An Investigation into Stereo Microphone Techniques and their Applications

Faculty of Computing, Engineering and Technology

Thomas Rawlinson

07001511

BSc (Hons) Music Technology

Supervisor: Charles Walker

May 2011

Abstract

There are a number of methods for creating a stereo recording, this report investigates and compares how a number of different stereo microphone techniques can be used to create a complete stereo field.

The report includes research from a number of sources examining how the human hearing system is able to identify the location of a sound source; through time and level differences between the two ears and the unique interaction of the waves with the listener's head and torso. A range of coincident, near coincident and spaced pair techniques were researched to compare how they captured and encoded the time and level differences caused by different sound source locations.

Test recordings were conducted to confirm the spatial cues produced by the various arrays matched the predictions made by formulae, using trigonometry to calculate the difference in path length. It was confirmed that, by varying the spacing and angles between microphones, the width and spacing of the stereo image could be adjusted to suit the recording. These tests also highlighted the shadowing, image distortion and difficulty in positioning the various techniques that occurs when too many microphones are setup simultaneously. Five stereo arrays were selected to create the final small ensemble recordings, each producing satisfactory, yet varied results confirming the research suggestion that no one microphone technique will be suitable for every situation. The recordings were subject to a small scale listening test, which concluded that; for this ensemble and recording environment, the Blumlein array produced the most pleasing recording with a wide stereo spread stretching from ear to ear and providing accurate imagery for each of the 4 performers.

Contents

Table of Figures

Acknowledgments

Thanks go to the following people and organisations for their assistance with the production of this report and the accompanying recording.

Charles Walker- For the supervision, direction and advice given throughout the course of the degree and this project.

Modern World Studios- For the opportunities and experience gained during placement and the inspiration for this project.

Stafford Performing Arts Centre Chamber Ensembles- For providing two excellent ensembles for recording, they were a pleasure to work with and incredibly talented.

Holly Clarke and **Joel Rawlinson-** For acting as assistant engineers during the recording sessions, without them it would have been impossible to set everything up in the short time available.

Throughout this document there are links to audio and video samples which help to explain and illustrate the points made in the text. To hear these clips ctrl/cmd-click on the links to open them. Alternatively, the files can be found labelled and indexed on the accompanying CD under the ‘Appendix Media’ folder.

Introduction

Early recordings, from the beginning of the last century, contained no spatial elements. Having been recorded in mono and reproduced over a single speaker it was impossible to create a sense of direction for elements within the recording. In 1881, at the Paris exhibition of electricity, Clement Ader used a pair of microphones to capture and transmit a performance of the Paris Opera to listeners 3km away. When listening over a pair of earpieces, listeners were given the '*impression that they too were sitting in the Paris Opera'* (Brice, 2001a). Although this demonstration illustrated the possibilities for realistic recreation of a source using two channel recording and reproduction it was many years before commercial value was seen in the technique.

In 1931 Alan Blumlein developed a method to encode two simultaneous channels of audio to vinyl allowing for future experimentation into methods of creating stereo recordings. Many microphone techniques have since been developed for capturing a sound source with directional cues encoded into the signal, however many stereo recordings are also created using close microphone techniques, taking advantage of 'pan pot stereo' to infer spatial cues for elements of the recording. Since then stereo recordings have become the standard release format, with many tracks now also released in multi-channel surround formats.

This project aims to investigate a number of stereo microphone techniques, using this research to create a stereo recording of a classical chamber group and comparing the results from the different recording techniques to select the most suitable stereo array for this scenario. To accomplish this, different stereo recording techniques will be researched and investigated to ascertain how they capture and encode a stereo signal and how this compares to the methods used by the human hearing system to discern the location of a sound source within a live environment. As the playback method used will affect the spatial image the different options should be considered and researched before selecting an appropriate playback method and encoding format. The final recording, designed from the products of research & experiments, will be assessed through small scale listening tests with listeners selected for their ability to make informed observations of the recordings spatial, spectral and artistic properties.

Glossary of Terms

ITD: (Inter Aural Time Delay) the time delay between the same sound arriving at each ear, caused by the extra distance travelled around the head.

ICTD: (Inter Channel Time Delay) the time delay between the same sound across two channels of audio, caused by spacing between microphones or artificially introduced delay.

ILD: (Inter Aural Level Difference) the level difference between the two ears for the same sound, caused by the shadowing effect of the head for the further ear.

ICLD: (Inter Channel Level Difference) the level difference between two channels of audio, caused by angling directional microphones apart or introduced artificially.

HRTF: (Head Related Transfer Function) the unique frequency response for every point around a listener caused by the combined interaction of the source with the listeners head, torso, pinnae and also the effect of ITD and ILD.

1. Research

1.1 How Humans Localise Sound

Humans localise sound through external cues creating level differences and time delays between the sounds arriving at each ear, allowing them to infer a direction to, and distance from the source. The ear closest to the sound source will detect a greater change in pressure than the other, however a sound source directly in front or behind the listener will give equal pressure at each ear. If the listener cannot initially ascertain a direction then we will 'automatically turn our head through about 45 degrees to point one ear in the *believed direction, our ear's most sensitive reception angle.'* (Bearment, 2001a). When the head is turned the far ear is effectively "shadowed" by the head, which absorbs much of the high frequency content. Due to the wavelengths and the size of the head lower frequencies will diffract around the head, therefore the direction of a source is discerned from the level differences between high frequency sounds.

The human hearing system isn't designed to distinguish between small level differences in isolation, as this doesn't occur naturally. In the real world, level differences are caused by one ear being closer to the source than the other, therefore introducing a time delay as well as the level difference. Bearment (2001b) concludes that

> *'If sound comes directly towards the head, it arrives at the two ears at exactly* the same time. As the head is rotated, it takes slightly longer to get to the ear further from the sound source and the time difference increases until, when the sound is arriving from one side, it takes the maximum amount of time to get to the other ear. Rotate the head further and the time gap decreases, to be zero again when the sound source is directly behind the head.'

This explains why neither differences in time or level indicates if a sound source is in front or behind. To differentiate between front and back, the reflections from the pinna, head and torso are used. When listening to a stereo source over loudspeakers the sound can be clearly placed in front of the listener however, when using headphones the reflections of the pinna are by-passed, creating ambiguity as to the position of the sound source although the listener will usually position the source in front of them, as that is where the sound is expected to come from. The combination of reflections provided by the pinnae, head and torso form the head related transfer function or HRTF. Every source position and angle of incidence around the listener will be have a unique affect on the frequency response perceived by the listener due to the unique shape of their head and torso.

These cues can be artificially created to re-position previously recorded sounds within the stereo field around the listener however, as the time and level differences used are increased they become noticeable as artificial, destroying the effect you are trying to create.

1.1.1 How Microphones Localise Sound

'A microphone is a device for turning acoustic energy, in the form of sound, into electrical *energy'* (Brice 2001b). A dynamic microphone converts the movement of a diaphragm, caused by the pressure wave, into a small voltage by moving a magnet attached to the diaphragm in and out of a coil of wire. They are more durable than other types of microphones but also less sensitive.

Capacitor microphones use sound pressure waves to move a diaphragm, connected to one plate of a capacitor. The movements of the diaphragm vary the distance between the two plates, which creates small changes in voltage that are then amplified within the microphone before being amplified by the microphone pre-amp. While capacitor microphones are more delicate than dynamic microphones and require a phantom power supply they have a wider frequency response, a lower self-noise and an increased sensitivity that makes them suited for use in stereo arrays.

'Microphones differ in the way they respond to sounds coming from different directions. Some respond the same to sounds from all directions; others have different output levels for sources at different angles around the microphone.*'* (Bartlett, 1991a) this can be shown on a plot of microphone output level against angle, known as the microphones polar pattern (see figure 1. RØDE Microphones). As this polar pattern shows, microphone sensitivity varies based upon frequency, becoming increasingly directional at higher frequencies hence why off axis sound sources tend to sound 'darker' as the microphone is less sensitive to high frequency content.

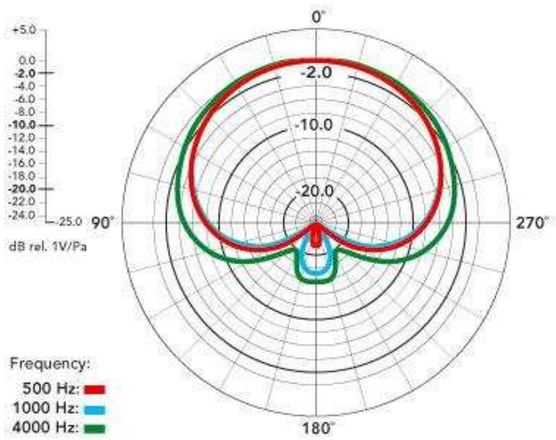


Figure 1 Rode NT1A Polar Pattern

http://www.rodemic.com/images/mics/nt1-a_polar.jpg

The three major classifications of polar patterns are omnidirectional, unidirectional and bidirectional.

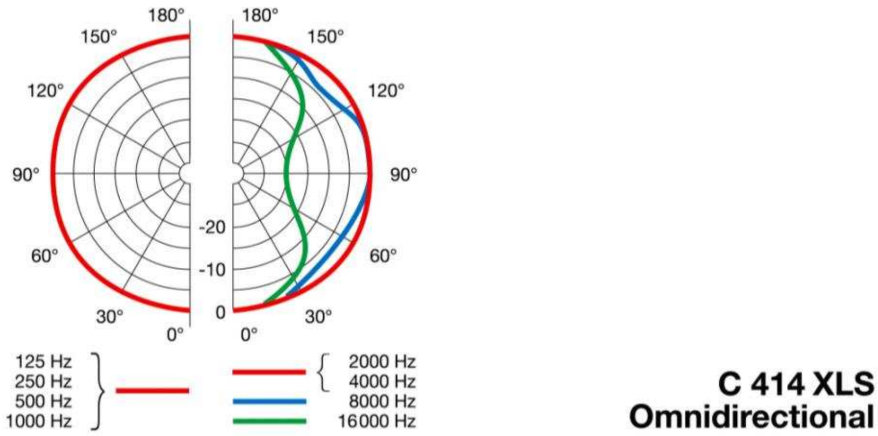


Figure 2 AKG C414 Omnidirectional Polar Pattern

http://www.akg.com/site/products/powerslave,id,1128,pid,1128,nodeid,2,_language,EN,view,specs.html

Omnidirectional microphones are equally sensitive to sounds arriving from all directions and are often pressure microphones as ‘pressure is a scalar quantity’ (Brice 2001b) meaning it has no directional component, allowing the microphone to theoretically be non-directional. However, the housing will have an affect, especially at higher frequencies. Omnidirectional microphones offer much better low frequency response than other polar patterns and will also capture more of the natural reverberation within the room.

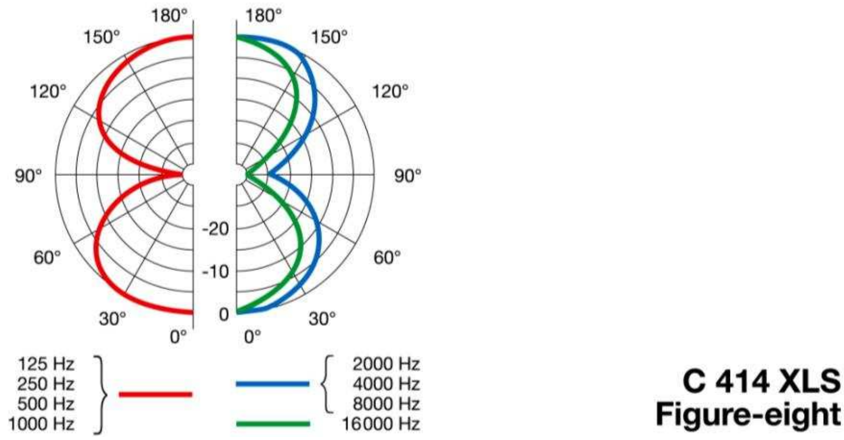


Figure 3 AKG C414 Bi-directional Polar Pattern

http://www.akg.com/site/products/powerslave,id,1128,pid,1128,nodeid,2,_language,EN,view,specs.html

Bidirectional microphones are sensitive to sound approaching from directly in front or behind but reject sound from the sides.

Unidirectional microphones are most sensitive to sounds coming from directly in front of the microphone and will discriminate against sounds entering the rear or sides of the microphone. The unidirectional polar pattern can be further subdivided.

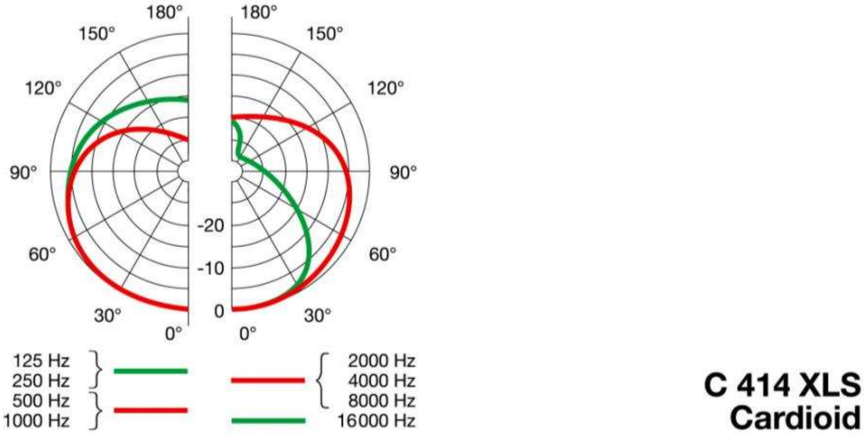


Figure 4 AKG C414 Cardioid Polar Pattern

http://www.akg.com/site/products/powerslave,id,1128,pid,1128,nodeid,2,_language,EN,view,specs.html

Cardioid is sensitive to a broad angle at the front of the microphone and according to Bartlett and Bartlett (2009a) about 6dB less sensitive at the sides and about 15 to 25dB less sensitive at the rear.

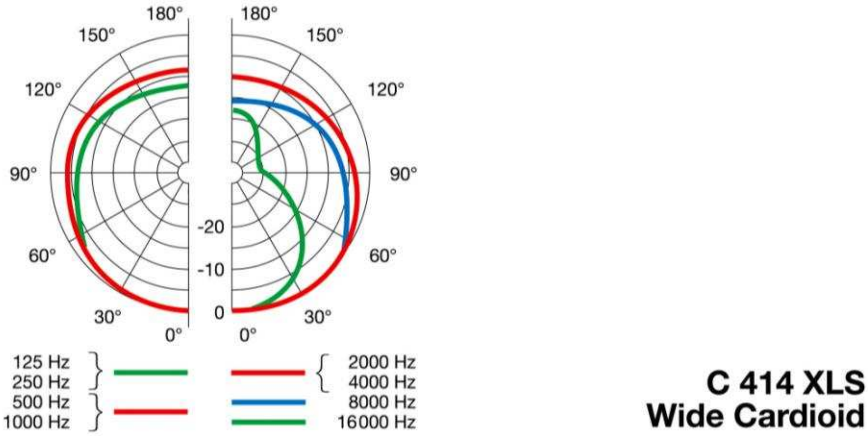


Figure 5 AKG C414 Super cardioid Polar Pattern

http://www.akg.com/site/products/powerslave,id,1128,pid,1128,nodeid,2,_language,EN,view,specs.html

Super (wide) cardioid is about 8.7dB less sensitive at the sides and has two areas where it is least sensitive at ±125° from the front of the microphone.

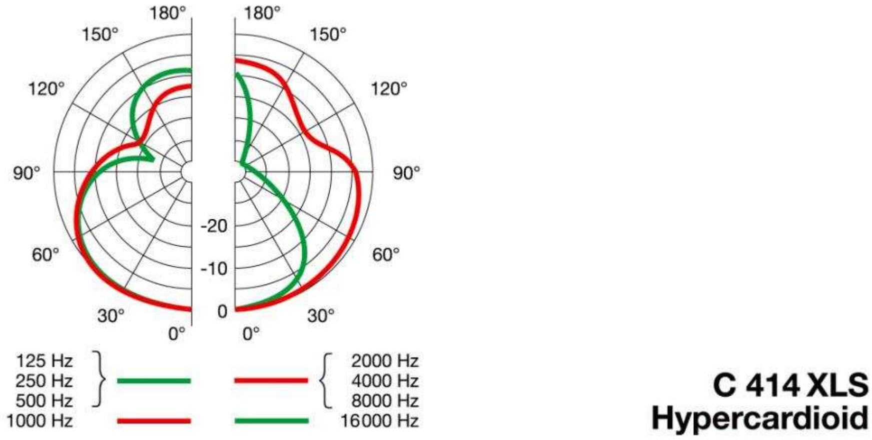


Figure 6 AKG C414 Hypercardioid Polar Pattern

http://www.akg.com/site/products/powerslave,id,1128,pid,1128,nodeid,2,_language,EN,view,specs.html

Hypercardioid is 12dB less sensitive at the sides and has two areas of least pickup at ±110° from the front. As the different unidirectional polar patterns become increasingly directional on axis, they also become less effective as rejecting off axis sound.

It is important that when multiple microphones are used for stereo recording they are the same model and, where possible, have been manufactured as a matched pair, as any difference in the polar response or sensitivity will cause the image to shift as the pitch changes. (Streicher & Dooley, 1985)

The choice of polar pattern is an important consideration when selecting microphones for stereo recording, as the degree of off axis rejection defines the ICLD that is created by an off axis source. The choice of polar pattern also affects the ratio of reverberant to direct sound, defining the maximum distance from source for microphone placement. Some techniques rely on omnidirectional microphones so as to eliminate any level differences for off axis sources, instead using time delays and level differences caused by the distance between microphones and source.

1.1.2 Interaural Time Delay & Interaural Level Difference

Before exploring specific stereo microphone arrays it is important to understand how microphones can be used to localise sound, using the same principles as seen in the human hearing system. To create an ICLD, two cardioid microphones can be placed with the capsules crossed so that one points at the left of the source and the other points at the right. Sources directly in front are picked up equally off axis by both microphones, whereas a source to one side will be on axis to one microphone but 90° degrees off axis to the other microphone producing a level difference of 6dB, as seen from the polar response chart.

Bartlett (1991b) explains that by introducing a horizontal distance of only a few inches between two microphones a time delay for off-centre sources is introduced. *'For example,* if the sound source is 45° to the right, and the microphones are 8 inches apart, the time difference produced between channels for this source is about 0.4 msec. For the same source, a 20-inch spacing between microphones produces a 1.5 msec time difference *between channels'*.

1.2 Microphone Techniques

1.2.1 Coincident Pairs

Intensity Stereo (XY)

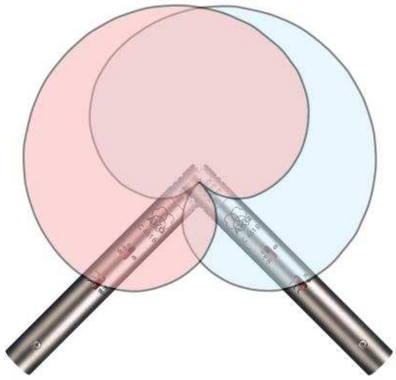


Figure 7 XY Array with Cardioid Microphones

Two unidirectional microphones are positioned so that the capsules cross at 90° in an X formation, one directed at the left of the sound source and the other at the right. Directional microphones have a reduced sensitivity off axis, so as the sound source moves towards one side of the sound field it will become increasingly on axis for one microphone while simultaneously becoming off axis for the other. At the far side of the sound field, the maximum output will be achieved from the on axis microphone while the other microphone will have a greatly reduced output since the source is positioned at its null point. The left side of the produced stereo field corresponds to the null point of the right capsules pickup, however Rumsey and McCormick (2006a) note that ‘psychoacoustically this point can be reached before the maximum level difference is arrived at’.

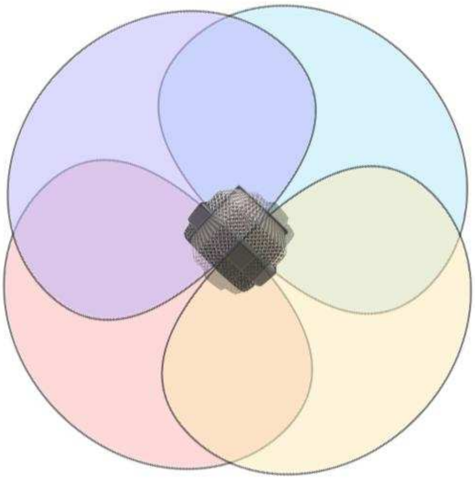


Figure 8 Blumlein Array with Bi directional microphones

Rumsey and McCormick (2006b) also comment that 'A coincident pair of figure-eight microphones at 90° provides good correspondence between the actual angle of the source and the apparent position of the virtu*al image when reproduced on loudspeakers'*. Additional room reverberation captured by the rear lobe, can help to lend the stereo effect greater realism, but can be overwhelming especially if the distance from the source is too great. Rumsey and McCormick (2006a) observe that a 'hypercarioid pattern is often chosen for its smaller rear lobes than the figure-eight, allowing a more distant placement *form the source'* because of the reduced level of room reverberation captured in the recordings.

The angle between the capsules can be varied to alter the width of the perceived sound stage. Increasing the angle between the capsules decreases the distance between the null points, causing the sides of the image to move towards the centre of the stage, so that sources that had been half way to one side will now appear closer to the centre of the sound field. A narrow angle between null points results in a wider sound field, as sources only have to travel a small distance from the centre to move to the far left or right of the image.

Bartlett (1991c) suggests a number of angles and polar patterns, including 180° co-incident cardioids which will give the will give the widest possible stereo separation but produce a weak central image due to the off-axis frequency response of both microphones, it also places the reverberation in the extreme left and right of the image which can be disorientating. Angling cardioid microphones at 90° places the reverberation in the centre but gives a narrow stage width unless the sound source is surrounding the array. However a pair of bi-directional microphones, placed at 90° to each other has been found to give sharp stereo images with a good sense of depth and reverberation while still allowing the listener to accurately place the sounds.

Mid-Side

Another arrangement of coincident microphones is called mid-side or MS, which uses a unidirectional microphone directly facing the source, summed and differenced with a bidirectional microphone positioned side on to the source so that each lobe captures a different side of the sound source.

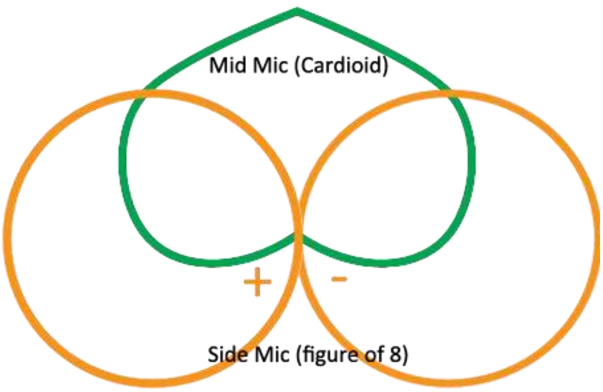


Figure 9 Mid Side Array

The MS array allows a unidirectional microphone to be on axis to the centre of the sound source giving a wider frequency response than in other arrangements where the centre of the sound source is partway off axis for both microphones.

Before this array can be monitored correctly it must first be summed to produce the stereo channels. The left channel is composed of the mid signal summed with the side, the right channel is the mid signal summed with a phase-inverted copy of the side. The far left or right of the image produced by a MS pair occurs where the level of the side signal is equal to the level of the mid signal. Because the ratio of the two microphones can be varied after recording, the size of the stereo field can therefore also be adjusted. If the side signal level is increased, then the points where the side signal is equal to the mid signal move inwards, bringing the edges of the sound field closer to the centre making the reproduced sound field wider.

