Lee, Hyunkook and Rumsey, Francis

Investigation into the Effect of Interchannel Crosstalk in Multichannel Microphone Technique

Original Citation

Lee, Hyunkook and Rumsey, Francis (2005) Investigation into the Effect of Interchannel Crosstalk in Multichannel Microphone Technique. In: Audio Engineering Society 118th Convention, 28-31 May 2005, Barcelona, Spain.

Audio Engineering Society
Convention Paper 6374
Presented at the 118th Convention
2005 May 28–3 Barcelona, Spain

Investigation into the Effect of Interchannel Crosstalk in Multichannel Microphone Technique

Hyun-Kook Lee and Francis Rumsey

Institute of Sound Recording, University of Surrey, Guildford, Surrey, GU2 7XH, England
Correspondence should be addressed to Hyun-Kook Lee: tonmeister1025@yahoo.co.uk

ABSTRACT

A series of subjective listening tests were carried out in order to investigate the effect of interchannel crosstalk in multichannel microphone technique. Perceived attributes of interchannel crosstalk images were first elicited, and then graded with various independent variables, including different types of microphone array (different combinations of time and intensity differences), sound source and acoustic condition. The results showed that the most dominant effects of interchannel crosstalk were an increase of source width and a decrease of locatedness. The ratio of time and intensity differences in microphone array was the most significant factor for both effects. Sound source type had a significant effect for source width increase, but not for the locatedness decrease. Acoustic condition was significant for locatedness decrease, but not for source width increase. This paper describes the experiment method, and presents and discusses the details of the result data.

1. INTRODUCTION

This paper examines subjective effects of interchannel crosstalk in multichannel microphone techniques for classical music recordings. Since multichannel stereophonic audio systems such as the so-called 5.1 surround became popular in recent years, a number of multichannel microphone techniques have been proposed corresponding to the requirement of the new system. 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 surround channels delivering a sense of spatial impression. However, the addition of extra channels inevitably gives rise to a question about the effect of interchannel crosstalk. If a three-channel microphone technique is used for recording a single sound source located at a certain position, each pair of microphones (L-C, R-C and L-R) will pick up the sound with a different relationship of time and intensity differences.

The influence of interchannel crosstalk on the resulting image quality has been an issue for debate. When talking about interchannel crosstalk, there is the implicit assumption that signals from microphones other than the pair primarily covering the sector of the recording angle in which a source lies can be treated as unwanted 'crosstalk'. This suggests that two-channel stereophonic principles are being applied to the analysis of multichannel microphone techniques, which is, by and large, the principle on which the arrays discussed in this paper have been designed. Theile [1] asserted that the ideal localisation curve for a three channel microphone array should be linear within the recording angle, similarly to that for a conventional two channel array. In order to achieve this ideal, the sound source located in one side of the recording angle must not be picked up by the microphone in the other side. Although this is practically impossible, Theile [1] suggested that a reasonably balanced localisation could be achieved by reducing the intensity of the interchannel crosstalk signal as much as possible. His claim is that if the interchannel crosstalk is not suppressed enough, triple phantom images will be created in reproduction, thus leading to a decrease in image focus and clarity. Intending to optimise the operation of three channel microphone technique with respect to interchannel crosstalk, a microphone technique called OCT (Optimised Cardioid Technique) has been proposed. The details of this technique are found in [2].

Theile's hypothesis of the perception of three separate images was claimed in [3]. Rumsey asserted that the listener tends 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 the need for further experiments regarding the effect of interchannel crosstalk. In fact, there is no experimental evidence available to support the triple phantom image hypothesis.

Williams [4] disagrees about the significant effect of interchannel crosstalk that was claimed by Theile. His argument is that the interchannel crosstalk is already reduced to a great extent using directional microphones, and therefore is not a great matter for localisation. He seems to suggest that it is more important to link the recording angles of each stereo segment without overlap in order to obtain a balanced localisation performance, rather than achieving the maximum suppression of interchannel crosstalk. Williams and Le Du [5] proposed a microphone technique based on their so-called 'critical linking' concept.

Although there is much debate on the issue of so-called interchannel crosstalk as shown above, to date there is no conclusive answer as to whether it really matters or not. More importantly, there seems to be no research of which the authors are aware that investigates the subjective effects of interchannel crosstalk in such contexts. Therefore, it is not clearly known what kinds of interchannel crosstalk effects listeners perceive and how those factors affect the stereophonic image quality. This is why the current experiment was designed. It is expected that the results of this experiment will provide recording engineers with useful guidelines for the design and application of multichannel microphone techniques. The primary research questions for this experiment were formulated as follows:

- What subjective attributes are perceived as a result of interchannel crosstalk?
- How audible are these attributes in general?
- Does the subjective grading for these attributes depend on type of microphone array (combination ratio of time and intensity differences), type of sound source, or acoustic condition?
- Is there a significant interaction between each of these variables?

2. EXPERIMENTAL DESIGN

2.1. General Experimental Methodology

The basic concept for the method used in the current experiment was inspired by a test method known as QDA (Quantitative Descriptive Analysis). A basic QDA consists of three stages [6]. Firstly, a group of qualified subjects are presented with stimuli and develop descriptive terms on the attributes of the product through discussion. Secondly, the elicited terms are grouped into common attribute scales through discussion based on the similarity of meaning of the terms. Finally, the stimuli are graded using the obtained common attribute scales. The advantages of using this method would be firstly that any experimenter's bias on the scales to be graded can be avoided, and secondly that the data obtained from a number of subjects can be statistically analysed together. However, it was considered that a full QDA, according to methods proposed in the food sciences, would be an unnecessarily time consuming and detailed process that would require gathering all subjects at the same time and place for a series of group discussions. Therefore, the QDA method was modified in several ways so as to select only the parts relevant to the problem in hand and to adapt it to the current experimental context in sound recording. Firstly, instead of using a full elicitation process, subjects were asked to select relevant attributes from a set of potential attributes that were provided. They were also asked to describe additional attributes using their own terms in case they perceived any other differences than the ones provided. The elicited subjective terms were carefully interpreted and unified by the experimenter through later informal discussions with individual subjects on the meanings of the terms they used. Subjects were also asked to grade the magnitude of audibility of the selected attributes. This was in order to weight the perceptual dominance of those attributes and accordingly reduce the number of attribute scales to be graded. Finally, a grading test was conducted using the selected attribute scales and statistical tests were carried out on the obtained data. From the above descriptions, it can be said that the whole experiment consisted of two stages: attribute selection test and grading test. More detailed descriptions of the test methods are presented in sections 3.1 and 4.1.

