Spatial Impression in Multichannel Surround Sound Systems

Jonathan Mark Hirst

Research Institute for the Built and Human Environment
University of Salford UK

Submitted in Partial Fulfilment of the Requirements of the Degree of Doctor of Philosophy. 2006

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANOVA	Analysis of variance
ASW	Apparent source width
BSI	Background spatial impression
CLL	Composite loudness level
CSI	Continuous spatial impression
DI	Index of preference field difference
DTS	Digital theatre systems
DVD-A	Digital versatile disc for audio
ERB	Equivalent rectangular bandwidth
ESI	Early spatial impression
HATS	Head and torso simulator
HRIR	Head related impulse response
HRTF	Head related transfer function
IACC	Interaural cross-correlation coefficient
IACCFF	Interaural cross-correlation fluctuation function
ICCC	Inter channel cross-correlation coefficient
ILD	Interaural level difference
ITD	Interaural time difference
LCRS	Left, centre, right and surround
LEV	Listener envelopment
LF	Lateral energy fraction
LFE	Low frequency effects
LG	Late lateral sound level

LSRD	Least significant rank difference
MLS	Maximum length sequence
MOA	Method of adjustment
PGIACC	Perceptually grouped interaural cross-correlation coefficient
SACD	Super audio compact disc
SDSS	Sony dynamic digital sound
SR	Spatial retention
VCA	Voltage controlled amplifier
VCP	Voltage controlled panner

Acknowledgements

I would particularly like to thank my parents for all their support throughout my studies.

I gratefully acknowledge the support and help I have received from the staff of the Acoustics, Audio and Video Discipline of the School of Computing, Science and Engineering, University of Salford. In particular, special thanks are owed to my supervisors, Dr. Bill Davies and Pete Philipson, whose help, support, guidance and input over the years has been invaluable. I would also like to particularly acknowledge the help I have received from Francis Wooff and Dr. Phil Duncan. The time and patience of the long-suffering participants in the subjective tests is also gratefully appreciated.

I am also thankful for the financial support provided by the Acoustics, Audio and Video Discipline of the School of Computing, Science and Engineering, University of Salford and the EPSRC.

The thesis is dedicated to the memory of my father.

Abstract

Spatial impression in both concert halls and reproduced sound has been identified as an important attribute of the listening experience. In this study, the synthesis and objective measurement of spatial impression in reproduced sound is examined.

A novel, multichannel spatializing technique for musical synthesis has been developed that entailed the separation of the individual harmonics of a musical note that were spatially distributed over multichannel surround systems. Subjective testing of the techniques revealed that the perceived degree of spatial impression significantly increased as the angular spread of harmonics increased, however, extending the spatial spread beyond 90° did not significantly increase the perception of spatial impression.

The concert hall measure of spatial impression, the interaural cross correlation coefficient (IACC) was used to objectively measure the effects of the spatializing techniques. The IACC measurements displayed a strong correlation to the subjective results. Further examination of the IACC measurement indicated the possibility of it’s adaptation to multichannel surround sound in general.

A method of adapting IACC to reproduced sound was further developed that involved comparing IACC measurements taken in a concert hall to IACC measurements taken in reproduced versions of the same concert hall. The method was first conducted as a simulation using basic auralisation techniques. Real concert hall measurements and reproduction systems were then employed. Results showed that the method was able to discriminate between the spatial capabilities of a number of different surround sound systems and rank them in a predictable order. The results were further validated by means of a subjective test.

In an attempt to sensitise the IACC measurement, the frequency dependency of IACC was investigated by means of a subjective test. The results indicated that a perceptually more accurate indication of spatial impression may be gained by applying a frequency-dependent weighting to IACC measurements. This may be useful in the spatial measurement of both reproduced sound and concert halls.

1 Introduction

1.1. Introduction

In this section, the research area is introduced and the research aims stated. The way in which the thesis is structured is outlined and the contributions to the research field are identified.

1.2. Spatial Audio and Spatial Impression

Since the invention of stereophonic sound reproduction, an extra sense of realism in the experience of listeners has been present. The added realism is due to the inclusion of spatial information in the audio reproduction. The listener is capable of perceiving different sound sources as emanating from different positions and experiencing a sense of the environment in which the recording was made.

As spatial audio systems have evolved since the early days of stereo, the (extent of the) sound field delivered by reproduction systems has extended to ‘surrounding’ the listener in both the horizontal and vertical planes. To fully achieve this, one approach is to surround the listener with a huge array of loudspeakers, each fed with its own audio channel. As this is not a practical approach, spatial audio systems usually rely on other methods to produce 'phantom images' which are sound sources that emanate from a point in space where there is not a loudspeaker. For some spatial audio systems, the goal is not to fully 'surround' the listener, rather to create a spatial effect, where only some of the aspects of an original sound field are reproduced.

As multichannel spatial audio systems become more commonplace, the ways in which the systems can be optimally utilised and the ways in which the spatial capabilities of different reproduction systems or recording methods can be measured has arisen.

With the advent of music-only surround sound formats, such as DVD-A and multichannel SACD, new potentials for spatial processing in music technology have occurred. In what ways can musical synthesizers take advantage of the spatial possibilities offered by surround sound systems? How can these technologies be utilised to create either a spatial effect or spatial realism?

Following on from this, the question arises of how effective are these spatializing techniques for musical synthesis in producing a spatial effect or spatial realism. Is it possible to measure the degree of spatial impression delivered by the techniques? This question could also be extended to spatial audio systems in general. One way of determining the degree of spatial impression delivered by reproduction systems is by means of subjective testing. This involves a time consuming process whereby a panel of subjects evaluate the spatial performance of a number of systems. A more efficient and perhaps consistent method of evaluation entails objective measurement.

Studies involving objective measures of spatial impression in concert hall acoustics have benefited from many years of research. Important contributors include Barron and Marshall [Barron and Marshall 1981] and Ando [Ando 1985]. These researchers have developed objective measures of spatial impression in concert halls that have been advanced and substantiated through psychoacoustic analysis. The objective measurements strongly correlate to the subjective experience of spatial impression brought about mainly by the design and architecture of the auditorium.

In reproduced sound, some of the spatial capabilities of a system may be assessed by the accuracy in which the system can position a sound in space (localization capabilities). However, as was pointed out by Rumsey [Rumsey 1998], localization measurements may not fully represent the spatial experience delivered by a reproduction system. A more appropriate measure of spatial impression in reproduced sound may be achieved by adapting the existing concert hall measures of spatial impression.

1.3. Research Aims

In general, the aims of the research were to investigate spatial impression in multichannel, reproduced sound by gaining knowledge of auditory perception, the technical issues in surround sound systems and objective and subjective measures of spatial impression. In particular the research aims were:

- To investigate methods by which a spatial ‘effect’ could be implemented in the multichannel reproduction of a synthesized sound.
- To develop an objective measure of spatial impression for use in both the aforementioned musical synthesis techniques and in reproduced sound in general.
- To align the objective measure to auditory perception through subjective testing.
- To sensitise existing objective measures of spatial impression by applying a frequency weighting.

1.4. Thesis Structure

The thesis has been structured in the following manner:

In Chapter 2, a literature review is presented that examines the subjective experience of spatial impression, spatial audio systems and the synthesis of spatial impression, concert hall measures of spatial impression and the adaptation of these measures to reproduced sound. Particular attention is paid to the descriptions of the spatial audio systems used extensively in experiments described in latter chapters of the thesis. The concert hall measurement of spatial impression, the interaural cross correlation coefficient (IACC), is also examined in detail and its possible adaptation to reproduced sound discussed.

In Chapter 3, a novel method of synthesizing spatial impression in musical synthesis through multichannel sound systems is reported. The method involved decomposing a musical note into its individual harmonics then distributing the harmonics over the loudspeakers of a circular array to produce a spatial effect. The effectiveness and limits of the techniques were evaluated by a subjective test.

Chapter 4 reports upon a method of objectively measuring the spatial capabilities of surround sound systems. The method is initially conducted as a simulation, then in a novel treatment, using real environments and reproduction systems. The method involves the comparison of IACC measurements taken in a real concert hall to those taken in a reproduced version of the same concert hall. The results of the experiment are discussed and varying methods by which the objective measurement is calculated are introduced.

In Chapter 5 the results obtained from the objective measurements are corroborated by means of a subjective test. The experiment involved the comparison of a number of surround systems in terms of their spatial capabilities. The results of the experiment are discussed and correlated to the results of the objective measurements.

In Chapter 6 the IACC measurement itself is examined. This involved the sensitising of the measurement by the application of a frequency weighting. A novel method of ascertaining the frequency weighting was achieved by the use of a custom designed mixing device in a subjective test. The results of the experiment are discussed and a method of applying the frequency weighting is proposed.

The thesis is summarised in Chapter 7, where the main findings of the research are reported, conclusions drawn and possibilities for further work discussed.

The end matter of the thesis follows Chapter 7. This entails the appendices and the list of references.

1.5. Contribution to the Research Field

The research undertaken for the thesis has resulted in a number of possible applications for the findings, some of which are novel. The investigation into spatializing a synthesised musical sound over a multichannel sound system has shown that a spatial ‘effect’, can be realised through use of the techniques. The basis for the techniques has been discussed in terms of psychoacoustics and may offer original areas for further research. This novel approach to spatial synthesis may be of interest to synthesizer developers as multichannel, surround sound consumer formats become more common.

An objective measure of spatial impression in reproduced sound has been developed. The measurement techniques were initiated as a computer simulation and then as a novel physical procedure, involving the comparison of concert hall spatial measurements to spatial measurements taken in reproduced versions of the same concert hall. Through subjective corroboration, the measurement techniques have been shown to successfully predict the degree of perceived spatial impression in reproduced sound.

Existing concert hall measures of spatial impression have been investigated by means of a controlled subjective test, involving a novel mixing device. The results of the test were analysed and an innovative refinement to the spatial measure was proposed based upon the frequency-dependent findings of the subjective test. The refined measure may be of use in both concert hall and reproduced sound measurement.

2 Spatial Impression in Concert Halls and Reproduced Sound

2.1 Introduction

Spatial impression has been recognised as an important attribute of concert hall acoustics for over forty years. Consequently, various methods for optimising and subjectively and objectively measuring spatial impression have evolved. With the advent of multichannel audio reproduction systems, moves towards enhancing the degree of perceived spatial impression in surround sound systems and developing objective measures of spatial impression in reproduced sound are underway.

The chapter commences with a number of descriptions of spatial impression in both concert halls and reproduced sound. Inherent in this discussion are brief insights into how spatial impression arises and how the auditory system determines spatial environments.

Spatial audio systems that strive to reproduce a natural or realistic sound field are extensively reviewed in terms of psychoacoustics and the technologies used to reproduce a spatial sound field. Ways in which spatial impression (or a spatial ‘effect’) may be created or enhanced in reproduced sound using signal processing or sound synthesis techniques are also examined.

Objective measures of spatial impression in concert halls are reviewed, with particular reference to the interaural cross correlation coefficient (IACC). The chapter concludes with a survey of existing objective methods of measuring spatial impression in reproduced sound and the possible adaptation of the concert hall measurement, IACC, to reproduced sound.

2.2 Perception of Spatial Impression

In order to create or measure spatial impression, a definition, or an understanding of what the perception of spatial impression actually is, needs to be established. The expressions ‘Spatial Impression’, ‘Spaciousness’, ‘Diffuseness’, and ‘Envelopment’ have been interpreted differently by various authors with some authors making a distinction between the above terms and others further dividing the expressions into sub-expressions. In general, spatial impression can be thought of as the auditory systems’ interpretation of information derived from the ear signals, in terms of the size, shape and type of environment a person is in. In anechoic conditions, the auditory system has no information about the environment in terms of its size or shape due to the absence of reflections from the boundaries of the environment. The interpretation of such an environment would be one of little or no spatial impression. In an enclosed and reverberant space, early and late reflections due to the reflective boundaries of the environment provide the auditory system with the necessary cues that help determine that the person is in a space with boundaries. The size and shape of the space may also be determined as the environment has certain spatial attributes or can be perceived as delivering a certain degree of spatial impression.

In concert hall listening, the desired degree of spatial impression is achieved by the careful architectural design of the hall, with early lateral reflections being considered as the most important contributor to spatial impression [Baron 1999]. The following descriptions and definitions of spatial impression were collected from articles relating to concert hall or enclosed space acoustics.

Blauert [1997] lists a number of terms that describe spatial impression and offers a definition as 'Auditory events…..perceived as being spread out in an extended region of space'. An often cited description is reported by Marshall [1967], who quotes an orchestra manager's interpretation of the sensation of spatial impression in concert halls as 'Corresponding to the difference between feeling *inside* the music and looking *at* it, as through a window'. Morimoto and Maekawa [Morimoto and Maekawa 1988] define spatial impression as being 'The width of an auditory event perceived temporally and spatially to be fused with the auditory event of a direct sound in an enclosure like a concert hall'.

Okano *et al* [Okano et al. 1998] went further by dividing spatial impression into three components, 'spaciousness', 'size impression' and 'reverberance'. It is assumed that the latter two terms are descriptors of the listening environment. The term 'spaciousness' is divided again into two subcomponents 'apparent source width' (ASW) and 'listener envelopment' (LEV). ASW is described as the 'apparent auditory width of the sound field created by a performing entity as perceived by a listener in the audience area of a concert hall'. LEV is described as the subjective impression of being enveloped by the sound field and is related to reverberant sound, but apparently not in the same way as the aforementioned perception 'reverberance'.

Griesinger, [Griesinger 1997] proposed that there are three types of spatial impression that are also dependent upon the sound source. Continuous spatial impression (CSI) occurs when early (<10ms) lateral reflected sound combines with a continuous sound source. CSI is dependent on the ratio of medial to lateral sound. Early spatial impression (ESI) results from lateral reflections arriving within 50ms after the end of an impulsive sound or music which consists of short discrete notes. Again, ESI is dependent on the ratio of medial to lateral sound and is not considered to be particularly enveloping. Background spatial impression

(BSI) arises when the source contains short musical notes or speech. The human auditory system assigns the notes or speech to a foreground stream, whilst sound energy that arrives in the gaps between individual notes or words are assigned to another stream, the ‘sonic background’. If these background sounds arrive, approximately 150ms after the direct sound and are spatially diffuse, BSI will occur. BSI is dependent on the amount of spatially diffuse reverberant energy and the level of the source.

In subjectively evaluating the spatial performance of audio reproduction systems, experimenters often require the subjects to report upon a perceived aspect relating to spatial impression. However, it can prove to be quite difficult to describe or report upon spatial impression in reproduced sound (with the possible exception of localization, where the position of the sound source can be indicted by angular location, pointing or more sophisticated methods). The perception of spatial impression in reproduced sound may be very different to spatial impression in real environments. As the name suggests, a surround sound system may envelop the listener, but this is no guarantee that the delivered spatial impression is accurate or realistic. The realism of the reproduced sound field depends upon the design criteria of the system; the goal may be to attempt to reproduce a real sound field as accurately as possible or the goal may be create a spatial ‘effect’ or a partial reproduction of a real sound field . The following are examples of ways in which experimenters have tried to convey aspects of spatial impression in reproduced sound to subjects taking part in listening tests.

Zacharov *et al.* [Zacharov et al. 1999], in testing the spatial performance of virtual home theatre systems asked the subjects to evaluate 'spatial sound quality', the aspects of which were described by Berg and Rumsey [Berg and Rumsey 1999] as 'Locatedness or localisation of the sound, how enveloping it is, it's naturalness and depth'. Rumsey [Rumsey 1999] asked subjects to compare the spatial impression qualities of a number of two to five channel surround sound synthesis (upmixing) algorithms. Spatial impression was referred to as 'The overall sense of acoustic space created by the reproduction'. In an experiment designed to compare the effects of loudspeaker directivity upon spatial impression, Zacharov [Zacharov 1998] simply asked the subjects 'Do you feel enveloped by sound?' However, the subjects underwent an extensive training period in which they were encouraged to discuss their awareness of spatial aspects. Unfortunately the contents of these discussions were not reported. Fredriksson and Zacharov [Fredriksson and Zacharov 2002] asked subjects to rate the performance of a number of surround sound systems in terms of naturalness. In particular the subjects were asked to consider 'How well could the direction of a sound be discriminated?' and 'Is the sound enveloping/surrounding you or coming from a particular direction?'.

Berg and Rumsey [Berg and Rumsey 1999] and Mason, as reported by Rumsey [Rumsey 2001], developed a detailed method of describing subjective attributes in reproduced sound. This involved firstly characterising spatial impression as two main perceptual areas; 'source' and ‘environment’. Source is further described by ‘position’, ‘dimensions’ and ‘diffuseness’, whilst environment is described by ‘envelopment’, ‘dimensions’ and ‘diffuseness’. These descriptors are then further subdivided.

Clearly, many different definitions and descriptions of spatial impression exist in the available literature. A general recurrent theme appears to be that spatial impression is comprised of attributes that are either source or environment related.

In this thesis the author has generally chosen to follow Barron’s suggestion [Barron 1999], that spatial impression refers to spatial attributes in general whilst other descriptions or definitions form subsets of spatial impression. However, to be more specific, spatial impression can be split in to two broad areas; ‘source’ and ‘environment’ [Rumsey 2001]. ‘Source’ refers to such attributes as position, depth width and focus/diffuseness. ‘Environment’ refers to attributes such as envelopment and dimensions (of the environment).

2.3 Spatial Audio Systems

In this section, the history and development of spatial audio systems are described. The systems include stereo, binaural, ambisonic and cinema/home cinema systems. Some systems are described in more detail than others as this reflects the level of usage and experimental testing of particular systems that are described in later chapters of the thesis.

Different spatial audio systems have different design goals in terms of their delivery of spatial impression. In describing the various spatial audio systems, an indication of the degree of delivered spatial impression of each system will be suggested.

2.3.1 Early Spatial Audio Systems

As reported by Hertz [Hertz 1981], the first use of spatial audio occurred in 1881 at the International Exhibition of Electricity in Paris and was developed by Clement Ader. This involved the placement of a number of spaced microphones in the footlights of the Paris Opera House stage. The microphone outputs were transmitted to the Palis de L'Industrie, three kilometres away where listeners auditioned a pair of microphone outputs (one in each ear) via a pair of telephone receivers. The resulting 'binaural' transmissions of operas proved very popular and paved the way for future spatial audio systems.

2.3.2 Stereo

Following Clement Ader's demonstration in Paris, the next significant development in spatial audio did not occur until fifty years later. In 1931, Alan Blumlein devised and patented a recording and reproduction system that could create a sense of space by means of 'phantom' imaging [Blumlein 1931]. A phantom image is an auditory event that is perceived as emanating from in-between two loudspeakers, at a *non-loudspeaker* location.

Stereophonic reproduction is based upon summing localisation. If the same signal is replayed through both loudspeakers of a standard stereo configuration (where each of the two loudspeakers and the listener are placed at the vertices of an equilateral triangle) a single, fused image is perceived at the centre point in between the loudspeakers. The image can be moved to any position in between the loudspeakers by introducing a time or level difference between the loudspeaker signals. Blumlein demonstrated that due to cross-talk between the loudspeaker signals (i.e. the left loudspeaker signal will be heard in both the left *and* right ears and vice-versa), an amplitude difference between the loudspeakers will result in a phase difference between the ear signals that approximates the phase difference associated with a real source. This is true for frequencies below approximately 800 Hz where interaural phase differences dominate and interaural level differences can be considered negligible.

Blumlein also developed a method of recording that would encode the signals with the appropriate amplitude differences required for stereo loudspeaker reproduction. This was achieved using a matched pair of directional microphones, with an angular spacing of between 90° and 180° and arranged as coincidently as possible.

Alternative stereo microphone techniques include the use of a spaced pair of omnidirectional directional microphones (direct encoding of time/phase differences) and Mid/Side recordings that add and subtract the outputs of coincident omnidirectional and figure of eight microphones to produce amplitude difference encoded signals. The Mid/Side technique forms the basis of ambisonic microphone recording, which is described in detail in Section 2.2.8.

Blumlein stereo has proved to be a successful and popular method of spatial audio reproduction. However, stereo is limited by the angular coverage of the soundfield which cannot extend further than the angle subtended by the loudspeakers at the listening position, which is 60°. If the loudspeakers are extended beyond this angle, phantom images tend to be pulled towards the nearest loudspeaker forming a 'hole in the middle' effect. In order to achieve a wider, more enveloping sound field, different approaches to spatial audio need to be adopted, some of which are discussed in the following sections.

2.3.3 Quadraphonics

Quadraphonics was a commercial surround sound system developed for home use in the early 1970s [Davis 2003]. The quadraphonic loudspeaker configuration is square in shape with the listening position at the centre of the square and a loudspeaker positioned at each corner of the square, resulting in an angular spacing of 90°. Mono signals were amplitude panned between pairs of loudspeakers in attempt to create phantom images. Due to the 90° subtended angle of the frontal loudspeakers, the phantom images tend to be pulled towards the loudspeakers and for lateral phantom imaging, amplitude panning for pairs of loudspeakers to side of the listener cannot be achieved.

Due to the poor spatial performance of quadraphonic systems and difficulties in affordably delivering a multichannel format to the domestic market, quadraphonic systems failed as a commercial surround system.

2.3.4 Multichannel Cinema Surround Sound

The film industry embraced surround sound at an early stage, with Walt Disney's 'Fantasia' being one of the first examples of surround sound being used in the cinema in 1939. More recently, 'Dolby Stereo' has become commonplace in both cinemas and the home. Dolby Stereo has evolved into a number formats but is based on a LCRS (left, centre, right and surround) channel format, where the channels can be either discrete or matrixed [Dolby 1999]. The left and right channels are used in a standard stereo format, whilst the centre channel is usually fed by a mono signal, such as the film dialogue. The centre channel overcomes the problem of non-centrally placed listeners being unable to appreciate central phantom images by providing a real source via a loudspeaker. The surround channel is also a mono signal but can be delivered to two or more rear loudspeakers. More sophisticated Dolby Stereo formats entail extra frontal channels and discrete surround channels.

A number of different cinema surround systems exist including SDDS (Sony Dynamic Digital sound) and DTS (Digital Theatre systems). These systems differ mainly in digital signal coding methods for storage and will generally use a LCRS-based loudspeaker configuration for delivery [Rumsey 2001].

In general, cinema-based surround systems were not designed to produce precise and localizable sound sources from the rear as anything apart from rear ’ambience’ was considered undesirable as this may distract the viewer from the (more important) images on the frontal screen.

2.3.5 5.1 Surround Sound Systems

As an extension to stereo and with a strong heritage developed from cinema surround systems, 5.1 surround systems are the commercial standard for both home cinema and music-only reproduction.

‘5.1’ refers solely to the loudspeaker configuration and does not necessarily imply that a particular encoding and / or decoding method is used with the system. A 5.1 system comprises of a stereo pair, a centre loudspeaker, two rear surround loudspeakers (the ‘5’ components) and a low frequency effects sub-woofer (the ‘.1’ component). This loudspeaker layout, without the sub-woofer, is also known as a 3/2 layout, where the ‘3’ refers to the front loudspeakers and the ‘2’ to the rear loudspeakers. The standard 5.1 (3/2) loudspeaker configuration places the frontal speakers in the usual stereo positions and the rear loudspeakers at an angular position of ± 100° - 120° relative to the listening position.

As pointed out by Rumsey [Rumsey 2001], 5.1 systems were developed to deliver a ‘cinema style’ of spatial reproduction, rather than an accurate representation of a real sound field. The three frontal loudspeakers provide standard stereo reproduction (with a real centre image if required), whilst the rear speakers can be used to introduce ambience and envelopment through lateral reverberation and other means, or weak phantom imaging behind the listener. Phantom imaging between the front and rear loudspeakers (between ± 30° and ± 115°), when using amplitude panning, is particularly poor as has been demonstrated by Philipson *et al.* [Philipson et al. 2002], amongst others.

The low frequency effects (LFE) channel is a spin-off from cinema sound systems and is reserved for material below a frequency of 120 Hz. This does not mean that the main programme material is to be low-pass filtered then routed to the LFE channel, rather the LFE channel is to be used for enhancing special effects such as explosions. Recently, some users of 5.1 systems have suggested *not* using the LFE channel for low frequency sounds, rather a better use of the audio track would be to provide a frontal elevated ‘height’ channel [Miller 2003].

2.3.6 Binaural and Transaural Surround Sound

Standard binaural technology involves the encoding of a source signal using either microphone recording or convolution. The encoded signals are then replayed through headphones so that ideally the listener can be presented with the same ear signals as if they were actually there and experience a realistic, three-dimensional representation of the original sound field.

Microphone encoding involves the use of a real or dummy head. Miniature omni-directional microphones are fitted either to the ear canal entrances or within the ear canals of both ears. The resulting recorded signals will contain cues that are essential to human spatial hearing (Interaural time and level differences and head related transfer functions (HRTFs)) that are not fully entailed in standard stereo recordings [Begault 1994].

Unlike stereo where crosstalk is essential to the process, successful binaural reproduction is dependent upon the complete separation of the left and right signals, therefore headphone reproduction is utilised.

The realism of spatial perception in binaural reproduction can be variable. This is mainly due to variations in individual pinna shapes. The pinna used in the recording may be very similar or dissimilar to the pinna of the listener. This will result in the delivered spatial cues being correctly or incorrectly interpreted by the hearing system of the listener [Begault 1991].

The encoding of binaural signals may also be attained through convolving a mono signal with a pair of head related impulse responses (HRIR). If a set of HRIRs for a number of sound source positions are collected, binaural synthesis, limited by the angular resolution of the HRIRs can be achieved. In a later chapter, binaural synthesis is utilised as a means of auralisation (room simulation).

The limitations of binaural surround sound are that HRTFs are (extremely) individualised which may result in poor reproduction performance or synthesis. Also, the frequency response of the headphones can affect reproduction as the subtleties of HRTFs may be lost in non-flat cases [Moller et al. 1995].

It is possible to replay binaural signals over loudspeakers; this technology is known as transaural audio. The problem that transaural reproduction attempts to overcome is that when binaural signals are played over loudspeakers, crosstalk occurs. The signal intended solely for the left ear will also appear (as a slightly attenuated and delayed version) in the right ear and vice versa. For this reason, crosstalk cancellation filters are implemented that produce an out of phase and slightly attenuated and delayed right ear signal in the left ear channel and vice versa [Griesinger 1989].

2.3.7 Ambiophonics

Ambiophonics is similar to transaural audio in that a form of dummy head recording is used to encode part of the audio and frontal loudspeaker cross talk is minimised upon reproduction [Glasgal 2003]. The 'source' and ‘environment’ aspects of ambiophonics are treated differently. The source is recorded using a type of dummy head that has no pinnae and is baffled to sound arriving from anywhere apart form the frontal regions. Upon playback, the recording is convolved with a concert hall impulse response and fed to lateral and rear speakers to provide ambiance, whilst the ‘dummy head’ recording is cross talk cancelled for frontal playback.

A variation on ambisonics (see next section) uses a similar approach to ambiophonics and is known as B+ format ambisonics [Chen 2001]. B+ format ambisonics uses a standard stereo pair of loudspeakers to provide the ‘source’ and ambisonic reproduction to provide the surrounding ‘environment’.

2.3.8 Ambisonics

Following on from stereo and early cinema sound systems, ambisonics was the first system that attempted to offer ‘realistic’ three-dimensional loudspeaker reproduction. Ambisonics differs in design goals from cinema-based systems as it attempts to reproduce a ‘realistic’ sound field with 360° horizontal localization with the (optional) added dimension of height.

Ambisonics is a two part process in which the directions and amplitudes of sound sources are firstly encoded, then decoded to loudspeaker signals, where the loudspeaker configuration can take any form from stereo to 360° *pantophonic* (horizontal only) reproduction through to full sphere *periphonic* (with height) reproduction. The underlying principles of ambisonics were discovered independently, in the early 1970s by Cooper and Shiga [Cooper and Shiga 1972] and Gerzon [Gerzon 1973].

2.3.8.1 Ambisonic Encoding

Ambisonic encoding can be realised synthetically, by processing a mono signal so that the required directional information is present, or by recording, using a Soundfield microphone [Farrer 1979a and 1979b]. A Soundfield microphone consists of four coincident directional microphones in a tetrahedral arrangement that can be combined to produce the required ambisonically encoded signals. No consideration of the reproduction system (decoding) is required at the encoding stage.

The encoding process ‘places’ sound sources on the surface of, or within the surface of a sphere, with the centre of the sphere being the reference point (and ultimately the listening position). The three dimensional position of a source can be described with increasing accuracy by increasing the order of the system. A zeroth order system will provide only non-directional information, namely, the amplitude of the pressure generated by the source. A first order system will provide directional information by means of the velocity of the source output. A first order system can be realised in terms of left/right (Y), front/back (X) and up/down (Z) coordinates, relative to the centre of the sphere (ambisonics uses a co-ordinate system that is shifted by 90° from the conventional system). These coordinates and their relative gains along with the pressure component are known collectively as B-format signals and form the ambisonic encoding process. The general, first order, ambisonic B-format encoding equations, as cited by Gerzon [Gerzon 1985] are shown below in Equation 2.1.

$$
\begin{aligned}
W &= \frac{1}{\sqrt{2}} \\
X &= \cos(\theta).\cos(\phi) \\
Y &= \sin(\theta).\cos(\phi) \\
Z &= \sin(\phi)
\end{aligned}
$$

Equation 2.1

θ is the azimuth angle and Φ the elevation angle. The 0.707 multiplication of the W component is applied to allow for approximately equal recording levels of all components. For horizontal-only transmission, the Z component can be discarded.

Even greater resolution in directional encoding can be achieved by introducing further directional components. Second order components can be realised using the following R,S,T,U and V equations [Furse and Malham 1999] in addition to the aforementioned zeroth and first order encoding equations. Second order horizontal-only transmission requires five channels (W, X, Y, U and V), whilst periphonic transmission requires nine channels (W, X, Y, Z, R, S, T, U and V). The additional second order components are listed below:

$$
\begin{aligned}
R &= 1.5\sin(\phi).\sin(\phi) - 1.5 \\
S &= \cos(\theta).\sin(2\phi) \\
T &= \sin(\theta).\sin(2\phi) \\
U &= \cos(2\theta).\cos(\phi).\cos(\phi) \\
V &= \sin(2\theta).\cos(\phi).\cos(\phi)
\end{aligned}
$$

Equation 2.2

The directional information encoded in zeroth and first order B-format signals can also be realised in recordings using zero (pressure) and first-order (velocity) microphones. The W component of B-format signals can be captured using an omnidirectional microphone, whilst the X, Y and Z components are captured using figure of eight microphones. X can be thought of as a front/back figure of eight microphone, Y, a left/right and Z an up/down. However, physically arranging an omnidirectional and three figure of eight microphones in a coincident or even near-coincident manner is a challenging operation.To overcome these difficulties an ingenious microphone design was developed by Gerzon and Farrar

[Farrar 1979a and 1979b] and manufactured by Calrec as the Soundfield microphone. The Soundfield microphone achieves the pressure and directional velocity responses by means of four matched and near-coincident sub-cardioid capsules that are mounted on the faces of a tetrahedron. The outputs from the microphones, which are collectively known as A-format, can be combined to create any microphone directional response, including B-format. The capsule configuration can be seen in Figure 2.1.

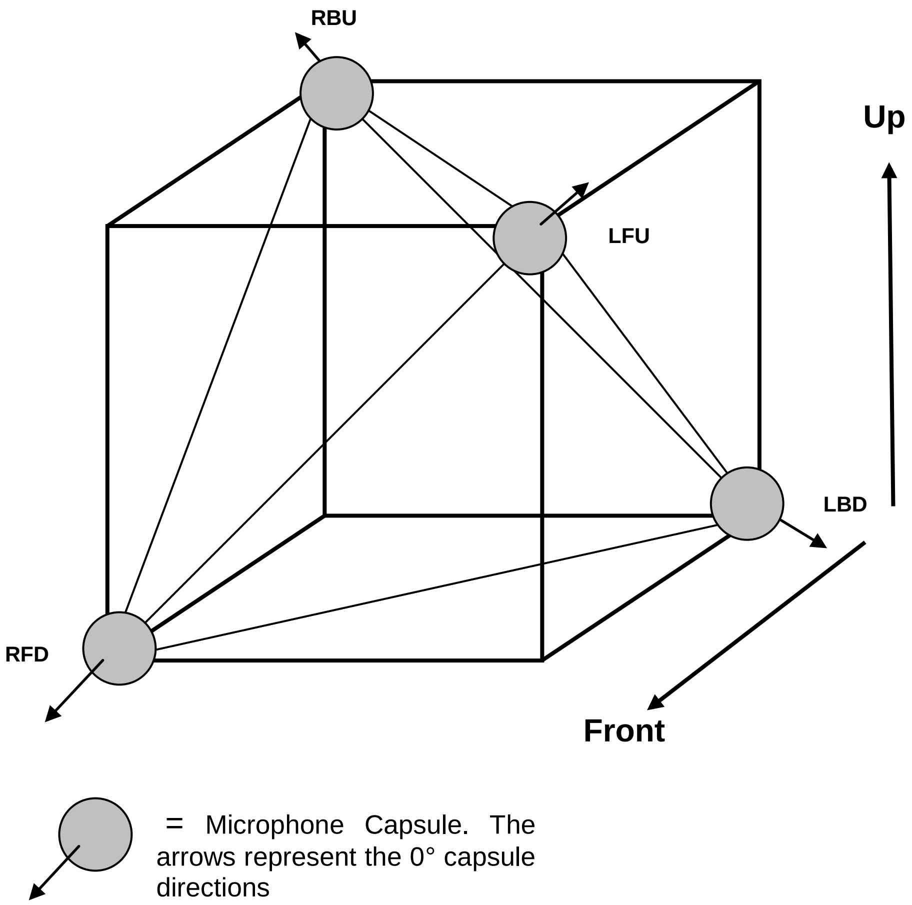


Figure 2.1 Soundfield microphone capsule configuration

The microphone outputs form the A-format signals, where LFU = Left Front Up, LBD = Left Back Down, RFD = Right Front Down and RBU = Right Back Up. To compensate for coincidence errors, due to the microphones being positioned on the faces of a tetrahedron, the microphone outputs are corrected electronically using a control box.

The B-format signals can be derived in the following manner:

$$
\begin{aligned}
W &= LFU + RFD + LBD + RBU \\
X &= LFU + RFD - LBD - RBU \\
Y &= LFU - RFD + LBD - RBU \\
Z &= LFU - RFD - LBD + RBU
\end{aligned}
$$

Equation 2.3

2.3.8.2 Ambisonic Decoding and Reproduction

For reproduction, B-format signals are converted to D-format signals which are the loudspeaker feeds. The decoding of B-format to D-format signals depends on the angular position of a particular loudspeaker, which in turn depends upon the loudspeaker configuration. The loudspeaker set-up can entail any number of loudspeakers, although the minimum number of loudspeakers should be greater than the number of B-format channels i.e. a minimum of four loudspeakers for pantophonic and six loudspeakers for periphonic reproduction [Gerzon 1985]. The loudspeakers should be arranged in a ‘regular’ layout (square, rectangle, cube etc.), where each loudspeaker is equidistant from the listening position and there is an equal angular spacing between each loudspeaker. Ambisonic reproduction through an irregular layout was investigated by Gerzon and Barton [Gerzon and Barton 1992]. In particular, a sophisticated decoder that partially corrected for the irregular layout of reproduction through a 5.1 system (see the next section) was developed. This became known as the ‘Vienna’ decoder and due to the current popularity of 5.1 systems has been realised as a commercial product. The performance of ambisonic reproduction through an irregular (5.1) loudspeaker layout (without Vienna decoding) is assessed subjectively and objectively in later chapters.

As pointed out by Gerzon [Gerzon 1983], the greater the number of loudspeakers the better. If a small number of loudspeakers are used, certain sounds tend to be ‘pulled towards‘ the nearest loudspeaker resulting in ‘speaker emphasis’. This point is investigated in later chapters, where ambisonic systems employing differing numbers of loudspeakers are compared both subjectively and objectively.

Whilst a number of variations in ambisonic decoding equations exist, throughout the work presented in this thesis, Gerzon's first order decoding equation is used [Gerzon 1985] and is shown below.

$$V_{LS} = W + \sqrt{2}\,(X\cos(\alpha) + Y\sin(\alpha) + Z\sin(\beta))$$

Equation 2.4

V_{LS} is the loudspeaker input, where α is the horizontal and β the vertical angle of each loudspeaker. It is the convention in ambisonics to measure the angles in an anticlockwise fashion.

A further enhancement to ambisonic decoding involves splitting B-format signals using phase-matched shelf filters to accommodate the different psychoacoustic mechanisms working above and below 500 Hz. As reported by Farina and Ugolotti [Farina and Ugolotti 1998], Gerzon's patents on Ambisonics include a frequency dependent decoder that is based on Equation 2.4 but includes different weightings at different frequencies. For periphonic decoding, Equation 2.4 is refined to:

$$V_{LS} = G_1 W + G_2\sqrt{2}\,(X\cos(\alpha) + Y\sin(\alpha) + Z\sin(\beta))$$

Equation 2.5

Where gains G_1 and G_2 are:

Frequency Range	G_1	G_2
< 500 Hz	1	$\sqrt{3}$
> 500 Hz	$\sqrt{2}$	$\sqrt{2}$

2.4 Spatial Impression in Musical Synthesis and Signal Processing

For stereo reproduction, a mono sound source, such as a synthesizer sound may be spatially processed to create a stereo ‘effect’. This is usually achieved by adding an ‘effect’ such as artificial reverberation to the sound source. A stereo reverberation processor will emulate a real room by distributing the room reflections over the stereo sound field and decorrelating the stereo signals to add a sense of space and realism. More recently, manufacturers have developed multichannel reverberation devices that take advantage of multi-loudspeaker systems, (such as 5.1 systems) by using multi-directional early reflection patterns. Convolution-based multichannel reverberation is also possible using multi-microphone impulse responses of existing acoustic spaces and convolving these with the input signal.

Other signal processing methods of a mono signal that result in a form of stereo output include stereo delay lines, that can pan the delayed repeating sound to any position in between the loudspeakers. Chorusing involves using a number of comb filters that delay the input signal by a small delay time (usually between 10 and 30 ms). The delay time of each comb filter is slightly varied over time. By using a number of comb filters, with different initial delay times and combining their outputs, an ensemble effect can be achieved, where more than one version of the original signal is perceived. A stereo (or perhaps multichannel) ‘effect’ can be achieved by spatially distributing the outputs of the comb filters over the reproduced sound field.

Synthesizers may incorporate reverberation or chorus devices to produce a ‘stereo’ output, however, other methods can be used to create a spatial effect. A synthesized sound may be formed by the combination of a number of individual sounds or ‘voices’. Each of the individual voices could be panned to different positions to create a sense of space. Similarly, a number of slightly detuned versions of a synthesised sound could be formed and again panned to different positions. These methods of producing a spatial effect for synthesizers could be extended or developed for multichannel use; this partially forms the subject matter of the next chapter.

2.5 Objective Measures of Spatial Impression in Concert Halls

In this section the development and implementation of concert hall measurements of spatial impression are discussed. In particular, the early lateral energy fraction, the late lateral energy fraction and the interaural cross correlation coefficient are examined.

2.5.1 Early Lateral Energy Fraction

Originally, spatial impression in concert halls was thought to depend solely upon reverberant sound. Barron [Barron 1999] reports that in the 1960s, through the works of Damaske [Damaske 1967], amongst others, a link between the directions of arrival of reflected sound and the subjective perception of spatial impression was made. In brief, for spatial impression, the arrivals of sound from the sides and rear of the listener as well as from the source were deemed critical.

Marshall [Marshall 1967] made an important breakthrough in suggesting that early lateral reflections were associated with spatial impression. Following on from this, Baron and Marshall [Baron and Marshall 1981] developed an objective measure of spatial impression by experimenting with a simulated sound field. A multi-loudspeaker set-up utilising a reverberation plate and delay lines was used to simulate the acoustics of a concert hall. The subjects were played a piece of music and could toggle between two conditions. They were then asked to state their preference. The variables in the conditions were, reflection delay time, reflection direction, relative reflection level, number of reflections, reflection spectrum and overall level.

The results showed that spatial impression was strongly related to the relative lateral reflection level, and to a lesser extent, the overall level. Lateral reflections in the 125 to 1000 Hz frequency range were found to be important in the creation of spatial impression, especially the lower frequencies. The arrival time of the reflections was also important. For arrival times between 8 and 90ms, spatial impression was found not to vary greatly. Because of this and the fact that echoes may be perceived at delay times greater than 80 ms, it was proposed that a time window of 5 to 80 ms was the important time period for spatial impression.

From the subjective results, an objective measure of spatial impression, the early lateral energy fraction (LF) was proposed. This can be seen as Equation 2.6. A single LF value is calculated by averaging the measurements in the four octave bands between 125 and 1000 Hz.

$$LF = \frac{\int_{0.005ms}^{80ms} Lateral\ Velocity(t)^2 dt}{\int_{0ms}^{80ms} Total\ \mathrm{Pr}essure(t)^2 dt}$$

Equation 2.6

Lateral velocity is measured using a figure of eight microphone with the null pointing at the source (usually an impulse). The total pressure is measured using an omni-directional microphone. Normally the source of the impulse would be placed on the stage of the concert hall and the microphone(s) in a number of positions in the audience area.

2.5.2 Late Lateral Energy Fraction

In a similar experiment to the above, Bradley and Soulodre [Bradley and Soulodre 1995] conducted subjective tests into perceived listener envelopment (LEV) using simulated sound fields in order to determine objective predictors of listener envelopment. A five loudspeaker semi-circular array was used to simulate the sound fields, in which reverberation time, early to late sound energy ratio (C_{80}), overall level and the angular distribution of the late arriving sound were varied. Subjects were presented with a piece of music and could toggle between two sound fields, one of which was a reference. The subjects then rated the difference in envelopment between the two.

The results showed that the angle of arrival and the level of the late lateral sound energy had the greatest influence on perceived envelopment. The balance between the early and late arriving sound had less influence and the variations in reverberation time had the least influence upon perceived envelopment.

In order to establish an objective measure of listener envelopment, the existing measure, LF was correlated with the results obtained from the subjective tests. Whilst LF correlated well with the results, it did not account for both the relative level and angular distribution of the late arriving sound. The late lateral sound level, LG_{80}, was proposed as a new predictor of listener envelopment. When compared to the subjective test results, LG_{80} correlated the most strongly. LG_{80} can be seen in Equation 2.7.

$$LG = \frac{\int_{80ms}^{\infty} Lateral \quad Velocity_{10m}(t)^2\, dt}{\int_{0ms}^{\infty} Total \; Anechoic \quad \Pr essure_{10m}(t)^2\, dt}$$

Equation 2.7

Lateral velocity is measured at a distance of 10 meters from the source (an impulse) using a figure of eight microphone with the null pointing at the source. The total anechoic pressure is measured at a distance of 10 meters from the source in an anechoic chamber using an omni-directional microphone.

2.5.3 Interaural Cross-Correlation Coefficient

The interaural cross-correlation coefficient (IACC) is a measure of similarity between the signals reaching the left and right ears in a sound field. The less similar (or less correlated) the signals are the greater the perception of spatial impression. In enclosed spaces, decorrelation occurs when a source and one or more reflections are present at the ears of the listener. The reflection(s) are delayed and attenuated versions of the source signal that arrive at the ears at slightly different times thus causing interference and creating dissimilarity between the two ear signals. This dissimilarity gives rise to the perception of spatial impression or in particular, as some researchers have described, apparent source width (ASW).

2.5.3.1 Cross-Correlation

IACC can be measured using a real or dummy head. For concert hall measurements, the impulse response is captured at each ear, with the head facing the source, then processed to extract the IACC [International Standards Organisation 1997]. In the literature the interaural cross correlation function (the maximum value of which is the IACC) is usually expressed in the full and normalised format (Equation 2.8).

$$IACF(\tau) = \frac{\left[\int_{t1}^{t2} x(t)y(t+\tau)dt\right]}{\left[\int_{t1}^{t2} x^2(t)dt\int_{t1}^{t2} y^2(t)dt\right]^{1/2}}$$

Equation 2.8

x(t) and y(t) are the left and right ear signals, τ is an offset between the two signals (usually ± 1 ms) and t1 and t2 are the time limits of the integration (usually 0 and 50 or 80 ms respectively). The offset τ is set to equal ± 1 ms to account for the maximum interaural time difference (due to the spacing of the ears) of an average listener.

2.5.3.2 Graphical Representation of IACC

Figure 2.2 depicts the IACC of the binaural impulse response for a source placed at 45° convolved with pink noise. The binaural impulse response was taken from Gardener and Martin's set of HRIRs [Gardner and Martin 1994], which were recorded using a Kemar dummy head with the microphones placed at inner ends of the ear canal.

