**KTH Speech, Music
and Hearing**

Objective Localisation Measures in Ambisonic Surround- sound

Karin Carlsson

Supervisor: Dr. Damian Murphy
Approved:
Examiner: Anders Askenfelt

……………………………
(signature)

Abstract: When designing sound spatialisation systems, localisation accuracy is of great importance, i.e. how well the reproduction of the sound source corresponds to the original recording particularly with respect to source positioning. Ambisonic surround-sound is one such spatialisation system. In this work the two major psychoacoustic mechanisms Interaural Time Difference (ITD) and Interaural Level Difference (ILD) have been chosen in order to investigate whether they may constitute objective measures for verification of a horizontal Ambisonic surround-sound system. The two mechanisms are effective in different frequency ranges, where ITD is most effective in the low frequency domain up to approximately 700 Hz and ILD works best above 700 Hz. Three recording sessions were carried out for seventy-two different source positions, i.e., a new position every 5° in the horizontal plane, in order to capture data that rendered the possibility to compare the interaural time and level differences in relation to source positioning. The first session A, constituted recordings of monophonic Impulse Response signals using a, for the Ambisonic system characteristic Soundfield microphone. In session B, the sound field generated from the same monophonic source, was binaurally recorded using a KEMAR (Knowles Electronics Mannequin for Acoustics Research). In the third session, C, the signals recorded under session A were decoded and reproduced over a square Ambisonic loudspeaker array with KEMAR positioned in the middle, for a binaural recording. In obtaining the ITDs for each angle of incidence, measurements on the binaural signals from first peak (left ear) to first peak (right ear) were carried out. Concerning ILD, peak level measures in the time domain on left and right ear respectively were carried out and then subtracted in order to obtain the difference. A second more detailed approach, spectral ILD, was also carried out. In comparing the interaural time and level differences for session C with session B, results reveal how well the captured sound-field for every angle of incidence in session C corresponded to the sound-fields captured in session B. For the ITD, good agreement was found between the field from the original point source and the field reproduced over the Ambisonics loudspeaker array. Hence it could be concluded that the psychoacoustic mechanism ITD may constitute a reliable objective measure for verification of horizontal Ambisonic systems in general. For the ILD it was found that ILD for a point source is adequately reconstructed on the far side of the head, but that the level as a function of incident angle was practically constant in the side of the head that is nearer to the source. While the explanation is not clear, ILD can still act as a directional hearing cue over the Ambisonics system.

Sammanfattning: Vid utformning av ljudsystem med rumslig återgivning är riktigheten av ljudlokalisering av stor betydelse, d.v.s. hur väl reproduktionen av en ljudkälla överensstämmer med det inspelade originalet, med avseende på ljudets placering. Ambisonics surround-sound är ett sådant rumsbestämmande system. I detta arbete har de två viktigaste psykoakustiska mekanismerna Interaural Tidsskillnad (Interaural Time Difference - ITD) och Interaural Nivåskillnad (Interaural Level Difference - ILD) studerats med syfte att bestämma huruvida dessa utgör objektiva mått vid verifiering av ett Ambisonics surround-sound system i det horisontella planet. De två mekanismerna är verksamma i olika frekvensområden, där ITD är mest effektiv för låga frekvenser, upp till 700 Hz, varifrån ILD sedan tar vid. För att möjliggöra jämförelser mellan tids - och nivåskillnader i relation till källans position utfördes tre inspelningar med sjuttiotvå olika infallsvinklar av källan, d.v.s. en ny position var femte vinkel i horisontalplanet. Under den första inspelningen, A, spelades monofoniska impulssvar in med en ljudfältsmikrofon. Ljudfältet genererades från samma monofoniska källa under inspelning B och spelades in binauralt med hjälp av en KEMAR (Knowles Electronics Mannequin for Acoustic Research). De inspelade signalerna från inspelning A avkodades och återgavs över ett Ambisonicssystem där högtalaruppställning var uppställda i fyrkant, med KEMAR placerad i mitten för binaural upptagning, inspelning C. De interaurala tidsskillnaderna uppmättes för varje infallsvinkel från vänster öras första amplitudtopp till höger öras första amplitudtopp. Beträffande ITD utfördes toppnivåmätningar på amplituden i tidsdomänen för vänster respektive höger öra, där nivåerna för höger öra därefter subtraherades från vänster för att uppnå skillnaderna. Ett andra, mer detaljerat tillvägagångssätt, spektral ILD, utfördes också. Vid jämförandet av den interaurala tids- och nivåskillnaden för inspelning B och C visade hur väl det inspelade ljudfältet för varje infallsvinkel för inspelning C följde ljudfälten inspelade under inspelning B. God överensstämmelse kunde noteras mellan ljudfälten från den ursprungliga punktkällan och det över en Ambisonicsuppställning återgivna ljudfältet. Följaktligen kan ITD sägas utgöra ett pålitligt objektivt mått för verifiering av ett generellt Ambisonicssystem i det horisontella planet. Beträffande ILD visade resultaten att nivåskillnaderna för en punktkälla är fullgott rekonstruerade på den sida av huvudet som befinner sig längst bort från källan, men att nivån som en funktion av infallsvinkel, visade sig vara praktiskt taget konstant på den sida av huvudet som befann sig närmast källan. Då förklaringen till detta ännu inte klarlagts, kan ILD ändock utgöra en ledtråd beträffande ljudets position över ett Ambisonicssystem.

Table of contents

List of abbreviations

ITD	Interaural Time Difference
ILD	Interaural Level Difference
HRTF	Head Related Transfer Function
KEMAR	Knowles Electronics Mannequin for Acoustics Research
SDDS	Sony Dynamic Digital Sound
DTS	Digital Theatre System
NRDC	National Research and Development Council
DVD	Digital Versatile Disc
LFE	Low Frequency Effect
ITU	International Telecommunication Union
SREV1	Sampling Reverberator
TSP	Time Stretched Pulse
DSP	Digital Signal Processing
IACC	Inteaural Cross-Correlation

1. Introduction

Everyday life is full of three-dimensional sound experiences. Even though the visual sense is looked upon as being the most important sense for perceiving the surrounding world, the ability of humans to understand and interact with the environment is strongly connected to and dependent on spatial awareness and hearing capabilities [1][2]. Natural sounds are perceived in terms of localisation and contain cues in all three dimensions: width, height and depth [2]. When designing sound spatialisation systems, all three dimensions need to be accounted for, in order to achieve a sound image that is as accurate as possible. Here, the emphasis is on how well the reproduction of the sound source corresponds to the original recording with respect to source positioning. There are many mechanisms involved when humans perceive, process and interpret the surrounding sound. The more that is known about how humans listen within a three-dimension space, and what mechanisms there are that co-operate within the ear and the brain, the more successful implementations of spatial illusion such as surround sound systems may be developed. Surround sound system designs are based on facts about these co-operating mechanisms, and in particular on knowledge concerning psychoacoustic cues such as the interaural time difference (ITD) and the interaural level difference (ILD). In order to improve the design schemes of surround-sound systems that are in use today, studies are necessary that verify the accuracy of these systems, using both objective and subjective measures. This will be an ongoing process, until good correlation has been obtained between a particular surround-sound system and the real source. Ambisonics is one of many surround-sound systems and has been developed since the early 1970's. One of the interesting features of Ambisonics is that both two- and three-dimensional formats are possible. This report examines the accuracy of the Ambisonic system where recorded pantophonic (horizontal only) B-format signals constitutes the foundation using a objective measures for verification. Experiments have been conducted to study the reproduction accuracy of two selected localisation parameters, namely ITD and ILD.

2. Background

As a background to the description of the present project, the following sections briefly describe (1) spatial hearing, with emphasis on those psychoacoustic mechanisms that are the topic of the project; (2) the early development of pioneer spatial audio systems; and (3) the basic principles of the Ambisonics approach to spatial reproduction.

2.1 Spatial hearing – Psychoacoustic Theory

The ability of humans to hear spatially has been thoroughly described by Blauert [2]. In particular, he defines two concepts to assist the understanding of spatial hearing: (1) Localisation – the rule by which the location of an auditory event[1] is related to an attribute in a sound event[2]; and (2) Localisation Blur – the smallest change in a specific attribute of a sound event. By using these two definitions in experiments in both the median and the horizontal plane, information concerning localising both distance and direction of a sound source has been made known and constitutes the foundation to the key psychoacoustic theories that are in use today [2]. There are many factors involved in understanding a three-dimensional sound event. According to many authors on the field [1][2][3] some of the most important parameters that contribute to a listener's perception are:

---

[1] Auditory event – sounds perceived auditorily [2]
[2] Sound event – describes the physical aspect of the phenomena of hearing. Sound source, sound signal and sound wave all constitute attributes used to describe the physical phenomena that characterise a sound event [2].

- the listener's *acoustical environment*
- the influence of the listener's *torso and head* on the affecting wavefront respectively
- the influence of the listener's *outer ears* on the total auditory wavefront
- the *nature of the source*

Concerning the listener's *acoustical environment* where the actual sound event takes place, it may involve varying conditions, for instance standing on a heath listening to shrieking sounds of birds, i.e., a so called free field space, or attending a music event within an enclosed space. In a free field a sound generated by an omnidirectional[3] source is radiated away without being reflected back. In consequence, the sound level experienced by a listener or registered by a sound pressure registration device, drops off rather quickly (~ 6 dB for every doubling of the distance). This drop off in level is caused when the sound energy is distributed over a sphere of ever-increasing surface area as it expands away from the source. The level will drop rapidly per unit of distance in the near field and more gradually in the far field, with increasing distance to the source.

A free field may be created artificially within an enclosed space, i.e., an anechoic room. Such a room or chamber has absorbing material on the walls, ceiling and floor, which prevents reflected sounds from interfering with the captured sound signals. In ordinary enclosed spaces, i.e., where little absorbing material is present, a certain amount of the radiated sound energy from sound sources is absorbed both by the air within the space and by the surfaces (walls, floor, ceiling, furniture, bodies (humans or animals) etc.). The non-absorbed energy is reflected back into the space again and continues to be either absorbed by the surfaces or reflected back and creates after a short period of time a dense and diffuse sound-field that eventually will decay into silence. As a consequence the sound level does not drop off as rapidly as in a free field since the reflected sound builds up and creates an unchanging level of diffuse sound [2]. The direct sound still tails off with distance in the same way as in a free field and eventually the reflective sound takes over [2]. According to Rumsey [1], reflections have the effect of modifying the perceived nature of discrete sound sources, where the early reflections strongly contribute to the listener's sense of the size and space of a room. The later reflections contribute more to a sense of spaciousness [1]. Research concerning the localisation of sound in rooms has so far suggested that reverberation improves the perception of distance but slightly degrades the perception of direction [4]. The fact that reverberation does have a slight impact on localisation indicates the need for using artificially made free fields i.e., anechoic chambers for best results when collecting data for certain measures of sound. In a reflecting space however it is found to be easier to detect cues that describe the distance than it is in a free field since the proportion of reflected to direct sound provides information of the sound source's location of distance [1]. According to Blauert [2] our knowledge about distance hearing is deficient in comparison to what is known about directional hearing and since it is considered being a rather complex subject the approach of this work will mainly focus on localising the direction of a sound source. Nevertheless, for completeness, the findings about localisation concerning distance will be discussed later in this section.

The influence of the *torso and head* on sound signals is important to our auditory interpretation of the surrounding three-dimensional space. The reflections from the body and

---

[3] Omnidirectional – radiates in all directions

shoulder and most importantly the reflections and resonance caused by the shape of the pinna, the **outer ear,** give rise to changes in the spectrum of the sound at the eardrums and help us to judge the location of a sound source [1][5][6]. Acoustically the pinna codes spatial attributes of the soundfield into temporal and spectral attributes and because of its whirly shape, incoming sound signals are distorted linearly and differently depending on their direction and distance. The acoustic effect of the pinna is based on physical phenomena such as reflection, shadowing, dispersion, diffraction, interference and resonance [1][2]. Since it is the spectral patterning of the sound at the eardrum that is of importance, the information provided by the pinna is most effective when the sound has spectral energy over a wide frequency range [6]. It has been shown that high frequencies above 5 kHz are particularly important, since only short wavelengths of sounds are sufficiently short to interact strongly with the pinna [6], but lower frequencies also contribute since their level at the eardrums is affected by the torso and head [6]. The disturbance of the sound-field due to the head has a great impact on the sound signals in the pinna and in the ear canal, since the head presents a considerable obstacle to the free propagation of sound [2]. According to Blauert [2] findings on measuring the impact the head has on the level differences between the two ears, suggest that the head may have an amplifying effect rather than an attenuating on the sound pressure level. This is due to the so-called head baffle effect, which refers to the comparative augmentation of high-frequency sound, caused by the acoustic diffraction of low-frequency sound by the head and pinna. The fact that we have an ear on each side of the head causes differences in the received signals at the eardrums depending on the nature of the source and the conflicting environmental cues [1]. From research on directional sound perception [1][2][6] it may be concluded that there are two primary mechanisms at work when humans "decode" or interpret a three-dimensional environment, namely ITD and ILD. This pair of mechanisms is often referred to as the Duplex Theory and will be discussed separately since they operate over different frequency ranges. According to Rumsey [1] and Moore [6] there exist monaural mechanisms as well, but those will not be *dwelled* upon in this work.