2.2. Choice of Microphone Technique

2.2.1. Basic philosophy

To date, a number of microphone techniques have been proposed for the recording and reproduction of multichannel surround sound. Rumsey [3] suggested a way of classifying the design concepts of the current multichannel microphone techniques, based on the purpose of the rear channels in the technique. According to his classification, there are two main groups: those that use a 'five-channel main microphone technique' and those that use a 'technique with front and rear separation'. 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. In other words, the microphone techniques in this group attempt to provide a satisfying directional image and spatial impression at the same time, with a fixed pattern of microphone placement. The fixed positions and polar patterns of the front and surround microphones might result in an inevitable compromise between the representations of optimised directional image 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 surround microphones [2]. The position and the polar pattern of the surround microphone array should not only be decided 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 the rear microphones. With this technique, interchannel crosstalk will be am 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 surround microphones.

The other group, on the other hand, uses a frontal main microphone array that is used specifically for accurate pickup of direct sound so that sources can be easily localized on reproduction, together with a separate rear microphone array that is 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 desired directional and ambience characteristics [2]. The distance between the front and the rear arrays can vary depending on different recording situations. Interchannel crosstalk between the front and rear microphones would not be taken into account much with this type of technique because of the sufficiently long distance between them. In this regard, it seems that this type of technique gives recording engineers more freedom to control the spatial impression and enables them to use their artistic and technical creativity more than the 5-channel main microphone technique. For this reason, a technique with separate treatment of front and rear was chosen as the basis for this experiment.

2.2.2. Simulation of microphone technique

It was considered that 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 3-channel frontal microphone technique was needed. Even though understanding the effect of interchannel crosstalk in anechoic recording conditions was the primary aim of this research, 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 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 ambience picked up by a rear microphone array was simulated by using an artificial reverberator.

2.2.3. Frontal microphone technique

The frontal microphone technique chosen for this experiment was the so-called 'critical linking' 3 channel microphone technique, proposed by Williams and Le Du (detailed descriptions of this technique can be found in [5]). The basic design concept of this technique aims to achieve a continuous distribution of phantom images across channels L-C-R by linking the recording angles of each stereophonic segment C-L and C-R without overlap. Within one segment, the psychoacoustic laws for localisation in conventional 2 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 degrees 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 a crosstalk to the channels C and R. Ideally, L should not be taken into account in localisation process since it is to be suppressed by the same effect or the precedence effect operating between C and L. It is hypothesised that even though the position of the phantom image can be solely determined by C-R without the aid of L, the presence of L will influence the spatial or timbral quality of the image to some extent. Reported studies on the perceived differences between phantom images created by the precedence effect and their corresponding mono images could be the basis for this hypothesis (e.g. the phantom image having 'greater spatial extent' [7], 'image extended toward the echo source' [8] and 'fuller tonal colour' [9]). 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 L-C-R) and that with the crosstalk channel turned off (C-R only).

The critical linking technique supposedly enables one to create various array styles having different distances and angles between microphones while keeping the recording angle across L-C-R constant. Therefore, the effect of the ratio between time and intensity differences between the crosstalk and the other channels can be investigated by comparing different microphone arrays sharing the same recording angle. Williams provided various examples of critically linked microphone arrays that can be created while keeping the same recording angle. For the current experiment, only four representative arrays were selected from the examples, as shown in **Figure 2.1** to **Figure 2.4**. The recording angles for these arrays were all 180°. The simulated direction of the sound source was 45 degrees from the centre line of the array, and the distance from the centre point of the array was 5 metres. The interchannel time and intensity differences between L and C and those between R and C calculated for each array according to these are shown in **Table 2.1**. In a conventional 2 channel stereophonic reproduction, the minimum delay time required between left and right channels for operating the precedence effect is normally 1.1ms for natural sound sources ([10], [11] and [12]). According to Theile's hypothesis [2], in a three-channel reproduction, the same effect between centre and left (or right) channels can be achieved with only a half of this delay time (0.55ms). From this, it might be possible to assume that the delay times between C and L in all the arrays shown in **Table 2.1** are long enough to operate the precedence effect between C and L.

	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 2.1: Time and intensity differences of the left and right channels to the centre channel for each array

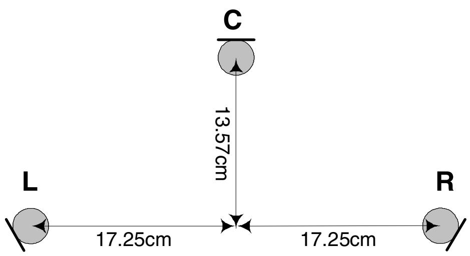


Figure 2.1: Configuration of microphone array 1: the angle between L and R is 100°.

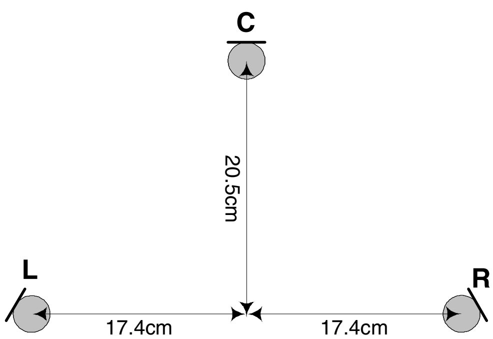


Figure 2.2: Configuration of microphone array 2: the angle between L and R is 80°.

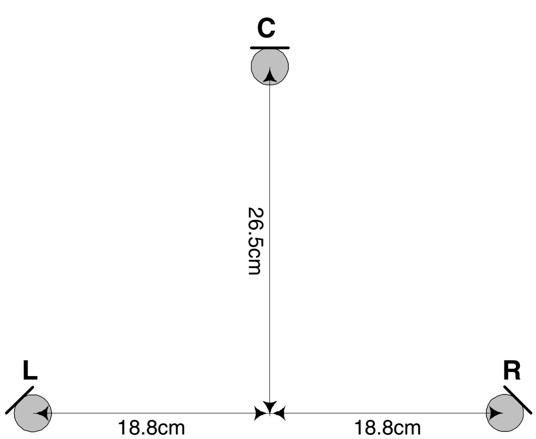


Figure 2.3: Configuration of microphone array 3: the angle between L and R is 60°.

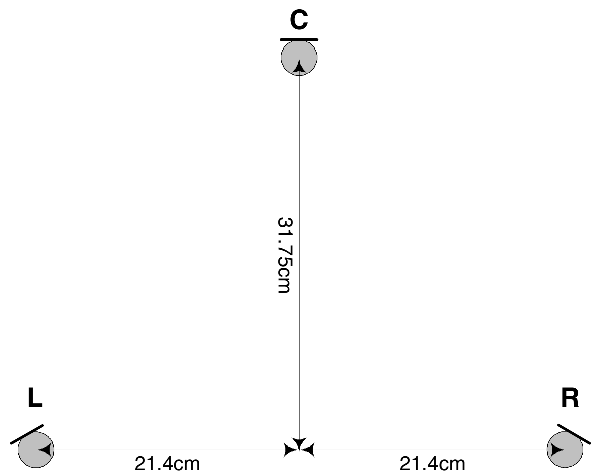
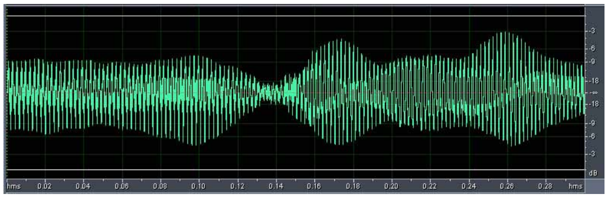


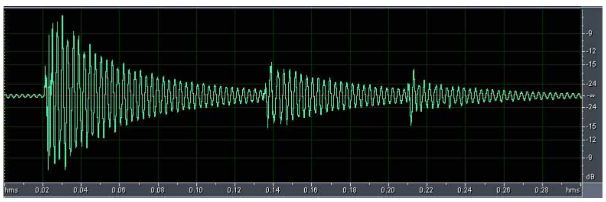
Figure 2.4: Configuration of microphone array 4: the angle between L and R is 40°.

