Ethan Kennedy

Student ID: xxxxxxxxxx

'A subjective comparison between panning algorithms for an ITU 5-speaker array.'

Bsc (Hons) Multimedia Technology and Music Production

University of Derby

Independent Project – 6EJ998

May 2006

Tutor: Bruce Wiggins

Acknowledgements

Thanks must be extended to Bruce Wiggins for his knowledge and support in implementing the Simulink models used in the listening tests, and helping to focus this report from such a broad subject area. Peter Lennox has been a source of support, advice, and wisdom throughout this year, and his enthusiasm inspires me continually. Final thanks go to my family for being there when I need them, and to my computer for not dying on me during the write up period!

Abstract

This project subjectively assesses several recent panning laws. Background research of traditional and newer multi channel audio formats and panning techniques has been undertaken. The understanding of the complex processes governing human spatial perception is becoming more and more important when designing surround sound formats that are subject to the limitations of current hardware and software. The low level processes that aid lateral localisation are explored, and three panning algorithms, specific to the ITU 5 speaker array are subjected to listening tests, and valid conclusions drawn.

Chapter 1 – Introduction

1.1 – Background

Multi channel 'surround sound' as we know it was first developed for use in motion picture audio presentations. As the term suggests, the idea was to create an aural 'environment' surrounding the listener using more than two loudspeakers. The development of storage media such as the Digital Versatile Disc (DVD), and digital television has meant that surround sound is no longer a thing of cinematic wonder, but an economically accessible consumer audio medium. (Holman, T, 2000 pg.11) The computer gaming industry and the music industry play an equally important part in the development of and demand for the closest thing to truly immersive surround sound. However, both industries have had to adapt to certain technological standards that have emerged from the prevalence of surround for motion picture presentation. One such domestic standard is the general adoption of the International Telecommunications Union (ITU)-R BS.775-1 recommended 5 speaker array, featuring loudspeakers positioned at 0˚, ±30˚, and ±110˚ on the horizontal plane around the listener. To this, a Low Frequency Enhancement (LFE) channel can be added to form the 5.1 set up most widely used in mass market surround audio.

The 5 speaker array described was developed to provide the listener with stable frontal imaging between ±30˚ and the surround channels at ±110˚ were intended to provide ambience information as well as special effects. This is almost ideal in a sound-for-picture situation, where accompanying visual information comes only from the front of the viewer, so only ambience and effects that give spaciousness are required of the rear surround speakers. For the music and computer games industry however, multi channel sound using this standard speaker set up is not ideal as it is necessary to create an immersive, psychoacoustically correct listening environment, in which the perceived direction of a source is not discriminatory to the limitations of the available loudspeakers (Wiggins, B, 2004, pg. 1-3). In this investigative study, several unconventional panning algorithms for multi channel audio will be compared analytically, and by subjective testing using the ITU speaker array in a controlled environment. Qualities related to the localisation of sound, including image focus, image stability, and image location will be tested using two recent panning algorithms proposed by Wiggins (Wiggins, B, 2004, 2006), and Peter Craven's 4th order circular harmonics surround decoder (Craven, P, 2003): The first two of which have never been subjectively tested before.

1.2 – Aims and objectives

• To conduct an extensive literature review and discussion of the following aspects of multi channel audio:

- Surround sound formats and history
- Traditional and recent panning laws for multi channel audio
- The ITU standard 5 speaker array

• To conduct extensive research into Psychoacoustic Theory and spatial perception and apply relevant findings to this project:

• To research subjective testing methods and current testing standards in order to assist in formulating suitable and ethically viable listening tests.

• To design listening tests that assess the subjective imaging qualities of each panning law, relating which to limitations in traditional panning algorithms, and to the limitations of the ITU standard 5 speaker array in reproducing truly immersive multi channel audio.

• To successfully implement subjective listening tests using an appropriate number of participants, to be carried out in The University of Derby's Multi Channel Sound Research Laboratory.

• To compile test result data coherently, and perform an analysis of the results to facilitate the drawing of conclusions as to how well each panning algorithm performs in each area tested.

• To use the analysis of the listening test results to attempt to make conclusions about the benefits or drawbacks of each panning algorithm, related to evidence of the performance of other types of panning algorithms and the limitations of the ITU 5 speaker array.

• To evaluate the project as a whole, and suggest further research or changes that could be made to strengthen or corroborate any conclusions drawn.

Chapter 2 – Psychoacoustics and spatial perception

2.1 – Introduction

In this chapter, the psychoacoustics of spatial sound perception is discussed. Without comprehensive research into this subject, it would be impossible to draw any valid conclusions about the panning methods under test. The tests being performed on each panning algorithm for this report will aim to judge qualities on an azimuth around the listener, as we are concerned only with the localisation of virtual sound sources from horizontally arrayed loudspeakers. The tests were carried out in a non-anechoic environment which has been acoustically treated and optimised for these types of tests. Azimuth can best be expressed as a projected circle encompassing the listener's head on the horizontal plane, which is measured in degrees: 0˚ at the front of the listener's head, covering a full 360˚ rotation back to 0˚. Figure 1 illustrates this:

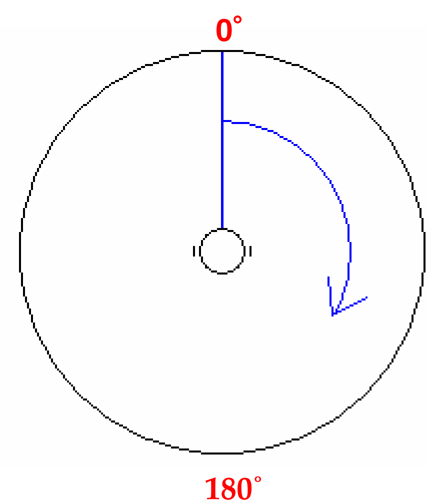

Figure 1. Azimuth around a listener

> 'Azimuth can be measured as an angle (in degrees) between a projection of a vector onto the horizontal plane and a second vector extending in front of the listener.' (West, J, 1998, pg. 13)

In this report, it is important to focus mainly on the areas of spatial perception relevant to judging qualities of panning decoders on the azimuth. However, still providing adequate information of the many other ways in which humans perceive the spatial sound field is beneficial in understanding the limitations of this research, and the possibilities of further testing or development of the particular panning algorithms studied.

2.2 – Localisation of sound

Humans are able to perceive spatial audio events and localise sound sources very effectively. The most obvious way this is possible is because we possess two ears on either side of our heads. Many cues are used by the brain to process and interpret very slight differences and variations in the qualities of the sound received by each ear. Among these cues, there are several that are relevant to this investigation, and several that are not. Consider a source, off-centre and in front of a listener on the horizontal plane, such as in Figure 2. The shortest signal path to each ear is marked (a) and (b). For the ear with the longest path, (b), the distance the sound has to travel will be longer. The signal will also be affected by the listener's head acting as an obstacle (Wiggins, 2004, pg.9-11). As the panning algorithms under test will be assessed for azimuthal imaging qualities, without concentrating on elevation or spatial envelopment qualities that aid localisation, the following topics will be discussed first, followed by an investigation of surround standards, and their relevance to this research:

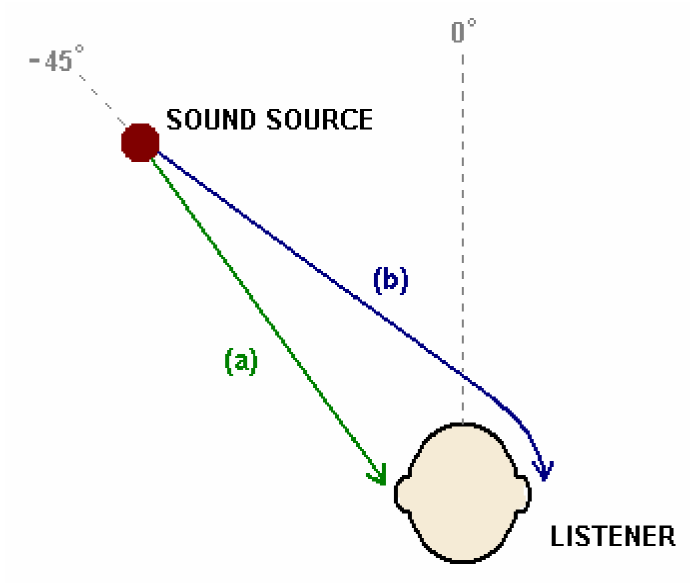

Figure 2. Paths of sound to each ear from a source at -45° on the azimuthal plane.

- Lateralisation.
- Inter-aural Level Difference (ILD).
- Inter-aural Time Difference (ITD).
- Inter-aural Phase Difference (ITD).
- Head Related Transfer Function (HRTF) and Motion theories.

Lateralisation can be expressed using Figure 2. The differences in source distance to each ear of the listener and the obstacle of the head will cause intensity and temporal differences to be perceived between the left and right ears (ILD and ITD respectively). When this effect is replayed over headphones (which exclude spaciousness and envelopment information), it gives the listener the feeling of an in-head sound source. The azimuthal location of the in-head sound is derived by the process of Lateralisation.

