THIS THESIS MAY BE READ ONLY IN THE LIBRARY

Reader's Declaration

I undertake not to reproduce any portion of, or use any information derived from this thesis without first obtaining the permission, in writing, of the Librarian, Trinity College. If permission is granted, I shall give appropriate acknowledgement for any portion of the thesis used or reproduced.

Date consulted	Name and address in block letters	University or institution	Signature

SPACIOUSNESS CONTROL FOR SOUND FIELD RECORDING AND RECONSTRUCTION

A dissertation submitted to the University of Dublin for the degree of Doctor of Philosophy

July 2015.

Marco Conceição

Financially supported by FCT under the QREN – POPH – Type 4.1 – "Formação Avançada", subsidized by the European Social Fund and national funds of MEC, Portugal; PhD grant SFRH/BD/45641/2008

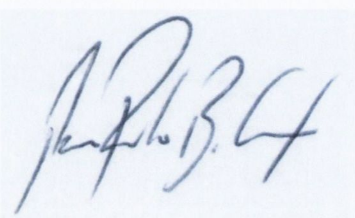
Music and Media Technologies
Department of Electronic and Electrical Engineering
Trinity College Dublin

DECLARATION

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and that it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgment.

Signed,



Marco Conceição

The *art* of recording lies in manipulating illusions.
The *science* of recording involves the tools and techniques
used to create these illusions.

(Streicher and Dooley, 1985)

ABSTRACT

"Auditory Spaciousness" is a perceptual characteristic which has been recognized as being an important aesthetic feature of music presentations of various types. Most previous research has been focused on the identification of the physical parameters which are causally involved in the generation of listener spatial impression when auditioning music. The presence of lateral reflections has been identified as contributing to room spatial impression. Significantly, it has also been recognised that lateral reflections generate an inter-aural dissimilarity which is of primary importance for a spatial impression, and that inter-aural cross correlation (IACC) can therefore be used as an indirect spaciousness index.

Stereo recorded music can also generate listener spaciousness, even though stereo is incapable of generating the lateral reflections which are usually regarded as being responsible for spaciousness. There is consequently an anomaly in this stereo perception, one that is worthy of study if the spaciousness experienced for other reconstructed formats is to be optimised.

The objectives of the study undertaken were: first, to establish perceptually significant metrics, based on those used in room acoustics, to be used in assessing 2-channel stereo and surround sound recordings; second to comparatively assess a number of microphone arrays by examining the reconstruction effectiveness of spaciousness delivered by them when played back through stereo and surround; and third to develop a spaciousness processor for use with a 5.1 multichannel array which will derive from standard 2-channel recording methods using perceptual reconstruction for spaciousness enhancement.

Anechoic chambers are usually used in the study of spaciousness, but the difficulty of access to such facilities can be a constraint. Because of this, in the present work, a variable experimental setup was introduced that made possible the control of spaciousness in different rooms. Measurements were performed under controlled conditions in which a dummy-head microphone system captured the signals for variable sound fields so that IACC could be measured. It was concluded that there is a similar trend evident in IACC results with

repetition of the experiments in different rooms. That is, measurement room acoustic details are not crucial to observed trends in IACC measurements resulting from recording and production parameter variation.

IACC measurements are used as a physical index which relates to listener spaciousness experience in a comparative study of the influence on spaciousness of different microphone arrays and microphone signal processing, thus allowing an objective approach to be adopted in the exploration of how microphone arrays affect the perceived spaciousness for stereo and surround sound reconstructions. That is, different microphone arrays recorded direct and simulated indirect sound components, and the recorded signals were played back in three different rooms. IACC measurements were made for the reconstructed sound fields using a dummy head microphone system. The results achieved showed how microphone array details influence the IACC peak, and lead to a better understanding of how spaciousness can be controlled for 2-channel stereo, 3-channel stereo, and 5.1 presentations. Parametric variation of microphone arrays details can therefore be employed to facilitate spaciousness control for reconstructed sound fields.

Having examined stereo reproduction variations in terms of auditory spaciousness alterations using IACC measurements as a spaciousness index, a surround sound spaciousness processor was developed which allows for auditory spaciousness control with 5.1 surround system music reproduction. A spaciousness processor Virtual Studio Technology (VST) plug-in was developed and assessed with regard to its influence on the auditory spaciousness experienced by reconstructed sound field listeners. It is suggested, based on the informal reports that are already in the literature and on the results achieved in this study, that this spaciousness processor can make contributions as an effective recorded audio production tool.

ACKNOWLEDGMENTS

I am deeply grateful to my supervisor, Dr. Dermot Furlong, for all of his support and guidance over the past years. Without his incredible knowledge and encouragement, this thesis would not have been possible. Thank you for making me appreciate, even more, the art of sound recording.

I would also like to thank all of my colleagues and former colleagues at ESMAE-IPP, MMT-TCD, and the Portuguese audio industry, in particular Mário Azevedo, Gustavo Almeida, Octávio Inácio, Jonathan Nangle, Aengus Martin, Marcin Gorzel, Gavin Kearney, Eamon Doyle and Joaquim Ramos, for their help in discussing sound related concerns, and for listening and discussing their informal listening test experiences. Also, thanks to the Aalto Department of Signal Processing and Acoustics, Helsinki, in particular to Dr. Ville Pulkki for allowing me to work in their Multichannel Anechoic Chamber.

A special thank you to the Baltar Cassola Guitar Duo and Orquestra Filarmonia das Beiras in the person of Maestro Antóno Vassalo Lourenço, for the opportunity to perform comparative recording sessions, and to Duarte Ferreira, Luís Alves, António Pedro Santos, and José Prata for all the help and discussions in the setup of such recordings.

A special thank you to Dr. Angelo Farina for, the clarification and help in using his software for the objective measurements of Inter-Aural Cross Correlation, and to Dr. Bruno B. Castro for the help and insightful comments on the statistical analysis of the subjective listening test.

I am very grateful for the support of all my friends from both Portugal and Ireland. Clare and Gilles Roy (and sons and daughter), Frances Mitchell, Jonathan Nangle, Les Keye and, Cathy and Dan Hegarty (and sons), thank you for always making me feel at home.

Finally to my family which I love very much, without your support this wouldn't be possible.

I dedicate this thesis to my wife Andreia, my sons João and Francisco. Thank you for giving meaning and joy to my life...I LOVE you very very much. This thesis is also yours!!!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

1 INTRODUCTION

As audio engineering evolved, it was responsible for the definition and development of a variety of processing tools which sought to enhance the artistry involved in the creation of a listener experience. Most of these tools initially took the form of analogue electronic devices. However, the introduction of the greater flexibility offered by digital techniques in the latter part of the 20th century led to an ever-increasing palette of sonic processing techniques for audio recording and production. Part of this has involved the exploration of a variety of multi-channel recording formats, many of which have not met with much commercial acceptance, perhaps because of a failure to identify techniques for aesthetic manipulation such as were available for stereo techniques.

Even for the case of stereo recordings, there has been a lingering confusion concerning the use of different microphone techniques and their influence on the perceptual impression created for the final listener. As a consequence, it is not surprising to find that definition of effective microphone techniques for multi-channel format recording exhibits considerable ambiguity in terms of their influence on the final perceptual impression created.

Because of this, it was decided for this thesis to revisit stereo recording techniques to attempt to unravel the factors which contribute to the perceptual impression created for the stereo listener, before considering how multichannel recording techniques could be used to enhance listener aesthetic perception. In particular, the perceptual attribute referred to as auditory spaciousness was focused on, as it is the case that spaciousness has been identified as one of the important perceptual features contributing to the perception of and preference for concert hall acoustics (Barron, 1971; Schroeder, Gottlob, & Siebrasse, 1974; Barron & Marshall, 1981; Ando, 1985; Blauert & Lindemann, 1986; Potter, 1993). If it is accepted that spaciousness is important for concert hall listener perceptual judgments, then transmission, or artificial creation, of auditory spaciousness is surely worthy of attention for reconstructed sound fields.

The term spaciousness will be used in this study specifically in relation to the perception of the spatial extent of the performance environment, *i.e.* that the sound field gives the impression of a large and enveloping space, in which a sound source is being presented.

1.1 Background to the research

The principal approach to the recording and reconstruction of "auditory perspective" has its origins in the inspiring work of Alan Blumlein (1933) and Harvey Fletcher's team at Bell Laboratories (Steinberg & Snow, 1934) in the 1930's. Blumlein's engineering approach to the problem of capturing and representing localisation information is both comprehensive and effective and has, over the years, proved extremely useful in the specification of both microphone arrays and signal processing devices for control of localisation information. In essence, Blumlein detailed the system requirements for 2-channel presentation of audio such that the cues for correct localisation were adequately preserved in the stereo recording/reconstruction chain. Stereo recording and reconstruction is surprisingly effective in re-presenting auditory perspective, and while only truly effective in terms of localisation accuracy for the centrally placed listener, the sound field presented is generally regarded as being preferable to mono presentation even for the non-optimally placed listener. It is noteworthy, however, that Blumlein's work is mainly concerned with frontal presentation only. That is, his modelling of audio reconstruction assumes a centrally placed listener facing two symmetrically organized loudspeakers. Later attempts to extend Blumlein's techniques to lateral and rear representation of audio in quadraphonic systems, such that a listener could be surrounded by "stereophonic stages", were not well thought out and ultimately not successful. While the failure of quadraphonic systems is often explained in terms of commercial mistakes, or 4-2-4 matrixing difficulties (*i.e.* downwards compatibility with 2-channel stereo), the truth of the matter is that lateral (and to a lesser extent, rearward) virtual imaging is not well supported by intensity stereo panning. Physical source localisation is not without ambiguity as there are image localisation uncertainties which originate in the imaging cone of confusion deriving from the non-uniqueness of localization cues for any defined azimuth. Details of the experienced cone of confusion can be found, for example in Tobias (1972). Similarly, stereo virtual images can suffer from similar localisation uncertainty, which uncertainties are more exaggerated for the case of non-frontal stage image locations (Rumsey, 2001). Blumlein would have immediately recognized that limitation, as would anybody who had followed his approach and way of thinking about and devising stereophonic recordings. One later contributor who engaged fully with the spirit and detail of Blumlein's formulation of the "auditory perspective problem" was Michael Gerzon. Gerzon was an early critic of quadraphonic approaches to surround sound, and significantly contributed to the specification and design of Ambisonics, which has proved to be an effective solution to full surround sound presentation. Ambisonics is based on spherical harmonic analysis and synthesis of wavefronts and seeks to optimally regenerate (at least) the zeroth and first order spherical harmonic components which contribute to the pressure and velocity wavefront dimensions, thereby allowing accurate localisation of "auditory images" in 3D space. Despite its proven workability, Ambisonics has not been, and is not likely to be, a commercial success. The requirements of a symmetrically disposed array of matched loudspeakers and a dedicated decoder has not been well received by the marketplace, perhaps because of prior, bitter experience with quadraphonics. Significantly, Ambisonics shares with Blumlein stereo a requirement of a centrally placed listener for optimal localisation reconstruction. It can be shown, both theoretically and practically, that effective off-centre reconstruction of localisation information demands a large number (*e.g.* more than 16) of loudspeakers in the array (Gerzon, 1974a). This is not typically a practical option for the domestic listener. Another approach for off-centre effective reconstruction would be to use higher order Ambisonics (Gerzon, 1973; Solvang, 2008).

In that the Bell Laboratories' spaced microphone approach to 2-channel recording and reconstruction does not provide robust inter-aural time differences (ITD) cues, which are of fundamental importance for auditory localisation, it would seem not to be an effective reconstruction technique. Blumlein's coincident microphone technique does generate correct ITD cues for frontal stage locations, and hence its effectiveness as a generator of "auditory perspective", and its general acceptance as a stereo reconstruction technique. However, in practice, it is noteworthy that both spaced and coincident techniques are used. It is a commonplace that coincident reconstruction is regarded as a *"they are here"* technique in the sense that it presents accurate source localisation cues, while spaced reconstruction is regarded as a *"you are there"* approach where source localisation reconstruction accuracy is compensated for by the provision of an enhanced sense of the recording space, or room impression. This is because spaced microphone recordings generate an "airy" or spacious impression suggesting that the listener is experiencing the performance space, as compared to the more focused impression of coincident recordings, which brings more attention to the performers rather than the room in which they are performing. This distinction is perceptual by definition. That is, it makes reference to the general perceptual impression created by both, and is not limited to effective recording and reconstruction of localisation information. As a general observation, it can be said that there is more to "spatial hearing" than merely localisation, and that recording and reconstruction techniques should address the other dimensions involved for effective aesthetic reconstruction. Also, there have been several works where spatial reconstruction is made an integral part of the compositional effort, either for music or for cinema. In the case of cinema, spatial presentation of movie audio has been in the mainstream of filmmaking since the 1930's onwards (*e.g.* with films such as Fantasia) (Holman, 2008). In the case of music, the artistic approach to compose, record and present works with more than the 2-channel approach is currently a "boom industry", meaning that the artistic and aesthetic generation of sound fields is evolving toward new concepts.