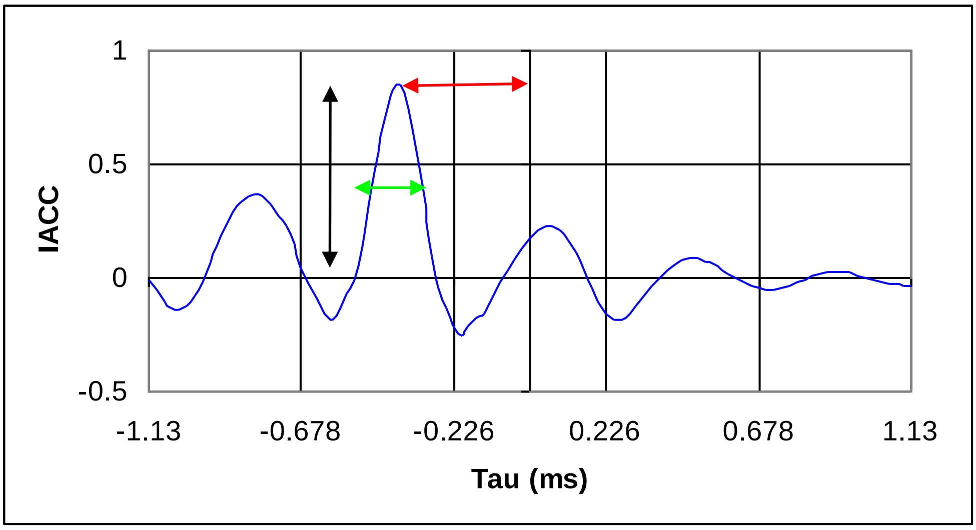
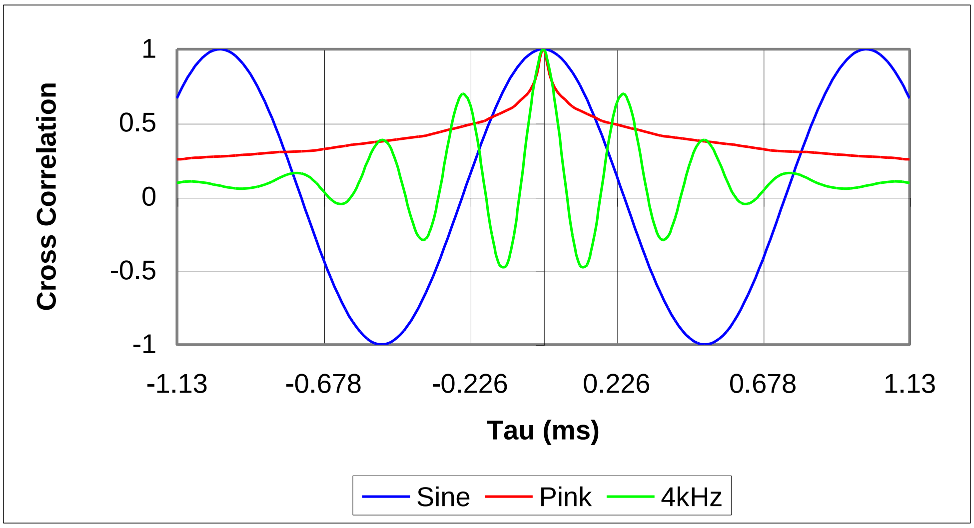


Figure 2.2 IACC of a HRIR for a source positioned at 45°

Highlighted on the graph are some of the important features of the IACC plot. $IACC_{MAX}$ (black arrow) is the maximum value of IACC and is the figure usually quoted as an indicator of spatial impression. τ_{IACC} (red arrow) is the offset time at which $IACC_{MAX}$ occurs. The value of τ_{IACC} is useful in localization studies as it corresponds to the interaural time difference arising due to the horizontal position of the source. W_{IACC} (green arrow) is a relatively new measure that is purported to relate to apparent source width (ASW). In an unpublished paper, Sato and Ando define W_{IACC} as the width of the largest peak within ten percent of the maximum value of the peak.

Example plots of cross correlation as a function of τ, for various sources can be seen in Figure 2.3. The signals used are a 1 kHz sine wave (labelled sine), pink noise (labelled pink) and pink noise with an 18dB boost at 4 kHz (labelled 4 kHz). Whilst these plots are actually autocorrelations (mono signals correlated with themselves), they are equivalent to cross-correlations captured using a pair of spaced omni-directional microphones in an anechoic environment with the source directly in front of and equidistant from the microphones.



Whilst difficult to see in the case of the sine wave, all sources exhibit a maximum cross correlation value of one at 0 ms offset only. This is an expected result of autocorrelation. Whilst the sine wave and pink noise with a 4k Hz boost display lesser peaks occurring at offsets other than zero, the pink noise signal displays only one peak at zero offset. This can be explained by periodicity. Due to the random nature (non-periodic) of pink noise only one peak occurs at $\tau = 0$ ms. In the case of the sine wave, the secondary peaks occur at $\tau = \pm 1$ ms which correspond to the period of a 1k Hz sine wave. Similarly the pink noise with a 4k Hz boost displays secondary peaks at $\tau = \pm 0.25$ ms and tertiary peaks at $\tau = \pm 0.5$ ms, again corresponding to the period of a 4k Hz sine wave. The pink noise with a 4k Hz boost example has been included to demonstrate that signals other than pure tones can exhibit periodicity in cross correlation plots.

2.5.3.3 IACC and the Subjective Perception of Spatial Impression

Initial investigations using cross-correlation to model the spatial aspects of human hearing were conducted by Jeffress [Jeffress 1948] and Sayers and Cherry [Sayers and Cherry 1957]. Jeffress theorised that the coincidence of nerve cell triggering between the two ears could be simulated using cross-correlation to extract interaural time differences. Sayers and Cherry investigated the degree of binaural fusion of a number of sources and found that the maximum value of cross-correlation corresponded to a signal that was perceived as being fused.

By creating broadband headphone signals with varying degrees of IACC, Chernyak and Dubrovsky [Chernyak and Dubrovsky 1968] were able to demonstrate the effects of the degree of correlation upon perceived spatial impression. Subjects were asked to sketch the extent of the perceived auditory event(s) on to a semicircular plan representing the frontal section of the head. The results showed that for a totally correlated signal (IACC = 1) a single and fairly narrow auditory event appeared in the centre of the head. As IACC was decreased the auditory event widened eventually resulting in two separate events appearing at either side of the head when IACC equalled zero.

A similar experiment, utilising loudspeaker signals, was conducted by Plenge [Plenge 1972]. The experiment involved using narrow band noise signals that varied in IACC from 1 to –1 and presented through loudspeakers arranged in the standard stereo configuration. The subjects were asked to draw the spatial extent of the signals with relation to the loudspeaker set up. The results showed that for a highly correlated signal a narrow auditory event was perceived in between the loudspeakers. As IACC was lowered the auditory event broadened and appeared closer to the subjects. When IACC equalled zero the auditory event broadened further with some subjects reporting two auditory events. In terms of distance, the auditory event appeared slightly in front or behind the listener. As IACC became negative, pairs of auditory events were perceived and at IACC = -1, a narrow 'in-head' perception was reported. Similar findings were reported by Kendal [Kendal 1995], in that the width of an auditory event increased with a decreasing value of positive IACC then decreased again as IACC became more negative. The perceived distance between the listener and the auditory event increased as IACC went from 1 to –1.

2.5.3.4 Variations in the Calculation of IACC

As IACC has been researched and developed, a number of variations and improvements to the calculation have evolved. These are outlined below.

Time Windows

The integration time limits, expressed as t1 and t2 in Equation 2.1 are usually set at 0 and 80 ms respectively. These values were selected to correspond to the time window used in the lateral energy fraction. For LF, the upper limit of 80 ms was chosen, as reflections that occur after this time can be perceived as distinct echoes. For an 80 ms window, IACC in terms of perception, is suggested to be an indicator of apparent source width and is termed $IACC_E$. A later and longer time window of 80 to 750 ms, termed $IACC_L$ has been proposed as an indicator of envelopment due to reverberation [Hidaka et al. 1995].

Frequency Range

Another variation in IACC involves the frequency range over which the calculation is made. At frequencies below 500 Hz IACC measurements tend to vary little and at frequencies below 200 Hz rarely fall below 0.8 [Tohyama and Suzuki 1989] which can be attributed to the wavelength of the sound becoming comparable to the ear-to-ear distance. This is demonstrated in Figure 2.4 which displays the IACC extracted from a binaural impulse response taken in a medium sized concert hall. The broadband IACC (represented by the black plot) demonstrates a high degree of decorrelation with an IACC of 0.12. The remaining plots are octave-band filtered IACCs with centre frequencies of 500, 250, 125 and 63 Hz with IACC values of 0.10, 0.62, 0.88 and 0.87 respectively.

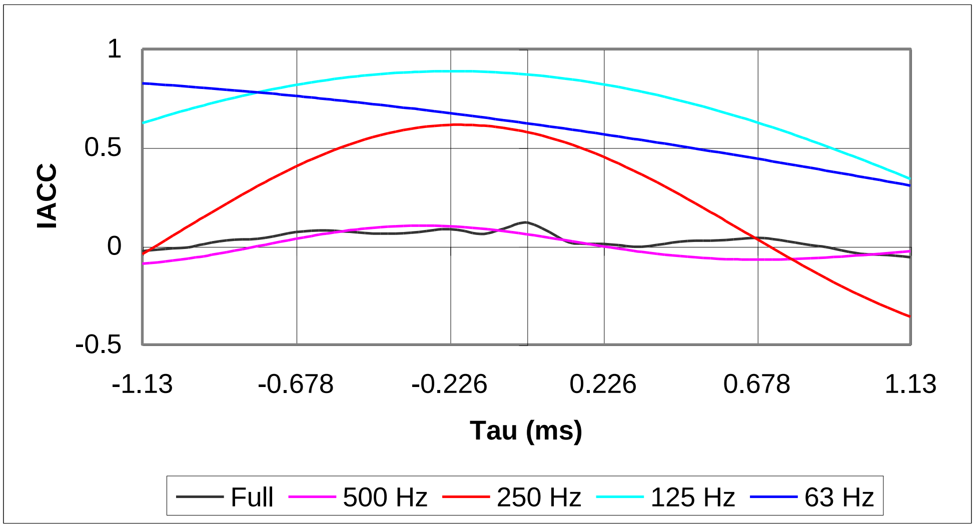


Figure 2.4 Broadband and octave band IACCs extracted from the binaural impulse response of a concert hall.

Hidaka et al. [Hidaka et al. 1995] proposed that the average of IACCs taken from octave bands with centre frequencies of 500, 1000 and 2000 Hz, termed $IACC_{E3}$ resulted in a more reliable measure of ASW (To give a positive correlation with the subjective perception of ASW, this measure is sometimes termed 1-$IACC_{E3}$). This is in partial agreement with the International Standards Organisation (ISO) recommendations that suggest that IACC be measured in octave bands ranging from 125 to 4000 Hz and then averaged [International Standards Organisation 1997]. Regarding frequencies above the upper limit, Blauert and Lindemann

[Blauert and Lindemann 1986] reported that the amount of source broadening perceived by subjects decreased with the rising centre frequency of band-passed noise whilst maintaining IACC at a constant level suggesting that high frequency components do not contribute to the perception of spatial impression.

Level Dependence

Another factor affecting the IACC measurement and perception of spatial impression is the presentation level of the signal. Morimoto and Iida [Morimoto and Iida 1995] presented anechoic music from a frontal loudspeaker along with a pair of reflections simulated by lateral loudspeakers in an anechoic chamber and asked subjects to judge the apparent source width under various conditions. One of the variables introduced was the overall level. By altering the ratio of direct to reflected sound, variations in IACC could be introduced. The results showed that for a presentation level of 50dBA, variations in IACC produced little difference in perceived source width. For a presentation level of 80dBA greater changes in source width were recorded as IACC was varied in the same manner. These results suggest that apparent source width has a dependence upon presentation level and therefore IACC (which is insensitive to differences in presentation level), as a measure of ASW, may not be adequate.

2.5.3.5 Changes in IACC over Time

The IACC measurement usually entails the first 80 ms of an impulse response. However, it could be argued that the auditory system may ‘update’ the perceived degree of spaciousness at regular intervals over time. The way in which IACC varies over time is discussed below.

The binaural impulse response of the medium sized concert hall used in the plots of Figure 2.4 is shown in Figure 2.5. The left ear response is shown in blue and the right in red. The impulse response displays the direct sound followed by a mass of early reflections that eventually form the reverberant tail. Beyond approximately 2.4 seconds the remaining signals could be considered as noise.

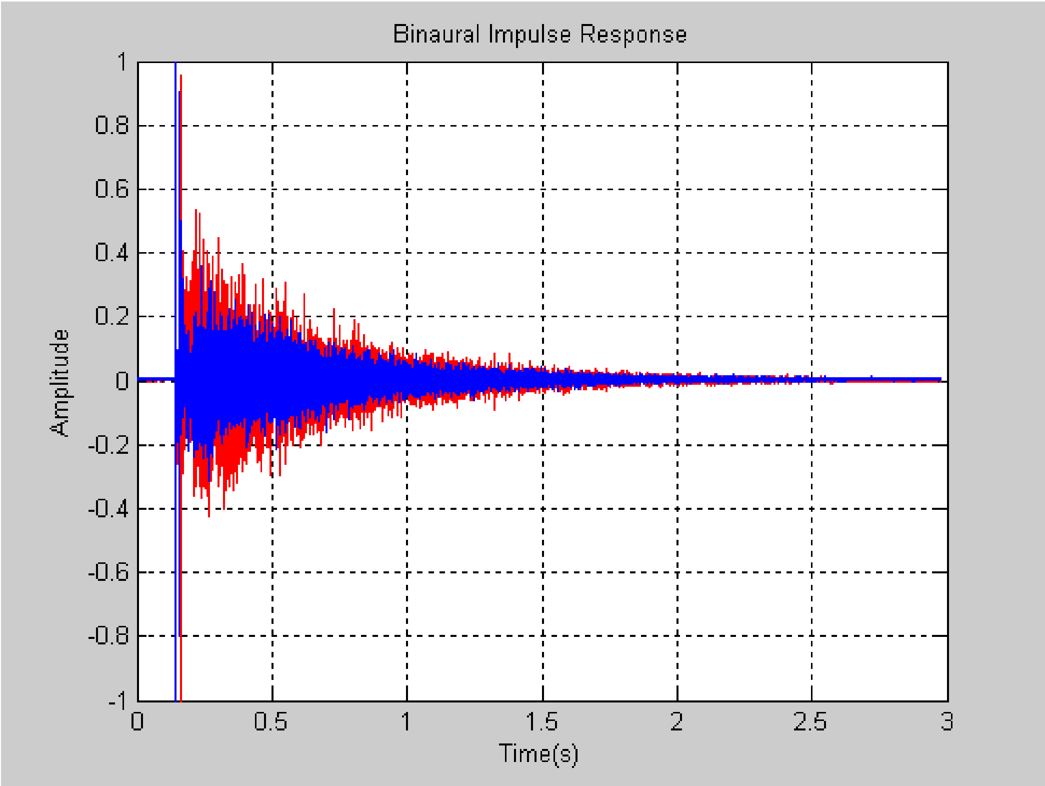


Figure 2.5 Binaural impulse response of a medium sized concert hall

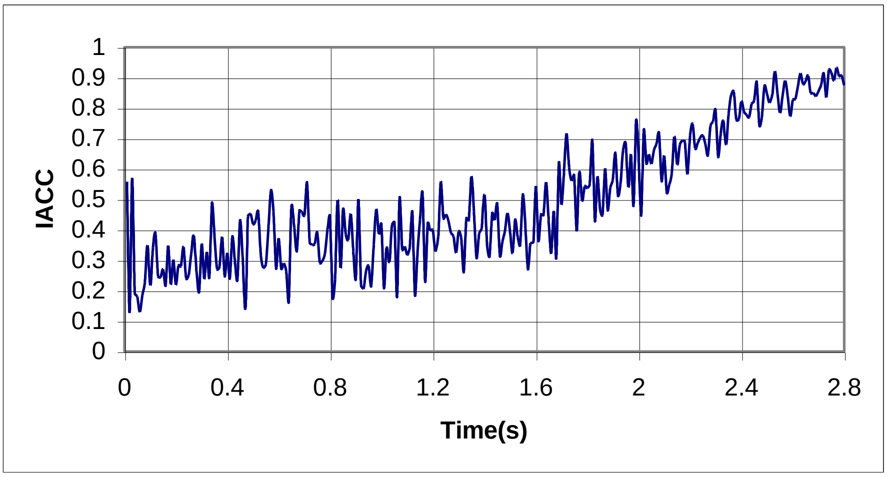


Figure 2.6 Variations in IACC over time of the binaural impulse response shown in Figure 2.5

In Figure 2.6 the variations in IACC over time are shown. This graph was generated by splitting the impulse response into 10 ms sections then taking IACC measurements of each section. Initially, in the presence of the direct sound, IACC remains fairly high then falls rapidly as early reflections arrive causing interference and dissimilarity between the ear signals. A slight rise in IACC is then observed followed by a period between approximately 0.4 and 1.6 s where a comparatively steady state is observed with IACC fluctuating around the 0.4 level. Beyond 1.6 s IACC rises reaching values above 0.9 in the later stages. This high value may be explained by common noise being present in both left and right ear signals.

From Figure 2.6 it can be seen that IACC fluctuates over time, but how is this perceived in terms of spatial impression? There is a considerable variation in the reported temporal aspects of binaural hearing. The ability of the auditory system to perceive variations in the cross-correlation of signals has been shown to deteriorate above a rate of variation of approximately 4 Hz [Grantham and Wightman 1978]. However, in another study [Pollack 1978] Pollack found that the variations could still be perceived up to a rate of approximately 250 to 500 Hz. Either way, it could be proposed that the auditory system utilises a type of sliding average to determine the spatial impression of the environment.

2.6 Objective Measures of Spatial Impression in Reproduced Sound

In assessing the spatial capabilities of audio reproduction systems, an accepted overall objective measure has yet to be developed. In this section, the success of the reproduction system in placing a sound source at an intended location and IACC-based measures of spatial impression are discussed as possible indicators of the spatial capabilities of the system.

2.6.1 Localization

Localization, in terms of the human hearing system refers to the ability to establish the physical location of a sound event through the perception of an auditory event [Blauert 1997]. In other words, how well does our perception of the direction and distance of a sound source compare to the *actual* location of the sound source? In the horizontal plane, the location of a sound event (known as lateralisation) is resolved by the hearing system by utilising interaural time and level differences (ITDs and ILDs) that occur due to non-central sound event signals arriving in the closest ear slightly before and slightly louder than in the furthest ear.

In surround sound systems, as a possible of measure of how well the system performs, a comparison between the ITDs and ILDs of real sources at particular azimuths to the ITDs and ILDs of sources reproduced by the system at the same intended azimuths could be made. This method of testing localization accuracy has been investigated by a number of authors. Mac Cabe and Furlong [Mac Cabe and Furlong 1994] devised a horizontal localization test system to investigate the performance of a number of surround sound systems. The test system measured the interaural level and time differences of ear input signals, by means of a dummy head. Measurements of three virtual reproduction systems were taken and compared to real source measurements. The ILDs were calculated by Fast Fourier transforming the left and right ear signals and finding the ratio between the two. The ITDs were evaluated by determining the IACC, where the ITD is given by the time offset at the maximum value.

Pulkki *et al.* [Pulkki et al. 1999] utilised a similar but more sophisticated binaural auditory model to estimate localisation cues and colouration generated by surround sound systems. HRTFs were used to model the outer ear, a 42 channel band-pass filter bank to model the basilar membrane of the inner ear and half wave rectification and 1 kHz low pass filtering of the filter bank outputs to simulate the hair cell and neural behaviour. By comparing the signals at both ears, IACC was calculated for each filter bank channel. From this, ITDs for each filter bank channel were calculated. Loudness levels at both ears for each channel were summed to form the composite loudness level (CLL) spectrum, from which colouration differences could be observed. From the differences in loudness level spectra between the ears, ILD spectra were formed. Various source production methods (amplitude and time delay panning, anti-phase stereo and HRTF processing) were tested against real sources and also compared to theoretical calculations. The main conclusions drawn were that the model was able to successfully predict a number of known localization phenomena in loudspeaker listening, including an increase in localization error at high frequencies and the fact that phantom images cannot be effectively produced between loudspeakers placed at the sides of a listener. However, the model produced ambiguous cues for the time delay panning and anti-phase stereo cases.

Both of these examples demonstrate that localization measurements may successfully indicate some of the spatial capabilities of reproduction systems, however, localization does not fully describe the listening experience in terms of spatial impression and other measures may need to be utilised.

2.6.2 Fluctuations in ITD

An alternative measure of spaciousness has been suggested that entails the fluctuations in ITD over time. For low frequency variations in ITD, source movement is perceived, however, as the fluctuations increase in frequency, a diffuse and broad source is perceived [Griesinger 1992]. Griesinger further developed these findings to propose the diffuse field transfer as a measure of envelopment in reproduced sound [Griesinger 1998]. The measure involved a refined determination of ITDs over time to establish the presence of important (for spatial perception) fluctuation frequencies and their relative strengths. The diffuse field transfer function, whilst showing possibilities as an objective measure, was not further developed or verified by Griesinger.

Mason [Mason 2001] and Rumsey [Mason and Rumsey 2001] extended this work by using IACC methods for extracting ITDs in a number of frequency bands, then weighting and filtering the output to produce a single figure ITD fluctuation magnitude. Through thorough subjective testing, the rates and magnitudes of ITD fluctuations were investigated. In general it was found that variations in the perceived width of a sound source were due to changes in the magnitude of the ITD fluctuations.

2.6.3 IACC Based Measures in Reproduced Sound

In addition to Mason and Rumsey's IACC-based procedure outlined in the previous section (which was termed the interaural cross-correlation fluctuation function or IACCFF), they have also investigated other objective methods based on the IACC in evaluating spatial attributes in reproduced sound [Mason and Rumsey 2000]. A subjective test investigated the performance of three virtual home theatre processors. The processors attempted to reproduce five-channel surround recordings over two loudspeakers. Subjects were asked to evaluate the systems in a number of spatial attributes. The results of the subjective test were found to be significant. A number of IACC-based measurements were then taken of binaural recordings of the stimuli presented to the subjects. These were IACCFF and IACC in octave bands, all frequency bands and mid-frequency bands. Correlation with the subjective results showed that the IACCFF measurement faired much better than the IACC measurements in predicting subjective evaluations.

Mason also proposed an IACC-based measurement termed the perceptually grouped interaural cross correlation coefficient (PGICCC) [Mason 2001]. The measurement firstly involves separating a binaural impulse response into its source and environment related segments. Each segment is then processed separately. The signals are frequency filtered, then cross correlated in a number of overlapping, consecutive measurements to give a time-varying IACC in each frequency band. The measurements are then weighted by signal strength to reduce noise and inverted. The data could then be analysed in each frequency band and over time. The maximum measurement of all frequency bands and at each point in time is then taken and then the mean and average over time is calculated. The PGIACC was shown to give reasonable results for test stimuli but still requires some development.

The use of IACC in assessing surround sound systems was carried out using a different approach by Furlong [Furlong 1989]. This method entailed comparing IACC measurements taken in an original environment to IACC measurements taken in a reproduced environment, where the reproduced environment was created by differing surround sound systems. The similarities of the IACC measurements in the original and reproduced environments were proposed as an indicator of spatial capabilities of the reproduction system. The results showed that the measurement method was able to discriminate between different systems and rank the systems in an expected order. This work is further discussed and expanded upon in Chapter 4.

2.7 Summary

In this Chapter, descriptions of spatial impression, a review of surround sound systems and objective measurements of spatial impression have been discussed.

Spatial impression is considered a very important feature in concert hall acoustics and its presence can be particularly attributed to the existence of early lateral reflections. The subjective aspects of spatial impression in concert halls can be subdivided into a number categories including apparent source width and listener envelopment. Spatial impression in reproduced sound can be described in terms such as naturalness, locatedness, envelopment and diffuseness.

Spatial audio systems have been present for the past eighty years and have evolved in a number of ways. Blumlein's pioneering work with stereo paved the way, where phantom images could be perceived at locations in between pairs of loudspeakers. Quadraphonics attempted to extend the stereo sound field by surrounding the listener with loudspeakers, however, due to the lack of consideration of psychoacoustics, the system did not work well and was a commercial faliure. Cinema surround sound systems added a new dimension to cinema-goers by incorporating multiple loudspeaker systems that provided a 'real' centre channel and surround channels that emanate from behind the viewer. Home cinema systems followed that allowed for cinema-type surround sound to be auditioned in living rooms. The popular 5.1 system has been utilised in this way and is also becoming the standard reproduction system for music-only surround sound. Binaural and transaural systems attempt to provide the listener with the same signals appearing at the entrance to the ear canal as if they were actually there. This is achieved by providing spatial cues that are essential to human hearing. Ambisonics is a multi loudspeaker system that can provide horizontal-only or three-dimensional (with height) reproduction. Ambisonic recordings can be synthesized or captured using a specialised microphone and replayed on loudspeaker systems incorporating four or more loudspeakers.

Spatial impression in surround sound systems may be synthesized using techniques that create a spatial effect. These include reverberation devices, multi-comb filter devices and panning of synthesizer voices.

Objective measures of spatial impression in concert halls were reviewed that included the early lateral energy fraction, the late lateral sound level and the interaural cross correlation coefficient. IACC was discussed in detail including, the subjective effects of varying IACC, variations in the way it is calculated, frequency dependency and variations in IACC over time. The deficiency of objective measures of spatial impression in reproduced sound was identified. Localization measurements are capable of determining some of the spatial capabilities of surround sound systems, but not all. The rate and magnitude of fluctuations in ITD have been proposed as a possible measure of spaciousness in reproduced sound, but have not been further developed. IACC-based measurements have also been investigated. These measurements involve the adaptation of the IACC to be better aligned to the auditory system by incorporarting critical band filtering and taking into account the way in which IACC varies over time. Another approach was to compare IACC measurements taken in original environments to those taken in reproduced versions of the same environment. The retention of spatial impression, as indicated by the similarity of IACC measurements may be used to gauge the spatial capabilities of surround sound systems.

3 Multichannel Spatializing Techniques for Musical Synthesis

In the last chapter, the perception and measurement of spatial impression, examples of spatializing techniques for musical synthesis and spatial audio reproduction systems were discussed. In this chapter some of these areas are expanded upon in the creation of a novel spatializing technique for musical synthesis.

3.1 Introduction

The audio outputs of a musical synthesizer are usually in the familiar stereo format. Stereo allows for the formation of phantom images in-between the loudspeakers, which gives rise to a limited perception of spatial impression. With the advent of cheaper digital storage, multichannel surround sound systems, which have the potential to produce a greater sense of spatial impression, have become more commonplace. Therefore, it is probable that synthesizers of the future will accommodate multichannel reproduction formats by increasing the number of audio outputs. In this chapter, spatializing techniques that take advantage of the increased number of synthesizer outputs, are developed and subjectively tested In terms of perceived spatial impression.

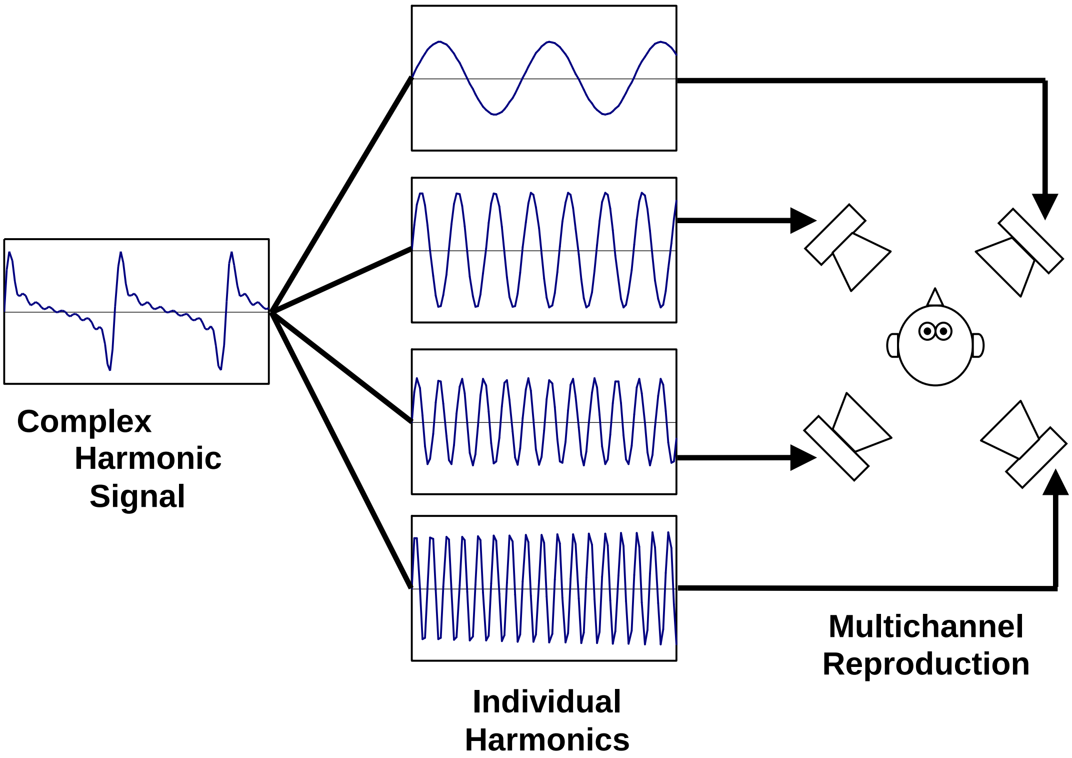


Figure 3.1 Overview of the spatializing techniques for musical synthesis

The chapter focuses on a non-room reflection (reverb) based multichannel spatializing technique for musical synthesis that involves decomposing a complex musical signal into its individual harmonics, then spatially spreading the harmonics over a circular or semi-circular loudspeaker array. This presents the auditory system with a potential perceptual conflict. Due to the spatial separation of the signal, a number of sources may be localised, however, the harmonic relationship and synchronous onsets of the signals provides the auditory system with strong grouping cues [Bregman 1999]. It was theorised that due to this conflict of cues, a spatial effect would be created. An overview of the technique can be seen in Figure 3.1.

The investigation formed two stages; firstly, an informal subjective test was undertaken to gauge the effectiveness of the technique in producing a sense of spatial impression. This was followed by a controlled subjective test that further investigated the spatializing technique utilising both real and virtual sources.

3.2 Spatially Separated Complex Tones

In this section, the perceived outcomes of the spatializing techniques are examined in terms of psychoacoustic theory. Whilst the author is unaware of similar spatializing techniques and directly related theory, the reported theory that follows may offer an insight to the expectations and outcomes of the techniques.

3.3 Localization and the Spatializing Techniques

Localization, in terms of human spatial hearing, refers to the relationship between the physical location of a sound, (the sound event) and the perceived location of the sound event (the auditory event) [Blauert 1997]. The auditory system utilises interaural time and level differences, which arise due to the path difference resulting from a sound event arriving at each ear, to localise a sound. By considering this auditory cue in isolation, it would be expected that the spatializing techniques would result in the perception of a number of individual auditory events. As the spatializing techniques present the auditory system with multiple, harmonically related and temporally coincident sound events, localization cannot be considered as the only cue available to the auditory system.

3.3.1 Perception of Complex Tones

The perception of complex tones has been widely reported and is well summarised by Moore [Moore 1997]. Moore reports on the hearing mechanism's ability to fuse complex tones, which consist of a number of harmonically related partials, into a single percept with a pitch equal to the fundamental frequency of the complex tone. Moore goes on to describe Schouten's work involving 'Residue pitch' or 'Fundamental tracking' [Schouten et al. 1962]. If the fundamental harmonic of a complex tone is removed, the perceived fundamental pitch of signal does not alter. Similarly, if all but few mid-frequency harmonics are removed, the perceived pitch remains the same, however the timbre of the signal is greatly changed.

Two main theories have been proposed to account for the phenomenon of pitch residue. Temporal theories propose that the pitch of a complex tone is related to time intervals between nerve firings emanating from a position on the basilar membrane where two partials are exciting the same critical band. Pattern recognition theories suggest that the complex tone is frequency analysed, then ‘matched’ to a pitch percept relating to the fundamental frequency of the matched pattern [Moore 1997].

The fusion of complex tones and fundamental tracking suggest that harmonically related partials provide the auditory system with a strong grouping cue. In the informal investigation, which is reported later in this chapter, the removal of the fundamental harmonic to observe if pitch residue would still occur when the signal was spatially separated was investigated.

3.3.1.1 Auditory Scene Analysis

In a natural environment, the acoustic energy produced by a number of concurrent sound sources arriving at the ears of a listener is mixed. Following basilar membrane ‘filtering’, the auditory system first analyses this mixture into a large number of frequency components. As extensively reported by Bregman [Bregman 1999], one of the problems addressed by auditory scene analysis is; which combination of frequency components should be attributed to each individual sound source? The analysis, which is dependent upon a number of cues, results in the perceptual fusion or segregation of sound sources and their components. The perceptual fusion or segregation of simultaneous components depends upon similarities or differences in harmonic relationships, regularity of spectral spacing, onset and offset synchrony, binaural frequency matching, parallel amplitude modulation, frequency, spectral envelope, amplitude and spatial location.

3.3.2 Interaction of Auditory Cues

Various auditory cues compete to form the perceptual grouping of sounds, however, cues do not operate in isolation. Interactions occur, with some cues reinforcing each other whilst other cues compete with each other. Of interest to this study are the interactions of onset synchrony, harmonicity and localisation cues.

By means of the rhythmic masking release paradigm, Turgeon [Turgeon 1999] examined the interaction of these cues using spatially separated concurrent complex tones of a short duration. Her main findings were that temporal (onset) synchrony strongly contributes to sound source grouping, whilst spatial separation and harmonicity contributed only weakly or not at all to the perceptual organisation of sounds. This was in partial agreement with earlier work [Buell and Hafter 1991] that suggested that harmonic structure is more important than commonality of spatial position for the grouping of complex sounds.

Regarding spatial impression, it is interesting to note that Turgeon also reports that whilst related sounds coming from different locations in space can be perceived as a single event, they are difficult to locate and can be described as ‘diffuse’.

3.3.3 Experiments Involving Spatially Separated Complex Tones

Previous investigations involving the spatial separation of harmonic signals have concentrated mainly upon psychoacoustics rather than the creation of spatial impression in reproduced sound. Bregman [Bregman 1999] cites a few examples of similar experiments. An unreferenced example involved a demonstration of how the hearing system fuses harmonically related signals. Two sets of partials of a synthesized speech sound, occupying different regions of the frequency spectrum were presented to different ears. When the two sets of partials shared the same fundamental frequency, the signal was perceived as being fused, when the sets of partials did not share the same fundamental frequency, two separate signals were perceived.

Another example cited by Bregman, involved a sound created for a piece of music by Reynolds and Lancino at IRCAM and reported upon by McAdams [McAdams 1984]. An oboe tone was synthesized with the odd and even harmonics separated into two channels, which fed loudspeakers on the left and right of the listener. Frequency micromodulation was applied to both signals. When the frequency fluctuations of the harmonics were synchronized, a single oboe was perceived in between the loudspeakers. When the fluctuations of the harmonics presented in the left loudspeaker were gradually made independent of those in the right loudspeaker, two separate sounds were perceived, one from each loudspeaker. This suggests yet another grouping cue important to spatially separated harmonics.

Bregman also describes an informal experiment he performed with Divenyi. Two harmonic signals were created, one consisting of tones of frequencies of 200, 400, 600 and 800 Hz, the other of frequencies of 300, 600, 900, and 1200 Hz. The signals were presented through headphones, with one signal panned 45° to the left and the other, 45° to the right. The signals were played at irregular intervals so that they overlapped for some of the time, but did not start or finish at the same time. The experimenters expected that in addition to the two complex tones, a third tone at 600 Hz (common to both complex tones) would be perceived at a central location. However, only the two complex tones were perceived. Bregman suggests that this was due to the 600 Hz tones always going on or off in synchrony with one of the complex tones thus accounting for the assignment of the 600 Hz tones to both complex tones, simultaneously.

3.3.4 Expectations of the Spatializing Techniques

These examples seem to suggest that unless there is some correlation between the spatially separated components of a harmonically related signal, fusion of the components will not occur. Harmonically related signals with synchronous onset times (such as the signals used in this experiment) provide the auditory system with a strong grouping cue even if the signals emanate from different spatial locations. In terms of spatial impression, a spatially separated signal is difficult to locate and can be described as diffuse. Therefore, it could be expected that the spatializing techniques would result in the perception of a fused sound, which is also spatially diffuse.

3.4 Description and Outcomes of the Informal Investigation

In an informal investigation, the spatialization techniques were tested in terms of spatial spread of the harmonics, differing sound sources and the inclusion and location of the fundamental harmonic. The purpose of the investigation was to gauge the effectiveness of the techniques and thus determine whether or not to pursue the work further.

3.4.1 Test Design

As the investigation was informal and not designed to allow for a complex statistical analysis, a simple verbal comparison between a test and reference signal was undertaken where the subjects were asked to describe *any* differences they could hear between the two. The reference signal was always (except for Comparison A7) a non-spatially separated, mono version of the test signal that was presented through a loudspeaker directly in front of the subject. Different types of test signals were used that were varied in degree of spatial spread and the position and inclusion (or non-inclusion) of the fundamental harmonic.

There were two stages to the informal test. In the first stage (Stage A), the test signals consisted of harmonics that were spread out over three or five loudspeakers. For the three-loudspeaker presentations different harmonics were routed to loudspeakers situated at 0° and ± 45°, relative to the listening position. The five-loudspeaker presentations included additional loudspeakers at ± 90°. In the second stage (Stage B), six equally spaced loudspeakers were arranged in a full circle around the listener, with loudspeakers positioned at 0°, ± 60°, ± 120° and 180°. Again, harmonics were distributed over the six loudspeakers.

3.4.2 Test Signals

For Stage A, a square wave consisting of a maximum of ten harmonics, with a fundamental frequency of 440 Hz was generated. Each harmonic was 2.5 s in duration and all harmonics started and finished at the same time. 40 ms linear fade ins and outs were applied to each harmonic.

For Stage B, two signals, both with a fundamental frequency of 150 Hz were generated. One signal was a sawtooth wave consisting of 18 harmonics, each with a duration of 2.5 s, the same start and finish time and 40 ms linear fade ins and out. The other signal was a synthesized piano note created using data obtained from Fletcher, Blackham and Stratton's work on piano note analysis [Fletcher et al. 1962]. The piano note was additively synthesized using 18 time varying (in amplitude) partials, but without phase or time varying frequency information. The combined note lasted 2.5s in duration. The original piano note data was collected for a note with a fundamental frequency of 98 Hz, however, to overcome possible low frequency deficiencies of the loudspeakers, the fundamental frequency was raised to 150 Hz.

A further two signals were generated which were identical to the saw-tooth signal and the synthesized piano note, except that the first harmonic (the fundamental) was omitted and a nineteenth harmonic added.

All the test signals were generated using a computer based unit generator (Csound), then inputted into a sample editor as both a combined waveform and as the separate harmonics of each waveform. By generating the waveforms using Csound, control of the frequency, amplitude and envelope of each harmonic was easily achieved. The sample editor allowed the order of test signal presentations to be easily arranged.

3.4.3 Subjects and Test Room Configuration

Ten subjects participated in Stage A and five in Stage B. All the subjects were untrained (although some had previously taken part in listening tests) and were either staff or students of The School of Acoustics and Electronic Engineering, University of Salford.

Both stages of the experiment were carried out in the semi-anechoic room of the School of Acoustics and Electronic Engineering, University of Salford.

In both stages of the experiment, the multichannel digital signals from the sample editor were routed to an ADAT recorder, which acted as a digital to analogue converter. The outputs from the ADAT recorder were connected to active loudspeakers.

3.4.4 Loudspeaker Configuration

For Stage A, five loudspeakers were arranged in an arc around the front of the listening position. All the loudspeakers were placed at a distance of 0.92 m from the listening position and at an equal angular spacing of 45°, with the centre loudspeaker remaining at 0° (directly in front) throughout. A diagram of the set up can be seen in Figure 3.2.

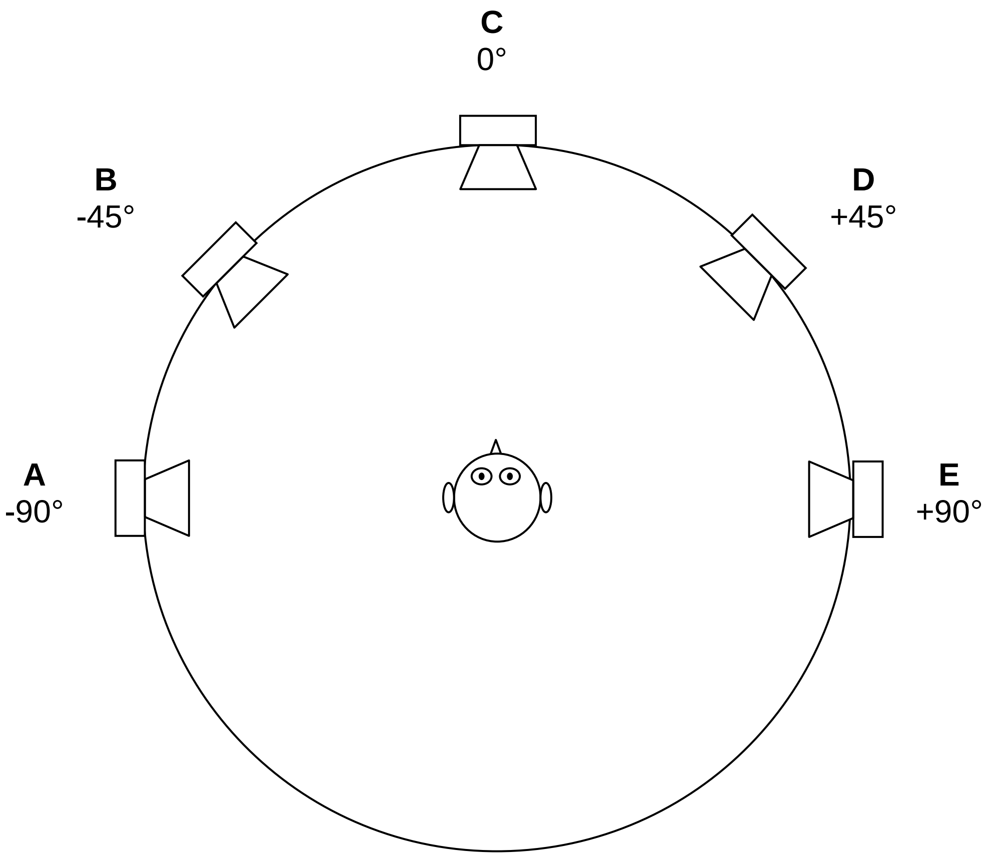


Figure 3.2 Loudspeaker layout for Stage A of the experiment

In Stage B, six equally spaced loudspeakers were arranged in a circle, with the listening position being at the centre. The loudspeakers were again placed at a distance of 0.92 m from the listening position and at an angular spacing of 60°. A diagram of the set up can be seen in Figure3.3.

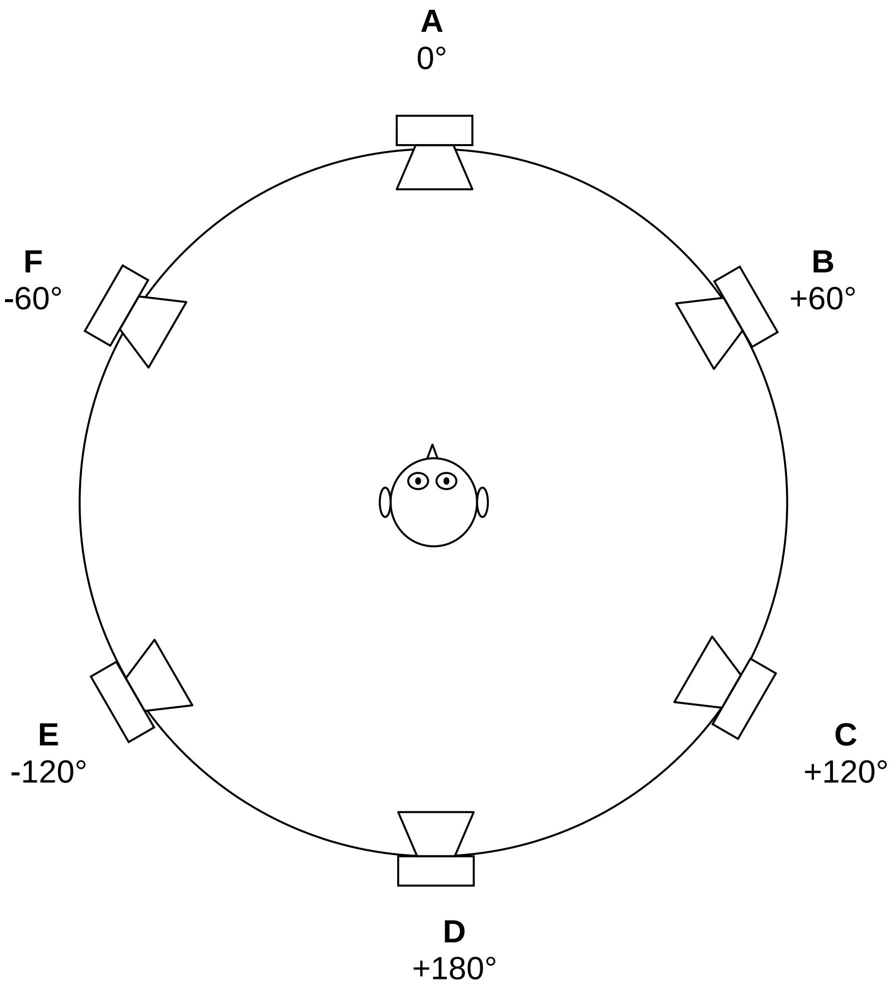


Figure 3.3 Loudspeaker layout for Stage B of the experiment

3.4.5 Experimental Procedure

In both stages of the experiment, the blindfolded subjects were presented with a combined (mono) harmonic signal replayed through the centre loudspeaker (the reference signal), followed by the same signal presented through between three to six equally spaced loudspeakers, with the harmonic components of the signal spread out amongst the speakers in various configurations (the test signal). In Comparison A7, instead of the reference signal appearing in the centre loudspeaker only, it appeared in all loudspeakers.

Each pair of test and reference signals were presented twice, or more times if the subject wished. The subjects were then asked to comment on *any* differences they could perceive between the reference signal and the test signal. In the second stage of the experiment, the subjects were again asked to report on differences and were also asked to verbally report the location of the test signal. The various configurations and spatial spreads of the harmonics that formed the test signals can be seen in Tables 3.1 and 3.2.

