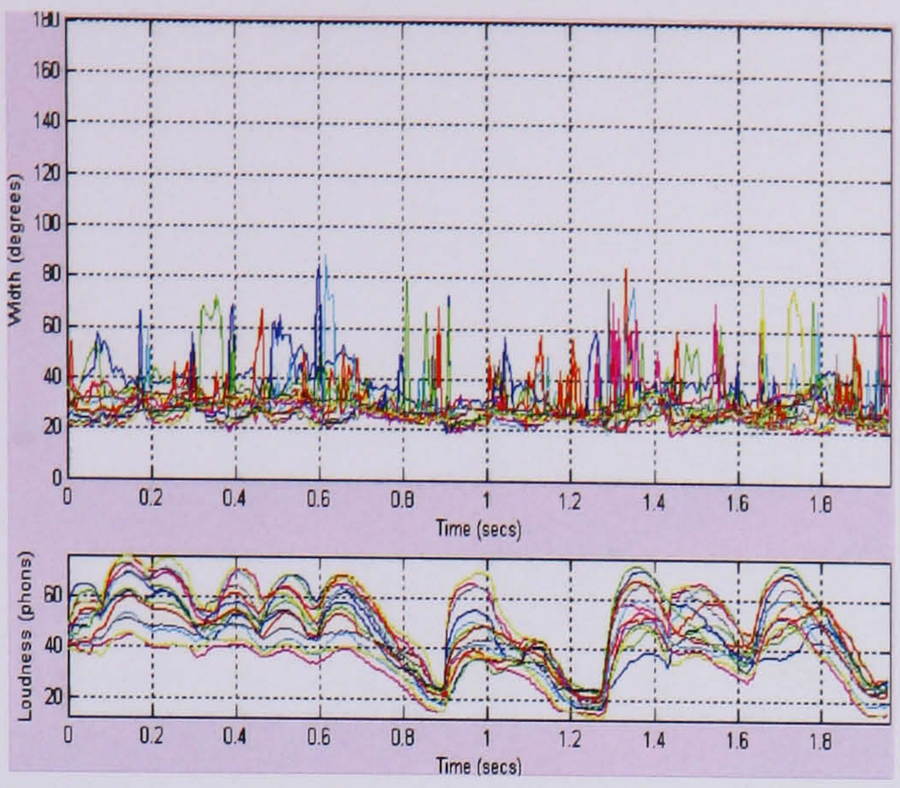


(b) Crosstalk-on

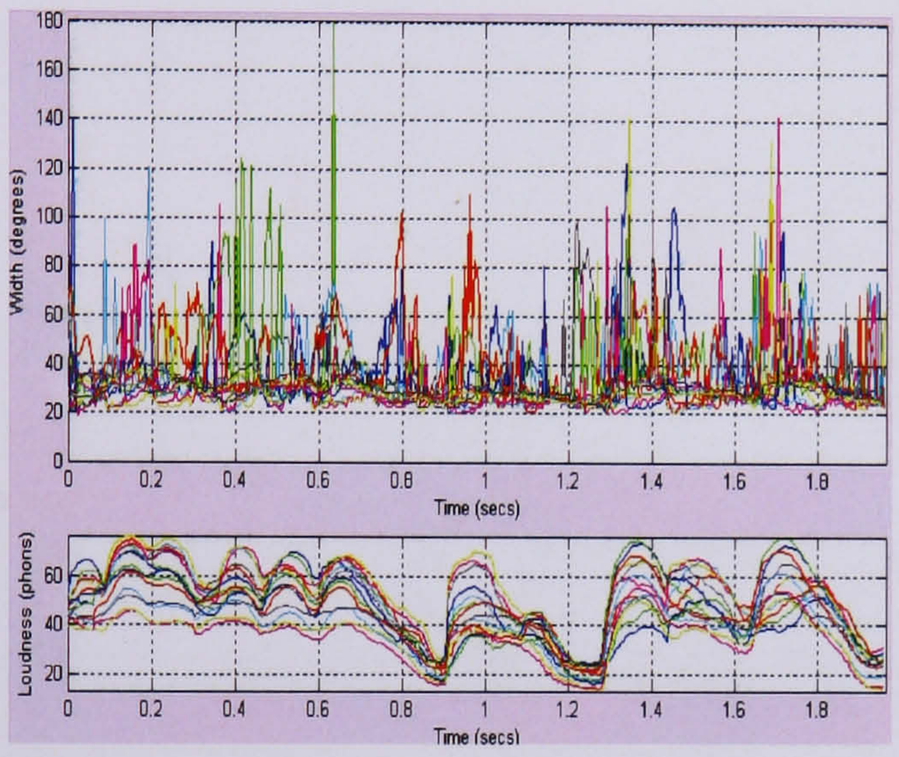


(c) Waveform

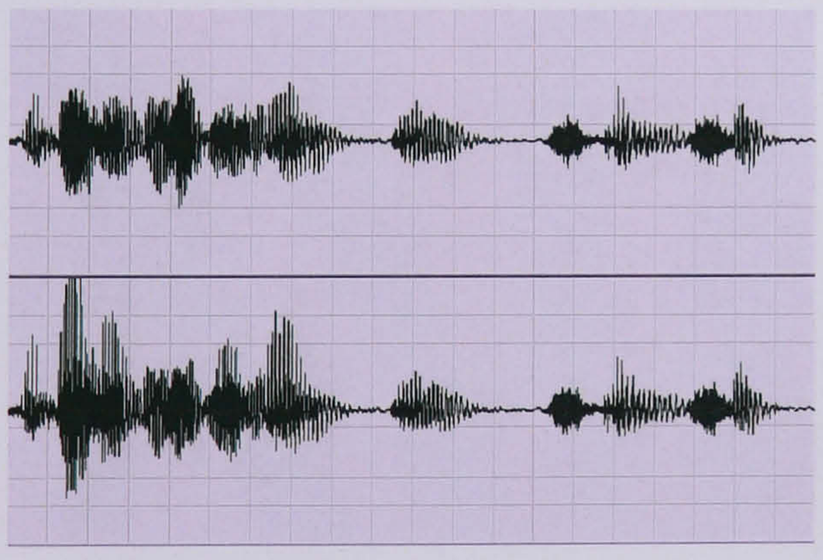
Figure B.53 Plots of the width measurement made for the anechoic speech stimuli of microphone array 4, with *fc* = 700, 845, 1000 Hz, and waveform of the binaural signal