Although true MS microphones come with a control box to matrix the two signals together it is possible to create a standard stereo signal using just 3 channels on a mixer. The Mid and Side signals are fed, in phase, into two separate channels on a mixer and a post fader send of the side signal is sent to a third channel where it is phased inverted. The Mid signal is panned centrally with the side signal panned left, since 2L=M+S, and the phase inverted signal panned right as 2R=M-S. An advantage of the MS arrangement is that because of the direct inverse relationship between the left and right channels, if the signal is combined to mono the side components of the left and right will cancel and the listener will only hear the mid signal.

1.2.2 Near-Coincident Pairs

Near-coincident pairs of directional microphones use a combination of the level differences caused by angling the microphones, with the time delay caused by spacing the microphones a small distance apart to capture the spatial cues of the sound source. The angle or spacing between the capsules can be varied to adjust the width and separation of the stereo field however, using too wide an angle or spacing can result in a weak central image and too narrow an angle or spacing will result in a narrow stereo spread. According to Rumsey and McCormick (2009c) this stereo microphone technique is particularly effective at creating an authentic stereo image when played back over headphones due to ‘the microphone spacing being similar to ear spacing’ creating a similar time delay caused naturally by the distance travelled by the sound before reaching each ear.

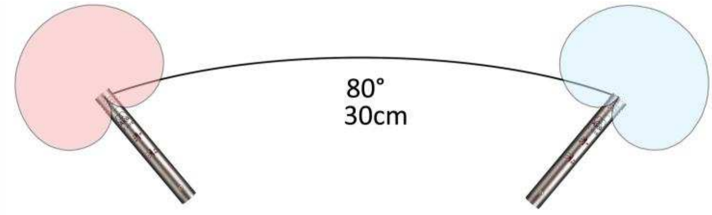


Figure 10 NoS Array with Cardioid microphones

This technique is also said to give an increased sense of spaciousness due to the lack of phase correlation between channels at high frequencies.

There are a number of different ‘named techniques’ where individuals or organisations, such as the Office de Radiodiffusion-Television Français (ORTF) and the Nederlande Omroep Stichting (NOS), have specified angles and distances between microphones for a set recording angle, as shown in Table 1.

Table 1 Specified near-coincident pairs. (Rumsey and McCormick, 2009, p.459)

Designation	Polar Pattern	Microphone Angle	Spacing	Recording Angle
NOS	Cardioid	±45°	30 cm	80°
RAU	Cardioid	±50°	21 cm	90°
ORTF	Cardioid	±55°	17 cm	95°
DIN	Cardioid	±45°	20 cm	100°

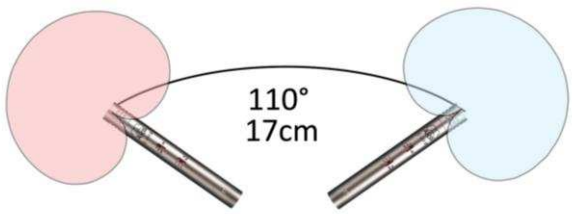


Figure 11 ORTF Array with Cardioid microphones

The distance of 17cm for ORTF was chosen because it provided the best image stability with head motion when the speakers are angled ±30°. The microphones are angled 110° apart because, with a spacing of 17cm, a smaller angle does not spread the image all the way across speakers and a larger angle leaves a gap in the centre of the stereo field.

1.2.3 Spaced Pair

The earliest documented stereo recordings conducted by Clement Ader used a spaced pair of microphones and the technique has been popular since then. The spaced pair or A-B technique uses two identical microphones, placed a few meters apart aiming straight ahead towards the ensemble. While any polar pattern can be used for this arrangement Bartlett (1991d) notes that the omnidirectional pattern is the most popular, especially as omnidirectional condenser microphones tend 'to have a flatter and more extended frequency response than their directional counterparts' (Rumsey 2001a).

A spaced pair relies on the precedence effect, where two similar waveforms arrive at each ear, one about 2-50ms before the other, the brain will localise the sound in the direction of the first wave rather than 'hearing' two separate sound sources in different directions. However, if the delay is greater than 50ms then the brain will interpret the second sound as an echo of the original source coming from a different direction. The time and level difference at each microphone will depend on the distance from the source; a more distant source will have a much smaller delay and level difference than a closer source but if the source is too close then less reverberant sound will be picked up by the microphone.

Although this technique can give the illusion of very natural sounding spaces Bartlett (1991e) notes that it is not very good for localising sound sources as 'stereo images produced solely by time differences are relatively unfocused.' It is possible to vary the width of the stereo image by increasing the distance between the microphones however a distance greater than 3 feet apart will lead to an exaggerated separation effect. The large time differences between microphones that produce the stereo effect correspond to phase differences between the two channels, which will cause phase-cancellation at a range of frequencies if the two microphones are combined to mono. It is also noted by Rumsey (2006c) that because of the lack of phase coherence, phase inverting one of the channels on playback has no noticeable effect on the stereo image unlike a coincident pair. It has been argued that the lack of complete phase coherence between the channels leads to the impression of spaciousness, as ‘the highly de-correlated signals which result from spaced techniques are also a feature of concert hall acoustics.’(Rumsey 2006c)

1.2.4 Binaural and Other Techniques

Figure 12 KU 100 Binaural Microphone

http://www.neumann.com/?lang=en&id=current_microphones&cid=ku100_description

Binaural techniques make use of an artificial or ‘dummy’ head with a flush mounted microphone in each ear complete with external pinna to record the sound, as heard at the ears of a listener. As well as dummy heads there are also ‘HATS’ (head and torso simulators) which incorporate the natural reflections from the torso into the recording. However the effect of the torso on the HRTF has been found to differ considerably between individuals and can cause a high degree of image confusion upon playback (Rumsey, 2001h). Bartlett (1991f) explains that binaural recording works on the premise that when we naturally listen to sound the input on our ears is just two one-dimensional signals, the sound pressure at our ears. If the same pressure can be recreated at the ears of the listener as would have occurred ‘live’ then we can reproduce the listening experience from the room, including the directional information and reverberation. The head acts as an obstacle to the high-mid frequency sound waves so the ear on the far side of the head from the source will have a reduced frequency response in this range compared with the other ear. The folds in the pinna affect the frequency response through the reflections interacting with the direct sound to create phase cancellations at certain frequencies depending on the location of the sound source acting as another localisation cue.

Most artificial heads mount the microphone flush to head rather than including the ear canal present in a human head, unless designed as a measuring device when the microphones are placed at the end of the ear canal where the eardrum would be (Rumsey, 2001h). Bartlett (1991g) comments that '*the ear canal's resonance does not change with* sound source direction, so the ear canal does not supply localization cues.' As well as using a dummy head Rumsey (2001h) explains it is possible to use the ears of a human to create a binaural recording, however 'it can be difficult to mount high quality microphones *in the ears and the head movements and noises of the owner can be obtrusive'*. A human head can be approximated using a sphere or disc to separate the two microphones, simulating the shadowing effect of the head, which helps to provide greater level differences at the higher frequencies however, a simplified system such as this does not include the filtering effects caused by the outer ear (Rumsey, 2001b). Rumsey (2001h) notes that binaural recordings made with approximations of the human head have 'reasonable loudspeaker compatibility as they do not have the unusual equalisation that results from pinnae filtering'.

Normal, un-equalised binaural recordings will be affected by two stages of pinnae filtering, once during recording, and then again upon loudspeaker playback. Listening to a binaural recording with headphones is the 'most spatially accurate method now known' (Bartlett, 1991f) as it recreates the spatial locations and room ambience remarkably well, producing sounds from all around your head. Binaural cues do not have to be generated through the use of dummy heads, provided that the HRTFs are known or can be approximated for the different angles of the sound sources, the appropriate delays and filters can be applied. However obtaining an accurate set of HRTF data for all angles and elevations of sound sources is difficult and time consuming, so often a small number of measurements are taken and then an approximate HRTF interpolated from this. (Rumsey, 2001h)

1.2.5 Comparing microphone techniques

Several comparisons of stereo microphone techniques have been undertaken with a wide degree of findings and results. The Unified theory of microphone systems for stereophonic sound recording (Williams, 1987) contains comparisons made of several systems examining the recording angle and geometric distortion, which leads to the effect of exaggerated separation, associated with each technique. However some of these results conflict with those of other authors such as Benjamin Benfield and Bennett Smith who's computer-aided models of image locations versus frequency for a number of techniques, determined that co-incident cardioids at 90° give a narrow stereo spread, while Williams observed a 180° recording angle.

Despite the theoretical advantages of binaural techniques combining time and intensity differences with HRTF frequency responses, Bartlett (1991h) explains that the location accuracy provided by a dummy head is less precise due to the difference in HRTF. However tests carried out by C. Huggonet and J. Jouhaneau using modulated tons bursts and a violin established it as giving the best depth perception. Carl Ceoen used a series of listening tests to compare the relative image sharpness offered by different techniques and found that 'the ORTF system was the best overall compromise, and that the MS system lacked intimacy.' (Ceoen, 1972 cited in Bartlett, 1991i)

Bartlett (1991j) summarizes that although each of the different experimenters has reached a number of different conclusions there some universal results, which are commonly found to be true. ‘Widely spaced microphones give poorly focused imaging’ while ‘the Blumlein technique gives sharp imaging’. The differing opinions of previous studies provide a useful starting point for this report, as the conclusions of previous experiments can now be tested, through new experiments, and evaluated for their relevance in this particular scenario before creating the final recording.

No one technique is suitable for every situation, familiarization with a number of different techniques can help to overcome any acoustic issues within the recording space and results can often be improved by tailoring the angle and spacing of the microphones to the situation.

1.2.6 Stereo Recordings of large ensembles

Bartlett (1991p) recommends condenser microphones, with wide flat frequency responses and a low self-noise specification, when creating a stereo recording of a large ensemble. A venue with good acoustics and an appropriately long reverb time are ideal, but if it is necessary to record in a room that is too ‘dead’ artificial reverb can be added at a later stage. The microphone array should be a few meters in front of the ensemble, raised above and angled down towards it. The position of the microphone array is important, as it will be this that will control the acoustic perspective and balance between the different elements as well as the affecting the stereo imaging. By raising the microphones above the ensemble you prevent the front row of instruments from dominating the rest of the ensemble.

When recording, ‘microphones must be placed closer to the musicians than in a good live listening position’ (Bartlett, 1991q). If the microphones are placed where the audience are, although the live sound is well balanced, the recorded sound will be overly “muddy” as it is dominated by reverberant sound. Conversely if the microphones are too close then the instruments will sound too detailed and the recording will lack the ambience of a natural space. By varying the distance from the ensemble to the microphones the ratio of direct to reverberant sound can be altered until a suitable balance is found.

Bartlett (1991r) states that Delos Recording Director John Eargle uses two stereo pairs one pair close to the ensemble to capture the clarity and then another 30ft further away to capture the ambience, blending the two to create the finished mix. Because of the large distance and different spectral characteristics of the two arrays, comb filtering effects are largely avoided however if the distance between the two arrays is too large then the time delay maybe great enough for the second pair to be perceived as an echo of the first.

For more complex recordings the use of spot microphones to allow for greater control over the balance of individual elements is recommended. The pan position of these spot microphones should be carefully matched to the location of the corresponding element within the stereo array. They can then be mixed in at a low level relative to the main pair to add clarity without destroying the depth provided by the original stereo pair. Bartlett also notes that, due to the close proximity, the timbre of the instrument picked up by the spot microphone may be excessively bright, which can be overcome with a high frequency roll off. To further integrate the spot microphones with the main pair, especially when the distance between them is considerable, the spot microphones signal can be delayed so that it coincides with the main pair.

When recording smaller groups or modern pop-orientated music there are a number of advantages to using a stereo microphone array. Although finding the correct positioning of performers and array can be time consuming the final recording will benefit from; an increased sense of ambience, a more natural timbral response from the instruments and the size and position of each element being encoded to the recording.

1.3 Predicting and Controlling the Stereo Image

1.3.1 Calculating the Stereo Image

Since the position of an auditory event within a sound field is based on a number of known cues, it follows that these factors can be calculated and the resultant position within the stereo field inferred. Breebaart & Faller (2007a) explain that with coincident microphones where there is no phase difference between the two signals, the position of a sound source is inferred purely from the intensity difference between the two channels, caused by angling the directional microphones. It can therefore be stated that the level difference is a function of the source angle, Φ which when played back over a pair of loudspeakers will appear at angle Φ' relative to the listener. In a correctly configured system Φ≈Φ'.

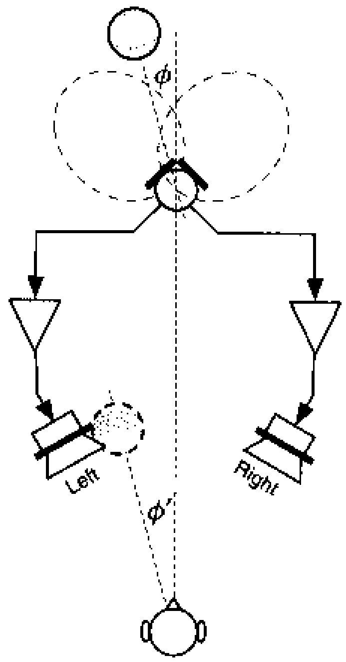


Figure 13 ILD vs ICLD for a sound source over speakers

Rumsey, F 2001, Spatial Audio, Focal Press, Oxford, Fig 3.2 p. 54.

The level difference between two microphones can be calculated using trigonometry and defined constants relating to the polar patterns of the microphones used. (Bartlett 1991l).

$$dB = 20\log\left[\frac{a + b\cos((\theta_m/2) - \theta_s)}{a + b\cos((\theta_m/2) + \theta_s)}\right]$$

a & b are given constants for different polar patterns
θ_m is the angle between the microphones in degrees
θ_s is the angle between the sound source and the centre in degrees

Omnidirectional	a=1,	b=0
Bi-directional	a=0,	b=1
Cardioid	a=0.5,	b=0.5
Supercardioid	a=0.366,b=0.634	
Hypercardioid	a=0.25,b=0.75	

When microphones are spaced apart, even by a small distance, a time difference between the two channels is introduced for off centre sound sources. Bartlett (1991b) explains how, as time delays between channels increase, the resultant phantom image moves further off centre. The time difference between channels can be calculated as a function of the distance from microphone to source, the spacing between the two microphones, the angle to the sound source and the speed of sound. This can be calculated using the following formula. (Bartlett, 1991m)

$$T = \frac{\sqrt{D^2 + [(S/2) + D\tan\theta_s]^2} - \sqrt{D^2 + [(S/2) - D\tan\theta_s]^2}}{C}$$

D= Distance from source to microphone (meters)

S= Spacing between microphones (meters)

θ_s= Source angle from centre (degrees)

C= Speed of sound (343ms^{-1})

For near co-incident microphone arrays, where the separation between the two microphones is small then the equation can be simplified to $T = \frac{S\sin\theta_s}{C}$

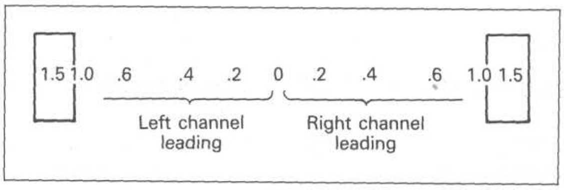


Figure 3-13 Approximate image location versus time difference between channels, in milliseconds (listener's perception).

Figure 14 Image location for various ICTD

Bartlett, B 1991, Stereo Microphone Techniques, Focal Press, Boston, 3.13 p. 42

Listening tests have been carried out by Bartlett locating the apparent position of a sound source when reproduced over loudspeaker while varying the time and level differences to map the required difference between cues to reproduce a sound at different points between the two speakers.

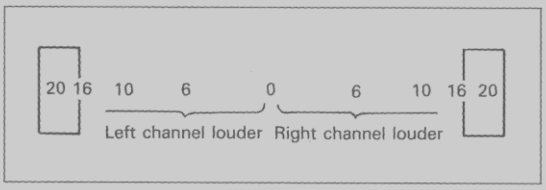


Figure 3-10 Stereo-image location versus amplitude difference between channels, in dB (listener's perception).

Figure 15 Image location for various ICLD

Bartlett, B 1991, Stereo Microphone Techniques, Focal Press, Boston, 3.10 p. 40

The phantom imaged produced can be further localised through a combination of these cues. If the level difference between two channels is sufficient to place the phantom image half way towards one loudspeaker and a time delay is then introduced, then the localisation shifts are added together moving the phantom image all the way to one side of the sound field. Where multiple cues are combined they must correspond, if the channel with the higher level is delayed then image confusion will occur as the two cues conflict with each other. The greatest degree of image localisation can be achieved in a system that uses both cues to create the final image.

1.3.2 Controlling the Stereo Image

Bartlett (1991k) explains that by varying the angling and spacing of the microphone pair the width of the stereo image can be adjusted to suit the recording environment and situation. 'The particular angle and spacing you use is not sacred' and will not give the same results when recording in different rooms with different performers, although they can be a useful starting point it is worth tailoring the angles to suit. Williams (1987a) states that the 'microphone position is generally a compromise between a good coherent stereophonic image and the required ratio of direct to reverberant sound.' Microphone arrays differ in the recording angle they are able to cover and result in varying degrees of deviation from the original location in the reproduced stereo field. Williams observed the recording angle of 90° coincident cardioids to be ±90° requiring the sound source to form a semi-circle around the array for the reproduced stereo field to stretch from speaker to speaker, while the similar Blumlein technique only offers a recording angle of 90°.

The balance of instruments within the field can be fine tuned by adjusting the microphone positioning; increasing the angle between microphones while decreasing the spacing between them will reduce the level of the central instruments while maintaining the width of the stereo sound field. This suggests a method for decreasing the effect of early reflections from the performance space, by reducing the level of the central instruments the stereo array can be moved closer to the instruments increasing the ratio of direct to reverberant sound. Additional reverberation could be added through the use of a pair of distant microphones or with artificial reverb. Angling the microphones further apart increases the ratio of reverberation in the recording, giving the impression that the source is further away, while increasing the spacing between microphones does not alter the sense of distance but does degrade the sharpness of the image.

1.4 Stereo Reproduction

Although two-channel stereo has now become the norm in most listening environments, many listeners are making the upgrade to multi-channel surround sound and it is important to remember that stereo wasn't always the predominant format. The use of 2 speakers for reproduction has become wide spread, since it is easy to implement in a wide variety of environments at relatively low cost while still providing a good phantom image for central sound sources. In the 1930s research by Steinberg and Snow was conducted into recreating the 'sound wavefront' using a large number of microphones and corresponding loudspeakers in another room. When reducing the number of microphones and speakers it was found that 'with two channels, central sources appeared to recede towards the rear of the sound stage and that the width of the reproduced soundstage appeared to be increased' (Rumsey, 2006g)

The objective of stereo reproduction is not always to recreate the original sound source and associated localisation cues exactly; sometimes it is preferable to create an approximation of the original sound source, creating a believable illusion of natural sources placed within a recording environment. It is also necessary to remember that the listener will not always remain static within the optimum listening position so the stereo image should remain effective for multiple listener positions.

1.4.1 Stereo reproduction over loudspeaker

Rumsey (2001d) states that based on 'formal research and practical experience it has become almost universally accepted that the optimum configuration for two loudspeaker stereo is an equilateral triangle with the listener just to the rear of the point of the triangle.' The clearest phantom images are established at the ideal listening position, with greater distance causing them to become less stable as spatial cues contradict each other; the effect of head movement on the image will also become exaggerated beyond this point.

Stereo reproduction over two channel loudspeaker is significantly different to listening to a natural sound source as the signals are now approaching the ears from 2 distinct locations rather than being heard from all around the head as in a natural listening environment, creating only a basic 3D illusion of a natural sound source. Rumsey (2006d) explains that in a natural situation a single source is 'heard' by both ears with the level and time localisation cues caused by the shadowing effect of the head and the time required for a wave to travel round the head to read the other ear. In a listening environment where two similar signals are produced from loudspeakers in different directions, the precedence effect localises the sound in the direction of the wave that arrived first, with subsequent waves arriving within a short time frame also being localised in this direction. There are also time & level delays caused by the original cues captured by the microphone array, which combine with those caused by the placement of the loudspeakers. Stereo sources encoded with a time delay between channels can often result in confusing and contradictory images for long continuous sounds while short transients are less likely to be affected by the phase differences between channels but the precedence effect is not as effective at identifying these separate signals as a single source. Time-difference stereo signals will cause phase cancellations if summed to mono, resulting in a distorted frequency response.

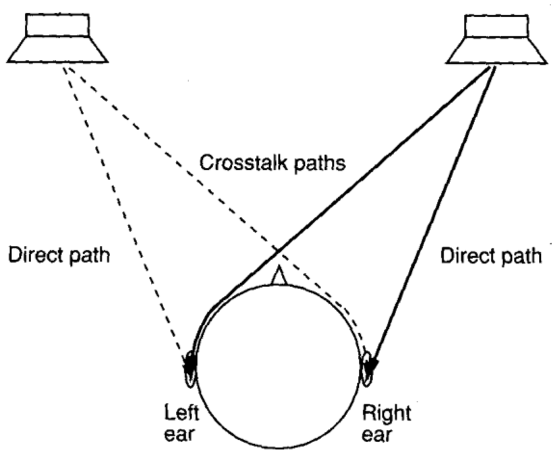


Figure 16 Acoustic Crosstalk

Rumsey, F 2001, Spatial Audio, Focal Press, Oxford, Fig 3.2 p. 54.

Rumsey (1989) explains that when two loudspeakers reproduce the similar signals but with level differences between them, either as a product of the microphone technique used or introduced artificially, the summing effect that occurs at each ear due to acoustic cross talk results in a phase difference between the signal at each ear. This is because the addition of vectors at each ear results in two signals that differ in phase angle proportional to the amplitude difference. If the level difference between channels remains constant then the phase angle will vary linearly with frequency according to Rumsey (2006e), although at higher frequencies the phase difference becomes less relevant as the shadowing of the head causes level differences between the signals at each ear. Due to this effect, artificial pan pot stereo is able to provide phase and level differences, when reproduced over loudspeakers. However if the signal is summed to mono there will be no phase cancellation as there is no actual phase difference between the two channels, only between the two signals produced by the loudspeakers.

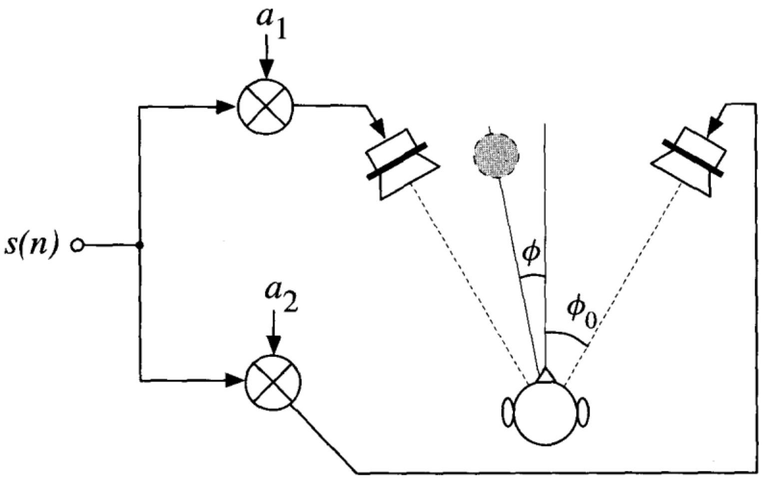


Figure 17 Stereo Panning using the law of sines

Breebaart, F and Faller, C 2007, Spatial Audio Processing, John Wiley & Sons Ltd, Chichester, Fig 2.2 p. 8.