Comparison Number		Input of Harmonics to Loudspeakers				
		A	**B**	**C**	**D**	**E**
A1	Reference			1 to 17		
	Test		3, 11,15	1, 7, 13	5, 9, 17	
A2	Reference			1 to 19		
	Test	3, 17	5, 15	1, 9	7, 13	11, 19
A3	Reference			1 to 17		
	Test		1, 7, 13	3, 11,15	5, 9, 17	
A4	Reference			1 to 17		
	Test		5, 9, 17	3, 11,15	1, 7, 13	
A5	Reference			1 to 19		
	Test	3, 17	5, 15	7, 13	11, 19	1, 9
A6	Reference			1 to 19		
	Test	1, 9	5, 15	7, 13	3, 17	11, 19
A7	Reference	1 to 19	1 to 19	1 to 19	1 to 19	1 to 19
	Test	3, 17	5, 15	1, 9	7, 13	11, 19

Table 3.1 Inputs of harmonics to loudspeakers A to E (See Figure 3.2 for loudspeaker positions) for Stage A of the experiment. Harmonic '1' refers to the fundamental harmonic, '3' to the third harmonic, and so on.

Comparison Number		Input of Harmonics to Loudspeakers					
		A	B	C	D	E	F
B1	Reference	1 to 18					
	Test	1, 7, 13	2, 8, 14	3, 9, 15	4, 10, 16	5, 11, 17	6, 12, 18
B2	Reference	1 to 18					
	Test	2, 8, 14	3, 9, 15	6, 12, 18	5, 11, 17	4, 10, 16	1, 7, 13
B3	Reference	1 to 18					
	Test	10, 11, 12	1, 2, 3	13, 14, 15	7, 8, 9	16, 17, 18	4, 5, 6
B4	Reference	1 to 18					
	Test	7, 8, 9	16, 17, 18	10, 11, 12	1, 2, 3	4, 5, 6	13, 14, 15
B5	Reference	2 to 19					
	Test	7, 13, 19	5, 11, 17	2, 8, 13	4, 10, 16	6, 12, 18	3, 9, 14
B6	Reference	2 to 19					
	Test	6, 12, 18	4, 10, 16	7, 13, 19	2, 8, 13	5, 11, 17	3, 9, 14
B7	Reference	2 to 19					
	Test	2, 3, 4	5, 6, 7	8, 9, 10	11, 12, 13	14, 15, 16	17, 18, 19
B8	Reference	2 to 19					
	Test	17, 18, 19	5, 6, 7	14, 15, 16	8, 9, 10	2, 3, 4	11, 12, 13
B9	Reference	1 to 18					
	Test	6, 12, 18	4, 10, 16	1, 7, 13	3, 9, 15	5, 11, 17	2, 8, 14
B10	Reference	1 to 18					
	Test	5, 11, 17	3, 9, 15	6, 12, 18	1, 7, 13	4, 10, 16	2, 8, 14
B11	Reference	1 to 18					
	Test	1, 2, 3	4, 5, 6	7, 8, 9	10, 11, 12	13, 14, 15	16, 17, 18
B12	Reference	1 to 18					
	Test	16, 17, 18	4, 5, 6	13, 14, 15	7, 8, 9	1, 2, 3	10, 11, 12
B13	Reference	2 to 19					
	Test	2, 8, 14	3, 9, 15	4, 10, 16	5, 11, 17	6, 12, 18	7, 13, 19
B14	Reference	2 to 19					
	Test	3, 9, 15	4, 10, 16	7, 13, 19	6, 12, 18	5, 11, 17	2, 8, 14
B15	Reference	2 to 19					
	Test	11, 12, 13	2, 3, 4	14, 15, 16	8, 9, 10	17, 18, 19	5, 6, 7
B16	Reference	2 to 19					
	Test	8, 9, 10	17, 18, 19	11, 12, 13	2, 3, 4	5, 6, 7	14, 15, 16

Table 3.2 Inputs of harmonics to loudspeakers A to F (See Figure 3.3 for loudspeaker positions) for Stage B of the experiment. Harmonic ‘1’ refers to the fundamental harmonic, ‘2’ to the second harmonic, and so on.

3.4.6 Results

In order to present the subject's responses to perceived differences between the test and reference signals in a clear and concise manner, each response was allocated to a particular difference category to allow for the formation of histograms. Tables of the subject's responses and the allocation to difference categories can be seen in Appendix A

In Stage A, the test signal was presented as a spatially separated harmonic series spread out over three or five loudspeakers arranged in an arc, in front of the listening position. In Comparisons A1, A3 and A4 the harmonics were distributed over three loudspeakers forming an angular spread of 90°. In Comparisons A2, A5, A6 and A7 there was a five loudspeaker distribution resulting in the harmonics being spread over 180°.

Figure 3.4 displays the subjects' responses to Comparisons A1 and A2. To the bottom of the histograms, a diagram displays the allocation of harmonics to loudspeakers. In these comparisons, the fundamental harmonic appeared in the centre loudspeaker, C.

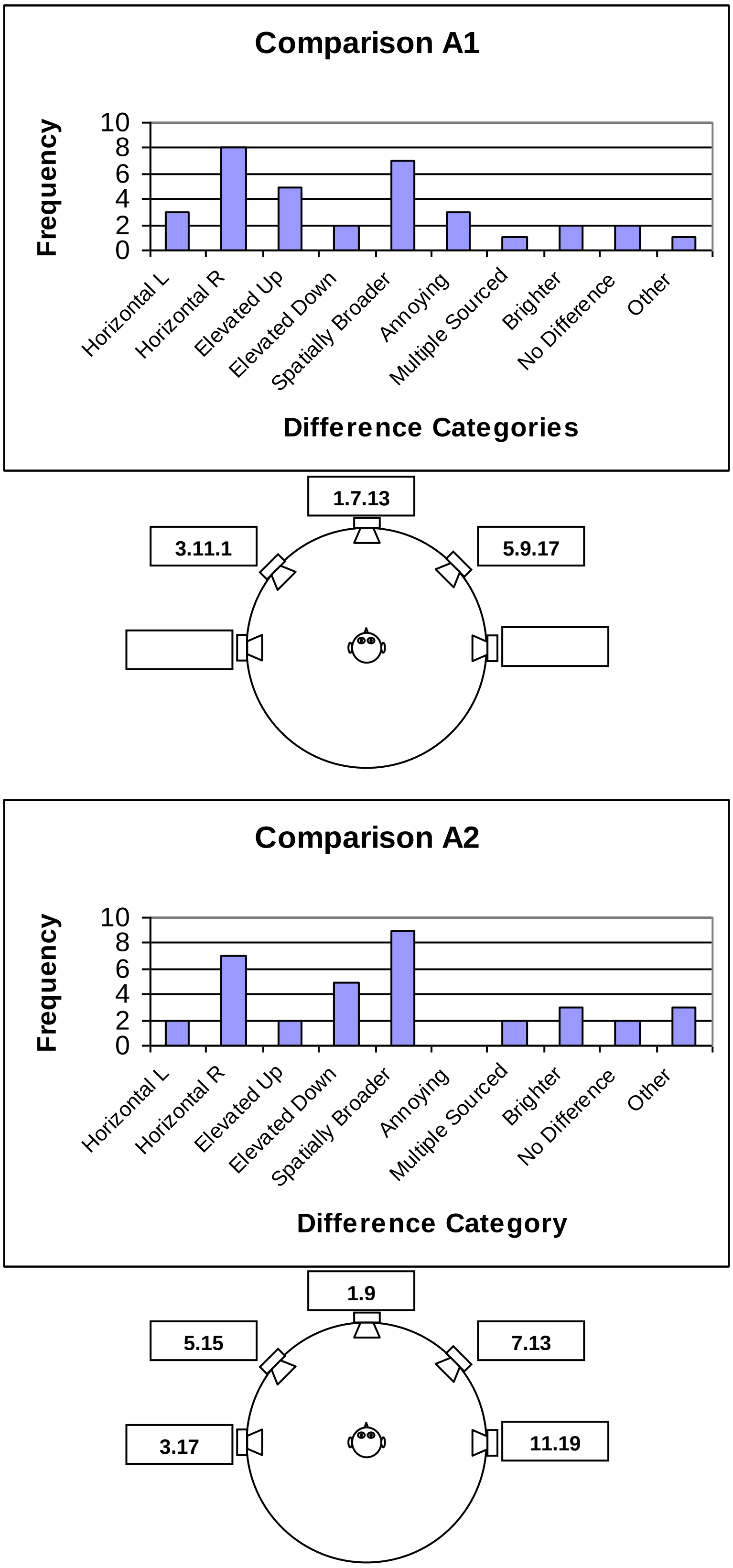
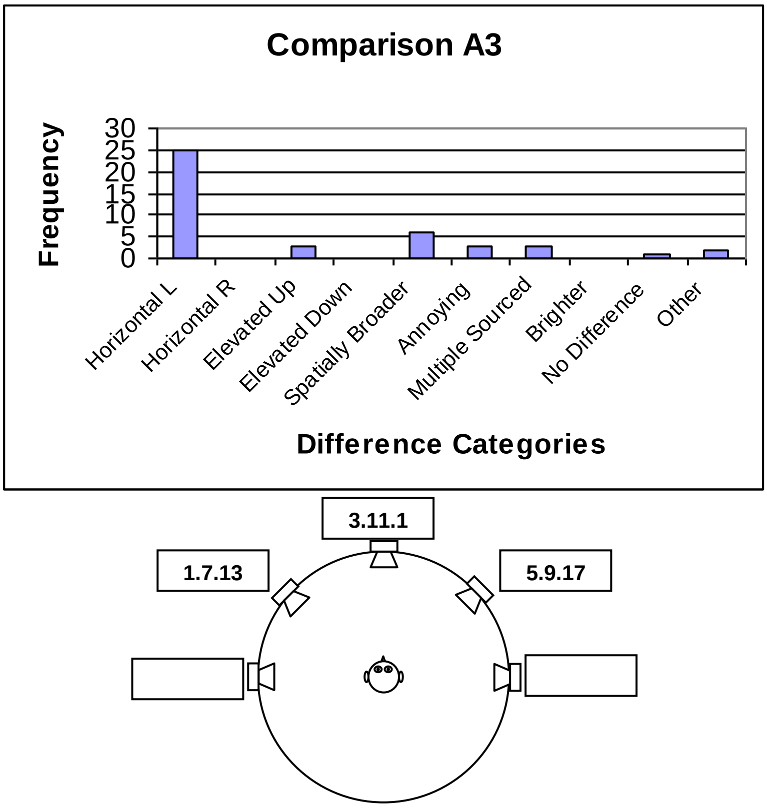


Figure 3.4 Frequency of responses to comparisons A1 (upper diagrams) and A2 (lower diagrams)



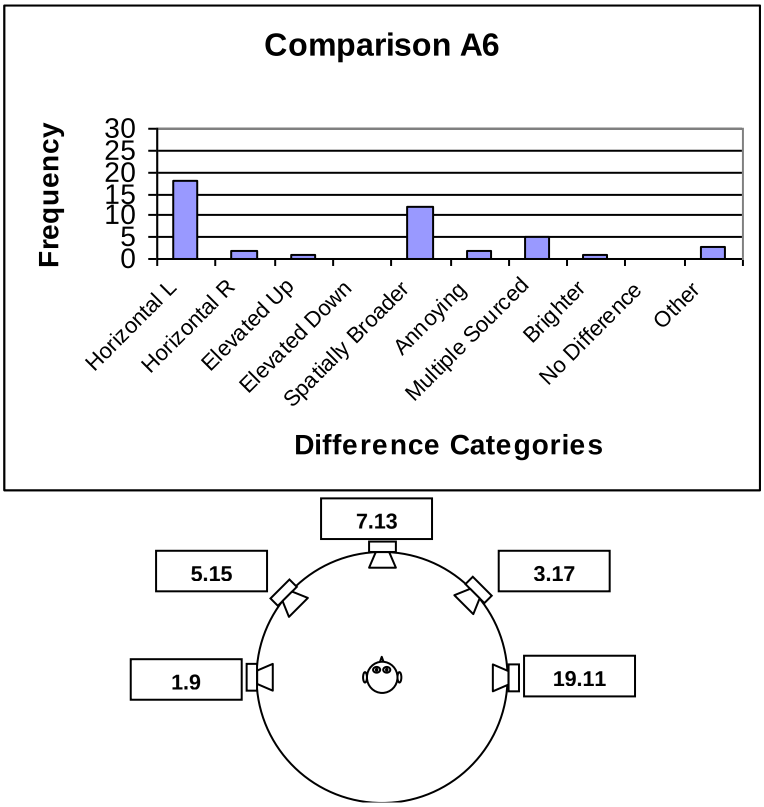


Figure 3.5 Frequency of responses to comparisons A3 (upper) and A6 (lower)

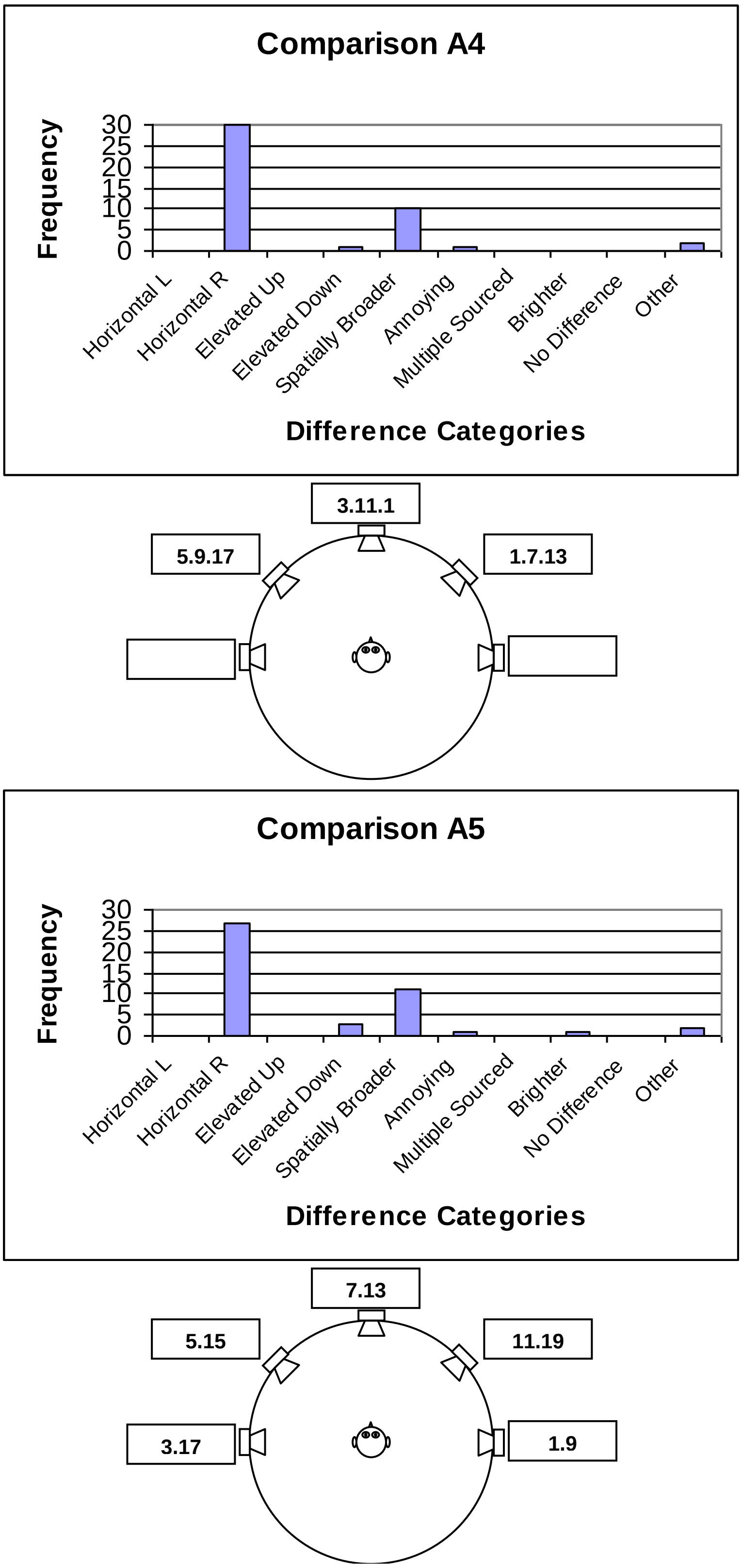


Figure 3.6 Frequency of responses to comparisons A4 (upper) and A5 (lower)

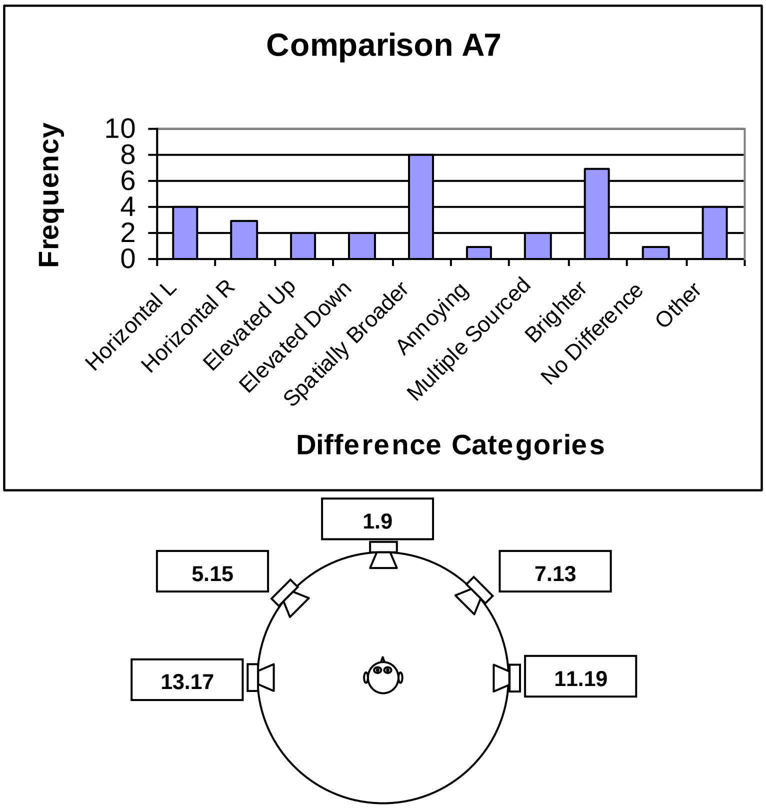


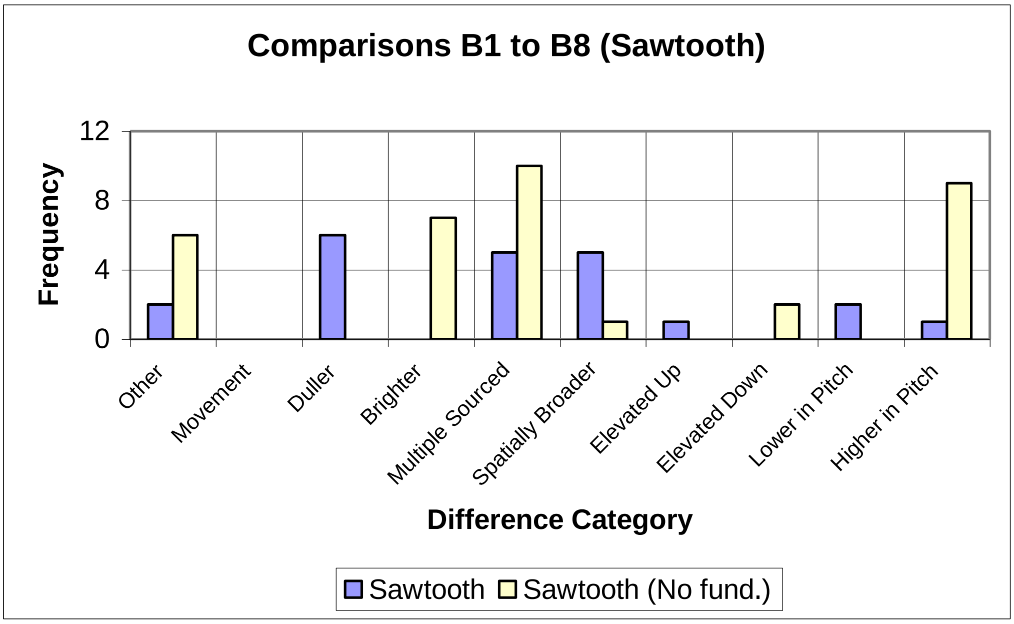
Figure 3.7 Frequency of responses to comparison A7

Figure 3.5 displays the responses to Comparisons A3 and A6, where the fundamental harmonic appeared to the left of centre, in loudspeakers A or B. Figure 3.6 displays the responses to Comparisons A4 and A5, where the fundamental harmonic appeared to the right of centre, in loudspeakers D or E. Figure 3.7 displays the responses to Comparison A7, which compared an unseparated square wave presented through all the loudspeakers to a spatially separated square wave.

In the second stage of the experiment six loudspeakers were arranged in a circular array, with the listening position being in the middle of the circle. As well as reporting any differences between the test and reference signals, subjects were also asked to report the location of the test signal. Two different samples were used in this part of the experiment, a sawtooth wave and a synthesized piano note.

Figure 3.8 displays the difference responses to Comparisons B1 to B16,. Comparisons B1 to B8 were for the sawtooth wave and B9 to B16, the piano note.

As the subjects were asked to describe the location of the test signal in their own words, different descriptors such as clock face locations, position in degrees or less defined descriptors such as ‘To the left’ or ‘To the rear’, were used. The location responses were then assigned to location categories. Figure 3.9 displays the frequency of each location category. The location categories ‘Left and right front’ and ‘Left and right rear’ describe instances where the subjects detected two source locations.



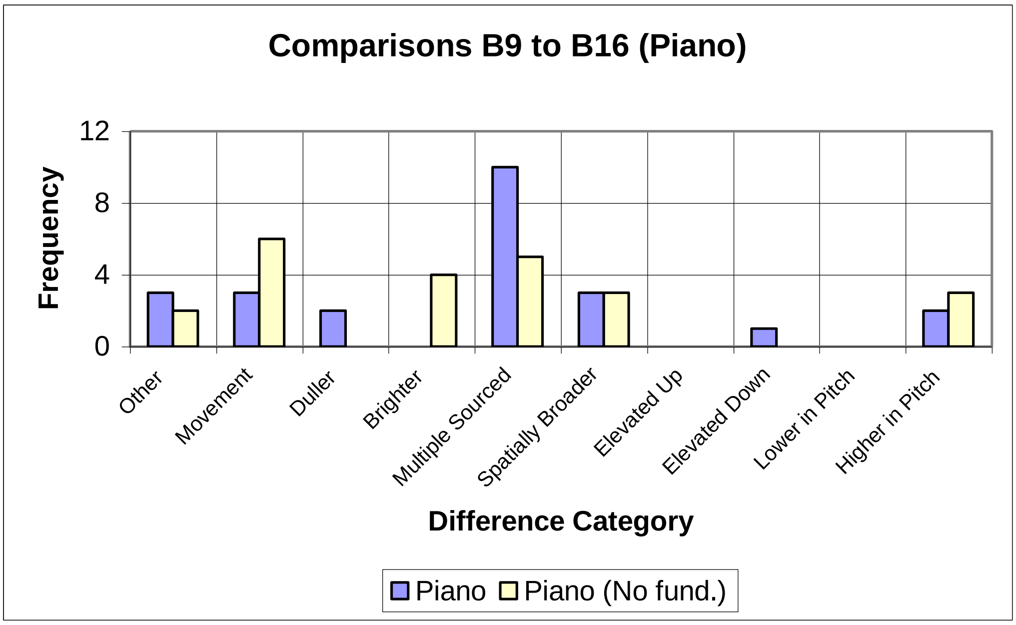
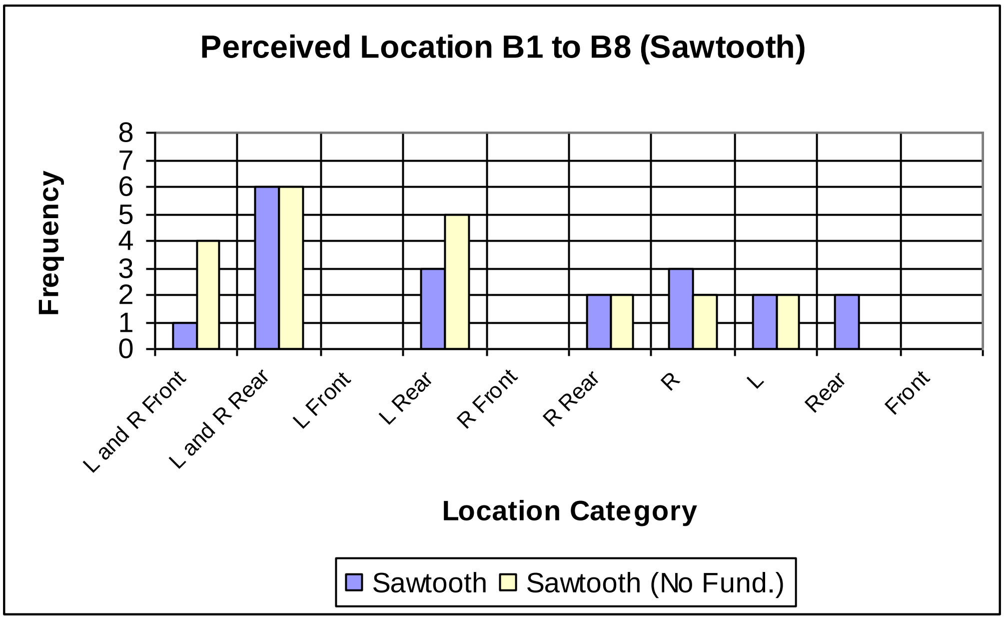


Figure 3.8 Frequency of responses to Comparisons B1 to B16



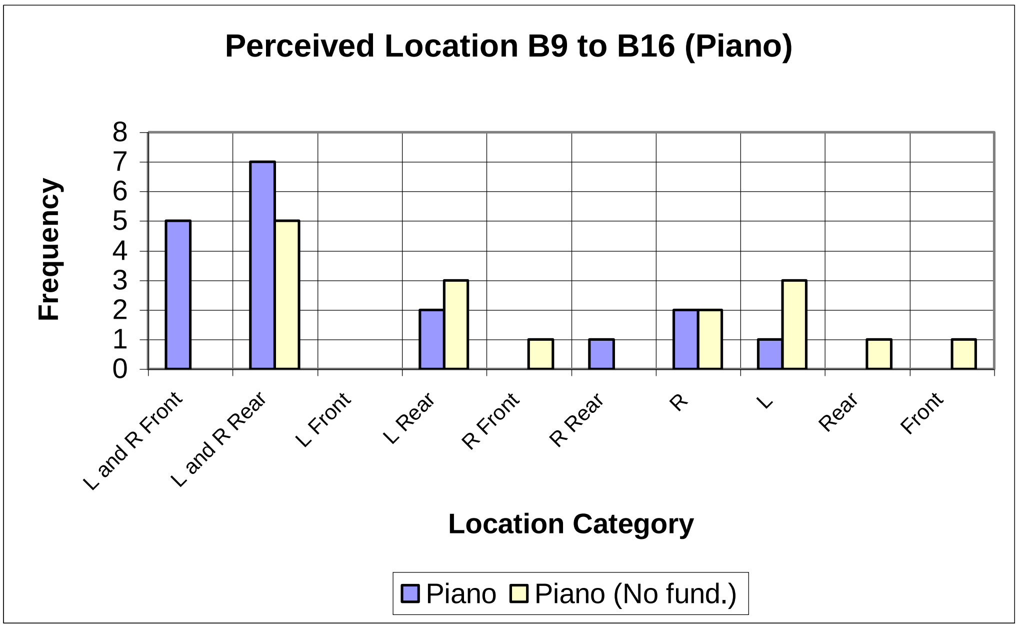


Figure 3.9 Perceived location of sound source in Comparisons B1 to B16

3.4.7 Observations

Stage A

Figure 3.4 displays the responses to Comparisons A1 and A2, where the test signal was distributed over either three or five loudspeakers arranged in an arc around the listening position, with the maximum extent of the arc covering a semi-circle in front of the listener. The fundamental harmonic remained in the centre loudspeaker. The most frequently reported differences for these tests were changes in horizontal or vertical position and spatial broadening.

The perceived changes in horizontal position were predominately to the right hand side. This was surprising as the author suspected any perceived changes in horizontal position may have corresponded to the positions of the lower harmonics. As can be seen in Figure 3.4, the positions of the third and fifth harmonics in Comparison A2 were both to the left of centre, yet the perceived location was predominately to the right.

The author cannot offer an explanation for the perceived differences in elevation, however, it is of interest to note that a number of subjects also reported that the *reference* signal was perceived as being at an elevation other than at head height.

As expected, the subjects tended to perceive the test signal as spatially broader than the reference signal, with the frequency of responses increasing with the angular spacing of the loudspeakers.

The ‘Brighter’ and ‘Annoying’ responses may be due to the higher harmonics ‘standing out’ more when presented as a spatially separated signal. This may suggest that due to spatial separation, the higher harmonics had not fused or had only partly fused with the perceived signal as a whole, thus forming a partially or wholly separate auditory event. The fundamental frequency of the signal used in Stage A was 440 Hz, which yielded higher harmonics in the region above 4 kHz. The author acknowledged that the presentations could be quite shrill (hence ‘annoying’) and for this reason the fundamental frequency of the test signals in Stage B was lowered to 150 Hz.

In Comparisons A3 to A6, the fundamental harmonic appeared in a loudspeaker other than the centre loudspeaker. From Figures 3.5 and 3.6, a clear correspondence between the position of the fundamental harmonic and the perceived horizontal position of the test signal can be seen. This occurred for both the three and five loudspeaker presentations. Again the frequency of spatially broader responses was greater for the five loudspeaker presentations than the three loudspeaker presentations.

In Comparison A7, a spatially separated square wave was compared to a combined square wave appearing in *all* loudspeakers. The main reported differences were ‘Brighter’, ‘Spatially Broader’ and ‘Horizontal left’. The differences in brightness may again be explained by the higher harmonics ‘standing out’. The ‘Spatially Broader’ differences demonstrate that the even though the mono signal presented in all loudspeakers *is* physically surrounding the listener, the spatializing techniques that are being presented in the same manner, contain extra auditory information (or conflicting cues) that results in a broader spatial perception.

Stage B

In Stage B, the spatially separated signals were presented through a circular, six-loudspeaker array, with the listening position being in the centre of the circle.

In Figure 3.7, which displays the combined results for Comparisons B1 to B16, the most frequently reported difference was ‘Multiple Sourced’. This suggests that with a signal that is spatially spread over a loudspeaker array that totally encircles the listener, the perceived signal is not fused into a single auditory event, thus implying a limit to the degree in spatial spread of harmonics for the spatializing techniques.

The predominance of the omitted fundamental signals in the ‘Brighter’ and ‘Higher in pitch’ differences suggest that due to fundamental tracking, the timbre of the signal had changed so that the perception of the fundamental harmonic was not as strong, or that the second harmonic was being perceived as a separate auditory event, with the other harmonics forming a residue pitch that was not as dominant as the second harmonic.

'Movement' accounted for a number of differences, especially with the synthesized piano note. As the piano differed from the saw-tooth wave in that time varying (in amplitude) partials were employed in the synthesis, the author suggests that this may be a reason for the perception of movement. As the general perception was one of multiple sources, it is possible that as one partial or groups of partials decreased in amplitude and another (in a different location) increased, so in an attempt to group the partials, the auditory system may have 'interpolated' the positions of the samples, resulting in a perception of movement.

From Figure 3.9, the most frequently reported location of the test signal was 'Left and Right Rear', which describes the signal as having two sources, often described by the subjects as being similar to hearing a stereo pair of loudspeakers placed behind the head. The second most frequently reported location was 'Left Rear'. Very few responses placed the signal to the front of the listening position. A possible explanation why the signal was predominately perceived as emanating from the rear of the listening position is that as the auditory system is confused by the conflicting cues, the 'unknown' sound perception could be thought of as ‘threatening’. For this reason, the perceived location of the sound is behind the listener so that a ‘fleeing’ response may be invoked in the listener to escape the potential ‘threat’.

From the results of Stage A, where the location of the signal corresponded well with the position of the fundamental, similar results were expected for Stage B. However, a relationship between the location of the fundamental harmonic and the perceived location of the test signal was not observed. A possible explanation is that again, the auditory system is receiving conflicting information about the signal and therefore cannot fuse or place the auditory event.

3.4.8 Summary of the Informal Investigation

- By distributing the harmonics of a musical signal over a spaced loudspeaker array, the signal was perceived as being spatially broader than a mono version of the same signal, thus indicating that the spatializing techniques are successful.
- As the angular spread of harmonics was increased from 90° to 180°, the signal was perceived as being spatially broader.
- There was a limit to the degree of angular spread in that the signal was perceived to be multiple sourced when the harmonics were distributed over a circular loudspeaker array (angular spread of 360°).

- When the angular spread was limited to a maximum of 180°, the perceived location of the spatially separated signal corresponded to the location of the fundamental harmonic.
- For some presentations, the higher harmonics may not have completely fused and the rest of the harmonics and tended to 'stand out'.
- When a spatially separated presentation was compared to a mono version of the presentation, replayed over all loudspeakers of a semicircular array, the spatially separated signal was perceived as being spatially broader.

As the outcomes of the informal experiment were encouraging, it was decided to develop the investigation of the spatializing techniques further by conducting a controlled and formal subjective test, which is reported upon in the next section.

3.5 Formal Investigation into Multichannel Spatialization Techniques for Musical Synthesis

3.5.1 Introduction

Following on from the outcome of the informal investigation, the premise behind the test design was that the spatializing techniques would deliver a spatial effect and as the spatial spread of harmonics common to a complex musical signal was increased around the listener, the perceived degree of spatial impression or envelopment would also increase. However, there might be a limit to the degree of spatial spread as the signal might be perceived as being multiple sourced when the spread becomes too great. The experiment also formally compares a spatialized signal to a mono signal replayed through an array of surrounding loudspeakers.

The basic spatializing techniques were similar to those used in the informal experiment in that the test signals were decomposed into their individual harmonics then spatially spread around the listener in increasing steps. The subjects were asked to rank order, in terms of spatial sound quality, four auditions (an ‘audition’ is defined as the playing of a single sample) of varying spatial spread and one audition comprising the original signal presented through all eight loudspeakers. To test the robustness of the spatializing techniques, the procedure was also tested using ambisonic reproduction.

3.5.2 Program Material

Two standard format (16 bit, 44.1 kHz) stereo samples were used in the test, both of which were downloaded from a website [Samplenet 2000]. The single note samples were of a 4.26 s, G4 (f_0 = 392 Hz) string ensemble and a 4.10 s, C4 (f_0 = 261 Hz, with a 130 Hz sub-harmonic also present) synthesizer sound. Using a sample editor (Cool Edit Pro), the stereo samples were converted into mono then narrow band pass filtered, using a Butterworth sixth order filter, to extract each harmonic. This yielded 28 harmonics for the string ensemble and 40 for the synthesizer sound. The harmonics were then assigned into groups for presentation.

3.5.3 Rank Order Arrangement

The assignment of harmonics to each group was dependent upon how many loudspeakers were being used in a particular audition. For example, the synthesizer sound consisting of 40 harmonics was split into eight groups of five harmonics for an eight-loudspeaker (360° spread) audition and five groups of eight harmonics for a five-loudspeaker (180° spread) audition. The assignments of harmonics to loudspeakers, for each audition are shown in Tables 3.3 and 3.4. For all auditions, the fundamental harmonic was assigned to the loudspeaker directly in front of the listening position.

	Harmonics Assigned to Each Loudspeaker							
Angular Spread of Speakers (degrees)	LS1	LS2	LS3	LS4	LS5	LS6	LS7	LS8
0	All							
90	2 5 8 11 14 17 20 23 26 29 32 35 38	3 6 7 12 13 18 19 24 25 30 31 36 37						1 4 9 10 15 16 21 22 27 28 33 34 39 40
180	2 9 12 19 22 29 32 39	3 8 13 18 23 28 33 38	4 7 14 17 24 27 34 37				5 6 15 16 25 26 35 36	1 10 11 20 21 30 31 40
360	2 15 18 31 34	3 14 19 30 35	4 13 20 29 36	6 11 22 27 38	8 9 24 25 40	7 10 23 26 39	5 12 21 28 37	1 16 17 32 33

See diagram below for loudspeaker numbering and angular spread. Harmonic 1 refers to a sub-harmonic, Harmonic 2 refers to the fundamental, Harmonic 3 refers to 2 x the fundamental etc.

Angular Spread (Degrees)	**Loudspeakers Active**
0	1
90	8, 1 and 2
180	7, 8, 1, 2 and 3
360	All

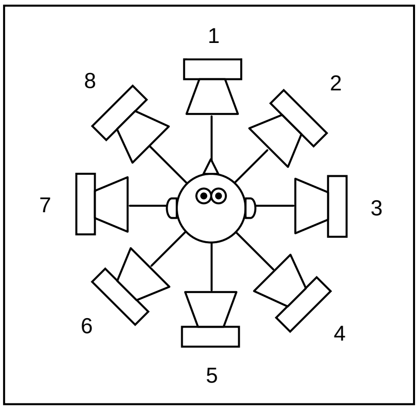


Table 3.3 Assignment of harmonics to loudspeakers (synthesiser)

	Harmonics Inputted to Each Loudspeaker							
Angular Spread of Speakers (degrees)	LS1	LS2	LS3	LS4	LS5	LS6	LS7	LS8
0	All							
90	1 4 9 10 15 16 21 22 27	3 6 7 12 13 18 19 24 25 28						2 5 8 11 14 17 20 23 26
180	1 10 11 20 21	3 8 13 18 23 28	4 7 14 17 24 27				5 6 15 16 25 26	2 9 12 19 22
360	1 16 17	3 14 19	4 13 20	6 11 22 27	8 9 24 25	7 10 23 26	5 12 21 28	2 15 18

See diagrams below for loudspeaker numbering and angular spread. Harmonic 1 refers the fundamental, Harmonic 2 refers to 2 x the fundamental etc.

Angular Spread (Degrees)	**Loudspeakers Active**
0	1
90	8, 1 and 2
180	7, 8, 1, 2 and 3
360	All

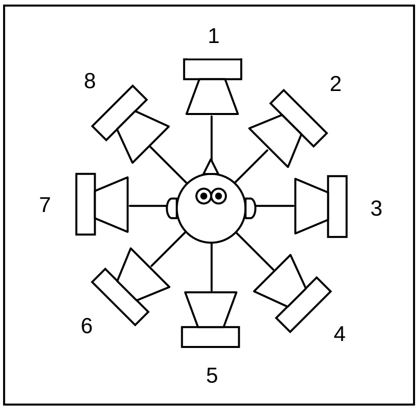


Table 3.4 Assignment of harmonics to loudspeakers (string)

Each presentation compared five auditions. Four of the auditions consisted of a decomposed signal with angular spreads of 0°, 90°, 180° and 360°. The other audition consisted of the original (not decomposed) mono signal, simultaneously replayed through all eight loudspeakers (henceforth referred to as 'Mo8'). This audition was included to determine if the distribution of the harmonics in space was the reason why an increased perception of spatial impression occurred or if by surrounding the listener with the same signal provided the spacious perception.

To negate any bias introduced by differences in perceived loudness, the overall level of the Mo8 audition was adjusted to be of the same perceived loudness as the decomposed auditions. This procedure was performed by the experimenter and confirmed by one of the subjects. With two source materials and two reproduction methods, this resulted in four rank order presentations of five auditions. All non-ambisonic auditions had listening levels of between 79 and 80 dBA. This listening level was chosen as an 'average' of preferred or most comfortable listening levels for music as determined by Mathers [Mathers 1979] as 83.5 dBA and Airo [Airo et al. 1996] as 69 dB.

For the ambisonic auditions the harmonics were assigned to the same groups as for the real sources then positioned around the listener, at the same angular positions as the real sources using the standard ambisonic encoding process.

3.5.4 Subjects

Twelve subjects, seven males and five females, participated in the experiment all of which were either staff or students of The School of Acoustics and Electronic Engineering, University of Salford. The majority of the subjects had previously participated in other listening tests (including the informal pilot study). Ideally the minimum number of experienced subjects is twenty [ITU 2003], however, due to difficulties in finding this number of willing participants, a compromise number of twelve was decided upon. All subjects attended a training session that involved an introduction to multichannel spatial audio and a trial run of the test procedure.

3.5.5 Test Room and Equipment Configuration

The experiment was carried out in the anechoic chamber of the School of Acoustics and Electronic Engineering, University of Salford. The working dimensions of the room were measured as 3.6m in height by 5.5m in length by 3.2m in width. The inner chamber was lined with 0.9m long fibreglass wedges to give an anechoic cut-off frequency below 100 Hz.

Eight loudspeakers, arranged in a circular array, were attached to an octagonal metal frame, with the listening position in the centre. The loudspeakers were placed at head height and at a distance of 1.41m from the listening position and at an angular spacing of 45°. An acoustically transparent curtain was hung between the listening position and the loudspeakers to facilitate blind testing. A computer keyboard (which was used as a switching mechanism), a loudspeaker and microphone (to

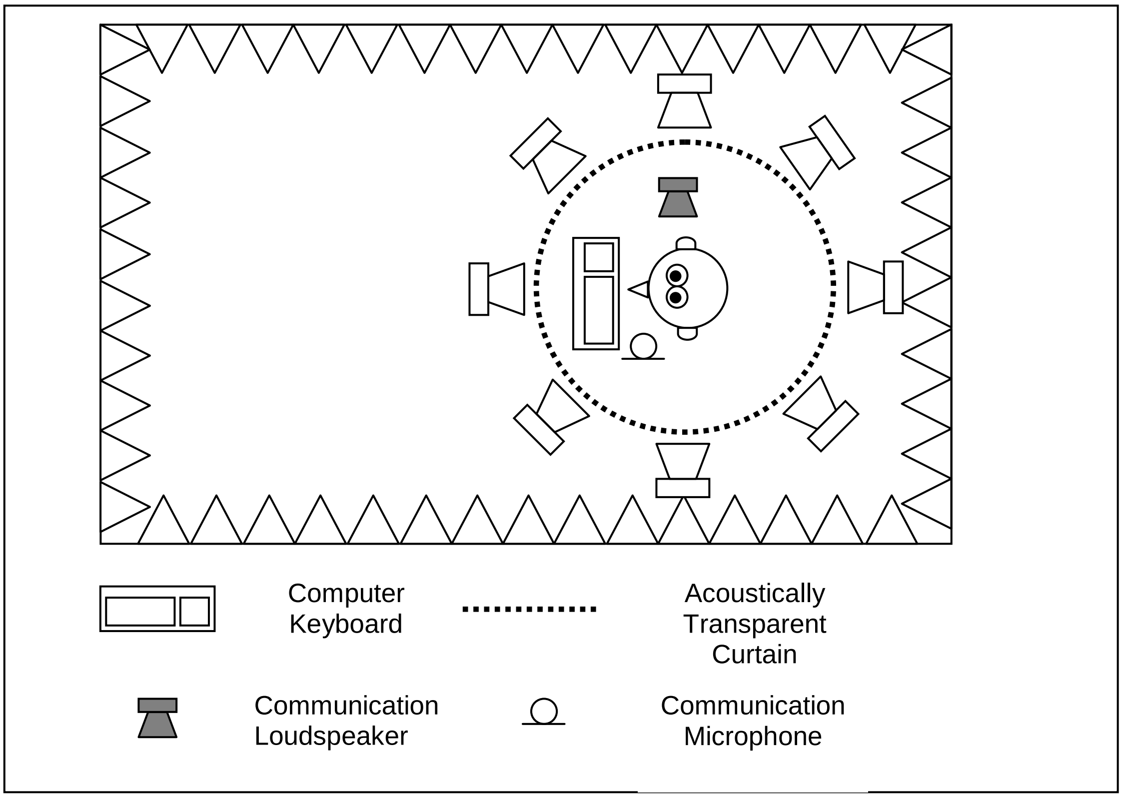


Figure 3.10 Test room configuration

enable communication between the subjects and the experimenter), were also present in the chamber. A diagram of the test room configuration is shown in Figure 3.10.

The test signals were loaded into a computer based audio sequencer (Cubase VST), the digital output of which (via a multichannel soundcard) was connected to an Alesis ADAT to allow for digital to analogue conversion. Balanced outputs from the ADAT fed the eight Genelec 1029A loudspeakers that were level aligned using pink noise and a sound level meter. By using the ‘Cue Point’ feature in Cubase and a computer keyboard acting as a remote control, the subjects could switch between each of the five auditions of a rank order presentation at will, thus

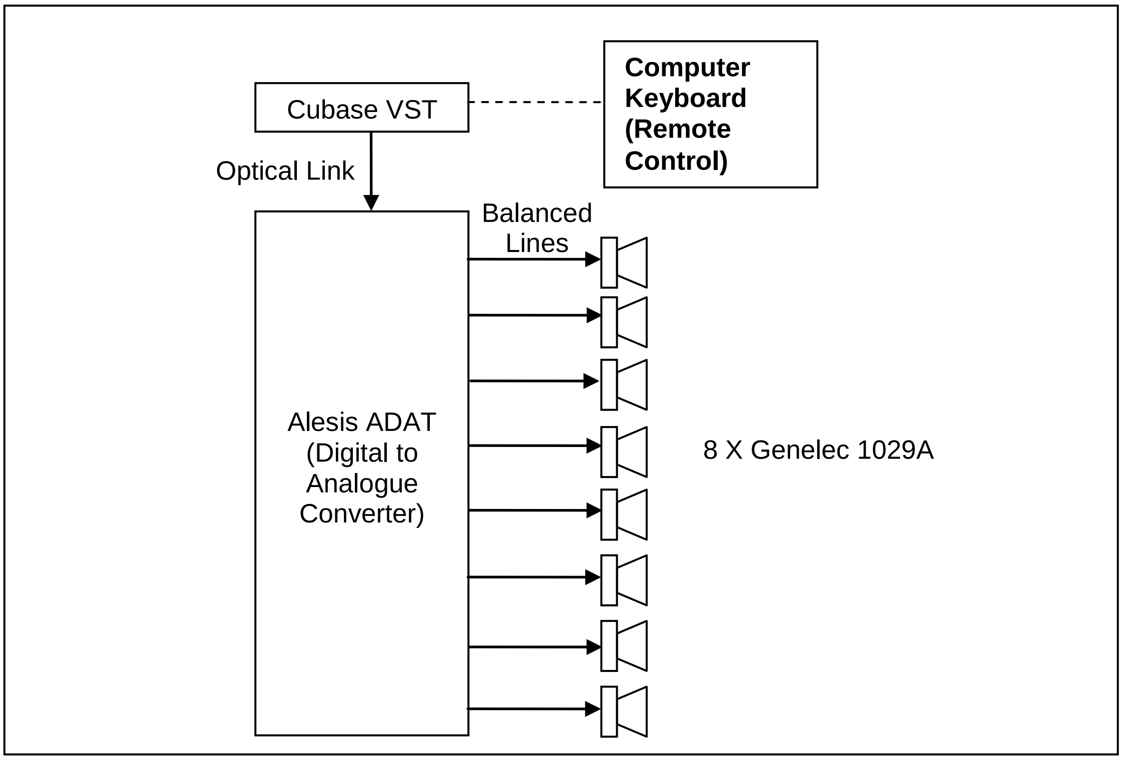


Figure 3.11 Equipment configuration

enabling quick comparisons that may result in less subject fatigue and hence smaller errors. A diagram of the equipment configuration is shown in Figure 3.11.