(a) Crosstalk-off

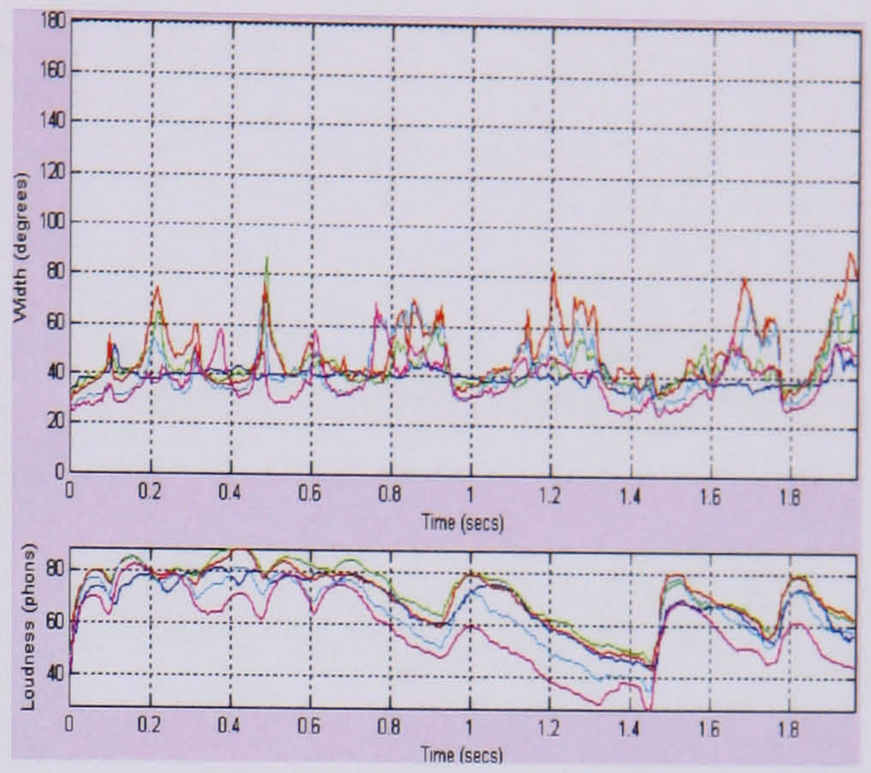


(b) Crosstalk-on

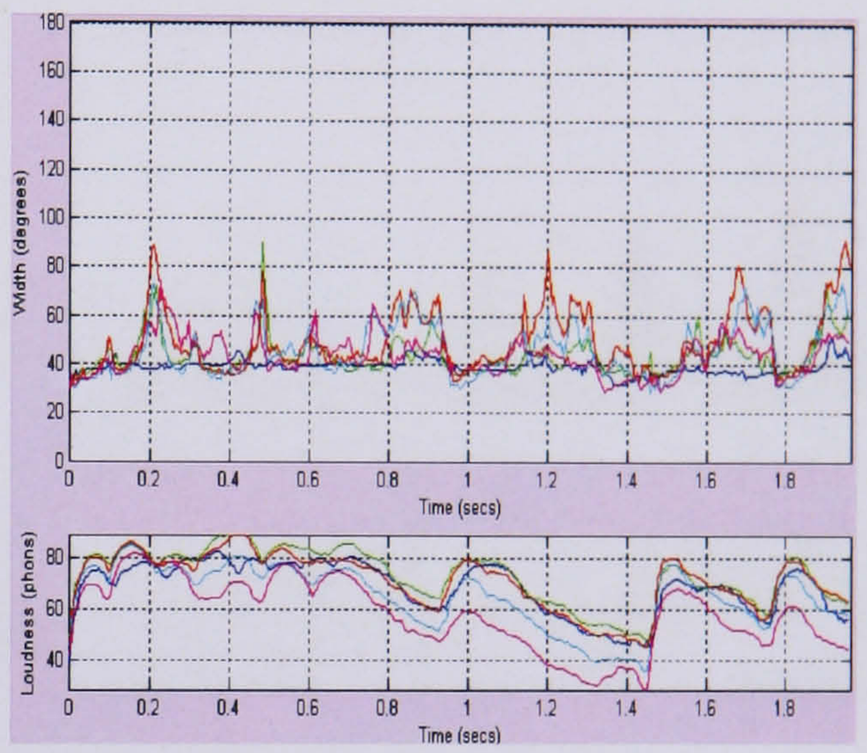


(c) Waveform

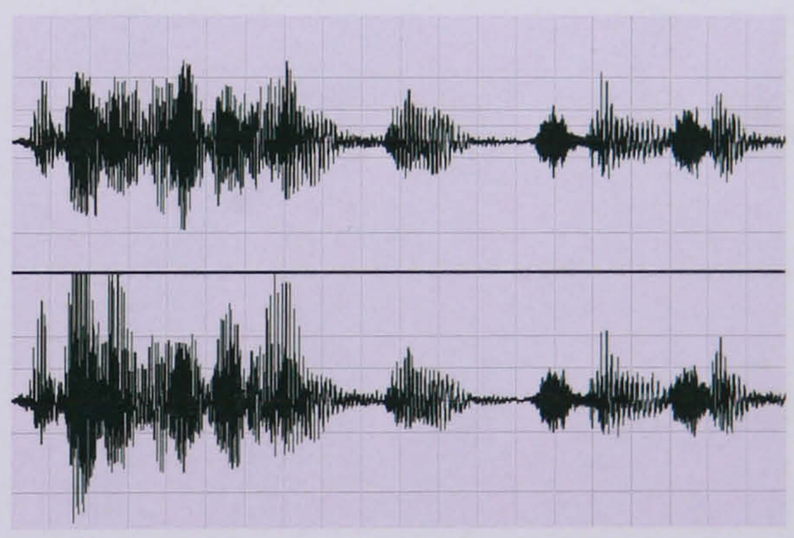
Figure B.54 Plots of the width measurement made for the anechoic speech stimuli of microphone array 4, with *fc* = 1175, 1375, 1600, 1860, 2160, 2570, 2925, 3425, 4050, 4850, 5850, 7050, 8600, 10750 Hz, and waveform of the binaural signal