An interaural time difference, ITD, is obtained when a sound wave originating off the centre front arrives at different times to the two eardrums of the listener [1][2][6]. The extra time the sound waves have to travel in order to reach the ear furthest away is related to the angle of incidence of the incoming sound [1][2][6]. Figure 2.1 shows the principle of ITD and equation (1) how the ITD can be calculated from the path difference between the two ears.
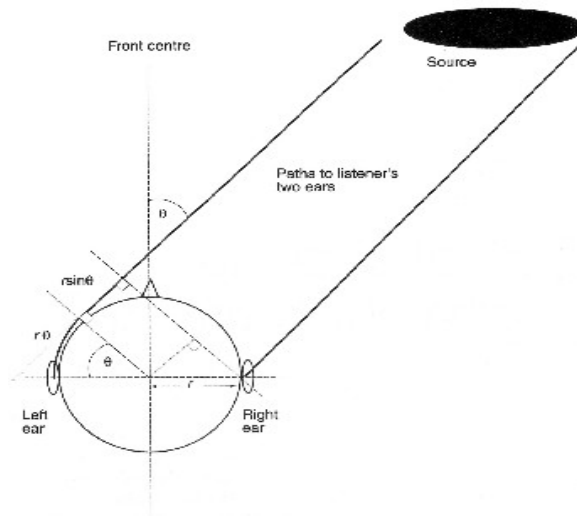
Figure 2.1 *The principle of ITD*

$$ITD = r \ (\theta + \sin\theta)/c \qquad\qquad\qquad (1)$$

r = half the distance between the two ears
θ constitutes the angle of incidence in radians
c = 344 m/s (20° C)

The ITD is based on the low frequency content up to approximately 700 Hz and is registered particularly at the starts and ends of the sound [1]. According to Blauert [2] this method enables humans to resolve the direction down to 1°. Furthermore, for sinusoidals a time difference is equivalent to a phase difference between the two ears [1][6][7] that depends on the frequency and the location of the source, since the distance between the ears is constant. For low-frequency tones the phase shift provides effective and unambiguous information about the location of the sound, but with increasing frequency, ambiguities start to occur when the period of time is approximately twice the possible ITD, i.e., when the period is about 1380µs and the frequency is 725 Hz [6]. A sinusoid of this frequency for an angle of incidence of 90° produces waveforms at the two ears which are in opposite phase [6], giving rise to an ambiguous localisation of the sound source since the wave form at the right ear might be either half-cycle behind or half-cycle ahead that at the left ear [6]. Moving the head solves the ambiguity, which is why there does not exist an abrupt upper limit for using the phase differences between the two ears [6]. For frequencies above 1.5 kHz, however, phase differences become very ambiguous [6]. As a result, high-pitched pure tones are considered harder to localise than complex waveforms since the interaural phase does not convey any information [3]. Luckily everyday real-world sounds are to a great extent timbral or broadband rather than sinusoids. They contain non-ambiguous transients, where the higher frequency content in the complex signals helps us localise the sound [6]. The fact that the size of the head constitutes a barrier to higher frequencies forms the basis for the second primary mechanism, the interaural level difference, ILD. High frequency sounds must diffract around the head in order to reach the eardrum furthest away [2][6]. This might best be described as an "acoustical shadowing" effect on one side of the head, were long wavelengths (low frequencies) are unaffected by the head but short wavelengths (high

9

frequencies) are shaded by the head, implying a "loss" of high frequencies at the distant ear, see figure 2.2, a and b. As a consequence it will be quieter at the "shadowed" ear, which enables the brain to localise the sound to the side where no shadow is taking place [5].
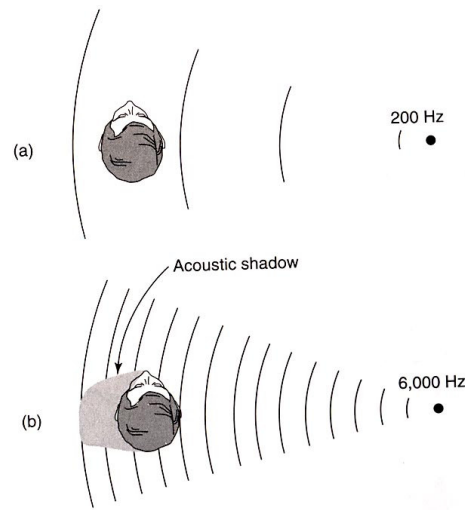

Figure 2.2 *The principle of ILD*

For sources that are very close to the head the difference in distance will also contribute to the ILD, but normally this effect is negligible.

Putting it all together, the unusual shape of the pinna, reflections from shoulders and body and the fact that the extra distance travelled to the eardrum furthest away give rise to level differences, all contribute to a modified spectrum of the incoming sound wave at the eardrums, which differs from the spectrum of the waveform of the original source. By measuring the spectrum of the sound reaching the eardrum (left or right ear respectively) and the spectrum of the sound source, and forming the ratio of the two spectra (in dB), one arrives at the Head-Related Transfer Function, HRTF. This function varies systematically with direction of the sound source relative to the head and is unique for every direction, since for every angle of incidence of a sound wave, the spectrum that reaches the eardrum will be different [1][4][6]. The shape of the pinna differs between individuals implying that HRTFs do too, which make it difficult to generalise the spectral characteristics [1][6]. HRTFs do not play such a significant role in the horizontal plane as in the vertical plane where the shape of the pinna, in particular, is of greater importance [1][2] (for further reading see Blauert [2], Moore [6]).

The ease of localising the direction to an omnidirectional source depends on the ***nature of the source*** (i.e. amplitude, frequency, complexity). Most sound sources have a directivity pattern of their own that represents the deviation from omnidirectional radiation at different frequencies [1]. The deviation difference is sometimes expressed as a number of dB gain compared with the omnidirectional radiation at different frequencies. According to Rumsey [1] sources tend to radiate more directionally with higher frequencies and more omnidirectionally at decreasing frequencies.

2.2 Surround-sound Systems

Before discussing surround-sound systems used today in more depth, a summary of the main contributions and most important developments in spatial audio systems throughout history will be summarised. According to Rumsey [1]:

- Clement Adler carried out the first known example of stereophonic transmission of music at the Paris exhibition 1881. By placing spaced telephone pickups (a substitute for microphones) in the footlights at the Paris Opera and then relaying the outputs of the telephone receiver at the exhibition, visitors could experience the opera "live".

- In the 1930's, both Bell Labs and Alan Blumlein independently of each other, developed stereophonic principles, where loudspeakers were used to create phantom images. Using totally different approaches, Snow and Steinberg, pioneers at Bell Labs, developed a three-channel system for large auditorium sound reproduction, meanwhile Blumlein by introducing only amplitude difference between a pair of loudspeakers showed that it would be possible to create phase difference between the ears. His system was mainly designed for source imaging over a limited angle and henceforth more suitable for domestic use.

- Early consumer stereo used a method similar to Blumlein's in the 1950's, where monophonic recordings were artificially reprocessed using techniques such as comb filtering and band-splitting techniques so as to create a stereophonic effect.

- Clark, Dutton and Vanderlyn of EMI revived Blumlein's theories in 1958 and showed in more detail how a two-loudspeaker system might be used in creating an accurate relationship between the original position of the source and the perceived position in the reproduction by controlling only the relative signal amplitudes between the loudspeakers. They also stated that the Blumlein concept did not take advantage of all the mechanisms of binaural hearing, especially the precedence effect, but used only a few to recreate the directional cues that exist in natural listening.

- Since the 1920's, the interest in encoding all of the spatial cues received by human listeners has been motivated by the success of the binaural method of recording stereo sound. In binaural stereo, two microphones are placed in the openings of the ear canal or further in, close to the eardrum, of a real person or on a dummy head - a KEMAR system (Knowles Electronics Mannequin for Acoustics Research). The term binaural stereo usually describes sound events that have been recorded to represent the amplitude and timing characteristics of the sound pressures that are present at the human eardrums.

- Multi-channel stereo formats for cinema became very popular in the late 50's and 1960's using three front channels and a surround channel. In the early 1970's Dolby Stereo was introduced which enabled a four-channel surround-sound signal to be matrix encoded into two optical sound tracks recorded on 35-mm film. This format is still in use, but gradually, modern cinema sound is moving towards digital sound tracks that incorporate either five or seven discrete channels of surround-sound plus a sub-bass effect. Low-bit-rate coding schemes used to deliver surround sound signals with movie film are for instance Dolby Digital, Sony SDDS, and Digital Theatre System (DTS).

- Quadraphonic – a surround sound system with a variety of competing encoding

methods where four channels of surround sound were conveyed in two analogue media. The system was configured for a square arrangement of loudspeakers. The lack of compatibility with two-channel reproduction gave a poor front image often with a hole in the middle. This format never gained commercial interest.

- Ambisonics – developed in the 1970's by a number of people (primarily Michael Gerzon and P B Fellgett) with the intention to capture the total sound field and then reproduce it with all its spatial cues intact over a minimum of four loudspeakers. It is partly based on an extension of the Blumlein principle to more channels. There are few recordings available in this format. Much of the work was supported by the NRDC (National Research and Development Council). The Ambisonic system has great potential, at least in theory, but has not yet been launched commercially because of several technical reasons.

- Since people have become used to surround-sound at the cinema alongside the development of new consumer audio formats such as DVD, Dolby Digital and DTS, the concept of installing an audio system at home that resembles the cinema audio system, for instance the 5.1-channel surround system, has gained popularity. This public interest (in the western world) shows that installing hi-fi equipment may benefit other media industries like the music business.

Ever since Clement Adler carried out the first known stereo transmission of music at the Paris Exhibition 1881, improvements on reproducing the recorded sound-field have been achieved over the years. Today we are exposed to many different systems for transmission of music, where the 5.1-channel surround constitutes the most popular system for both the public and the home entertainment audience. Since this thesis is concerned with the Ambisonic surround-sound system, an introduction to what "surround-sound" is all about and to the different systems that are in use will be given before thoroughly investigating the Ambisonics surround-sound system.

2.2.1 5.1 – channel surround (3-2 stereo)

Even though a stereo illusion may be very rewarding if well executed, only a small part of a bigger audience in front of a TV set may enjoy the sound experience. Surround-sound systems are designed to improve the spatial representation of the reproduced sound sources [7]. One of the most popular surround-sound systems is the 3-2 configuration (three front channels and two rear/side channels), also known as 5.1-channel surround. Originally 5.1-channel surround was developed for cinemas but has been standardised for home entertainment (television) as well. The 5.1-channel surround system consists of a front-orientated sound stage based on three-channel stereo accompanied by two rear/side channels that generate supporting ambience or spaciousness, where the front left and right channels preserve positional compatibility with two-channel stereo [1]. The .1 component indicates the Low Frequency Effect (LFE) channel, which is a separate sub-bass channel with a narrow bandwidth from 20 Hz up to 120 Hz. The 5.1 loudspeaker layout and channel configuration is specified in the ITU 1993 standard, which gives recommendations on how to best arrange the loudspeakers in relation to the position of the listener/s, but no definitions on how sound signals are represented or coded for [1]. A possible loudspeaker configuration for a 5.1-channel surround system can be shown in figure 2.3, where the LFE loudspeaker (the subwoofer) is situated behind the listener instead of at the front that is otherwise common.
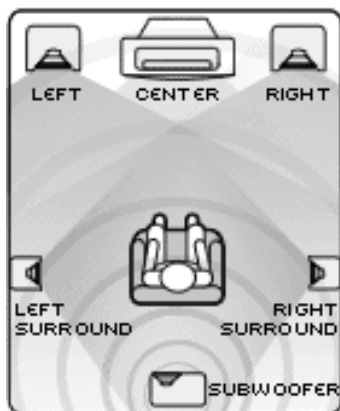
Figure 2.3 *5.1-channel surround*
for Home Entertainment System

According to the ITU standard the front left and right loudspeakers should both be mounted at an angle of 30° off centre front as in two-channel stereo and the two surround loudspeakers at the rear/side should be positioned at approximately ±110° off centre and be intended for generating supporting ambience and effects. It may be noted that, since the 5.1-channel surround already is a standard closely associated with the moving image media, i.e., television, cinema, games etc., and since the human visual sense is considered to be much superior in comparison with the auditory system, it is difficult to settle on the desired degree of accuracy for a 5.1-channel surround system, when recording other sound events, such as music performances [8]. As there are no given methods of sound-field representation or spatial coding for 5.1 there are a variety of recording techniques in use. All typical microphone recording techniques aim at creating as stable side images as possible, which is difficult to achieve since 5.1 does not completely support an accurate 360º phantom image [1]. Having personally experienced 5.1-channel surround, recorded for radio theatre within a 5.1 loudspeaker configuration at Sveriges Radio – Radio Sweden in Stockholm [9], it is easily stated that even though the performance was very well produced using all imaginable 5.1-channel surround microphone and panning technique possible, it did lack authenticity.

2.3 Ambisonics – "Surround-sound"

Where the 5.1-channel surround configuration fails to create a 360º -image location the Ambisonics system succeeds to a greater extent. Ambisonics is a system based on some key principles of psychoacoustics, ITD and ILD, and offers a complete hierarchical approach to directional sound pickup, storage or transmission and reproduction, equally applicable to mono, stereo, horizontal (pantophonic) surround-sound and full sphere (periphonic) reproduction [1]. The main idea of Ambisonics is to record the information of a sound-field and then reproduce the sound-field under the impression of hearing a true three-dimensional sound image over a number of equally spaced loudspeakers [10]. The ultimate goal of the system is to "transport" the listener to the actual recording venue instead of experiencing the music event in the less spacious "living room". This implies that the playback space, where the reproduced signal is presented, might be perceived as being larger or smaller depending on in what environment the recording took place [5].

2.3.1 The Soundfield Microphone and the formats of the Ambisonics system

The Ambisonic system builds on the early and original ideas of Blumlein who developed a technique consisted of a pair of microphones with figure-of-eight characteristics [1]. By adding an omnidirectional microphone to the orthogonally positioned pair of figure-of-eight units, the sound-field in the horizontal plane at a particular point may be captured, provided that the microphones are truly coincident [1]. Extending the principle to the third dimension does not pose any problems. Adding a third figure-of-eight microphone, orthogonally and vertically, to the pair of figure-of-eight units and to the omnidirectional microphone has been shown to capture adequately all the information needed [5]. It is difficult to achieve true coincidence using four separate microphones. Instead, a Soundfield microphone has been developed and is in use today for sound-field recordings [11], see figure 2.4.
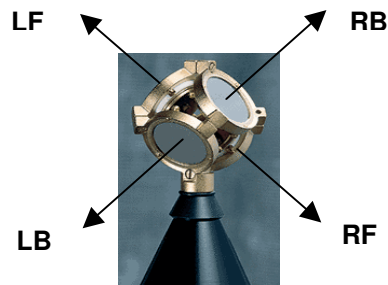

Figure 2.4 *The Soundfield Microphone*

The Soundfield microphone contains four capsules mounted in a tetrahedral array. In addition to the above mentioned microphone configuration (one omni plus three figure-of-eight units), it can provide a variety of conventional polar patterns by combining the signals from the individual capsules in different ways [11]. A Soundfield microphone system supplies four different signal formats, where A-format is for microphone pickup, B-format for studio equipment and processing, C-format for transmission and D-format for decoding and reproduction. There also exists a UHJ format (Universal HJ – letters denoting two earlier surround sounds) that is used when encoding Ambisonic information into two or three channels in order to retain good mono and stereo for non surround-sound listeners [1]. A sixth format developed by Gerzon and Barton [8] has been added to the list, the G-format, which is a technique that offers originally recorded Ambisonic B-format to be decoded into standard ITU 5.1-channel surround [10].