The surround sound format which has been adopted as the commercial standard to date, *i.e.* 5.1, has not been approached in a principled manner. Most microphone techniques (*e.g.* ORTF Surround, 5100 Mobile Surround Microphone) (SCHOEPS Gmbh, 2015; DPA Microphones A/S, 2015) are merely speculative suggestions with no objective basis for acceptance as the norm. To date there have been few detailed studies of the transfer of perceptual information between recording and playback environments (but see (Furlong, 1989)). Given that spaced stereo microphones do not handle localisation information very well, but do manage to "synthetically" generate a room-like listening experience, the research here is put upon to try to identify how this is achieved, before then extending the results of any spaced stereo study to 5.1 recording and reconstruction.

1.2 Aims of the research

A first objective of the proposed work would be to establish *perceptually significant* metrics, based on those used in room acoustics assessment, by which spatial hearing criteria, and the associated "auditory impression", could be assessed for a 2-channel stereo, 3-channel stereo, or surround sound recording and reconstruction context. The second objective of the proposed study will be to comparatively assess the reconstruction effectiveness of a number of "surround microphone arrays" by examining spatial hearing measurements for a primary (*i.e.* auditorium or recording) space and those for a secondary (*i.e.* listening room) space, under various recording arrangements. The third objective is to develop a spaciousness processor which derives from standard 2 channel recording methods using perceptual reconstruction as a metric. Objective assessment will be carried out through comparative study of primary (recording) and secondary (listening) space measurements of the identified perceptually significant parameters.

It is clear from the literature (Rumsey, 2001; Rayburn, 2012) that there are no well-established microphone array recording techniques for 5.1 reconstructions. As is often the case in audio engineering, the main difficulty is in the specification of exactly what the objective should be. Both Blumlein stereo and Ambisonics adopted sound field localisation and directionality accuracy as their objective. However, given that it is recognized that this is just not a possibility with a 5.1 loudspeaker configuration, the aims of this thesis are redirected toward a more perceptual, rather than physical wavefront, formulation. It is therefore here stated that an effective 5.1 system recording/reconstruction system should do the following:

- Preserve the perceived "auditory space" of a primary (recording) environment under reconstruction in a secondary (listening) space.
- Facilitate the synthesis of sound fields which can be emotionally pleasant and uplifting.

Given that the loudspeaker array side of the reconstruction chains is pre-defined, attention should be focused on the signal captured in the primary space, and its subsequent processing.

Approaches to 5.1 microphone techniques include stereo pair and flanking microphones arrays, Bruck array, Double MS array, and OCT Surround array, amongst others. While it seems, informally, that such arrays can sometimes prove operationally satisfactory with respect to the "perceptual" design goal stated above, there is little if anything, available by way of objective assessment of the performance capabilities of these approaches. How well, comparatively, do they maintain a "perceived room impression"? How do changes in microphone details and array configuration influence the perceived impression of spaciousness? Can the perceived spaciousness of surround sound systems such as 5.1 (ITU-R BS.775-1, 1992-1994) be controlled, and how? What does stereo recording and reconstruction tell us about control of listeners' spaciousness? There are no empirical studies available to answer such questions, to the author's knowledge. Therefore, the first objective, as previously described, of the proposed work would be to identify perceptually significant metrics based on those used in room acoustic assessment, by which spatial hearing criteria could be assessed in a surround sound recording and reconstruction context.

Having assessed (in a controlled environment) a suitable set of measurements, studies of parametric variation in primary recording spaces were then carried out in order to investigate the effect of such parametric variation on the chosen spaciousness metric, IACC. Continuing with the objectives proposed for this study, a comparative assessment of the reconstruction effectiveness of a number of "surround microphone arrays" was undertaken by comparing spatial hearing measurements for a primary (*i.e.* auditorium) space with those for a secondary (*i.e.* listening room) space, under various recording arrangements. This comparative assessment facilitated objective evaluation of surround microphone array capabilities. The goal here is to arrive at a fuller understanding of how the details of microphone arrangement influence listener perception of spatial attributes, and how this may or may not influence the aesthetic and artistic perception of the recorded/reconstructed sound field. Note that this is not limited to simple preservation of all localisation information. Significantly, it is here recognized that such acoustic preservation is simply not a possibility for a 5.1 reconstruction system. Instead, the focus is on perceptual, aesthetic and artistic attributes, rather than physical reconstruction.

1.2.1 Why investigate?

While it has been the norm in surround sound recording to assume that "more channels imply more microphones", this, actually, is not necessarily the case. If the focus is on "perceptual, aesthetic and artistic attributes" rather than wavefronts, attention should be drawn to the fact that listeners only have two ears. The fact of the matter is that all spatial information is encoded in the two signals arriving at the eardrums, although localisation information is also provided monaurally by HRTF spectral profiles (Blauert, 1997). While human auditory reception is much different from microphone capture, it remains true that a stereo microphone system receives information from all directions – it is just the case that much of the directional information is lost because of the microphone directivity characteristics and the angle between capsules of the arrays typically used. An exception is, of course, that of binaural recording where directional information is at least preserved in the recorded signals, but lost under loudspeaker playback if cross-talk cancellation is not used.

In all cases – stereo and binaural – the possibility exists of developing effective 5.1 reconstruction of spatial features. The prospect is, therefore, that established stereo or binaural microphone systems could be used to generate well-defined frontal stage images, while also allowing control of the other perceptually related parameters, such as spaciousness and envelopment.

There are, however, many unknowns involved in the approach. How do different stereo (*i.e.* coincident, quasi-coincident, and spaced) techniques operate in the context of artificial spaciousness generation? How does the microphone directivity influence the end results? As referred to in the previous section, it is proposed to engage in a comparative study of objective measures in order to explore these, and similar, questions. Objective assessment will be carried out through comparative study of primary (recording) and secondary (listening) space measurements of what are identified as being perceptually significant features.

Multi-channel presentations are now much exploited for artistic presentation of audio material. Typically, a multi-channel presentation is used to generate a spatially varying artistic experience for the audience. A recording and reconstruction of such a spatial audio experience should seek to maintain or synthesize those characteristics which are essential for the artistic impression. A further aspect of the proposed study is, therefore, to examine the potential of recording and reconstruction for spatial audio artistry by examining and comparing spatial hearing aspects of original and recorded spatial audio presentations. The results from such a study could then contribute to the definition of synthesized environments for use in the control of binaural signals that would allow manipulation of the spaciousness experience to suit any musical style. While it is recognized that auditorium acoustics contribute to musical experience, Michael Forsyth has analytically developed such casual observations (1985). If acoustic features contribute to the definition of musical style, then the control of these features would be important for any synthesized music types such as any recorded/electronic music as might be presented using cloud computing rather than concert hall performence, for example. That is, a synthetic music environment processor could be defined which would allow spaciousness control to suit any music style.

1.3 Thesis structure

The thesis has been structured in the following manner:

In Chapter 2 an overview of the art of sound recording is presented. A chronological presentation is used as a means to explain how audio engineering has evolved, and how performances have had to accommodate artistically to the technology of sound recording. It is important to note that without technological developments the art of recording would not evolve, and vice-versa. That is, recording techniques are only delivered if the art demands it. An account of how listening aesthetically to reconstructed sound fields has evolved will be presented.

Chapter 3 attempts to synthetize all the issues relating to auditory spaciousness in rooms, and in artificially generated sound fields. A review of the literature is presented which examines the perception of auditory spaciousness in rooms (*i.e.* auditoria). This is then related to sound field reconstruction, and to how spaciousness is an important feature for reproduced sound, parallel to that experienced in concert halls.

In Chapter 4 a discussion is presented of some of the most relevant existing methods for measuring spaciousness. An operational explanation is provided, and the mathematical background of these measurements techniques is outlined. Despite the confusion that these techniques can provide with their results, their continuous development has allowed for more consistent and reliable assessments of spaciousness measures.

In Chapter 5 an overview of the technology and techniques involved in true stereophonic sound recording is discussed. Stereo and surround microphone techniques are presented in terms of how they have developed throughout the years. Most of the proposed techniques have been aimed towards the physical reconstruction of a primary room sound field in a secondary listener environment. Despite the importance of such approaches, this chapter will examine the concerns of improving stereophonic reproduction with respect to other perceptual aspects apart from localisation, which ultimately will provide a "better sounding recording".

Chapter 6 provides methodology details on the experiments undertaken which will verify that the measurement trends for Inter-Aural Cross Correlation (IACC) of the setups used, are location independent. This is achieved through the repetition of the IACC generation and recording process in different environments, followed by a comparison of the results achieved.

Chapter 7 discusses the next set of experiments, giving methodology details of the setups and microphone arrays used. The effects of recording format details on perceived spaciousness are investigated. Since the previous chapter has verified that the measurement technique used was location independent, further measurement tests were conducted for microphone array reconstruction effectiveness with respect to perceived impression of spaciousness in a single, convenient environment.

In Chapter 8 a number of stereo and surround microphone techniques are investigated with respect to perceived spaciousness. Processing techniques such as stereo shuffling and 5.1 up-mixing techniques were used in order to investigate the possible control of spaciousness in stereo and 5.1 reproduction systems. The development of a spaciousness processor is presented, as it was exploited in the investigations undertaken for this thesis. A listening test was conducted to assess the effectiveness of the spaciousness processor and to further validate the findings of the IACC results. The subjective analysis was then correlated with the IACC results.

The thesis is summarised in Chapter 9, where the main findings of the research are reported and discussed, conclusion drawn, and possibilities for further work presented.

1.4 Contribution to the field

The research undertaken for this thesis has resulted in a number of findings, some of which are novel. The investigation of the perceptual features that relate to auditory spaciousness has been studied from the viewpoint of architectural acoustics, and how the findings from such research could relate to the perceptual impact of enhanced spaciousness on recording and reproduction format detail choice. Also, the study is novel in that it is concerned with the perceptual aspects at the listener's "side", and not just with the physical reconstruction of sound fields. The comparative study undertaken used physical measurements that relate to the perceived impression of spaciousness as an index for assessing how changes in physical attributes of the microphone arrays could influence the perceptual impression of spaciousness.

Michael Gerzon was an extraordinary audio engineer, and one who was perceptually oriented in his audio engineering developments. His design of stereo enhancement systems go from extensions of Alan Blumlein's original 2-channel stereo work, to multi-channel systems developments, all offering greater control possibilities to the music producer for the reconstructed sound field listener experience. The spaciousness study here, and ultimately the proposed processor, develops from the work of Gerzon, and facilitates independent user control of listener hearing features, including sound source localisation and spatial impression. Therein lies the originality of the work presented in this thesis. Music producers can employ technology to alter the spatial experience of the reconstructed sound listener, and create *art*, *i.e.* no production technique is entirely benign.

2 SOUND RECORDING: *AN ART*

2.1 Introduction

From the early days of sound recording - going back to Edison - it is possible to consider the technological achievement of sound field reconstruction as an *art*. During the period of recording using wax cylinders, a huge effort was made to position the performer in a proper manner, and to control the dynamics of the performance so that the diaphragm and stylus of the gramophone could mechanically register the sound waves into grooves on the wax. Not only did the performer accommodate to this "new" performance practice which was completely different from a performance, say, in a concert hall, but also the rooms in which these recordings were made were altered to accommodate for a dry acoustic that would ultimately help the recording process (Toole, 2008, pp. 13, 14). The adaptation to these new conditions made for changes in the performance of music, thereby contributing to the art of the music while at the same time delivering a new form of art (*i.e.* sound recordings) embedded in the medium in which it was delivered. Sterne argues "that sound reproduction is *always already* a kind of studio art" (Sterne, 2003, p. 223). The sound recording efforts made today are not that far removed from those of the early days; it is the "evolution" of the technology and recording techniques that accounts for changes in the aesthetic appreciation of the artistic performance which is associated with the *art* of sound recording.

Sound recording technology has evolved despite the repeated claim that the technological achievements of the equipment needed for capturing and delivering sound was already at the pinnacle of its achievement. If we look at advertisements from some of today's manufacturers of audio equipment it can be noticed that marketing lines such as "hear the truth with great sound" (www.jbl-com), "true-to-the original" (www.bostonacoustics.com), "realistic multichannel surround sound" (www.dolby.com) are not so different from advertisements of the Victor Talking Machine Co. from the 1920's where one could read "the human voice *is* human on the New Orthophonic". It has been the case that audio manufacturers have repeatedly claimed to have reached sound recording perfection, not only by comparing to rival companies, but also compared to their own line of products. This persistent pursuit of sound recording perfection has also been the case for the techniques used for recording the sounds. For example, sound engineers often claim that the use of certain microphone techniques will necessarily be better than others (Gerzon, 1971; Griesinger, 1985; Lipshitz, 1986).

Pursuing the best possible sound in recordings has not just been the *raison d'être* of the sound technicians and audio manufacturers, but also that of the performers who demanded a "perfect" sound in the recordings they made. The artistic achievements in the recording process should be as demanding as those of a live performance. It is worth pointing out that music recording has not generally evolved toward more-and-more exactness of physical reconstruction. Rather, it has developed toward facilitating more-and-more control over reconstructed sound field features (Sterne, 2003, p. 242; Swedien, 2009), the pursuit of which leads to the eternal question "how can it be made to sound better?"

