LOCALISATION OF SOUND SOURCES USING COINCIDENT MICROPHONE TECHNIQUES

B. Fazenda Music Technology, School of Computer and Engineering, University of Huddersfield

1 INTRODUCTION

Sound source localisation is an important aspect of monitoring and control engineering with various applications such as environmental monitoring, noise control, medical instrumentation and surveillance to name a few. In these applications, accurate localisation of source direction and distance has been sought, and techniques are varied and widespread[1 2 3 4 5].

On the other hand, in the professional audio context, the application of recording techniques based on microphone configurations has attempted to achieve the same accurate capture of source localisation in a complex field, and subsequently encode this into a realistic reproduction environment. From a range of techniques, coincident microphone configurations, employing two directional microphones at a single point in space, are well known for their performance. A concise coverage of literature regarding this is given in [6].

The work presented in this paper is motivated by an initial interest in source localisation techniques, mainly for engineering applications, that attempt to detect source location in complex environments. Extensive theory and practical application in this field has been developed. However, strangely enough, the vast knowledge generated from music recording applications appears to never have been explored in the more engineering based detection applications.

Moreover, the amount of published material existent on the accuracy of coincident microphones to encode the incoming direction of a given sound source appears scarce, particularly in practical applications, such as in built environments, and using common signals as the source.

This work attempts to initiate the transfer of knowledge between the two areas by studying the suitability of spatial recording techniques for source localisation. Empirical results from simple measurements and source angle extraction techniques are provided in Section 2.

In particular, the performance of two microphone configurations is studied in a series of measurements carried out to compare the Soundfield microphone (www.soundfield.com) and a common Mid-Side configuration. The accuracy in encoding the known direction of a sound source in an enclosed environment is investigated.

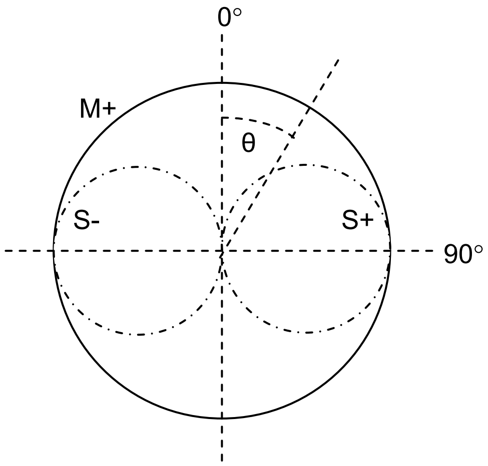
Finally, in Section 4, the performance of each technique is compared using listening tests designed to identify the ability of a subject in detecting the direction of a recorded sound source using common stereo reproduction technqiues. Two loudspeakers configurations are tested using 2 and 3 speaker decoding techniques.

2 THEORY

2.1 Mid-Side Configuration

A coincident microphone technique employs at least two directional microphones placed at a single point in space (as much as it is physically possible). Given their spatially coincident placement, the signals arriving at each microphones are considered to be in phase at least up to frequencies where the wavelength becomes comparable to the spacing between their diaphragms. The directional sensitivity of each microphone encodes a level difference between the two channels which can be read as directional information. For stereophonic reproduction, the technique relies on the well known auditory illusion where listeners localise the incoming direction of a sound source towards the direction of the louder sound. The ratio between the two channels – the Interaural Level Difference - controls the lateral displacement of the auditory illusion [7].

A simple configuration is to use a pressure transducer (omnidirectional polar pattern) placed coincidentally with a pure velocity transducer (bidirectional polar pattern) and extract the left and right channels of a stereo signal using a combination of the two microphone outputs (*Figure 1*). This technique is commonly known as a Mid-Side (MS) configuration.



L = 0.7071*(M-S)

R = 0.7071*(M+S)

Figure 1 - Mid-Side configuration using an omnidirectional microphone (M) and a bi-directional microphone (S)

Per definition, the amplitude of the Mid signal should remain constant for all angles, exhibiting an omni-directional pattern, whereas the amplitude for the Side signal should follow a *sin(θ)* function with origin at 0° and reaching a maximum at 90°, exhibiting a figure of 8 pattern side ways.