(a) Crosstalk-off

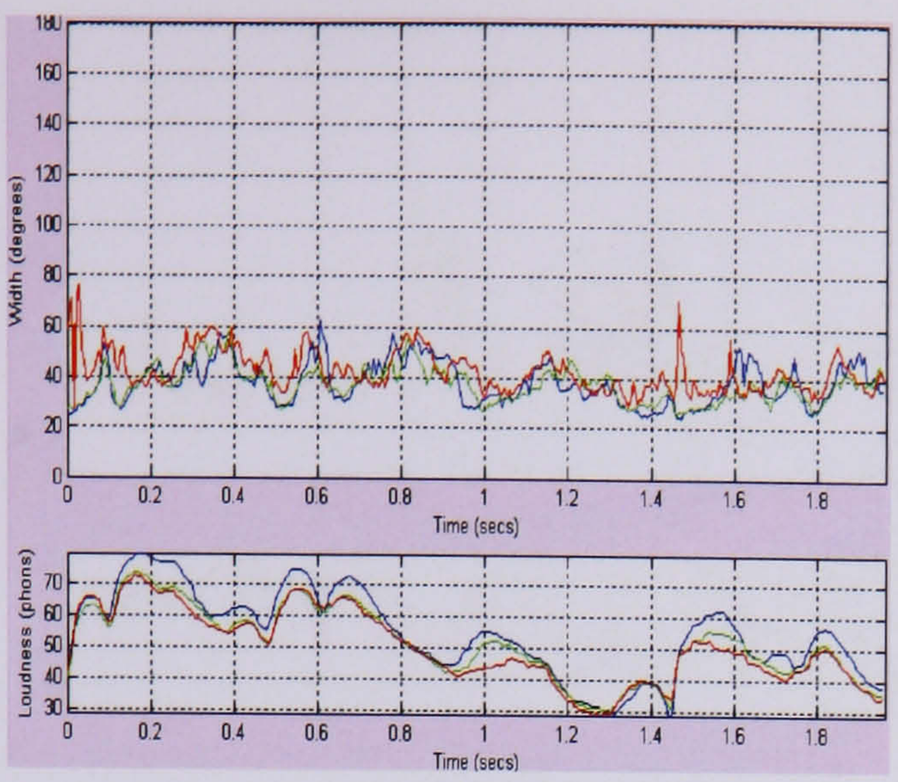


(b) Crosstalk-on

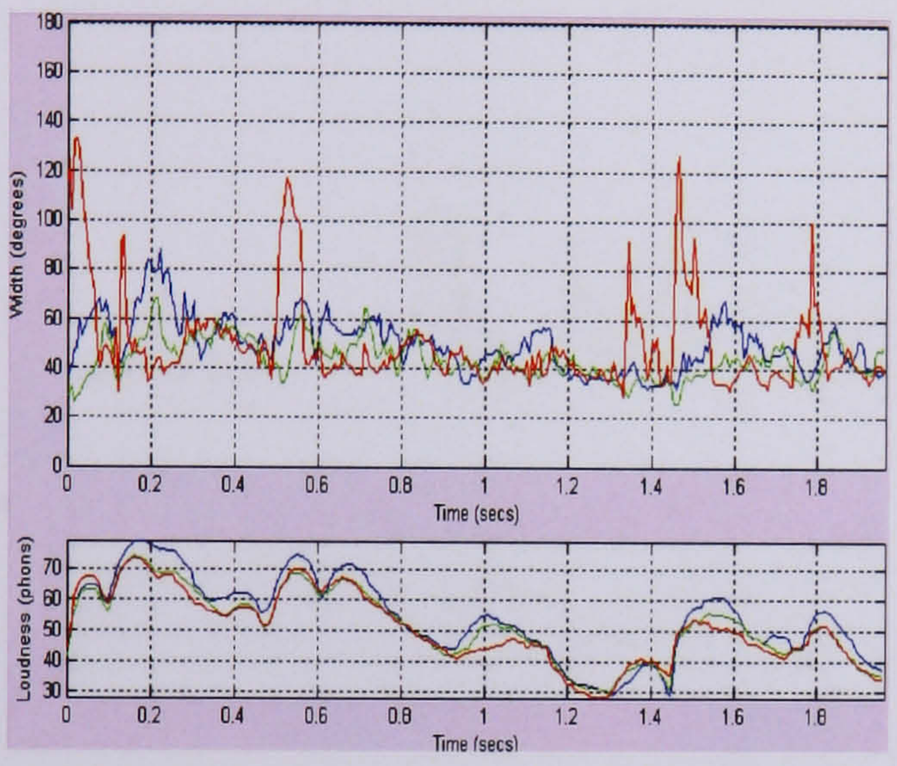


(c) Waveform

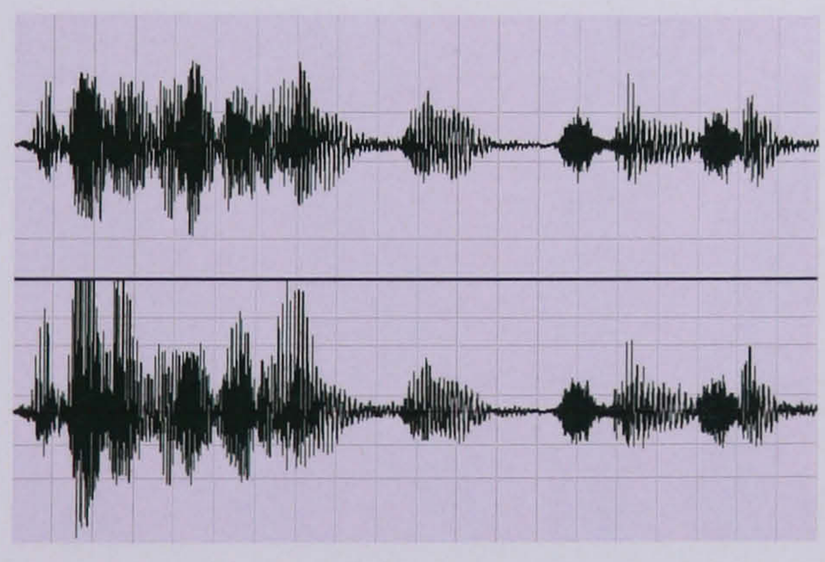
Figure B.55 Plots of the width measurement made for the room-reverberant speech stimuli of microphone array 4, with *fc* = 150, 250, 350, 455, 570 Hz, and waveform of the binaural signal



(a) Crosstalk-off

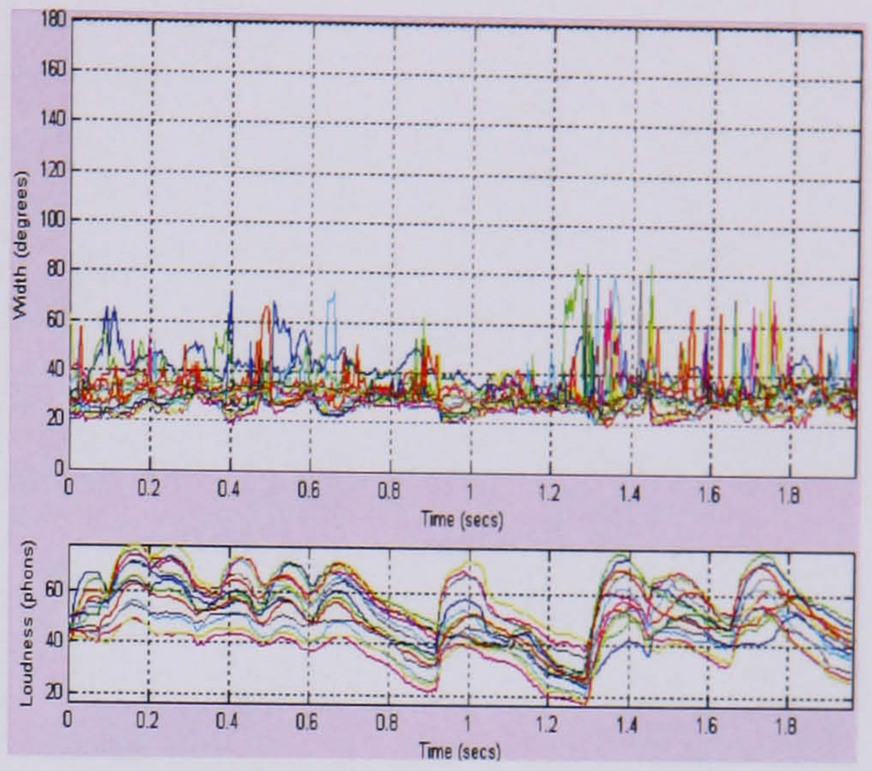


(b) Crosstalk-on

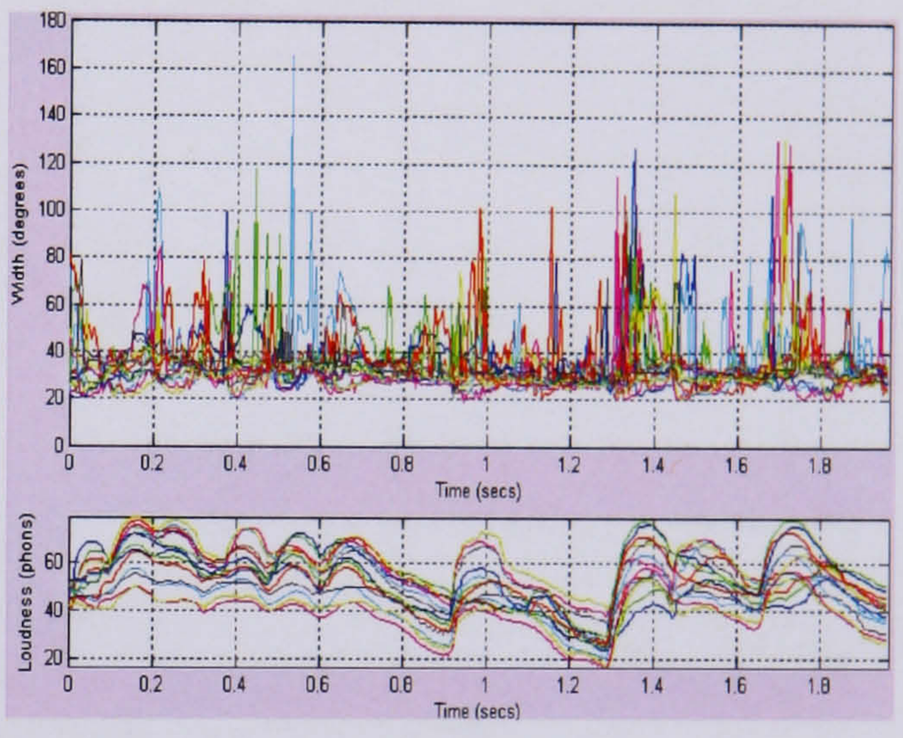


(c) Waveform

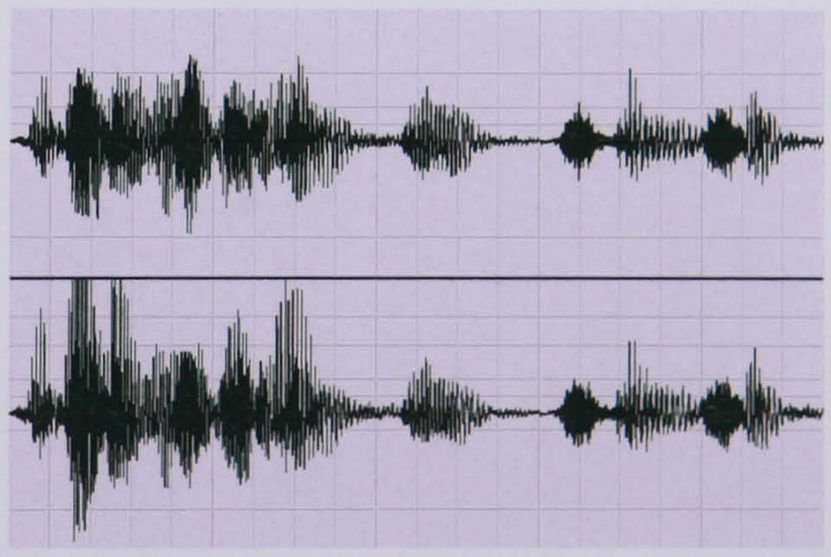
Figure B.56 Plots of the width measurement made for the room-reverberant speech stimuli of microphone array 4, with *fc* = 700, 845, 1000 Hz, and waveform of the binaural signal