2.2.1 – Inter-aural Level Difference (ILD)

The intensity of the sound experienced at each ear will rarely be identical, especially for an off-centre source, such as in Figure 2. The reduction in sound intensity caused by the extra distance that the sound has to travel to ear (b) is minimal, so the relative ILD will appear to be further reduced when the distance between source and listener is increased. This is because the angle of incidence of the signal to each ear will get more and more similar the further the source is from the listener (Gulick, 1989, cited in Wiggins, 2004). The inverse square law states that for every doubling in distance from a point source, a signal will lose a sound pressure of 6dB (Davis, G, Jones, R, 1987). This and the angle of incidence can be used to calculate the extent to which the difference in signal path length and source distance affects the ILD but, as mentioned, is minimal and applies only to point sources in the free field. The most substantial factor affecting the ILD is due to the physics of the listener's head and ears. Because the head acts as a barrier to incoming sound waves, the signal has to travel around the listener's head. This does not pose a problem for low frequency sound waves, which have a wavelength larger than the diameter of the listener's head. But higher frequencies, with a smaller wavelength than the diameter of the head will tend to be attenuated as they cannot diffract around the obstacle (Wiggins, 2004, pg.11). ILD is generally accepted to be most effective at frequencies above 500 Hz (Di Liscia, O, P, n.d.).

2.2.2 – Inter-aural Time Difference (ITD)

The temporal difference between the signals to each ear is known as the ITD (Note that 'Time' and 'Temporal' are used interchangeably). The sound from any source is usually closer to one ear than the other, thus a difference in distance can be perceived as a delay in the sound arriving at the later ear (e.g. (b) in Figure 2). Again, the physics of the listener play a part in catalysing the effect of this perceptual cue as the sound still has to travel around the head. When the source is located directly to the side of the listener's head, the farther ear may receive a signal that is delayed by approximately 680μs. For impulse sounds, it is easy to notice the difference in the time at which the signal arrives at each ear, and this alone is known as onset-time disparity, or put more simply, the inter aural time difference perceived at the start of the sound. For longer sounds, the ear uses the onset-time disparity, and also makes use of phase differences between the signals. As this disparity is ongoing, it provides the auditory system with continuous information to help judge the ITD. It has been shown that the brain largely uses these ongoing phase differences to assess the ITD, rather than just the onset-time disparity (Buell et al, 1991, cited in Plack, C, n.d, pg.4). The ITD is a generally more effective cue at frequencies below 1000Hz, although we can use phase differences at up to 1500Hz (Wiggins, 2004, pg. 35).

2.2.3 – HRTF and Motion theories

These perceptual cues are used quite effectively to determine the angular direction of an incoming sound. But used alone they would fail to deal with front-back reversals, which can be perceived when signals from the front and rear have the same level and time/phase disparities. Expressed by Begault (1994), the 'cone of confusion' best illustrates how this can occur (Figure 3) from sounds that are coming from the front and rear, as well as above and below a listener. For this report the testing concentrates on the horizontal azimuth, so only confusion between front and rear source location is considered:

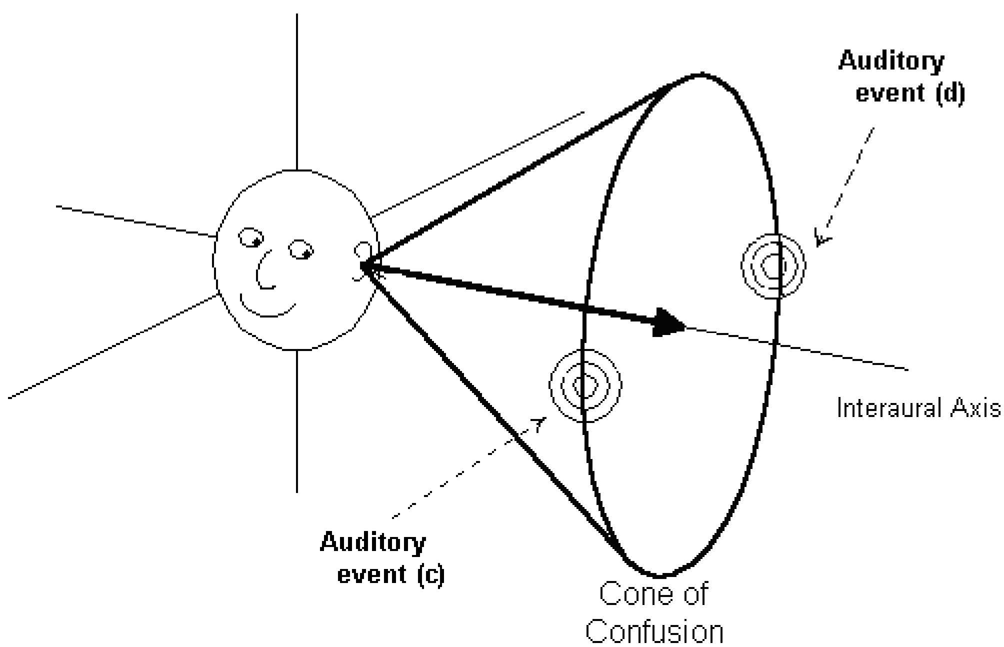

Figure 3. 'Cone of confusion'.

Auditory events (c) and (d) in Figure 3 both have the same ITD and ILD, so they would, in principle be subject to perceptual confusion. The way the auditory system deals with this phenomenon, which can occur when signals come from the extreme sides of the listener, is simple. The acoustic interaction between the sound and the listener's torso, head, outer ears and ear canals gives us vital information about the direction of the source (West, J, 1998, pg. 16-18). The measurement of the Head Related Transfer Function (HRTF) is a way of working out what the physics of an average listener's body and ears will impart on a sound signal arriving from any direction. It has been measured in the past by placing a microphone capsule into a person's ear canal, or alternatively the ear canal of a mannequin, and recording a series of test tones reproduced from different positions around the body of the subject. The outer ear (pinna) alone can give us enough information to discern the angular direction of sound sources monaurally, so is especially effective when utilized with all the other spatial auditory cues (and the other ear!) (Gulick, 1989, cited in Wiggins, 2004, pg. 17-18). The pinnae reflect sound best above approximately 4 kHz, and the head and torso are most effective below 2 kHz. Wiggins (2004) showed that: amplitude differences between the two ears are detected above around 700 Hz, because of the diffraction of lower frequency waves around the head; and phase cues are most effective below 800 Hz. It was concluded that phase cues are generally more effective in helping us lateralise sources than higher frequency cues such as those provided by the ear canal and pinnae.

West (1998, pg. 18) explains the way in which our auditory system exploits these disparities in time, phase, and amplitude to strengthen the cues received about the location of a sound source:

> Because the effects of the head, torso and ear are such strong functions of the directionality of a source relative to the head, head movement plays an important role in spatial hearing.

The movement of the head helps differentiate between front/rear sources, especially if they are coming from extreme angles. This has been simulated in Wiggins (2004, pg. 78-83, 153-160), who investigated the possibilities of optimising the binaural method of using HRTF data, to simulate spatial sound over loudspeakers. It was found that if head movement was simulated, it lead to problems whereby the phase difference between two angular head positions was not always strong enough to cue the corresponding source position on the azimuth, resulting in the effect of the sound source tracking the listener's head movements.

Angular head movement plays an important part in this report, and it has been shown by Rodgers (1981, cited in West, 1998) that phantom image instability occurs more under head movement. Phantom images are auditory sources that are being simulated by two or more loudspeakers, and are not necessarily located where the loudspeakers are located. In multi channel sound, when using the ITU standard 5 speaker set up, reproduction of phantom images is crucial to being able to simulate sources anywhere around the azimuth. Many audio engineers trying to create this effect using HRTF data and simulated head movement come across a common pair of problems: Every person has different physical dimensions, so it would be hard to standardise any such system (; and, more importantly, it is difficult to simulate reliable distance cues, let alone standardise them. So this type of optimisation is largely limited to binaural, in-head reproduction for headphones.

In the tests performed for this investigative report, the panning algorithms used were optimised 4th order Ambisonic decoders. Although they were optimised for irregular speaker layouts, such as the ITU 5-speaker array, they have never been subjectively tested together before. Detailed in Chapter 4, the listening tests were performed both restricting each participant's head movement, and also by allowing them natural lateral movement during the various tasks.

Chapter 3 – Surround sound

3.1 – Introduction

The development of digital audio technology has quickly paved the way for consumer multi-channel audio to sound more and more realistic. As Barbour (n.d, pg.17) puts it, things need to progress so as to maximise the limited capabilities to which the 5 channel domestic surround sound format falls hostage:

> It has been a goal of many audio practitioners to deliver the immersive, three-dimensional sound field of the real world, or of their creative imagination, to listeners in their home environment.