The A-format consists of four mono signals, the left front (LF), right front (RF), left back (LB) and right back (RB) which come directly from the four capsules mounted on the Soundfield microphone (see figure 1.4). Each signal is equalised to compensate for the inter-capsule spacing in order to create an output that truly represents the sound-field at the centre point [2]. The B-format is then derived from A-format using the sum and difference techniques (2):

$$X = 0.5((LF-LB) + (RF-RB))$$
$$Y = 0.5((LF-RB) - (RF-LB)) \qquad (2)$$
$$Z = 0.5((LF-LB) + (RF-RF))$$
$$W = 0.5((LF+LB) + (RF+RB))$$

The B-format signals together contain the complete 3D information for the sound pressure and the sound intensity vector at the centre of the microphone. In figure 2.5 a visual representation of the signals in all three dimensions can be seen.
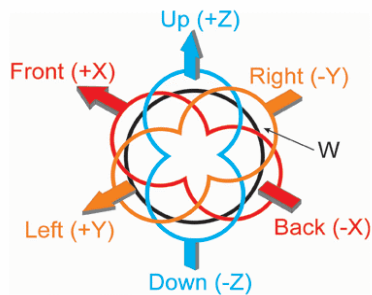

Figure 2.5 *B-format input components* W, X, Y *and* Z

2.3.2 Encoding

It is possible to encode a monophonic signal into B-format, such that the sound source will appear to have a given direction in 3D space when the B-format is reproduced over any appropriate loudspeaker configuration. The sounds are looked upon as being conceptually placed either on the surface or within a "unit" sphere [10] with the maximum radius of the sound-field as 1. Sounds outside of the sphere will not be properly decoded (*decoding* - see below) and the sounds tend to pull to the nearest loudspeaker [10]. In order to avoid this, the sound source coordinates must follow the equation (3):

$$(x^2 + y^2 + z^2) \leq 1 \tag{3}$$

where x is the distance along the X front-back axis, y the distance along the Y left-right axis and z the distance along the Z up-down axis [10] (see figure 1.5). The four B-format signals can then be derived by multiplying the monophonic signal with the polar coordinates for the desired direction, see equations (4), where θ is the angle of incidence in the horizontal plane and Φ the angle of incidence in the vertical [10]. The angles in parenthesis represent the polar patterns in horizontal only B-format. The resultant B-format encoding equations are shown in equations (5).

$$
\begin{aligned}
x &= \cos\theta\sin\Phi \quad (\cos\theta)\\
y &= \sin\theta\cos\Phi \quad (\sin\theta)\\
z &= \sin\Phi \qquad\quad (-)\\
w &= 0.707 \qquad\;\; \text{(the same).}
\end{aligned}
\tag{4}
$$

$$
\begin{aligned}
X &= \text{input } (\cos\theta\sin\Phi)\\
Y &= \text{input } (\sin\theta\cos\Phi)\\
Z &= \text{input } (\sin\Phi)\\
W &= \text{input } (0{,}707)
\end{aligned}
\tag{5}
$$

Even though figure 2.6 does not represent a sphere it does show how the polar patterns can be derived from both three dimensional and two-dimensional co-ordinate systems.
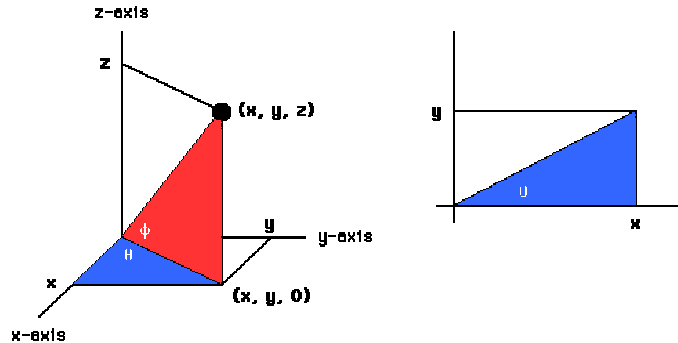
Figure 2.6 *Schematic plots over how polar patterns are derived both for full-sphere and horizontal only B-format encoding of a monophonic signal.*

The four (or three for horizontal recording) B-format signals represent between them the pressure and velocity components of the sound-field in any direction [1] and from these four signals any kind of directional related information and can be derived. This implies that whatever loudspeaker system will eventually be used does not need to be considered during the recording session [1][10].

2.3.3 Decoding

One of Ambisonic's great advantages is the two-part technological solution of first encoding the directions (and amplitudes) of sound signals into a specific format, B-format and then to decode them for reproduction over a loudspeaker array using an appropriate decoder. According to Gerzon [12] the aim of any surround-sound decoder is to provide the listener with a convincing illusion of sounds from all directions [12][13]. In order to optimise the accuracy of the Ambisonic experience, knowledge of some key psychoacoustic theories cast in a mathematical form has allowed decoder design to improve over the years [13]. According to Gerzon [13], a problem when designing decoders for surround-sound systems such as Ambisonics is the fact that the ears localise sounds by different methods depending on frequency, one at low frequencies below 700 Hz (ITD), and one at high frequencies above 700 Hz (ILD). Many different decoder designs have been independently developed throughout the years [14]. According to Gerzon [12], for low frequencies below 700 Hz, there are three important aspects of sound localisation when designing decoders. They are the "Makita" direction – the direction one turns to face the apparent sound source, the "velocity magnitude" – the degree to which the sound stays in its correct localisation as one turns to other direction and the "phasiness" – the degree to which unwanted components of sound, not in phase with the desired sound, are heard [12][15][16]. For high frequencies above 700 Hz, the theory of the 'Energy vector' localisation is applied and implies the direction the head has to face in order that there be no interaural amplitude difference [15][17]. In order to understand how these vector theories work one may, according to Gerzon [15], look at the amount of sound emerging from the loudspeakers being represented as vectors with a certain length pointing from the centre of the loudspeaker towards the middle of the speaker array. At low frequencies this amount of sound is the amplitude gain of the sound and for high frequencies the energy gain of the sound in each speaker. In adding the *magnitude* of the vectors, the total *amount* of the sound in the middle is derived. When adding *all the vectors*, one for each loudspeaker, the *direction of the resultant vector* is derived and is by Gerzon [15] introduced as the Makita localisation vector at low frequencies or the Energy vector localisation at high frequencies (see figure 2.7).
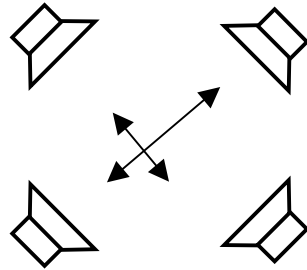
Figure 2.7 *The idea of vectors*

The image will be stable under rotation of the head only if the magnitude of the resultant vector is exactly the same amount as the total sound from the loudspeakers. The ratio of length between the resultant vector length and the total amount of sound is called the vector magnitude of the sound and should ideally equal one [15]. In a good decoder design the Makita and energy vector localisation should coincide, which implies that the image will be stable under head rotation only if the magnitude of the resultant vector is precisely the same as the total amount of sound from the loudspeakers [15][17]. Good decoder design consists of getting both the Makita and the energy vector localization correct for all sound directions at all frequencies, aiming at getting the low frequency vector magnitude equal to one and the energy vector as close as possible to one at high frequencies [15]. To achieve this, *shelf filters* are used. In figure 2.8 a schematic diagram of the decoding process is shown where the incoming signals to the left constitute encoded B-format signals that get modified by the shelf filters and then matrixed and derived to output signals, i.e., D-format signals, which feeds the loudspeakers [15][16].
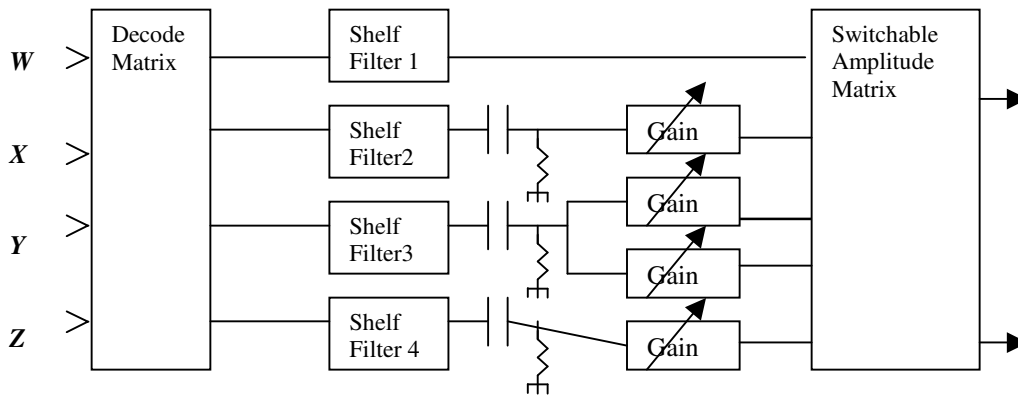

Figure 2.8 *Decoding process, full-sphere*

The shelf filters modify the B-format signals according to (6):

$$W'' = k_1 W \quad X'' = k_2 X \quad Y'' = k_2 Y \quad Z'' = k_2 Z \hspace{3cm} (6)$$

where k_1 for low frequencies is 1 and for high frequencies 1,4142 and k_2 for low frequencies is 1,732 and for high frequencies 1,4142. k_1 and k_2 constitutes shelf filter gains in dB. This information depends on what input signals to decode and is different depending on full-sphere or horizontal only decoding [18]. The function of the shelf filters is to modify the low frequency vector magnitude as the frequency increases so that the energy vector becomes

17

optimal at higher frequencies [15]. The amplitude matrix has to be adjusted to the shape of the loudspeaker layout [15]. In order to coincide the Makita and energy vector localisation, Gerzon [15] suggests that all loudspeakers should be located at the same distance from the centre of the loudspeaker layout, be placed in diametrically opposite pairs and that the sum of the two signals fed to each diametric pair is the same for all diametric pairs. It is important to choose a loudspeaker layout that may distribute the actual value of the energy vector magnitude in different directions that optimise the overall subjective results [15]. There are several loudspeaker layouts available depending on dimension and order of the system. For first order horizontal only Ambisonics, the minimum of loudspeakers to use is four, positioned either in a square or in a rectangle [10][15]. The loudspeakers are then fed with decoded B-format signals (D-format) each of which contain virtually all the elements of the recording, but with different relationships. The loudspeakers then work together to recreate the acoustics and the ambience of the original recording [8]. For a square loudspeaker array each individual loudspeaker is fed a combination of B-format signals corresponding to its position with respect to the centre of the array, see figure 2.9.
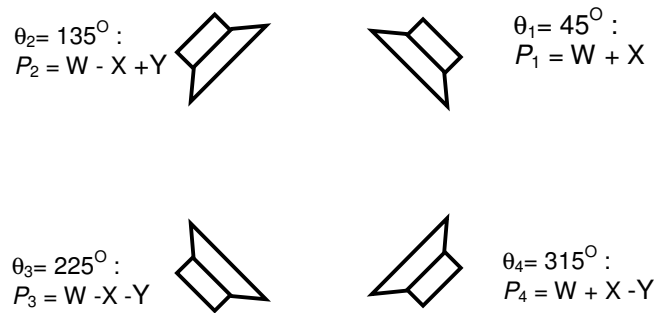

$\theta_2 = 135^{O}$:
$P_2 = W - X + Y$

$\theta_1 = 45^{O}$:
$P_1 = W + X$

$\theta_3 = 225^{O}$:
$P_3 = W - X - Y$

$\theta_4 = 315^{O}$:
$P_4 = W + X - Y$

Figure 2.9 *Decoding Equations for each loudspeaker, where 45° constitutes left front, 315° right front, 135° left back and 225° right back respectively.*

Infinitely many loudspeakers can be added as long as they are all evenly spread on an imagined circle around the centre point (the so-called sweet spot). For three-dimension, a cube with eight loudspeakers as a minimum is required in order to include the width-height dimension as well [10]. Each individual loudspeaker is then fed a weighted combination of B-format signals corresponding to the loudspeaker position in relation to the centre of the array [10].

So far only first order Ambisonics have been discussed, but second and higher order Ambisonics also exist. In incorporating additional directional components (see equations (7)) to the above mentioned polar pattern equations (4), an improved directional encoding may be derived which covers a larger listening area than first order Ambisonics [1] provided that an appropriate decoder is implemented that can deal with the second order components.

$$U = \sqrt{2} \cos(2\varphi)$$
$$V = \sqrt{2} \sin(2\varphi) \qquad\qquad (7)$$

Unfortunately there does not exist a second order (or higher for that matter) Ambisonic microphone system that produces the required polar pattern [1]. This restricts the system to use components above first order to be synthesised signals [1][19].

Furthermore Ambisonics may be looked upon as a holographic sound reproduction system. A related method is based on the Kirchhoff-Helmholtz integral, which relates to the pressure at a point inside a volume of space to the pressure and velocity on the boundary of the surface [20]. Where Ambisonics offer reconstruction of sound waves at a point, Holophonics[4] offer reconstruction over an area. Combining the two would imply a near full fidelity central for the Ambisonics system together with the enhanced area of coherency typical for a volume reconstruction solution [21]. It is most plausible that a full three-dimensional volume reconstruction solution will become technologically feasible in the near future for either home systems or for research [21].

3. Aim of the project

In order to evaluate the accuracy of existing surround-sound systems it is of great importance to find descriptive localisation parameters. Knowledge of how we humans derive and use the psychoacoustic cues within an experienced sound-field has played a crucial role concerning the design of the sound systems that are in use today. This holds for the Ambisonic system as well, since it is a system that has developed from knowledge on both of the two most important psychoacoustic cues, ILD and ITD.

This project aims at investigating objective localisation cues within a sound-field which define the success or otherwise of the horizontal-only Ambisonic surround-sound system. An objective approach without listening tests has the advantage that one does not need to involve humans. Still in this context localisation parameters that in particular help humans decide the direction of a sound are of great interest since parameters deciding direction of a sound source are easier to detect and extract than parameters concerning distance that are more complex. Therefore the main approach of this work will be based on localisation parameters that define direction. This approach will outline whether a first order Ambisonic set up correlates well to the real source using ITD and ILD as possible localisation parameters for verification.

---

[4] Holophony – this is an acoustical equivalent to holography and is the only solution that ensures a perfectly accurate reproduction, but its implementation is rather complicated. It is derived from the Huygens' Principle, which conveys the idea that an acoustical field within a volume can be expressed as an integral.

4. Method

It has been demonstrated that the psychoacoustic mechanisms ITD and ILD are the most important when localising direction of a sound event. Therefore they might together constitute appropriate measures when verifying the accuracy of a reproduced sound-field.

4.1 Test set-up

In order to measure the above stated localisation parameters, three recording sessions A, B and C were carried out. The four[*] B-format signals of a real monophonic sound source were captured using a Soundfield microphone, session A. In session B, the KEMAR system [22] was used instead of the Soundfield microphone. A KEMAR system records a sound-field binaurally with two microphones, each at the opening of the ear canal on the head of the mannequin. The system offers two different sizes of ears where the larger of the two were chosen since they corresponded best to the ears of the author. In the third session, C, the KEMAR system was placed in the middle of a square Ambisonic loudspeaker array, in order to record the reproduced decoded B-format signals that were captured in session A.