For ambisonic playback a square, pantaphonic four-loudspeaker configuration was used. This utilised loudspeakers 8,2,4 and 6 (See Table 3.4). The standard ambisonic decoding equation shown in Equation 2.4 was used and shelf filtering was not applied.

3.5.6 Experimental Procedure

The blindfolded subjects were individually escorted into the anechoic chamber, seated at the listening position then un-blindfolded. For each of the four presentations (two presentations for systems and two for program material), consisting of five auditions each, the subjects were asked to rank order the auditions in terms of spatial sound quality (1 = lowest rank, 5 = highest rank). In evaluating the spatial sound quality the subjects were asked to 'Consider all aspects of spatial sound reproduction. This might include the locatedness or localisation of the sound, the width of the sound, how enveloping it is or it's naturalness and depth' [Berg and Rumsey 1999].

A rank order procedure was chosen for three main reasons; the procedure is straight-forward for the subjects to understand and complete, the subjects do not have to interpret and gauge a scale and the statistical analysis of the data is relatively straightforward.

The subjects could freely switch between the auditions of each presentation and could take as long as they needed to determine the rank order. On average, the test took approximately twenty minutes to complete. When a particular rank order had been determined, the subjects verbally relayed their choice to the experimenter via a microphone.

3.5.7 Results

At the end of the training session, the subjects performed a trial rank ordering of the synthesizer sample. The data collected from this rank ordering were correlated with the actual test data for each subject. This was done to identify subjects that were unable to repeat the rank ordering with some level of consistence. Correlation analysis revealed that data collected from two of the subjects was particularly inconsistent and therefore was not included in the analysis.

For each of the four presentations the data collected from the remaining ten subjects was subject to the non-parametric Friedman analysis of variance test [Lawless and Heymann 1998] the output of which is seen in Table 3.5. In the first column of the table, the four presentations are denoted by ‘Ambi Str’ (string sound replayed over ambisonic system) or ‘Real Syn’ (synthesizer sound replayed over system using real loudspeaker sources) and so on. The ‘ASTRO0’ or ‘RSYN360’ labels refer to the type of presentation and the angular spread of harmonics. The analysis showed that the preference ranks for all four sets of data differed significantly at a maximum of the $p < 0.003$ level, thus indicating the results for all four presentations are statistically meaningful.

REAL SYN	N	Mean	Std. Deviation	Minimum	Maximum	Friedman
RSYN0	10	1.3000	.94868	1.00	4.00	N = 10
RSYN90	10	2.9000	1.10050	1.00	5.00	Chi Sq =
RSYN180	10	3.6500	1.10680	2.00	5.00	16.26
RSYN360	10	3.7500	.97895	2.00	5.00	df = 4
RSYNMO8	10	3.4000	1.50555	2.00	5.00	**Sig = .003**

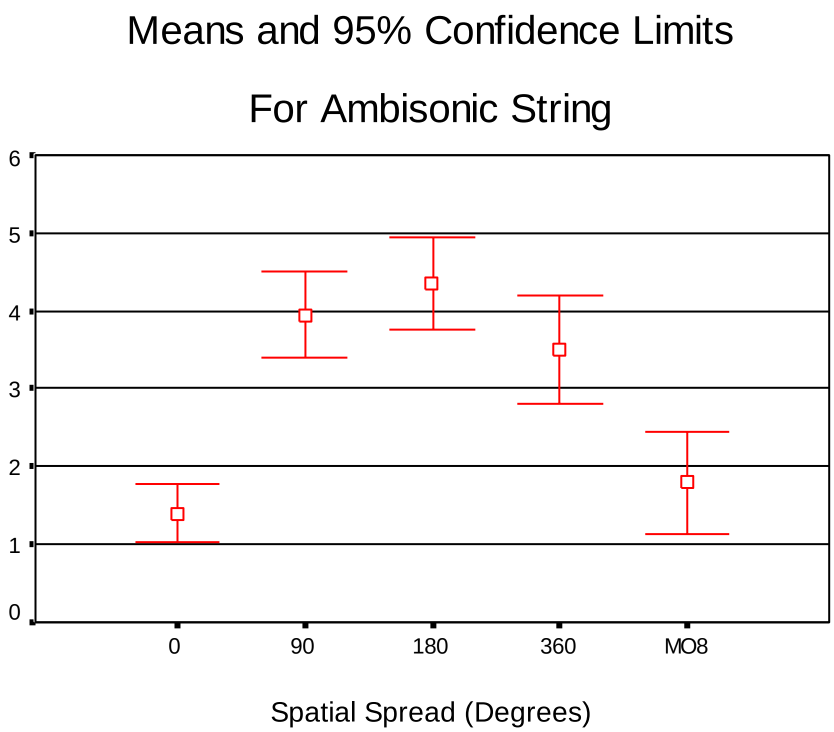
AMBI STR	N	Mean	Std. Deviation	Minimum	Maximum	Friedman
ASTR0	10	1.4000	.51640	1.00	2.00	N = 10
ASTR90	10	3.9500	.76194	3.00	5.00	Chi Sq =
ASTR180	10	4.3500	.81820	3.00	5.00	21.04
ASTR360	10	3.5000	.97183	2.00	5.00	df = 4
ASTRMO8	10	1.8000	.91894	1.00	4.00	**Sig = .000**

AMBI SYN	N	Mean	Std. Deviation	Minimum	Maximum	Friedman
ASYN0	10	1.3000	.48305	1.00	2.00	N = 10
ASYN90	10	2.9500	1.25720	1.00	5.00	Chi Sq =
ASYN180	10	4.4500	.49721	4.00	5.00	28.44
ASYN360	10	4.1500	.74722	3.00	5.00	df = 4
ASYNMO8	10	2.1500	.74722	1.00	3.00	**Sig = .000**

REAL STR	N	Mean	Std. Deviation	Minimum	Maximum	Friedman
RSTR0	10	1.1000	.31623	1.00	2.00	N = 10
RSTR90	10	4.0000	1.05409	2.00	5.00	Chi Sq =
RSTR180	10	3.6000	1.26491	2.00	5.00	21.04
RSTR360	10	3.5000	.70711	3.00	5.00	df = 4
RSTRMO8	10	2.8000	1.39841	1.00	5.00	**Sig = .000**

Table 3.5 Output of the Friedman analysis of variance test

Graphs depicting the mean rankings and 95% confidence limits for each individual presentation and the overall (averaged) spatial rank and 95% confidence limits can be seen in Figures 3.12 to 3.16.



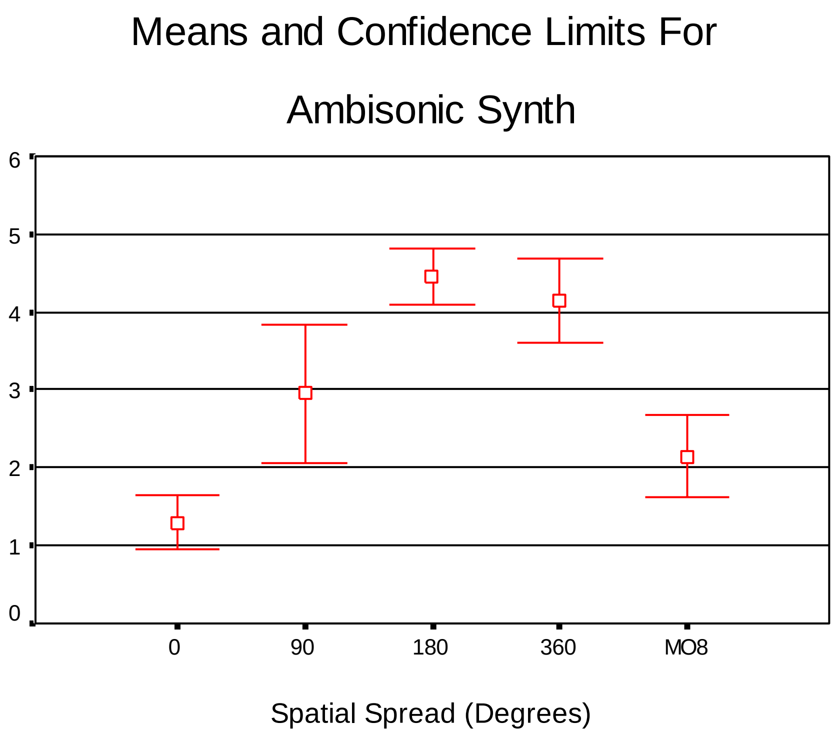


Figure 3.12 Mean subjective rank for ambisonic string

Figure 3.13 Mean subjective rank for ambisonic synth

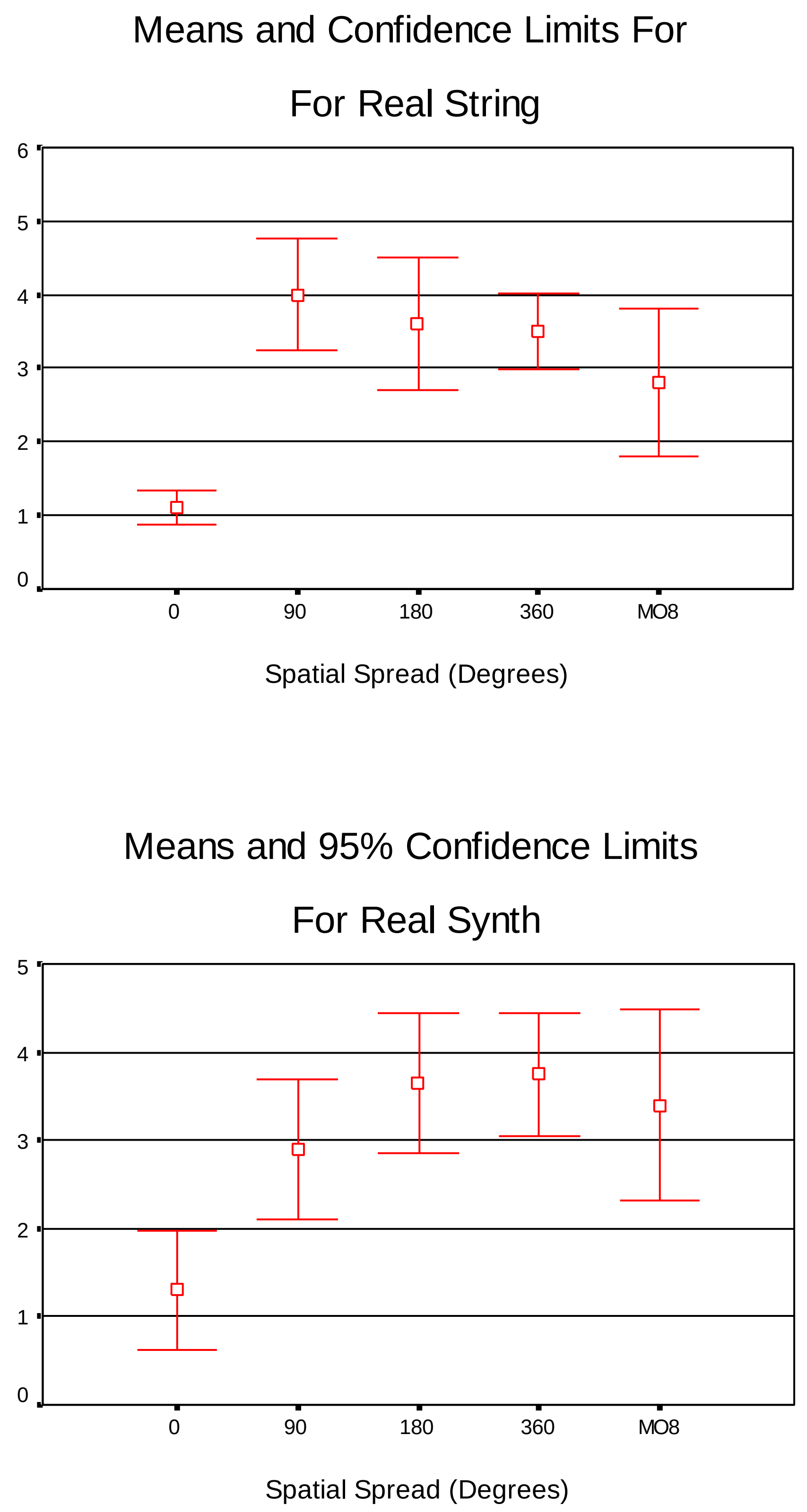
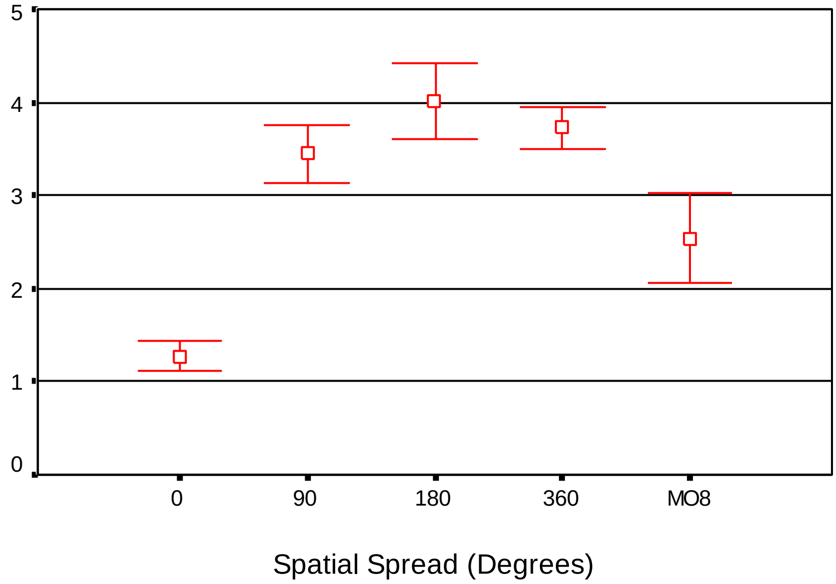


Figure 3.16 Overall mean subjective rank

Means and Confidence Limits

Overall



The validity of using of means and confidence limits in displaying rank order data has been established by Newson [Newson 2000]. Having established that preference ranks for each presentation differed significantly, the least significant rank difference (LSRD) for the Friedman test [Lawless and Heymann 1998] was calculated. This test determined which auditions were ranked significantly higher or lower in preference from one another for each presentation. The results can be seen in Table 3.6. Spatial spreads sharing the same colour-coded significance group letter do not differ significantly in ranked preference. In brief, the spatial spread of 0˚ was consistently ranked significantly lower than the 90˚, 180˚ and 360˚ auditions for all presentations. For the ambisonic synthesizer presentation, the 180˚ audition was ranked significantly higher than all auditions, apart from the 360˚ audition that was ranked significantly higher than the Mo8 and 0˚ auditions. For the remaining individual presentations, the ranked differences between the 90˚, 180˚ and 360˚ auditions were not significant.

	Spatial Spread (Degrees)				
Presentation	0	90	180	360	Mo8
Real Synth Rank Total	13	29	36.5	37.5	34
Significance group	A	B	B	B	B
Ambisonic Synth Rank Total	13	29.5	44.5	41.5	21.5
Significance group	A	BC	D	CD	AB
Real String Rank Total	11	40	36	35	28
Significance Group	A	B	B	B	B
Ambisonic String Rank Total	14	39.5	43.5	35	18
Significance Group	A	B	B	B	A
All (Average) Rank Total	12.8	34.5	40.1	37.3	25.4
Significance Group	A	BC	C	BC	AB

Table 3.6 Significance groups as determined by the least significant rank difference

To give an overall indication, the ranks for the spatial spreads of all presentations were averaged and the LSRD calculated. Again, the 0° audition was ranked lower significantly than all auditions apart from the Mo8 audition. The 180° audition was ranked significantly higher than the Mo8 audition.

3.5.8 Discussion

From the theory and the outcomes of the informal experiment, by presenting a complex musical signal in a harmonically decomposed and spatially distributed manner, it was expected that as the spatial spread of harmonics was increased, the perceived degree of spatial impression would also increase. The increase in spatial impression was expected to break down (resulting in the perception of more than one source) once the spatial spread of harmonics had extended beyond the sides of the listener as the conflicting cues presented to the auditory system could no longer be resolved as a single auditory event.

Additionally, by presenting a mono version of the signal over all the surrounding loudspeakers (a spatial spread of 360°), it was expected that when compared to the same signal that had been spatially processed, the latter signal would be perceived as being more spacious, even when the spatial spread was narrower than the former signal.

The graph for the overall data (Figure 3.16) shows that the 90°, 180° and 360° spatial spreads were ranked fairly similarly whilst the other auditions were ranked noticeably lower (especially the 0° audition).

From the LSRD test results shown in Table 3.6 it can be seen that a spatial spread of 0˚ was ranked significantly lower than all other spatial spreads in all four presentations. Apart from one presentation, there was no significant difference between the spatial spreads of 90˚, 180˚, and 360˚. This suggests that whilst the techniques deliver a spatial effect, the degree of spatial impression does not increase further as the spatial spread of harmonics is extended beyond 90˚, which is probably due to the auditory system being unable to fuse the individual harmonic components into a single perception.

As similar results were found for both program materials and both reproduction methods, the techniques appear to be robust. For the ambisonic presentations, the results were very similar. In both presentations spatial spreads of 180˚ and 360˚ were ranked significantly higher than 0˚ and Mo8. For one of the ambisonic presentations (Synthesizer), spatial spreads of 180˚ and 360˚ were ranked significantly higher than all other auditions. This may suggest that for ambisonic reproduction, extending the spatial spread beyond 90˚ results in an increase in the perceived degree of spatial impression and possibly without the perception of multiple sources.

3.6 Summary

In this Chapter novel spatializing techniques for musical synthesis were developed and investigated. The techniques involved distributing groups of harmonics, common to a complex musical signal, over a multi-loudspeaker array, in order to create a greater degree of spatial impression. Theory suggests that this presents the auditory system with conflicting cues. The localization cues would imply that there were a number of individual sources around the listener. However, these sources also possess a number of related cues, including harmonicity and similar onset times that would influence the auditory system in the grouping of the sources into a single perception.

A pilot study was conducted that investigated the effects of varying the spatial spread of harmonics around the listener and the effects of varying the position of the fundamental harmonic upon the perceived location of the signal. The outcomes suggested that the degree of perceived spatial impression increased with the increasing angular spread of harmonics although there may be a limit to the angular spread before the perception becomes multiple sourced. In terms of the localization of the signal, no obvious trend was recognised.

A formal subjective experiment further tested the hypothesis that perceived spatial impression increased with the angular spread of harmonics. Musical signals were presented over a circular array and a square pantophonic ambisonic system at varying degrees of spatial spread. A mono version of the signal simultaneously replayed over all loudspeakers was also presented. By means of rank ordering, subjects rated the presentations in terms of perceived spatial impression. The results of the experiment were shown to be significant. The perception of spatial impression increased as the angular spread of harmonics increased; however, extending the spatial spread beyond 90° may not significantly increase the perception of spatial impression. The techniques were also shown to deliver a significantly greater perceived degree of spatial impression than a multi-loudspeaker mono version of the signal.

The techniques offered a number of possible areas for further work, one of which was the objective measurement of spatial impression delivered by the techniques. In the next Chapter, these objective measurements of the spatializing techniques are reported upon and the application area expanded to accommodate other kinds of multi channel reproduction.

4 Adaptation of Concert Hall Measures of Spatial Impression to Reproduced Sound

4.1 Introduction

In the previous Chapter, spatializing techniques for musical synthesis were investigated. Having established that the techniques subjectively enhanced the perceived degree of spatial impression and as a result of pursuing further work, a change in the overall direction of the project came about in moving towards the development of objective measurements of spatial impression in reproduced sound.

This Chapter initially reports upon the use of IACC in objectively measuring the varying degrees of spatial impression delivered by the spatializing techniques discussed in the previous Chapter. The possibilities of adapting the IACC measurement to be used in reproduced sound in general are then discussed. As IACC is an objective measurement of spatial impression in concert halls, a number of problems arise in adapting the measurement and these form a large part of the discussion.

A previously investigated method of adapting IACC to reproduced sound [Furlong 1989] is used as a starting point. The method involves comparing IACC measurements taken in a concert hall to IACC measurements taken in reproduced versions of the same concert hall. The type of reproduction system can be varied and an indication of the systems' spatial performance may be gained from the comparison of original and reproduced IACC measurements for each system.

The method is first conducted as a simulation using basic auralisation techniques. Real concert hall measurements and reproduction systems are then employed in taking the method into new grounds. The method is further developed by introducing variations and refinements to the IACC measurement and to the methods in which the original and reproduced IACC measurements are compared.

4.2 Objective Measurement of the Spatializing Techniques for Musical Synthesis

The spatializing techniques for musical synthesis that were the subject of the previous Chapter were investigated using objective measurements. IACC measurements of the signals presented to the subjects were taken. The spatialized signals used in the experiment were described in detail in the last chapter. In brief, the signals consisted of spatialized synthesizer and string ensemble examples presented both through various components of an eight-loudspeaker circular array and ambisonically through a four-loudspeaker system. The angular spread of the harmonics of each signal was varied from 0° (mono) to 360°. The signals were also presented in multi-loudspeaker mono (the same signal was simultaneously replayed through all the loudspeakers of the circular array).

A binaural recording of each audition was made using the same equipment and in the same location as the subjective experiment, from which IACC measurements were taken. As the signals were not impulse responses, the usual 80ms time window was not applied to the signals (See Equation 2.8), instead the whole duration of the signal was used in the calculation. Whilst the measurements were not intended to result in 'absolute' IACC values, the measurements were hoped to be useful for making comparisons between the auditions of different harmonic spreads in terms of an objective measure of spatial impression.

By converting the IACC measurements to 1-IACC measurements, direct comparisons between the subjective results and the objective measurements for each set of auditions can be made (it is expected that the greater the subjective perception of spatial impression, the greater the 1-IACC measurement). Graphs displaying the mean subjective ranks and 1-IACC measurements versus angular spread can be seen in Figures 4.1 to 4.4. For all four plots a similar trend can be seen. As the angular spread is increased, the 1-IACC value increases up until an angular spread of 180°. For angular spreads above 180° (the 360° audition), the

1-IACC value decreased. The 0° (mono) audition and the Mo8 audition both resulted in 1-IACC measurements of close to zero.

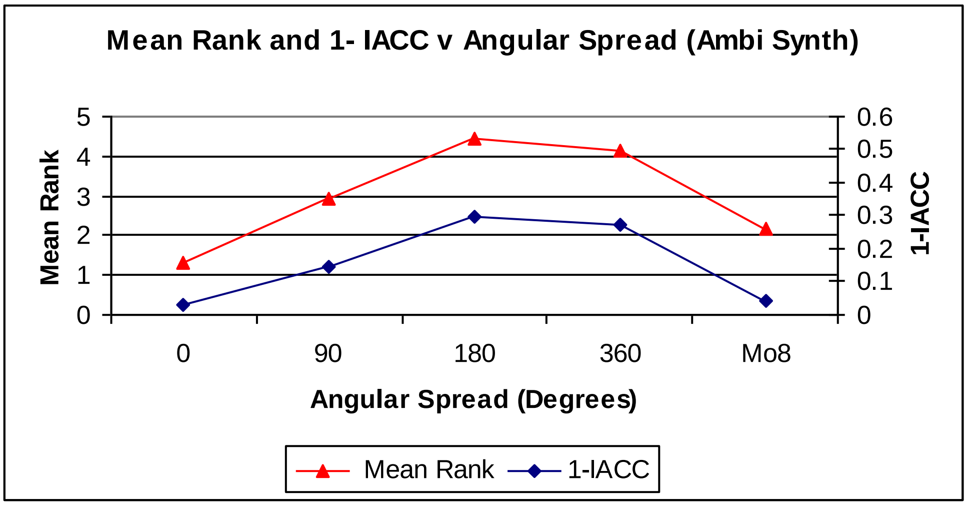


Figure 4.1 1 **–** IACC measurements and mean subjective rank for the ambisonic synthesizer sound

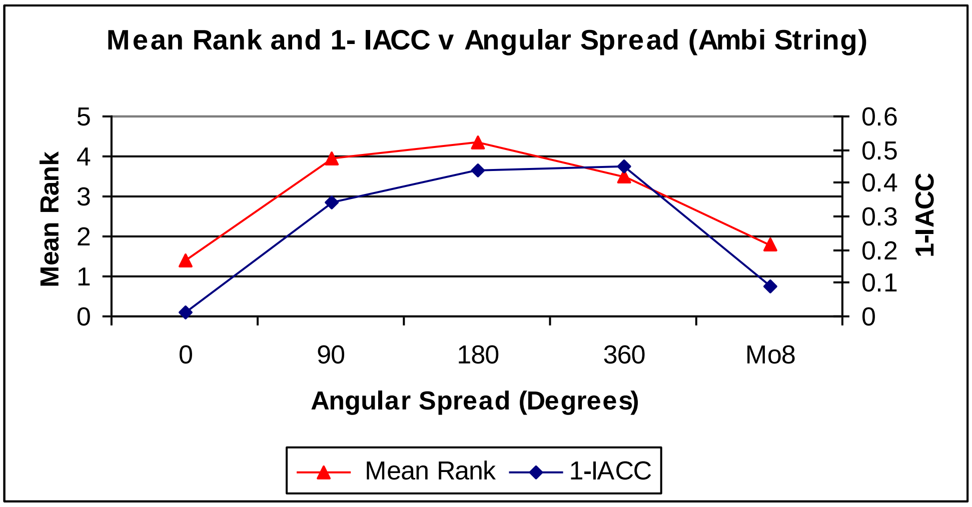


Figure 4.2 1 **–** IACC measurements and mean subjective rank for the ambisonic string sound

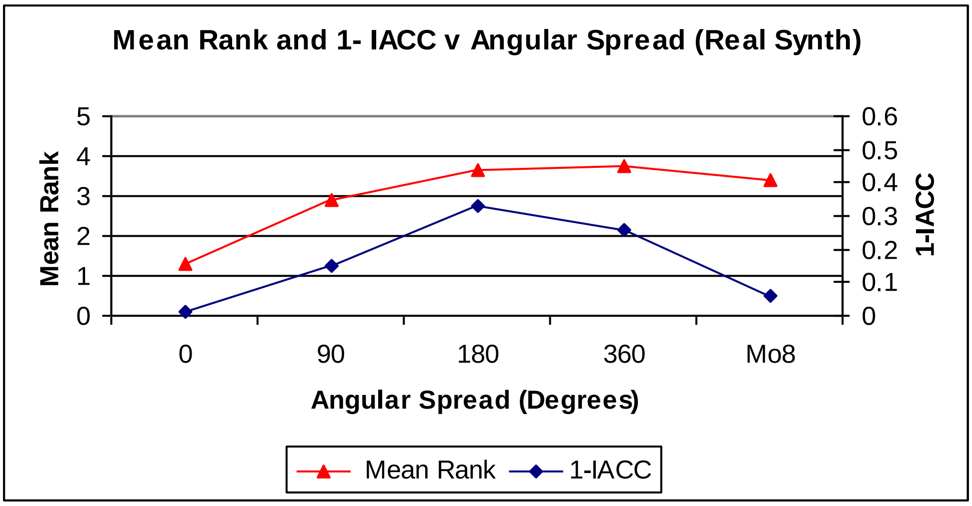


Figure 4.3 1 – IACC measurements and mean subjective rank for the synthesizer sound

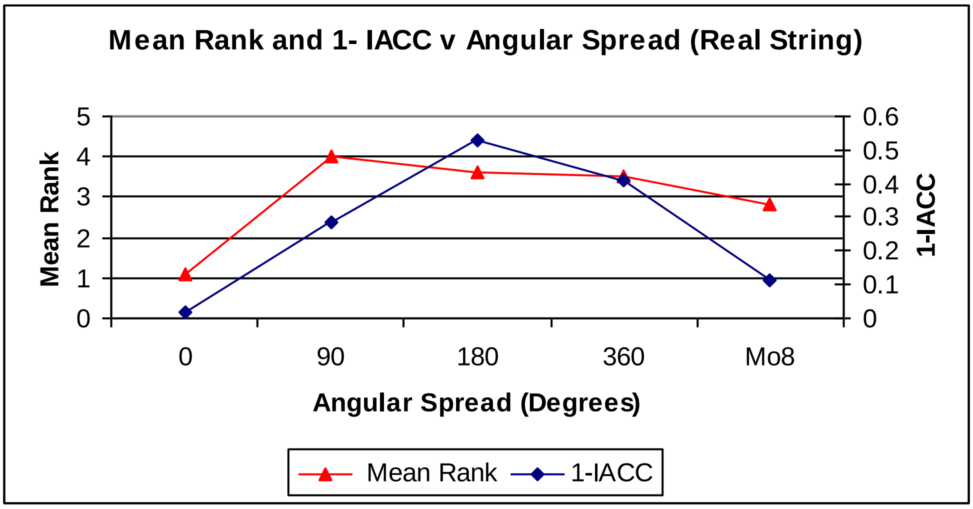


Figure 4.4 1 – IACC measurements and mean subjective rank for the string sound

A table displaying the correlation values and their associated significance levels between the mean subjective rankings and 1-IACC measurements can be seen in Table 4.1. Significant correlations are shown in red.

Audition	Correlation Coefficient	Significance
Ambisonic Synthesizer	0.978	0.004
Ambisonic String	0.942	0.017
Real Synthesizer	0.753	0.142
Real String	0.781	0.119

Table 4.1 Correlation coefficients between the 1 – IACC measurements and mean subjective rankings for the spatializing techniques

All of the subjective mean rankings exhibit a strong correlation to the IACC measurements with the ambisonic auditions exhibiting significance at the 0.05 level. This suggests that when used to compare spatial impression between related audio examples, the IACC measurement discriminates between examples of varying spatial impression and correlates well with subjective results.

Following on from these encouraging results, the possibilities of using IACC as an objective measure of spatial impression in other areas of sound reproduction were considered and are discussed in the next section.

4.3 Adaptation of IACC for Use in Reproduced Sound

Following on from using IACC as an indicator of spatial impression in the previous experiment, the possibility was explored of adapting the IACC measurement as an objective measure of perceived spatial impression in reproduced sound *in general*. In the previous example, IACC was used as a comparative measure between similar auditions. Could it be possible to adapt the IACC measurement to compare the spatial capabilities of a number of different reproduction systems or recording methods or other variable aspects in spatial audio?

Bearing in mind that the IACC measurement in concert hall acoustics is based upon the binaural impulse response of a room, when comparing the degree of spatial impression delivered by different reproduction systems, a problem arises in how to 'form' an impulse response of an audio reproduction system and its associated encoding and decoding methods.

In previous experiments using the IACC measurement, where like is being compared with like whilst a common attribute is varied, this problem has been avoided by obtaining the IACC from signals other than impulse responses. This was the case in comparing the spatial impression delivered by varying the spread of harmonics in the experiment reported previously. Various experiments have used other signals such as that conducted by Buck [Buck 1983]. Buck used IACC to determine the phantom imaging capabilities of various pairs of loudspeakers using a pink noise signal inputted to both loudspeakers of a pair. The premise was that the higher the IACC, the 'better' the phantom image. Cox *et al.* [Cox et al. 1993], in determining the difference limen for IACC used short pieces of music, whilst varying the acoustics of a simulated concert hall by altering the levels and delays of loudspeaker feeds. In both of these examples, IACC has been used as a comparative measure of a varied aspect within a particular system or set of circumstances, using signals other than an impulse response. In attempting to measure the spatial capabilities of differing spatial audio systems using IACC, the use of signals such as music or pink noise could be possible. However, the encoding and decoding of such signals (assuming them to be mono in the first place) through a spatial audio system would not necessarily result in *any* perception of spatial impression (apart from possible spatial positioning due to intended panning of the signal) as there is no spatial information present in the original signal.

A stereo signal, generated by stereo recording or amplitude panning (or both) *does* contain spatial information. If a stereo signal was used in comparing different systems, problems would arise in how to decode the signal to systems other than stereo. For example, if a stereo signal was decoded to a 5.1 system (without upmixing processing), the measured IACC would be exactly the same as for a stereo system because the centre and rear channels of the 5.1 system would not be in use. The same problem would arise using other spatial encoding methods with various decoding methods.

To overcome this problem, so long as the encoding methods of the systems under test are known in advance, the simultaneous encoding (recording) of a common sound source using all the encoding methods of the test systems could be undertaken thus allowing for a comparison of IACC measurements of the decoded source signals. This method could also be used to compare different encoding or decoding methods within a spatial audio system. For example, coincident and spaced microphone pair recordings (encodings) of a source could be simultaneously taken then replayed over stereo loudspeakers. The IACC for each encoding method could be measured and compared as an indicator of which microphone technique delivers the greater degree of spatial impression. To compare differing decoding methods in ambisonics, a Soundfield microphone recording of a source could be then decoded both periphonically and pantophonically for example. Again, the IACC measurements of the signals delivered by the two decoding methods could be compared as an indication of differences in delivered spatial impression.

To develop this comparative method further, consideration of the expectations of the comparative method of objectively measuring spatial impression in reproduced sound is necessary. In subjective appraisals of surround sound systems, the intended goal of the system needs to be realised. A surround sound system or process may have been designed to create a spatial ‘effect’ (as was the case in the spatializing techniques for musical synthesis) or to attempt to deliver a realistic, three dimensional soundfield that conveys as many aspects of the original recording space as possible. Ideally the listener will be presented with all the spatial cues that were present in the initial environment in a listening experience sometimes described as ‘You are there’ experience. In the latter case, the aforementioned method of comparing IACC measurements delivered by different systems could be developed further by comparing the IACCs of the different systems to the IACC of the original soundfield. In using this method it would be possible to use the impulse responses of the original and reproduced environments to calculate the IACC. The comparison of the original and reproduced IACC measurements could give an indication of how much auditory spatial information had been retained in the encoding / decoding process of a particular reproduction system and may therefore be used as an objective measure of spatial impression in reproduced sound.

The basis of the following investigation has been adapted from the previous work of Furlong [Furlong 1989]. Furlong compared primary (concert hall) and secondary (reproduction system) environments using, amongst other measurements, IACC in a computer simulation. In addition to IACC, Furlong measured and compared the listening level, delay times of early reflections and reverberation times of the primary and secondary environments. Ando [Ando 1985] theorised that these four parameters completely describe the acoustic properties of a concert hall and consequently (via subjective testing) assigned preferred values to each of the parameters. A total preference value, 'S', for a particular location within a concert hall can be calculated from summing the differences between the preferred and measured values of each parameter. Furlong calculated the similarity of the S values between the primary and secondary environments using a sum of squared differences approach, resulting in what he termed as the index of preference field difference (DI). The lower the DI value, the closer the secondary environment is to the primary environment. Furlong simulated mono, stereo and ambisonic reproduction and also varied loudspeaker and microphone directivities, stereo microphone techniques and absorption coefficients of the listening environment. In general, the lowest DI values were for ambisonic reproduction followed by stereo then mono. Whilst not verified by subjective testing, these outcomes are somewhat expected.

In the remainder of this chapter a method of objectively measuring spatial impression in reproduced sound is investigated in both simulated and real environments.

4.4 Outline of the Comparative Procedures Using Simulated Sound Fields

In the first part of this investigation, a simulation of the comparative procedures where the degree of retention of spatial information between an initial (real) environment and a reproduced version of the same environment is proposed as an indicator of spatial impression in reproduced sound. The simulation was carried out as a precursor to the bulk of the investigation where the procedure was conducted using a real concert hall and real reproduction systems. The outcomes of this investigation are later compared and correlated to the outcomes of a subjective test that evaluated the spatial performance of a number of reproduction systems (See Chapter 5).

4.4.1 Simulation of the Comparative Procedure Using Basic Auralization Techniques

In this section, a three-dimensional, first-order room simulation program is described. The program was used to 'record' the sounding of maximum length signals (MLS) in a virtual concert hall then 'replay' the signals through a number of reproduction systems in anechoic conditions. As an indication of the retention of spatial impression, the IACC measurements taken in both simulated environments are then compared.

In Section 4.5, the same procedures are carried out in real environments. However, due to the limitations of the simulation, direct comparisons between the output of the model described in this section and the results of the measurements taken using a real concert and real reproduction systems are not expected to be applicable as the model did not take in account the full effects of absorption and diffusion of the concert hall or higher (than first) order reflections. The main reason for conducting the simulation was to instil confidence in the methods, as the procedures in real environments would be resource intensive and time consuming. The simulation would also allow for experimentation within the procedures. The overall trends predicted by the model were expected to be reflected to some degree in the results obtained in the real environments.

4.4.2 Overview and Method

An overview of simulation procedure can be seen in Figure 4.5. The simulation involves binaural and B-format recordings of a MLS signal being taken at 24 seat positions in a simulated concert hall then replayed through simulated reproduction systems for further IACC measurements. The individual stages of the simulation procedure are outlined below. Stages 1 to 4 refer to the concert hall simulation, Stage 5 to the reproduction simulation and Stages 6 and 7 to both. All signal processing and calculations were undertaken using Matlab software. Program coding can be seen in Appendices B and C.

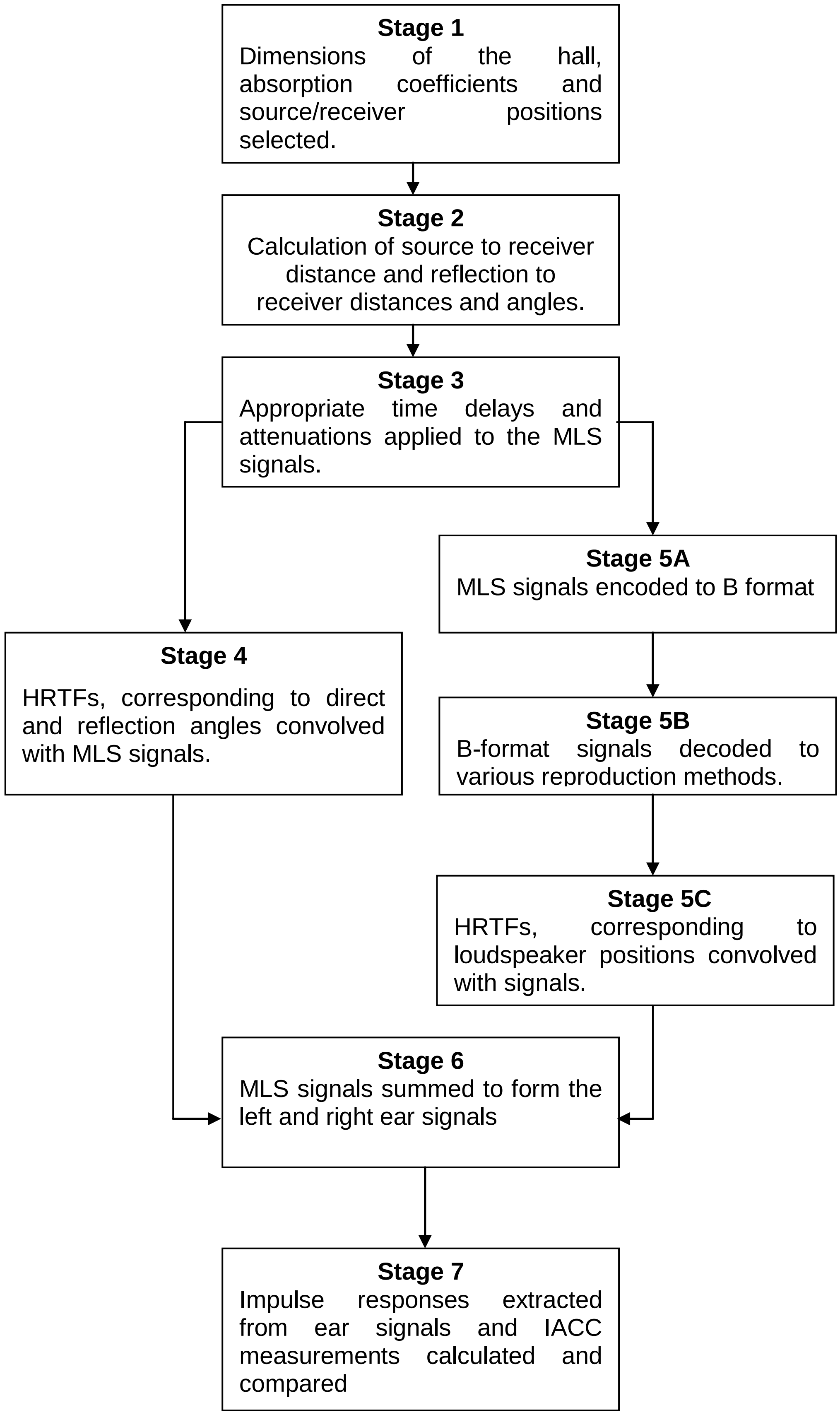


Figure 4.5 Overview of the simulated procedure for comparing IACC measurements in original and reproduced environments

Stages 1, 2 and 3 – Calcualtion of Time delays and Attenuation

The dimensions of the simulated concert hall could be specified in the program. These were set to approximate the dimensions of the concert hall later used for the ‘real’ measurements (see Section 4.5). This was approximately a ‘shoe-box’ shape; however other concert hall designs could have been investigated such as a reverse fan shape which would result in stronger lateral reflections and lower IACC measurements.

Absorption coefficients could be selected for each surface of the concert hall. Absorption coefficients that are typical of materials present in concert halls were selected that ranged from 0.5 to 0.95.

The source position was set to be in the centre of the stage and at a height of 1.6m. The source was assumed to be omnidirectional. As the simulated concert hall was symmetrical about its centre line, receiver positions were only required for one half of the hall. In order to generate a wide range of IACC measurements, evenly spaced receiver positions throughout all of one half of the concert hall were selected. For each of the 24 receiver positions, the source-to-receiver and reflection-to-receiver distances and angles were calculated using geometric methods. The receiver height was also set at 1.6m. For each receiver position, attenuations due to path differences and absorption and time delays due to path differences were calculated for the six, first-order reflections. In Furlong’s simulation a centrally placed receiver position was assumed and 25 measurements within a $1m^2$ area around this position were taken.

Whilst this method will result in some variations in the measured IACC, a wide variation in receiver positions (and hence IACC measurements due to differences in the proximity to side walls) was selected in the current simulation. This was chosen as a comparison between the ranges of IACC measurements taken in the two environments may indicate the limits of spatial impression delivered by varying reproduction systems. Certain systems may not be able to recreate spatial conditions below a certain IACC value. Furthermore, by selecting a potentially wide range of IACC measurements, the reproduction systems' ability to recreate a wide and varying range of spatial conditions may be determined. This may also indicate the spatial capabilities of the reproduction system.

Six, single period identical copies of a 16383-point maximum length sequence (MLS) signal were generated then delayed and attenuated accordingly. With only six reflections, a limited simulation of a concert hall was created. However, Ando shows that the measured degree of spatial impression of a synthetic soundfield can converge to a final value after only four reflections [Ando 1985].

Stage 4 – Convolution with HRIRs (Concert Hall)

In order to simulate pinna filtering and interaural time and level differences, each of the seven MLS signals (the direct sound and the six reflections), particular to a seat position, were convolved with a head-related impulse response (HRIR) that corresponded to the source-to-receiver or reflection-to-receiver angle. This method varies from Furlongs' in that in his simulation the binaural ear signals were derived from a widely spaced pair of omnidirectional microphones.

The HRIRs used in the simulation were taken from Gardner and Martin's set of anechoic KEMAR head measurements [Gardner and Martin 1994]. Whilst covering a large number of possible source positions, this set of HRIRs has a limited angular resolution ranging from 5° to 30° in azimuth and 10° in elevation. For each source and reflection-to-receiver angle calculated, the HRIR that was closest to the intended angle was selected for convolution.

Stages 5A and 5B – B Format Encoding and Decoding

Using the methods outlined in Stages 1, 2 and 3, a simulated B format microphone recording of the MLS signal was made. Having calculated the source-to-receiver and reflection-to-receiver angles, the velocity components (X, Y and Z) of the B format signals could be synthesized along with the omnidirectional W component.

For simulated loudspeaker reproduction, the B format signals were decoded to a number of reproduction systems. These were; mono, stereo, four and eight loudspeaker pantophonic ambisonic, eight loudspeaker periphonic ambisonic and 3/2 loudspeaker arrangement (with non-Vienna ambisonic decoding).

Stage 5C - Convolution with HRIRs (Reproduction System)

Having decoded the loudspeaker signals for each reproduction method, each individual loudspeaker signal was convolved with the closest HRIR corresponding to the angular position of the loudspeaker, relative to the listening position.

Stage 6 – Summation of Ear Signals

To simulate binaural recordings, the left and right ear signals created in Stages 4 and 5C were summed for both the simulated concert hall and simulated reproductions, respectively. From these binaural ‘recordings’ of the MLS signals, the impulse response was extracted and from this the IACC was calculated.

Stage 7 – IACC Comparisons

Impulse responses were extracted from the summed MLS signals and IACC measurements calculated for both the simulated concert hall and reproductions. Contrasts between the degree of spatial impression delivered by each reproduction method could be carried out by comparing concert hall to reproduced IACC measurements.

4.4.3 Simulation Results

Figure 4.6 displays the variations in IACC throughout the simulated concert hall and audio reproductions of the concert hall. The graphs can be thought of as spatial impression ‘maps’ of the concert hall. In calculating IACC the whole of the impulse response was used (no windowing) and frequency filtering was not implemented.