2.2 From "dead" to "live"

In the time of Edison, sound recordings were conducted in very controlled and acoustically "dead" rooms (*i.e.* with reflections reduced as much as possible). The fact that the presentation of reverberant sound over monophonic reproduction (see Chapter 5) sounds very muddy and overemphasised (*i.e.* less tolerant to reverberation levels) (Streicher & Everest, 2006), the need to control environmental noise was a contributing factor for the choice of such recording rooms. Also, the very insensitive recording systems, such as the early phonographs, made it so that the performer needed to be as close as possible to the microphone, resulting in recordings where the direct sound was dominant. Direct sound here is to be understood as the first arriving sound wave from a source to the ears of a listener or a microphone, travelling in a direct path without being reflected from any surface (Everest & Pohlmann, 2009). This technical and artistic approach was the norm and practice in the early days of sound recording. Although listeners, as will be discussed in following chapters of this thesis, enjoy listening to music in "good environments" (*i.e.* good acoustics), which contribute to making an uplifting sound for musical performances, it seemed that these acoustically "dead" recordings nevertheless created a pleasant illusion. The fact that this sonic illusion worked can be explained by the fact that the overall recording system was able to reproduce/communicate meaning and emotion. Toole explains that listeners have at some point felt "that tingling sensation that tells us we are experiencing something special and emotionally moving. What is 'real'? Was it 'reproduction'? Good sound or bad? Does it matter? The fact that these feelings happen confirms that the [overall recording] system works." (Toole, 2008, p. 5). However, if any sound recording can contribute towards an emotional feeling, will a more spatially complex sound recording contribute further in evoking a greater depth of feeling? Studies have shown that a performance in an acoustic environment has a more preferred impact than the same music played in a "dry" one. Listener preference has been found to relate more to a lively, spacious sounding performance environment (Ando, 1985). Consequently, there is an argument to be made, one that will be explored further in this thesis, that adding a spacious quality to sound recordings might lead to greater preference for the reconstructed sound fields that result. The audio industry has developed more and more tools and proposed more and more techniques for delivering complex sound field reconstructions which helps answer the question "how can it be made to sound better?". Toole (2008) proposes the following explanation of reconstructed sound field impact manipulation as follows: "by understanding the perceptual dimensions and the technical parameters that give control over them [sound recordings], it may be possible to give the artists tools that allow them to move into new creative areas by expanding the artistic palette." (p. 5)

The sound processing tools which exist today are numerous: from controlling dynamics, to filtering and modulating sound, from pure correction to the (re)-creation of new sounds, and from monophonic to stereophonic reconstruction where direction and spatial attributes of the sound field can be controlled and manipulated. Spaciousness, which will be discussed in the following chapters, is one of the perceptual parameters that can be controlled and which has been found to contribute to a heightened appreciation of the *art* of sound recordings. In a similar manner to Toole (2008), Read and Welsh in their book "From Tin Foil to Stereo" (1959) recount the statement written in 1951 by Edward Tatnall Candy in his "Saturday Review of Recordings":

> "Liveness," the compound effect of multiple room reflections upon played music, is—if you wish—a distortion of "pure" music; but it happens to be a distortion essential to naturalness of sound. Without it, music is most graphically described as "dead." Liveness fertilizes musical performance, seasons and blends and rounds out the sound, assembles the raw materials of overtone and fundamental into that somewhat blurred and softened actuality that is normal, in its varying degrees, for all music. Disastrous experiments in "cleaning up" music by removing the all-essential blur long since proved to most recording engineering that musicians do like their music muddied up with itself, reflected. Today recording companies go to extraordinary lengths to acquire studios, churches and auditoriums (not to mention an assortment of artificial, after-the-recording liveness makers) in order to package that illusively perfect liveness. (p. 378)

This statement helps draw attention to the fact that not all which seems measurably correct will be appreciated as art, or in the case of the theme here, as a good sounding recording. It is important, therefore, to understand the auditory features which are "missing" in recorded and reproduced sound such as "Liveness", a term that has been extended to the science of auditory spaciousness (Streicher & Everest, 2006, p. 12.1), and to provide the technological means to deliver such features.

2.3 Creating and recreating sonic "illusions"

Since the "birth" of stereo in the 1930's, the spatial experiences which could be conveyed in sound recordings have contributed towards a better sonic experience. Despite the fact that critics of stereo sound claimed that there was no need to have 2-channels since mono was capable of delivering a guaranteed impression of the recorded performance, stereo was enabled to develop with the persistence of but a few audio technicians, researchers and artists who were enthused by the capabilities of the stereo system. Swedien quotes one major recording label executive as saying "stereo is like taking a shower with two shower heads – and *you* wouldn't take a shower with two shower heads, would you? Ha! Ha! Ha!" (Swedien, 2009, p. 39). Such comment demonstrates the lack of vision for the potential that stereo could provide a more uplifting experience similar to that experienced in concert halls and also allow for a "sonic fantasy" where new "stereo spaces" could be created and new emotions experienced. Eighty three years after its introduction, stereo is still one of the most used recording formats, while, in the author's opinion, its potential has not yet been fully exploited (Lipshitz, 1986; Swedien, 2009).

Localisation or placement of sound objects in the stereo field is of great importance when creating sonic illusions. The fact that sounds can be perceived across apparent left to right locations in a stereo sound field is a marked improvement over that of monophonic sound reconstruction. But the sonic illusion is not only about accurate localisation. Bruce Swedien comments "...that really good stereo music reproduction was not merely one sound source coming out of one speaker and a different sound coming out of the other speaker", in fact, for him, the feeling of music can be reproduced "more emotionally by using stereo recording technique" (Swedien, 2009). It can be appreciated that if perceptual features of an acoustical environment, such as spaciousness, are conveyed within the sonic illusion, listeners are likely to enjoy the experience a lot more. Griesinger has stated that spaciousness is as important to sound recordings as it is in concert halls, and one of the major duties of a sound recordist should be to create the impression of spaciousness (Griesinger, 1985). In summary, auditory spaciousness (to be discussed further in Chapter 3) is the perceptual impression of sound in an enclosed space. The addition of early reflected sound to a discrete, direct sound source (*e.g.* a musical instrument) at the ears of a listener will create a sonic impression of a space which will differ according to the strength and details of the reflected sound (Barron, 1971).

In order to create and recreate illusions with sound recordings it is important that the features required for the sonic illusion to take place are fully understood. Accurate imaging, good sense of space, tonal quality and instrumental balance are but a few of the features which are important in a good sound recording. How can these features be controlled in a recording?

Should the recording space and the microphone techniques used during the production provide these features? Is it possible to artificially create these features? These questions and others have emerged since the beginning of the *art* of sound recording, and although over the years some solutions have been presented/suggested, many question remain, with yet more questions/further research following on from the answers provided.

2.4 Single point "micing" vs. "multi-micing"

Recording techniques have evolved in more or less two different styles. The first is the single point recording technique, which utilizes an array of microphones positioned at a spot in the room that aims at the sound source. Here the idea is to capture the sonic properties of the performance including the acoustics of the space in a somewhat minimal, almost "purist", approach. The second approach is to use a plethora of microphones which might include, or not, a main array and several accent microphones that aim at particular instruments or sets of instruments. The feeds from these microphones are then mixed in a (re)-creative fashion, where either a natural approach to the original sound stage or a "new" sonic stage might be the result. The two approaches discussed here relate to the recording of classical ensembles (*e.g.* orchestras and chamber ensembles) performing scored music. Although these different styles of recording techniques might also be used in the recording of jazz, pop and rock music, "multi-micing" is the preferred technique among sound engineers and producers of these styles of music (Bartlett & Bartlett, 1999).

Throughout the years, recording techniques have been debated and defended by those who prefer the single point and those who prefer "multi-micing". Those who defend the latter approach, as discussed by Gerzon, are more or less "objective", while those who defend single point arrays tend to make their claims on a more "subjective" analysis of the results being more "realistic " and "pleasant" (1971). The fact is that assessing the results using a purely objective analysis cannot solely be done, since the results from the recordings depend on the desired musical effect. If it is accepted that stereo is incapable of reproducing a realistic (*i.e.* close to the real performance in the acoustic space) sound stage, then it will always be a subjective judgment when judging one technique as being more realistic than the other (Gerzon, 1971).

What is important here is that the different features of the recorded sound are delivered to the listeners' ears so that a required feeling and emotion can be appreciated.

Is it then possible to deliver all perceptually significant aspects of a sound field from a primary room to a secondary listening environment, using either recording technique? If the perceptual feature of auditory spaciousness is examined, it is possible to discuss whether a recording is producing an appropriate feeling of spaciousness or not (Furlong, 1989), and this discussion is made without, sometimes, any knowledge of it (*i.e.* the recording) being a single point or a "multi-miced" recording. Some experiments made by Griesinger (1985), and later by Gerzon (1986), have rediscovered and provided insights into how it is possible to manipulate recordings produced with either technique, previously discussed, so that the perceived auditory spaciousness can be changed.

2.5 Summary

Despite the technical achievements in sound recording, there is still and always will be an artistic dimension involved in capturing sound. Audio equipment and techniques developed for use in the process of sound recording are only relevant if the necessary perceptual features are considered. The persistent question of "how can it be made to sound better" is motivated by the notion that sound recordings can communicate emotions and perceptual impressions. Therefore, it is important that research undertaken in the field of sound recording should also be focused on the control of the perceptual effect of the reconstructed sound field, and not merely on the physical effectiveness of the reconstruction. That is to say, physical sound field reconstruction should be approached using perceptually significant physical features.

The following chapters will examine one of the perceptual features of sound experience *i.e.* auditory spaciousness, which has been extensively studied in concert hall acoustics. Listeners in either concert halls or in reproduced sound fields appreciate this feature. It is therefore deemed to be worthy of detailed study.

3 SPACIOUSNESS

3.1 Introduction

According to the Merriam-Webster dictionary Spacious means:

1. Vast or ample in extent;
2. Large or magnificent in scale;

From these meanings, above, it can quickly be understood that the word spacious relates to sensory impressions that might seem spatially extended or large to observers. This is also true in relation to *auditory events* that perceptually sound big. When listening inside a room which is large, is not the experienced sound field heard as being sonically extended, or in other words spacious? Humans have the capability of interpreting sound fields in enclosed spaces, and appear to be very keen on doing so, even if this takes place at a subconscious level. But how are sound fields interpreted? What are the factors that contribute toward such expressions, (drawn from anecdotal evidence), as: "...this concert hall sounds very spacious", "...the cello has a spacious sound that fills the room", "...there's more space in the sound of this recording, when compared to that recording".

The study of auditory spaciousness has been thoroughly dealt with in concert hall research, which has made a great contribution to the better understanding of auditory perception of concert hall sound fields. Researchers such as Schroeder *et al.* (1974), Barron and Marshall (1981), Ando (1985) and Blauert and Lindeman (1986) have identified the auditory components which determine listeners' appreciation of the quality of concert hall acoustics, spaciousness being strongly correlated with the positive judgment of good concert halls acoustics. More recently, sound recording research has investigated the effects of spaciousness in either sound reproduction systems or stereo recordings, where the contribution from concert hall acoustics can be taken as a primary reference (Kendall, 1995; Tohyama, Suzuki, & Ando, 1995; Toole, 2008).

The objective study of concert hall acoustics has developed immensely since the pioneering work of W.C. Sabine in the early 20th century (Sabine, 1923). Reverberation time (RT) was the first objective parameter that allowed for the acoustical characterization of halls and rooms. Nevertheless, this parameter, which dictates the time a sound takes to decay by 60dB in a room once a sound source has been stopped, is but one of many parameters in concert hall acoustics. Over the last 100 years, there has been much concert hall acoustics research undertaken, all of which aimed at a better understanding of the subjective experience of halls for music, performing arts and cinema theatres. It has been found that there is more to judgments of "good acoustics" of concert halls than just optimal RT (Blauert & Lindemann, 1986). If two halls with identical RT are considered, a preference distinction between each one of them can still be made. In trying to address these kinds of issues, acousticians have engaged in time-consuming experiments so as to identify the perceptual subjective preferences for concert hall acoustics. One feature that has attracted particular interest has been *auditory spatial impression* (Potter, 1993). This feature is related to spaciousness, and can be described as the *sensation of space* (*i.e.* impression of an extensive sound field) a listener might have when exposed to a sound field which definition is not far from the meaning of the word spaciousness, previously identified. What is of great importance is that this *sensation* is a perceived effect, and it relates to a subjective understanding of the sound stimulus and the environment in which this stimulus originated.

3.2 Auditory Space Impression

There has been much discussion about what is the best term to describe the perception of auditory space. One problem is for a listener to try and verbalize their subjective impression of a sound field which they had been exposed to, and another problem is that different listeners might use different terms to characterize aspects of a sound field which are in fact similar. Terms like *spatial impression, ambience, apparent source width, feeling enveloped*, and *spaciousness* might all refer to the same, or different, characteristics which listeners are trying to express when characterizing either a concert hall or a reproduced sound field listening experience (Griesinger, 1985; Potter, 1993). Listeners' expectations and the context of their listening experience might be the cause of the verbal ambiguity which complicates the effort to improve sound fields in relation to listener experience. However, there is a general opinion among acousticians that "good acoustics" is strongly related to the overall spatial impression of a room (Schroeder, Gottlob, & Siebrasse, 1974).