Rumsey explains that the result of the mathematical phasor analysis provides a formula, which can ‘be used to determine, for any angle subtended by the loudspeakers and the listener, what the apparent angle of the virtual image will be for a given difference between left & right levels’. This is known as the “Law of Sines”

$$\sin \quad = \frac{(L \quad R)}{(L+R)} \sin \quad _0$$

Where apparent angle of offset from the centre of the virtual image, **L** & **R** are scaling factors and $_0$ is the angle subtended by the speaker and the listener

it can also be shown that $\tan \ _t = \frac{L \quad R}{L+R}$

where $\boldsymbol{\theta}_t$ is the true angle of offset of a real source from the centre-front of a Blumlein pair.

A coincident pair of bidirectional microphones, a Blumlein pair, produces level differences which provide a very close correlation between source angle and the angle presented at reproduction when played back over speakers 120° apart. However, Rumsey (2001e) states that ‘this angle of loudspeakers is not found to be very satisfactory for practical purposes for reasons such as the tendency to give rise to a hole in the middle of the image.’ As the angle between loudspeakers becomes narrower the apparent angle between sources and the loudspeaker remains proportional resulting in naturally proportioned but narrower image for the standard 60° separation between speakers.

Localisation is only possible if the polarity of the two loudspeakers is the same. Bartlett (1991n) explains that two signals with equal level but opposite polarity will result in an 'image that has a diffuse, directionless quality and cannot be localised.' If there is a level difference between two signals of opposite polarity then 'the image often appears outside the bounds of the speaker pair'. Signals could have reversed polarity due to an incorrectly wired speaker, cable or because of the polar pattern of the microphones, for example the rear lobe of a bi-directional microphone is in opposite polarity with the front lobe.

1.4.2 Transaural Loudspeaker Reproduction

The main differences between stereo headphone and loudspeaker reproduction is the separation of channels; with headphones the left channel is only heard by the left ear and the right channel by the right ear. Due to acoustic crosstalk reproduction over speakers leads to both channels being 'heard' by both ears. For binaural signals, recreating the natural cues recorded with in ear microphones or a dummy head, to be played back over loudspeaker correctly acoustic crosstalk must be cancelled.

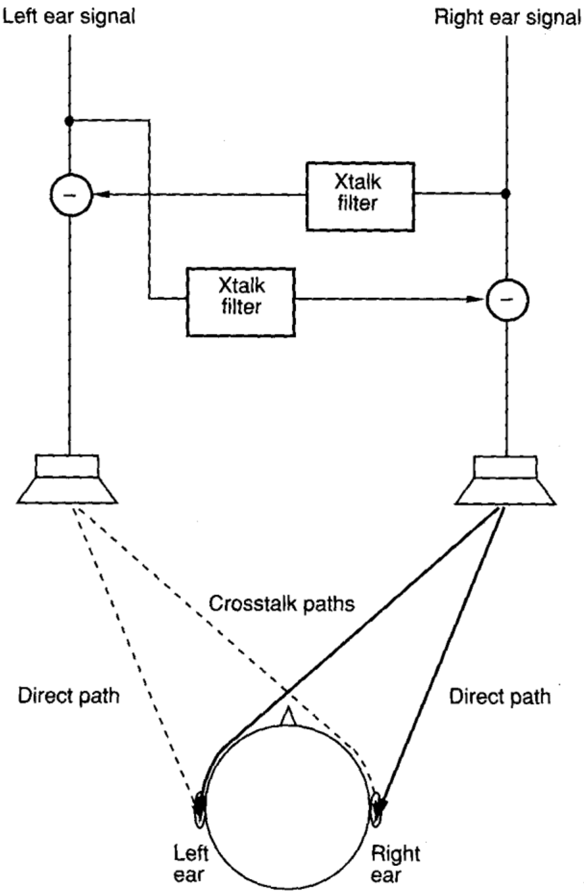


Figure 18 Crosstalk cancellation

Rumsey, F 2001, Spatial Audio, Focal Press, Oxford, Fig 3.2 p. 54.

Transaural stereo accomplishes this by sending an out of phase version of the left channel to the right speaker and vice versa, so that when the signals are summed at the ears the cross talk from the opposite speaker will cancel with the out of phase component of the near speaker. The 'cancellation' signal must be processed according to the HRTF characteristics of the path it will take. Rumsey (2006f) comments that this can produce 'fully three-dimensional virtual sources to be perceived including behind the listener (from only two loudspeakers located at the front).' However it should also be noted that the cross talk filters are only valid for a very narrow range of listening angles and, once the listener has moved more than a few 'tens of centimetres' beyond this point, the effect can disappear completely. Some listeners also find the effect very unnatural and experience fatigue when listening to it for extended periods of time.

The design of most binaural loudspeaker systems is a trade off between the localisation accuracy and the compatibility with a large number of listeners. A situation where transaural has been found to be highly effective is creating 3D sound for personal computers and home theatre systems where the position of the listener and speakers can be assumed to be fairly constant in relation to the screen, which is often attached to the speakers. Because the location of the listener can be predicted, the required filters can be calculated accurately without needing to account for the effect of listener movement. Also, due to the applications that make use of these techniques such as games and films, the localisation accuracy does not need to be as exact as the cues used can be exaggerated to ensure the binaural effect is heard to some degree even with low quality speakers.

1.4.3 Stereo reproduction over headphones

Stereo signals are usually optimised for loudspeaker playback, having been recorded, monitored and mixed via a number of different loudspeaker speaker systems by the recording engineers involved. The ICLD and ICTD parameters result in similar psychoacoustic effects regardless of whether the signals are played back over headphones or loudspeaker allowing the same signal to be used for both playback methods. However because the two channels are presented to each ear individually over headphones the width of the spatial image created is limited, presenting images localised within the head of the listener. Rumsey (2009a) states that Bauer showed that stereo signals designed for reproduction over loudspeakers had level differences too great for reproduction over headphones when compared with the cues presented by a real-life situation.

Loudspeaker stereo over headphones

Bauer suggested introducing a ‘measure of delayed crosstalk between the channels to simulate the correct interaural level differences at different frequencies’ (Rumsey, 2009a) as well as simulating the interaural time delays that would result from listening over a pair of loudspeakers at 45° to the listener. The circuit designed for this reduces the introduced delay at higher frequencies, as it would complicate the circuit and because localisation at high frequencies relies more on level differences than time delays.

1.4.4 Binaural Stereo

Binaural recording and reproduction is based on the premise that ‘the most accurate reproduction of natural spatial listening cues will be achieved if the ears of the listener can be provided with the same signals they would have experienced in the source environment or during natural listening’ (Rumsey, 2009b). By recording the signals exactly as they arrive at they arrive at the ear of the listener and then playing them back over headphones, a highly accurate stereo field can be reproduced.

For effective binaural reproduction the HRTF applied to the sound sources must be recreated at the ears of the listener. This presents a problem, in that the HRTF of each listener will be different, dependant on the shape and size of their head and torso. For truly natural recreation of the stereo field, the recording should be made with the HRTF of the end listener, by placing the microphones in their ears at the recording event. However more generalised HRTFs have been identified which, when applied to recordings, allow for a stable and reliable reproduction of directional cues for a wide range of listeners although lacking the absolute accuracy provided by using the listeners unique HRTF.

In natural listening situations localisation accuracy is improved by time and intensity differences that occur when the listener moves their head, allowing the brain to use the differences in these cues to refine the localisation accuracy. There is no real way to incorporate this into the recording process without it confusing the image, as the sound source appears to move while the listener’s head remains stationary. However in advanced systems, generally ones creating artificially synthesised binaural signals, it is possible to follow the listener’s head movements and adapt the signals being sent to the ears accordingly. Binaural recordings also prove difficult to localise without visual cues as there are few audible cues identifying the sounds position in front or behind the listener as well as its elevation relative to the listener.

Dummy heads or microphones in the ears of a human listener have a non-flat frequency response due to the heads diffraction of sound, usually resulting in a peak around 3kHz for frontal sounds (Bartlett, 1991o). This is why binaural recordings, played back over headphones or loudspeaker, sound ‘coloured’ unless they are equalised. There are several different equalisation schemes proposed to provide a flat frequency response. Diffuse-field equalisation compensates for the heads’ average response to sounds from all directions. Rumsey (2009a) states that this method of equalisation is preferable as it has *‘been found* to be quite suitable for both binaural and loudspeaker stereo signals’ as opposed to free-field equalisation which compensates for the heads’ response to sound sources in anechoic conditions. Møller says that for ‘binaural reproduction the headphones should be equalised to have a flat frequency response at the point in the ear where the binaural recording microphone was originally placed’ (Rumsey, 2001g). However because headphones are typically used to playback signals designed for loudspeaker reproduction they are often equalised to emulate a free field response or a diffuse field response at the listening position. It is suggested that a switchable form of equalisation is needed, ‘preferably tailored to the individual’.

1.5 Stereo Encoding

Due to the impractically large file sizes required by uncompressed recordings a number of methods for compressing audio files have been developed. Breebaart and Faller (2007b) define audio coding as the ‘process for changing the representation of an audio signal to make it more suitable for transmission or storage.’ In the example of a compact disc, a stereo signal is stored as two separate channels, each sampled at a frequency of 44.1 kHz.

Each sample is represented as a 16 bit integer resulting in a bit rate of 2 x 44.1 x 16= 1411kb/s.

Breebaart and Faller (2007b) define a perceptual audio encoder as ‘an audio coder which incorporates a receiver model, i.e. it considers the properties of the human auditory *system’*. This allows a perceptual encoder to reduce the required bit rate by not encoding signals that are beyond the range of the human hearing system. The details of exactly how perceptual encoders reduce bit rate is beyond the scope of this report, however the effect of data compression methods on the resultant stereo image needs to be considered.

1.5.1 Sum and Difference Coding

It is sometimes convenient to work with stereo in the *‘sum and difference*’ format which stores the sum of the left and right channels as ‘M’ (the sum or the main) and the difference of the L and R channels as ‘S’ (difference or side signal). The M signal is the signal that would be heard by a mono listener while the S signal encodes the side signals, adding width where stereo reproduction is possible. M/S signals can be created through the use of microphone techniques such as mid-side but it can also be derived from a standard LR stereo signal.

Audio Clip 1
Mid component of recording

Audio Clip 2
Side Component of recording

Audio Clip 3
Mid+side component

When creating the mid signal, a correction factor of between -3dB and -6dB is applied to compensate for the addition of similar material. If the left and right signals are identical the sum of the two amplitudes will be double the original, requiring a reduction of 6dB. Where there is no direct phase relationship between L & R then only a 3dB reduction is required. Rumsey (2001f) notes that ‘as most stereo material has a degree of coherence between the channels, the actual rise in level of M compared with L and R is likely to be *somewhere between'* 3dB and 6dB for actual programme material. The S signal is derived in a similar manner, using a similar correction factor except that it is the difference of L and R rather than the sum.

Rumsey (2001f) shows that because M=(L+R)-3dB and S=(L-R)-3dB it is possible to reconstruct the L and R signals from the MS signal as

$$(M+S)=((L+R)\quad 3dB)+((L\quad R)\quad 3dB)$$

$$=2L\quad 6dB \qquad \text{And}$$

$$(M\quad S)=((L+R)\quad 3dB)\quad ((L\quad R)\quad 3dB)$$

$$=2R\quad 6dB$$

1.5.2 Intensity Stereo Coding

Another method of reducing the required bit rate of a stereo audio file is intensity stereo coding. Above 1.6kHz the ICTD becomes redundant as the sound is localised by ICLD. To convey multichannel audio the high frequencies in each channel can be divided into bands and combined to form a common signal. This signal is then reproduced by both channels, using amplitude scaling to recreate the original level differences. According to Pohlmann (2011) it is *'particularly effective at encoding spatial information'* in a lossy format.

1.5.3 MP3 Stereo Coding

The most common form of lossy compression in commercial and personal audio is MPEG Layer III, more commonly known as MP3. To take advantage of redundancy in the signal the MP3 codec offers 4 different methods of encoding a stereo file:

- Standard stereo with independent L/R channels
- M/S stereo, where the entire spectrum is encoded with M/S
- Intensity stereo, where the lower spectral range is encoded with L/R and the upper spectral range is encoded as intensity stereo.

- Intensity stereo and M/S mode where the upper spectral range is encoded as intensity stereo and the lower spectral range is encoded as M/S.

Pohlman (2011) notes that this is more effective for stationary signals than transient signals as they ‘may have different envelopes in different channels’ which can lead to the introduction of artefacts.

Ogg Vorbis is another lossy compression codec, designed to offer higher audio fidelity than MP3 files of the same size. It is also open source and free to use, with no patents or licence fees, unlike MP3. There are a number of lossless encoders which offer reduced file sizes while maintaining high bit rates and allowing for the exact reproduction of the original waveform. To assess the effects of different encoding methods that could be applied to the final recording a number of experiments should be undertaken on reference files, and the effects on the stereo field noted.

2. Analysis of Research

2.1 Main Findings

From the research, clear parallels can be drawn between a number of microphone techniques and the human hearing system, and how the former emulate the latter to encode spatial cues into a stereo recording. There are also a number of key differences and distinctions between the individual techniques and the human hearing system. While in a natural situation a listener uses the time delay between ears, the level difference at the ears and their unique HRTF as well as visual cues to locate a sound source, microphone techniques are more restricted in the cues that they can capture.

Velocity microphones are, by design, more sensitive to sounds approaching from certain directions, dependant on their polar pattern. This fact allows a pair of microphones to encode an ICLD when angled away from each other. An ICTD can be introduced by separating the microphones by a small distance, as with the ears in a human head. The psychoacoustic limit for a sound to be localised entirely at one speaker is an ICTD of 1.5msecs or an ICLD of 20dB, where the sound is 10x louder at one speaker than the other. These two cues can be combined directly, so that a signal with a smaller ICTD and ICLD will be localised entirely at one speaker.

There are advantages and disadvantages for each of the stereo microphone techniques; coincident pairs can only capture ICLD although it has been shown that loudspeaker playback of these techniques can cause phase differences similar to those caused by ICTD, near coincident arrays introduce a time delay by creating a spacing between microphones which has been found to reproduce well over headphones. Spaced pair arrays create both time and level differences but in a manner dissimilar to both the other arrays and human hearing due to the large spacing between microphones causing the hole in the middle effect. The resultant signals have little phase coherence causing cancellation, but also a sense of spaciousness.

Binaural recording techniques attempt to capture the sound as it arrives at the ears of the listener and then recreate this exactly however there are a number of limitations. Recordings made with a real persons head tend to be highly customised to their unique HRTF providing confusing results to other listeners. While dummy heads are useful, designed around generic HRTF allowing for a greater degree of localisation for a greater number of listeners they are still limited, requiring substantial equalisation for headphone playback, cross talk cancellation for loudspeaker playback and even then listening for extended periods of time can be disconcerting and unnatural. The key difference between the two playback methods is that isolation between channels can be achieved with headphones whereas loudspeaker systems suffer from acoustic crosstalk, which creates useful phase differences in some arrays but will cause significant cancellation in others.

Recording with a stereo array depends on the source, environment and purpose of the final recording- no one technique will work in every situation, however they can each be tailored to suit the application. It is possible to calculate the time and level differences that will result from specific source angles and microphone arrays, this knowledge coupled with the perceptual thresholds required for localisation at specific points can be used to plan the locations of elements within the recording to create the desired final image. The angles and spacing between microphones can be adjusted to alter the width and distribution of the stereo image.

2.2 Impact on Final Recording

The research conducted has a significant effect on the design of the final recording; the choice of different stereo arrays presents a large number of options. To select the techniques that will be compared in the final recording, a combination of the research conducted and experimental results will be used.

Although research suggested some significant advantages of binaural recording, especially with localisation over headphones, it also highlights the difficulties of binaural reproduction over loudspeaker. Because the only available method of creating a binaural recording were a pair of in ear microphones which, according to the research conducted, will result in a signal highly tailored to one individuals' HRTF, with significant noise added by the subject. Since the group of performers require no conductor, there is no obvious subject in which to mount the microphones and since a dummy head was not available it was decided to forgo the use of this technique to avoid causing a significant distraction to the performers.

When choosing the microphone placement and spacing in the final recording, the formulae highlighted during the research will be used to decide the positioning required to produce the necessary time and level differences shown in the Williams curves and other listening tests. While it could be assumed that the coincident microphone techniques such as XY, Blumlein and MS would produce similar results, research suggests the rear lobe of the bi-directional microphone will provide a more diffuse sense of reverberation around the listener, while MS offers the ability to alter the size of the sound field after the recording.

The microphones that will be used for the final recording have been selected from the limited options offered by the university, where possible meeting the recommendations set out in the research. Small diaphragm microphones, which have a wider flatter frequency response, will be used with a cardioid polar pattern for coincident and near coincident techniques. To achieve a low self-noise and higher sensitivity for other techniques, which require other polar patterns and are simpler to position, large diaphragm condensers will be used.

2.3 Research Hypotheses to test

Based on the research a number of hypotheses have been stated. Before the final recordings are created it is important that these are tested to confirm their accuracy and assess the impact that these may have on the final recording. The researched techniques rely on the microphone array encoding an ICLD and ICTD similar to that which a human listener would experience. Experiments should be designed to confirm that the expected cues are encoded according to the research for each microphone technique. The research also discussed adjusting microphone arrays to vary the final size of the sound stage; this should be investigated to see how the width of the stereo field is related to the ICTD and ICLD values.

A number of equations were quoted which can calculate the time and level differences resulting from different microphone arrays capturing sources at varying angles. Before these equations can be used to inform the design of the final recordings the results of the formulae for calculating both ICLD and ICTD should be tested against experimental data. The research mentions the differences between headphone and loudspeaker playback and its effect upon the image created for the listener. The differences should be assessed before listening tests are conducted with the final recording.

Experiments should be conducted to test the effectiveness of the different microphone techniques of tracking a moving source to emulate the movement of musical parts between instruments in the final recording. The different microphone techniques can then be assessed on the size and clarity of the image created and from these results decisions over which techniques to use at the final recording can be made.

3. Experimental method

A number of experiments were conducted in university studios to validate hypothesis and questions posed by the research and to assess the viability and use of specific rooms, equipment and recording procedures.

3.1 Initial testing of microphone technique viability

The first experiment was designed to assess the localisation, and ability to track a moving source, for multiple stereo techniques. The following microphone techniques were tested:

Technique	**Polar Pattern(s)**	**Microphones**
Co-incident Pair: Blumlein	Figure of 8	AKG C414 x2
Co-incident Pair: XY	90° Hypercardioid	AKG C1000s x2
Co-incident Pair: XY	180° Cardioid	AKG C451 x2
Co-incident Pair: Mid-Side	Figure of 8 Cardioid	AKG C414 Neumann TLM103
Near co-incident Pair: ORTF	Cardioid 17cm 55°	AKG C414 x2
Near coincident Pair: NOS	Cardioid 30cm 45°	AKG C414
Spaced Pair	Omni-directional	Audix TR40 x2
Binaural		Soundman in ear microphones

Figure 19 Microphone Arrays

(Top – Bottom) Blumlein, Mid-Side, ORTF, XY, NoS

The room was measured to find the midpoint, and the microphone arrays positioned on this line 30cm from the wall. The different microphone techniques were then setup as close to this central point as possible, avoiding contact and shadowing between microphones where possible.

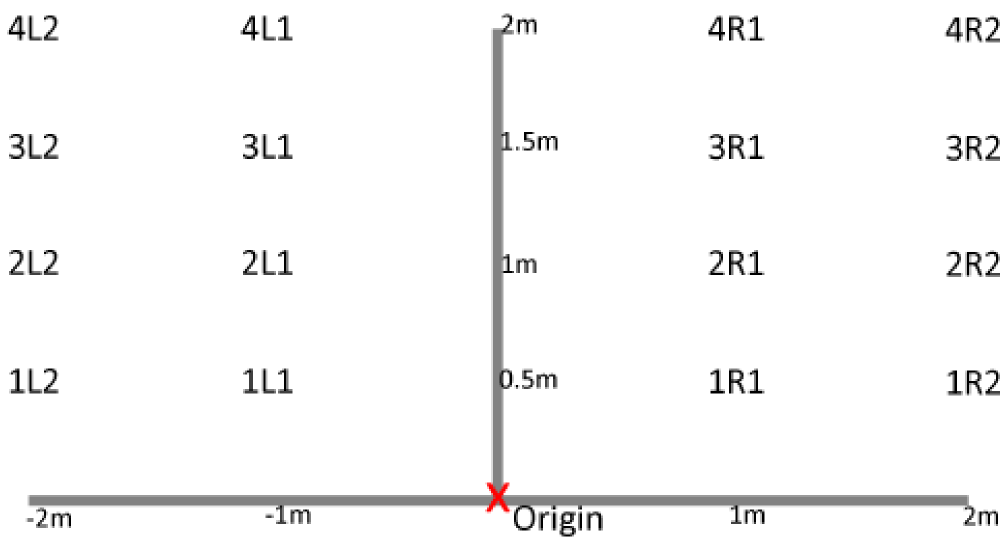


Figure 20 Test grid layout

Defining the microphone position as the origin, points were marked along the midline every 50cm up to 2m. Points were then marked off 1m and 2m to each side of these central points creating a grid pattern. A subject, using a tambourine as a sound source then walked from 1L2 to 1R2 repeating this for each distance to demonstrate the ability of the different arrays to track movement, and represent the width and depth of the stereo field. Two people then stood at 1L2 and 1R2 and spoke in a call and response pattern and then again at 1L1 and 1R1 to judge the width and distribution of the stereo field. This was repeated for each of the distances. Finally two people stood at 1L2 and 1R2 and walked across the stereo field, one reciting the alphabet and the other counting up, crossing in the middle and then continuing to the opposite side. This is repeated for each of the distances to show how the techniques track the movement of multiple sources simultaneously. The signals were recorded through the SSL super analogue pre-amps then directly into Pro-Tools bypassing any equalisation or compression that could affect the signal. The gain was set and then matched between microphones in their stereo pairs.

E1 was used as it has been acoustically treated so room modes and comb filtering effects will be less evident than in other available live rooms, however the dimensions of the room and its rectangular shape were a challenge in making sure that the microphones were placed far enough from the source while providing sufficient.

3.2 Measuring Level Differences between Channels

The second experiment was designed to test the formulae outlined by Bartlett (1991l). The CLIO acoustic testing kit was use to measure the ICLD produced by different microphone techniques for different source angles.