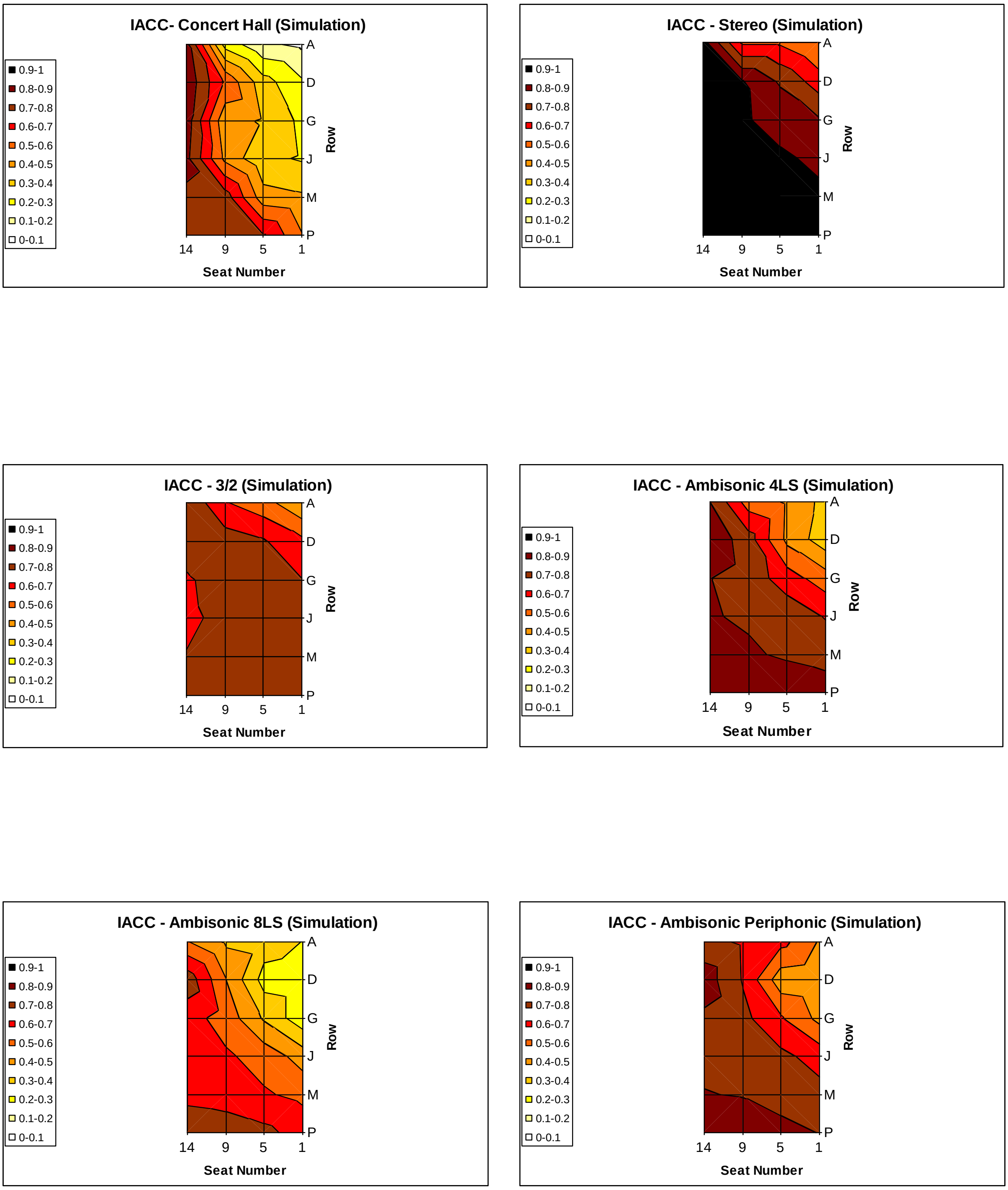


Figure 4.6 IACC measurements in the simulated original and reproduced concert halls

The graphs can be interpreted as a plan view of the right hand side of the concert hall. The stage is at the top of the graph and the rear of the hall at the bottom (x-axis). The centre of the hall is the extreme left of the graph (y-axis) and the right hand sidewall being at the extreme right. The lighter the shading in the graph, the lower the IACC measurement, thus the greater the degree of perceived spatial impression.

A plot of the mono reproduction has not been included as all the IACC measurements were either one or very close to one. The trend displayed in the simulated concert hall graph in the top left hand corner of Figure 4.6 shows that IACC tends to decrease with proximity to the sidewall and that IACC remains at a fairly high level over the central area. The decrease in IACC can be explained by the presence of strong lateral reflections, close to the sidewall, causing interference with the direct signal that result in a greater dissimilarity (lower IACC) between the left and right ear signals. Towards the centre of the hall, the left and right ear signals will tend to be similar, due to the symmetrical nature of the hall, thus leading to high values of IACC. The range of the IACC measurements in the simulated concert hall has a minimum of 0.083 and a maximum of 0.869.

From examination of the graphs, the visual similarity between the simulated concert hall and simulated reproduction systems is varied. The stereo and 3/2 graphs do not show as much variation in IACC as the concert hall graph and IACC is generally higher than the other reproduction systems. The stereo IACC ranges form 0.497 to 0.956 and the 3/2 from 0.416 to 0.719. The ambisonic four-loudspeaker, eight-loudspeaker and periphonic systems display similar variations in IACC as the concert hall graph, and the IACC values, whilst still higher than, are closer to the concert hall values than the stereo and 3/2 systems. This is particularly true for the eight-loudspeaker pantophonic graph. The four loudspeaker periphonic IACC ranges from 0.367 to 0.874, the eight loudspeaker periphonic from 0.291 to 0.762 and the pantophonic from 0.491 to 0.855.

The stereo, 3/2 and pantophonic reproduction methods all had relatively high minimum IACC measurements, with the other reproduction systems having lower minimums but still a long way from the concert hall minimum. The maximum IACC measurements for the reproduction systems were all close to that of the concert hall. The stereo and 3/2 systems also display little variation in measured IACC within their reproductions of the concert hall. Due to the non-optimised encoding methods for these systems (i.e. derived from a Soundfield microphone recording), the lower IACC measurements were not comparable to those of concert hall. This may be explained by directivities of some of the concert hall reflections becoming uniform (to some degree) thus reducing the interference effects that lead to decorrelation. This would help explain the high minimum IACC in stereo reproduction as the reproduced soundfield extends only to a subtended angle of 60°. In the case of the 3/2 system, the soundfield is extended, however the high minimum IACC and lack of variation in IACC may be explained by the irregular loudspeaker layout that is not suitable for uncompensated ambisonic decoding. Better results might of been obtained using a standard coincident pair for stereo and a recognised multi-microphone technique for the 3/2 system. The periphonic reproduction also exhibits a high minimum IACC value with limited variation in the IACC measurements when compared to the 4 and 8 loudspeaker pantophonic reproductions. This may be due to there being no loudspeakers in the horizontal plane in periphonic reproduction, thus reducing the strength of lateral reflections.

Having calculated IACC for the simulated concert halls and reproduction methods, a method of comparing the two measurements is required. Both correlation between the original and reproduced IACC measurements and an IACC-only version of Furlong’s index of preference field difference are considered.

A graph of IACC versus seat position can be seen in Figure 4.7. By arranging the seat numbers of the simulated concert hall in ascending value of IACC then doing the same for the reproduction systems, the two measures of IACC can be compared and correlated.

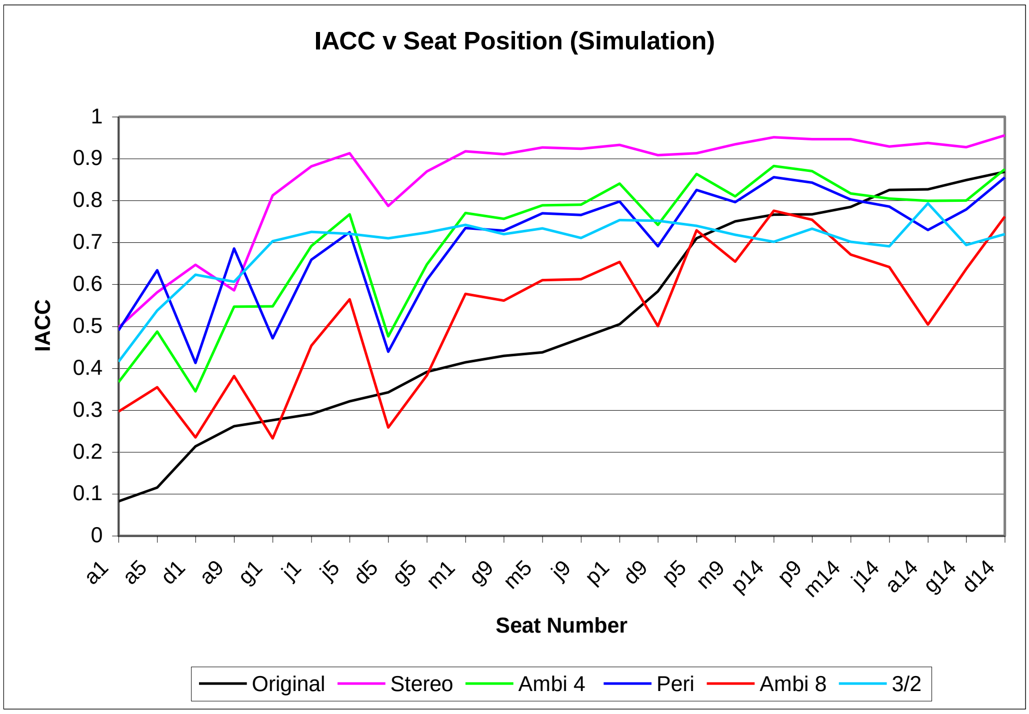


Figure 4.7 IACC measurements versus seat number in the simulated original and reproduced concert halls

In Figure 4.7 it can be seen that all of the reproduction systems measurements generally follow the same trend as the concert hall measurements with the plots displaying a rise in IACC as the seat numbers become more central (seat number 14 being at the centre of the concert hall). At low IACC values the reproduction system plots are furthest from the concert hall measurements, whilst at high IACC values the plots tend to converge. At the highest IACC values the reproduction systems generally have a lower IACC value than the concert hall measurements. This is also the case for pantophonic ambisonic eight-loudspeaker system at lower IACC values. Due to the limitations of spatial reproduction systems in general, it seems improbable that the concert hall IACC measurements should be higher than the reproduction systems IACC measurements as this would suggest that the degree of spatial impression was sometimes greater in the reproduced versions of the concert hall.

Table 4.2 displays correlation values between the IACC measured in the simulated concert hall and simulated reproductions. The correlation values shown in red are significant to the $p < 0.01$ level.

System	Mono	Stereo	3/2	Ambi 4 LS	Ambi 8 LS	Ambi Peri
Correlation Coefficient	0	0.76	0.58	0.81	0.80	0.75

Table 4.2 Correlation coefficients between IACC measurements in the simulated concert hall and reproduction systems

Whilst the degree of correlation demonstrates how closely *variations* in IACC in the concert hall are matched in the reproduction systems' version of the concert hall, the correlation value will say little about how close the *actual* concert hall and reproduced IACC measures are. This can be seen when comparing the correlation coefficients of Table 4.2 and the plots in Figure 4.7. The stereo correlation coefficient is 0.76, which is the third highest of the group. However, when comparing the original and stereo plots in the graph, the actual differences in IACC values can be seen to be large, especially for low IACC values. As the object is to indicate which reproduction system retains the greatest degree of spatial impression, as indicated by comparing IACC measurements, correlation whilst useful, does not fulfil this objective.

Shown in Table 4.3 are the IACC-only indexes of soundfield difference measurements. These have been named as spatial retention (SR) measurements. The SR values are calculated using Equation 4.1, where OR and RP are the IACC measurements taken in the original concert hall and reproduced concert hall respectively. The lower the SR value, the greater the degree of spatial retention.

$$SR = \left(\sum_{i=1}^{N} (OR_i - RP_i)^2 \right)^{\frac{1}{2}}$$

Equation 4.1 Spatial retention calculation

Shown in Table 4.3 are the calculated SR values for the simulated reproduction systems

System	Mono	Stereo	3/2	Ambi 4 LS	Ambi 8 LS	Ambi Peri
SR	2.68	1.88	1.36	1.22	0.74	1.26

Table 4.3 SR values of the simulated reproduction systems

The SR value is probably more useful than correlation in comparing the IACC measurements as it is a better indicator of how similar (as a result of fewer errors between the reproduced and original IACC measurements) the two sets of IACC measurements are. The similarity of the simulated eight-loudspeaker pantophonic ambisonic system to the simulated concert hall is reflected in the minimal SR value (0.74) recorded by the system. This was almost 0.5 SR points lower than any other system. Conversely, stereo recorded a high score of 1.88 and the periphonic system faired the worse of all the ambisonic systems. The difference between the stereo and the next highest scoring system (3/2) was over 0.5 SR points. This suggests that reproduction systems utilising more than two loudspeakers retain a greater degree of spatial impression.

The ranking of the systems in terms of spatial retention indicated by the SR values follows a somewhat expected pattern. The performance of the periphonic ambisonic system, in being ranked below the four and eight loudspeaker pantophonic ambisonic systems was slightly surprising to the author as the inclusion of the height dimension in reproduction would allow floor and ceiling reflections to be reproduced. The inclusion of these reflections in the formation of the impulse response would be expected to result in IACC measurements that are closer to the IACC measurements taken in the concert hall. As theorised by Gerzon, [Gerzon 1985] the use of greater numbers of loudspeakers in ambisonic systems results in improved reproduction. This is reflected by the SR values for the four and eight-loudspeaker systems.

4.4.4 Simulation Summary

IACC measurements in a simulated concert hall were compared to IACC measurements in reproduced versions of the same concert hall. As an indication of the spatial capabilities of the reproduction systems the IACC measurements were compared in terms of minimum IACC measurements, correlation and SR values (sum of squared errors). The comparisons suggest (ignoring mono reproduction) that in terms of the retention of spatial impression, stereo faired worse than systems using more than two loudspeakers. Eight loudspeaker pantophonic reproductions fared the best. Having extensive listening experience of the spatial capabilities of these reproduction systems, the author was not surprised by these outcomes. Ambisonic reproduction in comparison to stereo would be expected to have better spatial performance as a much larger soundstage (360° in azimuth) is encoded and decoded in the ambisonic process. As a consequence of the simulation outcomes, confidence in the procedures in general was enhanced and steps towards a non-simulation approach initiated.

4.5 Outline of the Comparative Procedures Using Real Sound Fields

4.5.1 Overview

Following on from the encouraging outcomes of the simulated procedure, it was decided to repeat the procedure using a real concert hall and real reproduction systems. The basis of the experimental method involved the comparison of IACC measurements taken at various positions within a concert hall to measurements taken in reproduced representations of the same concert hall. The method of collecting the IACC measurements conformed to ISO measurement procedures [BS EN ISO 3382 2000]. A flow chart of the procedure is depicted in Figure 4.8.

The overall procedure is similar to the previously outlined simulated procedure. For the comparative procedures described in this section a number of additions to the procedures outlined in Furlong's work have been introduced:

- Real concert hall and reproduction system impulse responses were used to calculate the SR measurements rather than a simulation.

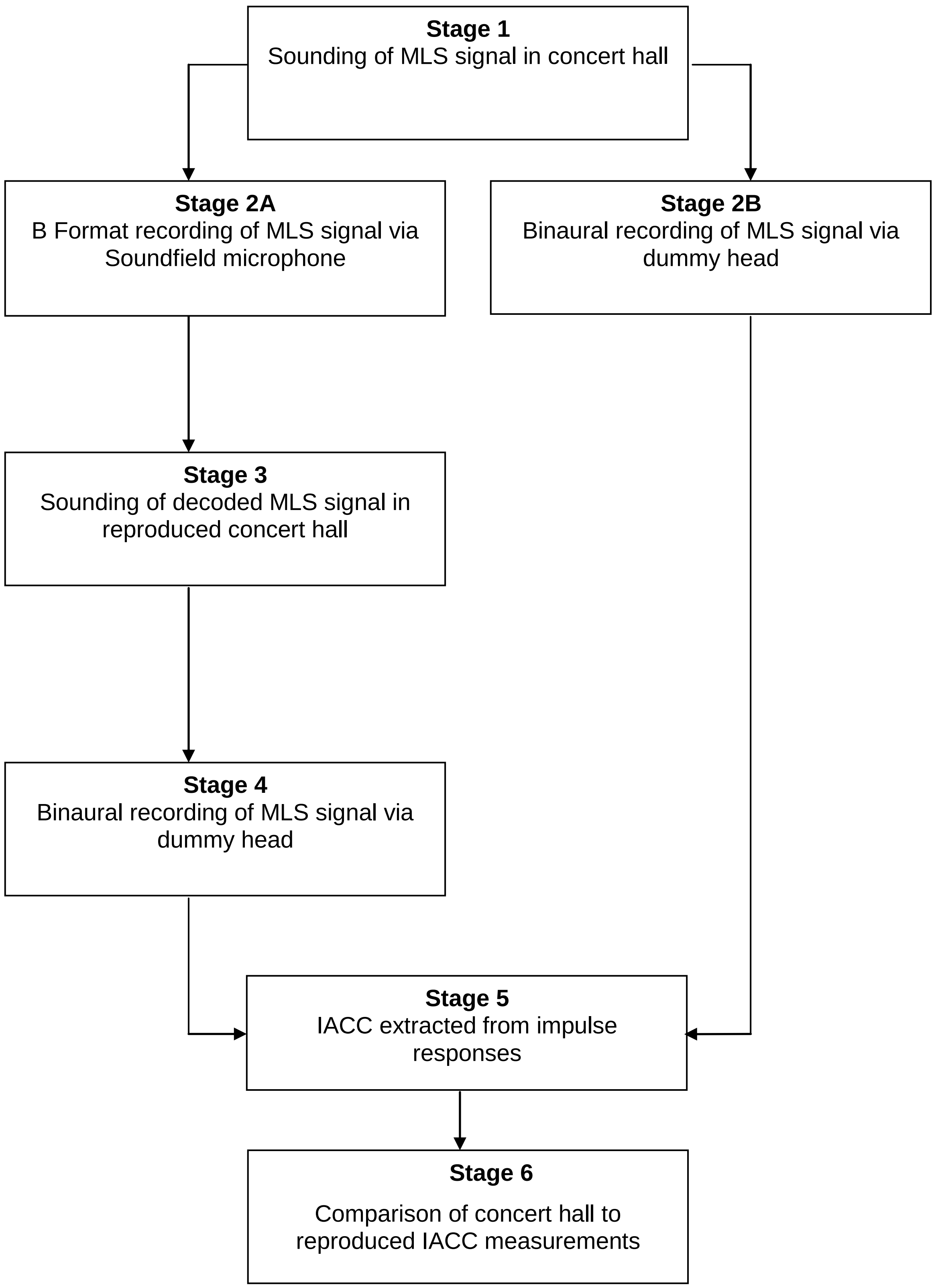


Figure 4.8 Overview of the procedure for comparing IACC measurements in original and reproduced environments

- IACC measurements were taken and MLS signals recorded across the whole concert hall rather than a small area, thereby allowing for large variations in IACC to be measured. The reproduction systems' ability to reproduce a wide variation of IACC measurements may be an indication of the systems' spatial capabilities.

- Variations in the IACC measurement (time windowing and frequency filtering) were used in calculating the SR measurements. Certain IACC variations may be better suited to reproduced sound measurements than others.

The following section details the stages outlined in the flow chart shown in Figure 4.8.

4.5.2 Stages in the Procedure

Stage 1 - Generation and Sounding of the MLS Signal

A 16383-point maximum length sequence (MLS) signal was generated at a sample rate of 44.1 kHz with a 16-bit resolution. On sounding the signal, eight periods of the MLS signal were generated to increase the signal to noise ratio. The signal was sounded using a PC with a multichannel soundcard connected to a digital to analogue converter. The analogue output signal was inputted to a power amplifier that was connected to a dodecahedral omnidirectional loudspeaker placed at the centre of the stage at a height of 1.6m.

The measurements and recordings were taken in Peel Hall, a medium sized (384 seats) concert hall situated in the campus of the University of Salford. The hall is rectangular in shape with a semi-circular rear wall. The dimensions of the concert hall are approximately 20m wide, 35m long and 20m high with the seats arranged in 16 rows of 24 seats. The seats are set on a sloping surface with the front seats being at stage level and the rear seats being approximately 10m above the stage. A photograph of the hall, looking from the rear of the hall to the stage can be seen in Figure 4.9.

Figure 4.9 Photograph of Peel Hall looking towards the stage

Stages 2A and 2B - Recording of the MLS Signal

The MLS signal was sounded twice to allow for two recordings of the signal. A Soundfield ST250 microphone was used for one of the recordings, allowing for B-format recording of the signal. A Bruel and Kjaer head and torso simulator (HATS) was used to record the binaural signals, allowing for the extraction of impulse responses from the MLS signals and ultimately for IACC measurements.

Using a multichannel soundcard, software and PC, simultaneous soundings and recordings of the MLS signal were taken at 24, evenly spaced seat positions and saved to the hard drive of the computer as 16-bit, 44.1 kHz PCM files. As the concert hall was symmetrical around the mid-line, measurements were only taken on the right-hand side (looking towards the stage) of the hall. Measurements were taken every fourth row (starting with the row nearest the stage and finishing on the last row) and every fifth seat (starting from the extreme right of the hall and finishing in the centre). This resulted in four measurements for each of the six rows.

Stage 3 – Sounding of the Decoded MLS Signal in the Reproduced Concert Hall

The reproduced measurements took place in the semi-anechoic chamber of the School of Acoustics and Electronic Engineering, University of Salford. The decoded signals for each reproduction system and for each seat position were sounded. Whilst general ambisonic decoding details can be seen in Section 2.3.8.2, details of the decodings of the B format signals for each system are shown in Appendix D.

The twelve loudspeakers required for auditioning the five non-periphonic systems were positioned upon stands, at head height and at distance of 1.15m from the listening position. The eight loudspeakers required for periphonic presentations were attached to a framework that surrounded the listening position. As the framework and the non-periphonic systems could not occupy the same space, the framework that supported the periphonic system arrangement did not form a perfect cube. In addition, the loudspeakers were placed a little further from the listening position (1.25m). To minimise reflections, the solid floor of the semi-anechoic room was covered with acoustically absorbent foam. All twenty of the loudspeakers used in the tests were Genelec 1029As that were level aligned using pink noise and a sound level meter. A photograph of the set up can be seen in Figure 4.10.

The decoded B format samples were replayed using a multichannel audio software package installed on a computer equipped with a multichannel soundcard that was connected to digital to analogue converters that, in turn, were connected to the loudspeakers. The computer and digital to analogue converters were located outside of the semi-anechoic chamber.

Figure 4.10 Photograph of the loudspeaker array used for replaying the MLS signals recorded in the concert hall. Also present is the dummy head used to binaurally record the loudspeaker output.

Stage 4 – Recording of Reproduced MLS Signal

The reproduced MLS signals were binaurally recorded in the same manner as described in Stage 2B.

Stage 5 - Extraction of the IACC Measures

Impulse responses were extracted from the binaural MLS recordings and then by utilising Equation 2.8 and octave band filtering, IACC and a number of variants of IACC were calculated for both the original and reproduced environments. These variations were included as certain

IACC measurement variations may be better suited to reproduced sound than others. The calculations were made using each IACC variant for the 24 seat positions in both the original and reproduced concert halls. The IACC variations (all of which have been previously used in concert hall acoustics) used in the test are listed below:

- $IACC_{FB}$ - full bandwidth IACC with no time window
- $IACC_{E}$ - full bandwidth IACC using a 0 to 80 ms time window
- $IACC_{3}$ - average of the 0.5k, 1k and 2k Hz octave bands with no time window
- $IACC_{E3}$ - average of the 0.5k, 1k and 2k Hz octave bands using a 0 to 80 ms time window
- $IACC_{L3}$ - average of the 0.5k, 1k and 2k Hz octave bands using a 80 to 750 ms time window

Stage 6 – Comparison of Concert Hall to Reproduced Sound IACC Measurements.

In order to evaluate the retention of spatial impression in each reproduction system, the original and reproduced IACC measurements were compared. A number of methods of comparison were implemented including correlation and SR. These are further discussed in the next sections.

4.5.3 Results of the Comparative Procedures

Figures 4.11 to 4.15 display the spatial measurements taken in the original and reproduced halls for $IACC_{FB}$, $IACC_{E}$, $IACC_{3}$, $IACC_{E3}$ and $IACC_{L3}$ respectively.

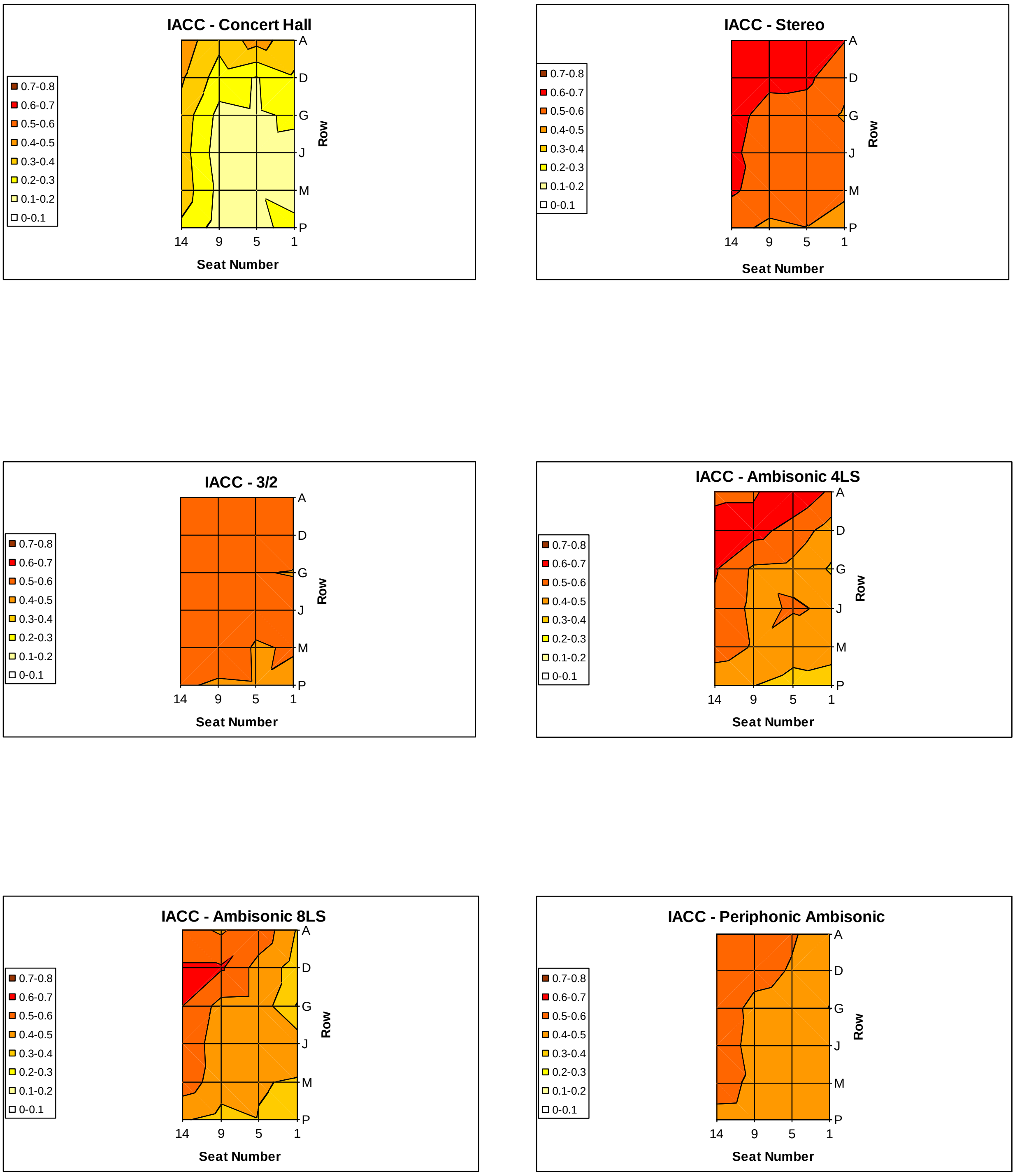


Figure 4.11 $IACC_{FB}$ measurements in the original and reproduced concert halls

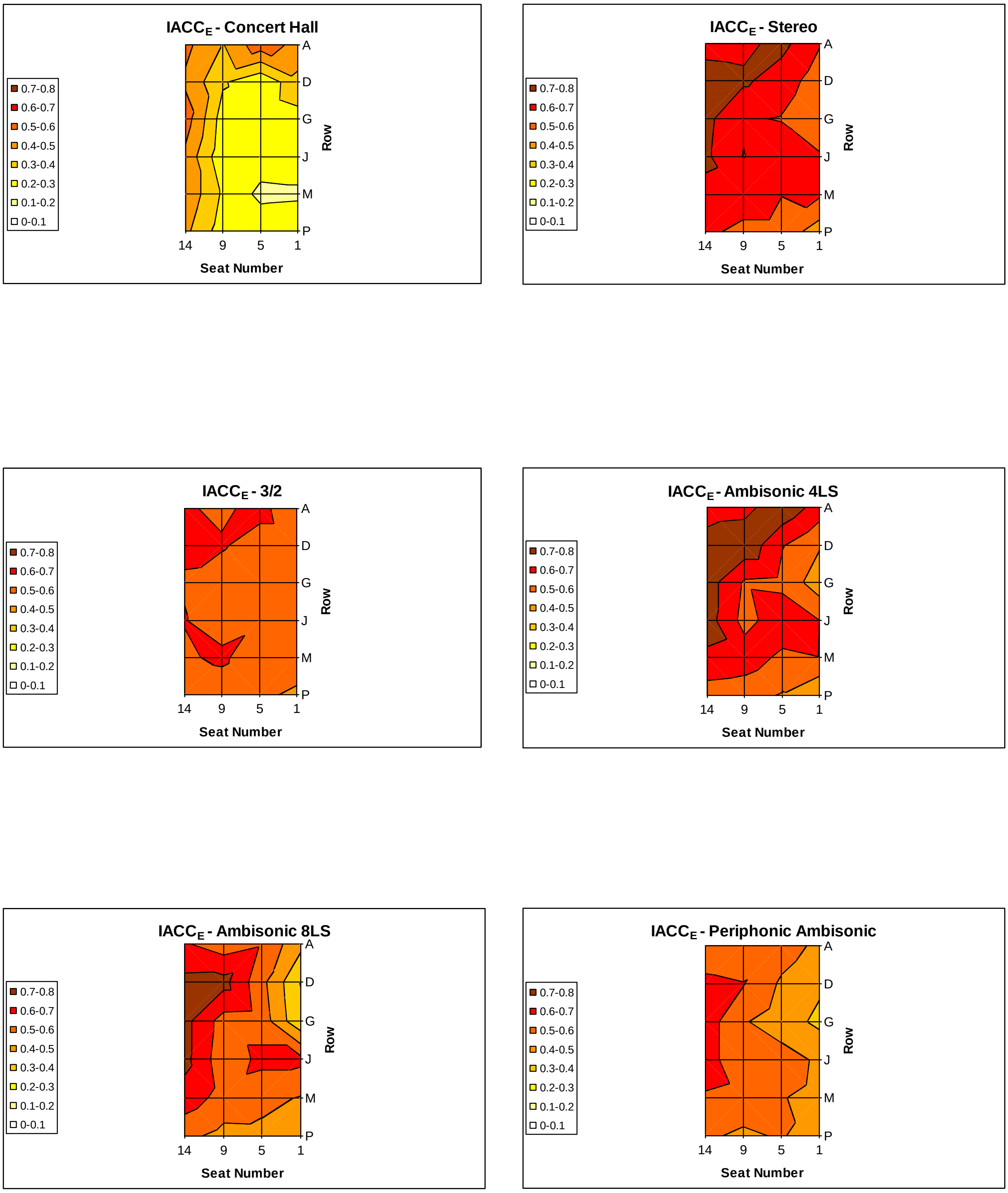


Figure 4.12 IACC$_E$ Measurements in the original and reproduced concert halls

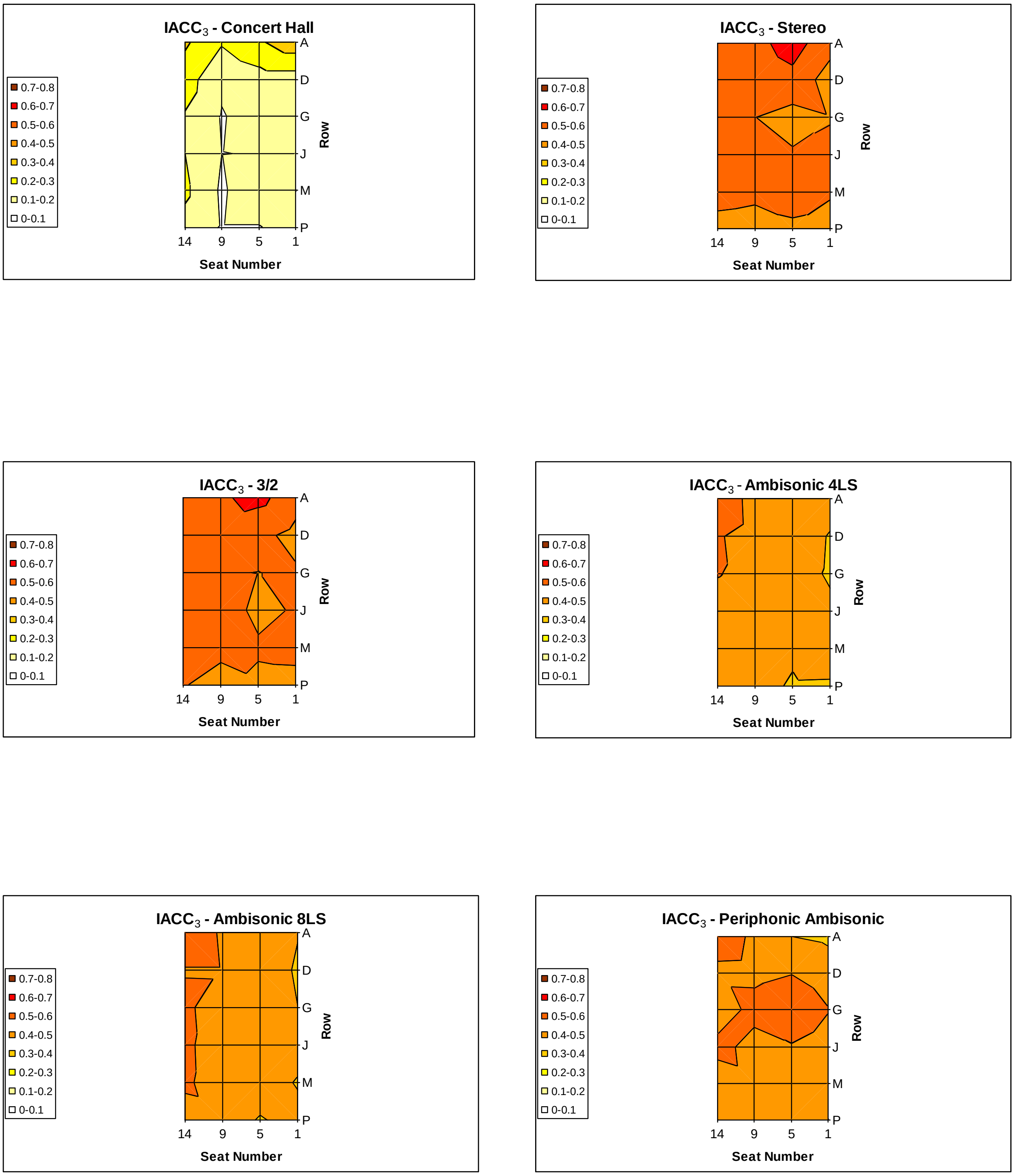


Figure 4.13 IACC$_3$ Measurements in the original and reproduced concert halls

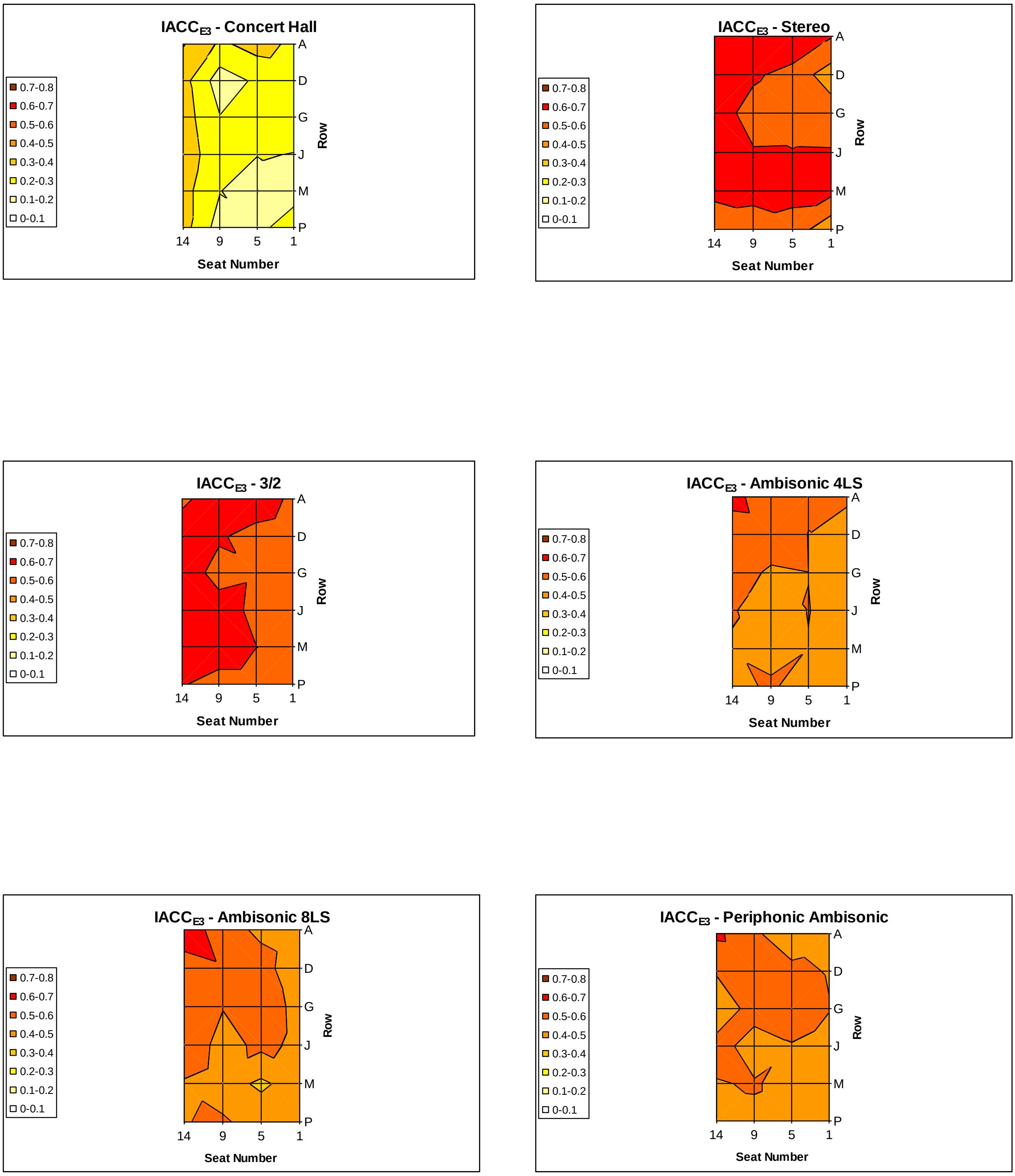


Figure 4.14 IACC$_{E3}$ measurements in the original and reproduced concert halls

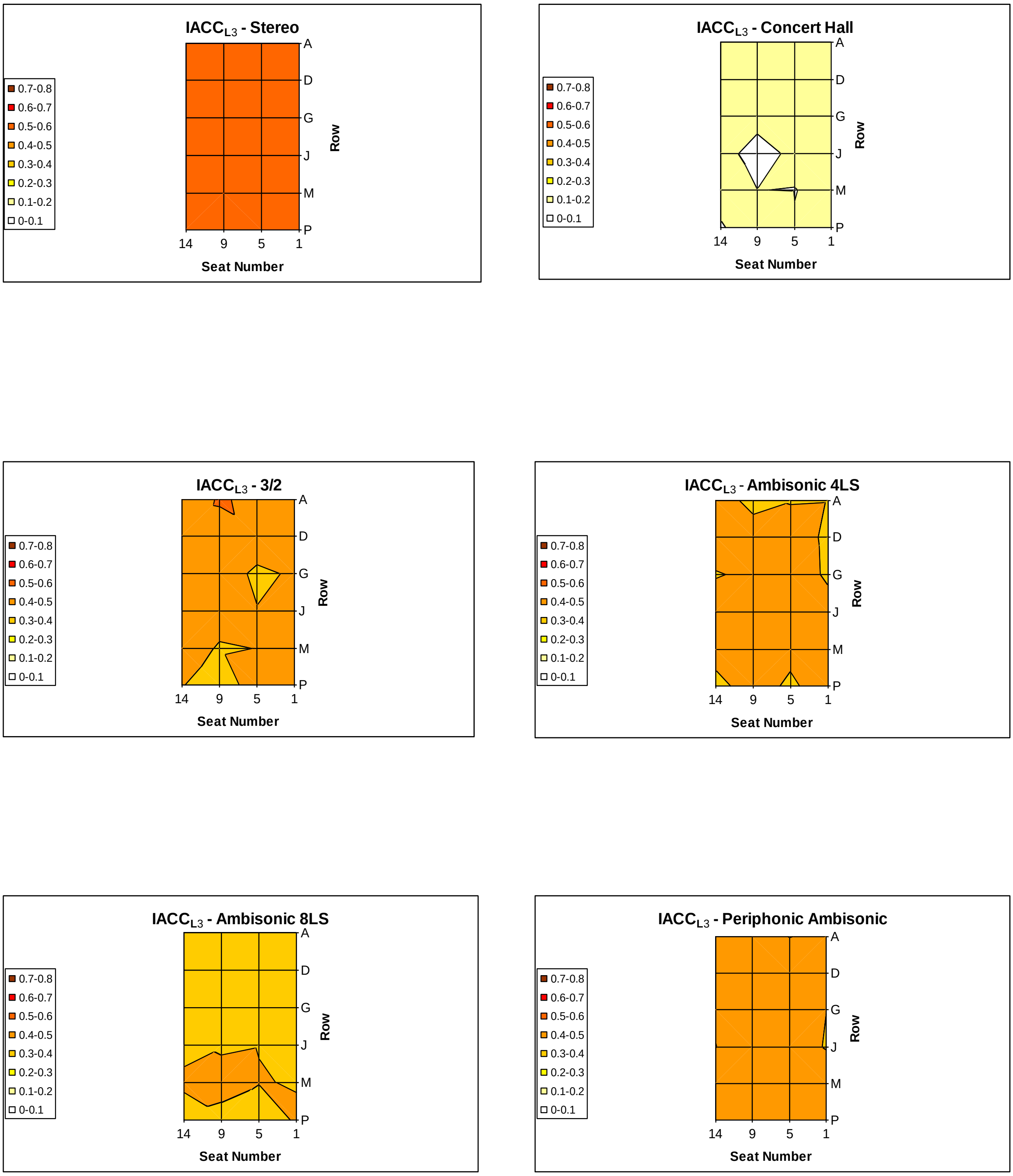


Figure 4.15 IACC$_{L3}$ measurements in the original and reproduced concert halls

The interpretation of the graphs was discussed in Section 4.4.3. An aisle that was approximately two seat spaces wide was present in between seats 5 and 6. Consequently, seat numbers 7 and 12 appear as 9 and 14 respectively in the graphs.

The trend observed in the simulation, in that IACC decreases with proximity to the sidewall, is reflected in the original concert hall graphs of Figures 4.11, 4.12, 4.14 and 4.14 ($IACC_{FB}$, $IACC_{E,}$ $IACC_3$ and $IACC_{E3}$, respectively) but to a much lesser degree in Figure 4.15 ($IACC_{L3}$). The insensitivity of $IACC_{L3}$ as a spatial measure is even greater in reproduced environments. The $IACC_{L3}$ measurements for stereo and periphonic ambisonic reproduction are almost unvarying throughout the reproduced concert halls. For this reason, $IACC_{L3}$ is not considered useful as an indicator of spatial impression for this investigation and is not further considered.

Also described previously, comparisons and correlations between the original and reproduced concert halls were made by arranging the data in order of ascending spatial measurement by seat position in the original concert hall, then plotting $IACC_{FB}$, $IACC_E$, $IACC_3$ and $IACC_{E3}$ measurements taken in the reproduced concert hall versus the same seat positions. These graphs are shown in Figures 4.16, 4.17, 4.18 and 4.19, respectively.