4.1.1 Test Environment

All sessions took place in a semi-anechoic chamber with a concrete floor at the Department of Electronics at the University of York. This free-field simulated environment was chosen in order to prevent first reflections from interfering with the recorded signals. In order to further reduce reflections from the floor, absorbing material, here four big mattresses were spread out during the recordings. The chamber measured 3.47 m × 3.47 m from absorbing peak to absorbing peak, and 2.73 m from absorbing peak to concrete floor. In placing additional equipment behind a temporary shelter made of absorbing material, noise and reflections were further prevented to interfere. Figure 4.1 shows the test environment and the shelter.
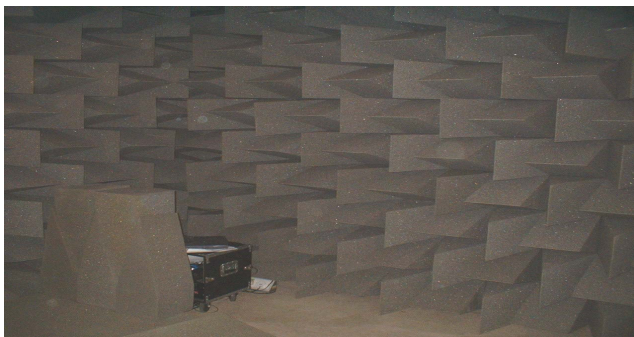

Figure 4.1 *The semi-anechoic Chamber at the University of York*

4.1.2 Angles of Incidence – Number of Microphone Positions

Humans are good at localising a sound source down to a difference of 1° of angle of incidence. In order to keep the number of positions within a reasonable and manageable range, a displacement in the horizontal plane of every 5° was chosen, resulting in seventy-two microphone positions. According to previous work on spatial resolution [1][2] a displacement

---

[*] Even though the Soundfield microphone provides the possibility to exclude the height dimension – z, it was retained, since four signals may be of use if this project will be extended to involve localisation parameters for full-sphere B-format in the future.

every 5º will be sufficient when deriving interaural differences. Since the recordings took place in a semi-anechoic chamber, the placement of the loudspeaker in session A and B was such as to try to prevent the first reflections from interfering with the direct sound. This led to the solution of placing the loudspeaker on its stand in one of the corners of the room. The microphones for session A and B were placed at a distance of 1,4 m and 1,5 m above ground. In a previous and similar investigation [3], a distance of 1,4 m has been used and proven to give good results. The aim is to prevent the sound from the loudspeaker to reflect on the KEMAR and back again, which otherwise would have an impact on the recorded signal. In order to capture the seventy-two source positions, the microphones were manually rotated clockwise. Time did not permit constructing a device for remote manoeuvring. The KEMAR system provided a stand with marks every 10º to which marks were added in between in order to obtain an as accurate location of the KEMAR as possible. For the Soundfield microphone, a disc made of hard paper with marks every 5º was mounted on the microphone. The KEMAR system and the Soundfield microphone can be seen in figures 4.2 and 4.3 respectively.
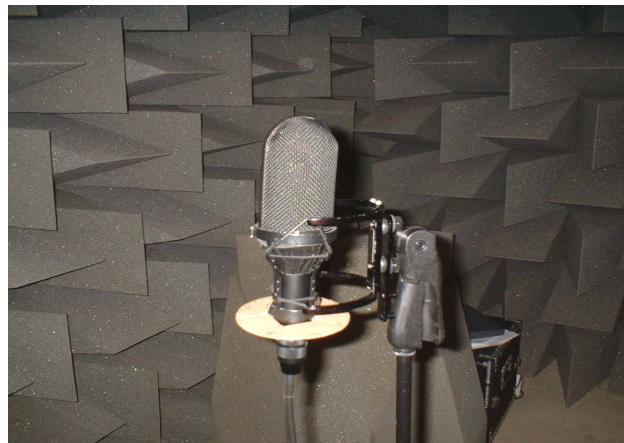

Figure 4.2 *KEMAR*

Figure 4.3 *The Soundfield Microphone with the hard paper disc*

4.1.3 Generating the Sound Source using the Yamaha SREV1 Sampling Reverberator

When investigating subjective psychoacoustic cues within a real or reproduced sound-field, sounds that are as authentic as possible for that particular surround-sound system should be used in order to develop the most accurate design schemes [23]. For this work no such requirement was included, since the method was entirely objective. The SREV1 Sampling Reverberator system was used in order to generate a monophonic sound source. See figure 4.4 and 4.5.
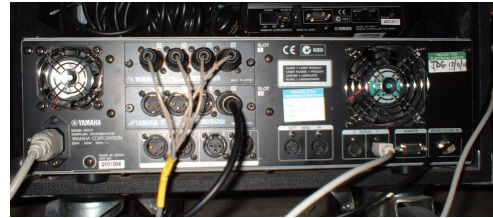
Figure 4.4 SREV1 front

Figure 4.5 SREV1 back.

This system can sample the reverberant characteristics of an acoustical space and apply them to any audio signal [24]. Sampling a sound-field involves the recording of the impulse response between the audio source that is sampled and a pickup point. Sound-field sampling consists of firing SREV1 test pulses into an acoustic space where microphones pick up the generated signal that returns to the SREV1 for processing. The data can then be saved onto a PC Card, edited and later be used to create reverb programmes. For this project the data were only saved onto a PC card. The method used here, to impose the characteristics of one signal onto another, is called convolution, where the SREV1 convolves the reverberation characteristic of a previously sampled acoustic space onto another audio source [24]. This results in the same overall sound that would have been heard if the audio source had been perceived in that actual acoustic space. Just as frequency shows how an audio circuit responds to a range of frequencies, "an impulse-response" shows how an acoustic space responds to impulse data [24].

The SREV1 can generate two types of test signals: an impulse signal, and a Time Stretched Pulse (TSP) where impulse signals are ideal for sound-field sampling since they have a very short play-back time and a flat frequency response at all frequencies. For this project the TSP method, also called the Swept Sine Method, was used [25] and has the same flat response as the impulse signal. It differs concerning frequency and it contains a sweep of all frequencies, providing a relatively high sound pressure level. The higher the level of the signals the better S/N ratio might be created which otherwise might be of a problem when using an impulse signal [25]. Figure 4.6 shows a schematic wiring diagram of how SREV1 functions and what equipment is needed. For this project no external D/A and A/D converters were needed and further no power amplifier to the loudspeaker since the loudspeaker had its own internal amplifier.
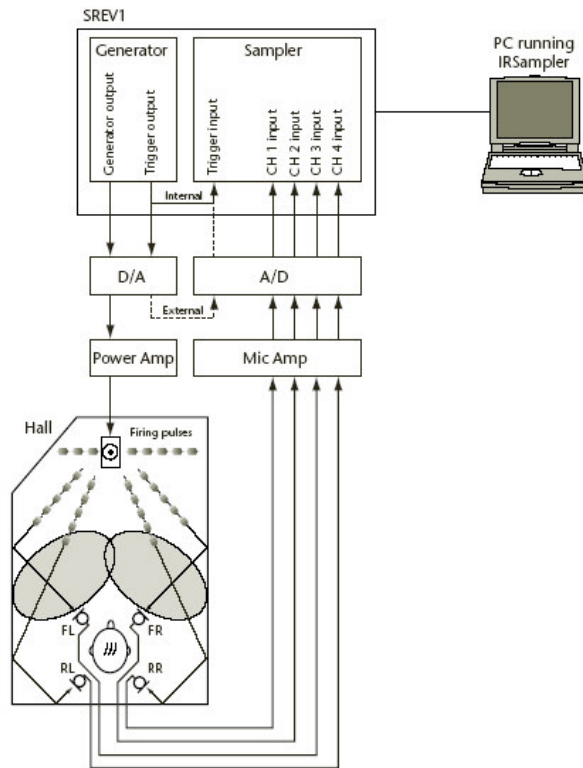
Figure 4.6 *SREV1 Wiring Diagram*

In order to generate a test signal using the TSP method, the sampler was configured according to the detailed instructions in the Yamaha SREV1 Sampling Guide. The generator source was set to 'TSP64k'. The loop interval had to be long enough so that the reverberation energized by each pulse had time to fade away properly before the next pulse was fired [25]. For this project the loop interval was chosen to 3000 ms (i.e., 144 000 samples) and the averaging was set to 4, i.e., four test signals were generated and then averaged. The SREV1 has to perform various DSP processes to the sampled sound so as to create the impulse-response data necessary for convolution.

4.2 Equipment Sessions A-C

4.2.1 Session A - Real-Source Data using a Soundfield Microphone

The monophonic test signal was recorded in seventy-two different angles of incidence using the Soundfield microphone with the amplifier connected to the SREV1. The recorded data, the B-format signals, were stored on a PC card (type II 48 MB) and then transferred and saved on a PC laptop, connected to the SREV1. The SREV1 software installed on the computer allowed for simultaneous generation of the test signals and four-channel acquisition of the microphone signal.

In figure 4.7 the schematic wiring diagram for session A can be seen, while figure 4.8 shows the placement of the equipment in the semi-anechoic chamber.
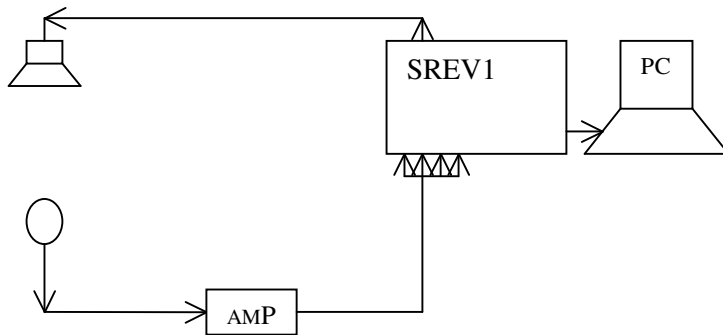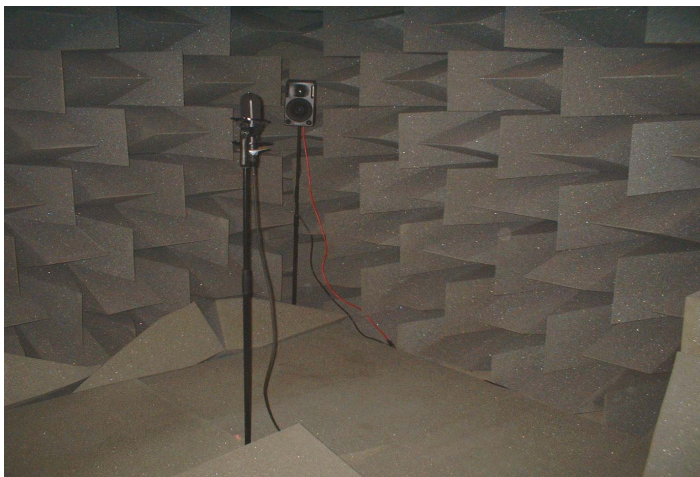
Figure 4.7 *Session A –
Wiring Diagram*

Figure 4.8 *Loudspeaker
and microphone
placement for session A*

4.2.2 Session B – Real Source Data using KEMAR – Binaural Recording

In session B the same equipment was used as in session A, but instead of the Soundfield microphone system the KEMAR system was used which, apart from the head and torso, involves two microphones and an amplifier. The microphones used here, Sennheiser, were omnidirectional and mounted at the openings of the ear canals. Unfortunately the microphones were slightly smaller than the openings of the ear canals. Blue tack was used to fill out the gaps between the microphones and the openings. The loudspeaker was placed in the corner as in session A, and the KEMAR system was placed where the Soundfield microphone had been situated. Once again the Yamaha SREV1 generated the test signal, and was connected to the Laptop with its PC card on which to store the data. Figure 4.9 shows the wire diagram and figure 4.10 the equipment in the chamber.
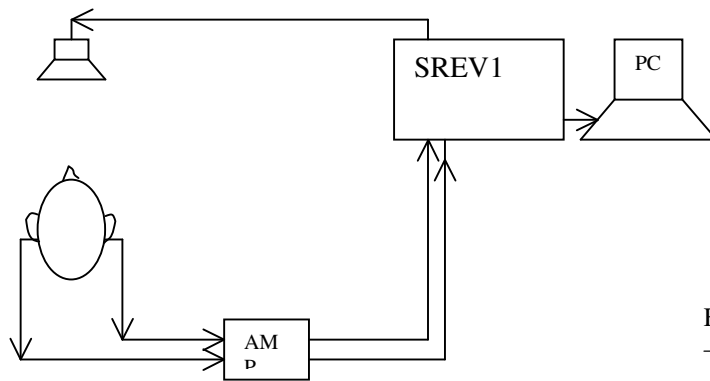
Figure 4.9 *Session B – Wiring Diagram*

Figure 4.10 *Loudspeaker and microphone placements for session B*

4.2.3 Session C – Recording the reproduced signals using KEMAR – Binaural Recording

Before reproducing the seventy-two signals captured in session A, the duration of the signals was edited to a total length of one second in Nuendo[5]. This was done in order to trim the audio files before reproducing the signals over a square Ambisonic loudspeaker array. Four loudspeakers with internal amplifiers were situated on a circle with a radius of 1.4 meters with the KEMAR and its amplifier positioned at the centre point. A part of the set up is shown in figure 4.11.

---

[5] Nuendo – Steinberg software for multi-track recordings and editing.

Figure 4.11 *Equipment set up, session C*

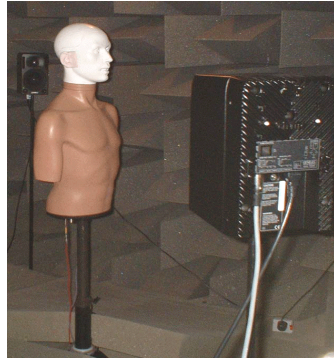 The MOTU 828 FireWire Audio Recording Interface [26], a single rack-space interface, was connected to a Laptop (Mac) on which Nuendo v. 1.5 was installed. The interface allowed for simultaneous input and output. As is shown in figure 4.12, the interface has eight channels of balanced 24-bit analog input/output. No PCI card was required since the MOTU 828 was connected to a FireWire-equipped Apple Macintosh computer, why all the recorded data were stored directly on the Laptop. The sampling rate was 48 kHz. In figure 4.13 the wiring diagram for session C can be seen.

Figure 4.12 *MOTU 828 FireWire Interface*
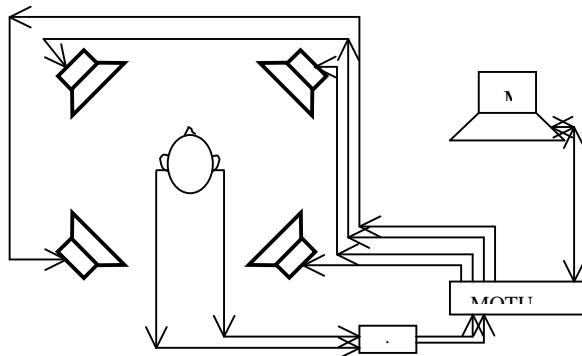
Figure 4.13 *Session C – Wiring Diagram*

For the decoding of B-format signals the 'B-dec' plug-in in Nuendo was used, a plug-in designed by Ambrose Field at Department of Music Technology at the University of York [27]. Since this plug-in is created for music reproduction, it does not contain the psychoacoustic filters required to obtain the correct decoding processes according to Gerzon. The configuration of the B-dec followed the B-dec plug-in scheme for a square Ambisonic loudspeaker array, where the multi-channel file B-format input, were two stereo files feeding

separate input channels that then fed the decoder. Because of problems with slight electrical noise the adjustment of the loudspeaker volume was not the ultimate; the level output of the four loudspeakers were adjusted as alike as possible, up to the middle of the total level range, and then modified in Nuendo.

4.3 Analysis methods

From the three sessions a vast amount of data was collected, of which a small part has been analysed. In the following, the methods used for deriving data, ILD and ITD, from sessions B and C, will be presented, alongside a brief survey of other commonly used methods. Furthermore, the B-format signals X and Y were inspected manually to verify the orientation of the microphone.