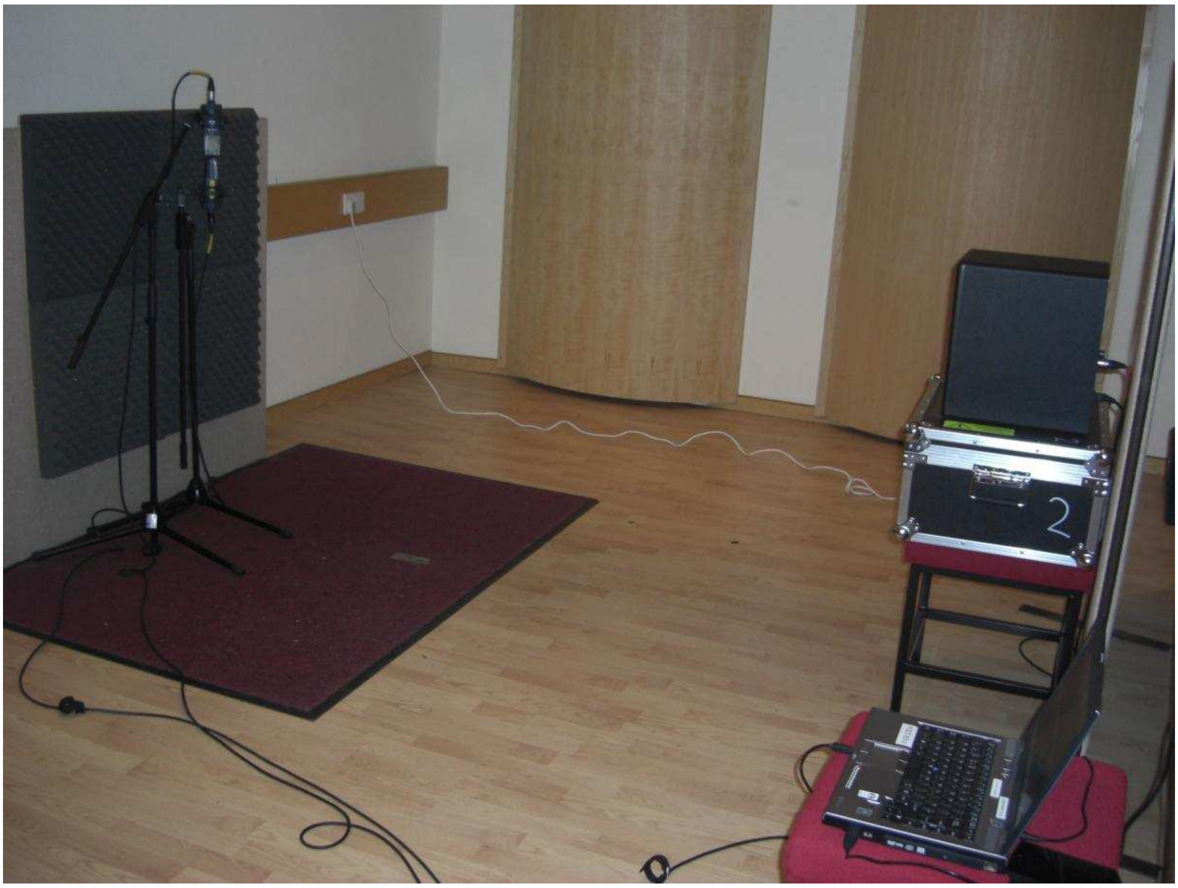


Figure 21 Measuring the ICLD of a Blumlein array using CLIO

The room was measured and the midpoint located, with the microphone arrays positioned on this midpoint 30cm from the rear wall. The loudspeaker output from CLIO was positioned 2.5m away from the microphone array along the midpoint line; this was defined at the central source position, with the left source position defined as 1.6m to the left of the central source position. To measure the ICLD both microphones in the array were connected to the CLIO interface then, using the Fast Fourier Transform (FFT) analyser

CLIO measures the dBSPL value for each microphone at a number of frequency bands. Using the centrally positioned loudspeaker a short burst of pink noise was played and the dBSPL values for each microphone recorded. This was then repeated with the left speaker position providing the noise source and again the dBSPL values for each microphone were measured and recorded. This process was repeated for the following microphone arrays and variations

Technique	**Microphones**	**Angle /Spacing 1**	**Angle /Spacing 2**	**Angle /Spacing 3**
Coincident XY	AKG C451b x2	90°	130°	60°
Coincident Blumlein	AKG C414 x2	90°	180°	-
Coincident MS	AKG C414 x2	-	-	-
Near coincident ORTF	AKG C451b x2	55° 17cm	90° 17cm	
Near coincident NOS	AKG C451b x2	45° 30cm	90° 30cm	
Spaced Pair	AKG C414 x2	0.5m	1m	

To correctly measure the dBSPL for each microphone it was necessary to enter the sensitivity for each microphone, multiplied by 1.465 (or 3.3dB) to account for the pre-amplifier in the CLIO hardware.

3.3 Measuring Time Delays between Channels

The 3rd experiment was designed to test the formulae quoted by Bartlett (1991m) for calculating the ICTD in a stereo microphone array. The room was once again measured to locate the midpoint, with the microphone array positioned on the midline 30cm from the rear wall. 4 points were then measured and marked, 1m and 2m from the microphone array along the midline and then 1.6m to the left of these two points. The microphone arrays were setup and the outputs routed through the SSL AWS into Pro-Tools for recording, a short tone was then played at each of the four positions. The time difference, in samples, could then be measured in Pro-Tools and then converted into a time difference. This was repeated for each of the microphone techniques and variations of angles as shown in the table.

Technique	**Microphones**	**Angle /Spacing 1**	**Angle /Spacing 2**	**Angle /Spacing 3**
Coincident XY	AKG C451b x2	90°	60°	120°
Coincident Blumlein	AKG C414 x2	90°	180°	-
Coincident MS	AKG C414 x2	-	-	-
Near coincident ORTF	AKG C451b x2	55° 17cm	90° 17cm	
Near coincident NOS	AKG C451b x2	45° 30cm	90° 30cm	
Spaced Pair	AKG C414 x2	0.5m	1m	

3.4 Localising Sound through Artificial Level Differences

This experiment was designed to test the ability to localise a recording of a source made using close microphone techniques to the position of a similar element, recorded with a stereo technique, through the use of 'pan-pot' level differences and also to explore the viability of artificially introducing a delay to create the same effect. This test was conducted in the larger live room of studio 4, due to the results from E1 showing the room to be too small to achieve a full stereo spread.

Figure 22 Blumlein and Spaced Pair microphone arrays

The room was measured, and the midpoint of the width found and marked. The Blumlein array was positioned on the central line, 50cm from the rear wall with a spaced pair positioned ±1m from this point. Due to the availability of microphones, and as the test was more concerned with creating accurate artificial stereo locations, it was decided to only use these two stereo techniques. Using a single large diaphragm C414 the sound source, an acoustic guitar, was recorded playing a short pattern using a close microphone technique. The same pattern was then played at 4 different positions throughout the room,

(-1.93m, 1.53m), (-1.03m, 2.82m), (1.03m, 2.82m), (-1.93m, 2.3m)

Which subtends an angle of ±40° for the outer positions and ±20° for the two inner positions while remaining 3m from the central microphone array at each point.

Figure 23 Positions for the acoustic guitar recording

Using the DAW Pro-Tools the mono signal, recorded using the close microphone technique, was sent to the main stereo output. By changing the relative level sent to the left and right output the apparent location of the sound was then changed. Using the formulae,

$\Delta dB = 20\log\left[\dfrac{a + b\cos((\theta_m/2) - \theta_s)}{a + b\cos((\theta_m/2) + \theta_s)}\right]$ and the four locations where the pattern was recorded by the stereo microphone technique, the relative levels sent to the left and right outputs were adjusted to place the mono signal in the same location as heard in the stereo signal. This was also attempted using the formulae to calculate the time delay that would have occurred for a near coincident array at the microphone point and introducing the calculated delay between channels of the mono signal.

3.5 Confirming the Precedence Effect

As discussed in section 1.2.2, research suggested that the spaced pair technique relies on the precedence effect to localise sound in the direction of the earliest arriving waveform when the same source sound arrives at both ears separated by a short delay of 2-50ms. To confirm the precedence effect the same sound source was presented simultaneously to both ears of a listener over headphones, at the same amplitude, resulting in a centrally positioned sound source. A delay was then introduced into the left channel, which was increased in 20ms increments resulting in the image location shifting to the right, until the delay was sufficient for the left channel to be observed as an echo or separate sound source. The research also indicated that the psychoacoustic phenomena that make up the precedence effect respond differently to different sources so the test was conducted with both a short percussive sound, a more musical phrase and a short percussive transient to assess the delay required before a separate source was identified.

4. Results

The results of the experiments, designed to test the research findings were used to assess the various methods of completing the final recording and draw conclusions as to which methods would offer the best results in this recording situation.

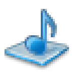
4.1 Discussion of test results

4.1.1 Initial Testing of Techniques

Despite the limitations of the experiment, the data collected does reveal some patterns and, when plotted on a goniometer, the phase correlation between the left and right channels of different techniques can be seen for the different source angles.

Binaural Goniometer Plots- Far left, Central and Far Right source positions

The in-ear binaural microphones provide a very clear distinction between the different positions of the sound source.



4. Binaural LR Tambourine.wav

Audio Clip 4
Binaural Tambourine Recording

Blumlein Goniometer Plots- Far left, Central and Far Right source positions

The original results from the Blumlein pair showed the effect of experimental error, as the right channel was always at a lower level than the left even at the far right source position suggesting either differences in the signal chain between the two channels, an error in their positioning, a fault with one of the microphones or the effects of another microphone shadowing this array. However when conducting the tests for section 3.4 Localising Sound through Artificial Level Differences a Blumlein pair was used to record a sound source moving across the field and the plots from this experiment show the expected clear separation for source positions.

5. Blumlein LR
Tambourine.wav

Audio Clip 5
Blumlein LR Tambourine

NoS Goniometer Plots- Far left, Central and Far Right source positions

The NoS technique shows a fairly widespread image but with the correct correlation for left, centre and right source positions.

6. NoS LR Tambourine.wav

Audio Clip 6
NoS LR Tambourine

ORTF Goniometer Plots- Far left, Central and Far Right source positions

The ORTF technique also seemed to suffer from a degree of experimental error, with the far left and central positions showing the expected spatial cues, while the far right position seems to remain fairly central in its positioning.

7. ORTF LR Tambourine.wav

Audio Clip 7
ORTF LR Tambourine

Spaced Pair Goniometer Plots- Far left, Central and Far Right source positions

The spaced pair gave a fairly dispersed image with more accurate imaging occurring towards the sides of the image where the microphones are located.

8. Spaced Pair LR Tambouri.wav

Audio Clip 8
Spaced Pair LR Tambourine

XY Goniometer Plots- Far left, Central and Far Right source positions

The XY pair gave accurate imaging effects, similar to the binaural pair, with a clear level difference between channels at the left and right source positions. It also retained a positive phase correlation, avoiding phase cancellation.

9. XY LR Tambourine.wav

Audio Clip 9
XY LR Tambourine

The video below demonstrates how a moving source can be tracked across a sound field with a Blumlein pair.

Video Clip 1
Guitar tracking L/R with Blumlein Pair

Setting up all the multiple microphone techniques simultaneously would in theory, allow for direct comparisons to be made between techniques as they tracked the same source. However it turned out to be impractical to have that many microphones and stands positioned in close proximity; as they caused shadowing, introduced reflections and interfered with positioning, especially as the stands couldn't hold the weight of microphones at a distance causing them to droop and disrupt the position of the other arrays. The SSL AWS900 and alpha link A/D converter also caused issues, refusing to interface correctly with Pro-Tools, causing large delays to the start of the experiment and reducing the overall time available for testing.

Despite the flaws in the method, the experiment still allowed for a number of conclusions to be drawn, which could then be explored more fully in further testing. It was shown that the use of multiple simultaneous microphone techniques, unless setup with much care and planning, could easily lead to compromised results. It is important that the signal chain between microphones in an array is closely matched as additional gain or processing to one half of the pair, or differences in microphone sensitivity and polar pattern will result in a distorted and unevenly distributed image.

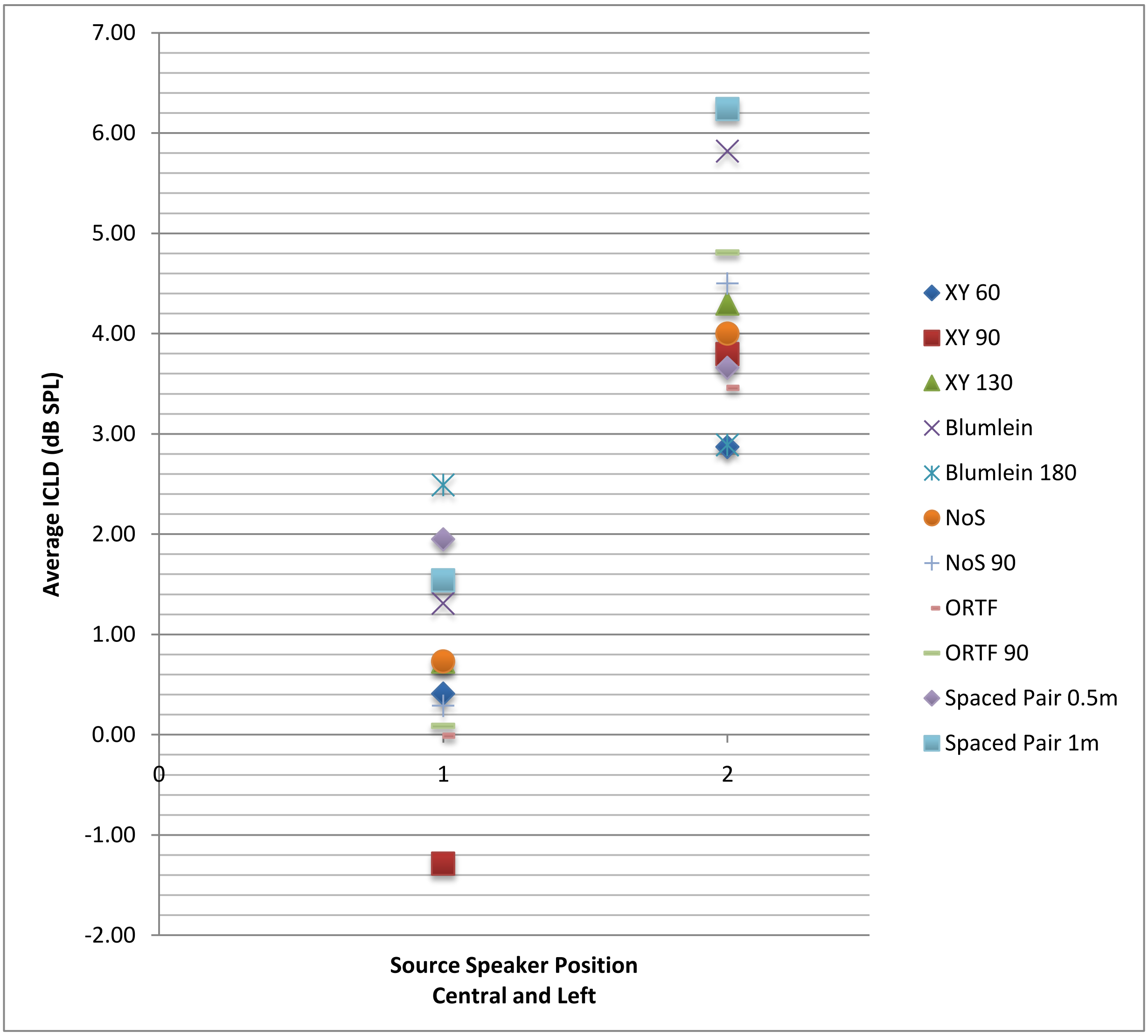
The results show that binaural recordings do have a very high degree of localisation accuracy as suggested from the research. The technique tracks the movement of the source very clearly however the results are less effective when played back over loudspeaker and noise from the subject, whose ears the microphones were mounted in, can be heard. The

Blumlein results demonstrate clearly the effects of flaws in the experiment, however when looking at the results from later experiments the accuracy of the Blumlein technique can be seen in the goniometer plots. The ORTF and NoS techniques both show the correct imaging cues however they are far more diffuse than those of the coincident techniques due to the combination of time and level differences between channels. The Spaced Pair recordings show increasingly clear localisation towards the edges of the image where the microphones are; the level differences between microphones becoming more obvious, confirming the ‘hole in the middle’ effect mentioned by research sources. The coincident techniques all maintain a positive phase correlation for most source positions while the spaced pair and near coincident techniques, which introduce a spacing between microphones suffer from increased phase cancellation as a result.

This experiment, while useful for assessing the localisation of different techniques and demonstrating the disadvantages of multiple simultaneous microphone techniques, does not consider the difficulties of capturing a spectrally complex musical source so should be considered alongside the results of other experiments to inform the design of the final recordings.

4.1.2 Measured Level Differences

By comparing the difference in sound pressure level between two channels in a stereo array for different sound source positions, a number of patterns can be observed. Although comparing the level differences at each frequency band provide interesting results it is easier to compare the results from multiple techniques when comparing the average level difference across all frequencies.



The average level differences show that the different techniques vary in their ability to locate sources based on level differences. According to the level difference formulae and research into localisation due to ICLD, a source 0° from the centre should have no difference in level between channels. However from the results it can be seen that many of the techniques vary from this, with the two spaced pair techniques both localising the two sources towards the left hand side of the image. When the speaker was moved 1.6m to the left this subtended an angle of 32° from microphone to source. The different techniques represented this movement by varying amounts, with the smaller angles of separation between microphones showing a smaller level difference between channels. The level difference between the pair of bi-directional microphones orientated at 180° to each other was an almost constant amount, as the sound source will always be nearer to the front lobe of one microphone and the rear of the other, creating a very small level different which will vary proportionally as the source moves. The two coincident ORTF techniques maintained a 0dB ICLD for the central source, even with the varied angle of microphone separation while the XY techniques varied wildly with the 90° separation having a 1dB increase in the right microphone channel for a centrally positioned source.

The second experiment was influenced by the previous results, which were highly subjective making it difficult to draw direct comparisons between techniques. This experiment was designed to give discrete, quantifiable data that can be used to: test the validity of the ICLD formulae, directly compare the spatial cues captured by different techniques and to assess the width of the stereo field possible in the room.

The results show that as the angle of separation between microphones increases, so does the resulting ICLD increasing the separation between positions within the recording angle. It was also observed that care needs to be taken with the orientation of microphone techniques as several of the results show an ICLD for the central source position, an error most likely caused by errors in the positioning of the stereo array.

By examining the individuals graphs of ICLD (see 10.1 Appendix 1. Full Results of Measuring Level Differences Between Channels) across the entire frequency range for different techniques more patterns begin to emerge. The ICLD can be seen to increase dramatically above ~1600Hz the frequency above which localisation is determined by level differences, as the phase differences that result from time delays become less noticeable at shorter wavelengths.

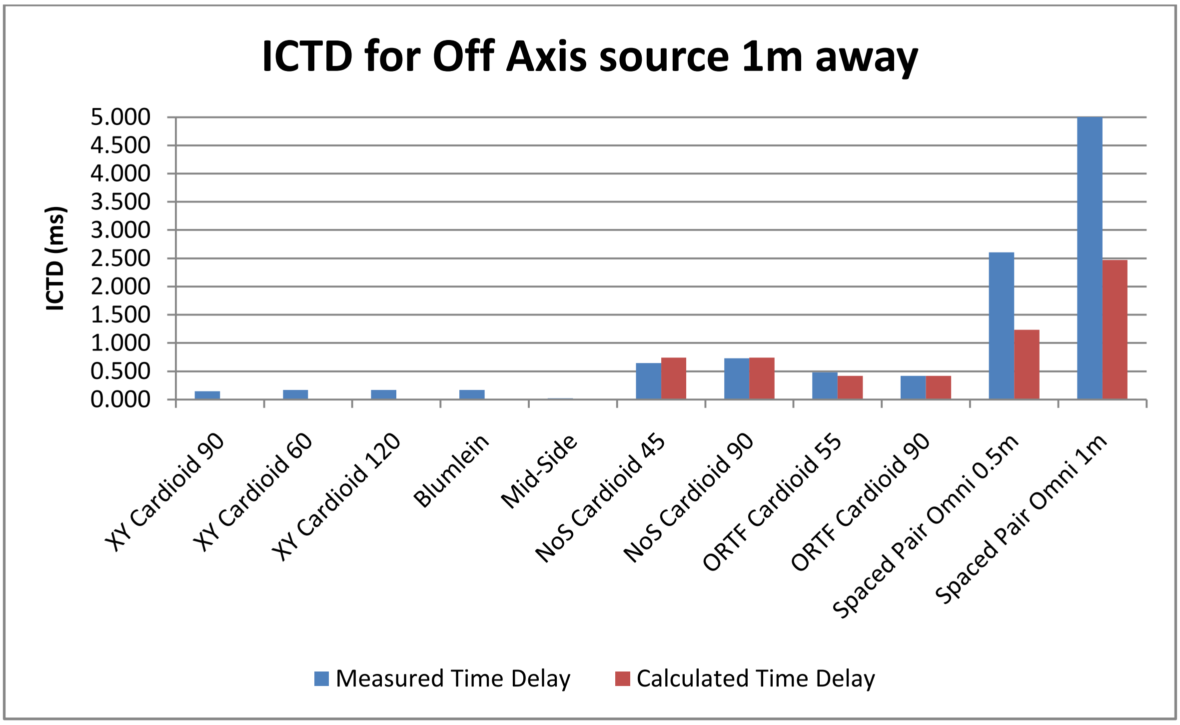
The difference in ICLD for different techniques is noticeable; the spaced pair with microphones positioned ±1m has an increase in ICLD of 5dB for the off centre source while the ±0.5m spaced pair shows an increase in ICLD of only 1.6dB for the same source position illustrating how small changed to microphone positioning can have a large difference on the spatial cues produced.

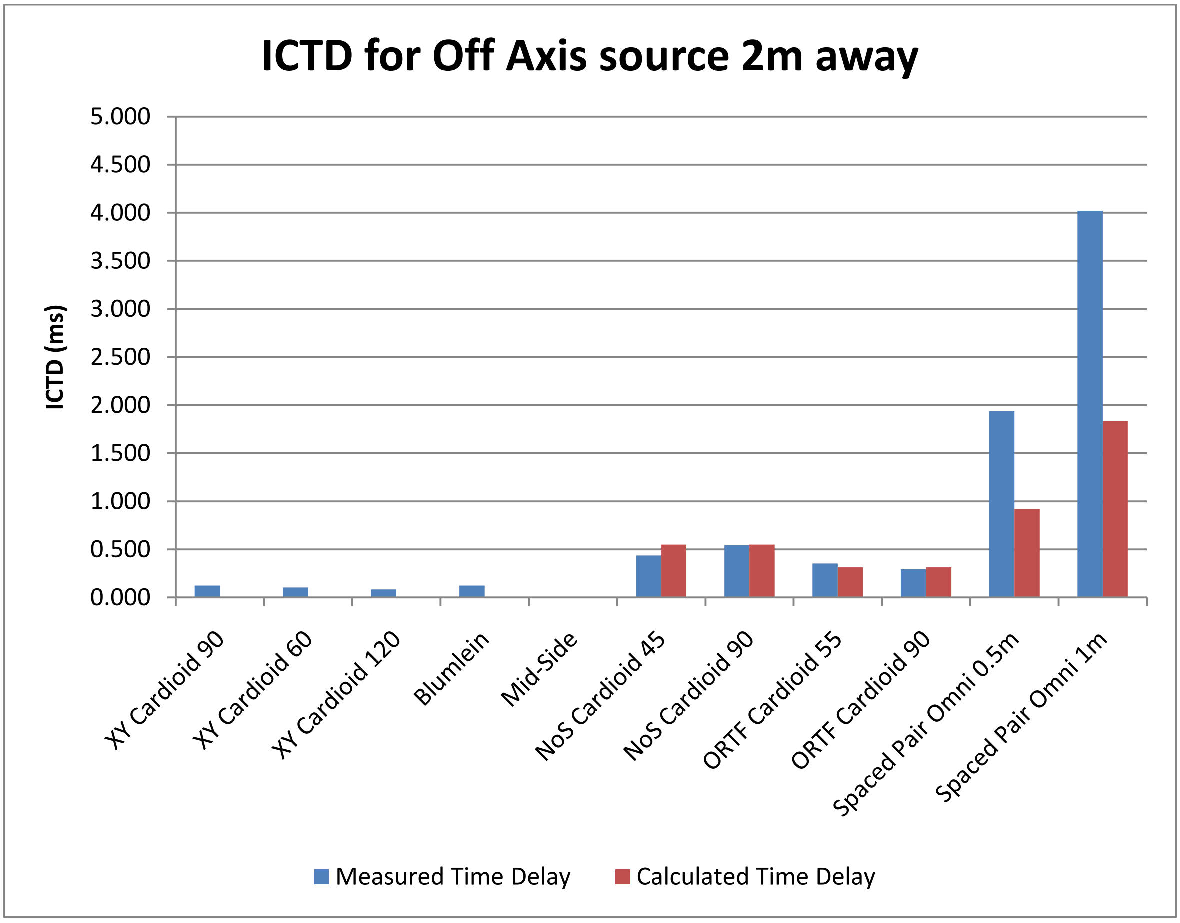
As well as confirming the validity of the ICLD formula this experiment helped to highlight that the room isn't ideally suited to creating stereo recordings. While the recording angle of the various techniques ranges from 40°-180°, using the space available it was only possible to subtend an angle of 32°, and this did not allow for much distance between microphone and source.