The angle of incidence, θ, for a given source signal may be extracted from the magnitude of the output of each microphone:

$$\theta = \sin^{-1}\left(S / M \right) \tag{1}$$

2.2 The Soundfield Microphone

The Soundfield microphone (SF) can be thought of as a 3 dimensional microphone, using a combination of 3 pressure sensitive microphones covering each Cartesian direction (x,y,z) and an additional pressure microphone with omni-directional polar pattern. This behaviour is achieved using 4 microphone diaphragms arranged in a tetrahedral configuration and placed as close as possible to each other to reduce any phase differences. According to the manufacturer, frequency dependent digital signal processing is further applied to compensate for the small distance that actually exists between the capsules (~12mm).

The microphone has 4 output signals, known as the B-Format, representing the velocity component in the three Cartesian directions X (front-back), Y (left-right) and Z(above-below) and one omni-directional signal, W, representing the pressure component. In this application, only the horizontal components of the B-Format have been used to extract source incidence angle in the horizontal plane.

Similarly to the Mid-Side configuration, the X signal, a figure of 8 polar pattern in the front-back direction, and the Y signal, a figure of 8 sideways, may be used to extract source angle (***Figure 2***). The W signal represents the measured pressure, thus not sensitive to direction and useful for magnitude normalization of the other three signals.

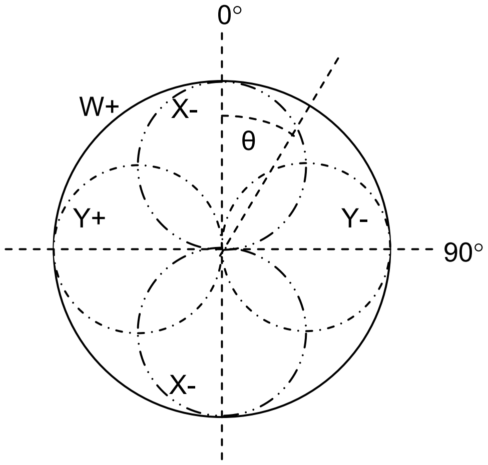


Figure 2 - Soundfield microphone configuration with X signal representing the front back velocity, Y representing left-right velocity and W representing the pressure component. The Z B-Format output representing the up-down velocity has been omitted.

In this case, for the axes defined in ***Figure 2***, the X signal may be represented by a *cos(θ)* function, whilst the Y signal may be represented by its orthogonal, *sin(θ)*. The magnitude of X and Y signals should match at 45°, and this may be used to calibrate microphone gains using a known signal source placed at this angle. The magnitude of the W signal may be used to normalize the amplitudes to within the –1 to 1 range.

Clearly, using configuration of ***Figure 2***, the angle of source incidence may be obtained from:

$$\theta = \tan^{-1}\left(Y/X\right) \qquad (2)$$

The technique described here may be developed to include estimation of source direction in true 3 dimensional coordinates, by introducing an elevation angle α dependent on the Z output of the B-Format.

Left and *Right* output signals for a common stereo reproduction system may be generated using a similar approach to that described in 2.1 for the Mid-Side technique, using the Y and W signals. Different weightings of X and Y may be used to emulate coincident cardioid techniques at various aperture angles.

3 MEASUREMENTS AND RECORDINGS

Measurements were taken in two different spaces – a small concert hall and a studio recording space. Genelec 8040A speakers were used as sound sources placed at discrete angles and at a radial distance of approximately 2 metres of the microphones.

The microphones used were an AMS Soundfield SPS422 with dedicated pre-amp and two AKG C414B P48 XLS with selectable polar patterns. The signals were routed through a Yamaha O2R digital mixing desk into a MOTU828mkII digital soundcard. The sound files were recorded on a PC using Steinberg Nuendo sequencing software.

The sources in the hall reproduced bursts of white noise. A real glockenspiel was also recorded at 3 different angles. In the studio space the signals used were Maximum Length Sequence signals, which exhibit the characteristics of noise whilst being deterministic, and a sample recording of a saxophone.