4.3.1 ITD

As has been discussed in *Background, Spatial Hearing - Psychoacoustic Theory,* and as Shinn-Cunningham [4] points out in her article Localising Sound in Rooms, for every single frequency there is an essential ambiguity in the interaural time difference that corresponds to a phase difference at that particular frequency. According to Shinn-Cunningham [4] the "true" interaural time difference is that value which yields approximately the same ITD at all frequencies, which for an anechoic space would show a flat line as a function of frequency. This would hold if no reflections, generated within that acoustic space, have interfered with the recorded signals. Since, in this investigation, precautions were made in using absorbing mattresses on the concrete floor no analysis concerning how big influence first reflections might have had on the recorded data has been carried out. In previous articles and papers that measured the interaural time difference the Interaural Cross-Correlation (IACC) method has been used [3]. This method calculates the time difference between the two discrete signals for the left and right ear through analysis of how well the signals correlate with each other, i.e., the degree of similarity between the signals may be evaluated. The cross-correlation function Φ_{lr} is defined as (8):

$$\Phi_{lr}(\tau) = \sum_{t=0}^{N-1} L(n)R(n + \tau) \tag{8}$$

where n is the length of the discrete ear sequences captured [3]. The maximum value is found at a time log τ, which corresponds to the time difference between the arrivals of the sound at the two ears. The ITD value is then the value of τ corresponding to the maximum value of Φ_{lr} [3]. At time τ = ITD, the value of Φ_{lr} is known as the $IACC_{max}$. This method produces precise calculations of ITD and is particularly useful when the received signals look very different from each other. Calculation of the cross-correlation function has not been calculated in this project, since the signals captured were considered to be similar enough. Instead the interaural time differences were derived by measuring the distance in milliseconds from first arriving peak of the signal of left ear to first arriving peak of the signal of right ear. Both the signal editing programmes Nuendo and Wavelab[6] were used. For further comparison and validation of the results obtained in session B, calculations of the ideal ITD using equation (1):

---

[6] WaveLab – Signal Editing Programme

$$ITD = r\ (\theta+\sin\theta)/c$$

were carried out. As the distance between the ears on KEMAR measured 15 cm (18 cm front lobe to back lobe.) r equal 7,5 cm. θ constitutes the angle of incidence in radians and has been calculated for the angles between 0° to 180°. For the angles between 185° to 360°, the same results derived for the "front" angles have been used but with a negative sign in front in order to create interaural time difference at the two ears. The recordings were carried out in a non-heated chamber with an average temperature of approximately 10° C, why the speed of sound, $c,$ was calculated to ~338 m/s.

4.3.2 ILD

ILD can be derived using several methods. Previous investigations have for instance derived ILD from magnitude spectra of different HRTFs [4]. Taking the frequency-dependant ratio of the Fast Fourier Transform (FFT) of the sampled left and right signals [3] produces a frequency-dependant ILD, which is defined in equation (9)

$$ILD_F = FFT[L(n)]/FFT[R(n)], \qquad\qquad (9)$$

where for this previous investigation the FFTs were preceded by a Hamming window. In this project the ILD's between left and right ear for both sessions B and C have been investigated using two analysis methods. Firstly **peak level measures** in dB on both left and right signals, were carried out for sessions B and C. The peak level measurements were carried out in both WaveLab and Nuendo. In order to obtain the ILDs for each session, the peak levels for right ear were subtracted from the peak levels of the left ear. The obtained data for session C were then compared with the data from session B. Secondly, in order to obtain more detailed data, magnitude spectra for each angle of incidence of both left and right signals were computed. By subtracting the spectrum of right ear from the spectrum of left ear the **spectral ILD** for every angle of incidence was obtained. The spectral ILD was first derived from data from session B so as to outline how well the spectra of the recorded signals corresponded to the expected spectra. This was done for the angles of incidence that are easiest to analyse in terms of interpretation of the spectrum, namely 0°, 90°, 180° and 270°. In order to outline how well session C followed session B, comparisons of the spectrum for the most relevant angles of incidence, mentioned above, were carried out in the frequency range 700–3000 Hz. All spectral calculations were performed in MatLab.

5. Results

Before discussing the results a few acknowledgement will be given. An error occurred while editing the four signals of B-format, W, X, Y and Z, for angle of incidence 95º. The signals were mixed in the wrong order, resulting in an improper signal that showed too low spectral information. The data for this angle was discarded. Furthermore the data for the angle of incidence 265º in session B, was corrupted during the formatting process between different formats (the original .tm4 to .wav). Neither of these data dropouts had an impact on the results. For completeness, all primary data are presented in Appendix I – ITD, Appendix II – Ideal ITD, Appendix III – Peak Level and Appendix IV – Spectral ILD. Further the reader will recall that the microphones, both the Soundfield microphone and the KEMAR system used, were turned clock-wise during the recordings. The signals reached the left ear first for angles of incidence between 0º and 180º, and the right ear first for angles between 180º and 360º. In always subtracting the right ear signal from the left ear signal, positive interaural differences for signals reaching the left ear first were obtained and negative differences when the right ear was reached first.

5.1 ITD

5.1.1 ITD Comparing Session B and Session C

From a quick look at diagram 5.1 it could be noted that the interaural time differences for session C, solid/dotted line, to a rather great extent followed the time differences for session B, dashed/triangular line. The shortest time differences in milliseconds for session B could be found at angles of incidences $0° = 0.016$, $180° = -0.01$ and $360° = 0.016$. For session C they were found at $0° = 0.06$, $175° = -0.016$ and for both $180°$ and $355° = -0.03$. As for the longest detectable time differences, they were for session B found at $90°$ and $100° = 0.87$ and at $275° = -0.97$ and for session C at $90° = 0.53$ and at $270° – 275° = 0.512$.

Diagram 5.1 *ITD, Sessions B and C*
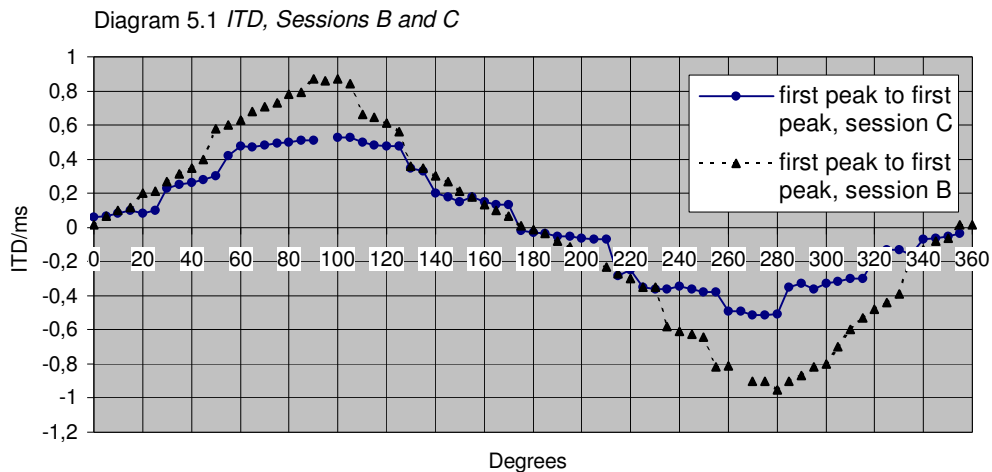
Furthermore, from diagram 5.1, it could be observed that the time differences for session C followed the time differences for session B at angles of incidence from 0º and up to approximately 60º. Beyond 60º, session C did not show as distinct an increase as could be seen for session B. Instead, from 60º to approximately 125º, the time differences had almost

stagnated. From there on to 180º, a decrease could be observed, which for session B started at 110º. Beyond 180º, when the location of the sound is on its "way back" to front position, it could be noted that the time differences for angles of incidence 185º - 355° for session C, were not as large as the time differences observed for angles of incidence 0º - 180º. As for session B, the time differences beyond 180º followed to a great extent, the differences observed for angles of incidence before 180º.

5.1.2 ITD Comparing Sessions B and C and Ideal ILD

By adding the ideal ITD (i.e., estimated ITD using equation (1)) to diagram 5.1, diagram 5.2 is obtained. The ideal ITD is here the dotted line. From the diagram it could be noted that the ITD:s for session B were longer than the ideal ITD:s for the "side" angles between 50˚ - 125˚ and 245˚ - 320˚ respectively. Furthermore, session C followed the calculated ideal ITD values remarkably well, in particular for the angles of incidence between 0˚ and 180˚. The correspondence is not as good beyond 180˚, but is still better than in comparison with session B. The average error value between session C and the ideal ITD were 0.059 and the standard deviation 0.057.

Diagram 5.2 *ITD, Sessions B and C and Ideal ITD*
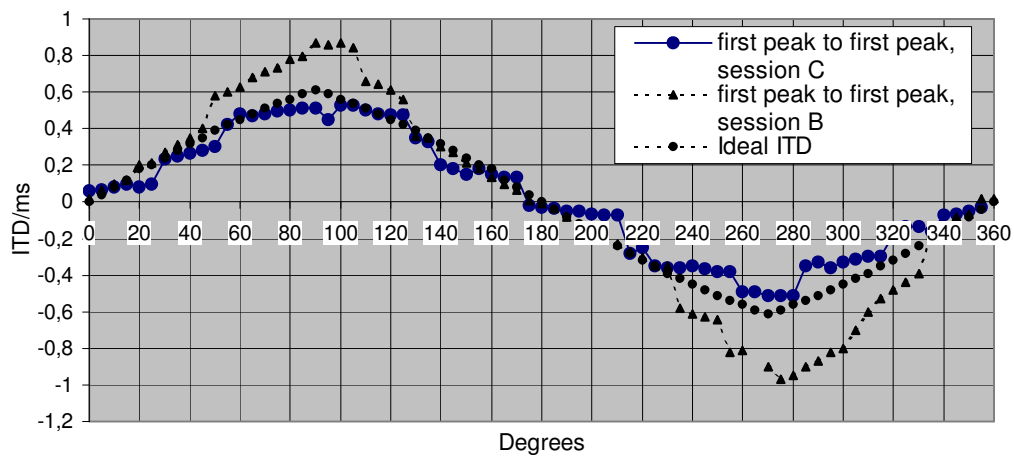
5.2 ILD – Peak Level Measurement

From diagram 5.3, it can be observed that the peak levels for session C, solid/dotted line, are not as distinctively associated with incoming angle of incidence as the peak levels for session B, dashed/triangular line. This is above all noticeable from ~215º and onwards, where the location of the sound is on its way back to the centre front. A slow descending decrease in level difference can be observed for session C.

Diagram 5.3 *ILD, Sessions B and C*
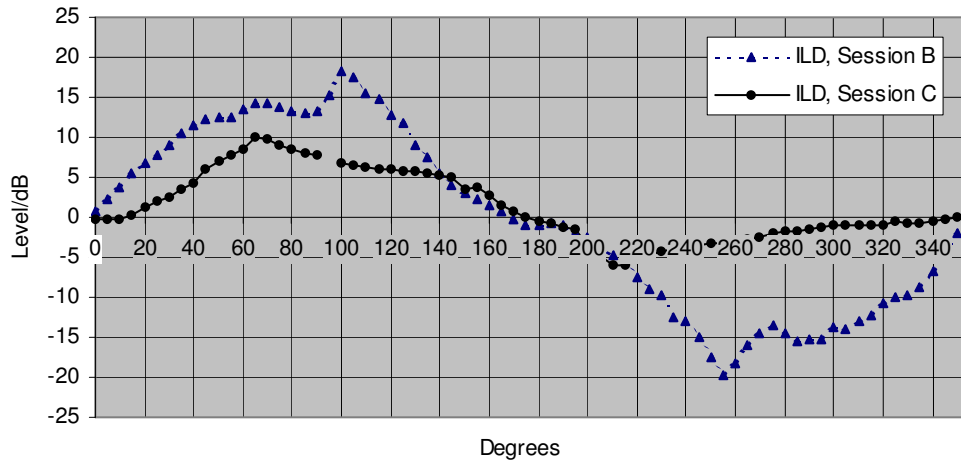
From the results one may get the impression that the equipment somehow failed after approximately 215°. In presenting the data as can be seen in the diagrams 5.4 and 5.5 (p. 32), yet another scenario is possible. For both diagrams the dashed/triangular line constitutes the peak levels for session B and the solid/dotted line, the peak levels for session C. From diagram 5.4, showing the right ear, i.e., the ear that was positioned farthest away from the source for angles of incidence between 20°-180°, it could be observed that the peak levels for session C, seemed to correspond well to changes in levels as a function of incident angle. Concerning the left ear, showing in diagram 5.5, for the same angles of incidence, a flat peak level response could be observed. This behaviour, a flat response, was also noticeable when the right ear was found closest to the source, between approximately 220° - 355°, see diagram 5.4. The left ear, now positioned farthest away, see diagram 5.5, showed the same pattern as for the right ear when positioned furthest away. The peak levels corresponded well to changes in levels as a function of incident angle, but not with the same prominence as could be observed for the right ear.

Diagram 5.4 *Peak Levels Right Ear, Sessions B and C*
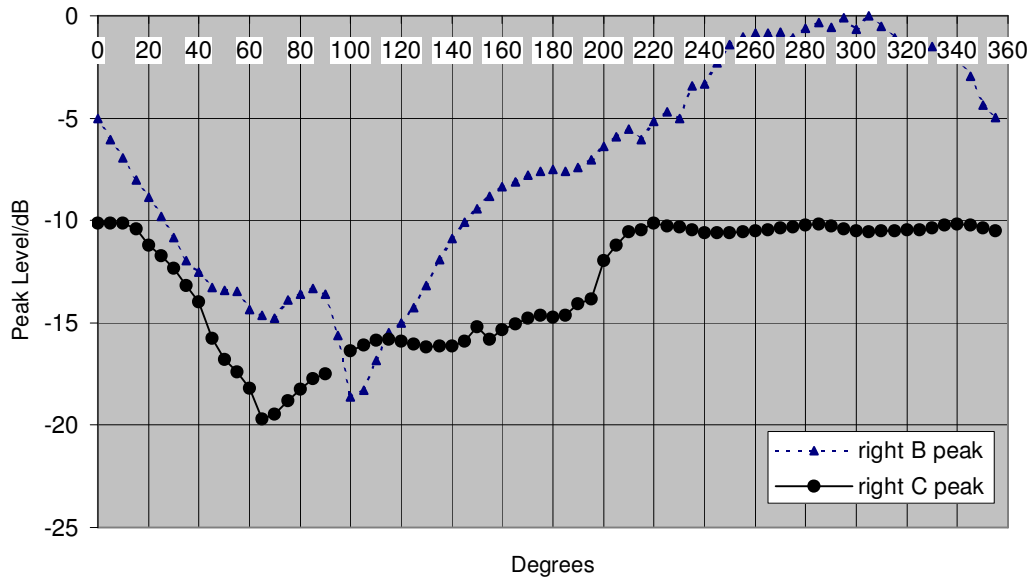
Diagram 5.5 *Peak Levels Left Ear, Sessions B and C*
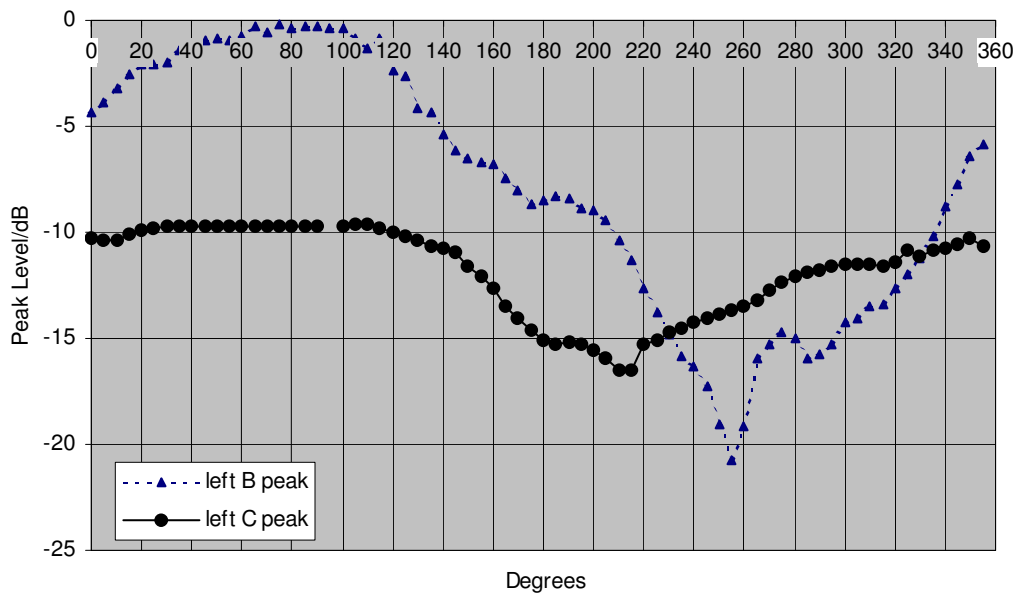
5.3 Spectral ILD

5.3.1 Session B

In diagrams 5.6 – 5.9 the spectral ILD:s for angles of incidence 0°, 90°, 180° and 270° are presented. In each diagram, the dashed line constitutes the spectral ILD, which underneath the magnitude spectrum for the left ear, solid line and for the right ear, dashed/dotted line respectively, are presented. The magnitude is expressed in dB. From the diagrams it could be observed that the spectral ILD:s, for frequencies between 700 and 3000 Hz, changes with angle of incidence. For angle 0º, see diagram 5.6, the two magnitude spectrum for left and right ears coincide; giving rise to a relatively flat spectral ILD that fluctuates around 0 dB.