Schroeder *et al.* (1974) identified a subjective preference that related to an objective measure which was nearly uncorrelated to reverberation time. In their findings, this subjective impression is described as follows: "...might be mediated by a more pronounced feeling – of being immersed in the sound...". Blauert and Lindemann (1986), Reichardt and Lehmann (1978) and Kuhl (1978) suggested that auditory spatial impression relates to listeners' exposure to appropriate sound fields which may occur in spaces that present listeners with surrounding reflected sound components. These reflections somehow give listeners an idea of the type and size of an actual or simulated space. Simulation here is of importance, as will be appreciated later. Spatial impression, according to these authors, is based on different perceptual attributes of the auditory events, including two of the primary attributes: reverberance and auditory spaciousness. Reverberance is the temporal slurring of auditory events that result from late arriving reflected energy from the walls of the enclosed space. Auditory spaciousness relates to the spreading of the auditory events, this effect being mainly caused by early lateral reflections. Spatial impression has also been described as the subjective sensation associated with early lateral reflections (Barron, 1971; Barron & Marshall, 1981). Prior to this, Marshall (1967) had already discussed the significance of early lateral reflections as a means to obtain the desired spatial effect, and therefore a greater subjective preference when exposed to such spatial effect. Other authors (Morimoto & Maekawa, 1989; Beranek, 1996; Potter, 1993; Bradley & Soulodre, 1995a) have suggested the division of spatial impression into three components: spaciousness, size impression (*i.e.* the notion of the type and size of a space) and reverberance. Spaciousness is then divided into apparent source width (ASW) and listener envelopment (LEV). ASW is described as the width a sound source might be perceived as having when performed in a concert hall, and LEV is described as the perceptual feeling of being enveloped by the sound field which is related to late arriving reflections. However, this is not related to the perception of reverberance. Contradicting this, Griesinger (1999) has suggested that the association between spaciousness and ASW should be abandoned, and that spaciousness and envelopment are synonymous. Griesinger (1996), also stated that it is the sound field which gives the impression of a large and enveloping space, *i.e.* "the sound field of an oboe can be spacious, but an oboe cannot".

So it seems that spatial impression may be caused by the fact that sound reaches listeners from all the surrounding boundaries after the direct sound is triggered and reflects from these boundaries. It is noteworthy that humans do not hear these components (*i.e.* direct sound and reflections from the walls) as being discrete elements. Instead, these are grouped into an overall spatial impression which relates to both the perception of spatial aspects of the sound source, and of the enclosed space (Bregman A. S., 1990; Blauert, 1997).

Although there has been thorough investigation over the years, the subjective impression of concert hall acoustics, and the terms used to describe these impressions, have not been uniformly systemized. A persistent ambiguity exists in terminology relating to the perception of spatial effects. Spaciousness or Spatial Impression might be used as a single descriptor for such perception. However, spaciousness is the perception of being surrounded by a large and enveloping space (Toole, 2008). It is therefore important to clarify that for this study spaciousness was to be dealt with in accordance with the perception of the spatial extent of the performance environment, *i.e.* that the sound field gives the impression of a large and enveloping space, in which a sound source is being presented.

3.2.1 Spatial impression in reproduced sound

In the previous section a description of spatial impression in concert hall acoustics was given. The aspects of spatial impression for reproduced sound will now be looked at. The concepts of ASW, image broadening, early spatial impression, spaciousness, and envelopment evolved within the concert hall context. The challenge of the work undertaken in this study relies on translating these descriptors into the context of reproduced sound (Toole, 2008).

The perception of spatial impression for reproduced sound may differ from the spatial impression experienced in concert halls. Contemporary surround sound systems claim to provide an impression of listeners being enveloped and surrounded by sound. But, a more general issue arises when dealing with reproduced sound in relation to the following questions: is it required that the spatial impression created by the reconstruction system be an accurate and realistic reconstruction of the spatial impression experienced in a performance space? Or is it more that the intention is to allow the creation of a spatial "illusion" and consequently to generate a different/new spatial impression which is not an accurate physical reconstruction of a primary, recording space?

There has been some investigation relating to spatial attributes of reproduced sound, some of which has been summarized in (Rumsey, 2001) and more recently has been systemized in (Toole, 2008). Following Rumsey, Nakayama *et al.* (1971) is one of the few examples of spatial subjective quality tests done for multichannel surround reproduction. The study was on the subjective effects of 1-8 channel reproductions in an anechoic chamber using recordings made from unidirectional microphone capture of performances in a concert hall. The arrangements of the microphone arrays were identical to the layout of the loudspeakers setups. These arrays were placed at three different distances from the orchestra. Other microphone setups were used such as an MS pair (see Section 5.3.1) in conjunction with a multichannel mix of close spot microphones.

The subjective assessment was based on two different approaches in which 13 different loudspeaker arrangements ranging from 1 to 8 channels were used to present recordings to listeners. The first approach was to present a single sound stimulus, where listeners made a preference judgement on a seven-point scale, ranging from "very good" to "very bad". The second approach was to use a comparison between paired sound stimuli, in which listeners were asked to judge the similarity between stimuli, also on a seven-point scale, ranging from "just the same" to "quite different". From the results a distance scale for preference was constructed and the similarity results were converted to similarity distances between all combinations. All of this data was then subjected to multidimensional analysis.

Important subjective factors such as *depth of sources, fullness* and *clearness* were interpreted from the results. An examination of their results suggests that *fullness* is very similar to what others have called *envelopment,* as it is strongly noticeable for loudspeakers set at the sides and rear of the listener, yet weak for two-channel stereo. *Depth of sources,* following Rumsey's interpretation of the author's results, seems to be related to *nearness* or *closeness* of sources. It changed greatly as the recording position of the microphones was moved closer to the orchestra. *Clearness* was found to be related to the measured concert hall acoustics parameter D50 (Definition) (*i.e.* it compares the sound energy arriving in the first 50ms with later arriving energy), which is clearly an indication of the direct to reverberant sound ratio (Rumsey, 2001).

From the work done by Nakayama *et al.* an equation was formulated that related the quality ratings of listeners to the three attributes, by weighting the factors appropriately. Their equation suggests that *fullness* was weighted most strongly, followed by *depth of sources*, with *clearness* being the least weighted attribute.

It is obvious that the perception of spatial impression in concert halls has a parallel in reconstructed sound fields. Listeners prefer an extended sense of space both in concert halls, and in stereophonic reproductions if more speakers-channels (*e.g.* lateral and rear speakers) are added to the basic stereophonic reproduction arrangement (Tohyama & Suzuki, 1989). If these speakers are fed with signals that are capable of providing differences in the ear signals, it is possible to induce illusions for listeners that suggest a more spacious sound. The fact is that providing lateral sound sources will create differences between the left and right ear signals which is the determining point in the perception of spaciousness (Blauert, 1997).

3.3 Effects of reflections

The investigation of reflections and how they affect the perception of a sound field has been ongoing for a few decades now. One of the first standard references, according to Toole (2008), is Haas' PhD thesis from 1949 and later translated from German to English in 1972 (Haas, 1972). His core experiment dealt with the investigation of the perceptual effect of a single reflection added to a direct sound.

In a semi-anechoic room, two speakers were used which, according to Haas (1972), were positioned at ± 45° to the left and right side of a listener. A speech signal, previously recorded, was sent to both speakers, and a delay could be introduced to the signal sent to one of the speakers. While leaving the sound level of both speakers equal, listeners were asked to judge the location of the sound source as a delay in one of the speakers was introduced. If no delay was applied, listeners judged the speech to come from halfway between the speakers (*i.e.* creating a centrally placed phantom image). As the delay was varied from 0 to 1ms, listeners judged the location of the speech to be shifting towards the earlier speaker, as if the source moved from the centre towards one-side. This is called summing localisation, and is the basis of phantom source imaging in stereo reproduction, assuming a listener is in a position where the speakers are equidistant from the listener (*i.e.* in the "sweet spot") (Blauert, 1997). However, if the delay was increased beyond 1ms up to about 35ms, then in most cases the source appeared to radiate from the earlier speaker, although this source localisation tendency can also be influenced by the nature of the source signal. The precedence effect has a number of parameters which influence its manifestation, but in general it can be simplified to saying that the first-arriving wavefront will dictate the perceived source location. It should be noted that, for Haas's study, this 35ms limit is for the case of speech, for equal level signals to both loudspeakers. When the 35ms limit is passed, listeners start judging two distinct sound sources, one following the other (*i.e.* an echo is perceived). Haas also tested to see, for each delay, how loud the delayed signal could be made relative to the non-delayed one, before the delayed signal was perceived as the source location. It was concluded that the delayed sound needed to be 10 dB higher if the delayed source was to be perceived as being a source location. This was described as an "echo suppression effect".

Haas also observed perceptual auditory effects that had nothing to do with localisation. He was able to determine that the addition of a second sound source, with a short delay, increased loudness, and that there were some changes to sound quality which he termed "liveliness" and "body" (Haas, 1972, p. 150), and a "pleasant broadening of the primary sound source" (p. 159) (Toole, 2008, p. 76).

Subsequent to Haas' work, Ando and Kageyama (1977) conducted subjective preference tests with a simulated single reflection in an anechoic chamber in order to learn of the preferred properties of sound fields. From their investigation they were able to plot the percentage of subjects who preferred the simulated sound, in this case speech, with a single echo as a function of the direction of the reflection. Following this, Ando correlated the subjective preference of early lateral reflections with the objective parameter of Inter-Aural Cross Correlation (IACC), as can be seen in Figure 8 of his paper (Ando, 1977, p. 1440), and he was able to determine that the magnitude of the IACC is almost independent of source signals used, including noise sources (Ando, 1977, p. 1437). The objective measures which relate to subjective preferences of spaciousness are discussed further in Chapter 4. Further experiments conducted by Ando and Gottlob investigated the effects of multiple reflections on preference tests of sound fields and concluded that a single reflection gave nearly the same results as were obtained with multiple reflections (Ando & Gottlob, 1979). Blauert (1997, p. 355) also states that spaciousness is clearly not significantly affected whether one reflection or multiple reflections were used.

Barron (1971) had already discussed the subjective effects of first reflections and their importance for good acoustics in concert halls. Figure 3-1 summarises his findings, where the range of levels and delays for a desirable effect of spaciousness in connection with musical signals, is shown.

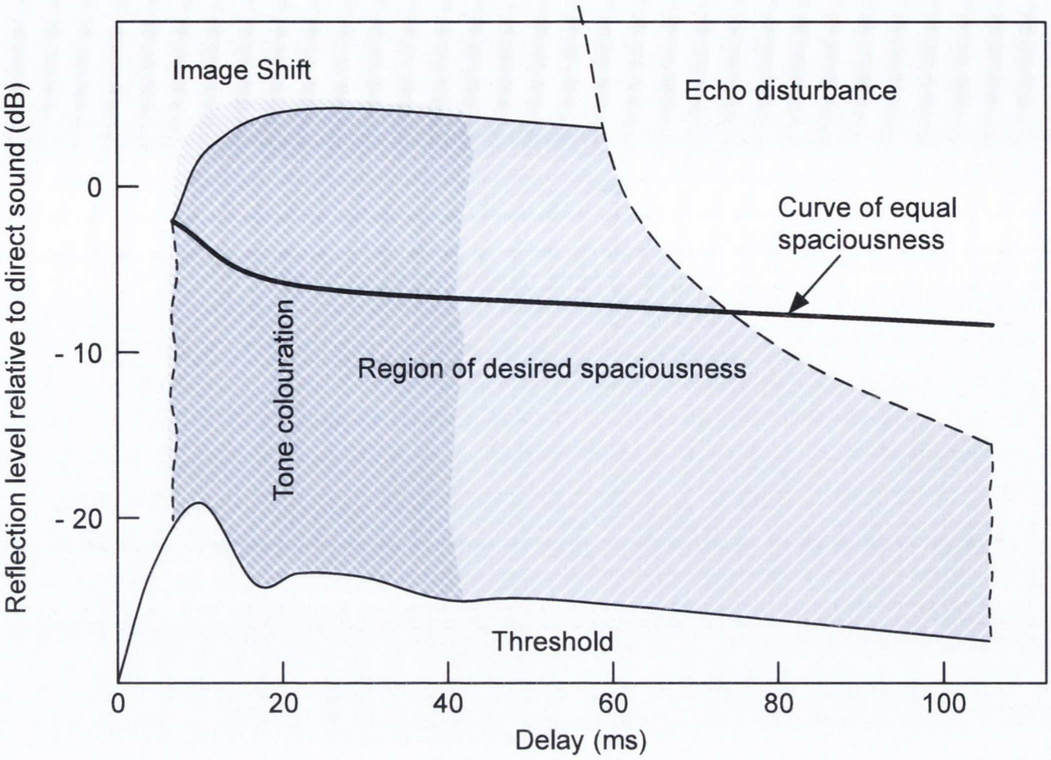


Figure 3-1: The subjective effects for music of a single side reflection (azimuth angle 40°) as a function of reflection level and delay relative to the direct sound (adapted from (Barron, 1971))

For Barron's experiment, only one early lateral reflection was used. The sensation of spaciousness occurs immediately above the masking threshold when the reflection delay exceeds about 5ms, and becomes stronger as the amplitude of the reflection is increased. However, if the reflection amplitude is increased such that it is made stronger than the direct sound, image shifts will occur. It can also be observed that the delay of the reflection does not significantly change the desired spaciousness effect. Beyond a reflection time of about 50ms, echo disturbances may occur if the reflection is strong enough. In Figure 3-1 a curve of equal spaciousness can also be seen as was evaluated by paired comparison using a reflection of 40ms delay and -6dB amplitude as a reference condition, in which for shorter delays the level of the reflection needs to be higher to produce the same perceived effect. A tonal coloration is also observed for shorter delay times.