Although the ICLD formulae allows for the calculation of angles to produce the required level differences stated by Bartlett's listening tests as placing the source fully at one loudspeaker, it makes more sense to ensure that the entire recorded stage fits within the recording angle of the microphone array, even if the far edges of the field do not provide the level differences to position it fully at one speaker according to the formulae. The current level differences place the edge of the sound field about half way towards the loudspeaker, however with the near coincident techniques the ICTD will combine with the ICLD to localise the sound further towards the edge of the sound field.

4.1.3 Measured Time Delays

The results of this experiment, for the most part, follow the expected pattern of results calculated with the ICTD formula. According to the formula the central source positions should result in no time delay between channels for any of the techniques however the measured results show time delays of up to 0.5ms some of which will be introduced due to small errors in the positioning of the microphone arrays and of the central sound source. Most of these time delays are well below the threshold of perceiving an effect on the image location however, the larger delays especially those produced by the spaced pair techniques may result in phase cancellations and a distorted frequency response. The large time delay between channels for the widely separated spaced pair demonstrates the 'hole in the middle' localisation effect that occurs with the technique.





The time delays for the off axis source confirm that, as predicted by the formulae, coincident techniques have no ICTD and will therefore not experience the effects of phase cancellation. The near coincident arrays, which use ICTD as part of their localisation process, follow the calculated results quite closely with discrepancies most likely caused by small differences in the distance between the microphones that occurred when adjusting microphone angles.

When the measured time delays are compared with the results predicted from the formulae they are found to be fairly accurate with the discrepancies within 1-2ms of the predicted results for most techniques, however the results from the spaced pair technique demonstrate the difficulty in predicting the time delay caused by larger distances of microphone separation.

Although the coincident microphone techniques should have no ICTD for any source angle or distance, the practicalities of positioning the two capsules at the correct angles causes a small separation leading to small time delays. Because the time delays are only very short they will not have any effect on the image localisation however they may lead to some limited phase cancellation.

Another limiting factor in this experiment may have been the choice of a test tone at the higher bounds of frequencies where ICTD is used for image localisation, above 1600Hz the phase differences resulting from the shorter wavelengths becomes less noticeable and so localisation is inferred from ICLD above this point. The use of a lower frequency test tone may have increased the accuracy of these results.

4.1.4 Results of Localising Sound through Artificial Level Differences

10. Position 1 Stereo Recording.wav	11. Position 2 Stereo Recording.wav	12. Position 3 Stereo Recording.wav	13. Position 4 Stereo Recording.wav
Audio Clip 10 Position 1 Stereo Recording	Audio Clip 11 Position 2 Stereo Recording	Audio Clip 12 Position 3 Stereo Recording	Audio Clip 13 Position 4 Stereo Recording
14. Position 1 Artificial Level Differe	15. Position 2 Artificial Level Differe	16. Position 3 Artificial Level Differe	17. Position 4 Artificial Level Differe
Audio Clip 14 Position 1 Artificial Level Difference	Audio Clip 15 Position 2 Artificial Level Difference	Audio Clip 16 16. Position 3 Artificial Level Difference.wav	Audio Clip 17 Position 4 Artificial Level Difference

It was found that although artificial level differences positioned the sound source at the desired angle, there were noticeable differences between this and the stereo recorded using the Blumlein technique. The close microphone technique captures a different, tonally coloured version of the source, the proximity effect causing a significant rise in the low frequency content. The microphones proximity to the source means its signal is dominated by direct sound and captures more performance noises, such as strings squeaking, the plectrum striking the strings and small rattles against the body of the instrument than would be heard at a natural listening position or where the stereo array is positioned.

The stereo array captures a natural balance of direct-reverberant sound, it's frequency response isn't skewed towards low frequencies and as well as localising the sound source at the correct angle; the frequency response, signal level and ratio of direct-reverberant sound all contribute to providing the source with a sense of distance. The distance between the source and the array however, does mean that the stereo microphone signal lacks the clarity and signal to noise ratio of the close microphone recording. For the two signals to be 'blended' together and retain the full effect of the stereo imaging it was found that the close microphone signal had to be brought in slowly until a level was reached where it could sit beneath the stereo signal, adding presence and clarity without overwhelming the sense of localisation, reverberation and distance provided by the quieter stereo pair.

Audio Clip 18
Position 1-4 Stereo Recording

Audio Clip 19
Positions 1-4 Stereo Recording and Artificial Level Difference

The use of close microphone recording techniques highlighted the limitations of the technique, producing an unnatural tonal balance with too much clarity and detail. Also, due to the proximity effect, there is a clear bias towards the low frequency content further colouring the signal. This is in contrast to the more natural signal captured at the stereo array, as the sound has had a chance to propagate through the air, interacting with reflections and other components of the instrument, which more closely matches the sound that would be heard by a listener in the room.

Although the ICLD and ICTD that occur with different techniques can be calculated using formulae and then applied artificially to the close microphone signal, actually matching the two signals together provides mixed results. Due to the distance between the close microphone array and the stereo array when the two signals are combined additional phase cancellation can occur further colouring the frequency response of the reproduced signal.

4.1.5 Results of Investigating the Precedence effect

From the experiments investigating the precedence effect, it was observed that as the time delay between channels increases the apparent location of the image shifts towards the earliest arriving channel, consistent with the research findings and experiments conducted in section 3.3 Measuring Time Delays between Channels. As the delay between channels became greater than the delay necessary to fully locate the sound at one side of the image, and moves beyond the limits of the precedence effect, the listener begins to hear the two channels as individual sound sources.

20. Guitar Strum 0ms Delay.wav
Audio Clip 20
Guitar Strum 0ms Delay

21. Guitar Strum 20ms Delay.wav
Audio Clip 21
Guitar Strum 20ms Delay

22. Guitar Strum 40ms Delay.wav
Audio Clip 22
Guitar Strum 40ms Delay

23. Guitar Strum 60ms Delay.wav
Audio Clip 23
Guitar Strum 60ms Delay

24. Guitar Strum 80ms Delay.wav
Audio Clip 24
Guitar Strum 80ms Delay

25. Guitar Strum 100ms Delay.wav
Audio Clip 25
Guitar Strum 100ms Delay

26. Guitar Strum 120ms Delay.wav
Audio Clip 26
Guitar Strum 120ms Delay

With the musical, complex sound of guitar strumming the two signals can be recognised as a single source up to a delay of 40-60ms, depending on the listener. With a more continuous source the delay required before the two separate sources are identified is further increased.

27. Guitar Chord 0ms Delay.wav
Audio Clip 27
27. Guitar Chord 0ms Delay.wav

28. Guitar Chord 20ms Delay.wav
Audio Clip 28
Guitar Chord 20ms Delay

29. Guitar Chord 40ms Delay.wav
Audio Clip 29
Guitar Chord 40ms Delay

30. Guitar Chord 60ms Delay.wav
Audio Clip 30
Guitar Chord 60ms Delay

31. Guitar Chord 80ms Delay.wav
Audio Clip 31
Guitar Chord 80ms Delay

32. Guitar Chord 100ms Delay.wav
Audio Clip 32
Guitar Chord 100ms Delay

Once the delay between channels moves beyond the perceptual threshold of the precedence effect then the second channel is perceived as an echo of the original source originating from the other side of the stereo field. With a short percussive signal like a snare the effect of the delay is noticeable much earlier, around 40ms, as the transient nature of the wave means the first signal has already begun to decay before the signal from the delayed channel is heard.

33. Snare 0ms Delay.wav Audio Clip 33 Snare 0ms Delay	34. Snare 20ms Delay.wav Audio Clip 34 Snare 20ms Delay	35. Snare 40ms Delay.wav Audio Clip 35 Snare 40ms Delay	36. Snare 60ms Delay.wav Audio Clip 36 Snare 60ms Delay
37. Snare 80ms Delay.wav Audio Clip 37 Snare 80ms Delay	38. Snare 100ms Delay.wav Audio Clip 38 Snare 100ms Delay	39. Snare 120ms Delay.wav Audio Clip 39 Snare 120ms Delay	

These results confirm that the short ICTD caused by the ear spacing in the human hearing system and the spacing between microphones in near-coincident arrays will be transparent to the listener. It also explains how early reflections from the performance environment can be used to inform the location of the sound source, rather than being perceived as an echo.

The perceptual limits of the precedence effect limit the maximum spacing between microphones in a spaced pair, as too great a spread will cause ICTDs long enough for the listener to perceive two, incorrectly positioned sound sources.

4.1.6 Comparison of audio codecs

Although the final mastered files used to create the audio CD are 1411 kbps lossless wav files there is a strong possibility that the tracks will be encoded to a lossy format at a later date. To assess the effect of lossy compression on the final tracks a number of experiments were conducted.

2. White Noise Panning Wav.mov Video Clip 2 White Noise Panning Wav	3. White Noise Panning OGG.mov Video Clip 3 White Noise Panning OGG Vorbis	4. White Noise Panning MP3.mov Video Clip 4 White Noise Panning MP3

Using a sample of white noise which pans from left to right it can be seen that MP3 and Ogg Vorbis encode a panned source differently to the original wav file. In the MP3 results the effects of the mid side and intensity stereo coding are obvious as the noise makes a ‘jump’ from side to mid and then back to side. By contrast the Ogg Vorbis file, which achieves a similar level of file size compression, maintains a smooth transition from left through right although the stereo image does spread slightly more than in the original wav file. The frequency bandwidth of the MP3 file is the smallest of the 3, with a sharp cut off at 15.5kHz while the Ogg Vorbis file has a drop at 16kHz but maintains a reduced frequency response all the way up to 22kHz although the level is little above the noise floor.

Video Clip 5
Clarinet Recording Wav Encoding

Video Clip 6
Clarinet Recording Ogg Vorbis Encoding

Video Clip 7
Clarinet Recording MP3 Encoding

5. Chamber Music Recording Experiments

5.1 Introduction

Using the information gathered through research and tested with experiments two recordings were made, one of a clarinet quartet and the other of a string quartet. The recording methods and techniques used were selected to create the most 'aesthetically pleasing', spatially accurate recording of the small ensemble performing in a small space.

5.2 Design of Final Recording

Based on the experiments conducted previously a number of parameters for the recording had already been decided. The chosen microphone techniques were positioned in their own space, with no more than 3 techniques setup simultaneously to avoid the shadowing and interference experienced with the initial experiments. Where possible AKG C414s were used as the large diaphragm, high sensitivity and switchable polar patterns made them suitable for most techniques, however with the XY and ORTF techniques AKG C451bs were used as the small diaphragm and housing design simplified the positioning of the microphones within the array.

Because of the previous experiments demonstrating that E1 was too small to create the necessary angles and distance between microphones and source the final recordings were designed around the live room of Studio 4. Although this room is not acoustically treated to the same standard as E1, and the focusrite pre amps in the Control 24 have a greater THD value and less headroom than those found in the SSL AWS 900, it was decided that because the focus of this project is the layout of the performers and recording of the stereo field, the larger live room justified the subsequent compromises.

The research and experiments highlighted a variety of microphone techniques however to simplify the recording and comparison process a smaller number of these techniques were selected. Due to the individualised, sometimes noisy recordings produced by the in-ear binaural microphones this technique was not used as it occupied a lot of space and interfered with the positioning of other techniques as well as having the potential to distract the performers by placing someone in the live room with the performers. Because of the similarity of the NoS and ORTF techniques it was decided to only use the ORTF technique as the experimental results showed good separation between sources with a wide stereo field. Although a near coincident technique with a baffle could have been attempted to roughly emulate the function of the human head, it was decided that this complicated the placement and setup too much in the limited time frame the performers were available.

The positioning of the performers within the room was designed around the chosen microphone techniques, research indicating that the techniques had recording angles of between 80°-110° so the position of the performers was calculated using trigonometry so that the outside performers were positioned ±40° and the inner performers at ±20° at a distance of 3m from the main microphone arrays, except for the XY array which required the source to be spread 180° around the array for a full stereo spread to be achieved on reproduction. The distance from the performers allowed the sound to propagate through the room first so that a balance of direct and reverberant sound would be captured, which is one of the advantages of using a stereo array over close microphone techniques.

Figure 24 Positioning of elements for final recording

Position No	X position	Y Position	Angle from Centre
1	-1.93m	2.3m	-40°
2	-1.03m	2.82m	-20°
3	1.93m	2.3m	20°
4	1.03m	2.82m	40°

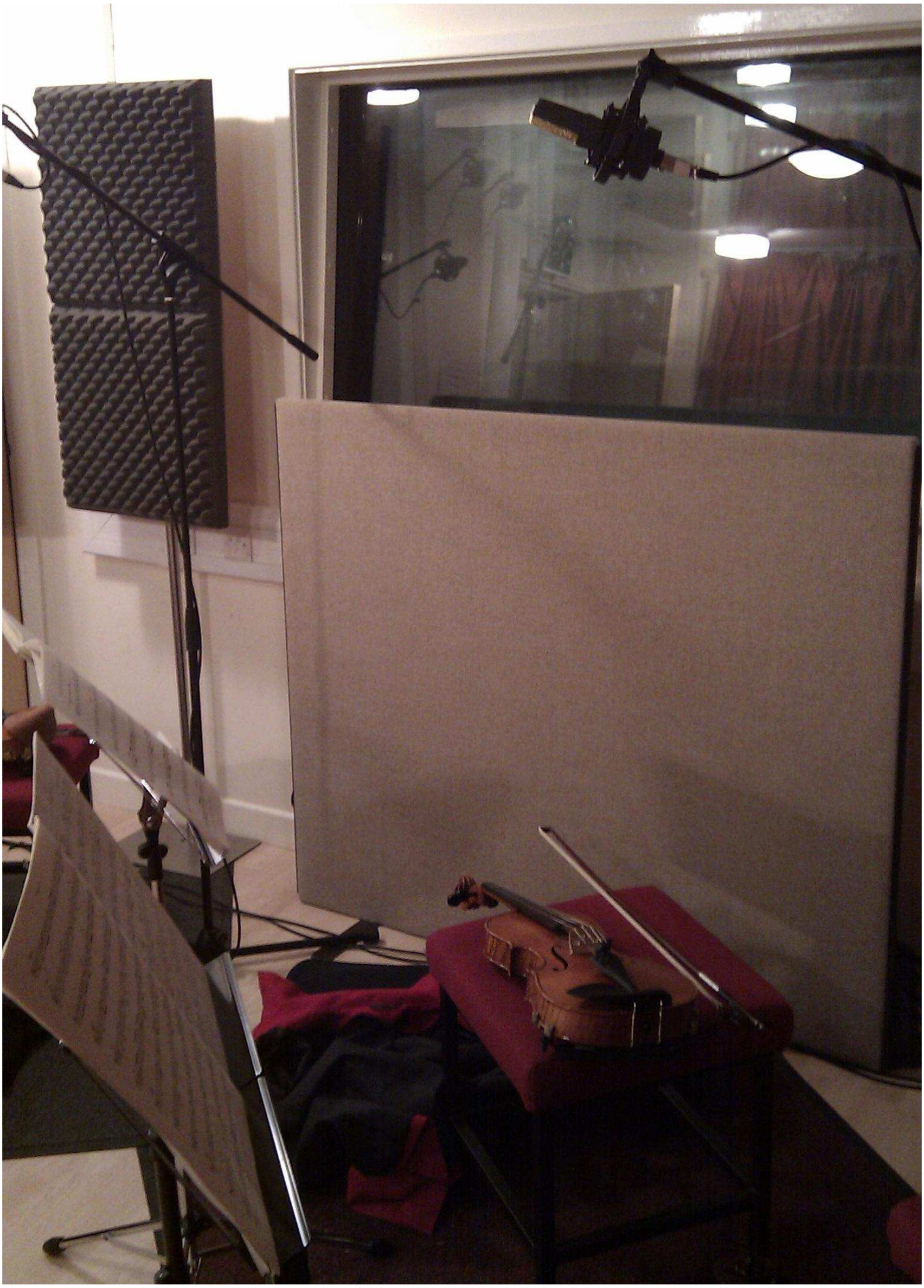


Figure 25 AKG C414s positioned overhead as spot microphones

To enable comparisons with close microphone, pan-pot stereo recording techniques spot microphones were setup close to the performers in the optimum positions, using C414s due to their high sensitivity and wide frequency response.

Figure 26 Spaced Pair, Mid-Side and ORTF arrays setup during the interval

To comply with health and safety concerns all cables were taped to the floor or placed out of the way, running parallel to each other where possible and crossing at 90° to minimize any interference that might be occur. To capture the highest quality recording the session was recorded at 96kHz sample rate allowing for the full range of audible frequencies to be encoded and reducing the noise and artifacts introduced by aliasing at lower sample rates. By recording at 24-bit depth, greater dynamic range can be encoded in the signal allowing for greater separation between the signal and noise floor. This is especially important with the stereo microphone techniques positioned at a distance, as pre-amps add more noise at the higher gain values, which would be more important for recording necessary signal levels at 16 bit. It also allows for sufficient headroom, as classical music is typically more dynamic than most other forms of program material. To synchronize the sample rate of the session and the two Digidesign 96 A/D interfaces the clock was set externally by the Digidesign Sync HD to ensure that sampling occurred at regular intervals, in sync between the two interfaces. The session was setup with mono tracks for each of the spot microphones and stereo tracks for each array, with each piece being tracked onto its own appropriately named playlist.

To ensure the smooth running of the session and to get the best from the performers, the live room was setup half an hour before they arrived. Several pieces were then tracked using the XY, Blumlein and Spaced Pair technique before giving the performers a 15 minute break to avoid fatigue and allowing for the setup of the ORTF, Mid Side and Spaced Pair techniques for tracking the final pieces.

5.3 Results of Final Recording

For the purposes of this report the comparisons and observations of the different techniques will be based on headphone monitoring, as loudspeaker monitoring introduces a number of uncontrolled variables; from the quality of the monitoring equipment, the modes and reflections of the listening environment and the relative positioning of the listener between the two loudspeakers- all of which impact the listening tests, introducing bias and unpredictability beyond the scope of this report.

The final recordings were limited by a degree of equipment failure, with only 1 A/D interface recognised during the clarinet recording, limiting the number of channels available for recording and subsequently no spot microphone recordings were captured for that session. The string recording was able to utilise spot microphones, however the room layout and microphone stand size severely limited their placement so they may have been in less than optimum positions. Comparing the different microphone techniques from each recording, it is highly subjective as to which produces the ‘preferred’ recording however observations about stereo spread, frequency response, phase coherence and quality of recording for each technique can be directly compared.

5.3.1 Blumlein Results

40. Clarient Piece
2_02 Blumlein.wav

Audio Clip 40
Blumlein recording of Clarinet Quartet

41. Strings Piece
2_01 Blumlein.wav

Audio Clip 41
Blumlein recording of String Quartet

The average RMS of the Blumlein recordings was between -35dB and -40dB for the two sessions, relatively quiet however within the limits of a 24bit recording. The captured frequency response is fairly even although there was a tendency towards rolling off above 5kHz possibly due to reflections and phase interference within the room. As expected from a coincident microphone technique the recording maintains strong positive phase coherence throughout. When listening back over headphones a clear separation between instruments is observed, especially for the performers at the edge of the group. The balance between performers within the group is even, with no one performer becoming more dominant unless stated within the score. Initially high gain settings caused some distortion on early clarinet recordings however recordings made with the reduced gain settings do not suffer from this. The string recordings were clear with strong low frequency content captured from the cello. When blended with the spot microphones clarity is added but somewhat at the expense of the natural feeling of cohesion, achieved from the stereo recording. The pizzicato sections benefited the most from the spot microphones, as these quiet sections are harder to capture with distance microphones.

5.3.2 XY Results

42. Clarinet Piece 2_02 XY.wav

Audio Clip 42
XY recording of Clarinet Quartet

43. Strings Piece 2_01 XY.wav

Audio Clip 43
XY recording of String Quartet

The average RMS level of the XY technique was between -37dB and -43dB across the two recordings, the reduced levels are a consequence of the lower sensitivity of the AKG C451b's that were used for this technique and increasing the gain to compensate increased the noise floor. The recordings from this array were strongly affected by a 50/60Hz hum, most likely as the result of a grounding fault or components not up to the same standards as those within the C414. The frequency response of this technique was similar to that of the other techniques although with slightly more high frequency content perhaps due to the proximity of the array to the performers, and lack of low frequency due to the response of the smaller diaphragm. The proximity of the array allows the image to spread around the listener however the separation between performers is less obvious and locating the individual performers within the image is not as easy. Another consequence of the proximity of the array is the increase in performer noise, evident in the clarinet recordings where the key and breath noise are far more prominent than in the other techniques. To reduce the effect of this, hypercardioid microphones could have been used to place the array further from the source. When combined with the spot microphones the low frequency response improves, due to the large diaphragm microphones used however, the close proximity of these microphones further increases the amount of performer noise captured. Again the coincident microphone spacing ensured a high degree of phase coherence, avoiding cancellation.

5.3.3 Spaced Pair Results

Audio Clip 44
Spaced Pair recording of Clarinet Quartet

Audio Clip 45
Spaced Pair recording of String Quartet

The spaced pair recordings had an average RMS level of between -35dB and -32dB across the two sessions with a similar high frequency roll off beginning around 5kHz further suggesting that the room acoustics are the cause rather than the microphone arrays. Although this technique creates a spacious sounding recording which, with a small amount of artificial reverb sounds pleasing and natural, it is difficult to derive more than a vague sense of direction and separation for each of the individual performers. This may be due to the varying phase relationship between the two channels, caused by the large spacing between microphones. The use of spot microphones with the spaced pair array seems to improve the location accuracy however it also removes some of the sense of diffuse 'spaciousness' that was present when exclusively using the stereo array.

5.3.4 ORTF Results

Audio Clip 46
ORTF recording of Clarinet Quartet

Audio Clip 47
ORTF recording of String Quartet

The ORTF near coincident technique had an average RMS of between -35dB and -45dB when recording the quieter string sound source suggesting that perhaps the less sensitive C451b is not suited for distance microphone techniques. Again there is evidence of a roll off at 5kHz and as with the other technique using the C451b the interference at 50Hz is noticeable, with the reduced sensitivity of the microphone giving a 'thin' sounding recording lacking the low frequency response of the C414. Although the phase correlation between channels isn't as strong as the coincident microphone techniques there is less cancellation than the spaced pair recordings, with this technique producing accurate, clearly separated imaging for each of the performers and their parts. Had this technique used more sensitive microphones then the results would have been drastically better.