Diagram 5.6 *Spectral ILD, Session B, θ = 0°*
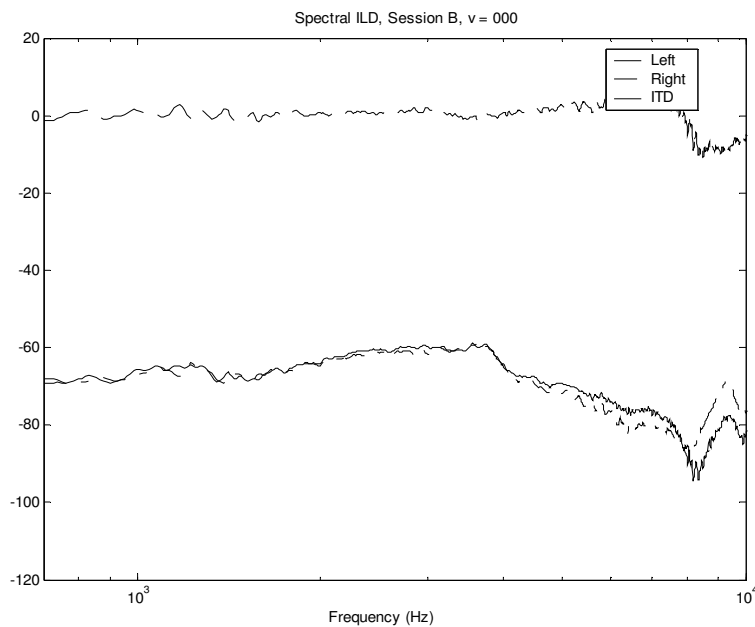
Concerning the angle of incidence 90º i.e., where the left ear is located closest to the source and the right ear is positioned furthest away from the same, spectral ILD of approximately 8 dB, could be observed, see diagram 5.7 below.

Diagram 5.7 *Spectral ILD, Session B, θ = 90 °*

Spectral ILD, Session B, v = 090
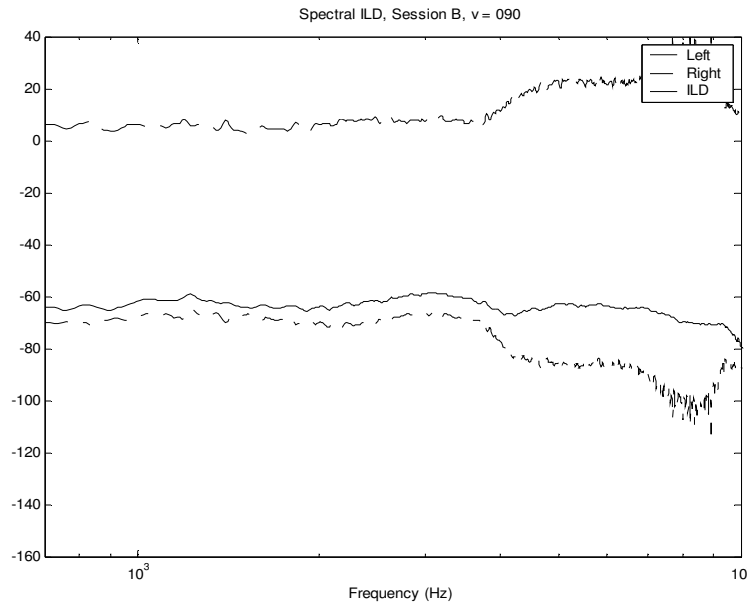
With the face fully turned away from the loudspeaker, at 180°, the spectral ILD was ~0 dB, see diagram 5.8. Both ears were positioned at the same distance from the source.

Diagram 5.8 *Spectral ILD, Session B, θ = 180 °*

Spectral ILD, Session B v = 180
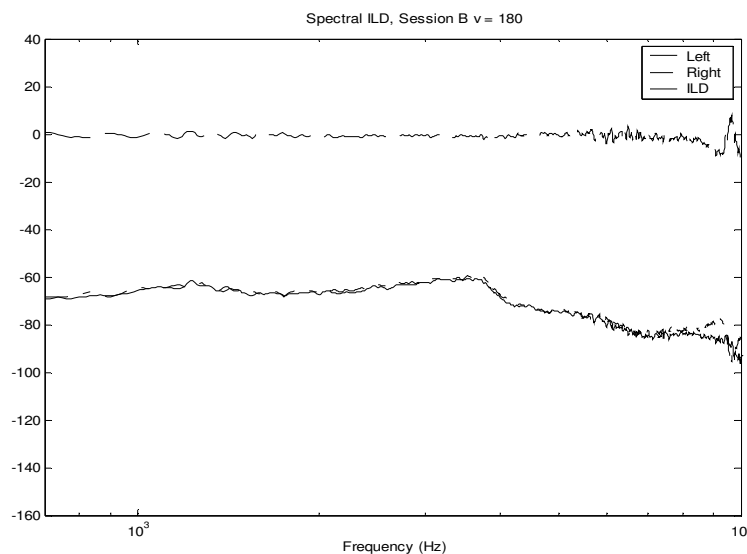
At an angle of incidence 270º, see diagram 5.9, the same spectral level difference as for angle of incidence 90º could be observed but with the opposite sign, approximately -8 dB. The minus sign indicates that the right ear was closest to the source.

Diagram 5.9 *Spectral ILD, Session B, θ = 270°*

Spectral ILD, Session B, v = 270
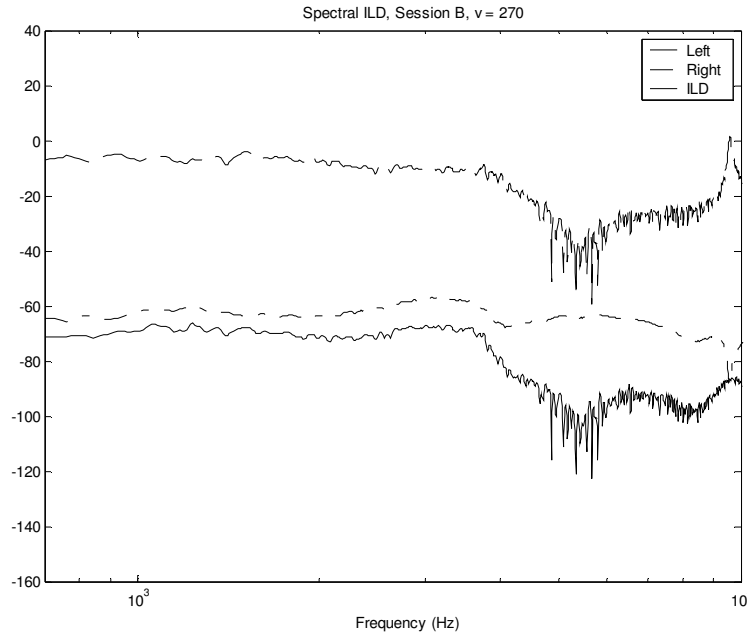
5.3.2 Comparing the spectral ILD for session B with session C

In order to outline whether the spectral ILD for session C was as satisfying as for session B, the two sessions were compared for the above-mentioned angles of incidence 0°, 90°, 180° and 270°. In diagrams 5.10 – 5.13 (p. 36-37), the spectral ILD for session B is represented by the solid line and for session C, by the dotted lined. The comparison between the two sessions was made in the frequency range 700 - 3000 Hz. Large spectral variations could be observed for session C for angles of incidence 0° and 90°, see diagram 5.10 and 5.11. For instance, at angle of incidence 90°, the spectral ILD for session C showed two remarkable spectral peaks at approximately 1,5 kHz and 3,8 kHz. Between the two peaks, at approximately 2000-3500 Hz, lower spectral ILD:s for session C could be observed. The extraordinary high peaks in combination with the lower levels deliver a pattern that resembles of a comb-filter effect. Furthermore for angle of incidence 180°, see diagram 5.12, a more reasonable result, than was found for angle of incidence 0°, could be observed.

Diagram 5.10 *Spectral ILD, Sessions B and C, θ = 0 °*

Spectral ILD, Session B an C, v = 000
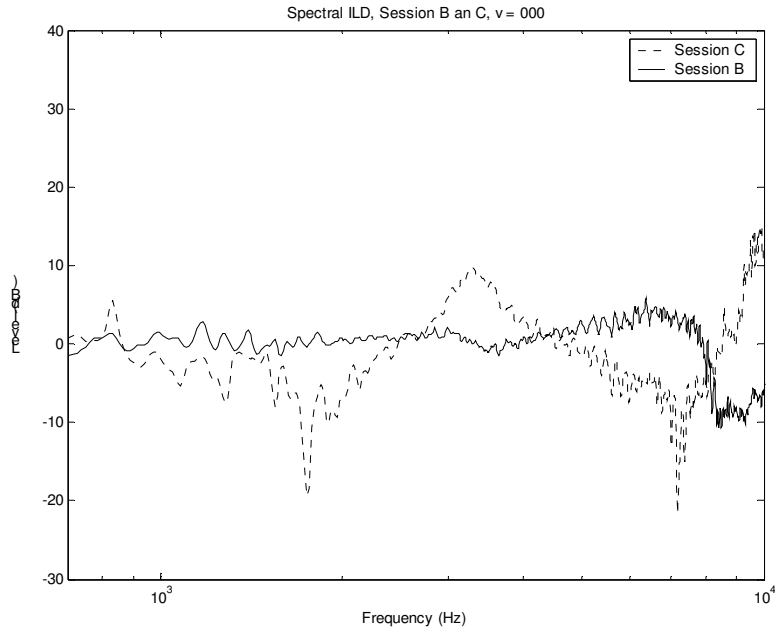
Diagram 5.11 *Spectral ILD, Sessions B and C, θ = 90 °*

Spectral ILD, Sessions B and C, v = 090
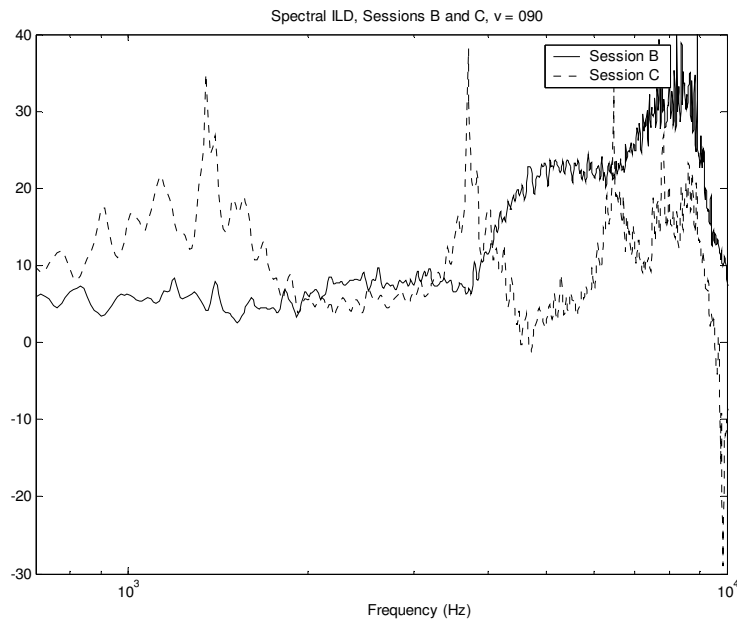
Diagram 5.12 *Spectral ILD, Sessions B and C, θ=180°*

Spectral ILD, Session B and C, v = 180
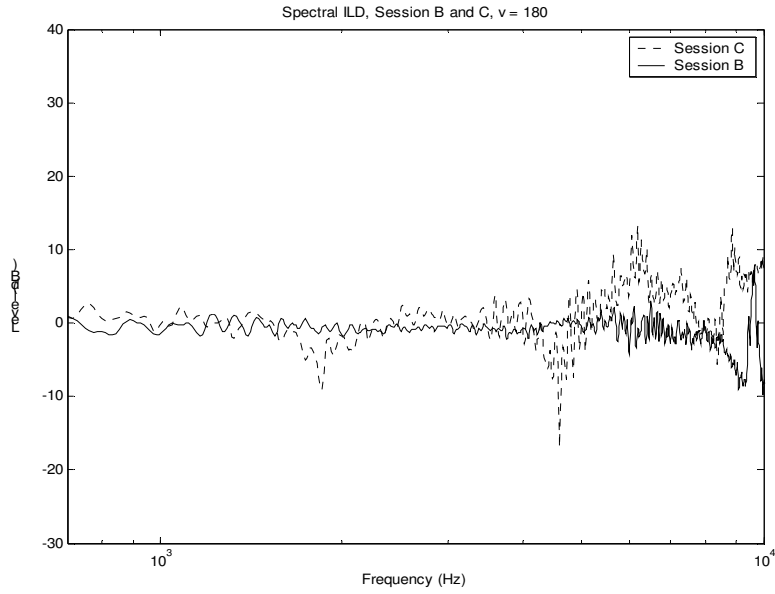
From the result in diagram 5.13, for angle 270°, the variations for session C was not as prominent as could be observed for angle 90°. No comb-filter effect was noticeable.

Diagram 5.13 *Spectral ILD, Session B and C, θ = 270°*

Spectral ILD, Sessions B and C, v = 270
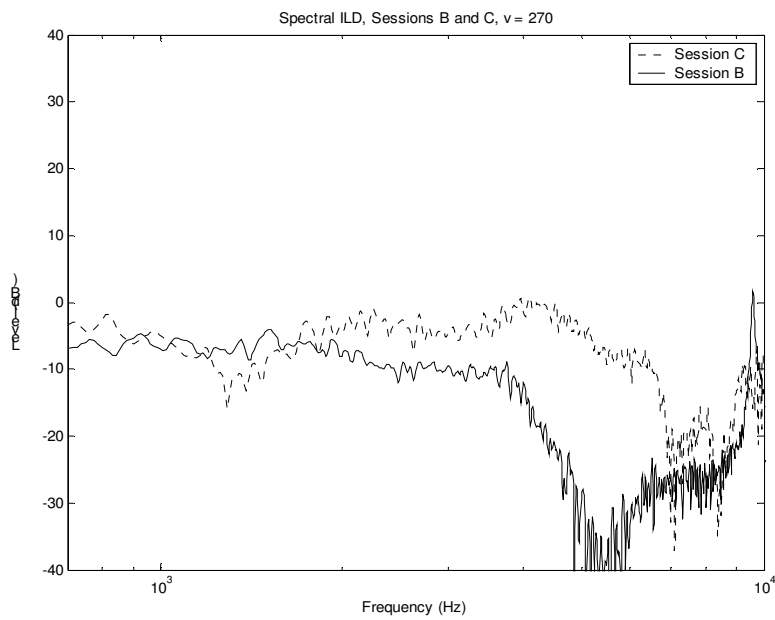
6. Discussion

On the basis of the above-presented results and as a first reflection, it may be stated that several more different analysis methods could have been used on the vast amount of data. Furthermore, it would have been preferable to analyse all of the captured data. Unfortunately time did not permit any of this, but still the analysed data is considered to be a sufficient enough basis to draw conclusions from.