2.3. Choice of Sound Source

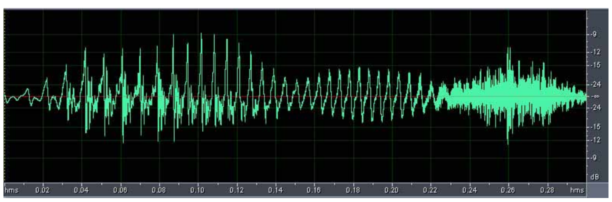
It was of interest to examine whether the effect of interchannel crosstalk depends on the use of different types of sound source. Three types of natural sound sources comprising cello, bongo and speech were chosen for this experiment due to their distinctive temporal and spectral characteristics, with 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 [13]. From a psychophysical view point, it might be claimed that the characteristics of natural sound source 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 on various aspects. However, the results obtained with strictly controlled stimuli might not be applicable to natural sound sources in the same manner because the characteristics of the latter are more complex, and 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 plot for each sound source are shown in **Figures 2.5** and **2.6**. The waveform shows temporal variations during specific 0.3 seconds taken from the whole performance, and the frequency analysis is a plot of the average intensity by frequency over the whole performance.



(a) Cello source

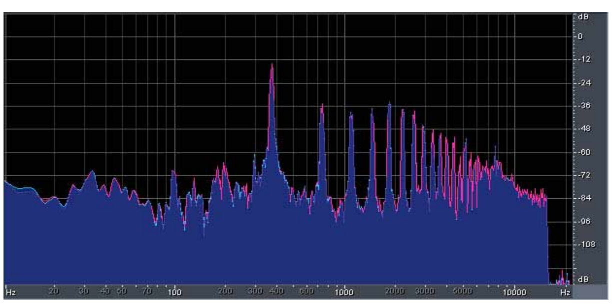


(b) Bongo source

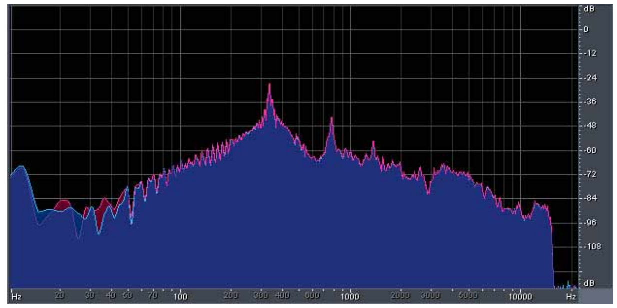


(c) Speech source

Figure 2.5: Extracts of waveforms of each sound source used for the experiment



(a) Cello source



(b) Bongo source

(c) Speech source

Figure 2.6: Long-term averaged frequency spectrum of each sound source used for the experiment

2.4. Acoustic Conditions

The acoustic conditions considered for 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 sounds 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. The room and hall conditions are chosen for their different acoustical characteristics. The detailed characteristics of the simulated room and hall conditions are described in the next section.

2.5. 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 2.7**. 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 2.1**. 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. Hiyama *et al* [14] investigated the number of loudspeakers required for the reproduction of optimum spatial impression of diffuse sound field. 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 4 loudspeakers, which were arranged in similar positions as the BS.775-1 recommendation, were required for listeners to perceive a similar spatial impression to the reference sound. For creating ambience 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 set up used for creating the room and hall ambience sounds are shown in **Table 2.2**. In general, the 'large room' set can be described as producing coloured and comb-filtered ambience sounds with slapping echoes. The 'large hall' creates ambience sounds that have longer reverberation time and are 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 2.2: Parameters of the 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 authors' aesthetic judgement as experienced balance engineers, and it was aimed 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.

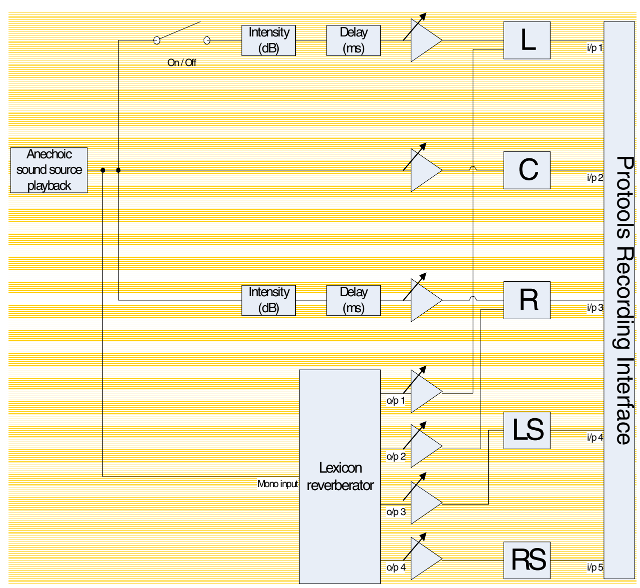


Figure 2.7: Diagram of signal processing for stimuli creation

2.6. Physical Setup

The experiment was conducted in an ITU-R BS.1116-compliant [15] listening room at the University of Surrey. According to the ITU-R BS.775-1 recommendation [16], five Genelec 1032A loudspeakers were set up at 0°, 30° and 100°, with a distance of 2m from the subject's seat. The levels of the loudspeakers were aligned to be equal, and the sound pressure level of all stimuli was calibrated at 75dB, A weighted, at the listening position. 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.

2.7. Test Subjects

A total of eight subjects took part in the experiment. All were experienced listeners, selected from staff members, research students and final year undergraduate students on the University of Surrey's Tonmeister course.

3. ATTRIBUTE SELECTION PROCESS

3.1. 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 the anechoic stimuli only was that they were considered to enable 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 had the freedom to control the playback of the stimuli. **Figure 3.1** shows the control interface used for this test, which was written using the MAX-MSP software. There were a total of six trial pages, and the buttons A and B in each page presented the images of C-R and L-C-R in random orders. The stimuli pair of A and B was synchronised and looped so that the subjects were able to switch between them freely, and to listen repeatedly.