The signals at the microphones were calibrated, as without this the correct angle cannot be properly defined. The Mid-Side (MS) configuration was calibrated so that the amplitude in each microphone was matched when the active source was at 90°. This represents the maximum output of each microphone. For one measurement, using a Blumlein pair (2 crossed bi-directional microphones), this calibration was done at 45°.

The Soundfield microphone signals were also calibrated so that a source at 45° generated the same amplitude in both X and Y signals. On axis was pointing towards the source at 0°.

The following sections show the results obtained from measurements taken in the 2 spaces using the techniques described above.

3.1 Studio

Measurements were taken in the recording space of a studio. The room has a fairly long reverberation time of around 0.7s at mid frequencies. No particular measures were used to attempt to control any reflections, as the aim was to test the technique in general conditions.

A set of MS and SF microphone were arranged at the origin of a Cartesian set of axes. 7 loudspeakers were positioned at 15º angles covering the range from 0º to 90º. This is illustrated in ***Figure 3***.

Figure 3 – Arrangement of 7 loudspeakers at a distance of 2 metres from microphones and ranging from 0º to 90º at 15º angles. Tweeter height is 1.3 metres.

Figure 4 – Soundfield and Mid-Side arrangement for measurements in recording space. Soundfield microphone diaphragms are at 1.3 metres from the floor.

The high frequency driver of each loudspeaker and the diaphragms of the SF microphone were matched at a height of 1.3 metres from the floor. The MS configuration was placed at the same vertical point (i.e. in the origin of the Cartesian space) and as close as possible to the SF microphone, as illustrated in ***Figure 4***.

Each loudspeaker generated a burst of MLS noise and a recorded musical phrase played on a saxophone. The results shown here are extracted using equations 1 and 2 from sections 2.1 and 2.2, using the calculated RMS level for each microphone output.

Figure 5 shows the results extracted from recordings done with the SF and with the MS microphone configured so that both microphones are set to bidirectional polar patterns, effectively creating a Blumlein pair. The Mid microphone is pointing on axis at 0º and the *side* microphone pointing to 90º. This represents an identical situation to that obtained with the X and Y signals of the SF microphone.

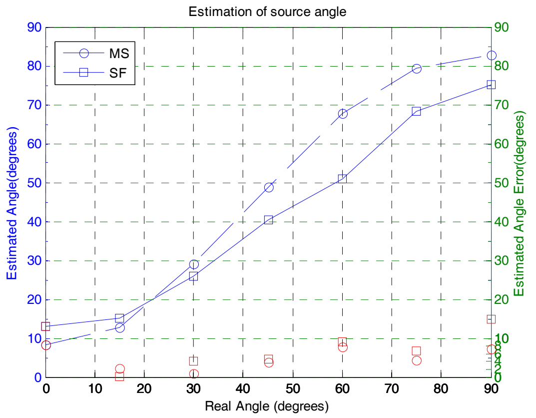


Figure 5 – Estimation of source angle from microphone signals. Comparison between a front-facing Blumlein pair (2x fig.of 8) and Y/X B-Format outputs from Soundfield microphone (SF). Error for each microphone set is indicated in red. Measurements taken in a studio recording space, using white noise.

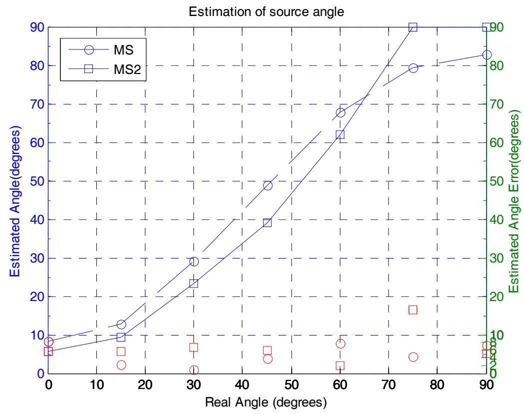


Figure 6 - Estimation of source angle from microphone signals. Comparison between Blumlein pair and conventional mid-side (omni-fig. of 8) configuration (MS2). Error for each microphone set is indicated in red. Measurements taken in a studio recording space, using white noise.