In a similar investigation, carried out by Mac Cabe and Furlong [3], three different surround-sound systems, Spectral Stereo, KEMAR reconstruction and Ambisonics for a reproduced 180º - image, were compared from a psychoacoustic point of view, where both ITD and ILD were in focus. In this previous investigation a source signal was fed through a square Ambisonic pan pot in order to synthesize B-format signals. The synthesised signals were then recorded by a KEMAR system positioned in the centre of the loudspeaker array. According to the results, concerning measurements on the Ambisonic system, the most interesting to note was that no calculated ITDs, using equation (8), were detectable between 0º and 45°. Between approximately 60° and onwards to 90°, the time difference were found constant at approximately 0,56 ms. This "flatness" followed by a decrease between 90° and 120°, which in turn was followed by an increase at 135° before a distinct decrease could finally be observed between approximately 150° and 180°. A better localisation performance concerning spectral ILD could be found according to the authors, at the early azimuths but were not as good as could have been expected. In this previous investigation, ILD were calculated using the above given equation (9) and the obtained results were specific for the room in which the tests were carried out. Since the aim of this investigation was to compare how well different surround-sound systems reproduced a real sound-field, the results concerning the Ambisonic system were not as thoroughly discussed. Still the results of this previous investigation have been taken into consideration when analysing the results in this latter project.

Since the original idea of the project, discussed in this report, was to find what objective measures there might exist in order to verify the horizontal only Ambisonic B-format, the use of a Soundfield microphone, was of importance. When the idea, during the project instead was transformed to focus on the accuracy of the Ambisonic system itself, the use of the Soundfield microphone consisted. A synthesized signal could have been used and encoded by using the encoding equations and then later on be decoded for reproduction over the Ambisonic system. But since the use of a Soundfield microphone in the future gives more opportunities, it was decided to continue as planned.

Furthermore, the best way of how to manoeuvre the microphones in session A and B into seventy-two different positions, were discussed, where the most desirable would have been to use an automatically manoeuvrable device. When a decision was to be made, it was considered being far too time consuming constructing such a device why the only solution left was to manually manoeuvre the microphones. This did not pose any problem concerning the KEMAR system since the torso was attached on a rotating stand with required marks every 5°. The manually manoeuvre of the Soundfield microphone was solved by mounting a one's own hand made hard paper disc, with marks plotted every 5°, on the Soundfield microphone. It later turned out, after a quick check on the recorded Soundfield microphone X- and Y- signals that these, for the angles of incidences 0°, 90°, 180° and 270°, corresponded to the expected differences for these positions. Instead a slightly dislocation could be found in the KEMAR signals. This displacement was probably due to the fact that

the KEMAR was not properly positioned in front of the loudspeaker in session B when the actual recordings were carried out. It could be argued that analysing the ITD on the basis of Interaural Cross-Correlation (IACC) would have performed a more accurate result concerning the displacement. But since it proved to be such a small displacement it has not been considered being of importance in the context of this investigation where accurate enough data has yet been derived.

Furthermore, since using Blue Tack, the perhaps not equally mounted microphones at the openings of the ears might have had a negative impact on the recorded signals. In particular concerning the ILD. On the other hand, in relation to the fact that the B-dec was used, which lacks the requisite psychoacoustic shelf filters, it is not very probable that Blue Tack could have had such a great impact.

Comparing the ILD results obtained from the methods chosen here, the peak level method and the spectral ILD method, it can be stated that the results are difficult to compare since the spectral ILD method shows more detailed information which gives the possibility to observe the spectral ILD for each frequency. Concerning the cancellations – the comb-filter effect - it might be argued that something within the semi-anechoic chamber gave rise to this. Since the recording equipment was hidden behind a shelter during the recordings and no other devices were interfering, as far as could be seen, the cancellations could have been due to the fact that the loudspeaker array was not accurately and symmetrically positioned on its circle with KEMAR in the middle.

Even though, according to the results derived so far, it may be concluded that the objective ITD measure is a good measure for verification of the accuracy of a horizontal only Ambsonic system, it would for future work be of interest to further validate the method using more loudspeakers, and then in particular investigate the performance of the ITD at the angles of incidence to the "side", and also analyse the ITD on the basis of IACC for more accurate data. Furthermore the decoder used in this project, the B-dec, designed mainly for musical production, lacks the, for a proper decoding scheme, psychoacoustic designed shelf filters that modifies the low frequency vector magnitude as the frequency increase so that the energy vector magnitude becomes optimal at higher frequencies. This might have given rise to the less well corresponding results, particularly concerning the ILD that differ from what could have been expected. However, this does not bars the possibility to obtain good ILD results using the same method in a horizontal only Ambisonic system, but correctly decoded. This should by all means generate a better ILD than was shown here. Furthermore, the data collected here contains information that has not yet been analysed and interpreted, not only concerning the already derived phenomenon, but other psychoacoustic mechanisms as well, such as HRTF and the shape of the pinna if the project will be expanded to a full sphere Ambisonic. The derived data might also be interesting to apply in a Digital Wave Guide Mesh model in order to verify the model itself.

7. Conclusion

In this project the two major psychoacoustic mechanisms, ITD and ILD, have been investigated whether they constitute powerful objective measures for verification of a horizontal Ambisonic surround-sound system. The two mechanisms are effective in different frequency ranges, where ITD works best in the low frequency domain up to approximately 700 Hz, and ILD beyond 700 Hz. The method chosen for capturing data involved recording a sound-field, generated by a real monophonic source, for seventy-two different angles of incidence in both using a Soundfield microphone (session A) and a KEMAR system (session B). The signals recorded in session A were reproduced over a square Ambisonic loudspeaker array, and then once again recorded by the KEMAR system (session C). For each angle of incidence, the results from session B and C have been analysed and then compared, where session B constituted the reference.

In analysing the data from session B, concerning both ITD and ILD, it could be concluded that the results corresponds well to what would be expected, i.e., a small or no change for angles of incidence at 0º and 180º, and maximum and minimum differences for 90º and 270º respectively. Furthermore in comparing the ITD in session B with the calculated ideal ITD it could be observed that session B, in particular for the "side" angles 60˚ - 120˚ and 245˚ - 320˚ respectively, exaggerated the ideal ITDs. These longer time differences for measurements on a real source could also be found in the results obtained by Furlong and Mac Cabe [3] for an 180˚-sound image. For this investigation the over all observed longer time differences for the real source might be explained by the low temperature in the semi-anechoic chamber when recording session B, giving a slower speed of sound. But the most reasonable is the fact that ideal ITD only constitutes an estimated value that gives an indication of how long the time difference might be expected to be.

Since the ITDs for session C could be found to follow, for 0º, 90º, 180º and 270º, both the ITDs in session B and to an even greater extent the calculated ideal ITD, it caould be concluded that measuring the ITD constitutes a good and reliable objective method for verification of the horizontal only Ambisonics system used here. The lack of larger differences for the "side" angles may be due to the fact that a square loudspeaker array was used, i.e., no loudspeakers were located at the angles of incidence 90º and 270º and 0º and 180º respectively. A source located at these positions needs to be reproduced through a combination of different loudspeaker feeds and since these angles are at the exact mid-points between the left/right and front/back speakers, then it can be stated that this imaging would not correspond that well to the real source, which was shown for the side angles but was not as legible for the front/back angles. In using more loudspeakers this would perhaps improve the result. Since no psychoacoustic shelf filters are accounted for in the B-dec, which primarily are designed in order to improve the high frequency content of a signal, it has been shown here that ITD does not seem to be as affected by the lack of this filters. This might be explained by the fact that ITD is mostly effective for the low frequency content of the signal and therefore more robust to a decoder that does not provide the psychoacoustic filters. This leads to the conclusion that the results concerning ITD obtained in this investigation may hold for not only the Ambisonic system investigated here, but for Ambisonic systems in general.

Concerning ILD, less clear and straightforward results were obtained. The peak level measure method, gave a first indication of whether the results were all right or not considering the level difference as a function of incident angle. A first glance at these results

may prove the contrary, in particular if only focusing on the results obtained in diagram 5.3 (p. 31), where an apparent "collapse" is noticeable after approximately 225°. From thereon the level differences are hardly noticeable, which might indicate that something went wrong with the equipment or that the decoder ceased providing the proper loudspeaker feeds. In presenting the data as can be seen in diagram 5.4 and 5.5 (p. 32), another conclusion may be drawn. From these diagrams it is clear that level differences are accounted for as a function of incident angles over a square Ambisonic loudspeaker array, but apparently only for the ear that is present furthest away, i.e., the ear that is shadowed. From the result it seems as if the Ambisonic system is incapable of producing higher levels. This is perhaps not desirable since the shadow effect is accounted for on the other side, i.e., the ear that is furthest away. Why the level, as a function of incident angel, is particularly constant on the side of the head where the ear is the closest to the source cannot be explained. Still this implies that the psychoacoustic mechanism ILD may constitute an objective measure for verification of a horizontal only Ambisonic system.

In the time domain the peak levels may be difficult to detect because of phase differences in the binaural signal. Therefore a more detailed spectral approach was carried out as well, resulting in much more complex and informative data. In choosing the angles of incidence 0°, 90°, 180° and 270° for both sessions B and C and then compare the spectral ILD:s between the frequency range 700 – 3000 Hz, great differences for session C could be observed. Where one would expect the spectral ILD curve for session C to follow the spectral ILD curve for session B at an angle of incidence 0° (see diagram 5.10, p. 36) the spectral ILD for session C showed irregularities such as a big dip at approximately 1800 Hz and a peak at approximately 3200 Hz and then a dip again at approximately 7000 Hz, giving rise to a typical comb-filter pattern. Such a pattern was even more noticeable at an angle of incidence 90° where for this angle the dips now had turned to peaks at approximately 1400 Hz and then at approximately 3400 Hz. This comb-filter pattern could also be observed at the angle of incidence 180° but was not as prominent as the former. As for angle of incidence 270° no prominent comb-filter effect could be observed.

From the above surveyed results it can be concluded the ITD may constitute a reliable objective measure for verification of horizontal Ambisonic systems in general. Concerning the ILD, while the explanation is not fully clear why the ILD for a point source is adequately reconstructed on the far side of the head but not on the side where the head is closest to the source, it can still act as a directional hearing cue over the Ambisonic system used here.

Appendix I- Table ITD

No. of position	Degrees	First peak to first peak, C	First peak to first peak, B
1	0	0,06	0,016
2	5	0,066	0,066
3	10	0,083	0,099
4	15	0,099	0,116
5	20	0,083	0,2
6	25	0,099	0,215
7	30	0,231	0,27
8	35	0,25	0,314
9	40	0,264	0,35
10	45	0,281	0,4
11	50	0,3	0,578
12	55	0,42	0,6
13	60	0,479	0,628
14	65	0,47	0,678
15	70	0,48	0,71
16	75	0,496	0,73
17	80	0,5	0,78
18	85	0,513	0,793
19	90	0,513	0,87
20	95		0,86
21	100	0,53	0,87
22	105	0,529	0,84
23	110	0,5	0,661
24	115	0,48	0,645
25	120	0,476	0,611
26	125	0,476	0,56
27	130	0,347	0,36
28	135	0,33	0,35
29	140	0,2	0,3
30	145	0,18	0,27
31	150	0,15	0,214
32	155	0,18	0,181
33	160	0,15	0,132
34	165	0,132	0,099
35	170	0,132	0,066
36	175	-0,016	0,01
37	180	-0,03	-0,01
38	185	-0,033	-0,033
39	190	-0,05	-0,083
40	195	-0,05	-0,116
41	200	-0,066	-0,15
42	205	-0,07	-0,2

43	210	-0,07	-0,231
44	215	-0,281	-0,27
45	220	-0,25	-0,3
46	225	-0,35	-0,35
47	230	-0,36	-0,35
48	235	-0,36	-0,579
49	240	-0,347	-0,612
50	245	-0,364	-0,628
51	250	-0,38	-0,645
52	255	-0,38	-0,82
53	260	-0,49	-0,81
54	265	-0,49	
55	270	-0,512	-0,9
56	275	-0,512	-0,9
57	280	-0,51	-0,95
58	285	-0,35	-0,9
59	290	-0,33	-0,87
60	295	-0,36	-0,82
61	300	-0,33	-0,8
62	305	-0,314	-0,7
63	310	-0,297	-0,6
64	315	-0,297	-0,53
65	320	-0,18	-0,479
66	325	-0,132	-0,44
67	330	-0,132	-0,39
68	335	-0,16	-0,165
69	340	-0,07	-0,132
70	345	-0,066	-0,083
71	350	-0,05	-0,066
72	355	-0,033	0,016
73	360		0,016

Appendix II – Table ideal ILD

Degrees	ms
0	0
5	0,04
10	0,08
15	0,12
20	0,18
25	0,2
30	0,24
35	0,28
40	0,32
45	0,35
50	0,39
55	0,42
60	0,45
65	0,48
70	0,51
75	0,54
80	0,56
85	0,59
90	0,61
95	0,59
100	0,56
105	0,54
110	0,51
115	0,48
120	0,45
125	0,42
130	0,39
135	0,35
140	0,32
145	0,28
150	0,24
155	0,2
160	0,18
165	0,12
170	0,08
175	0,04
180	0
185	-0,04
190	-0,08
195	-0,12
200	-0,18
205	-0,2

210	-0,24
215	-0,28
220	-0,32
225	-0,35
230	-0,39
235	-0,42
240	-0,45
245	-0,48
250	-0,51
255	-0,54
260	-0,56
265	-0,59
270	-0,61
275	-0,59
280	-0,56
285	-0,54
290	-0,51
295	-0,48
300	-0,45
305	-0,42
310	-0,39
315	-0,35
320	-0,32
325	-0,28
330	-0,24
335	-0,2
340	-0,18
345	-0,12
350	-0,08
355	-0,04
360	0