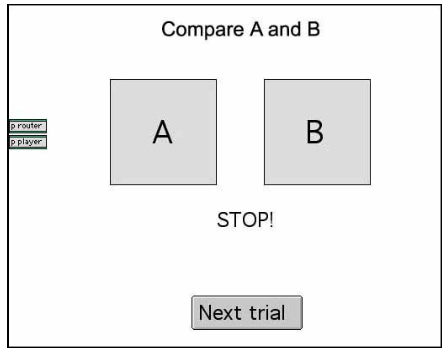


Figure 3.1: Layout of the control interface used for the pairwise comparison and elicitation of auditory attributes

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

1. To define the global set of auditory attributes for the perceived differences between the images of C-R and L-C-R

2. 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 discussed in section 1.1, 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 to be unified into the common terms by informal discussions between the subjects. The choice of the provided attributes was based on the results of the authors' previous experiment conducted to investigate the perceptible differences between monophonic and 2-channel stereophonic images [17]. In that experiment, a group of common attributes describing the perceived differences were elicited from subjects, and the effects of interchannel time difference, interchannel intensity difference and the type of sound source on the magnitude of those attributes were examined. The elicited attributes were three spatial and three timbral ones, comprising source width, source focus, source distance, brightness, hardness and fullness. A number of other spatial or timbral attributes are available to choose from various elicitation experiments ([18], [19] and [20]). However, due to the similarity of the experimental contexts, the attributes perceived between monophonic and 2-channel stereophonic attributes were considered to be the most appropriate basis for evaluating the differences between 2-channel (C-R) and 3-channel (L-C-R) stereophonic images. The choice of attributes and their definitions are shown in **Table 3.1**. For the attribute meaning the ease of localisation, the term 'source focus' from the result of the previous 2-channel experiment was replaced with 'locatedness' [21] 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 C-R and L-C-R in the authors' own informal 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 3.2** was used for the subjects to grade the magnitudes of audibility of the elicited attributes. The magnitude of audibility might vary for different stimuli, but the grading was to be made to the one having the greatest magnitude.

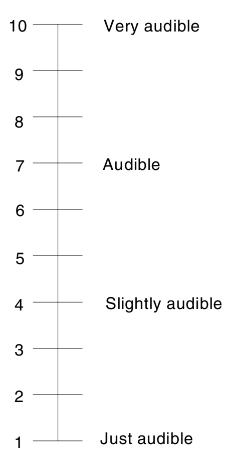


Figure 3.2: Scale used for grading the audibility of each attribute elicited

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 (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 3.1: Definitions of the auditory attributes provided to subjects for elicitation

3.2. Results and Discussions

As a result of the elicitation test, a total of eleven attributes were elicited from the subjects, comprising all of the seven provided attributes and four additional attributes. **Table 3.2** shows the attributes elicited, the number of their occurrences, and their audibility indexes. The audibility index represents the average magnitude 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' appears to be 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 audible more than 'slightly' 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 'slightly audible' range, were finally selected to be used for the next grading test.

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 3.2: Attribute group, number of occurrences and audibility index obtained for the differences perceived between the images of C-R and L-C-R with cello, bongo and speech sources.

4. GRADING PROCESS

4.1. Method

The grading test was conducted based on the result of the attribute selection experiment. This experiment was designed to enable subjects to grade the perceived magnitude of difference between the images of C-R and L-C-R.

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 similar due to a possible biasing effect between each other. 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.

There were a total of 36 stimulus-pairs created for comparison as described in section 2.5. 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 3 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 [6], 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.

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 rating scale was considered to be a more appropriate method since the data would be reliable for parametric statistical analysis due to the linearity of the scale, although the data would need to be normalised before statistical analysis because subjects might use different ranges of the scale. However, using a pure continuous rating scale without any labels, the subjects might have difficulties in maintaining consistency in testing through many trials individually. Therefore, for the scale used in this experiment, 7 point number labels from -30 to 30 were added to a classical continuous rating scale as guidelines for helping subject consistency. 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'.

The control interface written with MAX-MSP is shown in **Figure 4.1**. 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 images of C-R and L-C-R presented by the buttons 'A' and 'B' was randomised for each trial.

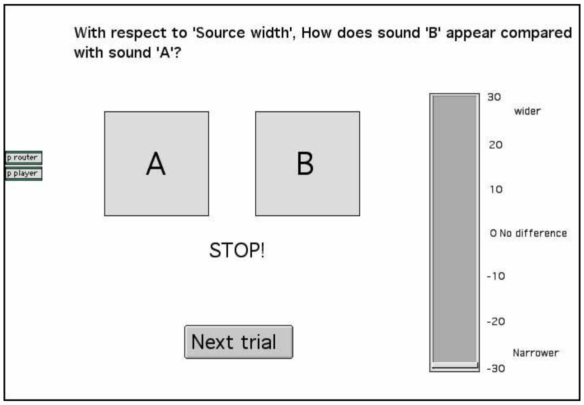


Figure 4.1: Layout of the control interface used for the pairwise comparison and grading for source width attribute

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 [6]. 6 representative stimuli comprising the extreme arrays of 1 and 4 combined with 3 sound sources were selected for the familiarisation trials.

4.2. Statistical Analysis

A repeated measure ANOVA (RM ANOVA) was carried out for statistical analysis of the data obtained from the grading test, 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 C-R and L-C-R on a scale of -30 to 30.

There were a total of 576 observations, consisting of 16 observations for each of the 36 acoustic-source-array combinations obtained from 8 subjects. 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 listener variability in use of the scale might cause inaccurate results from statistical analysis. Therefore, prior to the RM ANOVA test, the original grading data were normalised based on the ITU-R BS.1116 recommendation [15].

Tables 4.1 and **4.2** show the results of RM ANOVA for each attribute test. In the presentation of the results, each independent variable is termed 'acoustic', 'source' and 'array'. In order to interpret the results correctly, it was necessary to examine the 'assumption of sphericity' (equal variances of the differences between conditions) by using Mauchly's test of sphericity. **Tables 4.3** and **4.4** show the results of Mauchly's test for each attribute. 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 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.

4.3. Results and Discussions

4.3.1. Source width change

The results of an RM ANOVA test shown in **Table 4.1** 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 = 0.039). The acoustic*source interaction is shown to be insignificant (p = 0.714).