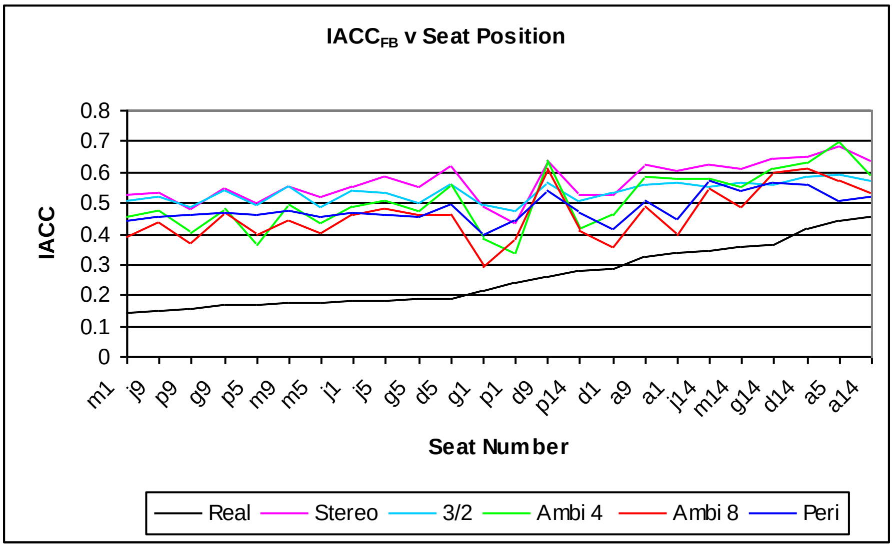


Figure 4.16 IACC$_{FB}$ Measurements versus seat number in the real and reproduced concert halls

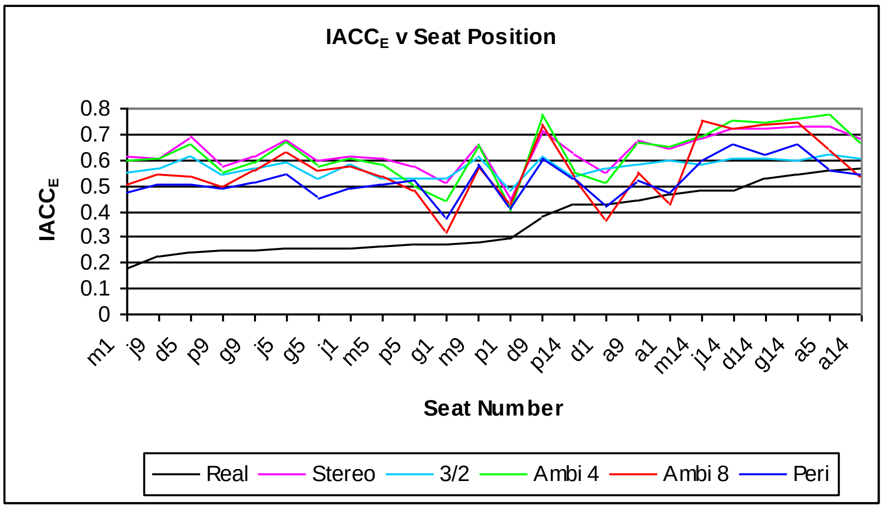


Figure 4.17 IACC$_E$ measurements versus seat number in the real and reproduced concert halls

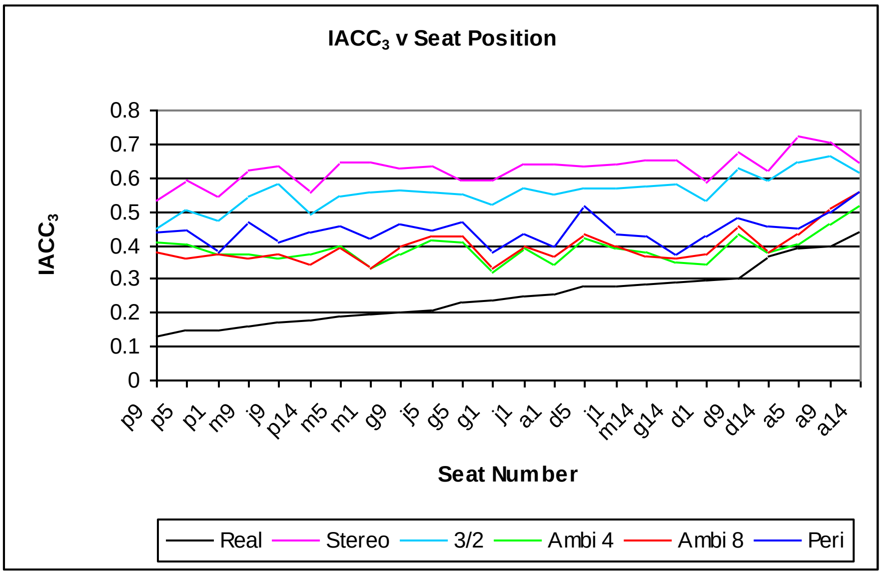


Figure 4.18 IACC$_E$ measurements versus seat number in the original and reproduced concert halls

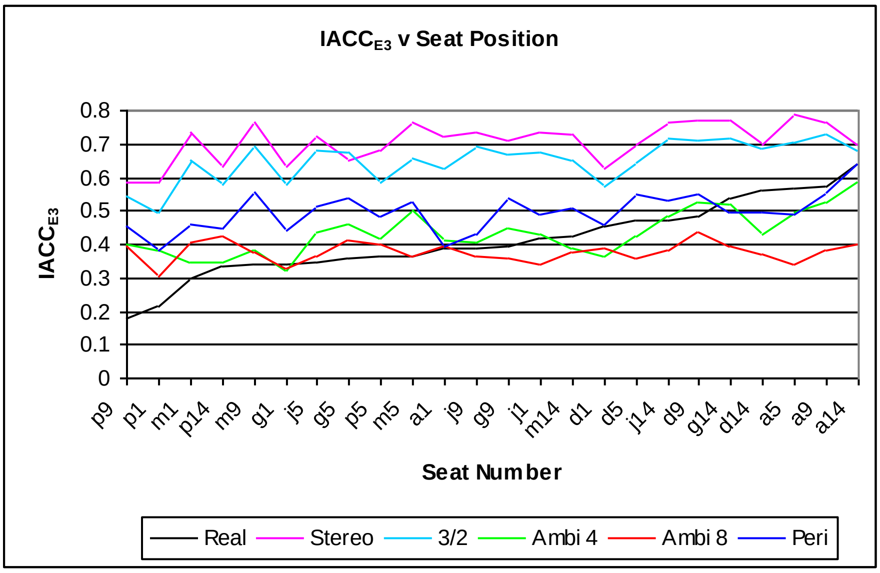


Figure 4.19 IACC$_{E3}$ measurements versus seat number in the original and reproduced concert halls

An important feature of these graphs is the expected rise in IACC as the plots for reproduction systems go from the left to the right hand side of the graph. For the concert hall plots (black line), the data has been arranged so that this is the case. Some of the plots for the reproduction systems display this feature whilst others only vary in IACC by a small amount for all seat positions or do not start to rise in IACC until approximately the mid-point on the x-axis. This suggests that the reproduction systems may have limits to the lowest (and highest) degrees of deliverable correlation in their reproduced soundfields. This could have a large effect on the SR values as if a reproduction system cannot deliver soundfields with an IACC of less than a certain amount, the errors between the IACCs measured in the original and reproduced environments could be large at low IACC values, therefore distorting the SR value somewhat. A table displaying the lowest measured IACC value for the concert hall and reproduction systems is shown in Table 4.4.

	Minimum $IACC_{FB}$	**Minimum $IACC_{E}$**	**Minimum $IACC_{E3}$**	**Minimum $IACC_{3}$**
Original Hall	0.142	0.175	0.179	0.128
Stereo	0.435	0.445	0.584	0.536
3 / 2	0.476	0.484	0.498	0.450
4 LS Ambi	0.338	0.405	0.328	0.333
8 LS Ambi	0.295	0.315	0.363	0.332
Peri Ambi	0.398	0.376	0.380	0.375

Table 4.4 Minimum IACC measurements taken in the original and reproduced concert halls.

The minimum IACC measurements shown in Table 4.4 display noticeable differences across systems with the ambisonic systems showing the lowest IACC measurements. However, there is a large difference between the minimum IACC measurements for the concert hall and all of the reproduction systems. For this reason, it may be preferable to only take into consideration the higher IACC measurements when calculating the SR values. This is further investigated in the next Section.

SR values for the six reproduction systems and four IACC measurement variations can be seen in Table 4.5.

	Mono	**Stereo**	**3/2**	**Ambi 4 LS**	**Ambi 8 LS**	**Ambi Peri**
SR ($IACC_{FB}$)	3.62	1.56	1.41	1.28	1.07	1.17
SR ($IACC_{E}$)	4.33	1.45	1.17	1.40	1.18	0.95
SR ($IACC_{3}$)	3.64	1.86	1.53	0.77	0.78	1.01
SR ($IACC_{E3}$)	2.88	1.50	1.23	0.41	0.44	0.60

Table 4.5 SR Values for the reproduction systems using different IACC measurements

4.6 Discussion

By comparing IACC measurements taken in the original concert hall with those taken in a reproduced version of the same concert hall, an indication of the spatial capabilities of the reproduction system may be made in terms retention of spatial impression. This procedure was initially conducted as a simulation. Spatial impression ‘maps’ of the concert hall were produced that demonstrated that the simulated procedure was capable of exhibiting expected trends (a reduction in IACC close to a side wall) and that these trends were reflected (to varying degrees) in the maps of spatial impression for a number of different reproduction systems. Whilst being a useful visual indicator of spatial impression, the maps do not give a numerical indication of the spatial performance of the reproduction systems. A measure of the differences between the concert hall map and the reproduction system map was calculated by summing the squared differences of IACC measurements between the two maps to produce SR values. The premise was that the smaller the summed differences between the two, the greater the retention of spatial impression by the reproduction system. The results of the simulation were encouraging as the SR measurements for the systems tested followed an expected pattern in that stereo out performed mono, non-Vienna decoded 3/2 ambisonics performed better than stereo but not as well as the regular (equally spaced) ambisonic systems. Of the ambisonic systems, eight-loudspeaker pantophonic fared the best, followed by four-loudspeaker pantophonic, then eight-loudspeaker periphonic.

The procedures were then carried out using a real concert hall and real reproduction systems. Variations in the type of IACC measurement used to calculate the SR values were also introduced as these may be better suited to reproduced sound and provide a more ‘sensitive’ measurement. The spatial impression maps again exhibited expected trends, although not as distinctly as was seen in the simulation. For all the IACC variation maps, stereo and 3/2 systems displayed fairly uniform measurements throughout their reproduced concert halls. For these systems, it could be expected that due to a limited spatial panorama (stereo) and non-optimal decoding (3/2), the measurable degree of spatial impression may be limited.

The SR values vary depending upon which IACC measurement has been used to make the calculation. For all IACC variations, the highest SR values were recorded by the mono system then followed by the stereo system. In all but one case, ($IACC_E$) the 3/2 system had the next highest SR value. Depending on the IACC variation used in the calculation, the SR values for the regular ambisonic systems varied, with the eight-loudspeaker system tending towards a low SR value for all IACC variations.

If the systems are rank ordered, according to the SR value obtained from $IACC_{FB}$ measurements, starting with the greatest difference between original and reproduced sound fields, the systems are ranked: mono, stereo, 3/2, four-loudspeaker ambisonic, periphonic ambisonic and then eight-loudspeaker ambisonic. In the results from the simulation, the rank order was almost the same, apart from periphonic and four-loudspeaker ambisonic exchanging ranks. The differences between systems as indicated by the SR values are also comparable (the average SR value for the simulation was 1.52 whilst the real concert hall SR averaged 1.68). The similarity of these results is encouraging as this demonstrates a degree of robustness in the procedures in that the 'real' and 'simulated' results are comparable.

The non-simulated procedure utilised four IACC measurement variations. By examining the output, an idea of which IACC measurement is best suited to reproduced sound may be gained. Additionally, the ways in which the original and reproduced spatial measurements are compared in the formation a spatial retention indicator are investigated.

Figures 4.11 to 4.15 display the plan views of the original and reproduced concert halls for the different IACC measures. $IACC_{FB}$ and $IACC_{E}$ plots in Figures 4.11 and 4.12 are fairly similar with $IACC_{E}$ perhaps showing a little more detail (variation) in the spatial measure of the reproduced concert halls. The general trend in comparing the original to reproduced plots for all four IACC measurements is that there is less variation in measurements displayed in the reproduced plots and that the measurements are generally higher than the original concert hall measurements.

In Figures 4.16 to 4.19, the variations in the four IACC measures over seat positions for the original concert hall and five reproduction systems can be seen. A difference can be noted between the graphs of the frequency filtered and non-frequency filtered IACC measurements. The graphs based on filtered IACC measurements, $IACC_{E3}$ and $IACC_3$, (Figures 4.18 and 4.19) tend to differentiate between systems better than the graphs of the non-filtered measurements $IACC_{FB}$ and $IACC_E$ (Figures 4.16 and 4.17) with the plots representing the different systems becoming more separated and distinct. The use of frequency filtered IACC measurements to produce SR values is discussed later in this section.

As shown in Table 4.4, the minimum IACC values measured using the reproduction systems are notably higher than those of the original concert hall. This has raised the possibility that low IACC measurements may distort the SR values calculated for each system. To investigate this, the IACC measurements taken in the real concert hall, that were arranged in ascending IACC, were split into two groups of the twelve lowest IACC measurements and the twelve highest IACC measurements. Correlations were made between the original and reproduced IACC measurements for 'Full' (all 24 measurements), 'Low' and 'High' measurements, for each system and each IACC variation, which can be seen in Table 4.6. Mono measurements have not been included due to unvarying IACC measurements close to one. Correlations significant to the $p = 0.05$ level are shown in red.

	Stereo	3/2	4 LS Ambi	8 LS Ambi	Peri Ambi	**Average**
$IACC_{FB}$ Full	0.72	0.71	0.70	0.59	0.64	**0.67**
$IACC_{FB}$ Low	0.23	0.13	0.08	-0.14	-0.28	**0.17**
$IACC_{FB}$ High	0.74	0.77	0.67	0.53	0.48	**0.64**
$IACC_{E}$ Full	0.56	0.52	0.56	0.42	0.54	**0.52**
$IACC_{E}$ Low	-0.13	0.01	-0.23	0.10	0.10	**0.12**
$IACC_{E}$ High	0.75	0.73	0.67	0.41	0.55	**0.62**
$IACC_{E3}$ Full	0.55	0.69	0.69	0.73	0.58	**0.65**
$IACC_{E3}$ Low	0.69	0.70	0.27	0.35	0.28	**0.46**
$IACC_{E3}$ High	0.16	0.41	0.71	0.65	0.46	**0.48**
$IACC_{3}$ Full	0.65	0.81	0.47	0.68	0.48	**0.62**
$IACC_{3}$ Low	0.44	0.60	-0.30	0.19	-0.02	**0.31**
$IACC_{3}$ High	0.40	0.73	0.69	0.74	0.66	**0.65**

Table 4.6 Correlations between original and reproduced concert halls for Full, High and Low IACCs

Almost all of the Low IACC correlations are not-significant and reinforce the observation that there is an upper limit to the degree of spatial impression that reproduction systems can deliver. All of the Full and almost all the High IACC correlations are significant. In comparing the average correlations over all systems, the Full and High correlations were fairly similar, with the Full correlations showing significance for all systems. From these correlations it appears that Low IACC measurements should not be included in the SR calculations.

In ascertaining the optimal manner in which SR values should be calculated, as well as the inclusion or non-inclusion of Low IACC measurements, the type of IACC measurement used ($IACC_{FB}$, $IACC_E$, $IACC_3$ or $IACC_{E3}$) also needs to be considered. Whilst the expected performance of the reproduction systems tested in terms of spatial capabilities could be proposed through listening experience and / or theoretical means, the utilization of SR values as an objective spatial measure of the systems could be better established by determining the optimal manner of calculating SR by a comparison to subjective perception.

By subjectively evaluating the spatial capabilities of the reproduction systems, a correlation between the subjective preferences and the objective SR values could be made. The optimal way in which the SR value is calculated could be found by comparing correlation coefficients. A subjective test and the correlation of the results to the SR values is the subject of the next Chapter.

4.7 Summary

In this chapter, the adaptation of the concert hall measure of spatial impression, IACC, was investigated for use in reproduced sound. This was initially investigated by measuring the IACC of the spatializing techniques for musical synthesis. It was shown that the spatial measure was able to discriminate between samples that had differing degrees of spatial spread. The IACC measurements for one type of synthesized sound correlated significantly to the results of the subjective listening test outlined in Chapter Three.

Following on from this, the possibility of using IACC as an objective measure of spatial impression in other aspects of reproduced sound was realised. A method of comparing IACC measurements made in a original concert hall to measurements made in a version of the same concert hall reproduced by a sound system was developed as an indicator of the spatial capabilities of the reproduction system. The greater the retention of spaciousness, the lower the SR value and the better the spatial performance of the reproduction system.

The procedure was initially conducted as a simulation using simple auralizations of a concert hall and reproduction systems which produced encouraging results. The procedure was then conducted using measurements taken in real a concert hall and real reproduction systems. An objective measurement of the spatial capabilities of six different reproduction systems was recorded by this method which rated the systems in an expected manner. A number of variations in the way that the objective measurement was calculated were introduced. The optimal method of calculation is to be ascertained by means of correlation to subjective preferences.

5 Subjective Evaluation of the Spatial Capabilities of Various Sound Reproduction Systems

5.1 Introduction

In order to reinforce the findings from the objective measurements outlined in the previous Chapter, a subjective listening test was conducted. The purpose of the test was to align the objective measures of spatial impression in reproduced sound to subjective perception. The differences in spatial impression between reproduction systems, as indicated by the SR measurements, were expected to be reflected in the results of a subjective test. Furthermore, an indication of the optimal type of IACC measurement used to calculate the SR values was to be established through correlation of the objective and subjective results.

As the objective measures compared real to reproduced IACC measurements, ideally the subjective test should also compare real to reproduced listening environments. As this would prove to be an impracticable test method, the subjects were presented with stimuli replayed over each of the six previously objectively measured reproduction systems and were asked to score the spatial attributes of each system in comparison to their own experiences of spatial listening in real concert halls.

In brief, the subjective test entailed the evaluation (in terms of spaciousness) of six different reproduction systems, using three different types of program material. The results of the test produced a mean score for each system. The scores were then correlated with the SR values from the objective measurement test.

5.1.1 Stimuli

Three types of programme material were used in the test; a female speech sample [Huopaniemi 2000], a sample of Mozart's overture, 'Le Nozze di Figaro' [Denon 1994] and a snare drum sample [Belschner 2001]. All samples were mono, anechoic recordings with a sample rate of 44.1 kHz and a 16-bit resolution. Each sample was edited to be approximately 10 seconds in duration (the snare drum sample was looped and repeated a number of times) using short fade ins and outs where necessary.

5.1.2 Processing of Stimuli

To enable each sample to be presented over each reproduction system, a set of the B format Peel Hall impulse responses, from a front row, off-centre seat position were convolved with the anechoic samples. A front row seat position was selected to avoid excessively reverberant presentations of the stimuli.

Following the convolution of the anechoic samples with the B-format W, X, Y and Z (if needed) impulse responses, the signals were decoded to the various reproduction systems using the equations outlined in Section 2.3.8.

5.1.3 Subjects

Ten listeners took part in the test, all of whom were staff, students or visiting students of the School of Acoustics and Electronic Engineering, University of Salford. The majority of the subjects had previously taken part in other listening tests and had an interest in audio and acoustics. Whilst not expert listeners, the subjects could be considered as 'selected assessors' [Bech and Zacharov 2006] and therefore be expected to produce reliable judgements. None of the subjects reported any known hearing defects. The tests were held over a ten-day period.

5.1.4 Physical Set-Up

The semi-anechoic chamber and loudspeaker arrangement described in Section 4.5.2 was used for the listening test and can be seen in Figure

4.10. The semi-anechoic chamber was chosen to negate the effects of room acoustics and for its low background noise. To accommodate all six reproduction systems, a total of twenty-one loudspeakers were utilised. All of the loudspeakers were level aligned using pink noise and a sound level meter. To avoid an inappropriate variable being introduced by having differing presentation levels for each system, the relative reproduction levels of the six systems were subjectively aligned by the author and one of the subjects. Level alignment by objective means (sound pressure level meter) was considered inappropriate due the presence of anti-phase signals in some of the ambisonic presentations. The subjective level alignment was achieved by comparing the perceived loudness of a sample replayed on each reproduction system (individually) to that of the same sample replayed in mono. The level of the non-mono reproduction system could be adjusted. When subjective equal loudness was attained, the channel levels of the non-mono system were noted. The relative input levels for each individual loudspeaker of a particular system (referenced to the mono system) are shown in Table 5.1.

System	Relative Input Level (dB)
Stereo	-1.5
Ambisonic 4 LS	-4.6
3/2	-6
Ambisonic 8 LS	-8.8
Ambisonic Periphonic	-9.4

Table 5.1 Relative loudspeaker input levels for the individual loudspeakers of each system. The levels are relative to the input level of the mono system (0 dB).

The samples were stored on a computer equipped with a multichannel soundcard that was connected to digital to analogue converters that, in turn, were connected to the self-powered loudspeakers. The multichannel samples were replayed using an audio sequencer. To allow for groups of samples to be graded (where a group contained six versions of the same sample, decoded to each of the six reproduction systems), a feature of the sequencer enabled the subjects to change between and compare reproduction systems by pressing specific keys on a computer keyboard that was present in the listening room. When a key was pressed, the newly selected audition would play from the beginning. This allowed for the subjects to make instant comparisons between the different systems.

An acoustically transparent curtain was hung between the listening position and the loudspeakers to eliminate any visual cues. The reproduction equipment used in the experiment was also screened from the subjects' view upon their entering and leaving of the semi-anechoic chamber. The solid (tiled) floor of the chamber was covered with acoustically absorbent tiles to reduce floor reflections.

5.1.5 Test Procedure

The subjects were presented with four groups (one for each programme sample (voice, music and drum) and one repeat) of six samples (one for each system). This meant that in evaluating systems, programme samples were not mixed. The group order was randomised with the presentation of the first group being repeated at the end. The presentation of the first group of samples was considered as a familiarisation period for the subjects, therefore the data was not used in the statistical analysis. The assignment of the keys of the computer keyboard to samples (which allowed for switching between samples) was also randomised.

The subjects were asked to grade each of the samples in terms of realism of spatial reproduction. The subjects were asked to consider spatial attributes of concert halls, such as apparent source width and envelopment, and to compare each reproduced sample to their own experiences of spatial listening in real concert halls. The subjects graded the six samples of each group by marking a 10-point linear scale. The extremes of the scale were 0 – 'None of the spatial attributes of concert hall listening were present in the example' and 10 – 'The spatial attributes of the example were identical or near identical to those of a concert hall'. The subjects could take as long as they wished to complete the grading and could audition each sample as many times as they needed. When a subject had finished grading a particular group of samples, the subject let it be known (by means of a microphone) that he or she was ready to grade the next group. The subjects took approximately between 10 and 25 minutes to complete the test. The subject instructions can be seen in Appendix E.

5.1.6 Results of the Subjective Test

A two-way, repeated measures analysis of variance (ANOVA) model was used for the statistical analysis. The first step was to check the data for conformation to the assumptions of the ANOVA model. Mauchly's test of sphericity was employed to check the data [Field 2000]. The data are spherical (and therefore useable in the ANOVA analysis) if the p-value is *non-significant*, which was the case for the main factors, 'System' ($p = 0.200$) and 'Sample' ($p = 0.161$).

A generalised linear model, using a type III sum of squares ANOVA was used to analyse the data. The results are shown in Table 5.2. As sphericity is assumed, the first row in each factor window of the table is employed. The output demonstrates that the factor 'System' was significant ($F = 36.3$, $p = 0.000$) whilst 'Sample' ($F = 0.53$, $p = 0.597$) and the interaction 'System*Sample' ($F = 1.46$, $p = 0.165$) were found not to be significant.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
SYSTEM	Sphericity Assumed	611.617	5	122.323	36.302	.000
	Greenhouse-Geisser	611.617	2.887	211.816	36.302	.000
	Huynh-Feldt	611.617	4.397	139.108	36.302	.000
	Lower-bound	611.617	1.000	611.617	36.302	.000
Error(SYSTEM)	Sphericity Assumed	151.633	45	3.370		
	Greenhouse-Geisser	151.633	25.987	5.835		
	Huynh-Feldt	151.633	39.570	3.832		
	Lower-bound	151.633	9.000	16.848		
SAMPLE	Sphericity Assumed	3.675	2	1.837	.531	.597
	Greenhouse-Geisser	3.675	1.464	2.510	.531	.546
	Huynh-Feldt	3.675	1.677	2.191	.531	.568
	Lower-bound	3.675	1.000	3.675	.531	.485
Error(SAMPLE)	Sphericity Assumed	62.325	18	3.463		
	Greenhouse-Geisser	62.325	13.176	4.730		
	Huynh-Feldt	62.325	15.095	4.129		
	Lower-bound	62.325	9.000	6.925		
SYSTEM * SAMPLE	Sphericity Assumed	27.458	10	2.746	1.466	.165
	Greenhouse-Geisser	27.458	2.458	11.171	1.466	.252
	Huynh-Feldt	27.458	3.451	7.955	1.466	.240
	Lower-bound	27.458	1.000	27.458	1.466	.257
Error(SYSTEM*SAMPLE)	Sphericity Assumed	168.542	90	1.873		
	Greenhouse-Geisser	168.542	22.122	7.619		
	Huynh-Feldt	168.542	31.063	5.426		
	Lower-bound	168.542	9.000	18.727		

Table 5.2 Anova output of the subjective test data

Figures 5.1 and 5.2 display the means and 95% confidence intervals for system and sample respectively. The non-significant difference between samples is apparent from Figure 5.2. Whilst the ANOVA model has shown that there are significant differences between the means of the system scores, upon inspection of Figure 5.1, apart from the mono system, there appears to be only a slight variation between the mean scores of some of the other systems. This was examined by performing a multiple comparison of factors using the Bonferroni procedure [Field 2000]. The output can be seen in Table 5.3. The results show that only System 1 (mono) differs significantly from the other systems.

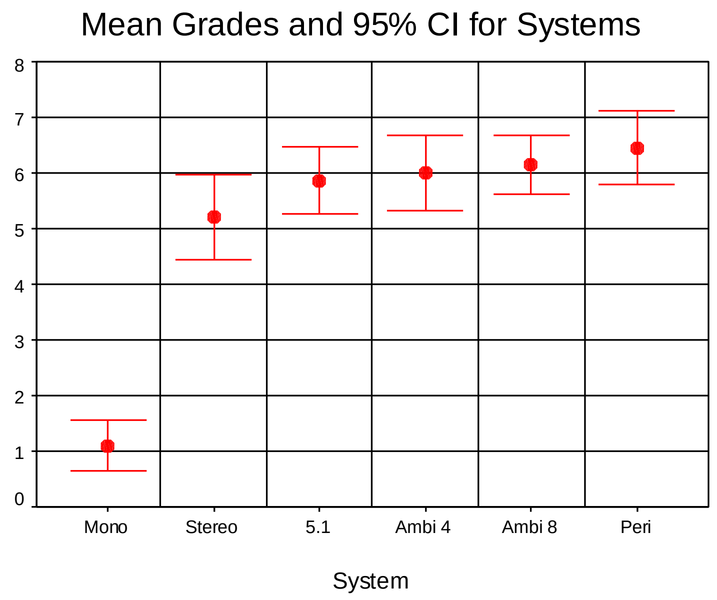


Figure 5.1 Mean subjective grading of spatial realism for reproduction systems

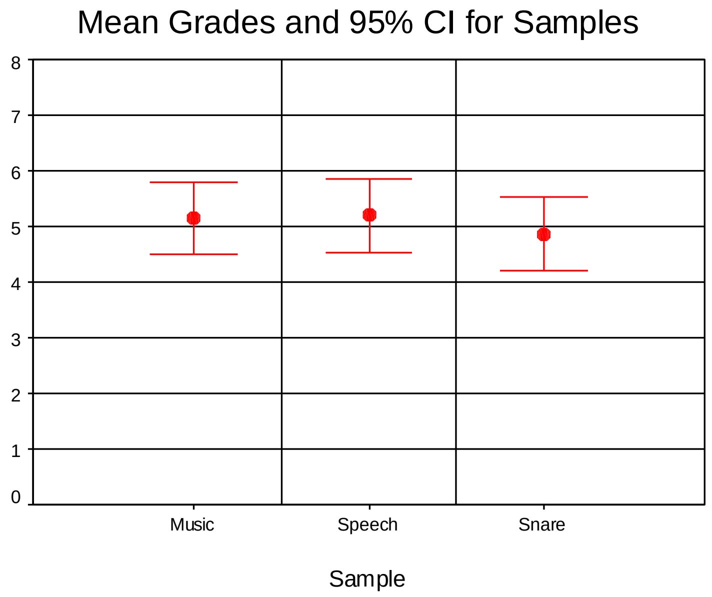


Figure 5.2 Mean subjective grading of spatial realism for sample types

Pairwise Comparisons

Measure: MEASURE_1

(I) SYSTEM	(J) SYSTEM	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a] Lower Bound	Upper Bound
1	2	-3.733*	.521	.001	-5.792	-1.674
	3	-4.767*	.451	.000	-6.549	-2.984
	4	-4.900*	.524	.000	-6.971	-2.829
	5	-5.050*	.429	.000	-6.746	-3.354
	6	-5.350*	.533	.000	-7.456	-3.244
2	1	3.733*	.521	.001	1.674	5.792
	3	-1.033	.594	1.000	-3.383	1.316
	4	-1.167	.437	.386	-2.896	.563
	5	-1.317	.616	.921	-3.754	1.121
	6	-1.617	.639	.483	-4.143	.909
3	1	4.767*	.451	.000	2.984	6.549
	2	1.033	.594	1.000	-1.316	3.383
	4	-.133	.386	1.000	-1.658	1.391
	5	-.283	.234	1.000	-1.210	.644
	6	-.583	.241	.581	-1.537	.370
4	1	4.900*	.524	.000	2.829	6.971
	2	1.167	.437	.386	-.563	2.896
	3	.133	.386	1.000	-1.391	1.658
	5	-.150	.455	1.000	-1.949	1.649
	6	-.450	.425	1.000	-2.132	1.232
5	1	5.050*	.429	.000	3.354	6.746
	2	1.317	.616	.921	-1.121	3.754
	3	.283	.234	1.000	-.644	1.210
	4	.150	.455	1.000	-1.649	1.949
	6	-.300	.414	1.000	-1.937	1.337
6	1	5.350*	.533	.000	3.244	7.456
	2	1.617	.639	.483	-.909	4.143
	3	.583	.241	.581	-.370	1.537
	4	.450	.425	1.000	-1.232	2.132
	5	.300	.414	1.000	-1.337	1.937

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Table 5.3 Table showing significance differences between pairs of systems.

System 1 = Mono, System 2 = Stereo, System 3 = 3/2, System 4 = 4-Loudspeaker Ambisonic, System 5 = 8-Loudspeaker Ambisonic and System 6 = Periphonic Ambisonic. On each row, one system is compared to the other five. The 'Sig' column indicates which pairs of systems differ significantly.

As the Bonferroni procedure indicates that only the mono system differed significantly between any pairs of systems, a second analysis of the data was performed that did not include the ratings for the mono system. This was undertaken to ensure that the significance of the results was not entirely due to the outlying mono data. In this analysis, Mauchly's test revealed that sphericity could not be assumed for the factor 'System', as $p = 0.046$ (however, a borderline case, very close to the 0.05 level). However, for data that violates the sphericity assumption, a Greenhouse-Geisser correction can be applied to produce a valid F-ratio. This resulted in the factor 'System' again being significant ($F=3.496$, $p < 0.048$). However, the Bonferroni procedure again failed to indicate which system(s) significantly differed from each other.

5.1.6.1 Subject's Comments

After completing the test, subjects were also encouraged to voice any comments. These included:

- Perceived localization of sources changed with system.
- Timbral differences were evident between systems.
- Spatial differences between some systems were very subtle.

- None-specified differences, other than spatial or timbral, were present between systems.
- A more spacious reproduction did not necessarily correspond to a more realistic concert hall listening experience

The comment that the subjects perceived the spatial differences between systems as subtle, may contribute to the ‘bunching together’ of the non-mono systems’ mean spatial ratings shown in Figure 5.1 (this may also apply to the final comment).

5.1.6.2 Results Summary

In summary, the subjective test has demonstrated that:

- The subjects were able to significantly identify differences between the reproduction systems in terms of realism of spatial attributes. However, due to the overlapping confidence intervals, firm conclusions regarding differences between the systems cannot be made.
- Differences between reproduction systems were independent of the type of program material presented.
- The systems were ranked in order of spatial realism (from least to most) as mono, stereo, 3/2, four-loudspeaker ambisonic, eight-loudspeaker ambisonic then periphonic ambisonic.

- Apart from differences between mono and all other systems, the differences between systems, in terms of mean subjective scores, were small.

5.2 Comparison Between the Objective Measurements and the Results of the Subjective Listening Test

The results of the subjective test were next used to validate the objective procedures by means of correlation. The SR values, calculated using different variations of the IACC measurement can be compared to indicate which IACC measurement is best suited to spatial measurements of reproduction systems. A graph displaying the mean subjective score and 5 – SR values for the four IACC variations are shown in Figure 5.3. (The SR values were subtracted from five to allow for a direct comparison with the subjective scores.)

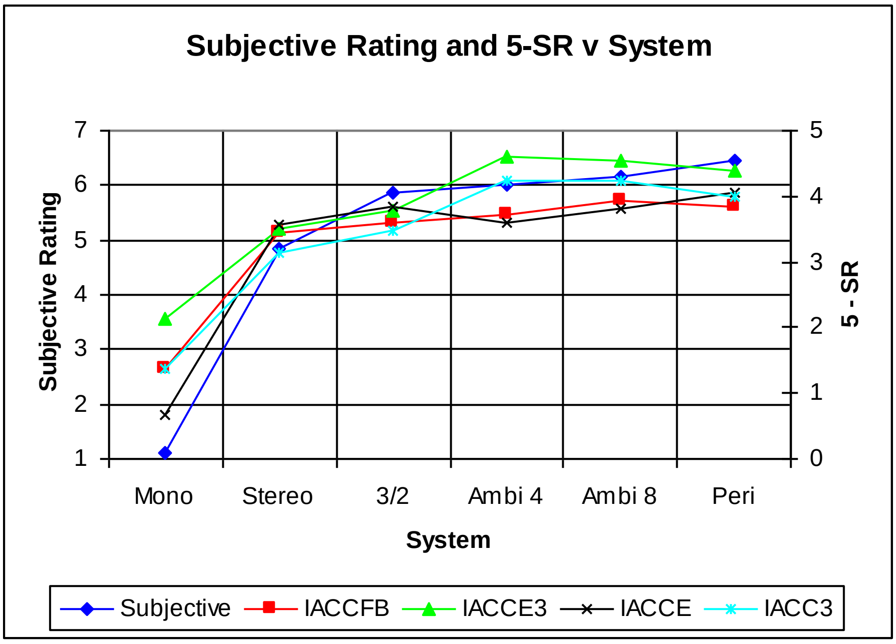


Figure 5.3 Mean subjective scores and 5-SR measurements versus systems

Upon initial inspection, all of the SR measurements correspond to the subjective ratings well. To determine how well the plots for the objective measurements vary with the subjective plot, the subjective results were correlated with the SR values calculated using the four different IACC measurements for the six reproduction systems. The results can be seen in Table 5.4.

SR Based On:	$IACC_{FB}$	$IACC_{E}$	$IACC_{E3}$	$IACC_{3}$
Correlation	0.99	0.98	0.95	0.97

Table 5.4 Correlation between SR measurements and subjective ratings

All of the correlations are significant to the p = 0.01 level, with the SR measurements based upon the $IACC_{FB}$ comparisons, correlating to the subjective results the best. The $IACC_{FB}$ based SR measurements ranked the systems (from least to most spatially retentive) as mono, stereo, 3/2, four-loudspeaker ambisonic, periphonic ambisonic then eight-loudspeaker ambisonic. The subjects ranked the systems in almost the same order, but with the periphonic ambisonic and eight-loudspeaker ambisonic systems exchanging ranks.

In the previous chapter, it was also suggested that the lower IACC measurements used to calculate the SR values may distort the values as the reproduction systems were incapable of producing versions of the concert hall that yielded IACC measurements below a certain value. By disregarding the lowest twelve IACC measurements in calculating the SR values, the subjective and objective results were again correlated. The results can be seen in Table 5.5.

SR Based On:	$IACC_{FB}$	$IACC_{E}$	$IACC_{E3}$	$IACC_{3}$
Correlation	0.99	0.96	0.98	0.96

Table 5.5 Correlation between SR values (calculated using High IACC measurements) and subjective ratings

Again, all the correlations are significant to the $p = 0.01$ level, with the SR measurements based upon the $IACC_{FB}$ comparisons, correlating to the subjective results the best. The 'High' $IACC_{FB}$ based SR measurements spatially ranked the systems in terms of spatial impression (from best to worse) as mono, stereo, four-loudspeaker ambisonic, 3/2, eight-loudspeaker ambisonic then periphonic ambisonic which differs from the subjective results in that the four-loudspeaker ambisonic and 3/2 systems have exchanged ranks.

5.3 Discussion

In this Chapter the objective SR measurements of spatial impression in reproduced sound were compared to the results of subjective test. The subjective test demonstrated that there were significant differences between reproduction systems in terms of spatial impression. These differences, for the non-mono systems, were however quite small. By examining the mean subjective scores for each system, there was a large and expected difference between mono (mean score of 1.1) and all other systems. Stereo had the next highest score (4.8) which was noticeably (but not significantly) lower than the ambisonic-based systems, where the next highest, the 3/2 system, scored 5.9. The highest score was for the periphonic system which was 6.5. This would suggest that increasing the number of loudspeakers used in a surround sound system does not necessarily result in a large increase in perceived spatial impression. This observation was also reflected in the results of the objective measurements reported in the previous chapter.

In comparing the objective and subjective results, in all of the correlations between the subjective and objective measurements, the correlation coefficients have been very high and the differences between the coefficients obtained from different IACC measurements have been small. Whilst the correlations have shown that SR values calculated using $IACC_{FB}$ measurements correlate best to the subjective results, the reliability of correlations calculated using only six pairs of data may be brought into question. It is also notable that whilst high correlation was achieved for all SR values, the rank ordering of the systems by subjective rating was not matched by *any* of the objective measurements, although for the SR value with the highest correlation coefficient, only the two highest ranked systems exchanged ranks.

Assuming that the high degree of correlation between the results of subjective test and the objective measurements can be accepted, it would appear that the SR values based on any of the IACC measurements can be used to accurately measure the spatial performance of reproduction systems. The SR values based on $IACC_{FB}$ measurements produce the (marginally) most accurate results. From the extensive research into the IACC measurement in concert hall acoustics, the author found it surprising that the SR values calculated using perceptually refined IACC measurements such as $IACC_{E3}$ did not produce significantly more accurate results. However, this could be partially explained by the way in which the subjects were instructed. The subjects were asked to consider 'spatial attributes' of concert halls, when it may have better to focus upon one particular spatial attribute such as apparent source width, which is particularly attributed to $IACC_{E3}$.

5.4 Summary

In this Chapter, a subjective experiment designed to validate the objective measurement procedures outlined in Chapter Four was reported upon. The subjective experiment involved the rating of six different sound reproduction systems in terms of their realism of spatial reproduction. The subjects rated the systems, in order of least to most spatially authentic as mono, stereo, 3/2, four-loudspeaker ambisonic, eight-loudspeaker ambisonic then periphonic ambisonic. The mean subjective ratings of the systems were found to be significant and independent of the type of program material presented.

The subjective mean ratings were correlated to the objective ratings as predicted by the SR values. A number of variations of the SR values were included that differed in the type of IACC measurement used in their calculation. The output showed that all the objective ratings correlated highly to the subjective results. The SR value based on the $IACC_{FB}$ measurements had the highest correlation coefficient.

The subjective results and their correlation to the objective measurements suggest that the validity of using the SR values based on $IACC_{FB}$ measurements for evaluating the spatial performance of reproduction systems has been reinforced. The refinement of the objective measurement to be more sensitive to small or subtle differences in spatial perception forms the basis of the following Chapter.

6 Refinement of IACC as a Spatial Measure by Means of Frequency Weighting

6.1 Introduction

In the previous two chapters, objective and subjective measurements of spatial impression in reproduced sound were examined. The objective measurements were based upon the comparison of IACC measurements taken in real and reproduced concert halls. In this chapter, the IACC measurement itself is investigated with a view to better aligning the measurement to the subjective perception of spatial impression. In particular, the frequency dependency of IACC is examined by means of a subjective test.

Previous work and theory proposes that a pair of filtered signals, covering different frequency regions but with the same IACC value may not be perceived as being equally spacious. This concept is investigated using a custom designed and built mixing device that allowed for an adjustable comparison of such signals whilst retaining a constant presentation level. This may demonstrate that IACC does not quantify spatial impression equally and requires a frequency dependent weighting.

6.2 Frequency Dependency of Spatial Impression

Concert hall measurements using IACC have addressed the question of IACC varying in different frequency regions by incorporating an average IACC value across frequency regions. The $IACC_{E3}$ measurement, introduced by Hidaka et al. [Hidaka et al. 1995] involves taking the average of the three IACC values in the 500, 1000 and 2000 Hz octave bands. The rationale behind this is that at lower frequencies, IACC varies little and tends towards a high value, whilst high frequencies are not considered important in the perception of spatial impression. Although this may generally be the case, the perception of spatial impression at different frequency regions and the corresponding IACC measurements may not be aligned.

For low frequencies in particular, where the IACC measurement is comparatively insensitive, the presence of low frequency components has been found to have a large effect upon the perceived degree of spatial impression when compared to components of higher frequencies [Morimoto and Maekawa 1988]. In the creation of spatial impression in concert halls, it has been cited that low frequencies in particular are very important [Barron and Marshall 1981]. However, in measuring IACC as a function of frequency in concert halls [Yanagawa et al. 1990], low frequency components tend to exhibit high IACC values whilst higher frequency components tend towards lower correlation values.

Potter et al. [Potter et al. 1995] pointed out this inconsistency and developed a number of subjective tests that examined the frequency dependency of spaciousness. One of the tests involved comparing octave band filtered noise signals (ranging from a centre frequency of 125 Hz to 4000 Hz) of a fixed IACC with a broadband signal of variable IACC. Subjects were asked to adjust the broadband signal to be of the same perceptual width as the filtered signal. The results showed that the lower frequency octave band noise signals were perceived as being broader than the high frequency band noise signals.

This experiment forms the basis of the present study. A related experiment is conducted, however, different approaches and methods are utilised. In the present experiment, the effects of varying IACC upon presentation level are addressed and an experimental approach that better attends to theories of spatial hearing is incorporated.

6.3 Pilot Study

In a similar manner to the formation of the Fletcher and Munson equal loudness curves [Fletcher and Munson 1933], a pilot study was devised as a rough indication of how ‘equal spatial impression’ curves may be developed. As with equal loudness experiments, an attribute of a test signal is adjusted until it is perceptually the same as a reference signal. In this case, the attribute is spatial impression with the reference signal being of a fixed degree of spatial impression.

Two, octave band-passed noise signals were compared, one being a reference signal centred on 1 kHz and the other a test signal centred on a number of different centre frequencies. Three filtered reference signals were created using a mixture of positively and negatively correlated independent noise signals to produce signals with inter channel correlation coefficient (ICCC) values of 0.8, 0.5 and 0.2 (where the 'channels' are the left and right signals feeding the headphones). The test signals were created in a similar manner with the mixing of the positively and negatively correlated signals being adjustable to vary the ICCC.

The creation and mixing of the test and reference signals was facilitated using a computer and a digital mixing desk and monitored on headphones. A looped cycle, consisting of the test and reference signals was implemented on the computer and outputted, resulting in five input signals appearing at the mixing desk (two for the reference signal and three for the test signal). The combination of the three test signals was varied by manually adjusting the levels of the input signals using the mixing desk faders until the perceived source width of the test signal was equal to that of the reference signal. The test signal was then recorded on to the computer to allow for ICCC measurements to be taken at a later stage.

This procedure was carried out by the author three times for each octave band comparison, then averaged and plotted. The results can be seen in Figure 6.1. As an initial observation, it was encouraging that when comparing samples centred on the 1 kHz octave band (comparing like with like), the measured ICCC values were all close to the reference values. In general, at low frequencies (< 1 kHz), the measured ICCCs of the test signals were all higher than the reference signals with the converse being observed (with some discrepancies) at high frequencies (> 1 kHz). This suggests that the perception of spatial impression is not quantified equally across frequency by the ICCC.

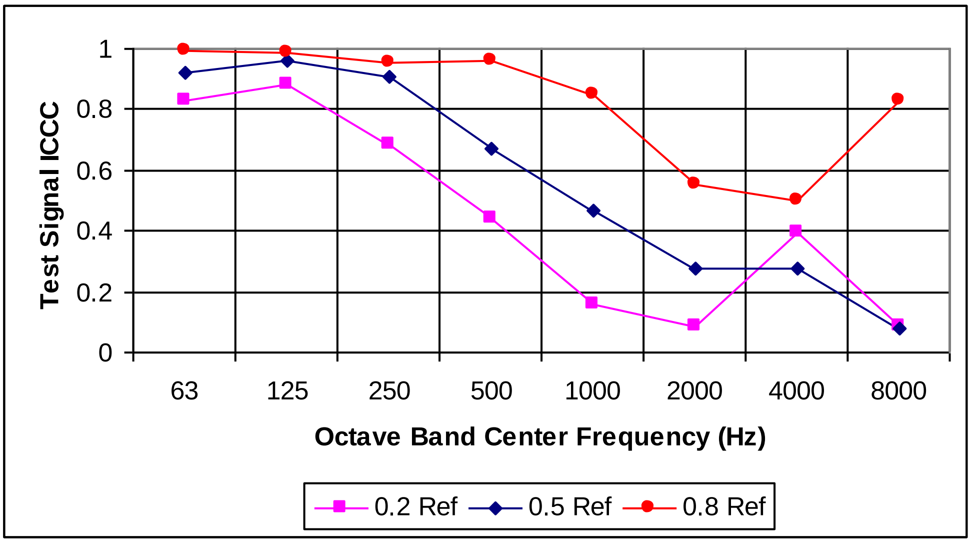


Figure 6.1 Test signal ICCC versus octave band centre frequency for reference signal ICCCs of 0.2, 0.5 and 0.8 (Pilot study).

Due to the encouraging results of the pilot study, a controlled subjective experiment was devised to investigate the apparent frequency dependency of spatial impression, as determined by the ICCC.

6.4 Experimental Method

In this section, the method of comparison used in the pilot study is further developed and refined for use in a controlled subjective experiment. The experimental method is similar to that of the pilot study, however, the effects of extraneous variables, such as differences in perceived loudness between samples, have been minimised.