Barron and Marshall (1981) conducted several physical measurements where artificially created early lateral reflections that changed in level and time of arrival with respect to the direct sound, were used. These reflections were created by means of loudspeakers positioned at an angle to the right and left of the loudspeaker reproducing the direct sound (with measurements being made in an anechoic chamber). The measurements were to evaluate the subjective effects introduced by early lateral reflected sound. Barron and Marshall's results showed that as the level of the reflection was increased, the listening test subjects noticed an increase in the perceived spaciousness. Their results also showed (confirming the previous results by Barron (1971)) that the change in delay of the reflection did not result in any significant change in the impression of spaciousness (see Figure 3-1). Barron and Marshall's studies engaged in subjective testing and assessment of the impression of spaciousness due to early lateral reflections, and used the physical measurement of Lateral Energy Fraction (LEF). However, such testing can effectively be made with the use of IACC peak values, since it has been well established by Okano *et al.* (Okano, Beranek, & Hidaka, 1998) and by others that the normalized peak value of IACC is an indication of the impression of spaciousness, where a value of 1 means a strong correlation between the ear signals of a dummy-head, hence meaning "NO" spaciousness, while a value of 0 means no correlation between the ear signals, hence indicating a high level of spaciousness. Analysing the signals that are received at the ears of a dummy-head permits objective assessment of spaciousness. Using IACC, it is possible to determine the degree of similarity of the signals at the ears. The available evidence would therefore support the idea that it is the differences between signals that arrive at listeners' ears that are of greatest importance for the perception of spaciousness.

The experimentation setup adopted for the studies undertaken for this thesis was based on the approach used by Barron (1971). However, as will be demonstrated, the first goal was to verify if the setup and experimentation could be justifiably made in any of the different rooms available. This issue will be discussed in Chapter 6.

It is worth noting that Ando's experiments were conducted in anechoic chambers, and that simulation of discrete reflections was added to a direct sound. The subjects' judgments of these simulated sound-fields was of a preference for the presence of early lateral reflections which gave an enhanced impression of the listener being in a three dimensional space.

3.4 Summary

It has been established that spaciousness is strongly appreciated as being a positive and desirable component in concert hall acoustics; this has been confirmed by extensive studies.

Strong early lateral reflections added to the direct sound are preferred by listeners and contribute to the perceptual sensation of spaciousness, a feature which is associated with listener inter-aural differences. It is certain that it is the reflections that increase listener "preference" when judging sound fields and when describing what led to that preference, which is commonly related to various spatial effects such as apparent lateral spread, broadening, of the sound source, or to the impression of being immersed in a large reflective space (Toole, 2008).

Spatial impression for sound reproduction systems seems to have a parallel to that experienced in concert halls. Listeners enjoy feeling surrounded by sound, which surround experience leads to a more spacious sonic impression. In order for sound reproduction systems to contribute to this effect, laterally placed speakers have been found to be important.

Although there is a lot of debate on the terminology that best describes the perceptual auditory sensations related to early and late reflections, the term spaciousness will be used in this study specifically in relation to the perception of the spatial extent of the performance environment, *i.e.* that the sound field gives the impression of a large and enveloping space, in which a sound source is being presented.

4 OBJECTIVE MEASURMENTS RELATING TO SPACIOUSNESS

4.1 Introduction

The following sections will discuss some of the existing techniques used in concert hall acoustics to objectively measure spaciousness. More attention will be given to Inter-Aural Cross-Correlation (IACC) as it is the measurement used for the work undertaken for this thesis.

4.2 Brief history

Most of the effort made to understand spaciousness has been undertaken in relation to concert hall research by acousticians seeking to understand this perceptual impression in order to improve the design of concert halls. Acousticians such as Beranek, Barron, Marshall, Schroeder, Ando and Blauert are but a few who have dealt with trying to understand how changes in concert hall acoustical parameters, such as reflection delay and amplitude, could affect the perceptual impression of spaciousness. As an overview which considers objective versus subjective attributes in concert hall acoustics and auditory perception, this thesis refers to Beranek's examination of the known acoustical attributes of concert hall acoustics (1996) to Blesser and Salter's on aural architecture (2007), and also to Blauert's (1997, pp. 347-358).

In 1967, Marshall investigated the importance of early lateral reflections for the creation of a subjective effect that he termed "spatial responsiveness", and which he defined as envelopment of the listener (Marshall, 1967). Barron and Marshall later revisited the early theories of Marshall and examined them in more detail (Barron & Marshall, 1981). From their work on the investigation of early lateral reflections, it was possible to better understand how these reflections relate to the perception of apparent source width and envelopment. They proposed measuring what they called "spatial impression" (see Chapter 3) by using the ratio of the lateral sound energy to the total energy, *i.e.* the lateral energy fraction (LF). This measurement is given in Equation 4.1 (Barron & Marshall, 1981).

$$L_f = \sum_{t=5ms}^{80ms} r\cos\phi \Bigg/ \sum_{t=0ms}^{80ms} r$$

Equation 4.1: Lateral energy fraction, L_f

In the above equation r is the sound intensity, and $\emptyset$ is the azimuthal angle of incidence measured from centre front on the lateral plane. To measure L_f an impulse response of the room is needed. The term $rcos\emptyset$ in Equation 4.1 is approximately a figure-of-eight microphone signal where its null axis is pointing towards the sound source, and r is, in practice, an omnidirectional microphone signal. LF can therefore be easily measured using a coincident setup of figure-of-eight (pointing towards the sides) and omnidirectional microphones.

Schroeder *et al.* (1974) also investigated audience subjective preference for concert hall listening, and how this correlated with objective measurements of geometric and acoustical parameters. One of the objective parameters highlighted was *C* which is the inter-aural "coherence", *C,* (the maximum of the cross-correlation function between the impulse responses at the two ears). Later, Ando (1977) investigated the subjective preference of a single echo added to sound fields, and how the degree of subjective preference was related to the objectively measured parameters of long-time autocorrelation function (ACF) of the sound source, and to the listener IACC. The work undertaken by Ando on the subjective preference for music sound fields has been discussed previously in Chapter 3. The normalized IACC can be calculated using Equation 4.2 (adapted from Peltonen (2000))

$$IACF_{t_1,t_2}(\tau) = \frac{\int_{t_1}^{t_2} p_l(t)\cdot p_r(t+\tau)\,dt}{\sqrt{\int_{t_1}^{t_2} p_l^2(t)\,dt \int_{t_1}^{t_2} p_r^2(t)\,dt}}$$

$$IACC_{t_1,t_2} = \max\left|IACF_{t_1,t_2}(\tau)\right|_{-1<\tau<+1}$$

Equation 4.2: Inter-Aural Cross-Correlation IACC

where the terms p_l and p_r designate the sound pressures measured at the left and right ears respectively, which are the signals whose correlation is to be measured. The temporal offset between the two signals being measured is represented by τ. Usually the offset difference is large enough to encompass the maximum inter-aural time difference, which for a typical human head (*i.e.* dependent on the physical separation of the ears), is approximately ±1ms. IACC is therefore a binaural measurement of the difference in the waveforms at the two ears, where the impulse responses at the ears can be captured using a subject or an artificial head with omnidirectional microphones placed in the ear canals. IACC can be used to estimate both ASW and LEV by changing the integration periods (Beranek, 1996).

Bradley and Soulodre (1995b), following Morimoto and Iida (1993), introduced the idea that the perceived sensation of spaciousness relates to two components, the first being ASW, and the second LEV. They established that the broadening of a sound source, or ASW, relates to the relative level and angle of arrival of early lateral reflections with up to 80ms delay, and that LEV relates to the level of later-arriving lateral reflections, *i.e.* later than 80ms delay. According to them, objective predictors used to measure LEV need to take into account both level and spatial or angular distribution, and should measure the late arriving energy from 80ms to ∞. In their work, they compared different measurement techniques and concluded that late lateral sound level LG_{80}^{∞} was better for the prediction of LEV, than LF and IACC.

The main difference between the LF and IACC is that IACC is a binaural measurement that relates to differences in signals arising at the ears of a listener or a dummy-head, whilst LF is a ratio between the signals reaching two different (pressure and velocity) microphones. LG can be considered as an extension of LF. While LF deals with the early lateral reflections up to 80ms, LG is related to the measurement of late lateral reflections beyond 80ms. Both approaches measure the total pressure with omnidirectional microphones, but in the case of LG the total pressure is measured in an anechoic chamber, which is not easy to achieve in all measurement situations.

4.3 IACC and localisation

One of the first measurements identified that related to auditory spatial perception was linked to the localisation of sound sources. As indicated by Mason (2002), Jeffress (1948) investigated the association of inter-aural time-based cues with the perception of sound source localisation in space. Although Jeffress' theory assumes that the cross-correlation coefficient may be applied to any two signals, for the purpose of predicting perceived location, the cross correlation coefficient is employed to analyse the differences between a pair of binaural signals.

The results obtained from the IACC calculation can be graphed as in Figure 4-1 (adapted from Hirst (2006)), where binaural impulse responses were recorded using a dummy-head in an anechoic chamber.

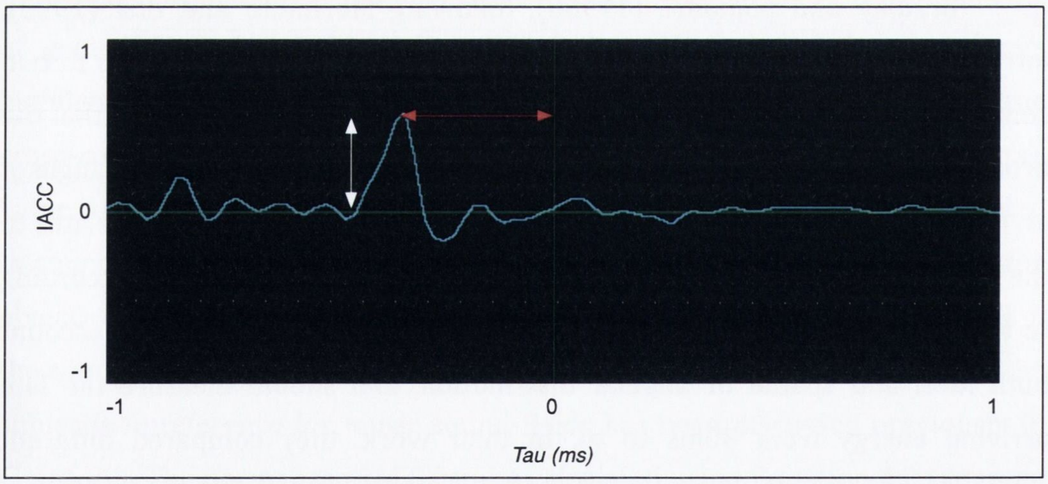


Figure 4-1: IACC for a binaural anechoic recording of a broadband noise source positioned at 45°.

On the graph, it is possible to identify the maximum peak value and the temporal offset of an IACC measurement, *tau*. The IACC peak (or maximum) value is indicated with a white arrow. This peak value can be used as an indicator of spaciousness (see Section 4.4). The offset of the maximum peak is represented on the τ axis, and is highlighted with a red arrow. By identifying the offset of the peak value, it is possible to determine the localisation of a sound source within the ±1ms window, which is the maximum inter-aural time delay for the average human ear spacing.

There is, however, some possible ambiguity when interpreting the IACC offset. Although it is a good indicator for sound source locations in the median plane, it is not capable of showing if the source is coming from front or back, above or below in relation to a person's, or an artificial, head. This confusion is more general that front-back discrimination. A source located anywhere on the surface of a cone with apex at the centre head will lead to the same inter-aural time difference. This "cone of confusion" gives rise to ambiguities when using IACC to determine sound source locations, as is indicated in Figure 4-2 regarding front-back confusion.

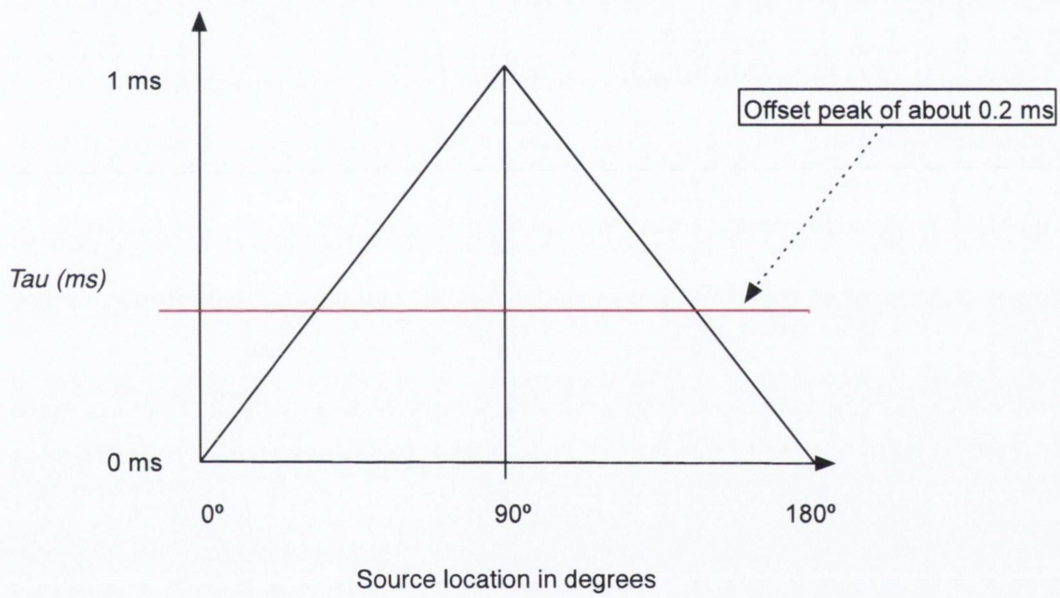


Figure 4-2: Demonstration of front-back confusion. Note that for an offset value of 0.2ms (red line), the source may appear to be coming either from 30° or from 150°.