(a) Crosstalk-off

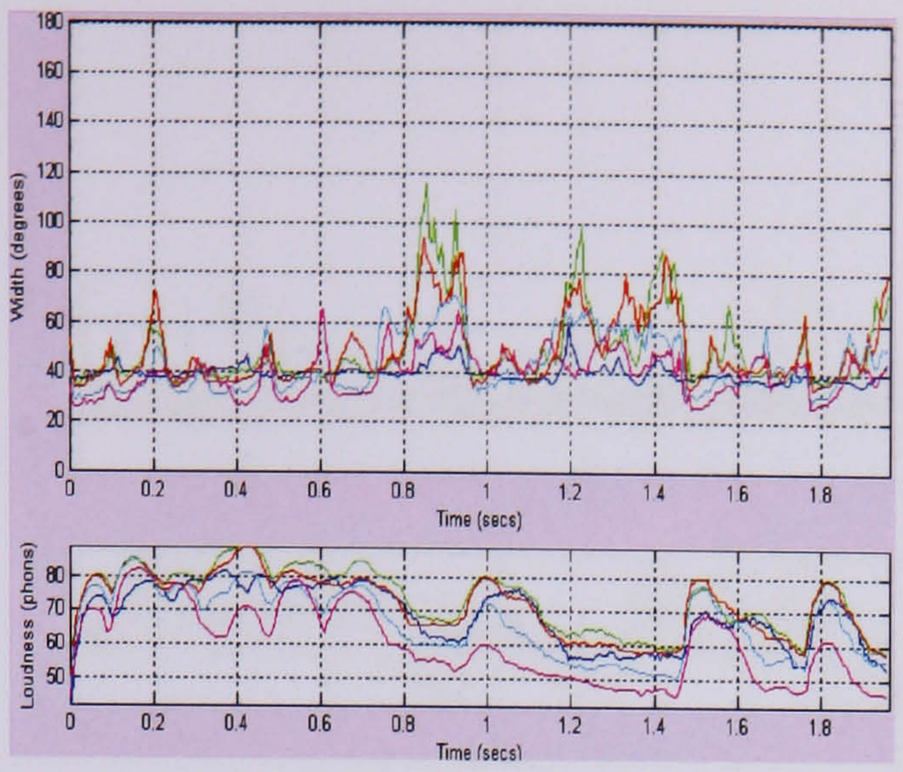


(b) Crosstalk-on

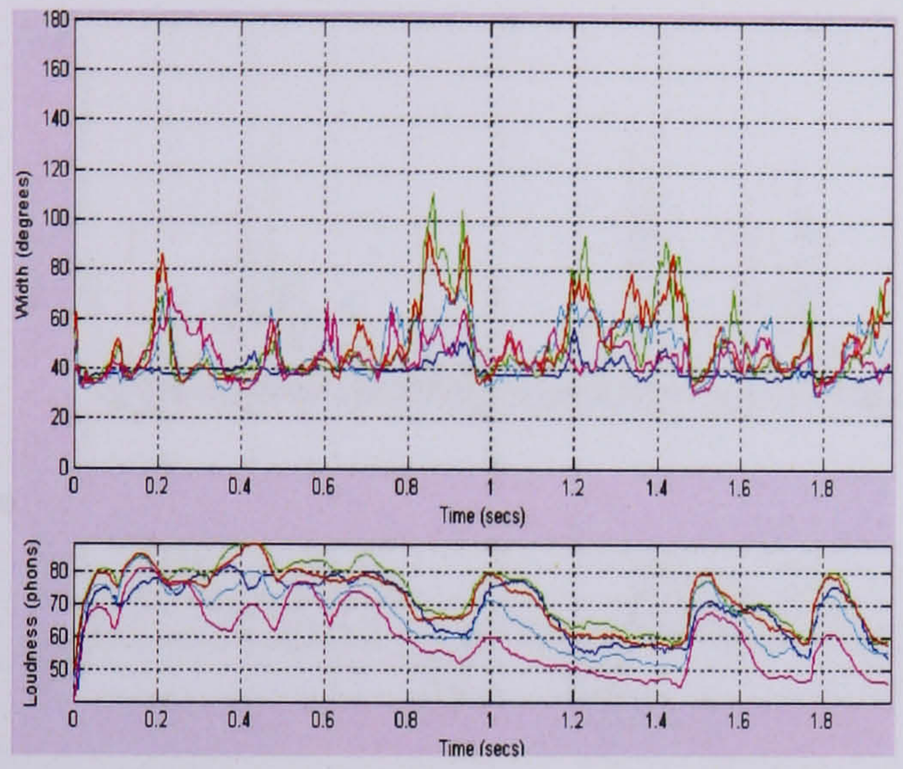


(c) Waveform

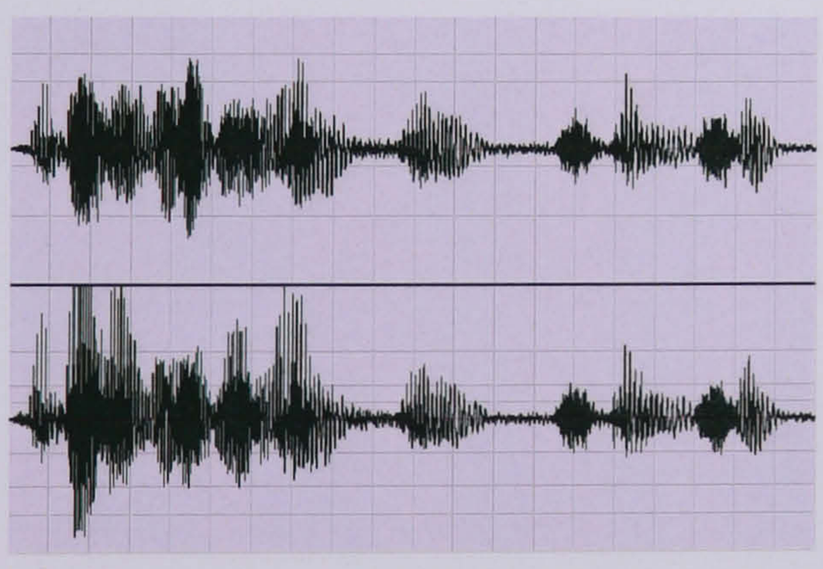
Figure B.57 Plots of the width measurement made for the room-reverberant speech stimuli of microphone array 4, with *fc* = 1175, 1375, 1600, 1860, 2160, 2570, 2925, 3425, 4050, 4850, 5850, 7050, 8600, 10750 Hz, and waveform of the binaural signal