Traditional panning laws often rely on the pair-wise, power panning methods developed from early Stereo. Based simply on level variation between channels, these 'amplitude panning' laws have proved to be effective when panning left to right in both stereo and 5 channel systems, front and rear (Holman, 2000, pg.122). This can be best understood by the fact that humans possess only two ears,-one on the right and one on the left! However, where the pair-wise laws tend to lack is in reproducing convincing side imaging in a 5 speaker set up. The nature of decoding multi channel audio to be ambiguous to the listener's loudspeaker set up means that the ability of a panning algorithm to produce consistent phantom images at any azimuth for an irregular speaker array is instrumental to its success. A number of criteria for assessing the quality of each panning algorithm tested in this report will be discussed in Chapter 4. In this chapter, the irregular ITU irregular 5 speaker array is analysed, and its limitations highlighted, relevant to the tests carried out in Chapter 4. Panning laws will be explored, outlining the need for a cross-platform, 5 channel panning algorithm. This will include investigation into the ways in which panning laws have been optimised using Gerzon's Ambisonic system, and how the laws under test were designed (Gerzon, 1974).

3.2 – Loudspeaker layouts and existing panning methods

3.2.1 – ITU 5 Speaker array

The International Telecommunications Union (ITU) recommended speaker setup forms the basis of the array used for the listening tests in this investigation. Known as 5.1 surround and 3/2 stereo, this format was designed without the intention being to 'provide a full surround imaging plane giving unlimited directional imaging of arbitrary events', as Günther Theile (2001) puts it. The idea was to strengthen the stereophonic image at the front with a centre speaker, and utilise the rear surround speakers to present effects and ambience information which can provide space cues. This was beneficial to the sound-for-picture industry because of the central, frontal placement of screens. The centre channel was designed distinctly with the purpose of providing increased stability of centrally panned images for off centre listening positions, such as for dialogue. Wiggins (2004) explains that with traditional stereo, as a listener moved off-centre, a centrally panned image will tend to be pulled to the speaker closest to the listener, due to the ILD and ITD between the speakers increasing. The centre channel at 0° reduces this problem.

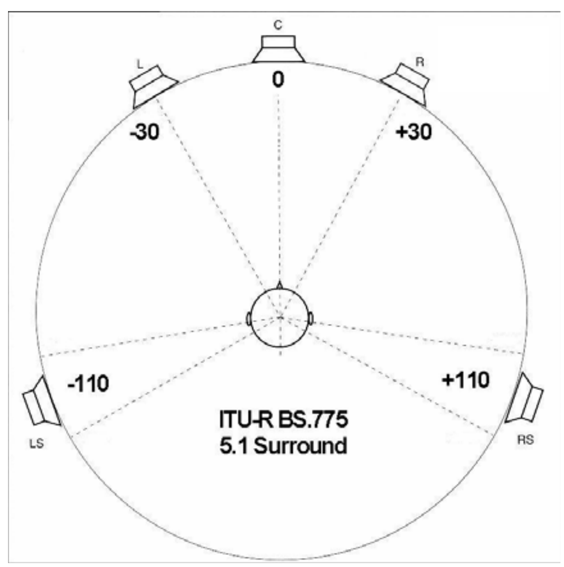

Figure 4. The ITU recommended 5 speaker layout

Figure 4 shows a standard ITU layout. The left (L) and right (R) speakers are placed at ±30˚, based on early stereo ideas that the set up should form a 60˚ equilateral triangle linking listener and speakers. This known to be the best configuration for stereo, as the relative width maximises the stereo perception, but is not too wide as to disturb the imaging qualities. Coupled with the centre (C), this makes for frontal lateral imaging qualities which are ideal for 5.1 surround (Holman, T, 2000).

The surround speakers, often referred to as 'left surround' (Ls) and 'right surround' (Rs), are at ±110˚, and complete the 5 speaker layout that has been used in this investigation. First designed for its potential use in motion pictures, several 'surround systems' were developed to maximise the ITU layout for domestic use, whilst maintaining economy in transmitting the signal in whatever form. These lossy codecs were specific to the ITU layout and were shipped on disc and tape formats to correspond with the standard 5.1 setup that was shortly to be in wide domestic use. Examples include Dolby Digital AC-3 and DTS systems. These are systems that define the speaker layout, and utilise perceptual data reduction techniques to provide signal to the rear surround speakers. To enable the format to be reproduced in 2 channel stereo by the consumer, in case they did not possess a surround system, Dolby stereo was delivered as two channels of matrix coded audio that could be up-mixed to 5 channel surround. This was done by summing the left and right channels to create the centre feed, and band limiting and phase shifting the left and right channels to send to the surround channels. The surround feeds could then be delayed to give the illusion of spaciousness (Wiggins, 2004). In 1992 further enhancements to this format were made, including abandoning the matrix codec for the delivery of 5 discrete channels, plus the addition of the <120Hz LFE (0.1 channel) we all know (Kyriakakis, 1998, pg.943).

3.2.2 – Panning techniques

Pair-wise amplitude panning is commonly used in consumer stereo and surround systems that pre-define the standard speaker layout. It literally means that, at the most, only a pair of speakers will be active at any one time. Also referred to as power panning, the method is also used in systems where rear speakers are required to produce direct sound images. When used with the ITU speaker layout several problems have been noticed: Pair-wise panning does not use any of the psychoacoustic principles explained in Chapter 2 to reduce the listening fatigue when a listener is trying to localise sound. For the ITU set up, this poses problems with the stability of phantom images panned to areas that have a wide angle between pairs of speakers. Between right and left speakers at the front and at the rear, the imaging does not suffer a great deal. It is at the sides that the system has most trouble. Either images, when panned centrally between Ls/Rs and L/R, appear to be further forward or will tend to jump or 'spectrally split' between the loudspeakers (Holman, 2000, pg. 103). This is obviously not desirable if one needs the perceived direction of a source to be indiscriminate to the limitations of the available loudspeakers, thus providing full 360˚ surround.

The codecs described previously differ from what is being investigated in this project. The panning algorithms tested for the purposes of this investigation define how the audio is panned, irrespective of what speaker layout is used.

3.2.3 – Ambisonics

The most relevant of which, the Ambisonic system, was developed after the failure of the first attempted domestic 'true surround' format - Quadraphonics. Whilst the motion picture industry hedged their bets on lossy formats that made the best use out of a limited speaker set up, Quadraphonics was being developed for home use. The idea behind it was to use stereo panning techniques between all four quadrilaterally placed loudspeakers, so as to create phantom images at any azimuth. It utilised the matrix system employed in formats such as Dolby Stereo to deliver the signal over two channels. However, testing showed that stable side and rear images were almost impossible to emulate with the Quadraphonics set up, and the larger frontal angle of ±45˚ between left and right was sub-optimal. The effective listening area was also very minimal:

> The systems only worked at all if the listener remained in a very small, highly impractical 'sweet-spot'… (outside of which) the imaging quickly collapses into puddles of sound at the speakers with the slightest movement either side of the central axis (Robjohns, 2001).

Ambisonic was developed by Gerzon utilising some of the failings of amplitude panned surround systems such as Quadraphonics. They proposed to create a new, psychoacoustically correct surround format that would be indiscriminate of speaker layout and in line with current and future transmission media. Unlike the speaker-dependent systems which deliver direct loudspeaker feeds to create phantom images, Ambisonic uses psychoacoustic principals to create the effect of phantom images encoded in the source material, which is played back over any speaker layout. Early Ambisonic formats were some of the first explorations into psychoacoustic-based panning for multi channel sound. Since the B-Format was developed with scope for periphonic (with height information) sound, as well as full 360˚ azimuthal sound, efforts were concentrated on just lateral 360˚ reproduction (without height), and the possibilities of matrix encoders to deliver the format in many different situations to the consumer (Gerzon, 1992). Because the coding and decoding process is separate, it makes it possible for Ambisonic codecs to be designed for different or irregular speaker layouts, such as the ITU array.

Panning algorithms can be represented as virtual microphone polar responses for a given speaker array. Following techniques adopted in Blumlein's 'shuffler' for stereo to binaural reproduction, Ambisonics took the idea of deriving custom microphone responses from figure 8 patterns for three dimensions (Figure 5- X, Y, and Z), and added an omni-directional microphone response ('w') to record the zero-th order signal, effectively the sound field pressure signal. Figure 5, which is from Gerzon's 1992 paper, illustrates the polar pickup pattern for a first order, B-Format (periphonic) system. B-Format was the starting block for Ambisonics, and could deliver full 3D sound through four channels. Since we are concentrating on only the azimuth, the 'z' channel can be assumed to have a response of zero. This reduces the channels to three instead of four. The microphone responses were not limited to a figure of 8 pattern, but could be extended all the way to sub-cardioid and omni patterns so the signal could be decoded to a speaker array. These are ways in which a signal is coded, but more importantly, how is the signal decoded? Before looking at this, it must be made clear that Ambisonics was designed and tested using regular arrays of equally spaced loudspeakers. As we are looking strictly at the ITU irregular 5 speaker array, it follows that the basic problems when decoding an Ambisonic signal over this array should be at least covered. Gerzon

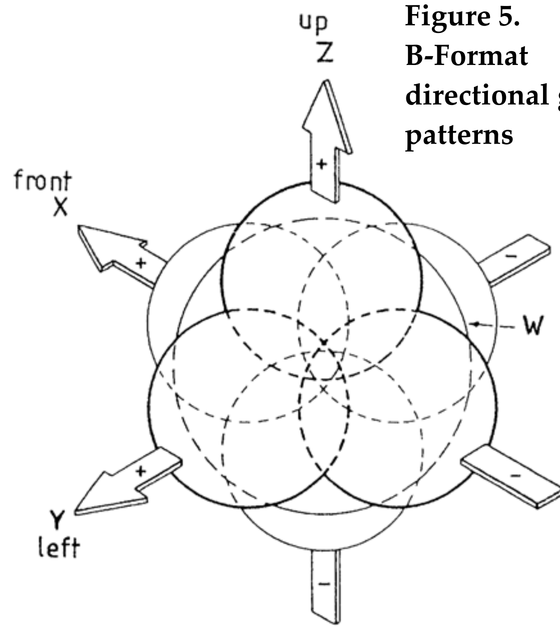

Figure 5. B-Format directional gain patterns

anticipated these problems in the 1970's, but was then unable to solve the various equations which needed to deal with the irregular layout and psychoacoustic factors involved in the placement of each speaker around the listener (Wiggins, 2004, pg.49-55). The psychoacoustic effects of sound measured at the listener position were generally used to assess the quality of localisation in multi channel systems.