5.3.5 Mid-Side Results

Audio Clip 48
Mid Side recording of Clarinet Quartet

Audio Clip 49
Mid Side recording of Clarinet Quartet

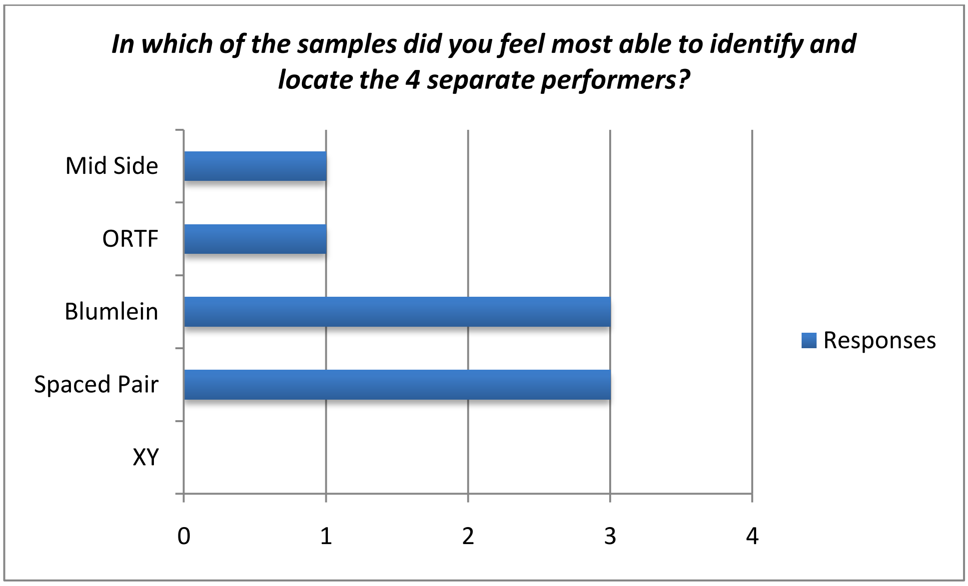
The mid-side pair, once decoded to left/right stereo had an average RMS of -38dB for both recordings. Again the 5kHz roll off is present but the frequency response for the main body of the recording should be better as the source is fully on axis to the mid microphone. There is, for the most part, a strong positive phase correlation however at some points within the recordings there is evident phase cancellation. If bounced to mono, the two side lobes should cancel leaving only the, perfectly in phase, mono mid signal. When listening to the L/R stereo recording there is a pleasant stereo width and the separation of source is evident however the location of these sources wasn’t as obvious as with other techniques. The string recordings using this technique seemed very harsh while the clarinet recordings came across as full bodied and clear most likely due to the on-axis mid microphone.

5.4 Listening Tests

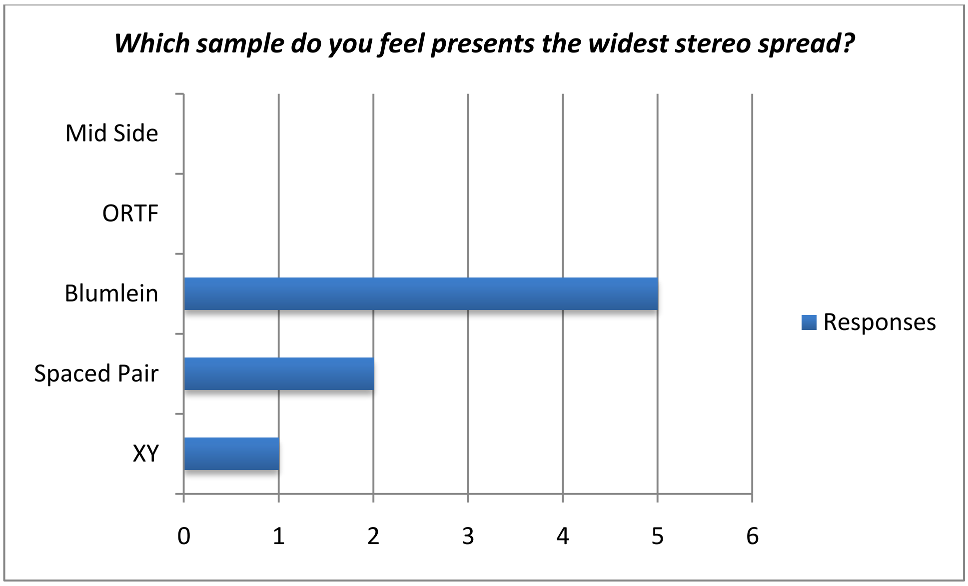
To assess the suitability and performance of the different microphone techniques a sample of each was played to a small listening audience over headphones, some locally and some over the Internet. The subjects ranged from experienced engineers with a technical understanding of the recording process to interested listeners with no specialist knowledge, allowing for a range of responses to be collected.

Because of the highly subjective nature of the investigation, a small sample size was questioned, as an impractically large sample size would be required to draw any 'definitive' conclusions.

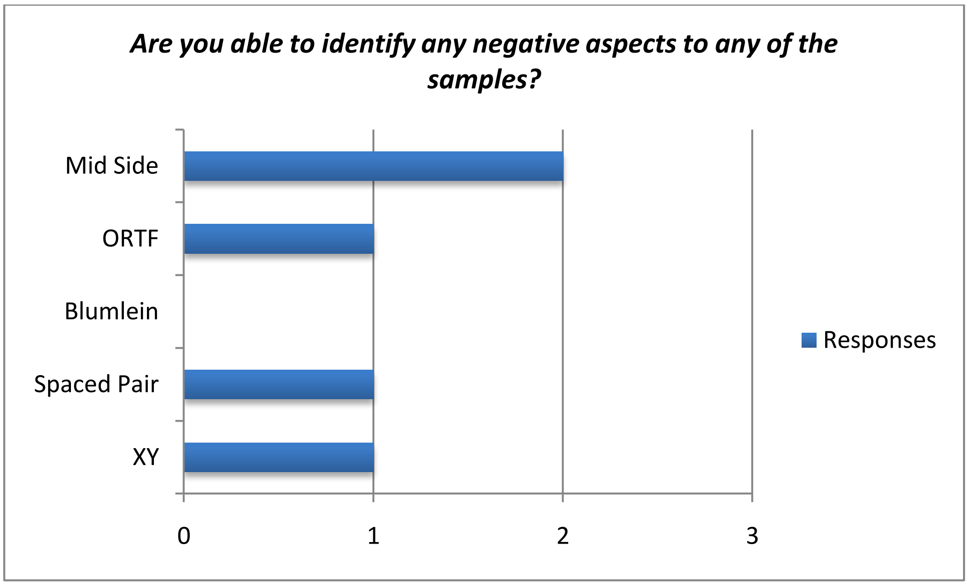
The subjects were asked to identify the sample they felt presented the clearest localisation for the 4 separate sources, which presented the widest stereo image, which contained any negative aspects and which gave them the most pleasant overall listening experience. The full questionnaire responses are available in Appendix 2 however the results are summarised below.



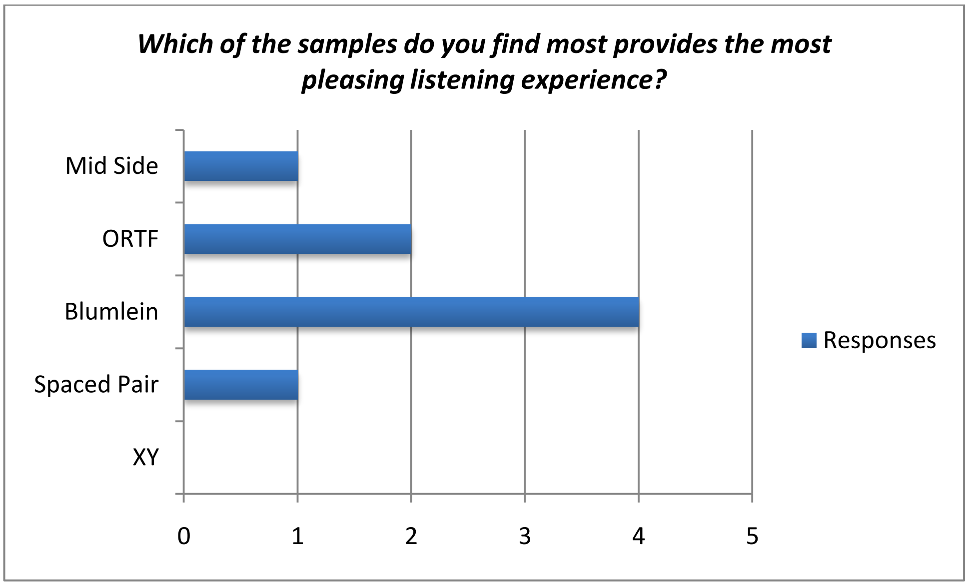
Surprisingly a large number of subjects identified the spaced pair sample as providing the most distinction between the 4 performers even though the research suggested that the spaced pair would create indistinct hard to localise images. While the image may have been difficult to localise exactly they obviously captured a clear sense of 4 separate instruments and performers rather than blurring them together. A large proportion of the subjects also responded that the Blumlein samples created highly accurate images, which more closely reflects the results expected from the research.



The widest stereo image was found to be given by the Blumlein technique, which accurately fitted the performance stage across the entire image. The spaced pair was also found to provide a wide stereo spread, and while the mid-side sample wasn't found to create the widest of stereo images this could have been improved by altering the ratio of mid to side signal in the mix.



Negative aspects were identified for each of the samples except the Blumlein samples. XY was found to sound too muffled and close to the performers, mid-side produced an image, which localised behind one subject. Another listener found both the Mid-Side and ORTF samples lacking in 'stereo definition' while the spaced pair had 'too much unwanted *background noise which didn't enhance the feel of the piece.*'



The majority of tests subjects found the most pleasant listening experience was provided by the Blumlein technique, with subjects reporting it 'sounding clearest', providing a stereo spread that is 'natural when listening through earphones, how it would sound in real-life' and 'providing the most full bodied stereo sound without unwanted noise'. The

ORTF samples were also found by some to be pleasing while listening over headphones, perhaps because of its approximation of the human hearing system.

5.5 Production of Mastered Tracks

To bring the final mixes up to ‘CD quality’ a number of processes were undertaken. At the mix stage a small amount of artificial reverb was added to help the recordings sit together in a more reverberant space. This added sustain and decay tails more representative of the environment the pieces were composed for.

A small amount of compression was applied across the tracks using a limiter to control the high intensity peaks, reducing the dynamic range slightly without squashing the tracks. The level of compression and track gain was controlled based on monitoring via the K system so that the ‘nominal’ output level was -20dB below 0dBFS allowing for 20dB of ‘headroom’ for louder dynamic parts of the performance.

Gentle EQ curves were used to boost the low frequencies, adding body to the instruments which otherwise felt thin and lacking. A gentle high frequency roll off was also applied to reduce the harsh shrill frequencies, which were cutting through. A sharp notch filter was applied at 50Hz for recordings made with the two techniques that used AKG C451b s to counter the hum introduced by these microphones.

5.5.1 Comparison of Audio Codecs

Comparing the results of the different codecs on the mastered mixes the effects of compression seem less noticeable than in the white noise experiment however some key differences can be observed. The MP3 file has less stereo correlation, with the phase meter showing a more diffuse plot than the original wav. Although separation can still be heard between individual performers, the position of the different elements seems to have been grouped together into discrete sections rather than the full stereo field reproduced by the wav file. The Ogg Vorbis file gives a good recreation of the stereo positioning, although slightly less correlated than the wav file the location of the performers seems more accurate than the MP3 file.

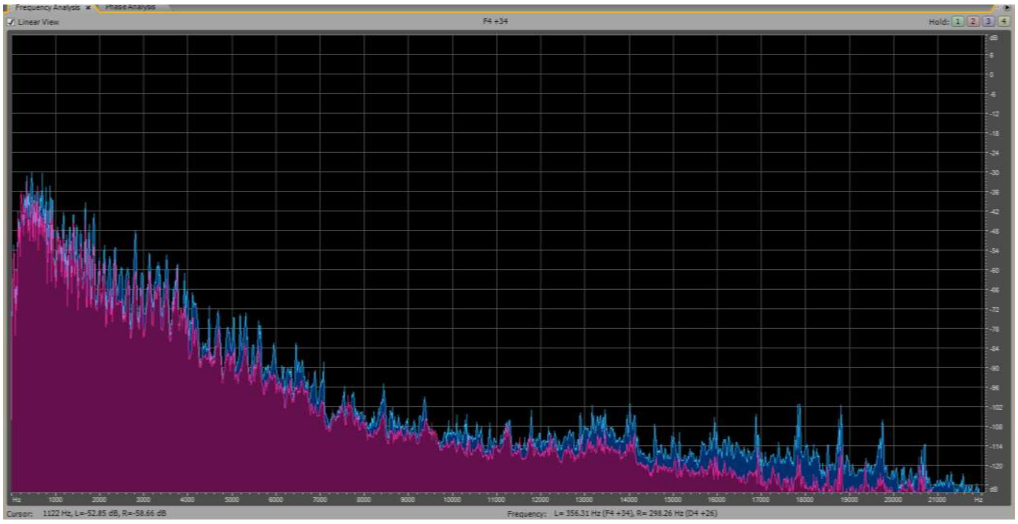


Figure 27 FFT analysis of Blumlein recording Wav file

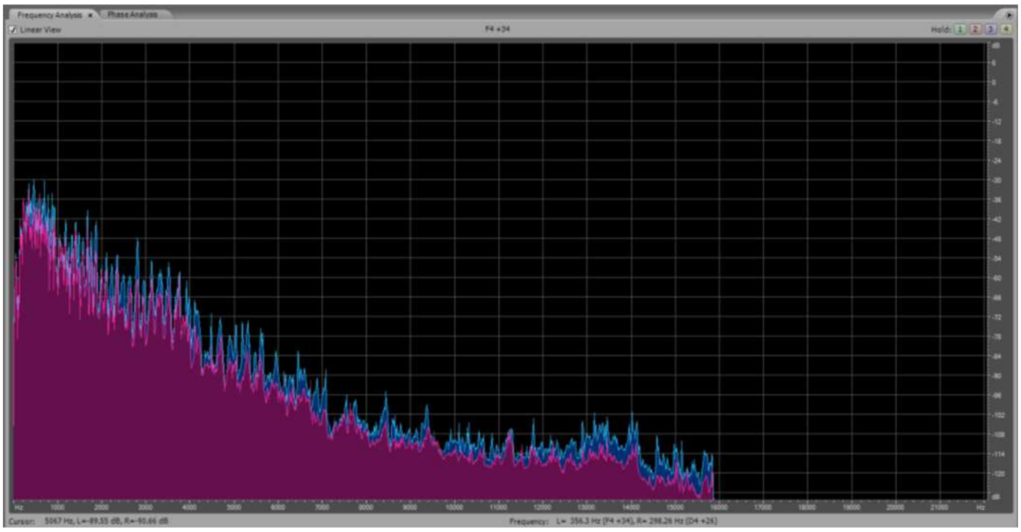


Figure 28 FFT analysis of Blumlein recording 128 CBR Ogg Vorbis file

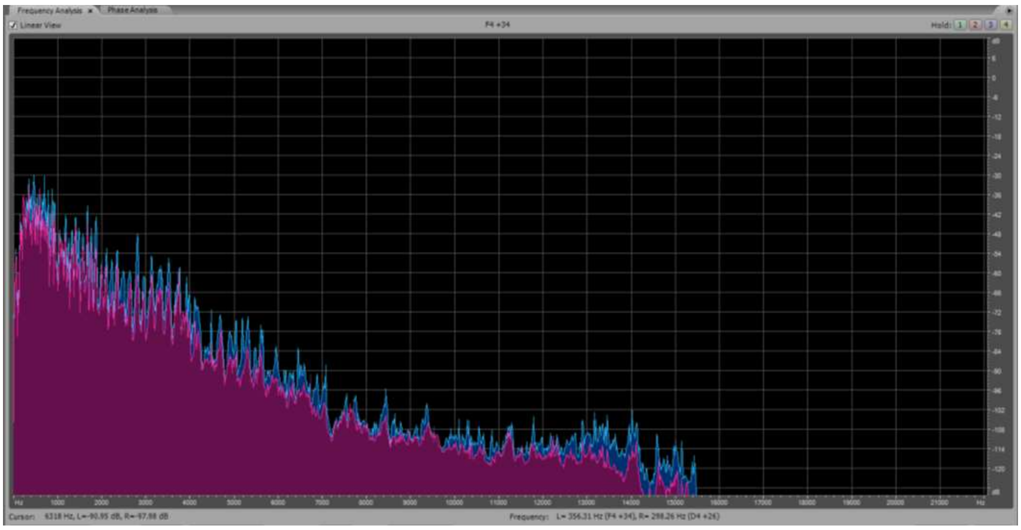


Figure 29 FFT analysis of Blumlein recording 128 CBR MP3 file

Comparing the FFT plots for the 3 formats it can be observed that the Ogg Vorbis file almost exactly mirrors the original wav up to just below 16kHz while the MP3 file shows some variation in the low frequencies and the response falls away swiftly after 14kHz.

50. Perptuum Mobile Blumlein Mastered Wa

Audio Clip 50 Blumlein Recording Wav Encoding

51. Perptuum Mobile Blumlein Mastered OG

Audio Clip 51 Blumlein Recording Ogg Vorbis Encoding

52. Perptuum Mobile Blumlein Mastered MP

Audio Clip 52 Blumlein Recording MP3 Encoding

53. Perptuum Mobile Blumlein Mastered FL

Audio Clip 53 Blumlein Recording FLAC Encoding

Audio Encoder	Bit Rate	File Size
Wav	1411 kbps	36.3MB
MP3	128 kbps	3.3MB
Ogg Vorbis	128 kbps	3.3MB
FLAC	1411 kbps	12.6MB

Ideally the finished tracks would be listened to, and stored as, a lossless format- either as the uncompressed wav files on CD or encoded using a lossless codec such as FLAC or Apple Lossless. Where this isn’t feasible due to storage restraints or compatibility with playback devices encoding to a lossy format should be done at the highest bit rate possible to preserve as much of the original signal as possible. If encoding to MP3 then mid-side encoding should be avoided as it limits the spread of the stereo field. If feasible then Ogg Vorbis should instead be used as it allows for higher fidelity than the equivalent bit rate MP3, making it ideally suited for embedding in webpages, supported personal music players and other bandwidth or storage limited applications.

6. Conclusions

The original objective of this project was to investigate the different methods of creating a stereo recording and explore how these could be used to record an ensemble, retaining the stereo cues upon reproduction. Research highlighted how the human hearing system localises sound based on time and level difference between the ears and how these cues are captured by the various microphone arrays to allow for the localisation of individual sources. The level differences and time delays that result from different angles and distances to the source can be calculated with a reasonable degree of accuracy as confirmed by experimental testing. Using these level differences it is possible to artificially position close microphone recordings to match the location of the elements within a recording made with a stereo array however the use of time delays proved more difficult, less reliable and was also found to cause phase cancellation. The level of close microphone signal had to be kept below the stereo array due to the obvious tonal differences between the two signals, the phase cancellation caused by the difference in distance from the source for the two arrays and to maintain the ratio of direct to reverberant sound.

The final ensemble recordings are the product of research and experiments creating a number of recordings of the two ensembles using a number of techniques, which could then be directly compared. From the results of the limited listening tests the Blumlein recordings were found to be preferred due to the clarity, giving a strong sense of stereo imagery with the performers seeming to surround the listener, as if they were in the performance space.

There are a number of areas for expansion, building on the findings of this report with research into ambisonics, binaural recording and surround sound recording and reproduction where the facilities and equipment are available. These areas all build directly on the principle of recreating a complex sound field at the ears of the listener although using different methods of capture and reproduction to accomplish this. The principles used when recording the two ensembles could also be scaled up to record a larger ensembles, observing the effects on the recordings produced to assess the suitability of the various techniques for use with larger ensembles.

With regards to the findings of this report, there are a number of improvements, which could be made to further investigate the microphone techniques. The XY and ORTF recordings both suffered adverse effects due to the smaller, less sensitive condenser microphones, which were used to simplify the positioning of the microphones. With the same microphones used for each array the recordings would have been more directly comparable. Despite the difficulties originally experienced when attempting this, it would be possible with a larger environment and the correct equipment to capture a signal source with all of the stereo arrays simultaneously to allow for true comparisons to be drawn.

7. Recommendations

The research, experiments and final recordings conducted as part of this report recommend the use of the Blumlein microphone array for recording a small ensemble within a small environment. High sensitivity, large diaphragm condenser microphones should be used in conjunction with low noise pre-amps to achieve a satisfactory signal level while using distance microphone techniques. If the recording environment has sufficient space, then spot microphones can be used and subtly mixed with the main stereo array to add clarity. Controlled listening conditions can be achieved with a high quality pair of monitoring headphones, however it is also advisable to monitor the recordings over a pair of correctly configured loudspeakers, as a large proportion of the end listeners will playback the final recordings over loudspeaker.

To capture time and level differences necessary to spread the stereo image from far left to right, it is important that the ensemble is distributed evenly across the entire recording angle of the chosen stereo array. To position elements of the recording within the stereo field the formulae outlined within this report allow for the prediction of the time and level differences that occur for any combination of source angle and distance. This can then be used in conjunction with previous listening tests to predict the apparent source location within the final stereo field.

8. Reference List

AKG AKG C414 Polar Pattern, 2009 [online], Available from: http://www.akg.com/site/products/powerslave,id,1128,pid,1128,nodeid,2,_language,EN,view,specs.html. last accessed: 08/04/2011.

Bartlett, B (1991a). Stereo Microphone Techniques. 1st ed. Boston: Focal Press. Page 1

Bartlett, B (1991b). Stereo Microphone Techniques. 1st ed. Boston: Focal Press. Page 41

Bartlett, B (1991c). Stereo Microphone Techniques. 1st ed. Boston: Focal Press. Page 56

Bartlett, B (1991d). Stereo Microphone Techniques. 1st ed. Boston: Focal Press. Page 64

Bartlett, B (1991e). Stereo Microphone Techniques. 1st ed. Boston: Focal Press. Page 22

Bartlett, B (1991f). Stereo Microphone Techniques. 1st ed. Boston: Focal Press. Page 96

Bartlett, B (1991g). Stereo Microphone Techniques. 1st ed. Boston: Focal Press. Page 98

Bartlett, B (1991h). Stereo Microphone Techniques. 1st ed. Boston: Focal Press. Page 99

Bartlett, B (1991i). Stereo Microphone Techniques. 1st ed. Boston: Focal Press. Page 79

Bartlett, B (1991j). Stereo Microphone Techniques. 1st ed. Boston: Focal Press. Page 81

Bartlett, B (1991k). Stereo Microphone Techniques. 1st ed. Boston: Focal Press. Page 48

Bartlett, B (1991l). Stereo Microphone Techniques. 1st ed. Boston: Focal Press. Page 44

Bartlett, B (1991m). Stereo Microphone Techniques. 1st ed. Boston: Focal Press. Page 45

Bartlett, B (1991n). Stereo Microphone Techniques. 1st ed. Boston: Focal Press. Page 34

Bartlett, B (1991o). Stereo Microphone Techniques. 1st ed. Boston: Focal Press. Page 100

Bartlett, B (1991p). Stereo Microphone Techniques. 1st ed. Boston: Focal Press. Page 112

Bartlett, B (1991q). Stereo Microphone Techniques. 1st ed. Boston: Focal Press. Page 116

Bartlett, B (1991r). Stereo Microphone Techniques. 1st ed. Boston: Focal Press. Page 117

Bartlett, B and Bartlett, J (2009a). Practical recording techniques : the step-by-step approach to professional audio recording. 5th ed. Burlington, MA: Focal Press. Page 83

Beament, J (2001a). How we hear music: The relationship between music and the hearing mechanism. Woodbridge: The Boydell Press. Page 128.

Beament, J (2001b). How we hear music: The relationship between music and the hearing mechanism. Woodbridge: The Boydell Press. Page 129.

Breebaart, J and Faller, C. (2007a). Spatial audio processing : MPEG surround and other applications, Chichester West Sussex England: Hoboken NJ: John Wiley & Sons. Page 7.

Breebaart, J and Faller, C. (2007b). Spatial audio processing : MPEG surround and other applications, Chichester West Sussex England: Hoboken NJ: John Wiley & Sons. Page 12.

Brice, R (2001a). Music Engineering. 2nd ed. Oxford: Newnes. Page 305.