(a) Crosstalk-off

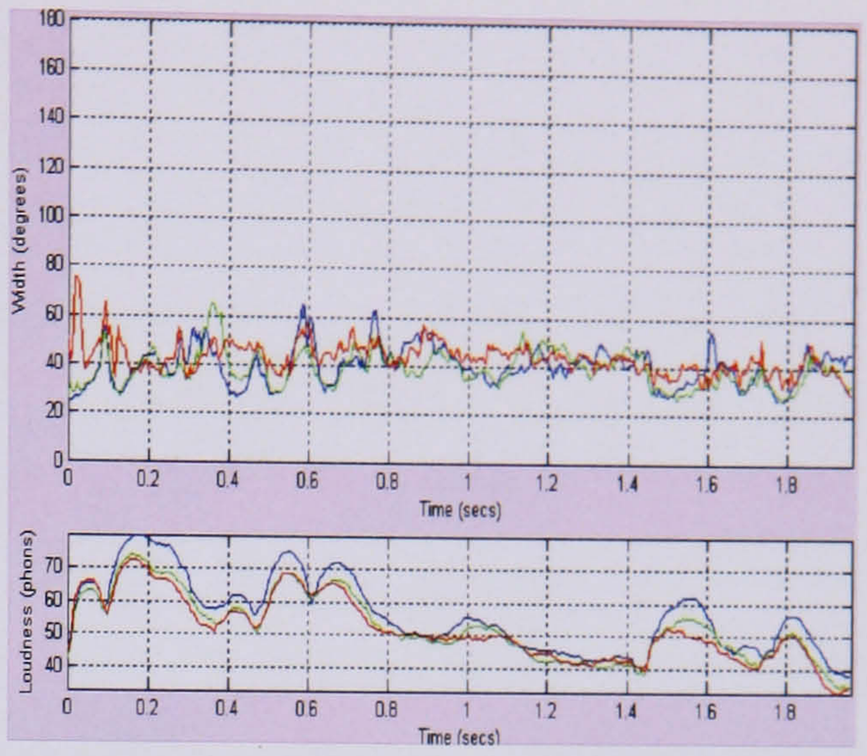


(b) Crosstalk-on

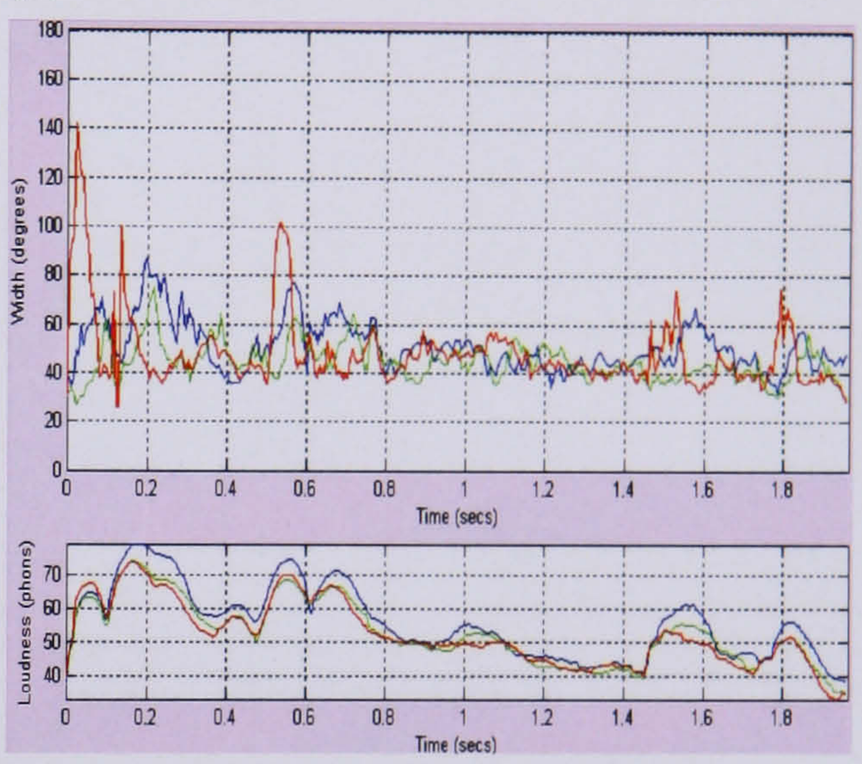


(c) Waveform

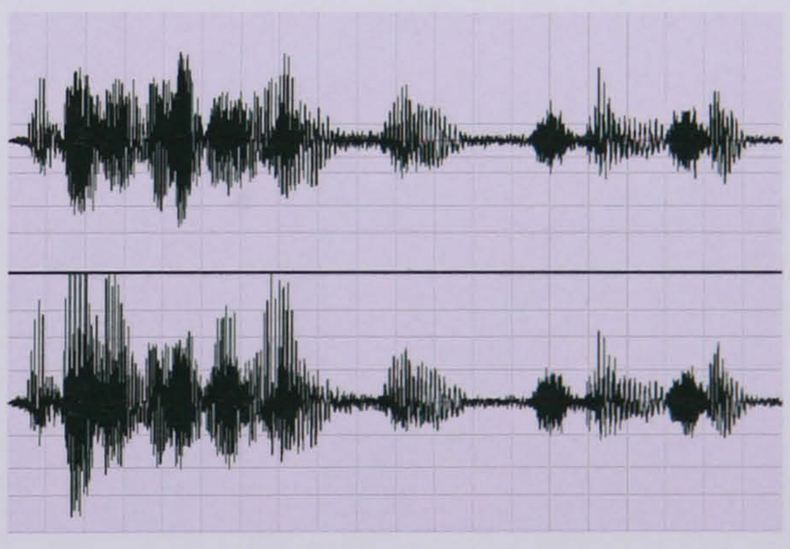
Figure B.58 Plots of the width measurement made for the hall-reverberant speech stimuli of microphone array 4, with *fc* = 150, 250, 350, 455, 570 Hz, and waveform of the binaural signal