Figure 4.2 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 4.5**). This result suggests that the effect of interchannel crosstalk on source widening becomes greater as a more spaced microphone technique is used, in other words as the ratio of time difference to intensity difference increased. 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 grounds for this discussion might be the result of the previous work by the authors [17], showing that 2 channel stereophonic images were perceived to be wider compared to the corresponding monophonic images, and the magnitude of this effect became greater as the ratio of time difference to intensity difference was increased. This result might be explained by the effect of interaural time difference fluctuations (ITD fluctuations) on the perceived width of a source. Mason and Rumsey [22] undertook research into interaural time difference fluctuations (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 between each signal can determine the magnitude of ITD fluctuations. A larger interchannel time difference will cause a higher degree of decorrelation between the interaural signals, therefore a greater magnitude of ITD fluctuations, which also means a smaller degree of interaural cross correlation (IACC). In this regard, a spaced microphone technique will produce a wider phantom image than a coincident technique due to the difference in the magnitude of ITD fluctuations. This seems to hold true in the case of the current experimental conditions. The interchannel time difference between L and C in the arrays 1 to 4 varies from 0.5ms to 1.1ms. The longest delay time of the crosstalk signal L in the array 4 might have caused the largest change in the magnitude of ITD fluctuations between the ear input signals of C-R and L-C-R, thus leading to the largest source width change; whereas the shortest delay time in the array 1 caused the smallest change. However, since the microphone techniques used in this experiment are near-coincident techniques, it should be noted that it might not only be the interchannel time difference that contributed to the significant difference between microphone arrays, but also the interchannel intensity difference. It was shown in **Table 2.1** that the interchannel intensity difference between L and C decreases from -20.5dB to -4.6dB as the microphone array moves from 1 to 4. The decrease in the intensity difference between L and C means an increase in the intensity of crosstalk signal. Therefore, this result seems to suggest that the magnitude of source width change increases as the intensity of the crosstalk signal becomes greater. Since the crosstalk signal has a similar form to a single reflection, the grounds for this suggestion could be found in concert hall acoustics research reporting that source width increases as the intensity of single reflection becomes greater ([23], [24]). From looking at the mean plots shown in **Figure 4.2**, it may be possible to ignore the crosstalk effect in the arrays 1 and 2 since the magnitudes of differences for those arrays fall within only a 10% region of the whole scale.

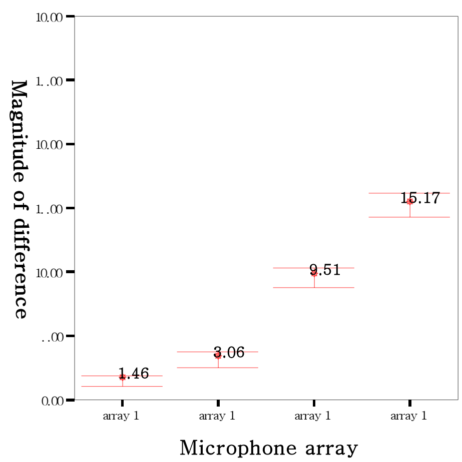


Figure 4.2: Mean value and the associated 95% confidence intervals of the grade of locatedness difference between the images of C-R and L-C-R separated by microphone array

The overall discussion might suggest that a more widely spaced 3-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 near-coincident microphone technique design requiring a trade-off between interchannel time and intensity differences. Conversely, it can also be suggested that in order to minimise the source width increasing effect of interchannel crosstalk in the design of 3-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. However, it is not yet known if this source widening effect of interchannel crosstalk can contribute to the improvement of perceived spatial image quality. Experiments studying the spatial impression in concert halls suggest that source widening by reflections is a desirable effect for improving the spatial quality of sound generated in the hall [25]. Regarding the crosstalk signal, L, as a single reflection to C and R could be a way of conceiving of the above question. However, those experiments mostly used much longer delay times of reflections (<80ms) than those of crosstalk signals that might be encountered in general microphone techniques, and therefore preference data obtained in a multichannel stereophonic reproduction are required. The subjective preference of the effect of interchannel crosstalk will be investigated in a future experiment.

Figure 4.3 shows the main effect of sound source on source width change between C-R and L-C-R. It appears that the speech source is outstanding compared to the cello and bongo sources. The multiple comparisons between each sound source indicated in **Table 4.6** 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). It seems to be the spectral content that caused the significant difference between speech and the other sources. The effect of spectral content in the signals on source widening has been widely investigated in the research related to concert hall acoustics ([24], [26] and [27]). Barron and Marshall [24] considered that source widening is mainly governed by middle frequencies around 1000 – 2000Hz in the signals while 'envelopment' is related to low frequencies. Hidaka *et al* [26] found that the strength factors *G* below 355Hz (*G*L : sound pressure level of the sound filed at low frequencies) increases source width in a concert hall. The importance of low frequency content in source width was also reported in [27]. In Morimoto and Maekawa's experiment comparing the source width of noise signals varying in lower cut-off frequency and IACC (interaural cross-correlation), it was found that keeping IACC equal, source width increased as the cut-off frequency decreased below 510Hz and frequency contents around 100 – 200Hz resulted in an especially remarkable increase of source width. These findings seem to provide a reasonable explanation of the current result. As can be seen in **Figure 2.6**, speech has greater low frequency energies especially at around 100 – 200Hz compared to the other sources, and this seems to be the main factor for the greater source width increase of speech compared to the others, based on the findings of [27].

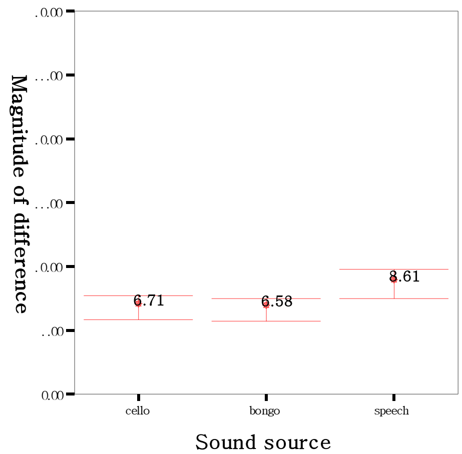


Figure 4.2: Mean value and the associated 95% confidence intervals of the grade of locatedness difference between the images of C-R and L-C-R separated by sound source

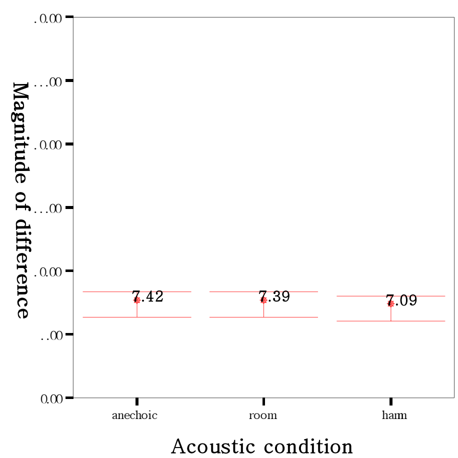


Figure 4.4: Mean value and the associated 95% confidence intervals of the grade of locatedness difference between the images of C-R and L-C-R separated by acoustic condition

The main effect of the acoustic condition on source width change is shown in **Figure 4.4**. Adding multiple reflections and reverberation to an anechoic sound might have increased the source widths for both images of C-R and L-C-R. 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 from the acoustic condition of recording space.

The source*array interaction is shown in **Figure 4.5**. 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, considering the estimated effect size is only 0.351, 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 keeps the same regardless of the acoustic condition (see **Figure 4.6**). Therefore, this interaction could be also ignored. The acoustic*source interaction, which was found to be insignificant is also shown in **Figure 4.7**.