3.2.4 – Velocity and Energy vector plots

Another proven way of assessing/presenting the overall quality of localisation for a speaker array is by the use of velocity and energy vector plots. Measured from the listener position, the velocity vector and energy vector lengths are a measure of localisation quality: Zero (quality) in the centre of the plot, which extends to 1 on the outsides of the plot, indicating a good localisation effect (and less listening fatigue). The vector angles represent the perceived angle from which the majority of the source appears to come, and the intended source angles are usually marked on a plot, providing a reference to assess this quality. For examples of vector plots, please see Chapter 3.3, which shows the plots for the three Ambisonic-based panning laws used in the subjective listening experiment. To optimise a panning law using vector plots would indicate that to get the best quality, both vector representations of the localisation qualities need to be as close to 1 as possible all around the 360° azimuth (Gerzon, 1992). The effect of an energy vector that is lower than 1 is that images are not as stable at listener positions off from the 'sweet-spot'. The effect of a velocity vector that is less than 1 is that images become quickly unstable when the listener moves their head. This is a result of HRTF's observed in Chapter 2, which supports the view that there will be a relatively wide tolerance for a sub-optimal vector length at front and rear locations in the ITU array, but it is very difficult to satisfy these functions at the extreme sides of a human listener (Gerzon, Barton, 1992, cited in Craven, 2003).

Velocity and energy vectors are not derived by the author for this investigation, but available plots will be subjectively assessed in order to inform the author and support findings from the listening tests.

3.3 – Panning laws under test

There are three panning algorithms that have been subjectively tested for this investigation. These laws will be detailed in this chapter, outlining any findings from past tests, concluding with a summary of the objective qualities of each algorithm, relative to each other and to the traditional pair-wise methods discussed in Chapter 3.2.

3.3.1 – Craven's 4th order panning law (2003)

This panning algorithm (Named '**C03**' for the purposes of the test section) uses 4th order circular harmonics (as used in Ambisonic systems) and is designed for reproduction over an irregular ITU 5 speaker array. Four channel, B-Format was defined in the 1970's as full 'first order' Ambisonics. Spherical harmonics added to the original W, X, Y, and Z configuration could raise this order, but there would also need to be more loudspeakers in order to fulfil Gerzon's own criteria for an Ambisonic system. For instance, the minimum number of loudspeakers in a 2nd order system was defined as 6 for horizontal reproduction, whilst 1st order systems were meant to utilise at least 4. Problems that arose from this were that calculations became more and more tedious beyond the 3rd order, and the normal way of optimising the decoders using vectors was not sufficiently accurate. Peter Craven designed this decoder to be frequency independent, which has been shown to be beneficial in terms of the compatibility with software/hardware platforms. However, the vector plot (Shown in Figure 6) shows that the high and low frequency vector angles are not ideal (Wiggins, 2004, pg.171).

Because the decoder was designed to be frequency independent and the velocity and energy vectors represent qualities at opposite ends of the frequency spectrum, a compromise between the qualities of the energy vector and the velocity vector is made.

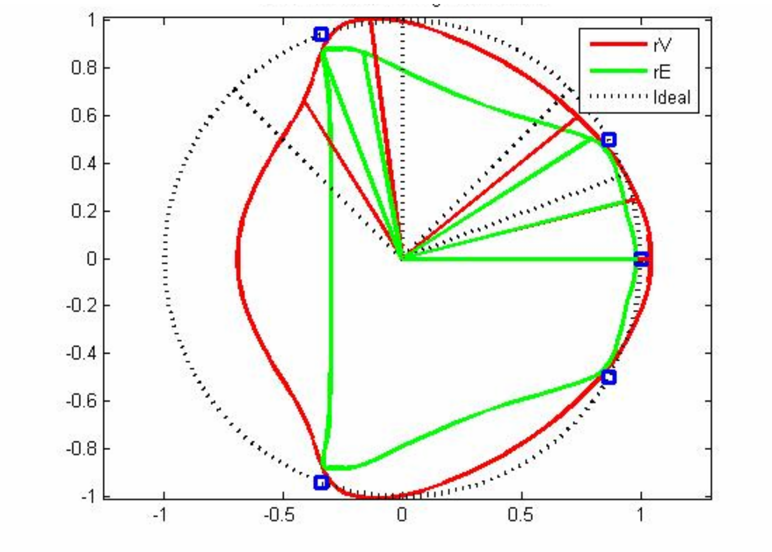

Figure 6. Energy and Velocity Vector Plot of a 4th Order Ambisonic decoder proposed by Craven (2003).

In Craven's (2003) report, **'C03'** is auditioned on an ITU 5 speaker irregular layout, and optimised using vector analysis. Several informal listening tests were performed using the 4th order decoder against a pair-wise amplitude panned decoder. The panning law was successful in many respects, and in some respects, the research highlighted

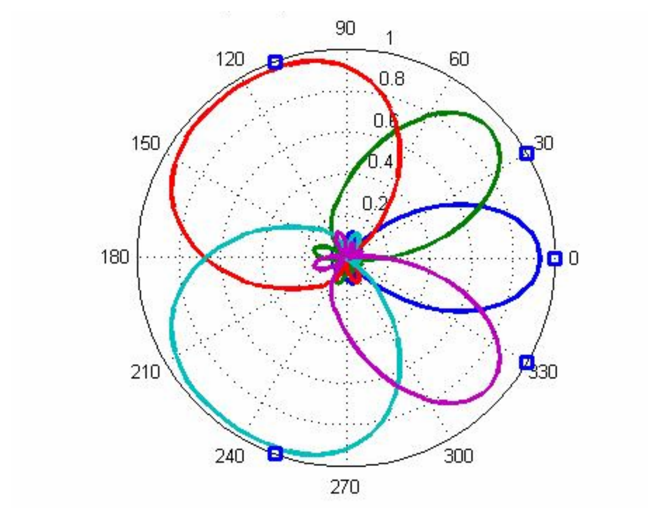

Figure 7. Virtual microphone response pattern of Craven's decoder (2003).

much potential in this type of 4th order circular harmonics decoder. Problems were found that were not wholly indicative of problems with the panner. The directional error (in both velocity and energy vectors) at ±50˚ on the azimuth was substantial for both the pair-wise and Craven's decoder, which suggests that the narrowly spaced front of the ITU layout may be exhibiting these characteristics. Observations noted from the subjective listening tests were that Craven's decoder, when panned across the frontal stage, was smoother, timbrally and spatially, than the pair-wise decoder. It was also observed that images became less stable at around ±40˚/±45˚ as sound was panned from front to rear, and side images tended to jump between the L/R and Ls/Rs speakers when testing Craven's decoder. Subjects also noted that side images were sensitive to head movement, but that images panned to 180˚ appeared more focussed than the pair-wise. This can be illustrated by the vector plots, as the velocity vector length at 180˚ is substantially closer to unity than in the pair-wise plot, due to phase-reversed signals being sent to the front speakers as sound is panned to the rear (Craven, 2003, pg.5). The majority of testing carried out on this algorithm is objective in nature. One aim of this project is to perform more subjective testing using this, and more recent 4th order optimised decoders for the ITU irregular 5 speaker array.

3.3.2 – Wiggins' optimised irregular Ambisonic laws (2004, 2006)

In his 2004 PhD, Bruce Wiggins proposed the optimisation of 4th order Ambisonic decoders (such as Craven, 2003) using the lateralisation parameters and velocity and energy vectors. It was by using a heuristic search algorithm that would satisfy the three 'fitness' criteria for the velocity vector and the energy vector, that the first of Wiggins' panning law under test in this investigation was designed. By manipulating the weighting of each vector quality, the decoder was be optimised. Because the energy vector covers a large frequency range for listeners at centre or off-centre (above 700Hz), Gerzon advises the use of the energy vector to plot the perceived amplitude of a decoder (Gerzon, 1977, cited in Wiggins, 2004, pg.172). The first decoder (Named '**W1'** for the purposes of this report) shows an improved low frequency velocity vector length, and improved correlation between these and the vector angles at both high and low frequencies (See Figure 8) (Wiggins, 2004, pg.173).