6.4.1 Varying Spatial Impression

In the pilot study, the spatial impression of the signals was crudely varied by adjusting the combination of a positively correlated noise source and an independent, negatively correlated noise source using the faders of a mixing desk. For the controlled experiment, a more precise method of mixing the signals was realised using a custom built mixer, however, the underlying method of creating signals of varying spatial impression was the same.

In a method similar to Yanagawa and Tohyama's [Yanagawa and Tohyama 1998], signals that could be varied in spatial impression were created in the following manner. Two independent, stereo pink noise sources, s_1 and s_2 were generated. The s_1 was in effect a mono signal as the left and right signals were the same. The s_2 was a signal where the right channel was a phase inversion of the left. The cross correlation of these signals would yield 1 in the case of s_1 and –1 in the case of s_2. In this experiment, the test signals were limited to a cross correlation range of 0 to 1, therefore the signals were combined in such a way that for a resulting cross correlation of 1, only s_1 was present and for a cross correlation of 0, s_1 and s_2 were present in equal amounts. The combination of s_1 and s_2 to produce signals of varying cross-correlation is shown in Equation 6.1.

$$\begin{aligned} \text{Left} &= Cs_1 l + (1-C)s_2 l \\ \text{Right} &= Cs_1 r + (1-C)s_2 r \end{aligned}$$

Equation 6.1

Left and Right are the ear signals, l and r are the left and right components of s_1 and s_2 and C is the combination level of s_1 and s_2.

6.4.2 Combining Correlated and Decorrelated Signals

By setting the combination level, C, to vary between 0.5 and 1, the resultant cross correlation measurements would range between 0 and 1 respectively. However, a problem arises in combining correlated and decorrelated signals as the level of the combined output signal is dependent upon C. This is similar to combining a sine and a cosine wave of the same amplitude, frequency and phase. When only one of the waves is present (in this case, when C = 1), the peak-to-peak amplitude of the combined signal is up by 3dB as compared to when the waves are combined in equal amounts (when C = 0.5). A plot of peak-to-peak amplitude versus Combination level, C can be seen in Figure 6.2.

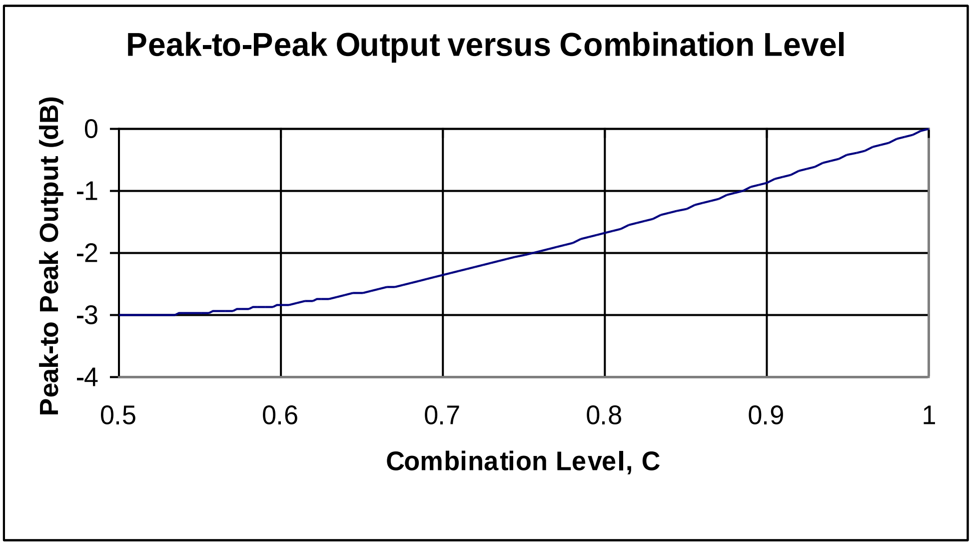


Figure 6.2 Peak to peak amplitude of combined correlated and decorrelated signals versus combination level.

To the ear, a rise in level with rising correlation introduces an extraneous variable when comparing signals of different correlations as the level of the signal has been shown to affect spatial impression judgements [Bradley et al. 1993]. The C-dependent rise in level needed to be counter-balanced and is the matter addressed in the following section.

6.4.3 Design of a Constant-Level Mixer for Combining Correlated and Decorrelated Signals

In order to combine signals s_1 and s_2 in the method shown in Equation 6.1 and to maintain a near-constant output level of the combined signals, a custom mixer was designed and built. The mixer had to be able to be able to produce signals with an ICCC that was continuously variable between 0 and 1 but without any undue variations in output level and be easy to operate (i.e. the ICCC could be varied by turning a single knob).

The design of the mixer was based around a voltage-controlled amplifier (VCA) integrated circuit that was configured as a voltage-controlled panner (VCP) circuit as described in the integrated circuit manufacturers' literature [Analog Devices 2003]. This part of the mixer allowed for Equation 6.1 to be realised and is depicted in Figure 6.3.

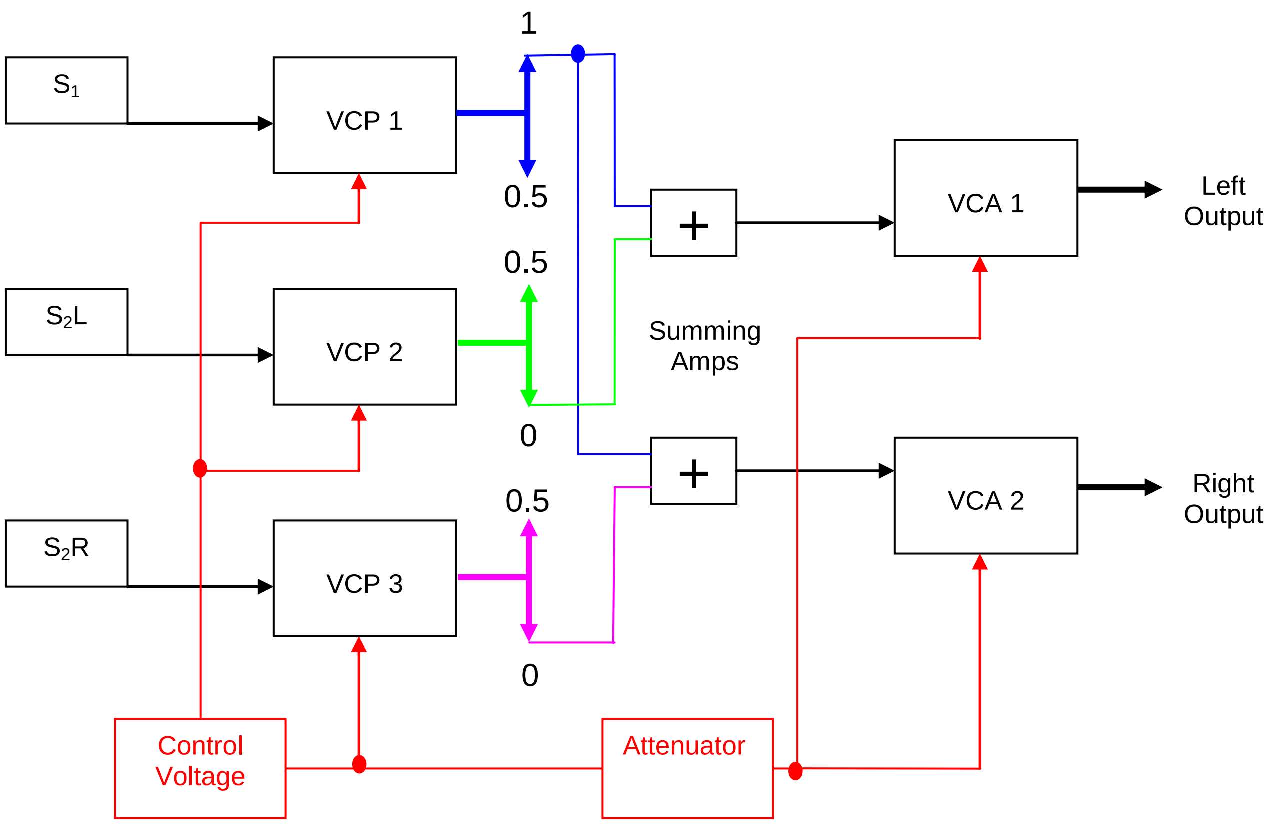


Figure 6.3 Simplified circuit diagram of the constant level correlated and decorrelated signal mixer

The potentiometer adjustable control voltage (which is analogous to C in Equation 6.1) and VCP circuits were set so that when the control voltage was at a maximum, the output from VCP 1 (controlling the amount of s_1 signal) was at a maximum and the outputs of VCPs 2 and 3 (controlling the amount of s_2 signal) were at zero. This is the setting of mixer shown in Figure 6.3. As the control voltage is lowered, VCP 1s' output is lowered whilst the output of VCPs 2 and 3 is increased until a minimum control voltage is reached, when all the VCPs have an equal output. The mixing of the VCP outputs was facilitated using an op-amp based summing amplifier and an op-amp based inverter.

To negate the level variations of combining s_1 and s_2, another VCA section was inserted after the summing amplifiers. The output of the VCAs were dependent upon a control voltage that had been passed through a potential divider and op-amp based attenuator so that the 3 dB increase observed in Figure 6.1 could be counter-balanced. Whilst the VCAs outputs are linear with respect to the control voltage, this is not the case for the combination of s_1 and s_2 with respect to C. By careful adjustment of the control voltage supplying the VCAs (achieved by altering the resistor values used in the potential divider) the counter-balancing of the effects of combining s_1 and s_2 were maximised. Using 1 kHz sine and cosine waves, a plot of the measured peak-to-peak voltage of the main output, the VCA attenuation and the summing amplifier output (all left channel only), as a function of control voltage (measured in 0.5 volt steps) can be seen in Figure 6.4.

The resulting maximum difference in main output level is 0.74 dB, which can be considered negligible. In terms of creating signals of varying degrees of ICCC, setting the mixer to the extremes and using broadband white noise as s_1 and s_2, resulted in ICCC measurements of 0.068 and 0.997 respectively.

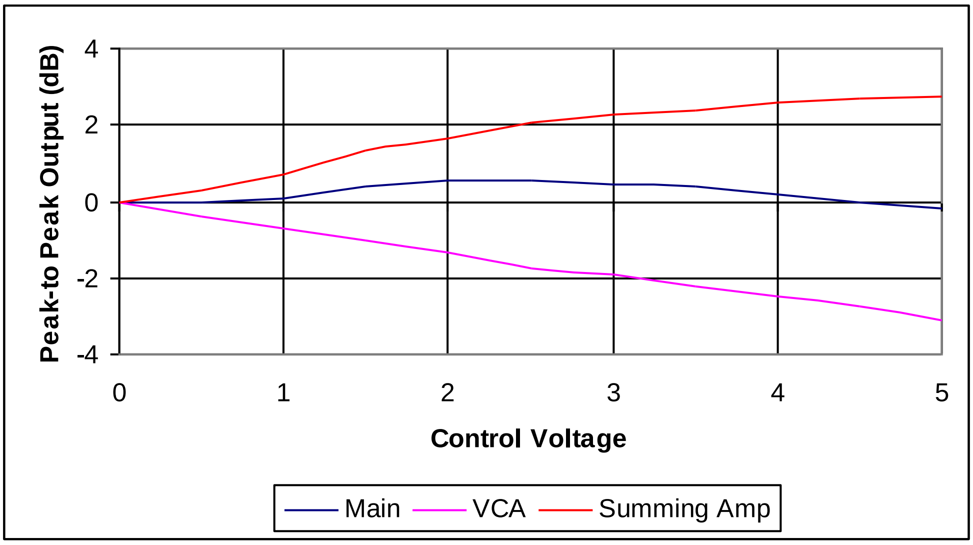


Figure 6.4 Main output, summing amp output and VCA attenuation versus control voltage of the correlated and decorrelated signal mixer

6.4.4 Test Design and Procedure

The purpose of the experiment was to compare the spatial impression of headphone presented signals covering different frequency regions by adjusting the degree of spatial impression of a test signal to be the same as that of a reference signal. This test procedure is known as the method of adjustment (MOA) [Cardozo 1965]. The advantage of using MOA is that the subjects' concentration levels are enhanced due to their active involvement in the adjustment process.

Seven band-passed test signals centred on different frequencies were compared to four band-passed reference signals all centred on 1 kHz but with different ICCCs of 0.2, 0.4, 0.6 and 0.8. This resulted in a total of 28 comparisons. For each comparison, the test and reference pair were looped and repeated until the subject had adjusted the spatial impression of the test signal to be the same as the reference signal by using the potentiometer of the mixer. The subjects could take as long as they wished to complete the task.

An oscilloscope that was placed in view of the subjects was connected to the test signal output. This helped the subjects to discriminate between the test and reference signal as the oscilloscopes' display only became active when the test signal was sounded. The subjects therefore knew which signal of the two was the adjustable one.

After each comparison had been completed, a recording of the test signal was taken to allow for the ICCC to be measured at a later time. After the recording had been made, the subjects were asked to move the potentiometer of the mixer to a random position, thus randomising the starting point of comparison for the next pair of signals. The order of the pairs of comparisons presented to each subject was also randomised.

6.4.5 Experiment Signals

In order to compare spatial impression in different frequency regions, two independent pink noise signals were band passed filtered to form the s_1 and s_2 signals. The centre frequencies of the band passed signals were the octave band centre frequencies ranging from 0.125 to 8 kHz. The bandwidths of the filters were calculated using Glasberg and Moore's equation for equal rectangular bandwidth [Glasberg and Moore 1990]. 50 ms fade ins and outs were applied to all the signals.

Four reference signals, of a fixed ICCC value and with a centre frequency of 1 kHz were generated. The ICCC, the IACC (measured using headphones placed on a dummy head) and the intended ICCC for the reference signals are shown in Figure 6.1. The maximum deviation from the intended ICCC is 0.0088. The similarity in the values of the measured IACC and ICCC signals (maximum difference of 0.0073) suggest that for this experiment, they can be considered equivalent.

Intended IACC	**Measured IACC (Dummy Head)**	**Measured ICCC (Mixer Outputs)**
0.2	0.2088	0.2015
0.4	0.4066	0.4031
0.6	0.6012	0.5995
0.8	0.8005	0.7997

Table 6.1 Intended and measured cross correlations of the reference signals as measured at the mixer outputs and at the ears of the dummy head

.

During the pilot test and the run up towards the experiment, the author and two of the subjects determined the preferred duration presentation time for both the test and reference signals. The signals were continuously looped using an audio software editor. The preferred presentation times were 1.5 seconds for the reference signal and 3.5 seconds for the adjustable test signal. The 1.5 second duration of the reference signal appeared to be long enough for the auditory system to 'store' the perceived degree of spatial impression to allow for comparison. The 3.5 second test signal was deemed adequate in duration to allow the subjects to adjust the spatial impression of the signal to match the reference signal by turning the knob of the mixing device.

In order to retain equal perceived loudness between signals, the frequency response of the whole reproduction system (described in Section 6.4.7) was taken into account. Additionally, the subjective equal loudness of the signals across frequency had to be addressed.

The frequency response of the reproduction system was compensated for by measuring the un-weighted equivalent continuous sound level (L_{eq}) of the system output in the left ear, over a 30 s period using a dummy head and a sound level meter. The output of the system was adjusted and set so that the L_{eq} of a correlated noise signal centred on 1 kHz was measured at 70 dB. The outputs of the test and reference signals were then sounded, measured and adjusted to also give a 70 dB reading. The output levels of the signals were further adjusted to allow for subjective equal loudness over frequency. A B-weighted frequency adjustment [IEC 61672-1 2002] was applied to the signals. A B-weighting was chosen as it is based upon the 70 phon equal loudness contours, which corresponds to the reference presentation level.

6.4.6 Subjects

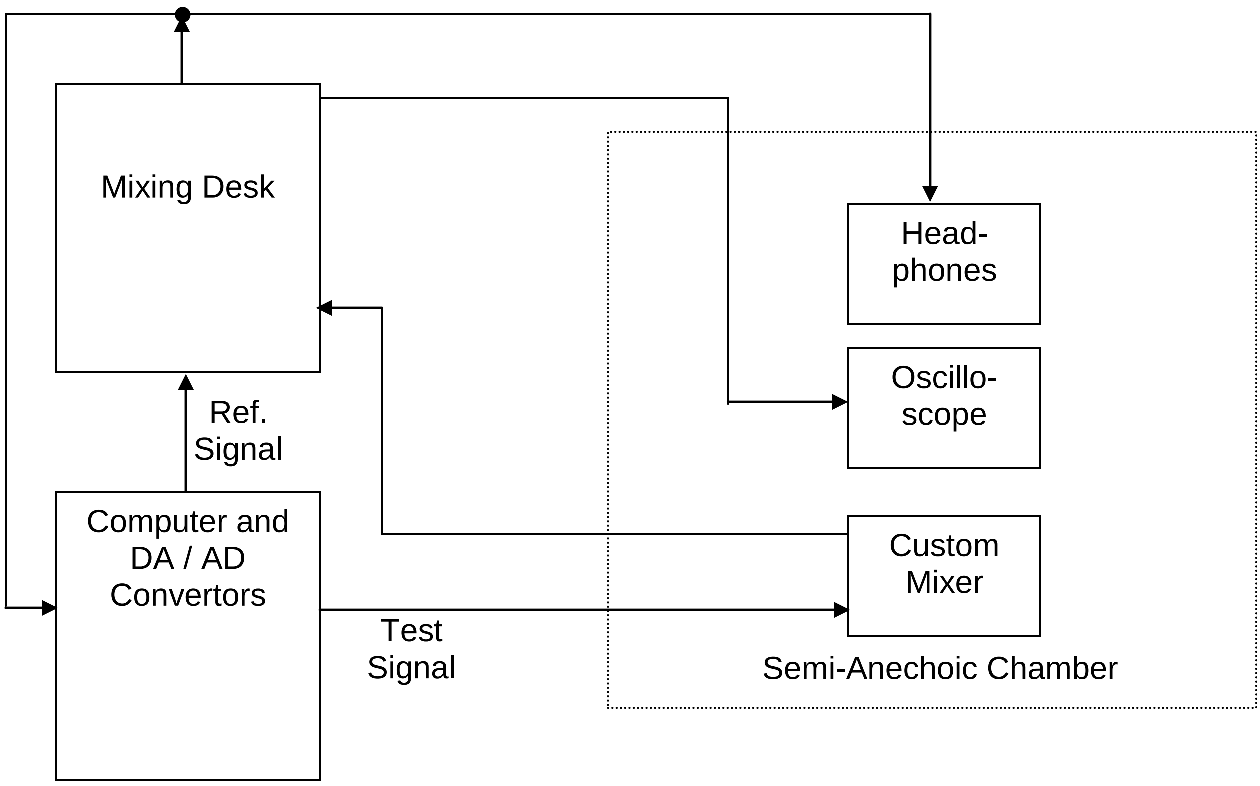
Thirteen subjects, all of whom reported no known hearing defects, took part in the test. All of the subjects were either staff or students of the School of Acoustics and Electronic Engineering, University of Salford. Before the test began, the subjects were introduced to and familiarised with the test signals, mixer and test procedure. To aid the subjects in detecting changes in spatial impression whilst altering the position of the mixers' potentiometer, Blauert and Lindemann's diagram of variations in average subjective apparent source width contours for four different IACC values [Blauert and Lindemann 1986] was shown to the subjects. Having read the subject instructions, the subjects began the test proper. The subject instructions can be seen in Appendix F.

The subjects attended three separate sessions, the first of which involved the introduction followed by the comparison of eight pairs of signals. In the other two sessions, ten comparisons were made in each. Spreading the tests over three sessions was deemed necessary as the subjects may have found a large number of comparisons during one session, too demanding. The testing covered a two-week period in total, with the subjects taking between 10 and 30 minutes to complete each session.

6.4.7 Equipment Configuration

The experiment took place in the semi-anechoic chamber of the School of Acoustics and Electronic Engineering, University of Salford. This room was chosen purely because of its sound isolation properties. The subjects were seated at a table with the mixer and oscilloscope facing them. A microphone was also present to allow for communication with the experimenter who was located outside of the semi-anechoic chamber.

The reference and test signals were arranged and looped using an audio software editor and outputted from the PC via a digital to analogue converter. The reference signal was routed to a mixing desk and the s_1 and s_2 test signals to the custom built mixing device, the output of which was connected to the mixing desk. The relative levels of the reference signal and the s1 and s2 signals were not affected. The main output of the mixing desk was routed both to the subjects' headphones and back into the PC, via an analogue to digital converter, to allow for the recording of the subjects' combination of the s1 and s2 test signals. An auxiliary send was tapped from the custom built mixing device channel and inputted to the oscilloscope. A diagram of the equipment configuration can be seen in Figure 6.5.



6.5 Results

The results from this subjective experiment are the measured values of the ICCC of signals adjusted by the subjects to be of the same spatial impression as a reference signal.

An overall view of the results is shown in Figure 6.6, where the average test signal ICCCs are plotted against the centre frequency of the reference signal for each reference ICCC.

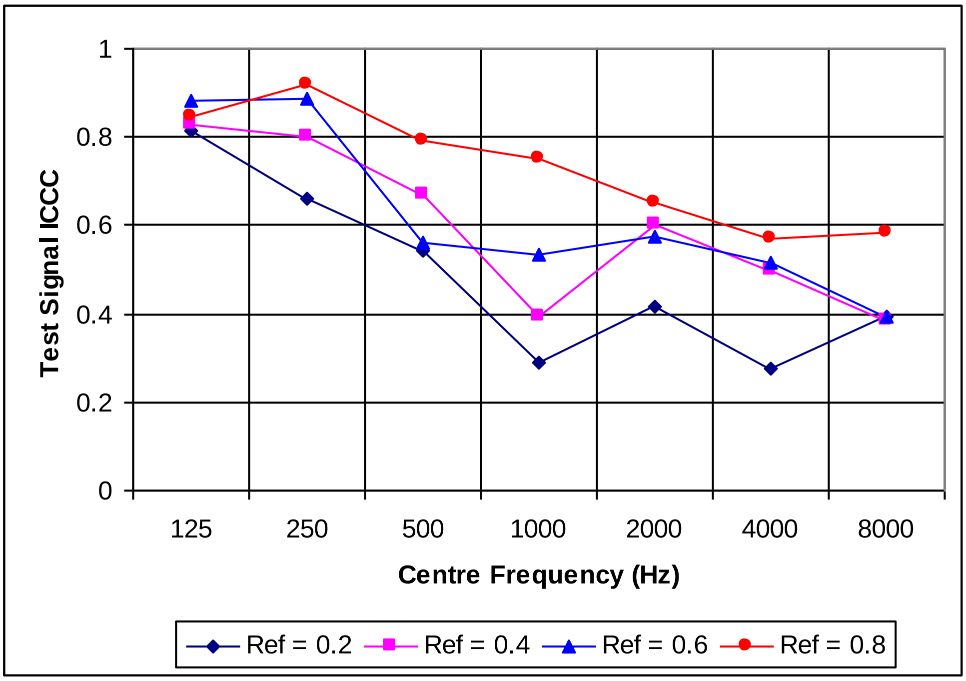


Figure 6.6 Mean test signal ICCC versus centre frequency for reference ICCCs of 0.2, 0.4, 0.6 and 0.8

A detailed analysis of the results was carried out by entering the data into a general linear model, repeated measures ANOVA. Separate analysis was carried out on each of the four reference ICCCs. Macauly's test for sphericity was employed and for three sets of data, since non-significance was found ($p > 0.05$), sphericity can be assumed. The data from the reference ICCC = 0.8 was found to have violated the assumption of sphericity ($p = 0.016$). However, for data that violates the sphericity assumption, a Greenhouse-Geisser or Huynh and Feldt correction can be applied to produce a valid F-ratio. The ANOVA output is shown in Table 6.2.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
C Freq (0.2)	3.066	6	0.511	12.407	0.000
C Freq (0.4)	2.551	6	0.425	10.938	0.000
C Freq (0.6)	2.799	6	0.467	12.395	0.000
C Freq (0.8) GG	1.385	3.26	0.425	5.427	0.003
C Freq (0.8) HF	1.385	4.62	0.300	5.427	0.001

Table 6.2 Anova results table with centre frequency of the test signals as the dependent variable for the four different levels of reference signal ICCC

.

The 'Source' column of the table displays the ICCC of the reference signal in brackets. For the ICCC = 0.8 results, both the Greenhouse-Geisser (GG) or Huynh and Feldt (HF) corrections are displayed. The results show that significant differences are present between test signals of different centre frequencies for all four sets of ICCC reference signals.

Figures 6.7 to 6.10 display the means and 95% confidence intervals for ICCC versus centre frequency for reference ICCCs 0.2 to 0.8, respectively.

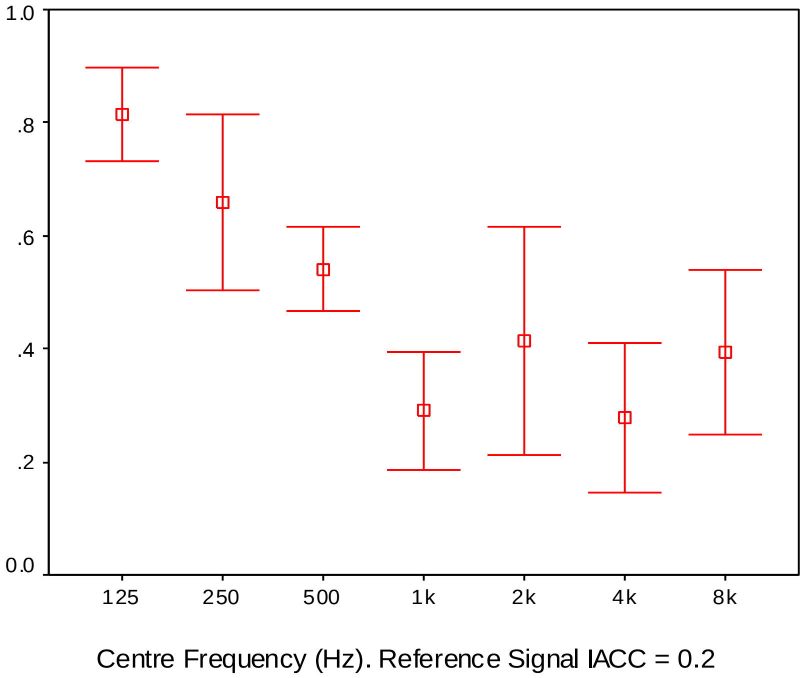


Figure 6.7 Mean test signal ICCC and 95% confidence intervals versus centre frequency for the ICCC = 0.2 reference signal.

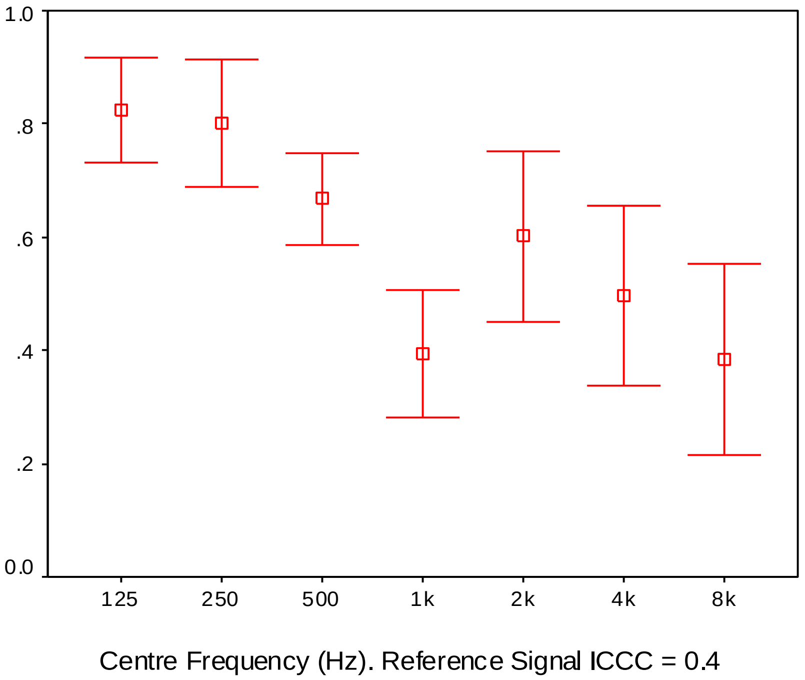


Figure 6.8 Mean test signal ICCC and 95% confidence intervals versus centre frequency for the ICCC = 0.4 reference signal.

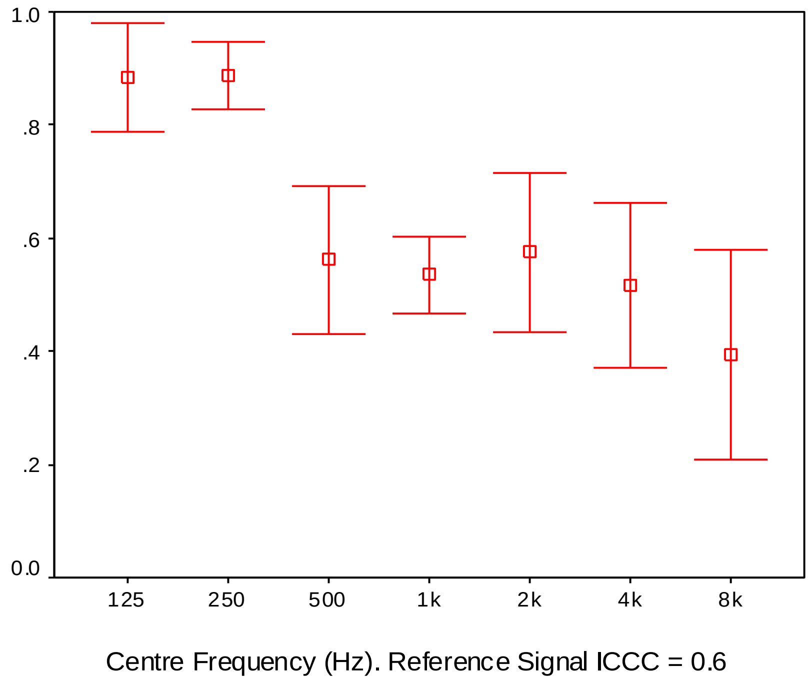


Figure 6.9 Mean test signal ICCC and 95% confidence intervals versus centre frequency for the ICCC = 0.6 reference signal

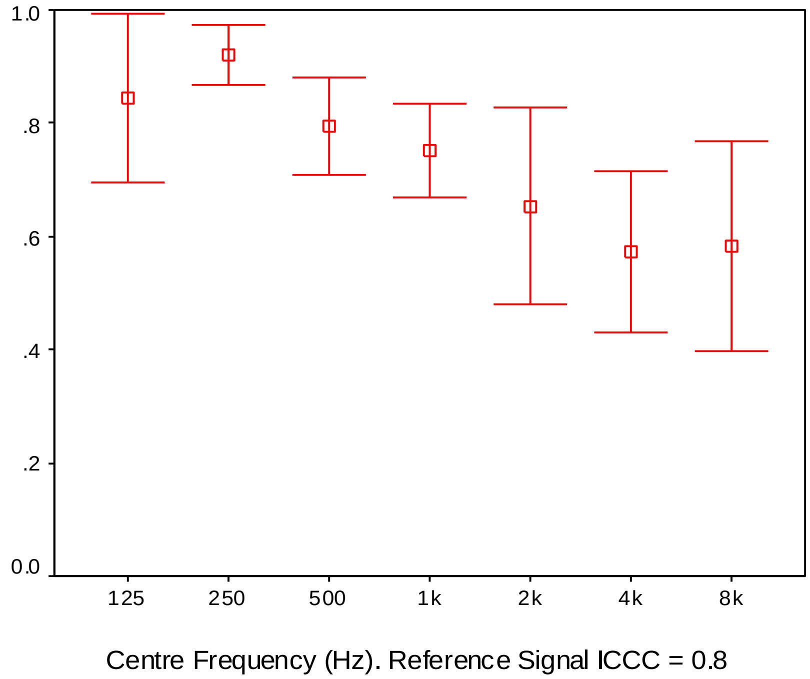


Figure 6.10 Mean Test Signal ICCC and 95% Confidence Intervals Versus Centre Frequency for the ICCC = 0.8 Reference Signal

6.6 Discussion

As an indicator of the reliability of the subjects' judgements, it is interesting to compare how accurately they could match the test and reference signals in terms of spatial impression when the signals were both of the same ERB centre frequency (i.e. 1 kHz). The ICCCs of the reference signals and the averaged test signals are shown in Table 6.3.

Ref ICCC	0.2	0.4	0.6	0.8
Test ICCC	0.29	0.39	0.54	0.75
Difference	+0.09	-0.01	-0.06	-0.05

Table 6.3 Reference and test signal ICCCs at the 1 kHz ERB

For the higher ICCC values the average subjects' ICCC was slightly less than the reference signal. For the lowest ICCC reference signal, the subjects' test signal ICCC was higher and had the greatest difference (0.09). Cox et al. [Cox et al. 1993], determined the difference limen for IACC to be 0.075 ± 0.008. As all but one of the differences between reference and test ICCCs are within the difference limen, the subjects' responses, in general, can be considered reliable.

For the test signals with ERBs centred on frequencies below 1 kHz and for all reference signal ICCCs, a higher test signal ICCC than the reference signal was recorded. In other words, for a constant ICCC value, the perceived degree of spatial impression will be greater at lower frequencies, which is comparable with previous results. In Figure 6.6, at low frequencies, the plots for all reference ICCCs follow a similar trend, apart from the 0.6 ICCC reference signal plot. At the 500 Hz ERB, the average test signal ICCC drops to the approximate same ICCC as the 0.2 ICCC reference signal. It is not obvious why this has occurred, however upon examination of Figure 6.9, the 95% confidence limits for the 500 Hz ERB appear greater than the other low frequency means.

At frequencies above 1 kHz, the results are not as consistent. At higher reference ICCCs (0.8 and 0.6), the test signal ICCCs are generally lower than the reference ICCCs, suggesting that for an equivalent degree of spatial impression, a lower ICCC value will be recorded at higher frequencies when compared to mid frequencies. For the lower reference ICCCs (0.4 and 0.2), the test signal ICCCs were higher than or equal to the reference signal ICCCs. These inconsistencies at higher frequencies may be due to uncertainty in the subjects' responses. Figures 6.7 to 6.10 display the means and 95% confidence limits of the subject's responses for each reference signal ICCC. For the 2, 4 and 8 kHz ERBs and for all reference signal ICCCs, the confidence limits are relatively large, which indicates a wide variation in the subject's responses. This is further implied by conducting a Boniferroni post hoc procedure upon the data. This procedure indicates which test signal ICCCs differed significantly within each set of reference ICCC data by means of pairwise comparisons. In brief, the procedure showed that for all reference ICCCs, all pairwise comparisons between any combination of ERBs centred on 1, 2, 4 and 8 kHz were not significant. Another explanation for the inconsistencies at high frequencies may be offered by the breaking down of phase locking at frequencies above ~ 2 kHz [Jeffress 1948], thus compromising the interaural cross-correlation process.

In general, the results suggest that in terms of the frequency dependency of ICCC, low frequency (<1 kHz) ICCC measurements are higher than mid frequency measurements, where both have the same degree of perceptual spatial impression. For high frequency (>1 kHz) ICCC measurements, the results of the subjective test appear ambiguous.

6.7 Application of the Test Results

Whilst the results concerning the high frequency comparisons may not be particularly reliable, assuming that the low frequency results are reliable, the application of the results is next discussed. Psychoacoustically, the foundation of the IACC is based upon the comparison of the ear signals. The implementation of the IACC calculation to model the coincidence of neural firings from each ear can be used to determine the location of a sound source [Jeffres 1948] or perceived spatial impression [Barron 1971]. Localisation models, such as the one proposed by Macpherson [Macpherson 1991] employs a number of IACC measurements (to extract the inter-aural time difference) over a range of different frequencies, where the bandwidth of the frequency ranges approximate the critical bands of the basilar membrane. The final IACC measurement is calculated by weighting and averaging the individual critical band IACC measurements.

The results of the current experiment could be processed in a similar manner to produce a more perceptually aligned IACC measurement. It could be argued that the auditory system may resolve the degree of spatial impression by analysing the correlation of the ear signals in different critical bands, which in this experiment were approximated using ERBs. The total perception of spatial impression could be determined by weighting and averaging the numerical outputs of the ERB correlations. In this experiment, signals in different ERBs, but with the perceived loudness were compared. A weighting could be applied that counteracts the inequality of the ICCC over different frequency regions as demonstrated by the test results. However, it has been shown that perceived source width is dependent upon the presentation level [Morimoto and Iida 1995]. In general, the greater the presentation level of a signal of fixed IACC, the greater the perceived source width. Whilst this is not considered in the following weighting discussion, an additional level-dependent weighting may be required for completeness.

A possible method of weighting using the low frequency results from the subjective test could be calculated in the following manner:

- The signals under test are ERB filtered, at centre frequencies of 125, 250, 500 and 1000 Hz. IACC measurements are taken in each ERB. These measurements are hence referred to as I_M.
- For each I_M, the I_M of the 1 kHz ERB is noted as I_{M1k}. I_{M1k} is then compared to the ICCC measurements at 1 kHz as given by the results of the subjective test (these are the four 'Test ICCC' values shown in Table 6.3). The ICCC measurement that is closest to the I_{M1k} is denoted I_C.
- I_C will have corresponding ICCC values in lower frequency ERBs that have equal degrees of perceived spatial impression, as determined by the subjective test. These ICCC values are denoted I_{EQ}.

The frequency-weighted IACC in each ERB could be calculated using Equation 6.2.

$$\text{IACC in each ERB} = I_C + (I_M - I_{EQ})$$

Equation 6.2 Calculation of frequency-weighted IACC in each ERB.

An overall frequency-weighted IACC could then be calculated from the average of the IACCs in the four low-frequency ERBs. An example calculation is shown in Appendix G.

The frequency-weighted IACC may be of use in both concert hall and reproduced sound measurements, including the SR measurement of spatial impression in reproduced sound procedure outlined in Chapter 4.

6.8 Summary

In this Chapter, the frequency dependency of IACC in measuring spatial impression was investigated. A pilot test indicated that when compared to mid-frequency signals of a fixed ICCC, lower frequency signals, that were adjusted to be of the same perceived degree of spatial impression, had a higher ICCC measurement. Consequently, a controlled subjective was designed that examined the frequency dependency further.

To facilitate the subjective test, a custom built mixer was designed and tested that allowed for the combination of correlated and decorrelated signals that were needed to produce signals of varying ICCC. The mixer outputted a signal of user-adjustable ICCC that was of a constant level. This was achieved by incorporating a compensatory gain stage into the mixer that counterbalanced the effects of combining correlated and decorrelated signals.

For the subjective test, ERB–filtered noise signals, adjusted to be of an equal loudness, were prepared. Four signals centred on the 1 kHz ERB with fixed ICCCs of 0.2, 0.4, 0.6 and 0.8 were the reference signals. Test signals centred on the 125, 250, 500, 1000, 2000, 4000 and 8000 Hz ERBs were presented to the subjects and compared to the reference signals. The test signals could be continuously adjusted between ICCCs of 0 and 1. The subjects were asked to adjust the test signals to be of the same perceived degree of spatial impression as the reference signals.

The results, that were shown to be significant, showed that for test signals with ERBs centred on frequencies below 1 kHz and for all reference signal ICCCs, a higher test signal ICCC than the reference signal was recorded. This shows that the perceived degree of spatial impression indicated by ICCCs at low frequency is greater than the same ICCC at mid frequency. At frequencies higher than the reference signal, inconsistencies in the results make it difficult to draw conclusions.

Due to the inequality of ICCC across frequency, a weighting procedure has been suggested that may better align the ICCC measurement to subjective perception in both concert hall and reproduced sound situations.

7 Conclusions and Further Work

7.1 Introduction

The experimental work presented in the thesis has involved multichannel spatializing techniques for musical synthesis, the adaptation of concert hall measurements of spatial impression to reproduced sound, the subjective assessment of a number of reproduction systems and the refinement of IACC as spatial measure by means of frequency weighting. In this Chapter, the main findings, implications, conclusions and areas for further work are presented.

7.2 Multichannel Spatializing Techniques for Musical Synthesis

The multichannel spatializing techniques for musical synthesis, which involved decomposing a complex musical signal into its individual harmonics, then spatially spreading the harmonics over a circular loudspeaker array were subject to a psychoacoustic preference test by means of rank ordering. To summarise, the results suggest the following:

- The Friedman test has shown that the subjective results are statistically significant and meaningful.
- A harmonic spatial spread of 90˚ and over resulted in a significantly higher degree of perceived spatial impression than a spatial spread of 0˚ for all presentations.
- In all but one presentation, increasing the spatial spread beyond 90˚ did not significantly increase the perceived degree of spatial impression.
- The techniques appear to be robust as the results were found to be similar for both real and ambisonic presentations.
- For ambisonic reproduction, increasing the spatial spread beyond 90˚ may further increase the perceived degree of spatial impression.

7.3 Adaptation of Concert Hall Measures of Spatial Impression to Reproduced Sound

A method of comparing IACC measurements taken in an original environment to those taken in a reproduced version of the same environment was developed. This was undertaken as both a simulation and by using real environments and reproduction systems. As an indication of the spatial capabilities of the reproduction systems, the degree of spatial retention (SR), as shown by the comparison of IACC measurements was calculated. Correlation between the two sets of IACC measurements was also considered. A number of different reproduction systems were tested. The main findings were:

- A basic, first-order room simulation and reproduction system simulation program produced realistic results. Expected variations in IACC measurements were recorded in the simulated concert hall.
- In the simulation, SR measurements were capable of discriminating between different systems in a predictable manner.
- Correlation coefficients between the sets of IACC measurements were not considered useful indicators of spatial impression.
- SR values showed that systems that utilised more than two loudspeakers fared better than stereo and mono systems. Eight-loudspeaker pantophonic ambisonic reproduction fared the best.
- Using a real concert hall and reproduction systems, similar results were recorded. SR values showed that the worst spatial performance was delivered by the mono system followed by stereo then 3/2. SR values for the three ambisonic systems were similar to each other.
- SR values were capable of differentiating between systems, however, the differences were sometimes small.

- The optimal type of IACC variant used for calculating SR is uncertain.
- The use of low IACC measurements in calculating SR may not be appropriate, however, their exclusion may hide reproduction system errors.

7.4 Subjective Evaluation of the Spatial Capabilities of Various Sound Reproduction Systems

In order to corroborate the objective SR measurement, a subjective test was conducted where the spatial capabilities of the same surround systems used in the objective test were evaluated. The results of the two tests were then compared. The main findings were:

- Statistical analysis showed that the results of the subjective test were significant.
- The subjects were able to identify differences in spatial realism between reproduction systems. This was independent of the type of program material.
- The systems were ranked in terms of spatial realism (from least to most) as mono, stereo, 3/2, four-loudspeaker ambisonic, eight-loudspeaker ambisonic then periphonic ambisonic.

- When compared to the objective measurements, the subjective results correlated highly, particularly with the SR values calculated using $IACC_{FB}$.
- In both the objective and subjective tests, the measured differences between non-mono systems were small.

7.5 Refinement of IACC as a Spatial Measure by Means of Frequency Weighting

As a means of sensitising the IACC measurement, frequency dependency was investigated through subjective testing. ICCC-variable test signals covering different frequency regions were adjusted to be of the same degree of spatial impression as mid-frequency signals of fixed ICCCs. To facilitate constant listening level comparisons, a custom designed mixed was built. The main findings were:

- Statistical analysis showed that the results were significant.
- For all reference signal ICCCs, test signals with ERBs centred on frequencies below 1 kHz had a higher ICCC value than the reference signal.

- The perceived degree of spatial impression, as indicated by ICCC measurements at low frequencies, is greater than the same ICCC measurement at mid-frequency.
- At frequencies above 1 kHz there were inconsistencies in the test results that made it difficult to draw inference.
- A method of frequency-weighting the IACC has been proposed that is based upon the results of the subjective test.

7.6 Further Work

7.6.1 Spatializing Techniques for Musical Synthesis

Regarding the spatializing techniques for musical synthesis, further work may entail continued subjective testing in order to establish greater confidence in the techniques and experimentation to develop the techniques further. As the success of the techniques appears to be limited by the degree of spatial spread, investigating the techniques using only two loudspeakers in the standard stereo configuration may be worthwhile.

Other areas involve developing the techniques to optimise spatial impression by investigating the grouping and positioning of the harmonics and creating asynchronous onsets by introducing short time delays to groups of harmonics.

By panning each harmonic or groups of harmonics over the loudspeakers of an array, perhaps from differing starting and ending points and at differing panning speeds, a more complex mixture of auditory cues will be presented to the hearing system that may result in spacious and 'moving' sound field.

In the investigation, the techniques have only been investigated in the horizontal plane. By distributing the harmonics over a three dimensional loudspeaker array, as either real or virtual sources, a greater sense of spatial impression may be achieved or this may result in a multiple sourced perception.

7.6.2 Objective and Subjective Evaluation of Surround Systems

In Chapter Four, the use of low value IACC measurements in calculating SR values was brought into question. Ignoring the low IACC measures in the SR calculations was suggested; however this may hide some of the errors of the reproduction systems thus reducing the saliency of the SR measurement. This area requires further investigation through re-examination of the data and better subjective alignment.

Whilst the results of the subjective and objective experiments correlate well (which reinforces the validity of the SR method of measuring the spatial capabilities of reproduction systems), the objective measurement method could be further optimised. The subjective results demonstrated that the subjects found the spatial differences between some reproduction systems to be subtle. Whilst this is somewhat reflected in the objective results, it would be desirable to 'sensitise' the objective measurement in order to better differentiate between small differences in the spatial performance of reproduction systems.

By re-examining the variations in and properties of the IACC measurement, using the knowledge gained form concert hall acoustics, it may be possible to facilitate the objective measurement to be able to better detect small spatial differences between systems. Even though the IACC variants used in the existing SR measurements ($IACC_E$, $IACC_3$ and $IACC_{E3}$) did not result more accurate measurements, it is proposed that the 'sensitising' of the procedures may be achieved by further investigating time windowing, variations in IACC over time and frequency filtering of the IACC measurement used to calculate the SR values.