The red line on Figure 4-2 represents an offset IACC peak of about 0.2ms, indicating a sound source radiating from an azimuth of about 30° from the centre front of a subject's or an artificial head. However, it could also indicate that the sound source is radiating from 150°, hence causing front-back localisation ambiguities in the measurement.

4.4 IACC and spaciousness

It is widely agreed by acousticians that spaciousness arises from the differences in the signals presented at left and right ears. Creating uncorrelated ear signals can cause these differences, which in the case of room acoustics is what early lateral reflections create when added to a direct sound. Since IACC is a binaural measure that represents the differences at the ears, it is understandable that many room acoustical measures for spaciousness are based on the inter-aural degree of coherence, *i.e.* IACC (Ando, 1985).

Chernyak and Dubrovsky (1968) investigated the degree of correlation between IACC and perceived spatial impression by creating broadband noise signals with varying degrees of IACC presented using headphones. In their experiments, they asked listeners to sketch where the perceived auditory event appeared to be on a graphical plan of a lateral cross-section section of the head, and also to differentiate the "magnitude of sensation" of the stimulus in each section of the plan. The results of these experiments showed that for an IACC value of 1 (totally correlated signals) the subjects represented the auditory event as appearing at the centre of the head, and that such an auditory event was fairly narrow in extent. As the IACC peak lowered, the auditory event experienced by the subjects became wider to the point where two separate signals appeared at either side of the headphones, this being the case for an IACC value of 0 (totally uncorrelated signals).

This work by Chernyak and Dubrovsky prompted further discoveries by others that followed. Room acoustics is an area of study where IACC has been used effectively as a physical measure which is correlated with the perceptual impression of spaciousness. Investigation of how the manipulation of room acoustics could lead to changes in perceived spaciousness has also been a concern of acousticians around the world.

Following Toole (2008), IACC correlates strongly with the perception of ASW, image broadening, spaciousness and envelopment and has been used by several researchers to assess sound field reconstruction from different playback systems.

While spaciousness studies have been primarily concerned with the understanding and development of concert hall acoustics, the results of such studies also have relevance for the perception of reconstructed sound fields, as exist in any sound field recording and reconstruction situation. The following chapters will deal with this, where IACC is used as a means to determine the effectiveness of microphone techniques, *shuffling* and up-mixing to 5.1 for control of spaciousness in reconstructed sound fields.

4.4.1 Just noticeable differences in IACC

Acoustic research from the second half of the last century has determined several objective measurements that relate to subjective preferences, IACC being one such objective measure as has been previously discussed. Such research has also lead to the identification of just noticeable differences (JND) for these physical measurements.

Pollack and Trittipoe (1959) investigated the JND of inter-aural cross-correlation for white noise stimuli; in their findings they were able to determine a small JND for the correlated noise stimuli (JND=0.02), and a larger JND for uncorrelated stimulus (JND=0.66). Later, Cox *et al.* (1993) determined JND values for changes in the intensity of the first early lateral reflection, and from their investigations they were able to convert the results to JND values for both LF and IACC. It is noteworthy that contrary to Pollack and Trittipoe (1959), where a noise stimulus was used, the stimulus signals used by Cox were musical motifs. It was noted by all that the smallest audible differences varied slightly depending on the stimulus used. The measurements they made were undertaken with a reference IACC of 0.33, and an average of the results was calculated which led to a JND of 0.075 (*i.e.* with an IACC of 0.33, an increase of 0.075 is necessary to be able to note a difference). Morimoto and Iida (1995) determined the JND of the inter-aural cross-correlation coefficient to be varying from 0.02 to 0.15 for an IACC range that varied from 0.9 to 0.5; such results were achieved by using two lateral reflections in a simulated sound field.

The JND for IACC seems to vary according to the different parametric conditions of the simulation (*e.g.* level of reflections, signal stimulus, and number of reflections). Okano (2002) gives a good insight into the JND for different physical measures used in concert halls and their relation to perceptual impression. From his research it was concluded that Apparent Source Width is the subjective attribute for which listeners are more sensitive to changes, and that for a range of IACC varying from 0.6 to 0.3, which is an accepted range for good concert halls acoustics (Beranek, 1996), the JND values vary between 0.05 and 0.08. The author is not aware of any studies of JND for the subjective attribute of spaciousness as used in this thesis. Consequently, for the work undertaken here an indicative IACC JND value of 0.065 was used, as the median value of the IACC JND range suggested by Okano.

4.5 Summary

In this chapter a discussion has been presented of some of the most relevant existing methods for measuring spaciousness. An explanation has been provided for their operational procedures, and their mathematical background was indicated. The uses of these measurement techniques have been very important in the characterization of spatial attributes in the perception of room acoustics. The variety of measures used can give rise to some confusion in their results, but the continuous development of these techniques has allowed for more consistent and reliable results in the measurement of spaciousness. The JND for IACC has been investigated and found to be source dependent. However, from the results available in the literature a JND for IACC of 0.065 has been proposed as a representative figure.

5 OVERVIEW OF EXISTING STEREO MICROPHONE TECHNIQUES

5.1 Introduction

Stereo recording techniques were introduced in the early 20th century by the pioneering work of Alan Blumlein (1933), Steinberg and Snow (1934) (under the direction of Harvey Fletcher), and others. Today's modern recording techniques make use of a variety of different microphone array configurations, which might include individual instrument spot micing, and also the use of other technologies that have been developed over the years (Bartlett & Bartlett, 1999; Streicher & Everest, 2006; Rayburn, 2012). It has been the intent of these stereo recording techniques to provide an acoustic illusion of virtual sound sources between two loudspeakers and, in some cases, to make this illusion appear to be beyond the physical span marked by the positions of the loudspeakers. Stereo reproduction is capable of producing listener cues which are the physical conditions required for the perception of source locations (phantom images) between the loudspeakers, and also other source impressions such as apparent size, diffuseness and space.

Stereo microphone techniques have been discussed minutely over the years (Audio Enginering Society, 1986). However, the perceptual impressions capable of being delivered by stereo recordings and reproductions have been debated (Gerzon, 1971; Lipshitz, 1986) but might have not been fully explored (Swedien, 2009). The fact that attention has been focused towards physical reconstruction of the localisation aspects of sound fields using stereo techniques has left a gap in the understanding of the physical parameters contributing to other perceptual impressions such as auditory spaciousness. It can be appreciated that, from the perspective of producers and sound recording enthusiasts, recordings (*e.g.* of music or sound for film) have not just evolved towards the localisation exactness of a physical reconstruction, but have also attempted to facilitate control over other characteristics of the reconstructed sound field, leading to the eternal question: "how can we make it sound better?"

5.2 From monophonic to stereophonic

Monophonic recording and reproduction is the simplest way of capturing and transmitting sound. Basically, sound is recorded on one channel using one microphone, and the recorded signal is then reproduced over one loudspeaker. The reproduction might be conveyed over more than one loudspeaker, but the signals being played are from the same microphone and transmission channel thereby making it the same in each loudspeaker, and therefore still a monophonic reproduction. The spatial cues delivered by such a reproduction system are very limited, since all sounds recorded (*i.e.* both direct and reflected sound) will be reproduced over one physical (loudspeaker) source, making it almost impossible to perceive the spatial origin of the sound as being anything other than the loudspeaker location. Thus, for example, the recorded reflections (*i.e.* the indirect sound) in the primary room are also reproduced as coming from the same location as the direct sound. Some physical attributes such as loudness, distance, depth and reverberation time might be conveyed to the secondary room listener, but most of the primary room binaural cues are missing in the secondary room reconstruction, which cues are, generally, important contributors to a listener's auditory spatial perception (see Figure 5-1)

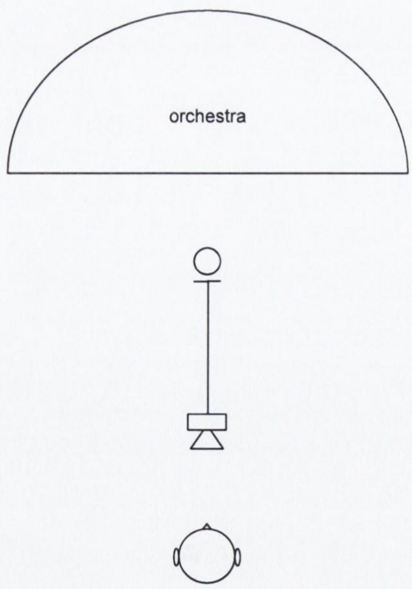


Figure 5-1: Monophonic representation of a recording/reproduction system

Stereophonic recording and reproduction, on the other hand, makes it possible to present a completely different secondary room listening experience.

By delivering important primary room (recording environment) physical cues to the secondary room listener, it is made possible for a secondary environment listener to perceive the source localisation (*i.e.* "phantom image") and other spatial aspects of an original performance. The physical cues provided are made possible because of the fact that, using a stereophonic system, there are two or more microphones and two or more loudspeakers used which are interconnected as separate channels, but which work together in creating the stereophonic illusion (see Figure 5-2).

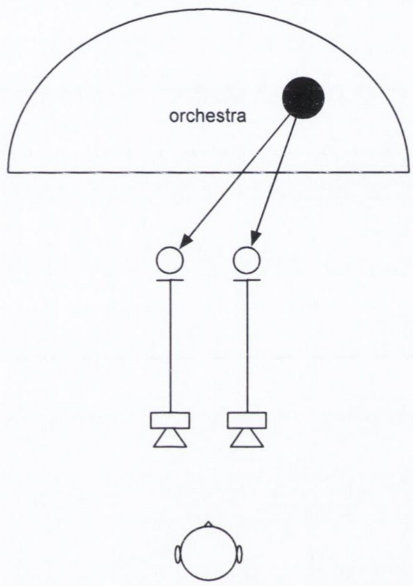


Figure 5-2: Stereophonic representation of a recording/reproduction system

By using two microphones and two reproduction channels it is possible to capture small differences of the sound in each microphone. Such small differences in intensity and time-of-arrival at each capsule are important for delivering some of the binaural cues necessary for the perception of a stereo illusion (*e.g.* the perception of a musical ensemble or orchestra spatially spread over the stereo stage). From Figure 5-2 it is possible to appreciate that a sound source to the right of a pair of microphones will reach the closest microphone earlier than the second microphone, while also the intensity of the captured signal will be lower (depending on frequency and microphone spacing) in the more remote microphone than in the closer one. These small nuances in both level and timing (*i.e.* phase differences) between the microphones are the backbone of the stereophonic illusion. Controlling the parameters involved in the recording setup (*i.e.* microphone angular positioning, spacing, and capsule type) will ultimately lead to alterations to the listener's overall spatial impression of the recorded content.

Localisation is an important feature of a stereo reproduction, but it is just one of the aspects of a good stereophonic recording. The perception of spaciousness in a stereophonic recording is also of importance, as it has been found that auditory spaciousness contributes to the aesthetic impression of a listeners' experience. Delivering physical cues, which might be the necessary conditions for the perceptual impression of spaciousness, allows for a better appreciation of a room's acoustics (see Chapter 3). Is it therefore possible to present and manipulate spacious sounding recordings? The work undertaken by Kurozumi and Ohgushi (1983) concluded that the main psychological factors governing sound image quality of white and band-limited noise sources with various cross-correlation coefficients reproduced via two loudspeakers (in a typical stereophonic arrangement with an included angle of 60°, see Figure 5-3), in either anechoic or echoic rooms, are the impressions of image width and distance. Some authors have discussed stereo image width as being equal, in terms of meaning, to spaciousness (Griesinger, 1985), but this is an issue requiring some clarification. For now, the possible relationship is simply being noted.

5.3 Intensity stereo

Blumlein's inspiring work relating to stereo was documented in his original patent (1933), where a system for recording, transmitting and reproducing sound which preserved source localisation information was introduced. The intent of such a system was to convey to the secondary room listener an impression of a spatially extended sound stage by delivering physical cues that gave the illusion that the sound is coming from a particular direction. Blumlein described several methods for capturing sound sources, where the recorded signals were then played back over a set of loudspeakers. The operation of the system can be understood by noting that inter-loudspeaker amplitude differences cause inter-aural phase differences for a listener, as described in Figure 5-3 following. Below approximately 700 Hz listeners identify the localisation of sound sources using the phase difference (binaural delay) between the ears. Above about 2 kHz, the wavelength of the signals is smaller than the dimensions of a listener's head which therefore causes shadowing, leading to an attenuated and filtered signal at the farther ear. These different cues reinforce each other and contribute to an ability to decode the localisation information delivered by a stereophonic system, thereby allowing listeners to perceive the illusion of spatially spread sound sources.