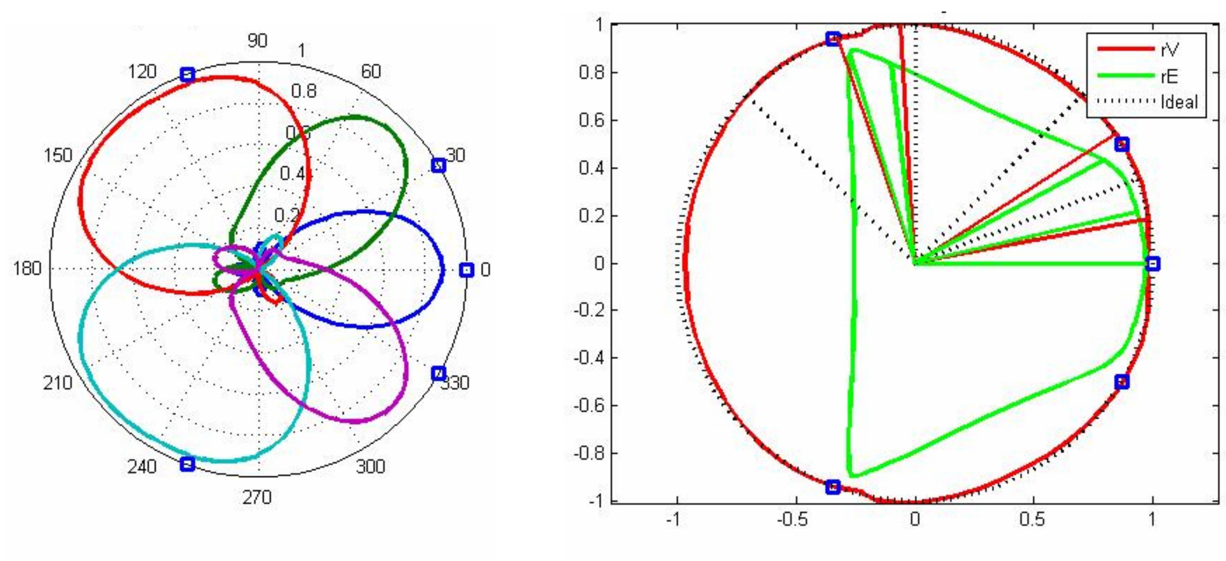

Figure 8. Energy and velocity vector plot, and virtual microphone response for 4th order optimised decoder 'W1' as proposed by Wiggins (2004).

Bruce Wiggins' 2004 Thesis describes the development of software based on a Tabu search algorithm that, once the five speaker positions were entered, would calculate optimized decoders automatically. The program was optimised so it could be adapted to use HRTF data (first proposed in Wiggins *et al*, 2001), both with and without head movement considerations. The use of HRTF data was shown to be flexible in that the effect of head movement could be quantified, which is not possible when just using energy and velocity vectors (Wiggins, 2004, pg. 269).

The second of Wiggins' panning algorithms under test (Named **'W2'** for this report) follows the research outlined in his 2004 PhD Thesis.

The decoder was provided to the author in early 2006 to be used in this investigation. It was proposed in 2004 that HRTF methods could be used to create an optimised decoder for an increased number of frequency bands. It can be seen from the energy and velocity vector plot and virtual microphone response in Figure 9 that it has been optimized from a law similar to '**W1**'. The directional disparity between the energy and velocity vectors is improved. This apparent 'localisation quality' has a much better perceived angle, at the cost of increased non-unity in the velocity vector. It was shown in Wiggins (2004) that the energy vector is of greater importance when designing a panner of this type. It appears that this was also the case when '**W2**' was designed, as the energy vector is particularly correlated with the velocity vector in perceived angle, and also shows optimal side and rear angular localisation quality. Not only this, but there is a better energy vector length at the rear, which is closer to unity than in '**W1'.** The apparently lower vector length at the front and sides is likely to be due to the advent of frequency bands in the design of the decoder, which should improve the quality of off-centre listening. This may explain why the 'average', low frequency, velocity vector length is closer to zero. That said, the velocity vector length at the rear is of similar magnitude to that in Peter Craven's 4th order law (Craven, 2003), but with improved directionality.

The extended velocity vector at the rear can be observed in the virtual microphone response plot for '**W2**' (Figure 10), which shows that Ls and Rs polar responses overlap considerably.

The intentions outlined in conclusion to the development of Ambisonic decoder optimisation in Wiggins (2004) were to use the software to take into account off-centre positions, potentially enlarging the 'sweet-spot' size, perhaps resulting in slightly sub-optimal results at the centre position.

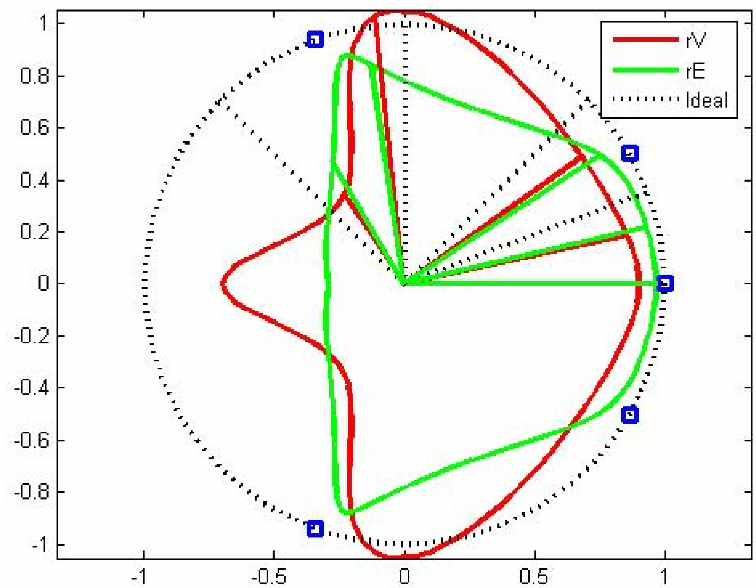

Figure 9. Energy and velocity vector plot for optimised 4th order decoder, 'W2' (Wiggins, 2006).

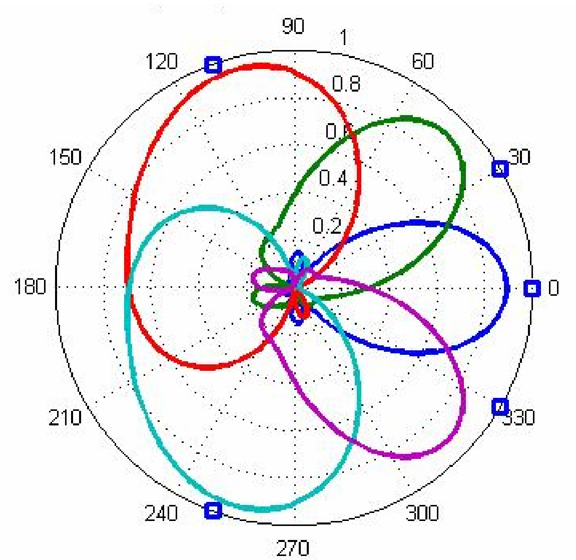

Figure 10. Virtual microphone response for 'W2' (Wiggins, 2006).

3.3.3 – Objective summary

When objectively compared with pair-wise amplitude panning methods it is clear that the above Ambisonic-type panning laws are very different in their approach to recreating the spatial environment. Amplitude panning has been used in stereo reproduction for a long time, simply because it is very effective. However, coupled with the ITU layout discussed, it has been shown to be sub-optimal, as side images will always suffer when compared to frontal images (Craven, 2003, Wiggins, 2004). As stated in Chapter 1, this is not a major problem in a sound-for-picture situation, where accurate side imaging is not essential. This is particularly true when a source is panned near to a speaker position, as there will be less localisation confusion caused by more than two speakers radiating simultaneously. Martin *et al* showed that Ambisonic systems also fall hostage to the effects of HRTF for images at the side, although it was tested on an 8 speaker system. Another observation from the subjective tests carried out in that paper was that the Ambisonic systems tested displayed the characteristic that the sound appeared to be closer to the listener than with the pair-wise panned method. This brings us back to what is being tested in this investigation. Amplitude panning will always give the listener more perceived distance, but this is not of concern in the tests that will follow, as it is mainly lateralisation parameters that are being studied (Martin *et al*, 1999, pg.10-12).

Chapter 4 – Subjective listening tests

4.1 – Methodology

The three panning laws described in Chapter 3.3 were subject to listening tests. The methods used and reasons for testing will be outlined in this chapter. It is hoped that by testing Craven's 4th order circular harmonic decoder (2003) with two more recent optimisations proposed by Wiggins (2004, 2006) it will be possible to assess the effectiveness of the methods used to optimise decoders using HRTF and search algorithms, and subjectively evaluate the characteristics of these Ambisonic based panning laws for localisation. Using the University of Derby's Multi Channel Sound Research Laboratory, it is easy to implement these tests, so suggestions for further research are made in Chapter 6.