Both microphone sets perform reasonably under these conditions, with maximum errors occurring at the extreme angles of 0º and 90 º. The SF appears to underestimate most angles except 0 º and 15 º. Both microphone sets exhibit large errors at 0 º (indicated in red), with estimated angles offset by at least 8 º. This inaccuracy in measuring very small angles is not unheard of even for techniques which are based on time of arrival estimation using omni-directional microphone arrays.

Figure 6 shows a comparison between estimates for the Blumlein pair and the conventional MS configuration. There is no evidence to suggest an improvement of performance with any technique except for 75º where the conventional MS configuration shows a large error. It may also appear that this latter technique can extract the incident angle of a sound source at 90º quite exactly. However, the reader is reminded that this is the angle used for calibration between the two microphone outputs, and as such, a nearly exact result should be expected.

The same degree of estimation accuracy can be obtained if the signal used is musical rather than noise. Figure 7 illustrates results estimated from musical source signals.

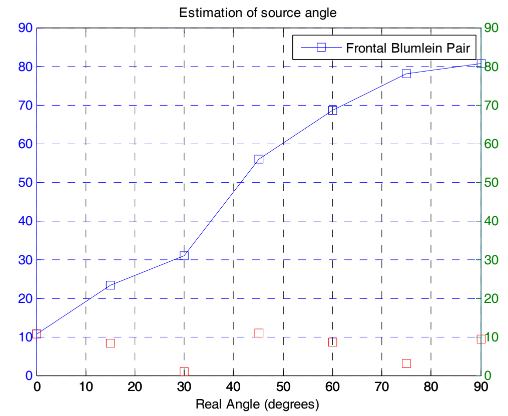


Figure 7 – Estimation of source angle using a Blumlein pair. Signal source is the recording of a musical phrase played on a saxophone.

Although not identical, estimated results are reasonably accurate, with errors capped at around 10 º.

3.2 Concert Hall

Measurements taken in a small concert hall are described below.

Sources were places at 0º, 15º, 30º, 45º and 90º angles. Similarly, a set of MS and SF microphones were arranged at the origin of an X,Y set of axes. The same equipment as in the studio room was used, except for data acquisition, where an Edirol R4 solid state digital recorder was used to record the B-Format. A PC laptop and an M-Audio FW410 soundcard were used to record the MS signals.

Figure 8 shows the results obtained.

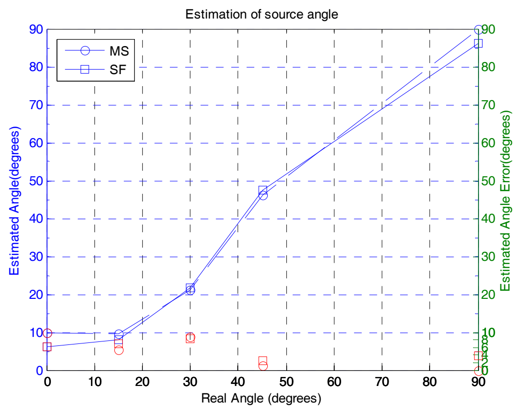


Figure 8 – Estimation of source angle from microphone signals. Comparison between mid-side (omni-fig. of 8) pair (MS) and Y/X B-Format outputs from Soundfield microphone (SF). Error for each microphone set is indicated in red. Measurements taken in a concert hall, using white noise.

In this case, it is apparent that the smaller angles are associated with larger errors (about 10º). There is good consistency of results between both microphone sets.

4 LISTENER LOCALISATION

A series of listening tests were conducted with the aim of identifying if any of the recording techniques used would offer better subjective localisation in stereo reproduction systems. Furthermore, the common 2 speaker stereo reproduction technique was contrasted with a 3 speaker stereo configuration, reportedly more robust in generating the phantom sources within the speaker field[8].

The aim of an accurate stereo recording technique is to create realistic phantom images where their position in the stereo sound-field corresponds to their original position in the recording stage.

Pilot listening tests were conducted in a post-production monitoring room, with speakers placed in an equilateral triangle with the listener sat at one edge. The distance between each speaker and the listener was 1.4 metres. The listener was positioned such that ear height matched tweeter height (1.4 m). A third central speaker was placed between the left and right stereo speakers and at the same radial distance to the listener.