Appendix III – Table ILD

No. of Position	Degrees	Left, B	Right, B	Left, C	Right, C	ILD, Session B	ILD, Session C
1	0	-4,34	-5,03	-10,28	-10,13	0,69	-0,15
2	5	-3,85	-6,04	-10,39	-10,11	2,19	-0,28
3	10	-3,2	-6,96	-10,34	-10,15	3,76	-0,19
4	15	-2,58	-8,03	-10,11	-10,42	5,45	0,31
5	20	-2,1	-8,85	-9,93	-11,19	6,75	1,26
6	25	-2,03	-9,78	-9,83	-11,73	7,75	1,9
7	30	-1,94	-10,84	-9,75	-12,35	8,9	2,6
8	35	-1,4	-11,94	-9,7	-13,18	10,54	3,48
9	40	-1,12	-12,53	-9,73	-13,98	11,41	4,25
10	45	-0,92	-13,26	-9,76	-15,77	12,34	6,01
11	50	-0,81	-13,43	-9,75	-16,78	12,62	7,03
12	55	-0,94	-13,46	-9,76	-17,42	12,52	7,66
13	60	-0,73	-14,33	-9,75	-18,22	13,6	8,47
14	65	-0,32	-14,62	-9,75	-19,69	14,3	9,94
15	70	-0,59	-14,77	-9,76	-19,47	14,18	9,71
16	75	-0,21	-13,87	-9,75	-18,79	13,66	9,04
17	80	-0,36	-13,62	-9,75	-18,26	13,26	8,51
18	85	-0,3	-13,32	-9,72	-17,75	13,02	8,03
19	90	-0,3	-13,59	-9,71	-17,48	13,29	7,77
20	95	-0,37	-15,64			15,27	
21	100	-0,37	-18,64	-9,68	-16,36	18,27	6,68
22	105	-0,82	-18,27	-9,65	-16,1	17,45	6,45
23	110	-1,3	-16,83	-9,66	-15,87	15,53	6,21
24	115	-0,83	-15,46	-9,78	-15,81	14,63	6,03
25	120	-2,32	-14,99	-10,02	-15,91	12,67	5,89
26	125	-2,6	-14,27	-10,22	-16,02	11,67	5,8
27	130	-4,14	-13,18	-10,41	-16,18	9,04	5,77
28	135	-4,38	-11,91	-10,64	-16,15	7,53	5,51
29	140	-5,34	-10,86	-10,78	-16,13	5,52	5,35
30	145	-6,13	-10,07	-10,93	-15,89	3,94	4,96
31	150	-6,53	-9,41	-11,57	-15,19	2,88	3,62
32	155	-6,69	-8,84	-12,09	-15,83	2,15	3,74
33	160	-6,81	-8,35	-12,63	-15,34	1,54	2,71
34	165	-7,44	-8,13	-13,5	-15,07	0,69	1,57
35	170	-8,04	-7,78	-14,01	-14,79	-0,26	0,78
36	175	-8,71	-7,59	-14,58	-14,64	-1,12	0,06
37	180	-8,45	-7,52	-15,14	-14,75	-0,93	-0,39
38	185	-8,26	-7,59	-15,33	-14,64	-0,67	-0,69
39	190	-8,42	-7,4	-15,23	-14,08	-1,02	-1,15
40	195	-8,91	-7,02	-15,25	-13,82	-1,89	-1,43

41	200	-9	-6,39	-15,54	-11,96	-2,61	-3,58
42	205	-9,47	-5,92	-15,96	-11,22	-3,55	-4,74
43	210	-10,4	-5,53	-16,54	-10,54	-4,87	-6
44	215	-11,36	-6,03	-16,55	-10,44	-5,33	-6,11
45	220	-12,68	-5,14	-15,33	-10,15	-7,54	-5,18
46	225	-13,81	-4,71	-15,07	-10,26	-9,1	-4,81
47	230	-14,78	-5,03	-14,71	-10,34	-9,75	-4,37
48	235	-15,87	-3,42	-14,52	-10,48	-12,45	-4,04
49	240	-16,31	-3,34	-14,23	-10,62	-12,97	-3,61
50	245	-17,24	-2,28	-14,1	-10,59	-14,96	-3,51
51	250	-19,01	-1,4	-13,9	-10,61	-17,61	-3,29
52	255	-20,77	-1,03	-13,71	-10,57	-19,74	-3,14
53	260	-19,16	-0,86	-13,49	-10,51	-18,3	-2,98
54	265	-15,96		-13,19	-10,46	-15,96	-2,73
55	270	-15,28	-0,81	-12,75	-10,37	-14,47	-2,38
56	275	-14,69	-1,07	-12,39	-10,3	-13,62	-2,09
57	280	-15,02	-0,63	-12,08	-10,23	-14,39	-1,85
58	285	-15,92	-0,35	-11,93	-10,2	-15,57	-1,73
59	290	-15,78	-0,58	-11,8	-10,29	-15,2	-1,51
60	295	-15,32	-0,09	-11,65	-10,4	-15,23	-1,25
61	300	-14,29	-0,65	-11,54	-10,51	-13,64	-1,03
62	305	-14,03	-0,02	-11,53	-10,55	-14,01	-0,98
63	310	-13,52	-0,51	-11,54	-10,53	-13,01	-1,01
64	315	-13,4	-1,1	-11,61	-10,49	-12,3	-1,12
65	320	-12,66	-1,79	-11,42	-10,45	-10,87	-0,97
66	325	-11,94	-2,06	-10,88	-10,48	-9,88	-0,4
67	330	-11,23	-1,51	-11,15	-10,36	-9,72	-0,79
68	335	-10,22	-1,58	-10,85	-10,22	-8,64	-0,63
69	340	-8,82	-2,06	-10,79	-10,2	-6,76	-0,59
70	345	-7,76	-2,95	-10,57	-10,24	-4,81	-0,33
71	350	-6,43	-4,37	-10,32	-10,38	-2,06	0,06
72	355	-5,89	-4,99	-10,69	-10,49	-0,9	-0,2
73	360	-5,23	-6,19				

Appendix IV – Spectral ILD, Diagram 1-4

Diagram 1. Spectral ILD, Session B, v = 100°

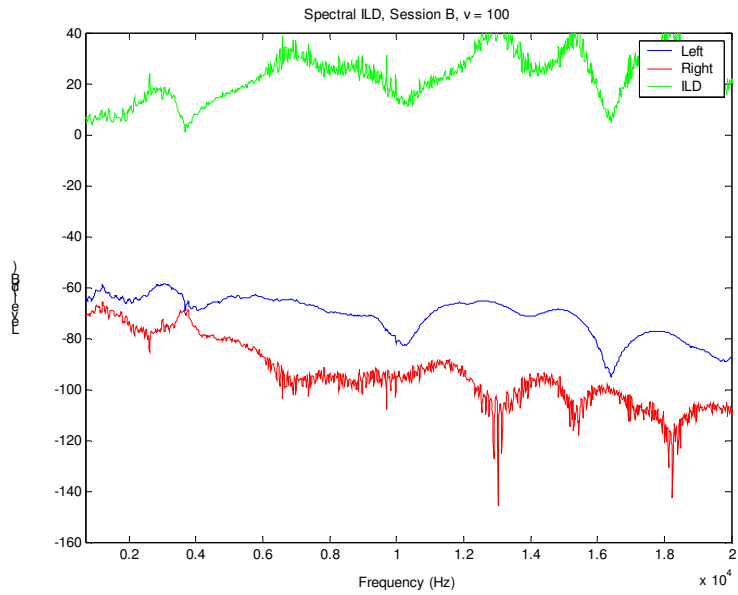Diagram 2. Spectral ILD, Session B, v = 255°

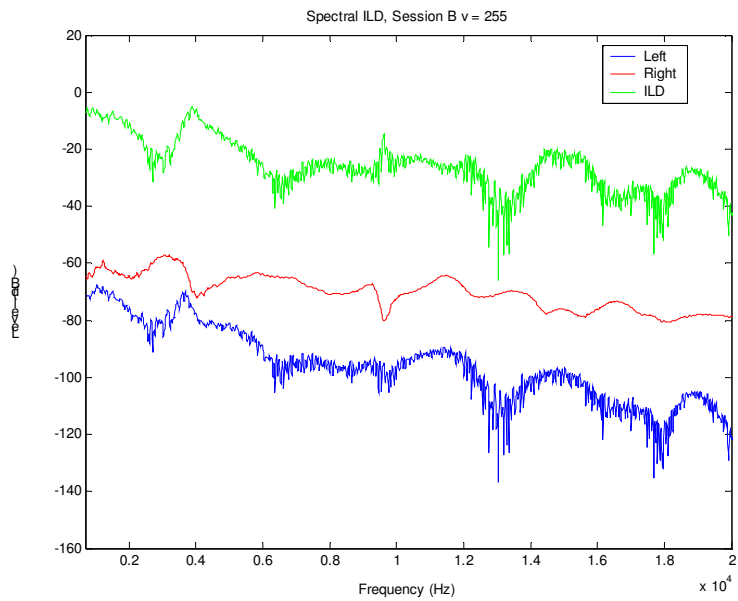Diagram 3. Spectral ILD, Session C, v = 000°

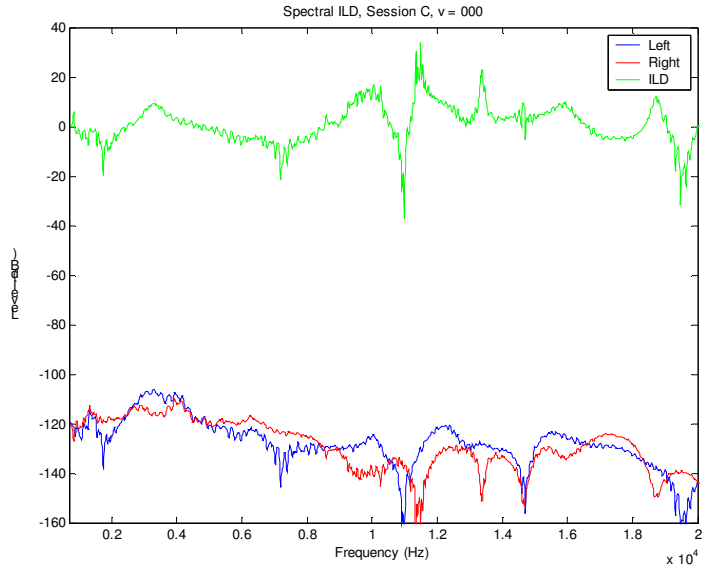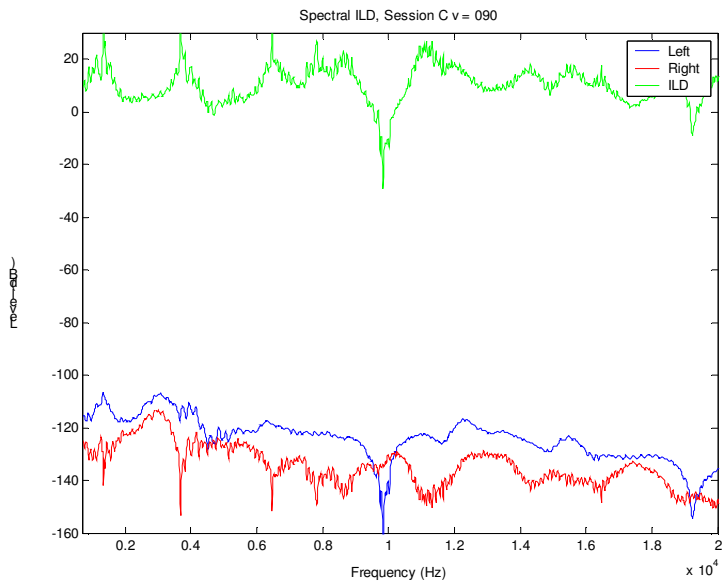Spectral ILD, Session C, v = 000

Diagram 4. Spectral ILD, Sessions C, v = 090°

Spectral ILD, Session C v = 090

References:

[1] Rumsey F., Spatial Audio, *Focal Press, 2001*.

[2] Blauert J., Spatial Hearing, 3rd printing 2001*, The MIT Press, Cambridge, Massachusetts*.

[3] Mac Cabe C., Furlong D.J, Virtual Imaging Capabilities of Surround Sound System*s. Department of Microekectronics and electrical Engineering, Trinity College, Dublin 2, Ireland. J. AudioEng. Soc., Vol. 42, No.1/2, 1994 January/February.*

[4] Shinn-Cunningham B., Localizing Sound in Rooms. *Department of Cognitive and neural Systems and Biomedical Engineering, Boston University. Proceedings of the ACM SIGGRAPH and EUROGRAPHICS Campfire: Acoustic Rendering for Virtual Environments,Snowbird, Utah, 26-29 May 2001, 17-22.*

[5] Malham D., Spatial Hearing mechanism and Sound Reproduction*,University of York*, 1998
ht.tp://www.york.ac.uk/inst/mustech/3d_audio/ambis2.htm

[6] Moore B., An Introduction to the Psychology of Hearing, *Academic Press 1996*

[7] Watkinson J., The art of Sound Reproduction, *Focal Press, 1998*

[8] Gerzon M., Barton, G. Ambisonic decoders for HDTV, 1993. *92nd AES Convention, Vienna.Audio Eng. Society.*

[9] Sveriges Radio – Radio Sweden http://www.sr.se/multikanal/index.stm

[10] Malham D., Anthony M., 3-D Sound Spatialization using Ambisonic Techniques. *Music technology Group, Department of Music, University of York, UK. Computer Musac Journal, 19;4, pp.58-70, Winter 1995.*

[11] http://www.soundfield.com/b_forum.htm, The Soundfield Microphone.

[12] Gerzon M. A.,Multi-system Ambisonic Decoder, *Mathematical Institute, Oxford. Wireless World 1977*

[13] Gerzon M. A., Ambisonics. Part two: Studio Techniques, Studio Sound, 1975

[14] http://www.geocities.com/ambinutter/Decoder.html List of Decoders

[15] Gerzon, M., Practical Periphony, the Reproduction of Full-Sphere Sound, *Mathematical institute, University of Oxford, Oxford OX1 3LB, England.*

[16] Gerzon, M., Design of Ambisonic Decoders for Multispeaker Surround Sound, *Mathematical institute, University of Oxford, Oxford OX1 3LB, England. 58th Audio Engineering Society Convention, New York, 1977*

[17] Gerzon, M., The Optimum Choice of Surround Sound Encoding Specification, *Mathematical institute, University of Oxford, 56th Audio Engineering Society Convention, Paris, Paris 1977.*

[18] http://www.geocities.com/ambinutter/SHELF_FILTERS.html

[19] Malham, D. G., Homogeneous and nonhomogeneous surround sound systems, Department of Music, University of York. http://www.york.ac.uk?inst/mustech/3d_audio/homogeneous.htm

[20] Nico, R., and Emerit, M., Reproducing 3D-Sound for Videoconferencing: a Comparison between Holophony and Ambisonic, *France Telecom CNET Lannion.* http://www.iua.upf.es/dafx98/papers/ROZ31.PS

[21] Poletti, M. A., AES Member. A unified Theory of Horizontal Holographic Sound Systems*, Industrial Reasearch Limited, Lower Hutt, New Zealand. J. Audio Eng. Soc. Vol. 48, No. 12, 2000.*

[22] http://www.knowlesacoustics.com/

[23] Hirvonen, T., Headphone Listening Methods*, Helsinki University of Technology, Department of Electrical and Communication Engineering, 2002.* http://www.acoustics.hut.fi/publications/files/theses/hirvonen_mst.pdf

[24] *Manual SREV 1 Sampling Guide* *http://www.yamahaproaudio.com/products/signal_pro/sampling/srev1*

[25] http://www.ymec.com/products/app/impulse2.htm

[26] MOTU 828 Fire Wire Audio Interface http://namm.harmony-central.com/WNAMM01/Content/MOTU/PR/828.html

[27] http://music.york.ac.uk