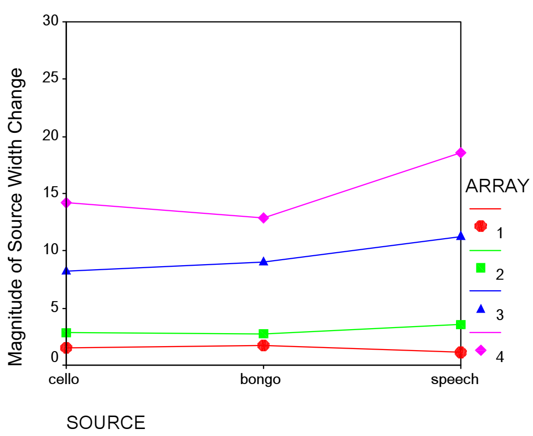


Figure 4.5: Interaction between microphone arrays and sound sources

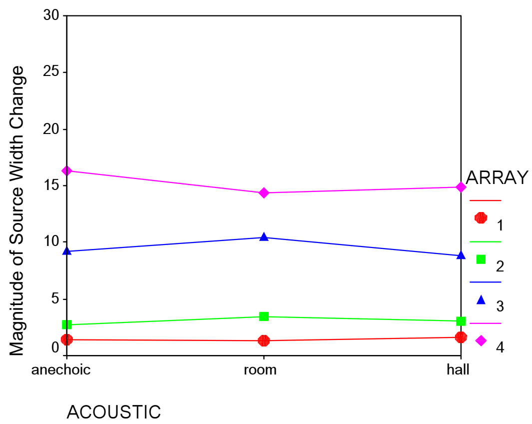


Figure 4.6: Interaction between microphone arrays and acoustic conditions

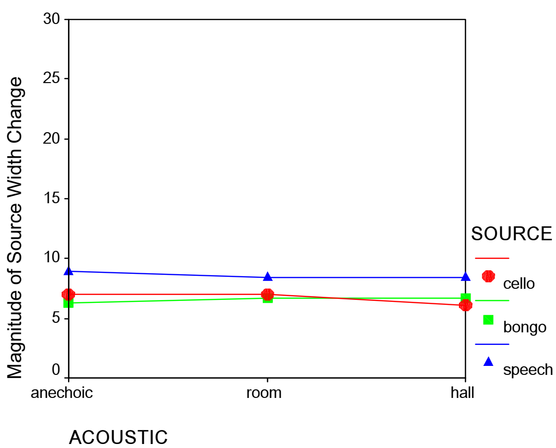


Figure 4.7: Interaction between sound sources and acoustic conditions

4.3.2. Locatedness change

Taking an overview of the results of the RM ANOVA test indicated in **Table 4.2**, 'microphone array' shows the most significant main effect to be a 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 of acoustic*array interaction can be judged differently depending on which corrected significance value is used because sphericity is violated (see **Table 4.4**). 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 magnitude of the significance value. The source*array interaction is shown to be insignificant (p = 0.058).

Figure 4.8 shows the mean value and associated 95% confidence intervals of the grade made for each microphone array. It can be firstly seen that the magnitude of locatedness change between the images of C-R and L-C-R 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 be also 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 4.7** (all p values are 0.000). This result suggests that the ratio of interchannel time and intensity differences in microphone technique is an important factor governing the perception of locatedness decrease resulting from interchannel crosstalk.

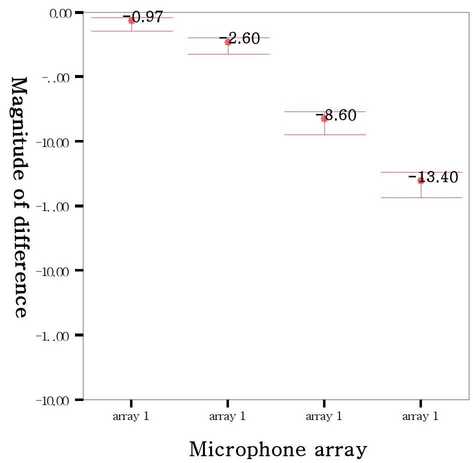


Figure 4.8: Mean value and the associated 95% confidence intervals of the grade of locatedness difference between the images of C-R and L-C-R separated by microphone array

The main effect graph for each acoustic condition is shown in **Figure 4.9**. 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 4.8** indicates that the only significant difference is between the anechoic and hall conditions (p=0.003). In other words, the perceived magnitude of locatedness change was significantly smaller in the hall condition than in the anechoic condition. This is likely to be because the effect of crosstalk was diminished in the hall condition due to the large influence of multiple reflections and reverberation in both images of C-R and L-C-R. This finding might lead to the hypothesis that the effect of crosstalk on locatedness change would become less audible in a more diffused recording space.

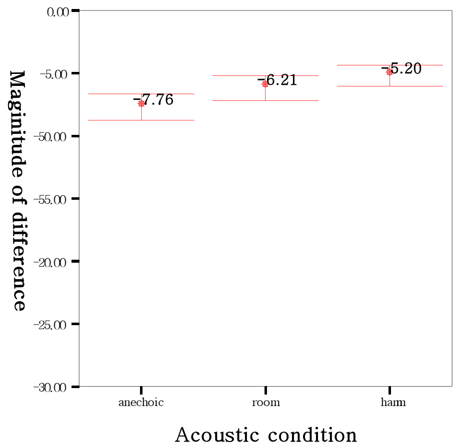


Figure 4.9: Mean value and the associated 95% confidence intervals of the grade of locatedness difference between the images of C-R and L-C-R separated by acoustic condition

The mean values and associated 95% confidence intervals of the normalised data for each sound source are shown in **Figure 4.10**. As can be seen, all sound sources have small differences in mean values and large overlaps in 95% confidence intervals. Many authors confirmed that a transient sound source is more important for operating the precedence effect than a continuous sound source ([28], [29] and [30]). Therefore, one may presume that the continuous nature of cello source would cause a greater locatedness decrease in the image of L-C-R than the transient nature of bongo source would. However, it should be noted that the characteristics of sound sources used in this experiment are different from those used in the classical studies on the importance of transient sound in localisation. That is, the latter used pure tones, while the former used natural sound sources having complex frequency spectra and waveforms as shown in **Figures 2.5** and **2.6**. It appears that all sound sources have sufficient transient information to retrigger the precedence effect. For example, the speech source has a fine structure of ongoing transients at every syllable. The cello source also has a continuous musical phrase containing ongoing fluctuations at every note or bow change. Every hit in the bongo source contains a rapid onset transient. Rackerd and Hartmann [30] pointed out that in the case of a complex signal such as noise, the precedence effect could be operated by continuous sounds also. Furthermore, Tobias and Zerlin [31] found that for noise signal, the continuous part became more influential on localisation than the onset transient as the duration of the signal increased. From the above literature, it is assumed that the series of ongoing fluctuations contained in the continuous cello source were strong enough to generate sufficient interaural time differences for retriggering the precedence effect, as well as the strong and rapid transients in the bongo source.