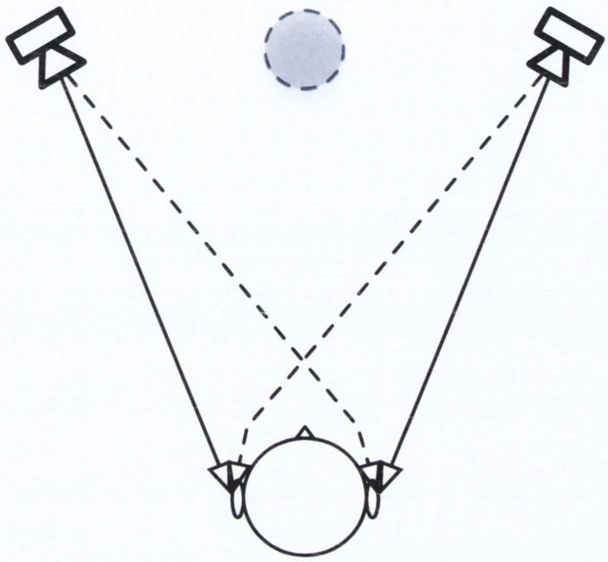


Figure 5-3: Two loudspeakers at ±30° creating equal signals at the ears of a listener, leading to a phantom sound source at the centre front. The introduction of level differences or delay to either Left or Right channel will cause the phantom image to shift to either side, depending on the relative level difference and/or on whichever channel is leading.

A physical interpretation of Blumlein's stereo arrangement recognizes that, since each loudspeaker communicates with both ears, differences in magnitude of the sound pressures at the loudspeakers at low frequencies produce phase and not magnitude differences at the ears, since the contributions from the two loudspeakers arrive at slightly differing time (Lipshitz, 1986). Using a pair of directional microphones, which must be arranged in a coincident fashion, to capture acoustic signals in a space will lead to signal outputs that are in phase with each other, but which will have magnitude differences determined by the directivity characteristics of the microphones and the location of the sound source in relation to each microphone. These signals, when played back through a pair of loudspeakers (Figure 5-3), will produce inter-aural phase differences at low frequencies and intensity differences at high frequencies due to the head shadowing effect. It has been claimed that this represents the nearest approach yet made to natural listening conditions (Clark, Dutton, & Vanderlyn, 1958).

5.3.1 Coincident microphone techniques

Coincident stereophonic microphone techniques are based on the intensity stereo principle described in Section 5.3. In order to achieve the stereo illusion, two microphones are used which must be arranged in a coincident array, where both microphones need to have their capsules as close together as possible. By achieving this arrangement inter-capsule phase differences are minimized since the recorded signals will arrive at both capsules at the same time. The output signals of the microphones will have only magnitude differences which are appropriate for delivering the necessary physical cues during the stereophonic reproduction. The choice of capsule and the angular span between capsules is of most importance in such techniques, since these factors are what account for the necessary left and right signal magnitude differences. This technique has the advantage, since it is based on intensity stereo, of maintaining angular accuracy in the stereo imaging regardless of the distance of the array from the source. As a disadvantage, it is pointed out that due to the lack of inter-channel time delay, the stereo image sometimes seems lacking in the "sense of space" or spaciousness (Streicher & Dooley, 1985), (Rayburn, 2012, pp. 217-224).

The most common coincident microphone configuration is that of the XY microphone arrangement, in which two directional microphones are positioned one above the other in a crossed configuration where one microphone points to the left, and the other points towards the right, of the sound stage. One of the most common choices of microphone type here is that of the cardioid polar pattern. The XY technique utilizing cardioid microphones which can be splayed over an angle of 90° to 135° is a common choice among sound recordists (Rayburn, 2012). Figure 5-4 illustrates an XY arrangement with an included angle of 90°.

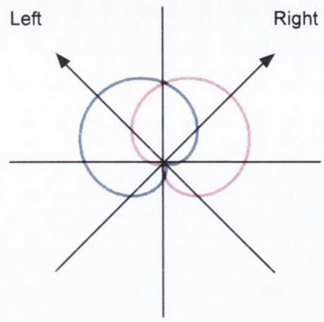


Figure 5-4: XY cardioid pair with an included angle of 90°.

The angle between capsules dictates the "apparent width" of the stereo stage, and can be set, depending on the suggestion used from the literature, from 60° to 180°. The choice of angle is usually determined by the individual preference of the sound recordist, which might relate to how "centre heavy" is the stereo image required to be, and to the direct-to-reverberant balance desired. Also, the choice of microphone capsule is taken into account. Crossed cardioids with an included angle of 90° might cause a very narrow stereo image. By splaying super-cardioids at 120° (Figure 5-5), the stereo image will became wider, and sound sources will be more naturally spread across the stereo span of the reproduction system (Rayburn, 2012). However, the choice of microphone capsule is of great importance not just for stereo imaging issues, but also for issues related to the spaciousness of the stereo field. By using super-cardioids at 90° instead of cardioids (Figure 5-6), it is possible to increase, due to inter-channel separation and the anti-phase lobe inherent in their polar patterns, the perceived impression of spaciousness, which issue will be discussed further in Chapter 7. Gerzon (1986) has also indicated that a small capsule separation of about 5 cm, in an XY technique with cardioids splayed at 115° to 120° apart could, in fact, not only provide for better image directionality but also increase the perception of spaciousness, especially when stereo "shuffling" is applied, as will be discussed further.

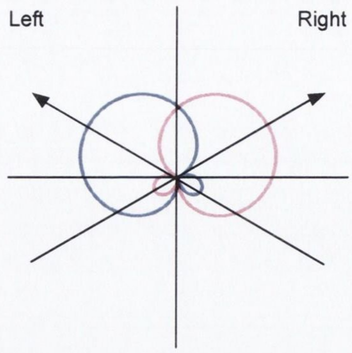


Figure 5-5: XY super-cardioid pair with an included angle of 120°.

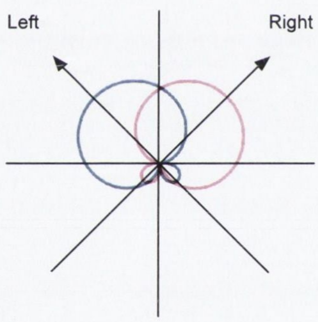


Figure 5-6: XY super-cardioid pair with included angle of 90°. The anti-phase lobe of the super-cardioid is noticeable in the rear-facing quadrant of the pair.

A crossed pair of figure-of-eight microphones is described in Blumlein's patent (1933) as a technique where the outputs of the microphones can be sent directly to the loudspeakers to which they will give the correct amplitude differences for the desired directional effect, if it is possible to neglect any phase differences due to time of arrival of the sound wave at each of the capsules. This will be the case, similar to the XY techniques already described, if the capsules are close together in a coincident fashion. Since each of the figure-of-eight microphones has its maximum positive lobe pointing at either the left or the right edge of the sound stage, the null part of each of these $cos\emptyset$ polar pattern microphone will be facing toward the opposite edge of the sound stage (Figure 5-7). This technique has the rear quadrant of the array picking up the indirect sound field (*i.e.* the room, hence the reflected, sound) with reverse polarity to that of front-quadrant which will be played back in a cross-channelled manner in the Left and Right loudspeakers. Since the sound field pickup at the rear quadrant has a diffuse quality, the sound imaging will not suffer because of this. In fact, it is often commented that this technique makes for a very natural sound (Gerzon, 1976) (Streicher & Everest, 2006). The left and right quadrants of this array will be picked-up out of phase (see APPENDIX I – Magnitude and Phase Response of a Blumlein Pair) which makes the placement of this array crucial in order to get a proper balance of direct-to-reverberant sound ratio, and to avoid excessive out-of-phase signals. It is recommended to avoid strong early reflections from the side walls because of the out-of-phase issues that might arise (Streicher & Dooley, 1985). However, the pick-up of strong early reflections with this technique can have a significant impact on the perceived impression of spaciousness, as will be demonstrated in Chapter 7.

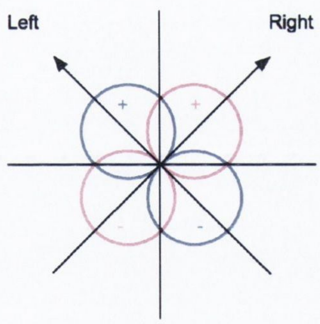


Figure 5-7: "Crossed figure-of-eight" pair, also known as Blumlein stereo microphone technique, with included angle of 90°.

Another popular coincident technique is that of the M-S stereo – also described in Blumlein's patent (1933). This stereophonic recording technique utilizes a forward facing microphone which is labelled the Mid component (M) and a side facing microphone which is labelled the Side component (S). In order to obtain utilizable Left and Right signals for playback, this technique needs to be decoded using a Sum and Difference matrix. The outputs from the matrix resolve into XY stereo signals, as can be observed in Figure 5-8. The decoding process involves the addition and subtraction of the M and S components: (M+S)=Left and (M-S)=Right. This technique has the best monophonic compatibility of all of the coincident techniques, since adding the Left and Right signals: (M+S) + (M-S)=2M, will result in the retrieval of just the forward facing microphone which is on-axis (*i.e.* the 0° direction of the M microphone) with the sound source. It is therefore important to notice that a proper monophonic recording can be balanced by a suitable choice of M microphone, and properly aiming it at the sound source before the resultant stereophonic recording is decoded. The side-facing microphone needs to be always of figure-of-eight polar pattern. This intensity stereo technique has no phase differences between capsules because of their proximity (one above the other).

M-S stereo also offers great advantages for the mixing process. By carefully adjusting the relative level between M and S components, it is possible to control the stereo width, the apparent distance, and the amount of ambience of the recorded image. If a remote-controllable-pattern microphone is used to adjust the M polar pattern, greater control over the stereo image can be achieved. The control of the sum and difference relationship can also lead to changes in perceived spaciousness. All of the adjustments made to the decoding process can be either done during the recording process, or later during the mixing session (Dooley & Streicher, 1982).

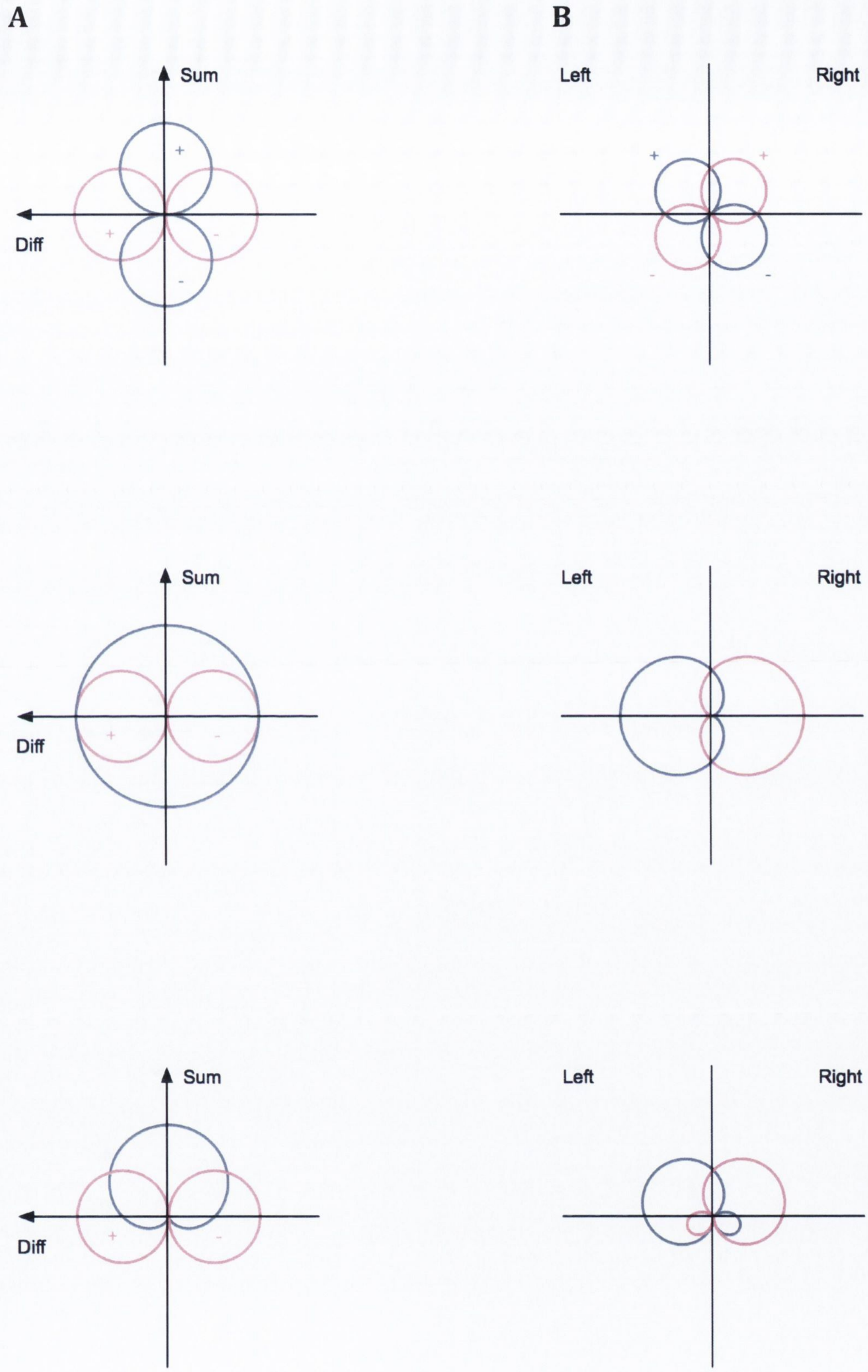


Figure 5-8: M-S stereo to equivalent XY stereo signals. (A) Sum and difference signals at equal level. (B) Resulting Left and Right signals after the decoding process.