Recorded sound samples, consisting of noise, glockenspiel and saxophone, were normalised to the same loudness level and played in a random order to 8 volunteer listeners. Only 1 audition of each sample was allowed.

For each sample the listener was asked to indicate the apparent position for the phantom image within the speakers. A diagram showing numbered locations for positions in 5º angles was provided. This included the left, right and centre speakers (positions 1,7 and 13 respectively) which corresponded to the recorded -90º, 0º and 90º respectively.

Since the recording stage spanned a 90º aperture and the stereo reproduction stage spans a 30º aperture, it is expected that each 15º step in the recording stage corresponds to a 5º step in the reproduction stage.

4.1 2 AND 3 SPEAKER DECODING

The recorded signals were decoded for presentation via 2 or 3 speaker configurations. The general 2 speaker configuration was obtained by generating left and right signals from the common Mid and Side and X,W B-Format signals as described in 2.1.

To decode the B-Format into 3 speaker signals, a set of signals in the M,S and T format has been generated. The MST format maintains the same properties as an MS format, and these are described in detail in reference 8. The following matrix is used to generate the MST format from X,Y and W:

$$\begin{aligned} M &= 0.7071(W+X) \\ S &= Y \\ T &= 0.7071(W-X) \end{aligned} \tag{3}$$

This can then be converted into Left, Centre and Right signals by calculating:

$$\begin{aligned} L &= 0.5*M+0.7071*S+0.5*T \\ C &= 0.7071*M-0.7071*T \\ R &= 0.5*M-0.7071*S+0.5*T \end{aligned} \tag{4}$$

The generation of L_3,C_3 and R_3 from the two L_2, R_2 output signals of a common MS configuration is obtained from:

$$\begin{aligned} L_3 &= 0.5*((\sin(\varphi)+1)*L_2 + (\sin(\varphi)-1)*R_2) \\ C_3 &= 0.7071*\cos(\varphi)*(L_2+R_2) \\ R_3 &= 0.5*((\sin(\varphi)-1)*L_2 + (\sin(\varphi)+1)*R_2) \end{aligned} \tag{5}$$

Where $\varphi = 45*pi/180$.

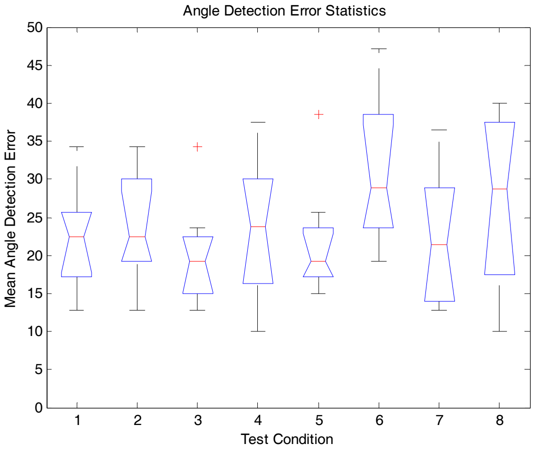
4.2 LISTENING TEST RESULTS

Listening test results are shown in Appendix 1 and present the group average for each angle, and each testing condition - Microphone, type of sound source and speaker setup.

The black dotted line in each chart represents the expected position of the phantom image. It appears that listeners are able to detect the general direction of a phantom image but this ability is somewhat erratic. In most cases, except for *instrument sound source* in concert hall, an increase in real source angle is associated with an increase in phantom image angular position. However, there is never exact match between real source angle and phantom image for any of the cases tested. None of the factors studied appears to induce any differences in the results.

4.3 STATISTICAL ANALYSIS OF RESULTS

The value of the *mean angle detection error* for each listener was calculated from the error at each real angle. This is shown for all the different testing conditions in ***Figure 9***.