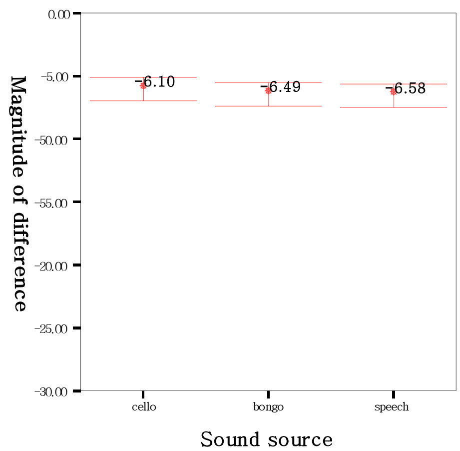


Figure 4.10: Mean value and the associated 95% confidence intervals of the grade of locatedness difference between the images of C-R and L-C-R separated by sound source

Figure 4.11 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). In terms of the interaction graph, 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 4.9**. 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. This means that in the anechoic condition, the locatedness decrease is significantly greater in bongo and speech compared with that of cello, but this significance disappears in the room and hall conditions. This result might be explained by the following assumption: the strong transient cues in the bongo and speech signals might have led to a weaker echo suppression effect than the relatively moderate transient cues in the cello signal. This assumption might be supported by Babkoff and Sutton [32]'s finding that if the intensity of a reflection is raised, a perceivable echo appears at a shorter delay time.

Figure 4.12 shows the acoustic*array interaction graph. Array 3 and array 4 have a significant difference when the room and hall conditions are compared. Also, array 2 and array 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 crosstalk signal does not change regardless of the acoustic condition of recording space.

The source*array interaction is shown in **Figure 4.13**. There was no significant interaction between sound source and microphone array. The cello and bongo in array 1 and array 2 are significantly different, but the experimental effect is minor.

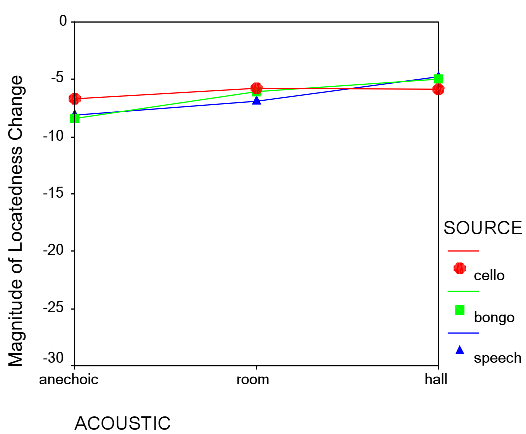


Figure 4.11: Interaction between sound sources and acoustic conditions

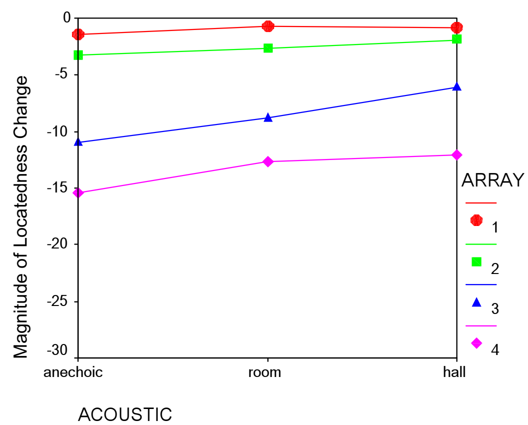


Figure 4.12: Interaction between microphone arrays and acoustic conditions

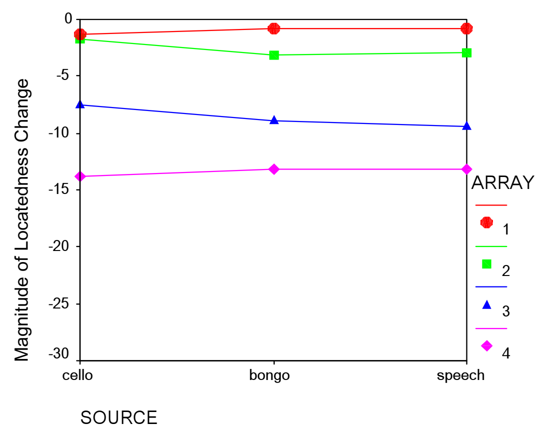


Figure 4.13: Interaction between microphone arrays and sound sources

4.3.3. Relationship between source width change and locatedness change

Table 4.12 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 in 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. It seems that the source width and locatedness attributes are often regarded as being negatively correlated. For example, in Berg and Rumsey's research [33], the 'source width' and 'localisation', although a different term was used for the same definition, were found to be negatively correlated at a moderate level.

However, the differences found between the locatedness and source width in the current experiment led to a hypothesis that the correlation between those attributes depends on sound source and acoustic condition. Therefore, a set of bivariate correlation tests were 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 4.13**. In general this result suggests that with respect to the effect of interchannel crosstalk in a microphone technique, a great source width increase by interchannel crosstalk does not necessarily mean a great locatedness decrease, or vice versa.

	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 4.12: Summary of significance values of the main effects and interaction effects for locatedness and source width changes

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

Table 4.13: Correlation value between locatedness change and source width change separated by microphone array

5. SUMMARY AND CONCLUSIONS

A series of subjective experiments were conducted in order to investigate the effect of interchannel crosstalk in multichannel microphone techniques. 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. The experiment employed two processes of attributes selection and grading, and was designed in order that subjects compared the perceptual differences between images with crosstalk and crosstalk-free images. The audible attributes of crosstalk images 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. The main findings obtained from the experiments are as follows.

1. 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, phasiness.

2. The source width and locatedness were found to be the only attributes that were audible more than 'slightly'.

3. In general, the interchannel crosstalk caused increase in source width and decrease in locatedness.

4. Statistically, the magnitudes of both source width increase and locatedness decrease significantly depended on the combination ratio of interchannel time and intensity differences in 3-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.

5. Sound source type was a significant factor for the source width effect but not for the locatedness effect.

6. Acoustic condition had a significant effect on the locatedness decrease, but not on the source width increase.

7. Interactions between microphone array type and sound source type, and between microphone array and acoustic array 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.

8. Interaction between sound source type and acoustic condition was significant for the locatedness effect, but not for the source width effect. The experimental effect for this interaction was very small, thus can probably be ignored.

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

6. ACKNOWLEDGEMENT

The authors wish to thank the staff and students at the University of Surrey's Tonmeister course who participated in the experiment, and Yeon-Koo Hong for helping with statistical analysis.

7. REFERENCES

[1] Theile, G. (2000): 'Multichannel Natural Recording Based on Psychoacoustic Principles', *108th Audio Engineering Society Convention*, Preprint 5156.

[2] Theile, G. (2001): 'Multichannel natural recording based on psychoacoustic principles', *Audio Engineering Society 19th International Conference.*

[3] Rumsey, F. (2001): *Spatial Audio* (Oxford: Focal Press)

[4] Williams, M. (2003): 'Multichannel sound recording practice using microphone arrays', *24th Audio Engineering Society International Conference.*

[5] Williams, M. and Le Du, G. (1999): 'Microphone Array Analysis for Multichannel Sound Recording', *107th Audio Engineering Society Convention*, Preprint 4997.