(a) Crosstalk-off

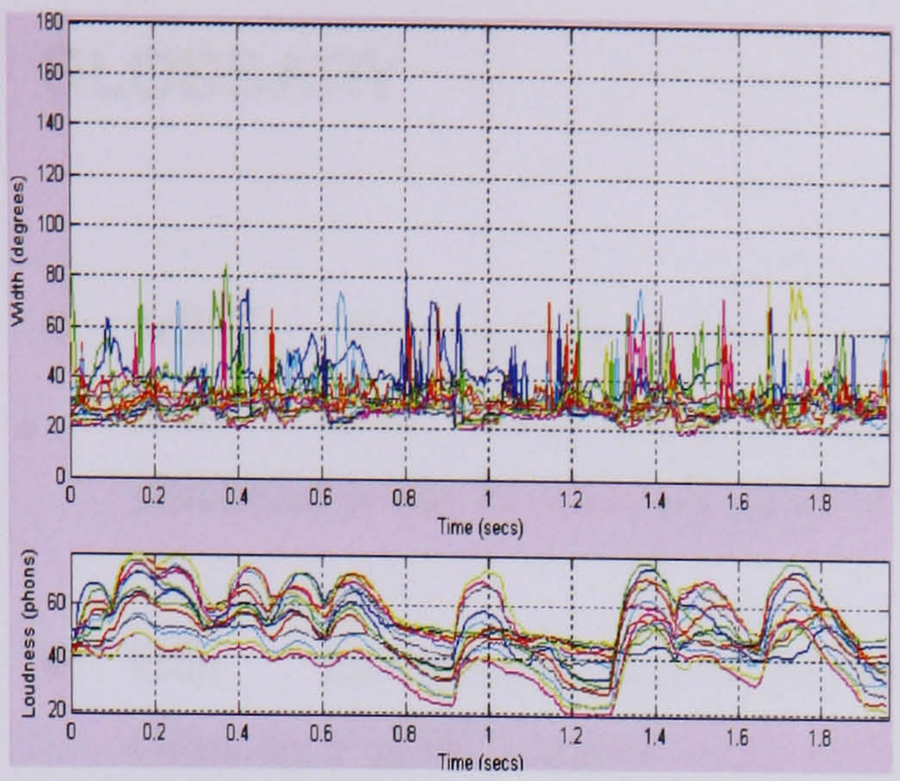


(b) Crosstalk-on

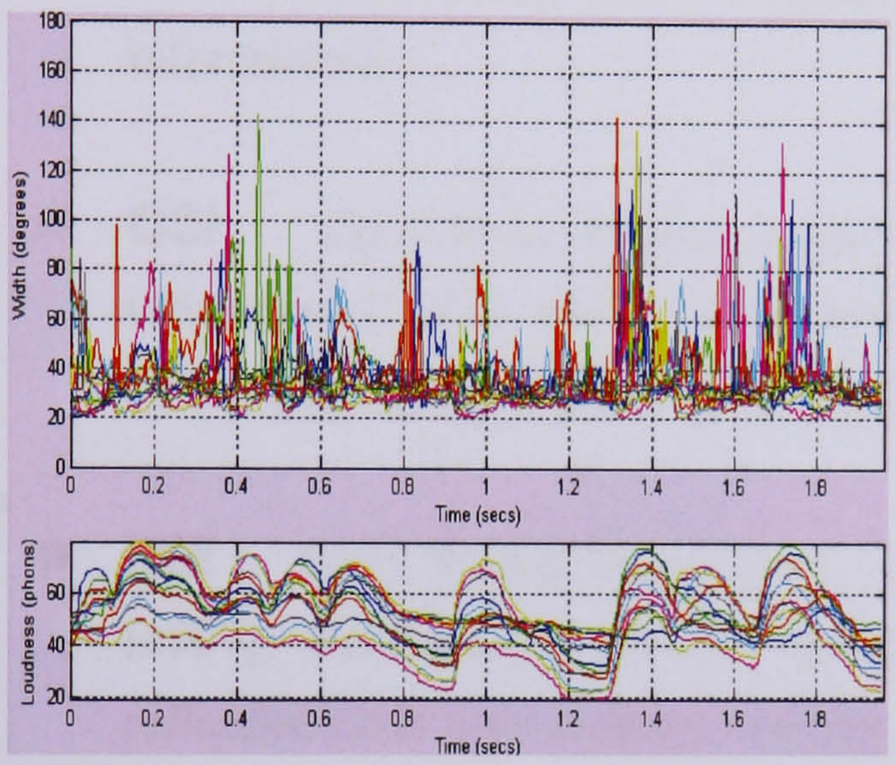


(c) Waveform

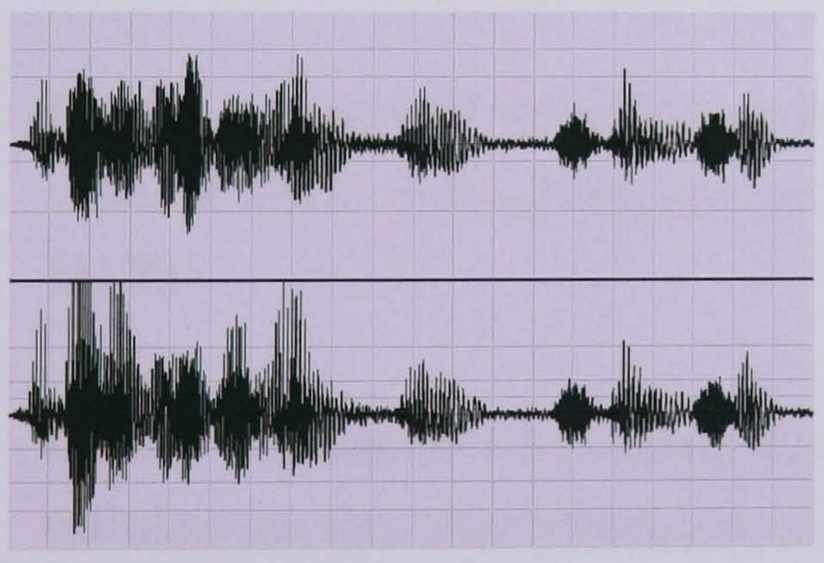
Figure B.59 Plots of the width measurement made for the hall-reverberant speech stimuli of microphone array 4, with *fc* = 700, 845, 1000 Hz, and waveform of the binaural signal



(a) Crosstalk-off



(b) Crosstalk-on



(c) Waveform

Figure B.60 Plots of the width measurement made for the hall-reverberant speech stimuli of microphone array 4, with *fc* = 1175, 1375, 1600, 1860, 2160, 2570, 2925, 3425, 4050, 4850, 5850, 7050, 8600, 10750 Hz, and waveform of the binaural signal

GLOSSARY

- **ASW** Apparent or Auditory Source Width. A spatial concept that was derived from concert hall acoustic research, which normally refers to the perceived width of sound image that is related to sound source.

- **BSI** Background Spatial Impression. A concept that was proposed by Griesinger [1997], which refers to the spatial perception that is associated with the reflections arriving in the foreground stream (see Section 2.3.1.2 for more information).

- **CSI** Continuous Spatial Impression. A concept that was proposed by Griesinger [1997], which refers to the spatial perception that is associated with continuous sound (see Section 2.3.1.2 for more information).

- **ESI** Early Spatial Impression. A concept that was proposed by Griesinger [1997], which refers to the spatial perception that is associated with the reflections and reverberation arriving 120ms after the end of all foreground sound events (see Section 2.3.1.2 for more information).

- **IACC** Interarual Cross-correlation Coefficient. The maximum absolute value of IACF over all frequencies in the range between -1ms and 1ms (see IACF).

- **IACF** Interaural Cross-correlation Function. A function that is used for calculate the similarity between the signals reaching each ear (see Section 2.3.2.3 for more information).

- **ICA-3** Ideal Cardioid Array – 3. A three-channel microphone technique using three cardioid microphones. This technique was proposed by Herrmann and Henkels [1998], based on the 'critical linking' design concept [Williams and Le Du 1999, 2000] (see Section 1.4.2 for more information).

- **ICID** 'Interchannel' Intensity Difference. Phantom imaging of a coincident microphone technique relies on the intensity difference between two-channel signals.

- **ICTD** 'Interchannel' Time Difference. Phantom imaging of a spaced omni microphone technique relies on the time-of-arrival difference between two-channel signals.

- **IID** 'Interaural' Intensity Difference. For a non-median sound source, the ear-input signals will have a difference in intensity at frequencies above around 1kHz due to the head-shadowing effect.

- **ITD** 'Interaural' Time Difference. For a non-median sound source, the ear-input signals will have a difference in time at frequencies below around 1kHz.

- **LEV** Listener Envelopment. A spatial concept that was derived from concert hall acoustics research, which refers to the subjective impression of being surrounded by the reverberant sound field.

- **MAA** Minimum Audible Angle. Listener's ability to distinguish the directional change of the sound source decreases as the direction of the source moves from the front to the side.

- **OCT** Optimised Cardioid Triangle. A three-channel microphone technique proposed by Theile [2001] aiming to obtain the maximum reduction of interchannel crosstalk. This technique employs two super-cardioid microphones for the side channels and a cardioid microphone for the centre channel (see Section 1.4.2 for more information).

- **Ψ** Degree of phantom image shift in stereophonic reproduction.

- **SI** Spatial Impression. A spatial concept that was derived from concert hall acoustics research, which is normally considered to include two sub-attributes of ASW and LEV (see ASW and LEV).

- **SRA** Stereophonic Recording Angle. The sector of the sound field in front of the microphone array that is localised at fully left or right between the two loudspeakers (See Section 1.2.1 for more information).

REFERENCES

Ando, Y. and Kageyama, K. (1977): 'Subjective preference of sound with a single early reflection', *Acustica*, 37, pp.112-117

Banks, S. and Green M. (1973): 'Localisation of High- and Low-Frequency Transients', *Journal of the Acoustical Society of America*, 53, pp.1432-1433.

Barron, M. (1971): 'The subjective effects of first reflections in concert halls – the need for lateral reflections', *Journal of Sound and Vibration*, 15, pp.475-494.

Barron, M. and Marshall, A. (1981): 'Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure', *Journal of Sound and Vibration*, 77, pp.211-232.

Bech, S. (1999): 'Methods for subjective evaluation of spatial characteristics of sound', *Proceedings of the Audio Engineering Society 16th International Conference*, pp.407-504.

Beranek, L. (1996): *Concert and Opera Halls – how they sound* (New York: Acoustical Society of America)

Berg, J. and Rumsey, F. (1999): 'Spatial attribute identification and scaling by repertory grid technique and other methods', *Proceedings of the 16th Audio Engineering Society International Conference*, pp.51-66.

Berg, J. and Rumsey, F. (2002): 'Validity of selected spatial attributes in the evaluation of 5-channel microphone techniques', Audio Engineering Society 112th Convention, Preprint 5593.

Blauert, J. (1972): 'On the lag of lateralisation caused by interaural time and intensity differences', *Audiology*, 11, pp.265-270.

Blauert, J. (1997): *Spatial Hearing. The psychophysics of Human Sound Localisation* (Cambridge: MIT Press)

Blauert, J. and Divenyi, P. (1988): 'Spectral selectivity in binaural contralateral inhibition', *Acustica*, 66, 267-274.