1	2 Channel MS with INSTRUMENT
2	3 Channel MS with INSTRUMENT
3	2 Channel SF with INSTRUMENT
4	3 Channel SF with INSTRUMENT
5	2 Channel MS with NOISE
6	3 Channel MS with NOISE
7	2 Channel SF with NOISE
8	3 Channel SF with NOISE

Figure 9 – Mean angle detection error for each test condition. Mean is determined from the apparent angle error for each subject at each angle. Test conditions correspond to Source type (noise/instrument); Microphone configuration (MS/SF); number of channels decoded (2/3).

It appears that test conditions 3 and 5 (SF and MS decoded into 2 channels) exhibit the lowest angle detection error. However, the analysis of variance presented in Table 1, shows that there are no significant differences between any of the means obtained. This further confirms the results shown in appendix 1, and suggests that for this group of listeners and under these test conditions the translation of apparent source angle from recording to listening condition has been somewhat inaccurate.

Table 1 – Analysis of Variance of Mean Error Detection Angle for each test condition

ANOVA Table

Source	SS	df	MS	F	Prob>F
Columns	698.82	7	99.8315	1.29	0.2743
Error	4348.5	56	77.6518		
Total	5047.32	63			

5 DISCUSSION AND CONCLUSIONS

The results for the work presented may be distinguished into two categories:

5.1 Source Angle Detection

The results obtained from microphone signal magnitude show that Mid-Side and Soundfield microphone configurations can be useful for the detection of the angle for a sound source in a complex environment as long as signal gain is calibrated with known sources. Under the conditions tested, estimation accuracy is associated with a 10º or less error which would make it unsuitable for the most demanding applications, where TOA techniques are capable of achieving accuracy levels of around 2º [2]. However, for general applications the systems tested are shown to perform acceptably.

There doesn't appear to be a significant difference between the performance of a Soundfield microphone and a generic Mid-Side configuration using high quality studio microphones. Likewise, both Mid-Side and Blumlein configurations appear to provide similar estimates for source angle.

Using the principles described here, the source angle may be estimated equally well from wide frequency range test tones, such as noise, as well as from general music signals. This is encouraging for more general applications where the source characteristics cannot be controlled.

5.2 Listening Tests for 2 and 3 speaker stereo

The results obtained for listener detection of phantom images show that the systems tested have not provided the right cues to ensure reliable measurement of apparent source angles. Although listeners appear to be able to distinguish the general direction of a phantom image, the exact match between real recorded angle and its reproduced counterpart has not been demonstrated. Some factors may be responsible cause for this discrepancy:

1. The mismatch between real aperture angles (90º) and reproduction angle (30º).
2. A small physical distance between the loudspeakers that spatially compresses the phantom image positions and makes the task more difficult.
3. The existence of recorded room reflections that may interact with the temporal structure of the ILD at the reproduction stage and make the detection task more difficult.
4. The testing technique has unanticipated flaws that prevent listeners from performing accurately.

All these are interesting points to warrant future research avenues.

6 CONCLUSIONS

It has been shown that coincident techniques relying on directional properties of microphones are sufficient to encode signal level ratios that enable the estimation of source angle.

Under the testing conditions, no significant differences have been found between Mid-Side techniques using common studio microphones and the more specialised 3 dimensional Soundfield microphone.

Source angle estimation is possible in complex environments and from real sound sources, such as instruments, as well as from test signals, such as noise.

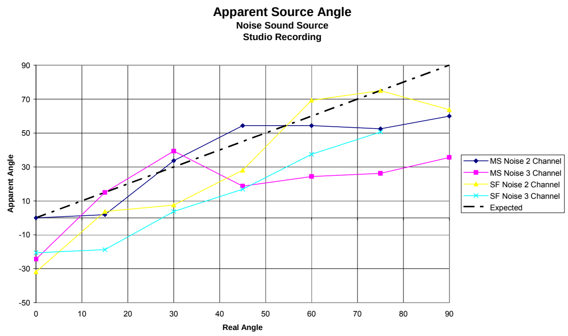
The accuracy of these techniques in transferring a representation of the real stage into a realistic reproduction of it has not been proven. None of the various microphone and speaker techniques tested have led to significantly accurate detection of phantom images in a reproduction environment.