4.1.1 – Design of tests

As mentioned, the subjective listening tests performed for this report have concentrated on the horizontal azimuth (without height information). This specifically refers to the azimuth at ear-level around a seated listener. A number of qualities have been assessed for each panning algorithm, and these can be related to some of Gerzon's criteria for designing pan-pots. The factors relevant to this report are outlined as follows: 1) The localisation angle equals the panning angle; 2) Images are stable with off-centre listener positions; 3) Smooth and uniform movement as the control is moved, and approximation to constant power behaviour for uniform loudness. In summary, these factors mean that ideally a panning law will represent sources as close to the true location consistently, when factors such as off-centre listening, head movement and source movement are introduced. The other quality of localisation that has been tested is the perceived source image 'focus'. This quality has some 'grey area' associated with it, as not everyone will understand image focus in the same way. For this investigation, it can be explained as the perceived 'width' of a phantom source – if the image appears larger than the anticipated source it is judged as less focused. Martin *et al* (1999) suggest that the focus of an image is defined by the perceived width of the source, as well as increased distance and less low frequency spread. The tests in this report concentrate on just the perceived width of a phantom source as the governing attribute. The 6 participants used were all University undergraduates of roughly the same age and hearing ability, and were all studying a related Music Technology or Computing degree programme. This was beneficial because they already possessed a foundation of knowledge of the terminology and effects being created for the test.

Details of the three localisation qualities of each panning law tested are as follows:

1) **Image stability** of a panned source, looking particularly at the consistency of the movement, the consistency of volume, and overall perceived 'quality' of the panned sound: To test the consistency in movement and amplitude of the source, broadband pink noise was panned 360˚ around the listener so that a complete rotation took 8 seconds. Pink noise is useful for a test of this type because 'it (pink noise) is considered to excite all frequency bands and both ITD and ILD mechanisms evenly' (Pulkki, n.d, pg.3). It was intended to make it easier for the listener to notice discrepancies in the movement and loudness of the signal as it was panned. Each of the three panning laws was utilised, with short pauses in between when the noise was played. The last test in this section was explained to the listener to be a subjective opinion of the overall quality of the panned source. They were asked to listen to the three panning laws again, this time with an anechoic trumpet piece as the sound source, panning around the azimuth. For this section, it was important to

obtain useful results that accurately reflected what the listener thought and perceived, so a system of ordering results was used. Rather than scoring each panner, they were asked to order them from perceived worst to perceived best. This removed any absolute value of the scores, and made them relative only to the other scores in the test (Mason *et al*, 2000, pg. 380). This made it easier for the listener to express their opinions quickly and confidently on the test sheet.

2) **Image focus** or angular width of a source: For this test, a sample of an anechoic female human speech recording was used and played to the listener from agreed angles around the azimuth. Human speech is especially good to test for image focus because we are all trained listeners, and can judge easily whether a speech source appears to be focussed or not. It is relatively broadband and is comfortable to listen to (Martin *et al*, 1999). The source sound, which was 6 seconds long, was played and the listener asked to record the perceived angular width of the sound, from left to right. This was confusing at first, and most of the subjects could not grasp the concept from just reading the test paper notes. The use of verbal techniques helped clarify what they were listening for, and may have helped inter-result accuracy. This is suggested because the way in which perceived source focus/width was explained remained the same for every participant. This was that they should draw the source width in degrees from the left to right points of the 'widest' sounding part. To make the interpretation and recording of results easier for the author and the participants, a three-coloured pen was provided to record the perceived source width of each of the three panning laws. The participants were reminded frequently that the indicated location of the loudspeakers on the diagram, and in the laboratory, was for reference only, and that they may not necessarily be in use.

3) **Image accuracy**, which measures how well the perceived localisation angle corresponds with the panned angle of a source: This test was hoped to measure how accurately each panning algorithm could reproduce sources at specific angles. The material used was an anechoic sample of a rhythm being played on a djembe drum. This sound had quite narrow frequency content and had a lot of attack, which would make it easier for the participant to localise the angular direction of the sound. The participants were asked to mark their answers in the same colours they used for the previous test. To give the subjects a reference point, the sound was played at the location of a speaker before the test began. This was obviously subject to the effects of the decoder used to place it at the speaker location, but it was vocalised by most participants that it was useful that they heard the sound at all, before having to locate its virtual source position.

Each of the three sections (above) was split into a further three. The panning laws were tested under the following conditions:

The **first** test sheet required the participant to sit on the chair at the central position and voluntarily restrict their head movement. They were asked to try to hold their head in the direction of 0˚ at all times during this section. Angles at a resolution of 5˚ were marked on the test sheets for reference.

The **second** test sheet of each section required that the participant stay seated at the central position, but no longer restrict their head movement. It was made clear verbally, that participants were free to move their heads to aid localisation, but strictly lateral movement (rotation about the azimuth) only. This was not physically enforced, but as each subject understood what was required, it was of no great concern. Natural head movement to assess elevation properties would have inadvertently occurred, but by assuming that it would be of similar magnitude for all tests, it would be relative.

The **third** test sheet of each section required the participant to be seated off-centre (specifically 2 ft left of the central position), but still allow head movement. For each of these circumstances the test material and instructions remained the same.

The use of non-verbal and verbal elicitation techniques has been studied by Mason *et al* (2001) among many others. The subjective listening tests for this investigation utilised both ways of communication, and were informally monitored by the experimenter. Prior to the listening tests starting, the participants were encouraged to make notes on the test paper, and ask questions if necessary. For each test, the sheet was read through with the experimenter beside the subject. If the participant did not understand something, it was explained using visual metaphors. The author was careful, however, not to affect the results in any way, so made sure to explain any specific requirements of the participant in the same way (wherever possible) for every participant. It has been suggested that the use of non-verbal techniques, such as drawing, can be favourable for tests such as these, because of the similarity of the neural processes linking auditory and visual senses (Mason *et al.* 2001, pg.380). To this end, listeners were encouraged to note any observations on the test sheet. It may be observed (See appendix 1) from the test sheets that they showed the layout of the speakers in the Multi channel Research Laboratory.

The order in which the panning laws were auditioned as the tests were carried out was, as far as the participant knew, randomised. Actually, they were swapped around by the experimenter at points, and the order was noted down prior to tests taking place. Although this is not technically randomised, it was suggested to the listener that the panning laws may not be auditioned in any order, which may insinuate that in fact the order was randomised, thus further hindering any 'learning' of the order that may have taken place. For details of the order of presentation of the panning laws, exact parameters used, and actual test sheets please see accompanying Logbook (Appendix 2).

4.1.2 – Test Material

The material used for the listening tests was selected specifically. All sound files were sampled at 44100 Hz and were in 16 bit Windows Wave file format (Microsoft Corporation, 1990). All sources were mono, and royalty free. Each sound file was edited to be 6 seconds long, and was uncompressed. All source sound was recorded in a near anechoic environment, apart from the pink noise used for the panned image stability test.

4.1.3 – Multi channel sound research laboratory

The tests took place in The University of Derby's Multi Channel Sound Research Laboratory over two days. The acoustic properties have been obtained from Wiggins (2004), which states that:

> The room has been acoustically treated and a measurement of the ambient noise gave around 43 dBA in most 1/3-octave bands. The RT60 of the room is 0.42 seconds on average.

The reverberation time and ambient noise level are relatively low for a room of this size and in this location (a busy technology department).

4.1.4 – Matlab and Simulink

Using a PC and a multi channel soundcard the tests were modelled in Simulink, which is a 'block-diagram based approach to the Matlab programming language' (Schillebeeckx, 2001, pg.1). The code was kindly provided by Bruce Wiggins, who had previously designed a 5 channel output block that utilizes the Soundscape Mixtreme 16 channel soundcard. This block was simply used as the output, and real-time input blocks could be added that allow the experimenter to alter parameters on the fly. Between these input blocks and the output block, a model was added that allows the user to switch between panning decoders in real-time, simply by clicking a dedicated button. The source sound files were inputted using the 'wave from file' block in the Simulink library. The other variables (panned angle, and rotated pan) were also modelled using library blocks that, once the parameter was defined, could be altered in real-time. Variables within these functions included defining the panned angle in degrees (or radians), defining the speed of the rotary panned source, and how far around the azimuth it was to travel.

Chapter 5 – Results of listening tests

5.1 – Interpretation of results obtained

The raw results for each test can be observed by referring to the accompanying Logbook (Appendix 2). A statistical analysis has not yet been performed on the results of the listening tests. However, approximations have been made as to the performance of each panning law in each localisation quality tested, taking into account anomalous results and limitations of the test procedure. Participants often made notes of observations during testing, and this information is often used to support any results or trends in the performance of each panning algorithm. For the 'Image stability' test (where pink noise is panned around the participant), ordering rather than scoring was used to record results (Projekt Verdi, 2002) Because this method was used, it was simple to quantify results.