Blauert, J. and Lindemann, W. (1986): 'Auditory spaciousness: Some further psychoacoustic analyses', *Journal of the Acoustical Society of America*, 80, pp.533-542.

Bradley, J. and Soulodre, G. (1995): 'Objective measures of listener envelopment', *Journal of the Acoustical Society of America*, 98, pp.2590-2597.

Chernyak, R. and Dubrovsky, N. (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. A3-A12.

Clark, H., Dutton, G.. and Vanderlyn, P. (1958): 'The stereophonic recording and reproducing system: a two-channel system for domestic tape records', *Journal of the Acoustical Society of America*, 6, pp.102-117.

Clifton, R. (1987): 'Breakdown of echo suppression in the precedence effect', *Journal of the Acoustical Society of America*, 82, pp.1834-1835.

Clifton, R. and Freyman, R. (1989): 'Effect of click rate and delay on breakdown of the precedence effect', *Percept. Psychophys.*, 46, pp.139-145.

Clifton, R., Freyman, R., Litovski, R. and McCall, D. (1994): 'Listeners' expectations about echoes can raise and lower echo threshold', *Journal of the Acoustical Society of America*, 95, pp.1525-1533.

Damaschke, J., Granzow, M., Riedel, H. and Kollmeier, B. (2000): 'The equivalence relation between interaural time and level differences for short stimuli', *Zietschrift fur Audiologie*, vol. 39, pp. 40-52.

Damaske, P. (1971): 'Head-related two-channel stereophony with loudspeaker reproduction', *Journal of the Acoustical Society of America*, 50, pp.1109-1115.

David, E., Guttman, N. and van Bergeijk, W. (1959): 'Binaural interaction of high-frequency complex stimuli', *Journal of the Acoustical society of America*, 31, pp.774-782.

Deatherage, B. and Hirsh, I. (1959): 'Auditory localisation of clicks', *Journal of the Acoustical Society of America*, 31, pp.486-492.

de Boer, K. (1940): 'Stereophonic Sound Reproduction', *Philips Technical Review*, 5, pp.107-114.

Dooley, W. and Streicher, R. (1982): 'M-S stereo: A powerful techniques for working in stereo', *Journal of the Audio Engineering Society*, 30, pp.707-718.

Eargle, J. (2001): *The microphone book* (Boston: Focal Press)

Feddersen, w., Sandel, T. Teas, D. and Jeffress, L. (1957): 'Localisation of high-frequency tones', *Journal of the Acoustical Society of America*, 29, pp.988-991.

Fletcher, H. and Munson, W. (1933): 'Loudness, its measurement and calculation', *Journal of the Acoustical Society of America*, 5, pp.82-108.

Freyman, R., Clifton, R. and Litovski, R. (1991): 'Dynamic process in the precedence effect', *Journal of the Acoustical Society of America*, 90, pp.874-884.

Fukada, A., Tsujimoto, K. and Akita, S. (1997): 'Microphone techniques for ambient sound on a music recording', Audio Engineering Society 103rd Convention, Preprint 4540.

Fukada, A. (2001): 'A challenge in multichannel music recording', *Proceedings of the 19th Audio Engineering Society International Conference*, pp.439-446.

Gabrielsson, A. and Sjogren, H. (1979): 'Perceived sound quality of sound-reproducing systems', *Journal of the Acoustical Society of America*, 65, pp.1019-1033.

Grantham, W. and Wightman, F. (1978): 'Detectability of varying interaural temporal differences', *Journal of the Acoustical Society of America*, 63, pp. 511-523.

Glasgal, R. (2001): 'Ambiophonics: Achieving physiological realism in music recording and reproduction', Audio Engineering Society 111th Convention, Preprint 5426.

Griesinger, D. (1992): 'IALF – binaural measures of spatial impression and running reverberance', Audio Engineering Society 92nd Convention, Preprint 3292.

Griesinger, D. (1996): 'Spaciousness and envelopment in musical acoustics', Audio Engineering Society 101st Convention, Preprint 4401.

Griesinger, D. (1997): 'The psychoacoustics of apparent source width, spaciousness & envelopment in performance spaces', *Acta Acustica*, 83, pp.721-731.

Haas, H. (1972): 'The influence of a single echo on the audibility of speech', *Journal of the Audio Engineering Society*, 20, pp.146-159.

Hafter, E. and Carrier, S. (1972): 'Binaural interaction in low-frequency stimuli: The inability to trade time and intensity completely', *Journal of the Acoustical Society of America*, 51, pp.1852-1862.

Hafter, E. and Dye, R. (1983): 'Detection of interaural differences in time in trains of high-frequency clicks as a function of interclick interval and number', *Journal of the Acoustical Society of America*, 73, pp.644-651.

Hartmann, W. (1993): 'Localisation of sounds in rooms', *Journal of the Acoustical Society of America,* 74, pp.1380-1391.

Henning, G. (1974): 'Detectability of interaural delay in high-frequency complex waveforms', *Journal of the Acoustical Society of America*, 55, pp.84-90.

Herrmann, U. and Henkels, V. (1998): 'Main microphone techniques for the 3/2-stereo-standard', www.hhton.de

Hamasaki, K., Fukada, A., Kamekawa, T. and Umeda, Y. (2000): 'A concept of multichannel sound production at NHK', *Proceedings of the 21st Tonmeistertagung.*

Hamasaki, K. (2003): 'Multichannel recording techniques for reproducing adequate spatial impression', *Proceedings of the 24th Audio Engineering Society International Conference.*

Hansen, V. and Munch, G. (1991): 'Making recordings for simulation tests in the Archimedes project', *Journal of the Audio Engineering Society*, 39, pp.768-774.

Hartmann, W. (1983): 'Localisation of sounds in rooms', *Journal of the Acoustical Society of America*, 74, pp.1380-1391.

Hartmann, W. (1993): 'Auditory localisation in rooms', Proceedings of the Audio Engineering Society 12th International Conference, pp.34-46.

Hartmann, W. and Rakerd, B. (1989): 'Localisation of sounds in rooms, IV: The Franssen effect', *Journal of the Acoustical Society of America*, 86, pp.1366-1373.

Hidaka, T., Beranek, 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*, 98, pp.988-1007.

Hiyama, K., Komiyama, S. and Hamasaki, K. (2002): 'The minimum number of loudspeakers and its arrangement for reproducing the spatial impression of diffuse sound field', Audio Engineering Society 113th Convention, Preprint 5674.

Howard, D. and Angus, J. (1996): *Acoustics and Psychoacoustics* (Oxford: Focal Press)

ITU-R (1993): 'Recommendations ITU-R BS.775-1: Multichannel stereophonic sound system with or without accompanying picture', *International Telecommunications Union.*

ITU-R (1994): 'Recommendations ITU-R BS.1116: Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems', *International Telecommunications Union.*

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.

Kjeldsen, A. (1998): 'The measurement of personal preference by repertory grid technique', Audio Engineering Society 104th Convention, Preprint 4685.

Klepko, J. (1997): '5-channel microphone array with binaural head for multichannel reproduction', Audio Engineering Society 103rd Convention, Preprint 4541.

Kuhl, W. (1978): 'Spaciousness (spatial impression) as a component of total room impression', *Acoustica*, 40, pp.167-181.

Kuhn, G. F. (1977): 'Model for the interaural time differences in the azimuthal plane', *Journal of the Acoustical Society of America*, 62, pp. 157-167.

Leakey, M. (1959): 'Some measurement of the effects of interchannel intensity and time differences in two channel sound systems', *Journal of the Acoustical Society of America*, 31, pp.977-986.

Lee, H.K. (2004): *M.Phil-Ph.D transfer report*, University of Surrey, England.

Martin, G., Woszczyk, W., Corey, J. and Quesnel, R. (1999): 'Sound source localisation in a five channel surround sound reproduction system', Audio Engineering Society 117th Convention, Preprint 4994.

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

Mason, R. (2002): 'Elicitation and measurement of auditory spatial attributes in reproduced sound', *Ph.D thesis*, University of Surrey, England.

Mason, R., Brookes, T. and Rumsey, F. (2004): 'Development of the interaural cross-correlation coefficient into a more complete auditory width prediction model', *Proceedings of the 18th International Congress on Acoustics*, pp. 2453-2456.