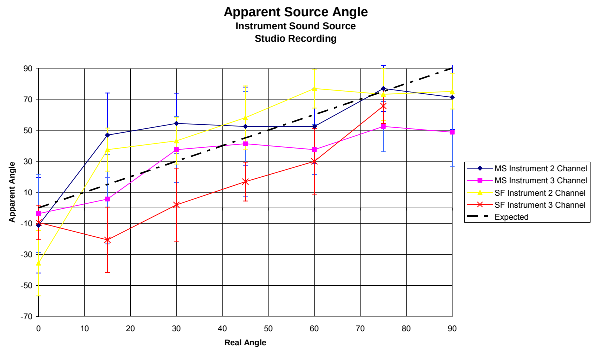
Taking into consideration that this type of investigation is in its very early stages, so far there does not appear to be any significant benefit in decoding signals for 3 speaker reproduction. This is, of course, an interesting subject for future research.

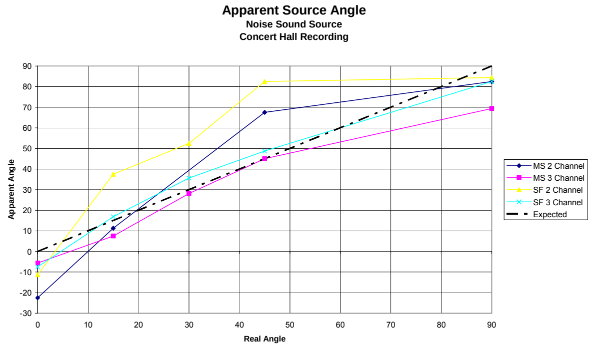
<u>Acknowledgements:</u>

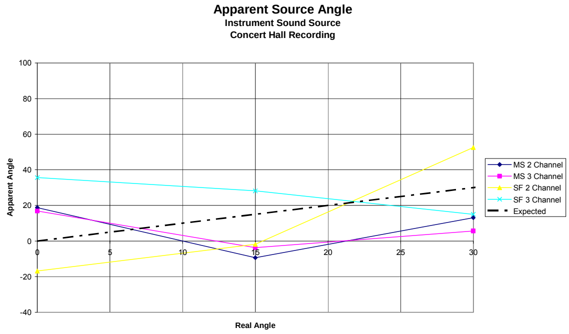
The author would like to thank all those who have participated in the listening experiments and especially the precious help of Mr. Matthew Wankling and Mr. David Moore, researchers at the School of Computing and Engineering, for their help in preparing the tests.

7 APPENDIX 1 – LISTENING TEST RESULTS









8 REFERENCES

[1] J. Chen, L Yip, J. Elson, H. Wang, D. Maniezzo, R.E. Hudson, K. Yao, D. Estrin, "Coherent acoustic array processing and localisation on wireless sensor networks", *IEEE Proceedings,* vol. 91, issue 8, pp. 1154 -1162, Aug 2003.

[2] T. Scott Brandes and Robert H. Benson, Sound source imaging of low-flying airborne targets with an acoustic camera array, *Applied Acoustics*, *Volume 68, Issue 7*, *July 2007*, *Pages 752-765*

[3] Reinhard Blumrich and Jürgen Altmann, Medium-range localisation of aircraft via triangulation *Applied Acoustics*, *Volume 61, Issue 1*, *September 2000*, *Pages 65-82*

[4] YC Choi and YH Kim, "Near field impulsive source localization in a noisy environment", J. Sound and Vibration, 303, 209-220, 2007.

[5] K. Nakadai, H.G. Okuno and H. Kitano, "Real-time sound source localization and separation for robot audition", Kitano Symbiotic System Project, ERATO, Japan Science and Technology Corp., Tokyo, Japan.

[6] Eargle, J., "Stereophonic Techniques – An anthology of reprinted articles on stereophonic techniques", Audio Engineering Society, 1986

[7] Blauert,*Spatial Hearing: The Psychophysics of Human Sound Localization,* MIT Press, Cambridge, MA, 1983

[8] Gerzon, M., "Microphone Techhniques for 3-Channel Stereo", Journal of the Audio Engineering Society, Preprint 3450 (S-3), Presented at the 93rd Convention, San Francisco, October 1992