For both the objective and subjective tests, different reproduction systems and encoding / decoding methods could be introduced. For example, transaural systems, ambiophonic systems and shelf-filtered ambisonic decoding methods.

7.6.3 Frequency Weighting of IACC

In Section 6.7, an application of the results of the subject test that examined equal spatial impression over frequency bands was suggested. To verify and refine this application of frequency weighting to the IACC, subjective testing would be necessary. In particular, the frequency-weighted IACC could be used in sensitising the SR measurement procedure.

The subjective test resulted in contradictory data at high frequencies. This could be further investigated to examine if frequencies above 1 kHz do contribute towards the perception of spatial impression and to find the maximum frequency limit of their contribution.

8 Appendices

Appendix A. Allocation of Subject's Responses to Preference Categories (Chapter 3).

Horizontal Left	Horizontal Right	Elevated Up	Elevated Down	Spatially Broader
Slightly to the left.	2 o'clock.	Higher.	Lower.	Image goes wide.
To the left.	3 o'clock.	Sources at different heights.	A little down.	One source, spread wide.
11 o'clock.	Right.	Up.	Down-wards.	Image broadens.
130º to the left.	Slightly to the right.	Quite high.	Lower in height.	Spatial, stereo effect.
Left.	From the right hand side.			Stereo.
30º to the left.	Just forward of 90º.			Wider source.
Pretty left.	Very right.			Spacious, like reflections.
	Extremely right			Very wide image.
				Spaced quality.
				More broader.
				Wider direction and source.
				Two loudspeakers placed at 90º.
				Creates space, like reverb.

Annoying	Multiple Sourced	Brighter	No Difference	Other
High frequency buzz. High pitched ring. High frequency ringing. High frequency tone interfering. Painful in right ear. Buzz in ear. More resonance, like feedback. Shriek in ear, not pleasant.	Two sources. Two points left and right. From both sides. Not like a single source. More than one source. Coming from either side. Two events.	Slightly brighter. Brighter/ Harsher. More brighter. Timbre cleaner. Brighter.	No difference. The same.	Louder. Quieter. 6 o’clock. Quality difference. Different balance. Closer. Further. Nasal. Similar in distance. Different pitch. Lower in pitch. Different timbre. More angular.

Appendix B. Matlab Coding for Room Simulation (Chapter 4)

```
%Simple room simulation. Here we take a signal, convolve with 0
degree HRTF for
%direct sound. Reflections are calculated using path difference
to give path diff and
%time delay,
%attenuation to give absorbtion and HRTF convolved to give angle
of reflection.

close

linevalue = 1; %enter 0 to switch off the wavwrite section, 1 to
switch on.Use 0 to
%select HRTFs before running program for proper.

rswitch=1; % switches on (1) or off (0) the right reflection
lswitch=1;
fswitch=1;
bswitch=1;
flswitch=1;
cswitch=1;
gain=0.5; %attenuates read in wavs to stop clipping

source=wavread('m14a');%enter sound source here - usually 'pink'

%the room is described in x and y terms where x is the breadth
and y the length. If a
%source was in the bottom right hand corner sy and sx = 0. For a
source was towards
%the left and back of the room it could be sy = 2 and sx = 1.
Have to know
%dimensions of the room, say 15 length and 10 breadth.

l=35; %length of room
b=20; %breadth of room
h=15; %height of room
sy = 33 ;%source length position
sx = 10;%source breadth position
ry =  32;%receiver
rx =  10.5;
sh=1.6; %source and receiver height from ground
rh=1.6;
StoR = sqrt((abs(sy-ry))^2+(abs(sx-rx))^2);%source to receiver
distance


%This is for first reflection
%off the right hand wall

if sx>rx
   opr=(sx-rx)+(2*(b-sx));
else
   opr=(2*(b-sx))-(rx-sx);
end
adjr=abs(sy-ry);
hypr=sqrt((opr^2)+(adjr^2));
```

```
RefDisR=hypr;%PathDifR= RefDisR-StoR;%for working out attenuation
PathDifR=RefDisR-StoR;
TimeDifR= PathDifR/343;%for working out time delay
AngR=((asin(opr/hypr))/pi)*180;%angle of receiver to reflection.
For selecting hrtf


%This is for second reflection
%off the left hand wall
if sx<rx
   opl=(rx-sx)+(2*sx);
else
   opl=(2*sx)-(sx-rx);
   end
adjl=(abs(sy-ry));
hypl=sqrt((opl^2)+(adjl^2));
RefDisL=hypl;
PathDifL=RefDisL-StoR;
TimeDifL=PathDifL/343;
AngL=360-((asin(opl/hypl))/pi)*180;


%this is for a reflection from the front (top) wall (l=max)

c=((sy-ry)+2*(l-sy));%distance between s and r plus twice the
distance between s
%and rear wall
d=(abs(sx-rx)); %x distance between s and r
RefDisF=sqrt(c^2+d^2);
PathDifF=RefDisF-StoR;
TimeDifF=PathDifF/343;
AngF1=((atan(d/c))/pi)*180;
if sx<rx
AngF1=360-AngF1;
end
if sx>rx
AngF1=AngF1;
end
AngF=AngF1;

%this is for the back (bottom) wall l=min

e=(abs(sx-rx));
f=(2*sy-(sy-ry));
RefDisB=sqrt(e^2+f^2);
PathDifB=RefDisB-StoR;
TimeDifB=PathDifB/343;
AngB1=((atan(e/f))/pi)*180;
if sx<rx
AngB1=AngB1;
end
if sx>rx
AngB1=-AngB1;
end

AngB=180+AngB1;

%this is for s to r angle
op=abs(sy-ry);
adj=abs(sx-rx);
```

```
AngSR1=((atan(op/adj))/pi)*180;
if sx<rx
   AngSR=360-(90-AngSR1);
else
   AngSR=0+(90-AngSR1);
end

%this is the floor

RefDisFl=sqrt((2*sh)^2+StoR^2);
PathDifFl=RefDisFl-StoR;
TimeDifFl=PathDifFl/343;
opf=2*sh;
AngFl1=((atan(opf/StoR))/pi)*180;
AngFl=-AngFl1;

%this is the ceiling
RefDisC=sqrt((2*(h-sh))^2+StoR^2);
PathDifC=RefDisC-StoR;
TimeDifC=PathDifC/343;
opc=2*(h-sh);
AngC1=((atan(opc/StoR))/pi)*180;
AngC=AngC1;


%attenuation delay and angles

Right_Atten = (StoR/RefDisR);
Left_Atten = (StoR/RefDisL);
Front_Atten = (StoR/RefDisF);
Back_Atten = (StoR/RefDisB);
Floor_Atten= (StoR/RefDisFl);
Ceil_Atten = (StoR/RefDisC);
Right_Delay = TimeDifR
Left_Delay = TimeDifL
Front_Delay = TimeDifF
Back_Delay = TimeDifB
Floor_Delay=TimeDifFl
Ceil_Delay = TimeDifC
Direct_Ang=AngSR;
Right_Ang = AngR;
Left_Ang = AngL;
Front_Ang = AngF;
Back_Ang = AngB;
Floor_Ang = AngFl;
Ceil_Ang = AngC;

%need to round angles to allow for hrtf select

rang=round(Right_Ang/5)*5
lang=round(Left_Ang/5)*5
fang=round(Front_Ang/5)*5
bang=round(Back_Ang/5)*5
srang=round(Direct_Ang/5)*5
flang=round(Floor_Ang/5)*5
cang=round(Ceil_Ang/5)*5
%set absorbtion

labsorb=0.8;%left wall
rabsorb=0.8;
```

```
fabsorb=0.95;
babsorb=0.5;
flabsorb=0.7;%floor
cabsorb=0.95;
if linevalue > 0.5%this 'swiches off' this particular chunk of
the programme


%create wavs for direct and reflections

%DIRECT
direct=source*gain;%attenuated to stop clipping

%REFLECTIONS - ATTENUATION AND DELAY
%Right Reflection
rdelay=round(Right_Delay*44100);%converts time diff into number
of samples
zr=zeros(1,rdelay)';%zeros to match number of samples
rig=source*rabsorb*Right_Atten*gain;%read in wav and attenuate
right=[zr;rig];%add zeros (delay) to wav
%Left Reflection
ldelay=round(Left_Delay*44100);%converts time diff into number of
samples
zl=zeros(1,ldelay)';%zeros to match number of samples
lef=source*labsorb*Left_Atten*gain;%read in wav and attenuate
left=[zl;lef];%add zeros (delay) to wav
%Front Reflection
fdelay=round(Front_Delay*44100);%converts time diff into number
of samples
zf=zeros(1,fdelay)';%zeros to match number of samples
fro=source*fabsorb*Front_Atten*gain;%read in wav and attenuate
front=[zf;fro];%add zeros (delay) to wav
%Back Reflection
bdelay=round(Back_Delay*44100);%converts time diff into number of
samples
zb=zeros(1,bdelay)';%zeros to match number of samples
bac=source*babsorb*Back_Atten*gain;%read in wav and attenuate
back=[zb;bac];%add zeros (delay) to wav
%Floor Reflection
fldelay=round(Floor_Delay*44100);%converts time diff into number
of samples
zfl=zeros(1,fldelay)';%zeros to match number of samples
flo=source*flabsorb*Floor_Atten*gain;%read in wav and attenuate
floor=[zfl;flo];%add zeros (delay) to wav
%Ceiling Reflection
cdelay=round(Ceil_Delay*44100);
zc=zeros(1,cdelay)';
ce=source*cabsorb*Ceil_Atten*gain;
ceil=[zc;ce];



%select hrtfs and convolve with signals.
%the floor and ceiling need to include the hdir azimuth
%as well as calculated height angle - but check resolution of
%mits hrtfs

hdir=wavread('H0e335a');
hr=wavread('H0e085a');
hl=wavread('H0e275a');
```

```
hf=wavread('H0e335a');
hb=wavread('H0e180a');
hfl=wavread('H-40e315a');%this will depend on s to r angle also
hc=wavread('H90e000a');

directL=conv(direct,hdir(:,1));%convolve direct with left hrtf
directR=conv(direct,hdir(:,2));
rightL=conv(right,hr(:,1));
rightR=conv(right,hr(:,2));
leftL=conv(left,hl(:,1));
leftR=conv(left,hl(:,2));
frontL=conv(front,hf(:,1));
frontR=conv(front,hf(:,2));
backL=conv(back,hb(:,1));
backR=conv(back,hb(:,2));
floorL=conv(floor,hfl(:,1));
floorR=conv(floor,hfl(:,2));
ceilL=conv(ceil,hc(:,1));
ceilR=conv(ceil,hc(:,2));

%to sum signals we have to make them all the same lenght so pad
ends with zeros

pad=270000;%make sure this is longer than the max length of
signals
paddirect=pad-length(directL);%number of zeros we need for each
sample
padright=pad-length(rightL);
padleft=pad-length(leftL);
padfront=pad-length(frontL);
padback=pad-length(backL);
padfloor=pad-length(floorL);
padceil=pad-length(ceilL);

zpd=zeros(1,paddirect)';%creates zeros for each sample to equal
'pad' when added
zpr=zeros(1,padright)';
zpl=zeros(1,padleft)';
zpf=zeros(1,padfront)';
zpb=zeros(1,padback)';
zpfl=zeros(1,padfloor)';
zpc=zeros(1,padceil)';

dL=[directL;zpd];%adding zeros to signals to make them all the
same lenght
dR=[directR;zpd];

zz=length(dL); %for use in switching refs on and off

if rswitch >0.5;%
   rL=[rightL;zpr];
else
   rL=zeros(1,zz)';
end


if rswitch > 0.5;%
   rR=[rightR;zpr];
else
   rR=zeros(1,zz)';
```

```
end


if lswitch > 0.5;%
   lL=[leftL;zpl];

else
   lL=zeros(1,zz)';
end


if lswitch > 0.5;%
   lR=[leftR;zpl];
else
   lR=zeros(1,zz)';
end



if fswitch > 0.5;%
   fR=[frontR;zpf];
else
   fR=zeros(1,zz)';
end


if fswitch > 0.5;%
   fL=[frontL;zpf];
else
   fL=zeros(1,zz)';
end



if bswitch > 0.5;%
   bL=[backL;zpb];
else
   bL=zeros(1,zz)';
end



if bswitch > 0.5;%
   bR=[backR;zpb];
 else
   bR=zeros(1,zz)';
end


if flswitch > 0.5;
   flL=[floorL;zpfl];
else
   flL=zeros(1,zz)';
end



if flswitch > 0.5;
   flR=[floorR;zpfl];
else
```

```
   flR=zeros(1,zz)';
end


if cswitch > 0.5;
   cL=[ceilL;zpc];
else
   cL=zeros(1,zz)';
end


if cswitch > 0.5;
   cR=[ceilR;zpc];
else
   cR=zeros(1,zz)';
end




%sum Letfs and rights to create ear signals

LeftEar=(dL+rL+lL+fL+bL+flL+cL)*0.3;
RightEar=(dR+rR+lR+fR+bR+flR+cR)*0.3;

wavwrite(LeftEar,44100,'14aL');
wavwrite(RightEar,44100,'14aR');


%iacc and itd

cor=xcorr(LeftEar,RightEar,50,'coeff');
tau=[-1.13:1.13/50:1.13];
[C,I]=max(cor);
iacc=max(cor)
itd=tau(I)

else

end
```

Appendix C. Matlab Coding for B Format Encoding and Decoding (Chapter 4)

```
%Simple room simulation. Here we take a signal, convolve with 0
degree HRTF for
%direct sound. Reflections are calculated using path difference
to give path diff and
%time delay,
%attenuation to give absorbtion. Angles of reflection are
calculated then the
%pick up pattern of a SF mic is simulated.

close

gain=0.4; %sets gain to prevent clipping of wavs

%the room is described in x and y terms where x is the breadth
and y the length. If a
%source was in the bottom right hand corner sy and sx = 0. For a
source was towards
%the left and back of the room it could be sy = 2 and sx = 1.
Have to know
%dimensions of the room, say 15 length and 10 breadth.

l=35; %length of room
b=20; %breadth of room
h=15; %height
sy = 33 ;%source length position
sx = 10;%source breadth position
ry =  2;%receiver
rx =  10.5;
sh=1.6;%source and receiver height
rh=1.6;

StoR = sqrt((abs(sy-ry))^2+(abs(sx-rx))^2);%source to receiver
distance

%This is for first reflection
%off the right hand wall

if sx>rx
   opr=(sx-rx)+(2*(b-sx));
else
   opr=(2*(b-sx))-(rx-sx);
end
adjr=abs(sy-ry);
hypr=sqrt((opr^2)+(adjr^2));
RefDisR=hypr;%PathDifR= RefDisR-StoR;%for working out attenuation
PathDifR=RefDisR-StoR;
TimeDifR= PathDifR/343;%for working out time delay
AngR=((asin(opr/hypr))/pi)*180;%angle of receiver to reflection.
For selecting hrtf


%This is for second reflection
%off the left hand wall
if sx<rx
   opl=(rx-sx)+(2*sx);
else
```

```
   opl=(2*sx)-(sx-rx);
   end
adjl=(abs(sy-ry));
hypl=sqrt((opl^2)+(adjl^2));
RefDisL=hypl;
PathDifL=RefDisL-StoR;
TimeDifL=PathDifL/343;
AngL=360-((asin(opl/hypl))/pi)*180;


%this is for a reflection from the front (top) wall (l=max)

c=((sy-ry)+2*(l-sy));%distance between s and r plus twice the
distance between s
%and rear wall
d=(abs(sx-rx)); %x distance between s and r
RefDisF=sqrt(c^2+d^2);
PathDifF=RefDisF-StoR;
TimeDifF=PathDifF/343;
AngF1=((atan(d/c))/pi)*180;
if sx<rx
AngF1=360-AngF1;
end
if sx>rx
AngF1=AngF1;
end
AngF=AngF1;

%this is for the back (bottom) wall l=min

e=(abs(sx-rx));
f=(2*sy-(sy-ry));
RefDisB=sqrt(e^2+f^2);
PathDifB=RefDisB-StoR;
TimeDifB=PathDifB/343;
AngB1=((atan(e/f))/pi)*180;
if sx<rx
AngB1=AngB1;
end
if sx>rx
AngB1=-AngB1;
end

AngB=180+AngB1;

%this is for s to r angle
op=abs(sy-ry);
adj=abs(sx-rx);
AngSR1=((atan(op/adj))/pi)*180;
if sx<rx
   AngSR=360-(90-AngSR1);
else
   AngSR=0+(90-AngSR1);
end

%this is the floor

RefDisFl=sqrt((2*sh)^2+StoR^2);
PathDifFl=RefDisFl-StoR;
TimeDifFl=PathDifFl/343;
```

```
opf=2*sh;
AngFl1=((atan(opf/StoR))/pi)*180;
AngFl=-AngFl1;

%this is the ceiling
RefDisC=sqrt((2*(h-sh))^2+StoR^2);
PathDifC=RefDisC-StoR;
TimeDifC=PathDifC/343;
opc=2*(h-sh);
AngC1=((atan(opc/StoR))/pi)*180;
AngC=AngC1;

Right_Atten = (StoR/RefDisR)
Left_Atten = (StoR/RefDisL)
Front_Atten = (StoR/RefDisF)
Back_Atten = (StoR/RefDisB)
Floor_Atten = (StoR/RefDisFl)
Ceil_Atten = (StoR/RefDisC)
Right_Delay = TimeDifR
Left_Delay = TimeDifL
Front_Delay = TimeDifF
Back_Delay = TimeDifB
Floor_Delay=TimeDifFl
Ceil_Delay=TimeDifC
Direct_Ang=AngSR;
Right_Ang = AngR;
Left_Ang = AngL;
Front_Ang = AngF;
Back_Ang = AngB;
Floor_Ang=AngFl;
Ceil_Ang=AngC;



%set absorbtion

rabsorb=0.8;
labsorb=0.8;
fabsorb=0.95;
babsorb=0.5;
flabsorb=0.7;
cabsorb=0.95;

%create wavs for direct and reflections

%DIRECT
direct=wavread('m14a')*gain;%gain is to avoid clipping

%REFLECTIONS - ATTENUATION AND DELAY
%Right Reflection
rdelay=round(Right_Delay*44100);%converts time diff into number
of samples
zr=zeros(1,rdelay)';%zeros to match number of samples
rig=wavread('m14a')*rabsorb*Right_Atten*gain;%read in wav and
attenuate
right=[zr;rig];%add zeros (delay) to wav
%Left Reflection
ldelay=round(Left_Delay*44100);%converts time diff into number of
samples
zl=zeros(1,ldelay)';%zeros to match number of samples
```

```
lef=wavread('m14a')*labsorb*Left_Atten*gain;%read in wav and
attenuate
left=[zl;lef];%add zeros (delay) to wav
%Front Reflection
fdelay=round(Front_Delay*44100);%converts time diff into number
of samples
zf=zeros(1,fdelay)';%zeros to match number of samples
fro=wavread('m14a')*fabsorb*Front_Atten*gain;%read in wav and
attenuate
front=[zf;fro];%add zeros (delay) to wav
%Back Reflection
bdelay=round(Back_Delay*44100);%converts time diff into number of
samples
zb=zeros(1,bdelay)';%zeros to match number of samples
bac=wavread('m14a')*babsorb*Back_Atten*gain;%read in wav and
attenuate
back=[zb;bac];%add zeros (delay) to wav
%Floor Reflection
fldelay=round(Floor_Delay*44100);%converts time diff into number
of samples
zfl=zeros(1,fldelay)';%zeros to match number of samples
flo=wavread('m14a')*flabsorb*Floor_Atten*gain;%read in wav and
attenuate
floor=[zfl;flo];%add zeros (delay) to wav
%Ceiling Reflection
cdelay=round(Ceil_Delay*44100);%converts time diff into number of
samples
zc=zeros(1,cdelay)';%zeros to match number of samples
cei=wavread('m14a')*cabsorb*Ceil_Atten*gain;%read in wav and
attenuate
ceiling=[zc;cei];%add zeros (delay) to wav



%to sum signals we have to make them all the same lenght so pad
ends with zeros

pad=270000;%make sure this is longer than the max length of
signals
paddirect=pad-length(direct);%number of zeros we need for each
sample
padright=pad-length(right);
padleft=pad-length(left);
padfront=pad-length(front);
padback=pad-length(back);
padfloor=pad-length(floor);
padceiling=pad-length(ceiling);


zpd=zeros(1,paddirect)';%creates zeros for each sample to equal
'pad' when added
zpr=zeros(1,padright)';
zpl=zeros(1,padleft)';
zpf=zeros(1,padfront)';
zpb=zeros(1,padback)';
zpfl=zeros(1,padfloor)';
zpc=zeros(1,padceiling)';


d=[direct;zpd];%adding zeros to signals
```

```
r=[right;zpr];
l=[left;zpl];
f=[front;zpf];
b=[back;zpb];
fl=[floor;zpfl];
c=[ceiling;zpc];


%convert angles into rads
dang=(Direct_Ang/180)*pi;
rang=(Right_Ang/180)*pi;
lang=(Left_Ang/180)*pi;
fang=(Front_Ang/180)*pi;
bang=(Back_Ang/180)*pi;
flang=(Floor_Ang/180)*pi;
cang=(Ceil_Ang/180)*pi;


%encode to b format and write wavs

W=((1/sqrt(2))*(d+r+l+f+b+fl+c));
X=(d*cos(dang))+(r*cos(rang))+(l*cos(lang))+(f*cos(fang))+(b*cos(
bang))+(fl*cos(dang)*cos(flang))+(c*cos(dang)*cos(cang));
Y=(d*sin(dang))+(r*sin(rang))+(l*sin(lang))+(f*sin(fang))+(b*sin(
bang))+(fl*sin(dang)*cos(cang))+(c*sin(dang)*cos(cang));
Z=(fl*sin(flang))+(c*sin(cang));

wavwrite(W,44100,'wp12');
wavwrite(X,44100,'xp12');
wavwrite(Y,44100,'yp12');
wavwrite(Z,44100,'zp12');

%ambidec - this decodes b format signals to 4 ls signals then
creates binaural signal
%Using CLOCKWISE ENCODE/DECO

w=wavread('wp12')*0.5;%W component
x=wavread('xp12')*0.5;%X component
y=wavread('yp12')*0.5;%Y component


%LS Decode - assumes clockwise encoding and decoding

lf=w+x-y;
lb=w-x-y;
rb=w-x+y;
rf=w+x+y;

%ls signals to ears - hrtf conv

hlf=wavread('H0e315a');
hlb=wavread('H0e225a');
hrb=wavread('H0e135a');
hrf=wavread('H0e045a');

lfL=conv(lf,hlf(:,1));
lfR=conv(lf,hlf(:,2));
lbL=conv(lb,hlb(:,1));
lbR=conv(lb,hlb(:,2));
rbL=conv(rb,hrb(:,1));
```

```
rbR=conv(rb,hrb(:,2));
rfL=conv(rf,hrf(:,1));
rfR=conv(rf,hrf(:,2));

L=0.4*(lfL+lbL+rbL+rfL);
R=0.4*(lfR+lbR+rbR+rfR);

wavwrite(L,44100,'p12l');
wavwrite(R,44100,'p12r');
```

Appendix D. Decoding of the B Format Signals (Chapter 4)

Reproduction System	Decoding
Mono	Omnidirectional Mic = W
Stereo	L = W + 1.22X + 0.71Y R = W + 1.22X - 0.71Y
3 / 2 Layout	L = W + 1.22X + 0.71Y C = W + 1.42X R = W + 1.22X - 0.71Y SL = W – 0.71X + 1.22Y SR = W – 0.71X – 1.22Y
Ambisonic 4 Loudspeaker Pantophonic	LF = W + X + Y RF = W + X – Y LR = W - X + Y RR = W - X - Y
Ambisonic 8 Loudspeaker Pantophonic	CF = W + 1.42X LF = W + X + Y LS = W + 1.42Y LR = W – X + Y CR = W – 1.42X RR = W - X – Y RS = W - 1.42Y RF = W + X – Y
Ambisonic 8 Loudspeaker Periphonic	LFU = W + 1.1X + 0.9Y + 0.8Z LBU = W -1.1X + 0.9Y + 0.8Z RBU = W - 1.1X - 0.9Y + 0.8Z RFU = W + 1.1X - 0.9Y + 0.8Z LFD = W + 1.1X + 0.9Y - 0.8Z LBD = W -1.1X + 0.9Y - 0.8Z RBD = W - 1.1X - 0.9Y - 0.8Z RFD = W + 1.1X - 0.9Y - 0.8Z

The periphonic rig was not an exact cube. The elevation angles were ± 35° and the azimuth were in 40° steps.

Appendix E. Subject Instructions (Spatial Impression of Reproduction Systems – Chapter 5)

Thank you for agreeing to participate in the experiment. Your task is to judge a number of audio examples in terms of their spatial attributes.

You will be presented with audio examples in groups of six. A computer keyboard will act as a switch allowing you listen to each of the examples in a group. You listen to each of the six examples by pressing the 3, 4, 5, 6, 7 or 8 keys on the keyboard. You can listen to the examples as many times as you wish and can take as long as is needed to make your judgements.

You are asked to judge each example **solely in terms of spatial impression.** How realistic is the spatial reproduction? You may wish to consider some spatial attributes that have been used to describe concert halls:

- Source width. Does the sound source (instrument/voice) appear to be broad?
- Envelopment. Does the sound appear to come from all around you?
- Environment width. Does the environment appear to be broad?

Try to recall your experiences of listening in real concert halls. How well does the audio example, in terms of spatial attributes, compare to your experiences? To grade this you are asked to mark a ten point grading scale. The extremes of the scale are described as:

0	10
None of the spatial attributes of concert hall listening were present in the example.	The spatial attributes of the example were identical or near identical to those of a concert hall

For each example of the group, please draw a vertical mark at any point on the grading line at the position that corresponds to your judgement.

Your preference judgements should be recorded on the sheets provided. Once you have recorded your answers for a group of examples, please say ‘Finished’ or words to that effect so that the next group may be presented. Please allow about 10 seconds before pressing the numerical keys again.

Please feel free to ask any questions before the test starts and thank you again for your time.

Appendix F. Subject Instructions (IACC Frequency Dependency – Chapter 6)

Firstly, thank you for agreeing to participate in this experiment. Your task is to compare two sound samples in terms of spatial impression, in particular the apparent source width of the sound. You will be presented with a number of looped pairs of samples consisting of a reference signal and a variable test signal.

You will be able to vary the apparent source width of the test signal by adjusting the dial on the box. To help you differentiate between the test signal and the reference signal;

1. The test signal is the longer of the two samples.
2. The display of the oscilloscope will illuminate when the test signal is sounding.

You are asked to adjust the apparent source width of the test signal so that it is the same width as the reference signal. You can take as long as you like to compare the two signals. When you have achieved equal source width, please remove your hand from the dial and say ‘Finished’ or words to that effect. After a short pause, the experimenter will ask you to reset the dial to a random position, ready for the next pair of samples.

In adjusting the degree of source width, if you are unable to further adjust the source width because you have reached either of the end positions of the dial, leave the dial at the nearest end position.

Please feel free to ask any questions at this point.

Enjoy yourselves and thank you for your time!

Appendix G. Example Calculation of Frequency-Weighted IACC (Chapter 6)

Frequency Band cf (Hz)	125	250	500	1000
IACC	0.8942	0.7932	0.4687	0.4241

IACC, in ERB frequency bands. Taken from a Peel Hall impulse response.

Frequency Band cf (Hz)	125	250	500	1000
Ref IACC 0.2	0.8139	0.6590	0.5414	0.2899
Ref IACC 0.4	0.8259	0.8007	0.6676	0.3933
Ref IACC 0.6	0.8845	0.8866	0.5614	0.5356
Ref IACC 0.8	0.8455	0.9207	0.7937	0.7512

IACCs of equal spatial impression taken from the results of the subjective test.

- The signals under test are ERB filtered, at centre frequencies of 125, 250, 500 and 1000 Hz. IACC measurements are taken in each ERB. These measurements are hence referred to as I_M.
- For each I_M, the I_M of the 1 kHz ERB is noted as I_{M1k}. I_{M1k} is then compared to the ICCC measurements at 1 kHz as given by the results of the subjective test (these are the four IACC values in 1000 Hz column of the lower table). The ICCC measurement that is closest to the I_{M1k} is denoted I_C.
- I_C will have corresponding ICCC values in lower frequency ERBs that have equal degrees of perceived spatial impression, as determined by the subjective test. These ICCC values are denoted I_{EQ}.

$$\textbf{IACC in each ERB} = I_C + (I_M - I_{EQ})$$

For the 125 Hz IACC measurement:

I_M = 0.8942
I_{M1k} = 0.4241
I_C = 0.3933
I_{EQ} = 0.8259

IACC in 125 Hz ERB = 0.3933+(0.8942-0.8259) = 0.4616

IACC in 250 Hz ERB = 0.3858

IACC in 500 Hz ERB = 0.1944

IACC in 1 kHz ERB = 0.4241

Average = 0.3665

9 References

[Airo et al. 1996] Airo, E., Pekkarinen, J. and Olkinuora, P. : ’Listening to music over headphones: an assessment of noise exposure’, **Acta Acustica united with Acustica**, vol. 82, pp. 885-894.

[Analog Devices 2003] Analog Devices, : ‘Voltage-controlled amplifier/OVCE SSM-2018, a voltage-controlled panner’, **Analogue Devices**, http://www.analog.com/.

[Ando 1985] Ando, Y. : ‘**Concert Hall Acoustics’**, (Springer-Verlag, Berlin).

[Bareham 1996] Bareham, J.R. : ‘Measurement of spatial characteristics of sound reproduced in listening spaces’, **Audio Engineering Society Preprint**, 101st Convention, preprint no. 4381.

[Barron 1999] Barron, M. : ’Spatial impression and envelopment in concert halls”, **Proceedings of the Institute of Acoustics**, vol. 21, Part 6, pp. 163-170.

[Barron and Marshall 1981] Barron, M. and Marshall, A.H. : ‘Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure’, **Journal of Sound and Vibration**, vol. 77, no. 2, pp. 211-232.

[Bech and Zacharov 2006] Bech, S. and Zacharov, N. : **‘Perceptual audio evaluation – theory method and application’**, Wiley and Sons, Chichester, UK.

[Begault 1991] Begault, D. :'Challenges to the successful implementation of 3D sound', **Journal of the Audio Engineering Society**, vol. 39, pp. 864-870.

[Begault 1994] Begault, D. :**'3D sound for virtual reality and multimedia**', (Academic Press).

[Belschner 2001] Belschner, T. : 'Snare drum sample', **DAFX Sound Library** , http://echo.gaps.ssr.upm.es/COSTG6/sound.php3.

[Berg and Rumsey 1999] Berg, J and Rumsey, F. : 'Spatial attribute identification and scaling by repertory grid technique and other methods', **Audio Engineering Society Preprint**, 16th International Conference, paper no. 16-005.

[Blauert 1997] Blauert, J. : **'Spatial Hearing'**, Revised Edition, MIT Press, Cambridge USA.

[Blauert and Lindemann 1986] Blauert,J. and Lindemann, W. : 'Spatial mapping of intercranial auditory events for various degrees of interaural coherence', **Journal of the Acoustical Society of America**, vol. 79, pp. 806-813.

[Blumlein 1931] Blumlein, A. :'Improvements in and relating to sound transmission, sound recording and sound reproducing systems', **British Patent Specification**, 394325.

[Bradley and Soulodre 1995] Bradley, J.S and Soulodre, G.A. :'Objective measures of listener envelopment', **Journal of the Acoustical Society of America** , vol. 98 pp, .2590-2597.

[Bradley et al. 1993] Bradley, J.S., Soulodre, G.A. and Popplewell, N. : 'Pilot study of simulated spaciousness', **Journal of the Acoustical Society of America**, vol. 93, p 2283.

[Bregman 1999] Bregman, A.S. : **'Auditory scene analysis'**, MIT Press, Cambridge USA.

[BS EN ISO 3382 2000] British Standards Institute. : 'Acoustics – measurement of the reverberation times of rooms with reference to other acoustical parameters', **British Standards**.

[Buck 1983] Buck, M. :'Measurement of spatial reproduction quality in interaural cross-correlation techniques', **Audio Engineering Society Preprint**, 74th Convention, paper no. 2044.

[Buell and Hafter 1991] Buell, T.N. and Hafter, E.R. : 'Combination of binaural information across frequency bands', **Journal of the Acoustical Society of America**, vol. 390, pp. 1894-1900.

[Cardozo 1965] Cardozo, B.L. : 'Adjusting the method of adjustment: SD vs DL', **Journal of the Acoustical Society of America**, vol. 37, pp. 786-792.

[Chen 2001] Chen, T. :'Surround sound beyond 5.1', **Ambisonic.Net**, http://www.ambisonic.net/articles.html

[Chernyak and Dubrovsky 1968] Chernyak, R.I. and Dubrovsky, N.A. : 'Pattern of the noise images and the binaural summation of loudness for the different interaural correlation of noise', Proceedings of the 6th International Congress on Acoustics. pp A53-A56.

[Cooper and Shiga 1972] Cooper, D and Shiga, T. :'Discrete matrix multichannel stereo', **Journal of the Audio Engineering Society**, vol. 20, pp. 346-360.

[Cox et al. 1993] Cox, T.J., Davies, W.J. and Lam, Y.W. :'The sensitivity of listeners to early sound field changes in auditoria', **Acustica,** vol.79, pp. 27-41.

[Damaske 1967] Damaske, P. :'Subjektive untersuchungen von schallfeldern (subjective investigations of soundfields)', **Acustica**, vol.19, pp. 199-213.

[Davis 2003] Davis, M.F. :'History of spatial coding', **Journal of the Audio Engineering Society**, vol. 251, pp. 554-569.

[Denon 1994] Denon. : '**Anechoic Orchestral Music Recording**', Performed by the Osaka Philharmonic Orchestra and conducted by Enkoji, Kawaguchi and Tsubonou. ASIN: B0000034M9.

[Dolby 1999] Dolby Laboratories Inc. ;'Surround sound past present and future', **Dolby.Com** http://www.dolby.com/

[Farina and Ugolotti 1998] Farina, A and Ugolotti, E. : 'Software implementation of B-format encoding and decoding', **Audio Engineering Society Preprint**, 104th Convention, preprint no. 4691.

[Farrer 1979a] Farrer, K. :'Soundfield microphone: Design and development of microphone and control unit', **Wireless World**, vol 85, pp. 48-50 (Oct.).

[Farrer 1979b] Farrer, K. :'Soundfield microphone 2: Detailed functioning of control unit', **Wireless World**, vol 85, pp. 99-103 (Nov.).

[Field 2000] Field, A. : '**Discovering Statistics Using SPSS for Windows**', Sage Publications, London.

[Fletcher et al. 1962] Fletcher, H., Blackham, E.D. and Stratton, R. : 'Quality of piano tones', **Journal of the Acoustical Society of America**, vol. 34, pp. 749-761.

[Fletcher and Munson 1933] Fletcher, H and Munson, W.A. : 'Loudness, it's definition, measurement and calculations', **Journal of the Acoustical Society of America**, vol. 5, pp. 82-108.

[Fredriksson and Zacharov 2002] Fredriksson, M. and Zacharov, N. :'Natural reproduction of music and environmental sounds', **Audio Engineering Society Preprint**, 112th Convention, preprint no. 5581.

[Furlong 1989] Furlong, D.J. : 'Comparative study of effective soundfield reconstruction', **Audio Engineering Society Preprint**, 87th Convention, preprint no. 2842.

[Furse and Malham 1999] Furse, R.W.E. and Malham, D. 'First and second order ambisonic decoding equations', **3D Audio**, http://www.muse.demon.co.uk/ref/speakers.html.

[Gardner and Martin 1994] Gardener, B. and Martin, K. : 'HRTF measurements of a Kemar dummy-head microphone', **MIT Media Lab**, http://sound.media.mit.edu/~kdm/hrtf.html.

[Gerzon 1973] Gerzon M.A. : 'Periphony: with height sound reproduction', **Journal of the Audio Engineering Society**, vol. 21, pp. 2-10.

[Gerzon 1983] Gerzon M.A. : 'Ambisonics in multichannel broadcasting and video', **Audio Engineering Society Preprint**, 74th Convention, preprint no. 2034.

[Gerzon 1985] Gerzon, M.A. : 'Ambisonics in multichannel broadcasting and video', **Journal of the Audio Engineering Society**, vol. 33, no. 11, pp. 859-871.

[Gerzon and Barton 1992] Gerzon, M.A. and Barton, G.J. : 'Ambisonic decoders for HDTV', **Audio Engineering Society Preprint**, 92nd Convention, preprint no. 3345.

[Glasberg and Moore 1992] Glasberg, B.R and Moore, B.C.J. : 'Derivation of auditory filter shapes from notched-noise data', **Hearing Research,** vol. 47, pp. 103 – 138.

[Glasgal 2003] Glasgal, R. ; 'Surround ambiophonic recording and reproduction', **Audio Engineering Society Preprint**, 24th International Conference, preprint no. 24-042.

[Grantham and Wightman 1978] Grantham, W.D. and. Wightman, F.L. : 'Detectability of varying interaural temporal differences', **Journal of the Acoustical Society of America**, vol. 63, pp. 511-523.

[Graziano and Raulin 1997] Graziano, A.M. and Raulin, M.L. : **'Research methods a process of inquiry',** Longman Publications, New York, USA.

[Griesinger 1989] Griesinger, D. : 'Equalization and spatial equalization of dummy-head recordings for loudspeaker reproduction', **Journal of the Audio Engineering Society**, vol. 37, no. 1/2, pp. 20-29.

[Griesinger 1992] Griesinger, D. : 'Measures of spatial impression and reverberance based on the physiology of human hearing', **Audio Engineering Society Preprint**, 11th International Audio Engineering Society Conference. Paper 11-016.

[Griesinger 1997] Griesinger, D. : 'The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces', **Acustica**, vol. 83 pp. 721-731.

[Griesinger 1998] Griesinger, D. : 'General overview of spatial impression, envelopment, localization and externalization', **Audio Engineering Society Preprint**, 15th International Audio Engineering Society Conference. Paper 15-013.

[Hertz 1981] Hertz, B.F, :'100 years with stereo: The beginning', **Journal of the Audio Engineering Society**, vol. 25, no. 5, pp. 368-372.

[Hidaka et al. 1995] Hidaka, T., Beranek, L.L. and Okano, T. : 'Interaural cross-correlation, lateral fraction and low and high frequency sound levels as measures of acoustical quality in concert halls', **Journal of the Acoustical Society of America**, vol. 98, no. 2, pp. 988-1007.

[Huopaniemi 2000] Huopaniemi, J. : 'Anechoic speech sample', **Reverberation Modelling**, http://www.acoustics.hut.fi/~riitta/.reverb/.

[IEC 61672 – 1] International Electrotechnical Commission. : 'Electroacoustics - Sound level meters - Part 1: Specifications', **International Electrotechnical Commission,** Recommendations IEC 61672 – 1.

[International Standards Organisation 1997] International Standards Organisation, : ‘Acoustics – Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters’, **International Standards Organisation,** ISO 3382:1997(E).

[ITU 2003] ITU Recommendation BS.1534-1, : ‘Method for the subjective assessment of intermediate quality level of coding systems’, **International Telecommunications Union Radiocommunication Assembly**, 2003.

[Jeffress 1948] Jeffress, L.A. : ‘A place theory of sound localisation’, **Journal of Comparative Physiology and Psychology**, vol. 61, pp. 523-527.

[Kendal 1995] Kendall, G.S. : ‘The decorrelation of audio signals and its impact on spatial imagery’, **Computer Music Journal**, vol. 19 pp.71-87.

[Kuhn 1987] Kuhn, G.F., : **‘Physical acoustics and measurement pertaining to direction hearing’**, in W. A. Yost and G. Gourewitch, Eds., *Directional Hearing,* Springer, New York, NY.

[Lawless and Heymann 1998] Lawless, H.T. and Heyman, H. : **‘Sensory evaluation of food’**, Chapel and Hall, London.

[Mac Cabe and Furlong 1994] Mac Cabe, C.J. and Furlong, D.J. : ‘ Virtual imaging capabilities of surround sound systems’, **Journal of the Audio Engineering Society**, vol. 42, no.1/2, pp. 38-49.

[Macpherson 1991] Macpherson, E.A. : ‘A computer model of binaural localization for stereo imaging measurement’, **Journal of the Audio Engineering Society**, vol. 39, no.9, pp. 604-621.

[Marshall 1967] Marshall, A.H. ‘A note on the importance of room cross-section in concert halls’, **Journal of Sound and Vibration**, vol. 5, pp. 100-112.

[Mason 2001] Mason, R. : ‘Elicitation and measurement of auditory spatial attributes in reproduced sound’, Doctorial Thesis, University of Surrey.

[Mason and Rumsey 2000] Mason, R and Rumsey, F : ’An assessment of the spatial performance of virtual home theatre algorithms by subjective and objective methods’, **Audio Engineering Society Preprint**, 108th Audio Engineering Society Convention. Preprint no. 5137.

[Mason and Rumsey 2001] Mason, R and Rumsey, F : ’Interaural time difference fluctuations: their measurement, subjective perceptual effect, and application in sound reproduction’. **Audio Engineering Society Preprint**, 19th International Audio Engineering Society Conference. Preprint no. 1894.

[Mathers 1979] Mathers, C.D and Landsdowne, K.F.L : ‘Hearing risk to wearers of circumaural headphones: an investigation’, Technical report BBC RD 1979/3, **The British Broadcasting Corporation**.

[McAdams 1984] McAdams, S. : **‘Spectral fusion, spectral parsing and the formation of auditory images’**, Doctorial Thesis, Stanford University, USA.

[Miller 2003] Miller, R.E. : ’Scalable tri-play recording for stereo, ITU 5.1/6.1 2d, and periphonic 3d (with height) compatible surround sound reproduction’. **Audio Engineering Society Preprint**, 115th Audio Engineering Society Convention. Preprint no. 5934.

[Moller et al. 1995] Moller, H., Hammershøi, D., Jensen, C.B. and Sørensen, M. F. : 'Transfer characteristics of headphones measured on human ears'. **Journal of the Audio Engineering Society**, vol. 43, no.4, pp. 203-217.

[Moore 1997] Moore, B.C.J. : **'Introduction to the psychology of hearing'**, Fourth Edition, Academic Press, London.

[Morimoto and Iida 1995] Morimoto, M. and Iida, K. : 'A practical evaluation method of auditory source width in concert halls', **Journal of the Acoustical Society of Japan,** vol 16, no. 2, pp. 59-69.

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

[Newson 2000] Newson, R. : 'Somersd – Confidence intervals for rank order statistics and their differences', **Stata Technical Bulletin,** 55: 47-55. Reprinted in Stata Technical Bulletin reprintsa vol 10, pp. 312-322.

[Okano et al. 1998] Okano, T., Beranek L. and Hidaka, T. : 'Relations among interaural cross-correlation coefficient ($IACC_e$), lateral fraction (LF_e), and apparent source width in concert halls', **Journal of the Acoustical Society of America** vol 104, pp. 255-265.

[Philipson et al. 2002] Philipson, P.J., Hirst, J.M. and Woolard, S. : 'Investigation into a method for predicting the perceived azimuth position of a virtual image', **Audio Engineering Society Preprint**, 112th Convention, preprint no. 5597.

[Plenge 1972] Plenge, G. : 'Uber das problem der im-kopf-lokalisation (On the problem of in-head localization)', **Acustica**, vol. 26, no. 5, pp. 241-252.

[Pollack 1978] Pollack, I. : 'Temporal Switching Between Binaural Information Sources', **Journal of the Acoustical Society of America** vol 63, pp. 550-558.

[Pulkki et al. 1999] Pulkki, V., Karjalainen, M. and Huopaniemi, J. : 'Analyzing virtual sound sources using a binaural auditory model', **Journal of the Audio Engineering Society**, vol. 47, no.4, pp. 203-217.

[Rumsey 1998] Rumsey, F. : 'Subjective assessment of the spatial attributes of reproduced sound', **Audio Engineering Society Preprint**, 15th International Conference, paper no. 15-012.

[Rumsey 1999] Rumsey, F. : 'Controlled subjective assessments of two-to-five-channel surround sound processing algorithms', **Journal of the Audio Engineering Society,** vol.47 pp. 563-581.

[Rumsey 2001] Rumsey, F. : **'Spatial Audio'**, Focal Press, Oxford.

[Samplenet 2000] http://www.samplenet.co.uk., **Samplenet**

[Sayers and Cherry 1957] Sayers, B and Cherry, E.C. : 'Mechanism of binaural fusion in the hearing of speech', **Journal of the Acoustical Society of America**, vol. 29 pp. 973-987.

[Schouten et al. 1962] Schouten, J.F., Ritsma, R.J. and Cardozo, B.L. : 'Pitch of the residue', **Journal of the Acoustical Society of America**, vol. 34, pp. 1418-1424.

[Tohyama and Suzuki 1989] Tohyama, M. and Suzuki, A. : 'Interaural cross-correlation coefficients in stereo reproduced sound fields',. **Journal of the Acoustical Society of America,** vol. 85, pp.780-786.

[Turgeon 1999] Turgeon M. : 'Cross-spectral auditory grouping using the paradigm of rhythmic masking release', Doctoral Thesis, McGill University, Montreal.

[Yanagawa and Tohyama 1998] Yanagawa, H. and Tohyama, M. : 'Sound image broadening by a single reflection considering temporal change of interaural cross-correlation', **Acustica**, vol. 87, pp. 247-252.

[Zacharov 1998] Zacharov, N. : 'Subjective appraisal of loudspeaker directivity for multichannel reproduction", **Journal of the Audio Engineering Society,** vol. 46 pp. 288-303.

[Zacharov et al. 1999] Zacharov, N., Huopaniemi, J. and Hamalainen, M. : 'Round robin subjective evaluation of virtual home theatre sound systems at the 16th International Conference', **Audio Engineering Society Preprint** , 16th International Conference, paper no. 16-049.