5.1.1 – Image Stability

With the listener at the central or 'sweet-spot' position, with restricted head movement, it was seen that the consistency of speed (*which will highlight any panning laws that are subject to the sound 'jumping' between adjacent speakers, in particular at the sides and rear of an ITU layout*) was perceived to be best when using the '**C03**' (Craven, 2003) decoder. This was not unanimous, and '**W2**' (Wiggins, 2006) had almost as many 'best' results. For consistency of amplitude, '**W2**' seemed to be the preferable decoder. The panning law that performed worst on average in these tests was '**W1**' (Wiggins, 2004). This was not expected as Craven's panning law has been shown to exhibit 'jumping' source images between front and rear speakers. This decoder has shown good results when images are placed at 180˚ (Craven, 2003). However, one participant noted that '**C03**' caused the sound to jump between Ls and Rs for every test. This was not observed by the other participants, who perceived '**CO3**' to perform best in this respect.

When head movement and off-centre listening was introduced to the tests, '**W2**' performed best for the majority of participants. The optimisations made when designing this decoder were specifically intended to improve the off-centre listening and cater to the psychoacoustic effects of natural head movement when localising a source (Wiggins, 2004). As Gerzon's criteria for evaluating the quality of panning laws states, the requirement is that a good panning law exhibits 'smooth and uniform movement as the control is moved, and approximation to constant power behaviour for uniform loudness'. This may not necessarily provide the best results when tested in a real situation, as the psychoacoustic effects do not always favour this method. '**W1**', on average, showed the worst image stability for an off-centre position with the allowance of head movement.

In the last part of the test, for 'Overall quality', each panning algorithm was tested this time with a well recorded (anechoic) trumpet piece, which was again rotated around the azimuth. The participants were asked use their experience with the pink noise tests to compare the overall consistency and stability of the trumpet piece for each panner (which may not necessarily have been in the same order). '**W2**' performed best for the tests where head movement was allowed, for both centre and off-centre listening positions. The results for the central position, with head movement restricted were similar for all the panning laws.

5.1.2 – Image focus

The tests to judge the perceived phantom source width, (or 'focus' of the source) involved the participants drawing what they perceived. This coupled with the possible margin of error caused by differences in how individual people perceive aural 'dimensions', may invalidate the results somewhat. However, upon analysing the data, apparent trends emerged. '**W2**' performed the best (i.e. was more focussed) at frontal locations, while '**C03**' averaged to be the least focussed (for centre, and off-centre locations with head movement allowed). '**W1**' was the least focussed at the front when the listener's head movement was restricted.

The other tests had image source positions around the 'trouble areas' at the sides and towards the Ls/Rs speakers. On average, for the tests where the listener was in the centre and head movement was restricted, '**W1**' was the least focussed for most participants. When head movement was allowed (in the centre position), both '**W1**' and '**C03**' showed to be equally the least focussed. At the off-centre position, test results indicate that '**C03**' performed worst, and was noticed by several participants to be very diffuse and sometimes hard to localise. Tests show that '**W2**' exhibited the most focussed source images in all tests, and was noticed to sound more 'direct' than the other decoders for one participant. Where any of the panning laws showed a lack in image focus, it tended to appear that the images were being pulled towards the nearest loudspeaker. This said, there were a number of anomalous results that did not correspond with other participants results. For instance, '**W2**' was observed to be quite 'diffuse' at the rear/side for one off-centre participant, but performed the best in the same test for every other participant.

5.1.3 – Image accuracy

Similar to methods used in Nielson (1991), this test required the listeners to record their perception of the angular placement of a panned virtual source on a test sheet with marked azimuthal angles. The interpretation of this information was difficult, which may have been due to sub-optimal test methods (This is discussed further in Chapter 6). The average, overall error (in degrees) was calculated for each panning law, and is recorded below. The average error exhibited by each law in just the off-centre listening tests is also shown below.

Average overall percentage error:

Wiggins (2004), (W1)	Wiggins (2006), (W2)	Craven (2003), (C03)
57°	42°	34°

Average percentage error for off-centre tests:

Wiggins (2004), (W1)	Wiggins (2006), (W2)	Craven (2003), (C03)
98°	65°	55°

The number and area of perceived reversals (either between the front and rear, or the left and right) for off-centre tests was also noted:

Wiggins (2004), (W1)	Wiggins (2006), (W2)	Craven (2003), (C03)
2 – Left/Right	2 – Left/Right	2 – Front/Rear 1 – Left/Right

Using this crude averaging and ordering method for this analysis still shows that some decoders performed better than others for the accuracy of panned source tests. Although it can be seen that '**W1**' was subject to a higher average angular error and '**C03**' was most accurate overall, there were many anomalous results that would have affected the averages, potentially leading to inaccurate results. In some cases, for instance when '**C03**' was panned to 290˚ for off-centre listening, it was difficult for the listener to even determine where the source was coming from at all. In these extremes, it was said by several participants that they had to use the perceived amplitude of the source to localise its direction, just drawing the angular direction where it appeared to be loudest. The average overall error perceived for the off-centre tests was much higher for '**W1**' than either of the other laws. Because the test material chosen was not a broadband signal, the frequency content of the material used in the image accuracy tests may have affected the results. '**C03**' performed best in this test, but it should be noted that this law was designed to be frequency independent. '**W1**' was expected to perform sub-optimally for off-centre positions, but '**W2**' was expected to out-perform the other decoders at this listener position. Although close, '**C03**' appears to be the most accurate panning law in both cases. When '**C03**' was inaccurate, it often exhibited the perception of 'reversals'. The number of reversals perceived for each decoder for just the off-centre tests is shown on the previous page. '**C03**' performed the least accurately in this respect, with both front/rear and left/right reversals occurring. It was expected that '**C03**' would cause reversals when the participant was free to use head movement to aid localisation, and that '**W2**' would remain stable, but this was not apparent in the results. When head movement was allowed, each panning law had a marked accuracy improvement – but it was roughly equal for all the decoders. Another characteristic noted was that the perceived angular direction of the sources was pulled towards the nearest speaker in almost every instance for every panning law. It is unknown whether this was an accurate representation of the function of each decoder, or whether it was influenced by the presence of marked speaker locations on the test sheets and in the laboratory.

Chapter 6 – Conclusions

Referring to the interpretation of results, conclusions can be made regarding the success of each of the panning algorithms under test. When objectively analysing each panning law against a more traditional law, such as pair-wise panning the advantages can be theorised: Both velocity and energy vectors for pair-wise decoders will always be sub-optimal at the sides and rear when using an ITU layout. Craven outlined that non-ideal vectors will cause the decoder to exhibit problems under head movement and listener position, especially at the sides. It was also stated that the energy vector in a pair-wise amplitude panned system will always be better than for any other type of panning method (Craven, 2003).

It has been shown that pair-wise panning requires speakers to be no more than 60° apart for successful imaging. This is not possible with the ITU layout, so this panning method always shows weakness at the sides (As proven by the Quadraphonics experiments conducted by Gerzon). This is obviously not a problem in stereo reproduction, and pair-wise panning is often shown to be superior at the front than many other techniques. This can be partly explained by the use of cross talk cancellation to create many of the Inter aural difference cues used to localise sources in stereo. When objectively assessing a system of the type tested in this investigation, there are some obvious differences to note. It is much harder to design a non-pair-wise panning law that satisfies the complex pinna cues we use to localise sound, because there will simply be more speakers radiating at one time, and the direction-dependent filtering applied by the shape of the outer ear is hard to model. Wiggins (2004, pg. 272) explains that objective testing is not always the be all and end all of assessing the quality of a panning law. One that may satisfy the vector model and theoretically (and in practice) show better spatial accuracy may not perform as well when tested artistically with music. Below is a summary of what was achieved in this investigation, and observations from the subjective listening tests.

• An extensive literature review and discussion of the following aspects of multi channel audio was successfully carried out.

- Surround sound formats and history
- Traditional and recent panning laws for multi channel audio
- The ITU standard 5 speaker array

• Research into Psychoacoustic Theory and Spatial Perception was carried out successfully, although it proved very hard to focus the research on relevant areas early on in the project, because knowledge in certain aspects of multi channel audio was limited.

• Research into subjective testing methods and current testing standards was undertaken, to the end that if the listening tests were not based on these foundations, improvements could easily be suggested using this knowledge.

• Subjective Listening tests were successfully devised that assess the imaging qualities of each panning law. The results of which were analysed, and conclusions drawn as to the reason for any trends or anomalous results.

• Summarized Conclusions of results obtained for each panning algorithm:

- **Craven's (2003), (C03)** circular harmonic decoder produces stable moving images, but only when head movement is restricted.

- **Wiggins' (2006), (W2)** optimised Ambisonic decoder produces noticeably more stable moving images when head movement occurs, as well as for off-centre listening.

- **Wiggins' (2004) (W1)** optimised Ambisonic decoder performed worst under allowed head movement, and for off-centre listening.