Brice, R (2001b). Music Engineering. 2nd ed. Oxford: Newnes. Page 41.

Neumann KU100 Binaural Microphone, 2000 [online], Available from: http://www.neumann.com/?lang=en&id=current_microphones&cid=ku100_description last accessed: 12/4/2011.

Pohlmann, K (2011). Principles of digital audio 6th ed. New York: McGraw-Hill/TAB Electronics. Page 409

RØDE Microphones. NT1-A Polar Pattern. Available: http://www.rodemic.com/images/mics/nt1-a_polar.jpg. Last accessed 07th November 2010.

Rumsey, F (1989). Stereo sound for television. Boston: Focal Press. Page 8

Rumsey, F (2001a). Spatial Audio. Oxford: Focal Press. Page 171

Rumsey, F (2001b). Spatial Audio. Oxford: Focal Press. Page 172

Rumsey, F (2001c). Spatial Audio. Oxford: Focal Press. Page 173

Rumsey, F (2001d). Spatial Audio. Oxford: Focal Press. Page 55

Rumsey, F (2001e). Spatial Audio. Oxford: Focal Press. Page 56

Rumsey, F (2001f). Spatial Audio. Oxford: Focal Press. Page 62

Rumsey, F (2001g). Spatial Audio. Oxford: Focal Press. Page 69

Rumsey, F (2001h). Spatial Audio. Oxford: Focal Press. Page 70

Rumsey, F & McCormick, T (2006a). Sound and Recording. 5th ed. Oxford: Focal Press. Page 449.

Rumsey, F & McCormick, T (2006b). Sound and Recording. 5th ed. Oxford: Focal Press. Page 447.

Rumsey, F & McCormick, T (2006c). Sound and Recording. 5th ed. Oxford: Focal Press. Page 458.

Rumsey, F & McCormick, T (2006d). Sound and Recording. 5th ed. Oxford: Focal Press. Page 476.

Rumsey, F & McCormick, T (2006e). Sound and Recording. 5th ed. Oxford: Focal Press. Page 436.

Rumsey, F & McCormick, T (2006f). Sound and Recording. 5th ed. Oxford: Focal Press. Page 444.

Rumsey, F & McCormick, T (2006g). Sound and Recording. 5th ed. Oxford: Focal Press. Page 431.

Rumsey, F. & McCormick, T (2009a). Sound and recording 6th ed., Oxford: Focal Press. Page 484

Rumsey, F. & McCormick, T (2009b). Sound and recording 6th ed., Oxford: Focal Press. Page 482

Streicher, R & Dooley,W (1985). Basic Stereo Microphone Perspectives- A Review. Audio Engineering Society . 33 (7/8), 548-556.

Williams, M (1987). Unified theory of microphone systems for stereophonic sound recording. In Proceedings of the 82nd Conference of the Audio Engineering Society. London. Page 1

9. Bibliography

Bartlett, B. (1991). Stereo Microphone Techniques. Oxford: Focal Press.

Brice, R (2001). Music Engineering. 2nd ed. Oxford: Newnes.

Holman, T. (2007). Surround Sound, Second Edition: Up and running 2nd ed. Oxford: Focal Press.

Blauert, J (1996). Spatial Hearing: Psychophysics of Human Sound Localization. Cambridge, USA: MIT Press.

Blumlein, A. (1933). Improvements in and relating to Sound-transmission, Sound-recording and Sound-reproducing Systems. British Patent 394,325 June 14th 1933

Cooper, D & Bauck, J. (1989). Prospects for Transaural Recording. Audio Engineering Society. Vol 37 No 1/2.

Owsinski, B. (2009). *The Recording Engineer's Handbook* 2nd ed. Boston: Course Technology PTR.

Pulkki, V. & Karjalainen, M., (2001). Localization of amplitude-panned virtual sources I: Stereophonic panning. Journal of the Audio Engineering Society, 49(9), 739–752.

Faller, C. (2005). Pseudostereophony Revisited. In Audio Engineering Society Convention 118.

Rumsey, F. (1989). Stereo sound for television, London ;Boston: Focal Press.

Rumsey, F. (2001). Spatial Audio, Oxford: Focal Press.

Rumsey, F. & McCormick, T. (2006). Sound and Recording, Fifth Edition: An Introduction 5th ed. Oxford: Focal Press.

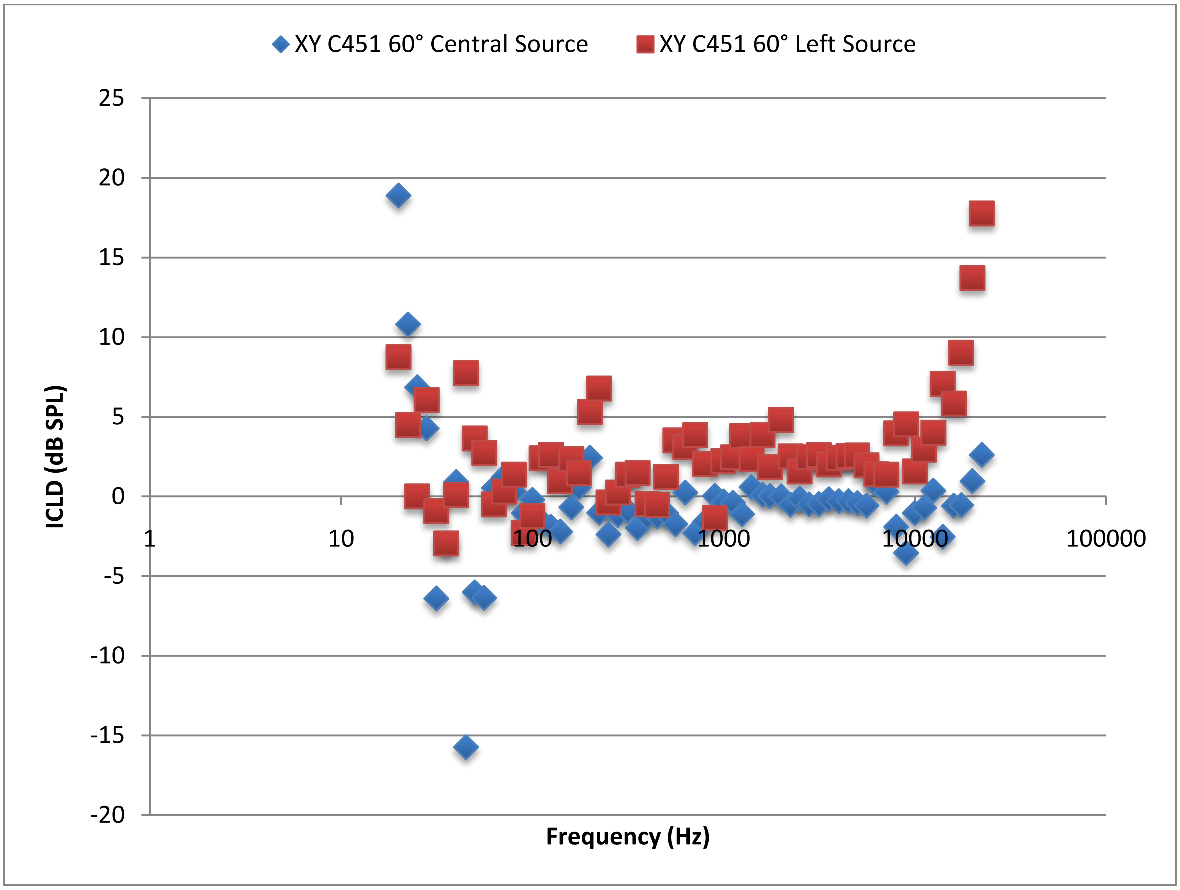
Streicher, R & Dooley,W (1985). Basic Stereo Microphone Perspectives- A Review. Audio Engineering Society . 33 (7/8), 548-556.

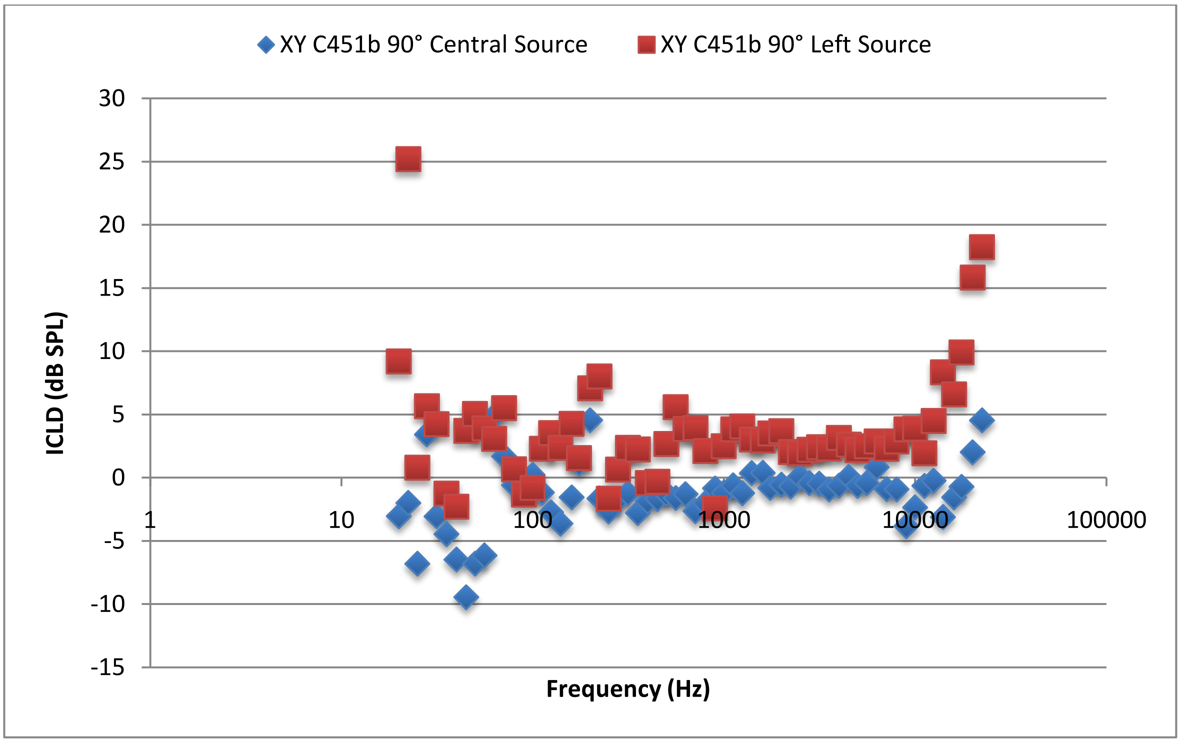
Williams, M. (1987). Unified theory of microphone systems for stereophonic sound recording. In Proceedings of the 82nd Conference of the Audio Engineering Society. London.

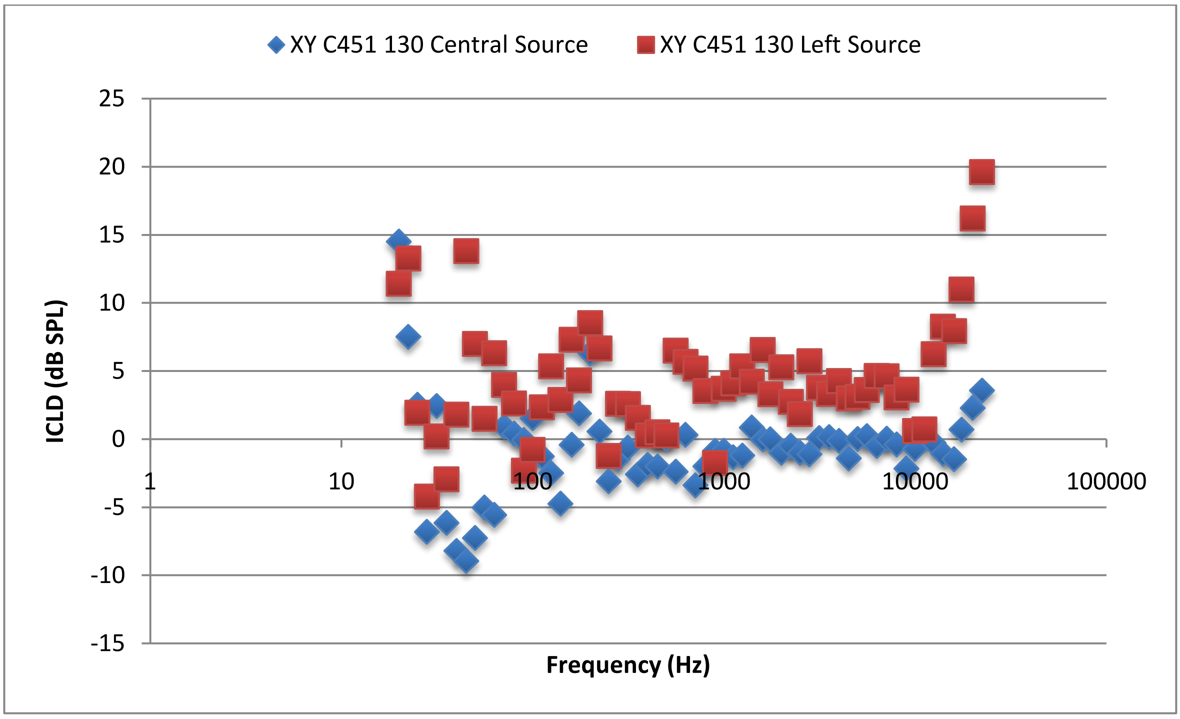
10. Appendices

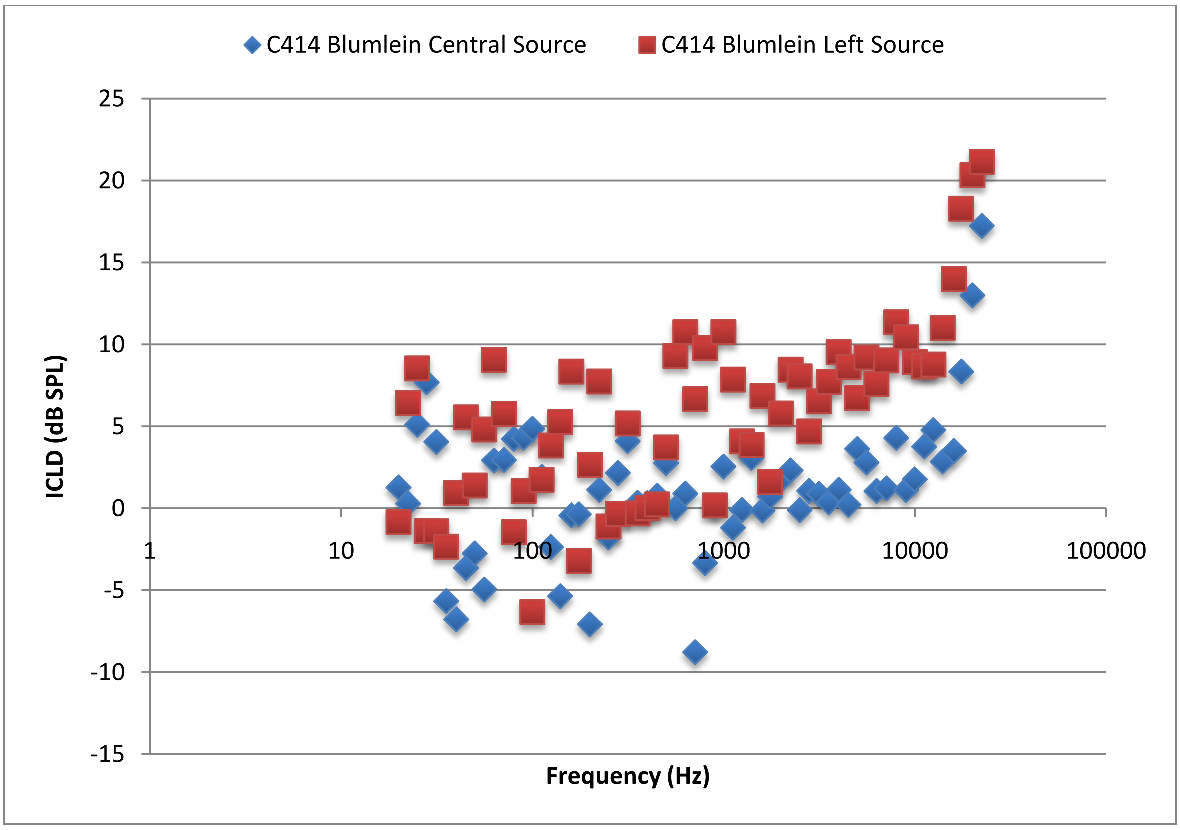
10.1 Appendix 1. Full Results of Measuring Level Differences

Between Channels

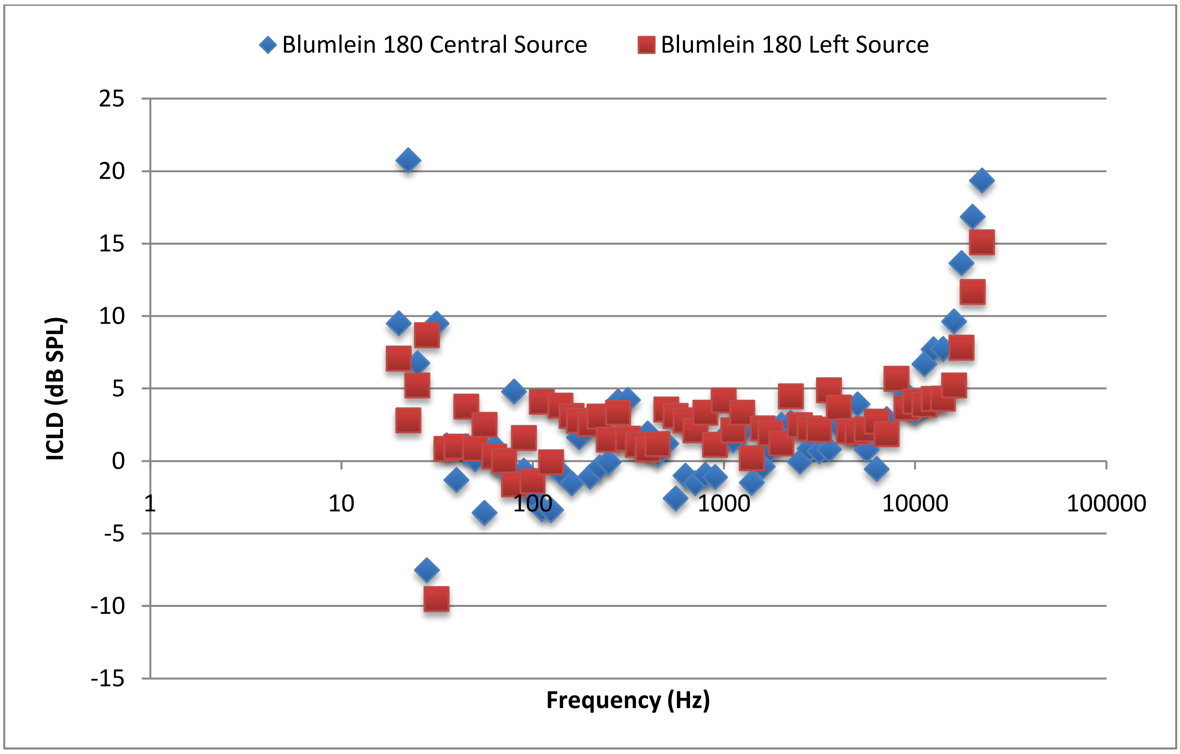



XY C451b 90° Central Source
XY C451b 90° Left Source
ICLD (dB SPL)
30
25
20
15
10
5
0
-5
-10
-15
1
10
100
1000
10000
100000
Frequency (Hz)

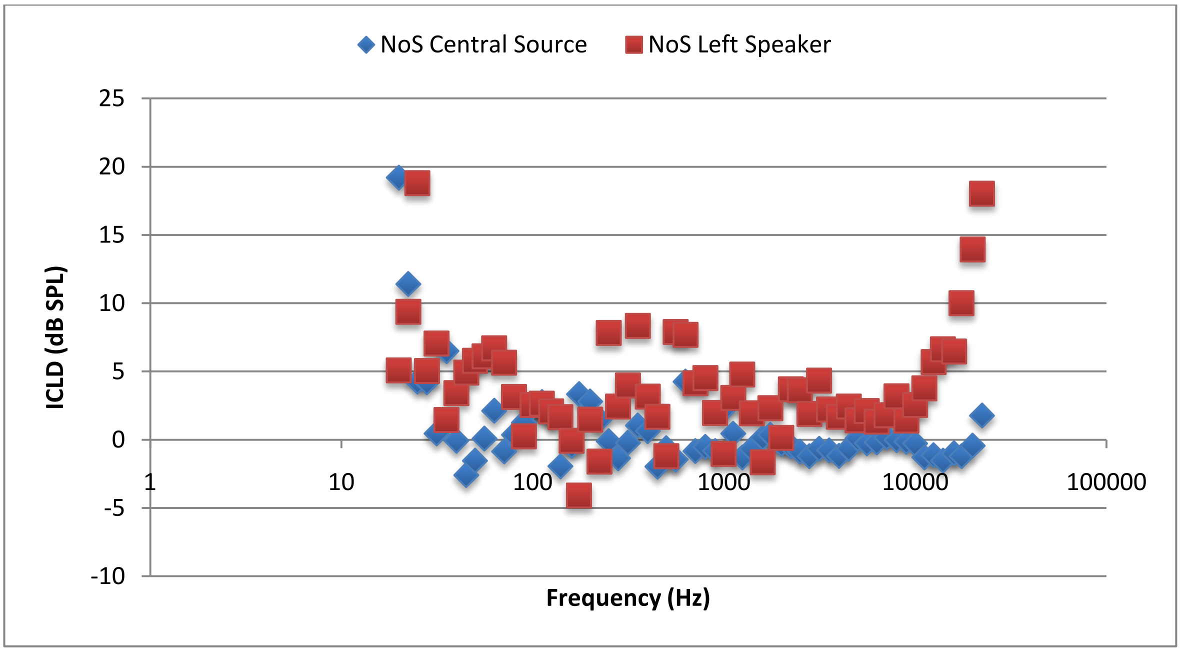

XY C451 130 Central Source
XY C451 130 Left Source
ICLD (dB SPL)
25
20
15
10
5
0
-5
-10
-15
1
10
100
1000
10000
100000
Frequency (Hz)