[6] Stone, H. and Sidel, L. (2004): *Sensory Evaluation Practices* (London: Academic Press)

[7] Freyman, R.L., Clifton, R.K., Litovski, R.Y. (1991): 'Dynamic process in the precedence effect', *Journal of the Acoustical society of America,* 90, pp.874-884.

[8] Perrott, D.R., 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.

[9] Streicher, R. and Everest, F. A. (1998): *The New Stereo Soundbook, 2nd Ed.* (CA: TAB Books).

[10] Simonsen, G. (1984): Master's thesis, Technical University of Lyngby, Denmark.

[11] Wittek, H. (2001): 'Directional imaging using L-C-R microphones: theoretical and practical investigations', *Audio Engineering Society 19th International Conference.*

[12] Lee, H.K. (2004): *M.Phil-Ph.D Transfer Report*, University of Surrey, England.

[13] 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.

[14] 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', *113th Audio Engineering Society Convention*, Preprint 5674.

[15] ITU-R BS.1116 (1994): 'Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems'. *International Telecommunications Union.*

[16] ITU-R BS.775 (1993): 'Multichannel stereophonic sound system with or without accompanying picture'. *International Telecommunications Union.*

[17] Lee, H.K. and Rumsey, F. (2004): 'Elicitation and grading of subjective attributes of 2-channel phantom images', *116th Audio Engineering Society Convention*, Preprint 6142.

[18] Berg, J. and Rumsey, F. (1999): 'Spatial attribute identification and scaling by repertory grid technique and other methods', *Audio Engineering Society 16th International Conference*.

[19] Zacharov, N. and Koivuniemi, K. (2001): 'Unravelling the perception of spatial sound reproduction: Analysis & external preference mapping', *111th Audio Engineering Society Convention*, Preprint 5423.

[20] Gabrielsson, A. and Sjogren, H. (1979): 'Perceived sound quality of sound-reproducing systems', *Journal of the Acoustical Society of America*, 65, pp.1019-1033.

[21] Blauert, J. (1997): *Spatial Hearing. The psychophysics of Human Sound Localisation* (Cambridge, Massachusetts, London: MIT Press)

[22] Mason, R. and Rumsey, F. (2001): 'Interaural time difference fluctuations: their measurement, subjective perceptual effect, and application in sound reproduction', *Audio Engineering Society 19th International Conference*.

[23] Keet, W. (1968): 'The Influence of Early Lateral Reflections on the Spatial Impression', *6th International Congress on Acoustics*, pp E53 to E56.

[24] Barron, M. and Marshall, A. (1981): 'Spatial impression due to the early lateral reflections in concert halls: the derivation of a physical measure', *Journal of Sound and Vibration,* 77, pp.211-232.

[25] Barron, M. (1971): 'The subjective effect of first reflections in concert halls – The need for lateral reflections', *Journal of Sound and Vibration*, 15, pp.475-494.

[26] 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 of concert halls', *Journal of Acoustical Society of America*, 98, pp.988-1007.

[27] Morimoto, M, and Maekawa, Z. (1988): 'Effects of low frequency components on auditory spaciousness', *Acustica*, 66, pp.190-196.

[28] Wallach, H., Newman, E.B. and Rosenzweig, M.R. (1949): 'The precedence effect in sound localisation', *American Journal of Psychology*, 52, pp.315-336.

[29] Yost, W.A., Wightman, F.L. and Green M.D. (1971): 'Lateralisation of Filtered Clicks', *Journal of the Acoustical Society of America*, 50, pp. 1526-1531.

[30] Rackerd, B. and Hartmann, W.M. (1986): 'Localisation of sound in rooms, III: Onset and duration effects', *Journal of the Acoustical Society of America*, 80, pp.1695-1706.

[31] Tobias, J.V. and Zerlin, S. (1959): 'Lateralisation thresholds as a function of stimulus duration', *Journal of the Acoustical Society of America*, 31, pp.1591-1594.

[32] Babkoff, H. and Sutton, S. (1969): 'End point of lateralisation of dichotic clicks', *Journal of the Acoustical Society of America,* 39, pp.87-102.

[33] Berg, J. and Rumsey, F. (2002): 'Validity of selected spatial attributes in the evaluation of 5-channel microphone techniques', *112th AES Convention, Munch*, Preprint 5593

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-Geisser	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-Geisser	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-Geisser	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-Geisser	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-Geisser	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-Geisser	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-Geisser	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-Geisser	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-Geisser	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-Geisser	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-Geisser	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-Geisser	1168.601	39.948	29.253			
	Huynh-Feldt	1168.601	49.378	23.667			
	Lower-bound	1168.601	15.000	77.907			

Table 4.1: Result table 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-Geisser	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-Geisser	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-Geisser	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-Geisser	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-Geisser	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-Geisser	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-Geisser	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*SOURCE)	Sphericity Assumed	843.229	60	14.054			
	Greenhouse-Geisser	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-Geisser	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*ARRAY)	Sphericity Assumed	1433.750	90	15.931			
	Greenhouse-Geisser	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-Geisser	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-Geisser	1036.302	53.615	19.329			
	Huynh-Feldt	1036.302	72.454	14.303			
	Lower-bound	1036.302	15.000	69.087			

Table 4.2: Result table of repeated measure ANOVA test for locatedness change

Measure: MEASURE_1

Within Subjects Effect	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 4.3: Mauchly's test of sphericity for source width change

Measure: MEASURE_1

Within Subjects Effect	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 4.4: Mauchly's test of sphericity for locatedness change

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 4.5: Result of multiple pairwise comparison between each microphone array for source width change

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 4.6: Result of multiple pairwise comparison between each sound source for source width change

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 4.7: Result of multiple pairwise comparison between each microphone array for locatedness change

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 4.8: Result of multiple pairwise comparison between each acoustic condition for locatedness change

Paired Samples Test

		Sig. (2-tailed)
Pair 1	anechoic+cello - anechoic+bongo	.016
Pair 2	anechoic+cello - anechoic+speech	.048
Pair 3	anechoic+bongo - anechoic+speech	.678
Pair 4	room+cello - room+bongo	.611
Pair 5	room+cello - room+speech	.145
Pair 6	room+bongo - room+speech	.250
Pair 7	hall+cello - hall+bongo	.241
Pair 8	hall+cello - hall+speech	.125
Pair 9	hall+bongo - hall+speech	.686
Pair 10	anechoic+cello - room+cello	.192
Pair 11	anechoic+cello - hall+cello	.132
Pair 12	room+cello - hall+cello	.892
Pair 13	anechoic+bongo - room+bongo	.000
Pair 14	anechoic+bongo - hall+bongo	.000
Pair 15	room+bongo - hall+bongo	.131
Pair 16	anechoic+speech - room+speech	.171
Pair 17	anechoic+speech - hall+speech	.000
Pair 18	room+speech - hall+speech	.010

Table 4.9: Result of paired T-test for acoustic condition and sound source