- **Wiggins' (2006) (W2)** was judged by participants to have the best 'overall quality' when a source is moved around the listener, at both off-centre and centre positions with allowed head movement.

- **Wiggins' (2004) (W1)** and **Craven (2003), (C03)** produced diffuse images at most azimuthal positions when head movement was restricted.

- **Wiggins' (2006) (W2)** was noted to produce the most focussed images of the three panning laws under every test condition, and participants commented that '**W2**' and '**W1'** sounded more 'direct' than the other decoder at the central position. When asked about this during the write-up of this report, it was said that the sound appeared to be closer to the listener than the other decoder. This was also observed in Wiggins (2004, pg. 272) with earlier optimisations of the same type of panner. This said, it may have been due to the specific anechoic material that was used for this investigation.

- **Wiggins' (2004) (W1)** was shown to exhibit a larger average error over the course of the tests for image location accuracy, at off-centre positions.

- **Craven (2003) (C03)** performed best in image accuracy tests for the central listener position. But when tested with listeners off-centre, reversals were perceived where the image jumped, or was wholly perceived as a reversal (F/R and L/R).

6.1 – Evaluation of listening tests

The listening tests carried out were subject to a number of limitations which have been identified during the analysis of the results. A brief summary of changes that could be made to improve reliability are as follows:

6.1.1 – Panning algorithms

The tests were carried out on three relatively newly developed panning laws. They were all essentially based on the same 4th order Ambisonic method. The latter two of which were psychoacoustically optimised versions of the first, which utilised new methods of distributing and testing attributes in the stages of objectively modelling the decoder. It would have been beneficial to have tested them all along side a standard pair-wise amplitude-panned decoder, a commercially available decoder such as from Pro Tools or Waves, and an un-optimised B-Format Ambisonic decoder. This would have given a broader idea of the performance of the newer algorithms relative to current standards.

6.1.2 – Test design

The tests, if carried out again, would utilise a greater number of participants, preferably over 10. Also, the design of the sheets would be revised in the following ways: It was shown that sources that were to be directionally localised tended to get pulled towards the speakers. It was suggested that this may have been somewhat due to the fact that the participants had the layout of the speakers marked on the sheets, and also that the speakers were clearly visible in the test room. The speakers were marked on the diagram for reference purposes, but the marked azimuthal angles alone would have been just as effective a reference, as long as 0˚ was clearly marked. The only way of concealing where the speakers were in the room would have been either to cover them with material screens, or to not divulge which of the 14 or so speakers in the Multi Channel Research Laboratory were going to be used. For the image focus tests, there were problems with the design which were only apparent after it was too late. The system of ordering, as was used for the image stability tests, should have been used, whereby the listeners would be played the material and asked to draw their perceived source widths for each decoder, but after wards asked to repeat the test and order the decoders from 'more focussed' to 'more diffuse' or something similar. The angles used should have been more consistent, and the order in which the decoders were auditioned should have been more random. The most notable revisions would be made to the third test, the localisation accuracy of a static image. The decoders were not all auditioned at the same angle, so consistent and analytical results were difficult to obtain. Also, it was suggested in Wiggins (2004) that pulsed pink noise was better for localisation, and I would propose that for the accuracy tests, this pink noise would be used next time, so as to satisfy a broad range of frequencies. This would have provided more usable results between the frequency independent panning laws and the psychoacoustically optimised ones.

The use of verbal and non-verbal techniques to elicit results from the participants was effective, so in the future, specific techniques would be recorded, so that they could be subjectively assessed with other test methods. To further strengthen the results, each tested variable would have a calibration test sound played to the participant before the test began, so that they knew what to expect, and to give a reference with which to order the actual observations.

The particular variables that the listener was subject to (of-centre, head restriction etc) would be more precise, in particular the head movement variable would be altered so that they rotated their heads about a certain horizontal angle, which was consistent for all subjects.

6.2 – Evaluation of project

The project was a success in many ways, and in the others a valuable learning experience for the author. Considering the limited knowledge of surround systems, in particular the objective testing and optimisation of hierarchical panning methods, much was achieved in the way of research. Again, the author was only experienced in simple subjective testing, more suited to psychology, so this project provided the opportunity to design a listening test and go about analysing the (very) subjective results. The proper methods for a thorough statistical analysis were not tackled in this project due to the subjective nature of the results obtained. However, if the above recommendations were made to the design of the tests, this would be possible.

It is concluded that each panning law possesses characteristics that can be objectively qualified to a point, but which need subjective testing to make the research whole. If no subjective testing took place, the ability of the panning law to simulate complex psychoacoustic effects that are instrumental to human localisation would be limited.

6.3 – Further research

This project has highlighted a number of other areas which would require further research and testing to support this area of the qualification of panning laws for a 5 speaker array:

- The ability of a panning law to be down mixed to stereo, binaural and transaural recordings (and vice versa), so the speaker layout is not such an affecting factor in the quality of the sound.

- Further testing of psychoacoustically optimised Ambisonic decoders, which have been designed using HRTF and perceptual cues.

- Research into testing methods, in particular objective methods.

- Disc formats, hardware and storage media essentially define the audio quality that can be obtained by the domestic user, and thus limit the progression of some technologies into the home entertainment domain. Extensive research into these areas, and perceptual coding would be beneficial to this area of study.

Chapter 7 – References

Barbour, J. (n.d.), School of Social and Behavioural Sciences, Swinburne University: Australia.

Begault, D, R. (1994), '3-D Sound for Virtual Reality and Multimedia', Academic Press: Boston

Buell, T, Trahiotis, C, and Bernstein, L. (1991) 'Lateralization of low-frequency tones: Relative potency of gating and ongoing inter-aural delays', Acoustic Society of America

Craven, P. (2003), 'Continuous Surround panning for 5-speaker reproduction', AES 24th International Conference: Banff, Canada.

Davis, G, Jones, R. (1987), 'The sound reinforcement handbook', (2nd Edition), Hal Leonard Corporation: Milwaukee.

Di Liscia, O, P. (n.d.) 'Spatial listening and its computer simulation on electronic music', Universidad Nacional de Quilmes: Education Ministry of the National Govt. of Argentina.

Gerzon, M. (1992), 'Psychoacoustic Decoders for Multi speaker Stereo and Surround Sound', Proceedings of the 93rd AES Convention, October 1992.

Holman, T. (2000) '5.1 Surround Sound Up and Running', Focal Press: Oxford.

Howard, D.M, Angus, J. (1996) 'Acoustics and Psychoacoustics' (2nd Edition), Focal Press: Oxford.

ITU-R, 1993, Recommendation BS.775: 'Multi-channel stereophonic sound system with or without accompanying picture', International Telecommunications Union.

Kyriakakis, C. (1998), 'Fundamental and Technological limitations of immersive audio systems', Vol. 86, May 1998, No. 5, IEEE.

Martin, G. *et al* (1999), 'Controlling phantom image focus in a multi channel reproduction system', Presented at the 107th International convention of the AES, September 1999, New York, USA.

Mason, R. *et al.* (2000), 'Verbal and Nonverbal Elicitation Techniques in the Subjective Assessment of Spatial Sound Reproduction', Presented at the 109th Convention of the AES, September 2000, Los Angeles, USA.

Microsoft Corporation, (1990). Available from [www] http://www.microsoft.com, (Accessed 24/02/06).

Multi Media Projekt Verdi (2002), 'Design of the Listening Test'. Available from [www] http://www.stud.tu-ilmenau.de/~proverdi/indexen.html. (Accessed 12/05)

Nielsen, S. (1991), 'Depth Perception: Finding a Design Goal for Sound Reproduction systems', 90th Convention of the AES, 1991, Paris, France.

Plack, C. (n.d.), 'Perception of sound, speech and music': MSc Course material, University of Essex, available from [www] http://privatewww.essex.ac.uk /~cplack/courses/PS454_hbk.html (Accessed 12/05/06)

Pulkki, V. (n.d.) 'Compensating displacement of amplitude panned virtual sources', Presented at the AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio.

Robjohns, H. (2001), 'You are surrounded: Surround sound explained, part 1', February 2001, SOS Publications Group: Cambridge, available from [www] http://www.soundonsound.com (Accessed 11/05/06)

Schillebeeckx, P. *et al.* (2001) Using Matlab/Simulink as an implementation tool for Multi-Channel Surround Sound. Proceedings of the 19th International AES conference on Surround Sound, Schloss Elmau, Germany, 21 – 25 June.

Theile, G. (2001), 'Multi channel Natural Music Recording Based on Psychoacoustic Principles' presented at the AES 19th International Conference, May 2001.

West, J. (1998), 'Five channel panning laws: An analytical and experimental comparison', University of Miami: Florida.

Wiggins, B. *et al.* (2001), 'The analysis of multi-channel sound reproduction algorithms using HRTF data'. AES 19th International Surround Sound Convention, Germany.

Wiggins, B. (2004), 'An investigation into the real-time manipulation and control of three-dimensional sound fields', PhD Thesis, University of Derby.

Chapter 8 – Appendix 1

Appendix 1 – Example test sheet used.