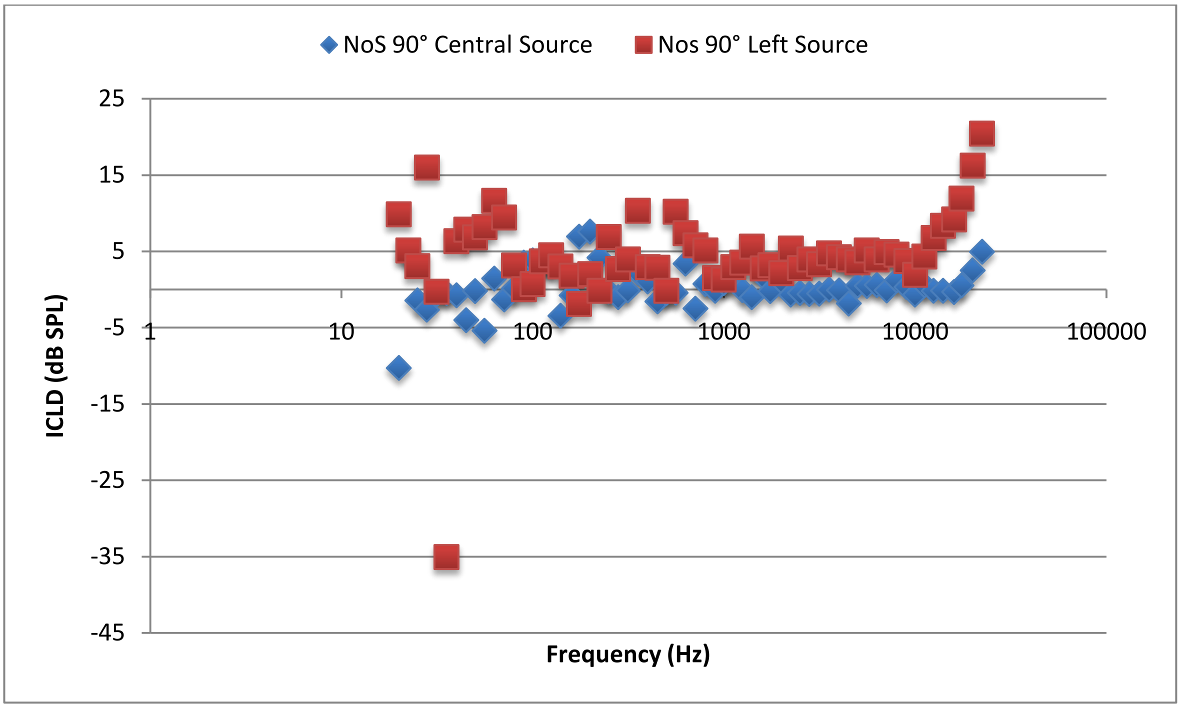


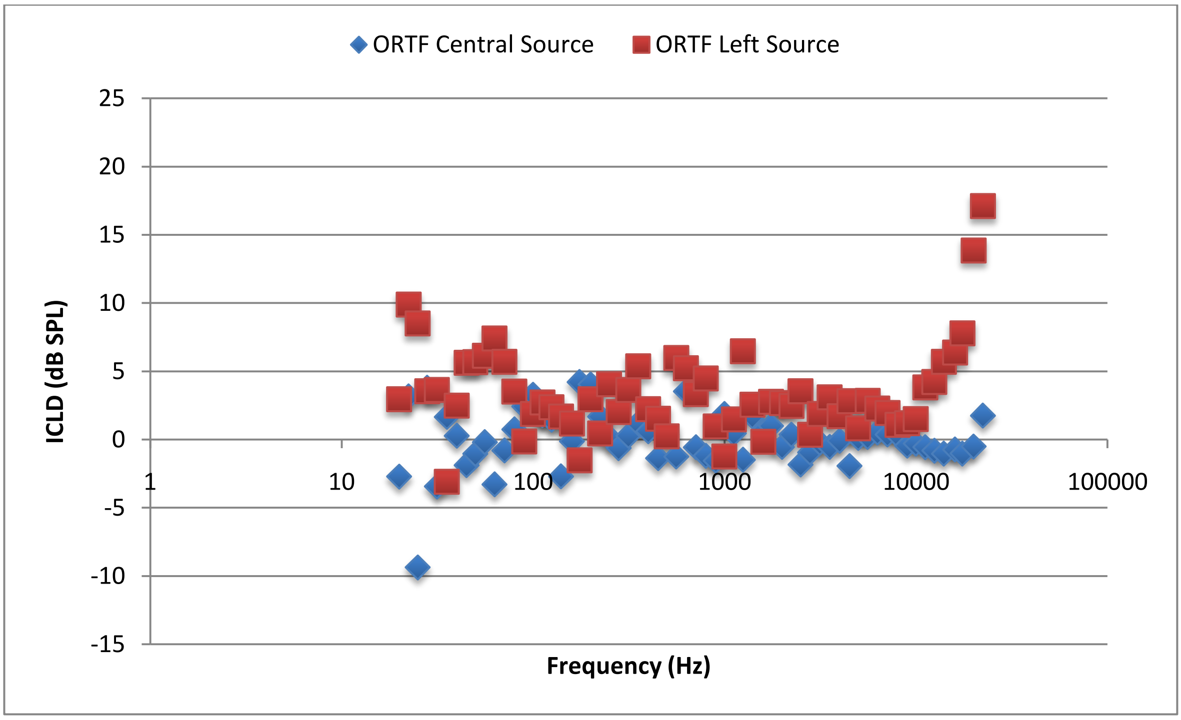
C414 Blumlein Central Source
C414 Blumlein Left Source
ICLD (dB SPL)
25
20
15
10
5
0
-5
-10
-15
1
10
100
1000
10000
100000
Frequency (Hz)



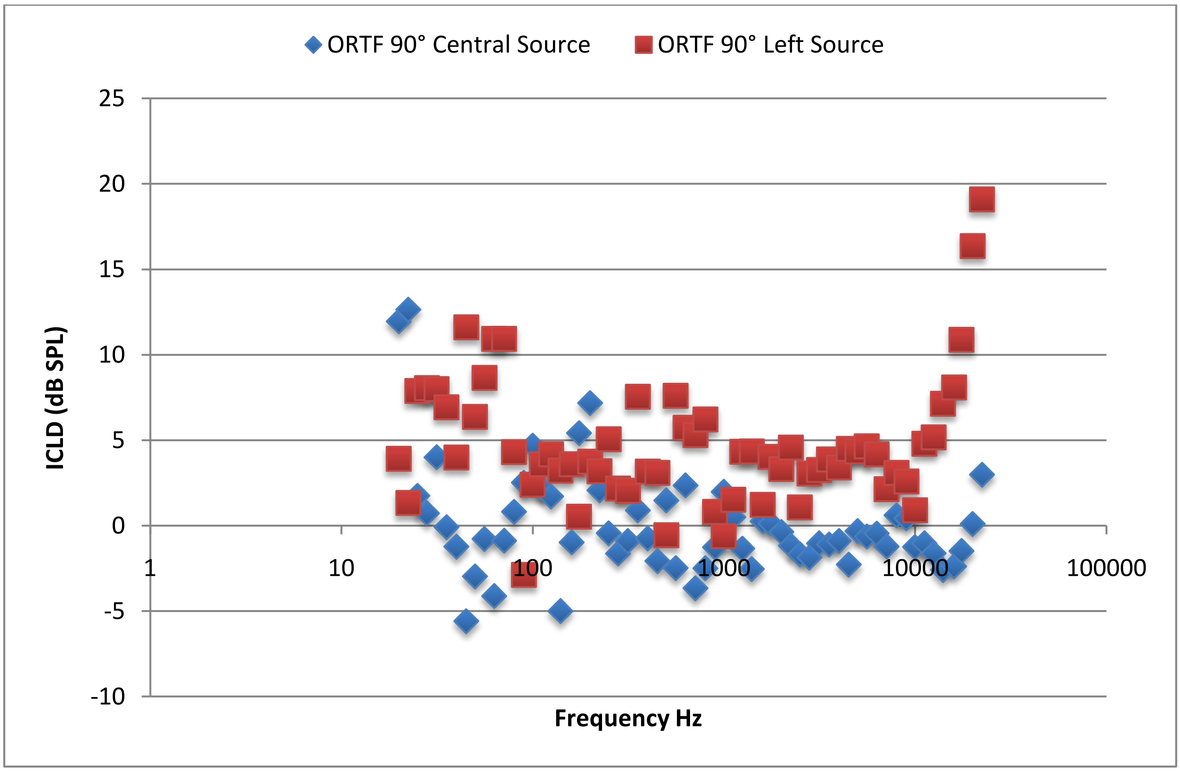
Blumlein 180 Central Source
Blumlein 180 Left Source
ICLD (dB SPL)
25
20
15
10
5
0
-5
-10
-15
1
10
100
1000
10000
100000
Frequency (Hz)


NoS Central Source
NoS Left Speaker
25
20
15
10
5
0
-5
-10
ICLD (dB SPL)
1
10
100
1000
10000
100000
Frequency (Hz)

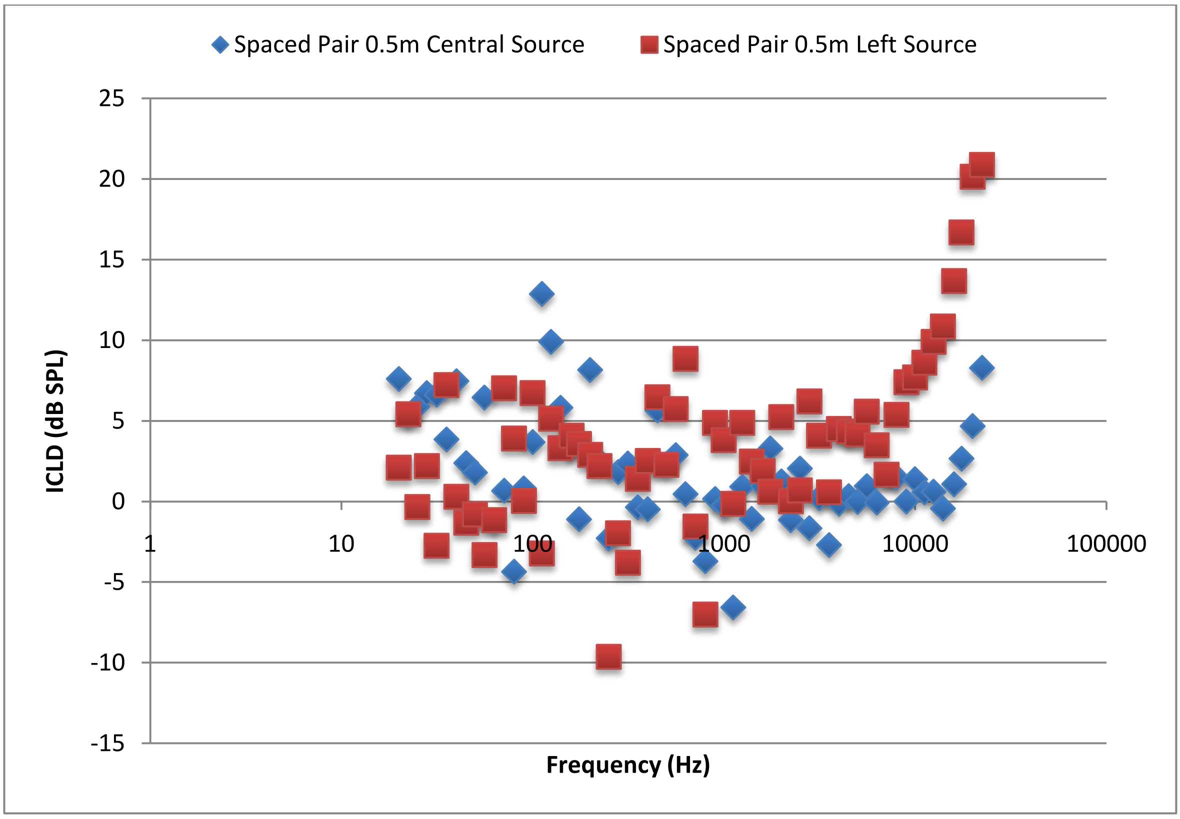

NoS 90° Central Source
Nos 90° Left Source
25
15
5
-5
-15
-25
-35
-45
ICLD (dB SPL)
1
10
100
1000
10000
100000
Frequency (Hz)

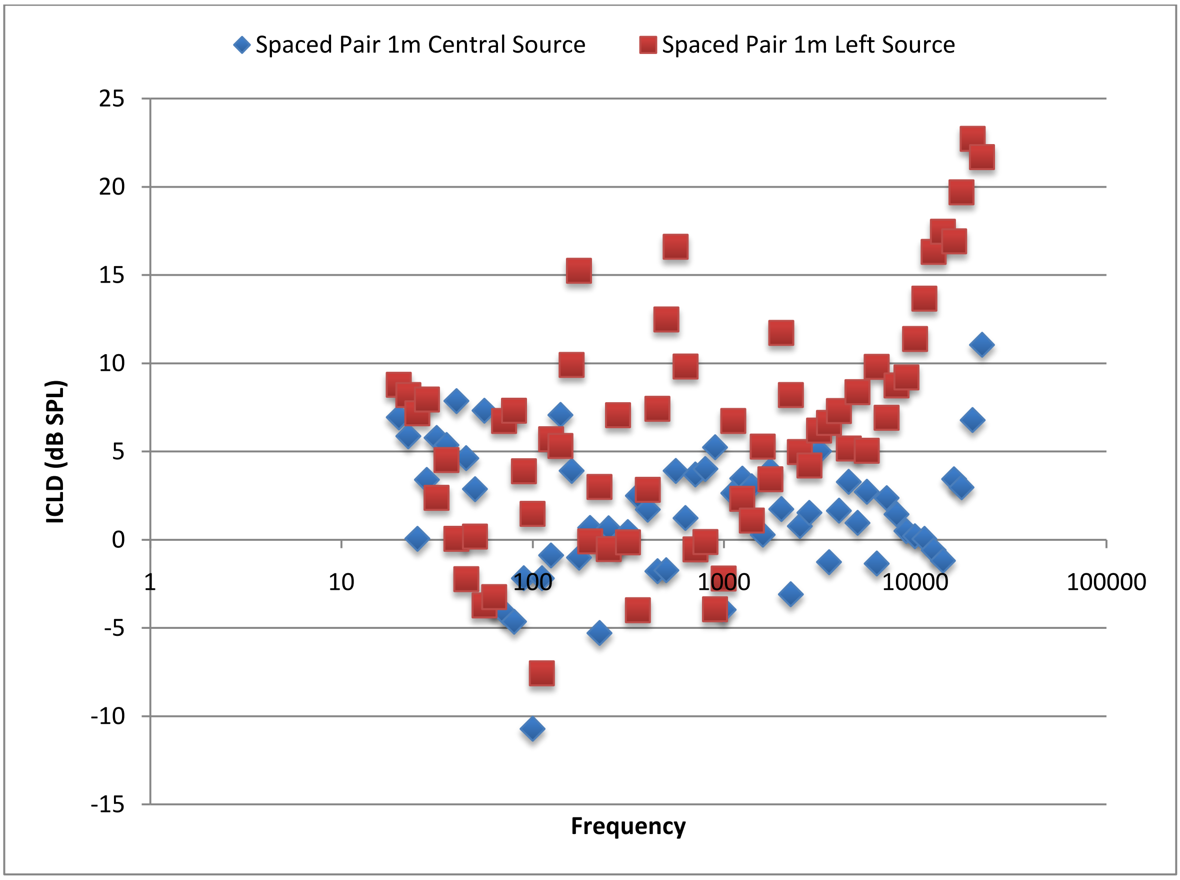


ORTF Central Source
ORTF Left Source
ICLD (dB SPL)
25
20
15
10
5
0
-5
-10
-15
1
10
100
1000
10000
100000
Frequency (Hz)



ORTF 90° Central Source
ORTF 90° Left Source
ICLD (dB SPL)
25
20
15
10
5
0
-5
-10
1
10
100
1000
10000
100000
Frequency Hz


Spaced Pair 0.5m Central Source
Spaced Pair 0.5m Left Source
ICLD (dB SPL)
25
20
15
10
5
0
-5
-10
-15
1
10
100
1000
10000
100000
Frequency (Hz)


Spaced Pair 1m Central Source
Spaced Pair 1m Left Source
ICLD (dB SPL)
25
20
15
10
5
0
-5
-10
-15
1
10
100
1000
10000
100000
Frequency

10.2 Appendix 2 Responses to listening test questionnaires

Listening Test for Stereo Microphone Technique Recordings

Response No: 1

My name is Tom Rawlinson and I am currently a student at Staffordshire University studying for a degree in Music Technology BSc. As part of this award I am investigating the different methods of creating a stereo recording, specifically recording a small classical ensemble in a small studio.

This listening test is designed to allow me to asses the results of these recordings. Your participation in this study is completely voluntary and you have the right to withdraw at any time. Please note that any data collected up to the point of withdrawal may be used within the study. No personal information about you will be stored and any information collected from this project will remain confidential and will be destroyed after the project has been assessed and the marks confirmed.

Please listen to the 5 samples using the headphones provided and then answer the following questions:

In which of the samples did you feel most able to identify and locate the 4 separate performers?

Sample 1 [ ] Sample 2 [ ] Sample 3 [ ] Sample 4 [ ] Sample 5 [✓]

Which sample do you feel presents the widest stereo spread?

Sample 1 [ ] Sample 2 [✓] Sample 3 [ ] Sample 4 [ ] Sample 5 [ ]

Are you able to identify any negative aspects to any of the samples? (i.e. phase cancellation, recording noises etc)

Sample 1 [✓] Sample 2 [ ] Sample 3 [ ] Sample 4 [ ] Sample 5 [ ]

sounds muffled and too close to the microphone, making the stereo field small.

Which of the samples do you find most provides the most pleasing listening experience? Please briefly describe why.

Sample 1 [ ] Sample 2 [ ] Sample 3 [✓] Sample 4 [ ] Sample 5 [ ]

sounds clearest.

18/04/2011

Listening Test for Stereo Microphone Technique Recordings

Response No: 2

My name is Tom Rawlinson and I am currently a student at Staffordshire University studying for a degree in Music Technology BSc. As part of this award I am investigating the different methods of creating a stereo recording, specifically recording a small classical ensemble in a small studio.

This listening test is designed to allow me to asses the results of these recordings. Your participation in this study is completely voluntary and you have the right to withdraw at any time. Please note that any data collected up to the point of withdrawal may be used within the study. No personal information about you will be stored and any information collected from this project will remain confidential and will be destroyed after the project has been assessed and the marks confirmed.

Please listen to the 5 samples using the headphones provided and then answer the following questions:

In which of the samples did you feel most able to identify and locate the 4 separate performers?

Sample 1 [ ] Sample 2 [✓] Sample 3 [ ] Sample 4 [ ] Sample 5 [ ]

Which sample do you feel presents the widest stereo spread?

Sample 1 [ ] Sample 2 [ ] Sample 3 [✓] Sample 4 [ ] Sample 5 [ ]

Are you able to identify any negative aspects to any of the samples? (i.e. phase cancellation, recording noises etc)

Sample 1 [ ] Sample 2 [ ] Sample 3 [ ] Sample 4 [ ] Sample 5 [ ]

no

Which of the samples do you find most provides the most pleasing listening experience? Please briefly describe why.

Sample 1 [ ] Sample 2 [ ] Sample 3 [✓] Sample 4 [ ] Sample 5 [ ]

like the stereo in it.

18/04/2011

Listening Test for Stereo Microphone Technique Recordings

Response No: 3

My name is Tom Rawlinson and I am currently a student at Staffordshire University studying for a degree in Music Technology BSc. As part of this award I am investigating the different methods of creating a stereo recording, specifically recording a small classical ensemble in a small studio.

This listening test is designed to allow me to asses the results of these recordings. Your participation in this study is completely voluntary and you have the right to withdraw at any time. Please note that any data collected up to the point of withdrawal may be used within the study. No personal information about you will be stored and any information collected from this project will remain confidential and will be destroyed after the project has been assessed and the marks confirmed.

Please listen to the 5 samples using the headphones provided and then answer the following questions:

In which of the samples did you feel most able to identify and locate the 4 separate performers?

Sample 1 [ ] Sample 2 [ ] Sample 3 [✓] Sample 4 [ ] Sample 5 [ ]

Which sample do you feel presents the widest stereo spread?

Sample 1 [ ] Sample 2 [ ] Sample 3 [X] Sample 4 [ ] Sample 5 [ ]

Are you able to identify any negative aspects to any of the samples? (i.e. phase cancellation, recording noises etc)

Sample 1 [ ] Sample 2 [ ] Sample 3 [ ] Sample 4 [ ] Sample 5 [X]

Felt like the sound was coming from behind

Which of the samples do you find most provides the most pleasing listening experience? Please briefly describe why.

Sample 1 [ ] Sample 2 [ ] Sample 3 [ ] Sample 4 [✓] Sample 5 [ ]

~~Sounded~~, Generally easier to listen to than the others

18/04/2011

Response Started:
Monday, April 18, 2011 8:08:20 AM

Response Modified:
Monday, April 18, 2011 8:23:41 AM

1. In which of the samples did you feel most able to identify and locate the 4 separate performers?

Sample 3

2. Which sample do you feel presents the widest stereo spread? (i.e. where the sound stretches fully one side of the headphones to the other rather than the sound being located in the middle)

Sample 3

3. Are you able to identify any negative aspects to any of the samples? (i.e. phase cancellation, extraneous recording noises etc) Please select all that apply and detail below

No Response

4. Which of the samples do you find most provides the most pleasing listening experience?

Sample 3

It seems more as though the musicians are in an arc rather than just a straight line. Feels more like actually being in a concert.

Response Started:
Monday, April 18, 2011 8:25:25 AM

Response Modified:
Monday, April 18, 2011 8:27:40 AM

1. In which of the samples did you feel most able to identify and locate the 4 separate performers?

Sample 4

2. Which sample do you feel presents the widest stereo spread? (i.e. where the sound stretches fully one side of the headphones to the other rather than the sound being located in the middle)

Sample 3

3. Are you able to identify any negative aspects to any of the samples? (i.e. phase cancellation, extraneous recording noises etc) Please select all that apply and detail below

No Response

4. Which of the samples do you find most provides the most pleasing listening experience?

Sample 4

Response Started: Monday, April 18, 2011 8:38:44 AM
Response Modified: Monday, April 18, 2011 8:41:52 AM

1. In which of the samples did you feel most able to identify and locate the 4 separate performers?

Sample 2

2. Which sample do you feel presents the widest stereo spread? (i.e. where the sound stretches fully one side of the headphones to the other rather than the sound being located in the middle)

Sample 2

3. Are you able to identify any negative aspects to any of the samples? (i.e. phase cancellation, extraneous recording noises etc) Please select all that apply and detail below

No Response

4. Which of the samples do you find most provides the most pleasing listening experience?

Sample 3

The spread seems most natural when listening through earphones how it would sound in a real life experience

Response Started: Monday, April 18, 2011 8:38:44 AM
Response Modified: Monday, April 18, 2011 8:42:20 AM

1. In which of the samples did you feel most able to identify and locate the 4 separate performers?

Sample 2

2. Which sample do you feel presents the widest stereo spread? (i.e. where the sound stretches fully one side of the headphones to the other rather than the sound being located in the middle)

Sample 2

3. Are you able to identify any negative aspects to any of the samples? (i.e. phase cancellation, extraneous recording noises etc) Please select all that apply and detail below

No Response

4. Which of the samples do you find most provides the most pleasing listening experience?

Sample 3

The spread seems most natural when listening through earphones how it would sound in a real life experience

Response Started:
Monday, April 18, 2011 10:18:11 AM

Response Modified:
Monday, April 18, 2011 10:55:14 AM

1. In which of the samples did you feel most able to identify and locate the 4 separate performers?

Sample 2

2. Which sample do you feel presents the widest stereo spread? (i.e. where the sound stretches fully one side of the headphones to the other rather than the sound being located in the middle)

Sample 1

3. Are you able to identify any negative aspects to any of the samples? (i.e. phase cancellation, extraneous recording noises etc) Please select all that apply and detail below

Sample 4

Sample 5

Less stereo definition

4. Which of the samples do you find most provides the most pleasing listening experience?

Sample 2

Depth of stereo sound was far clearer.

Response Started:
Monday, April 18, 2011 12:35:08 PM

Response Modified:
Monday, April 18, 2011 12:50:20 PM

1. In which of the samples did you feel most able to identify and locate the 4 separate performers?

Sample 3

2. Which sample do you feel presents the widest stereo spread? (i.e. where the sound stretches fully one side of the headphones to the other rather than the sound being located in the middle)

Sample 3

3. Are you able to identify any negative aspects to any of the samples? (i.e. phase cancellation, extraneous recording noises etc) Please select all that apply and detail below

Sample 2

More (unwanted) background/performance noise, not enhancing the 'feel' of the piece with Spaced Pair.

4. Which of the samples do you find most provides the most pleasing listening experience?

Sample 5

It was between Mid Side and Blumlein, both offer a more full bodied stereo sound with less unwanted noise.