Mason, R., Brookes, T. and Rumsey, F. (2005a): 'The effect of various source signal properties on measurements of the interaural cross-correlation coefficient', *Acoustical Science & Technology*, 26, pp.102-113.

Mason, R., Brookes, T. and Rumsey, F. (2005b): 'Frequency dependency of the relationship between perceived auditory source width and the interaural cross-correlation coefficient for time-invariant stimuli', *Journal of the Acoustical Society of America*, 117, pp.1337-1350.

Mason, R., Brookes, T. and Rumsey, F. (2005c): 'Perceptually motivated measurement of spatial sound attributes for audio-based information systems', documentation on website www.surrey.ac.uk/soundrec/PMMP/

McFadden, D. and Pasanen, E.G. (1976): 'Lateralisation at high frequencies based on interaural time differences', *Journal of the Acoustical Society of America*, 59, pp.634-639.

Mckinnie, D. (2004): 'The influence of the factors 'spatial envelopment' and 'localisation accuracy' on perceived quality of sound reproduction", *Ph.D thesis*, University of Surrey, England

Mertens, H. (1965): 'Directional hearing in stereophony: Theory and experimental verification', *European Broadcasting Union Review. Part A*, 92, pp.1-14.

Mills, A. (1958): 'On the minimum audible angle', *Journal of the Acoustical Society of America*, 30, pp.237-246.

Morimoto, M. and Maekawa, Z. (1988): 'Effects of low frequency components on auditory spaciousness', *Acustica*, 66, pp.190-196.

Morimoto, M. and Iida, K. (1995): 'A practical evaluation method of auditory source width in concert halls', *Journal of the Acoustical Society of Japan* (English Translation), 16, pp. 59-69.

Morimoto, M. (2002): 'The relation between spatial impression and the precedence effect', *Proceedings of the 8th International Conference on Auditory Display*, pp.297-307.

Neher, T. (2004): 'Towards a spatial ear trainer', *Ph.D Thesis*, University of Surrey, England.

Okano, T., Beranek, L. and Hidaka.T. (1994): '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*, 104, pp.255-265.

Perrott, D. and Baars, B. (1974): 'Detection of Interaural onset and offset disparities', *Journal of the Acoustical Society of America*, 55, pp.1290-1292.

Perrott, D., Marlborough, K. and Merrill, P. (1988): 'Minimum audible angle thresholds obtained under conditions in which the precedence effect is assumed to operate', *Journal of the Acoustical Society of America*, 85, pp.282-288.

Rakerd, B. and Hartmann, W. (1985): 'Localisation of sound in rooms, II: The effect of a single reflecting surface', *Journal of the Acoustical Society of America*, 78, pp.524-533.

Rakerd, B. and Hartmann, W. (1986): 'Localisation of sound in rooms, III: Onset and duration effects', *Journal of the Acoustical Society of America*, 80, pp.1695-1706.

Ratliffe, P. (1974): 'Properties of hearing related to quadraphonic reproduction', BBC R&D 38.

Lord Rayleigh (1907): 'On our perception of sound direction', *Phil.Mag.*,13, 6th series, 214-232.

Rosenzweig, M. and Rosenblith, W. (1950): 'Some electrophysiological correlates of the perception of successive clicks', *Journal of the Acoustical Society of America*, 22, pp.878-880.

Rumsey, F. and McCormick, T. (1997): *Sound and Recording, an Introduction* (Oxford: Focal Press)

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

Schroeder, M., Gottlob, D. and Siebrasse, K. (1974): 'Comparative study of European concert halls: Correlation of subjective preference with geometric and acoustic parameters', *Journal of the Acoustical Society of America*, 56, pp.1195-1201.

Schubert, P. (1966): 'Wahrnehmbarkeit von Einzelruckwurfen bei Musik', *Electro-Acoustique*, 10, pp.39-44.

Simonsen, G. (1984): Master's Thesis. Technical University of Lyngby, Denmark.

Snow, W. (1953): 'Basic principles of stereophonic sound', *Journal of the Society of Motion Picture and Television Engineers*, 61, pp.567-589.

Soulodre, G., Lavoie, M. and Norcross, S. (2002): 'Investigation of listener envelopment in multichannel surround systems', Audio Engineering Society 113th Convention, preprint 5676.

Stone, H. and Sidel, J. (1993): *Sensory Evaluation Practice: 2nd Ed.* (Academic Press: New York)

Streicher, R. and Everest, F. A. (1998): *The New Stereo Soundbook, 2nd Ed.* (CA: TAB Books)

Theile, G. and Plenge, G. (1977): 'Localisation of lateral phantom images', *Journal of the Audio Engineering Society*, 25, pp.196-200.

Theile, G. (2000): 'Multichannel natural recording based on psychoacoustic principles', Audio Engineering Society 108th Convention, Preprint 5156.

Theile, G. (2001): 'Multichannel natural recording based on psychoacoustic principles', In *Proceedings of the Audio Engineering Society 19th International Conference*, pp.201-229.

Thurlow, W. and Parks, T. (1961): 'Precedence suppression effects for two-click sources', *Percept. Motor Skills*, 13, pp.7-12.

Tobias, J. and Zerlin, S. (1959): 'Lateralisation thresholds as a function of stimulus duration', *Journal of the Acoustical Society of America*, 31, pp.1591-1594.

Wallach, H., Newman, E. and Rosenzweig, M. (1949): 'The precedence effect in sound localisation', *American Journal of Psychology*, 52, pp.315-336.

Whitworth, R. and Jeffress, L. (1961): 'Time versus intensity in the localisation of tones', *Journal of the Acoustical Society of America*, 33, pp.925-929.

Williams, M. (1987): 'Unified theory of microphone systems for stereophonic sound recording', Audio Engineering Society 82nd Convention, Preprint 2466.

Williams, M. and Le Du, G. (1999): 'Microphone array analysis for multichannel sound recording', Audio Engineering Society 107th Convention, Preprint 4997.

Williams, M. and Le Du, G. (2000): 'Multichannel microphone array design', Audio Engineering Society 108th Convention, Preprint 5157.

Williams, M. (2003): 'Multichannel sound recording practice using microphone arrays', *Proceedings of the Audio Engineering Society 24th International Conference*

Williams, M. (2004): *Microphone arrays for stereo multichannel sound recording. Vol.1* (Segrate: Editrice Il Rostro)

Wittek, H. (2000): 'Untersuchungen zur richtungsabbildung mit L-C-R hauptmikrofonen', *Master's Thesis*, Institut für Rundfunktechnik, Germany.

Wittek, H. (2001a): 'Image assistant', JAVA Applet and documentation on website www.hauptmikrofon.de

Wittek, H. (2001b): 'Studies on main and room microphone optimisation', In *Proceedings of the Audio Engineering Society 19th International Conference*, pp.448-455.

Wittek, H. and Theile, G. (2002): 'The Recording Angles: based on localisation curves', Audio Engineering Society 112th Convention, Preprint 5568.

Yost, W., Wightman, F. and Green M. (1971): 'Lateralisation of filtered clicks', *Journal of the Acoustical Society of America*, 50, pp. 1526-1531.

Zacharov, N. and Koivuniemi, K. (2001): 'Unravelling the perception of spatial sound reproduction: Techniques and experimental design', *Proceedings of the Audio Engineering Society 19th International Conference*, pp.272-286.

Zurek, P. (1980): 'The Precedence effect and its possible role in the avoidance of interaural ambiguities', *Journal of the Acoustical Society of America*, 67, (March), pp.952-964.

PUBLICATIONS

Lee, H.K., Rumsey, F. (2004): 'Elicitation and Grading of Subjective Attributes of 2-Channel Phantom Images', Audio Engineering Society 116th Convention, Preprint 6142.

Lee, H.K., Rumsey, F. (2005): 'Investigation into the effect of interchannel crosstalk in multichannel microphone technique, Audio Engineering Society 118th Convention, Preprint 6374.

Kassier, R. Lee, H.K., Brookes, T., Rumsey, F. (2005): 'An informal comparison between surround microphone techniques', Audio Engineering Society 118th Convention, Preprint 6429.