5.4 Stereo "Shuffling"

Blumlein (1933) did not just devise stereo as being Left and Right signals. Rather, he in fact devised stereo as being L+R and L-R. To think of stereo as Sum and Difference channels creates new dimensions regarding what can be accomplished when stereophonic signals are reproduced over loudspeakers (as will be further discussed in Chapter 7). Any stereophonic recording can be encoded into Sum and Difference channels, which makes it possible to adjust the signals in a manner that is different from that which results from adjusting the Left and Right signals independently. By processing the Sum and Difference channels separately it is possible to make changes to the different components of the sound field. One can, for example, adjust the relative levels of the information concerning the middle of the stereo stage (*e.g.* direct sound) from that of the information concerning the sides of the stereo stage (*e.g.* reflections). Also, these signals can be filtered differently in a manner that can produce controllable responses at the ears of the listener.

When "shuffling" was first conceived it was to convert phase differences from a pair of spaced pressure microphones into appropriate magnitude differences which could then be reproduced over loudspeakers working in a manner similar to intensity stereo. These findings were revised later by Clark *et al.* (1958) and Gerzon (1994). However, in creating such a network device, Blumlein discovered that if an increase in the level of the difference channel were to take place, the width of the stereo image would increase as well. Reducing the level of the Difference channel would, therefore, decrease the stereo width. It was noted that the Difference channel could not only be merely altered in level, but it would be possible to control its level in a frequency-dependent manner. By doing this, it was discovered that not only could the directional quality of particular stereo techniques (*e.g.* a spaced pair of omnis) be improved, but also that of the perceived impression of spaciousness (Griesinger, 1985) (Gerzon, 1986).

Stereo "shuffling", as discussed by Gerzon (1986) without the phase shifts as initially proposed by Blumein (1933), is exploited in this thesis as a contributor to increasing the perceived spaciousness of the reconstructed sound fields of stereo recordings. The perceptual impression created by such an effect can be objectively measured using Inter Aural Cross Correlation (IACC), as is discussed further in Chapter 7.

5.5 Spaced arrays

Stereo microphone techniques are also implemented using pairs of spaced microphones (Figure 5-2). The most commonly used spaced stereo format is the AB technique where, usually, two pressure microphones are spaced apart in which the distances between capsules can be varied according to the width of the sound stage to be recorded, or to the reconstructed stereo stage spread desired (Bartlett & Bartlett, 1999). With this technique, stereo signals are captured which feature a time-of-arrival difference at each of the capsules. Sounds located closer to one capsule will arrive later at the opposite capsule. Depending on the distance separating the capsules, the "Inverse Square Law" (Everest & Pohlmann, 2009) contributes to the Left and Right imaging, since sounds will have greater amplitude at one capsule relative to the other (Streicher & Everest, 2006). However, if the spacing between capsules is too far apart discrete echoes may result which generally degrade the perceived stereo sound field. Experiments investigating this technique were first undertaken at Bell Laboratories in the 1930's (Steinberg & Snow, 1934), (Snow, 1953).

Many authors (Streicher & Everest, 2006) (Rayburn, 2012) have presented these spaced arrays as providing an increased sense of "space" around the performers. Informal comments regarding such techniques adding more "air", delivering a more "open" sound, or providing a more "spacious" sound have been at the origin of several disputes regarding stereo performance, with one faction claiming that such techniques result in inferior stereo imaging accuracy when compared to coincident techniques (Lipshitz, 1986), and the other faction claiming that the loss of imaging accuracy is compensated for by the more "spacious" and "natural" sound which results (Theile, 1991). This effect of a perceived increased spaciousness is somewhat related to the fact that phase anomalies (*i.e.* comb filtering effects) are introduced by the time-of-arrival differences between the two capsules. However, this "phasiness" has sometimes been appreciated as something which is pleasant to the ears, and is considered to be an improvement over coincident techniques. It is worth mentioning that most of the stereo orchestral recordings undertaken in America throughout the 1950's and 1960's were made using such spaced techniques, and that there is also a strong tradition of using this recording technique to capture the sound of small musical ensembles (Rayburn, 2012).

5.5.1 Near-Coincident techniques

Microphone techniques that are categorized as near-coincident are usually arrays that effectively act as a coincident technique up to a certain frequency, below which the wavelengths are larger than the inter-microphone distance. From that frequency onwards there will be some evident time-of-arrival differences between the capsule signals. Such microphone techniques will have a small inter-capsule separation which, in most cases, use distances that are close to the natural separation of the human ears. Since the spacing between capsules is small (*i.e.* 10cm - 30cm), the in-phase relationship of intensity stereo (Section 5.3) is still preserved for frequencies up to approximately 1000 Hz, which make these techniques largely dependent on intensity differences for their stereophonic information. However, the introduction of small time-of-arrival differences due to capsule separation and orientation of the microphones will result in small phase differences, and such minimal phase differences are perceived as a more "airy" or "spacious" impression. As with the AB techniques, informal comments that relate to the quality of the stereo sound created are subject to much controversy, but in the case of near-coincident techniques this is minimized because near-coincident techniques are considered as being a hybrid format (Lipshitz, 1986). Figure 5-9 to Figure 5-11 illustrates some of the most commonly used near-coincident techniques.

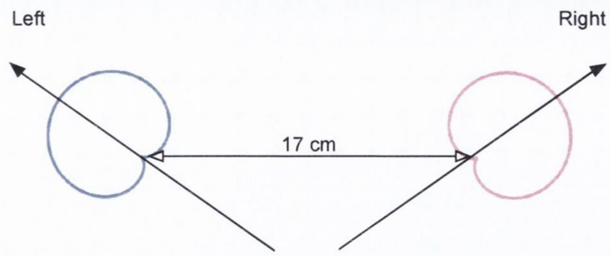


Figure 5-9: ***Office de Radiodiffusion-Télévision Française*** **(ORTF) stereo microphone technique. This near-coincident technique utilizes two outward facing cardioids with an inter-capsule spacing of 17cm between them with an included angle of 110°.**

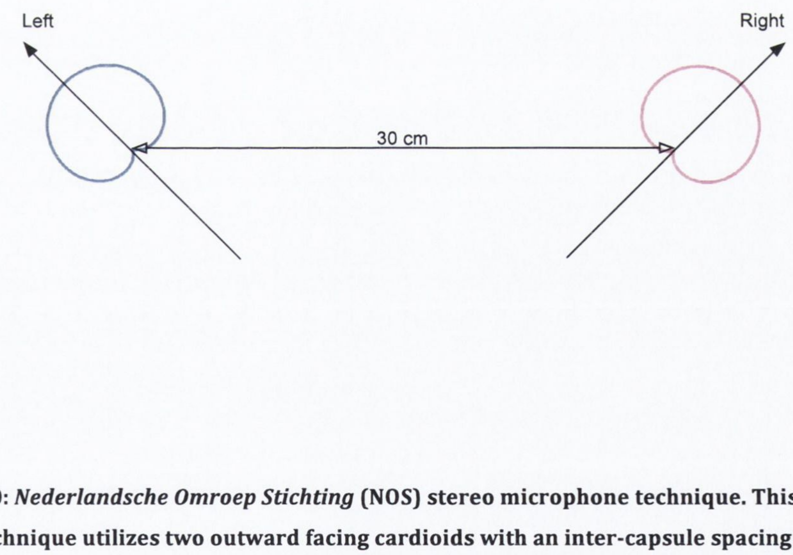


Figure 5-10: ***Nederlandsche Omroep Stichting*** **(NOS) stereo microphone technique. This near-coincident technique utilizes two outward facing cardioids with an inter-capsule spacing of 30 cm between them with an included angle of 90°.**

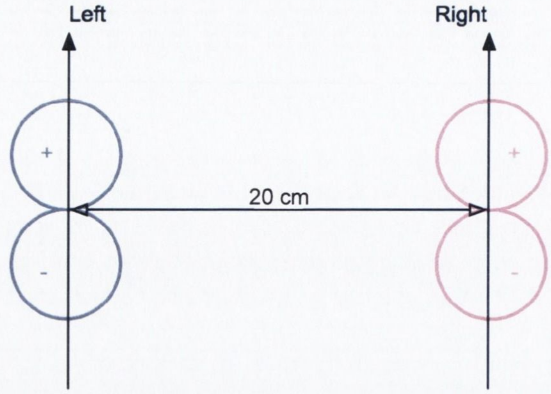


Figure 5-11: Faulkner array stereo microphone technique. This near-coincident technique utilizes two forward facing fig-of-eights with an inter-capsule spacing of 20cm between them.

The spacing, angle of splay, and choice of capsule for these types of arrays can be varied according to the creative choice of the producer/engineer, and according to the recorded program. Since the introduction of such techniques, as in ORTF for example, there has been extensive experimentation and documentation supporting the choice of angle, separation, and polar pattern. Michael Williams has proposed data for a useful range of angle and spacing for these recording techniques (Williams, 2013). The Image Assistant 2.1 developed by Helmut Wittek is a JAVA applet which allows for the calculation of localisation curves for stereophonic recordings when reproduced over Left and Right loudspeakers, and over Left, Centre and Right loudspeakers, based on psychological and mathematical fundamentals (Wittek, 2011).

5.6 Surround sound

From the last quarter of the 20th century, research in surround sound technology, including microphone techniques, has generated considerable interest. But, what is surround sound or “surroundphonic” reproduction? By looking at the meaning of the word stereophonic – a solid or firm sound – one could raise the issue that, based on its meaning, stereo sound could already potentially present the listener with a surrounding 3-D sound. However, stereophony is still regarded, in most cases, as simply a 2-channel reproduction system. Nevertheless, if Blumlein’s British Patent 394,325 (1933) is examined, it is there proposed that a plurality of microphones and loudspeakers might be used to convey a complete directional "sound picture" (*i.e.* including horizontal and vertical directions) whereby a full three-dimensional location sound system can be accomplished. It seems, therefore, important to consider the possibilities of true stereophonic recording and reproduction to produce the desired surround sound which has become so popular in more recent years.

Surround channels (*e.g.* loudspeakers placed to the rear of a listener) have been proposed either for motion pictures, or for musical presentation in both consumer (home), and in public presentations. The introduction of these surround channels is considered to be an improvement in the reproduction of de-correlated signals. It has been shown that at least four channels are needed to reproduce similar IACC trends of de-correlated signals as those created in a reverberant chamber (Tohyama & Suzuki, 1989). The added channels might lead to a subjective preference for the sound field so created if suitable signals are fed to these channels.

There have been several possible surround sound system layouts proposed, and in more recent years even more have been suggested, especially to render height information. However, the ITU recommendation (ITU-R BS.775-1, 1992-1994) is the most commonly used format for mixing and production of music and film sound for home presentation. There have been claims, by many workers in the field, that such recommendations are appropriate for surround sound presentation, but the included angle of 60° between front Left and front Right speakers may be too wide for "normal stereo" presentation for listening in the home (Rayburn, 2012). Although the ITU norm is oriented towards music production, it is not a popular choice for music presentation, despite the fact that 5.1 systems have become widely available for consumers. Why music production is still mostly presented in 2-channel stereo is a question yet to be answered. Can it be because there is insufficient "added value" for 5.1 music presentation to justify the overall system cost as compared to 2-channel stereo, which is quite good and very adaptable in terms of playback systems. This then raises the question: "how can extra quality be added to 5.1 presentation?" Figure 5-12 demonstrates the layout of a surround sound monitoring system in accordance with ITU-R BS.775-1. It is important to note that in using this layout for a surround sound mix, the 3 frontal speakers (*i.e.* Left, Right and Centre) are important in stabilizing stereo images, and in providing an enlargement of the listening area, causing less listening fatigue as well as enhancing stereophonic reproduction. In general, these observations have been offered by Michael Gerzon based on his extensive stereo listening experiments (1990; 1992a; 1992b). These frontal channels are used for delivering most of the physical localisation cues; localisation of phantom images to the sides and rear is less precise. The use of the surround sound channels is mainly for delivering early reflections and reverberation, because of the difficulty in localizing phantom sources to the sides and rear (Rayburn, 2012). The most common use of surround sound in natural stereophonic recordings is therefore to reproduce the reflected sound of a primary recording environment, which will produce a more spacious sound. In some pop mixes and electroacoustic music, the use of surround channels might be to convey some instruments or parts which, depending on the style of the music, might feel awkward.

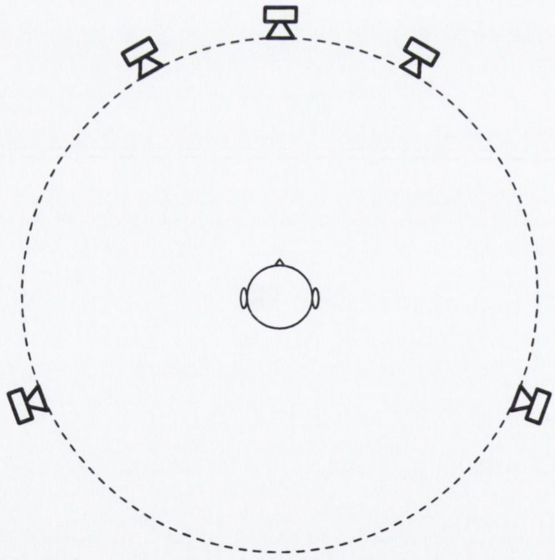


Figure 5-12: ITU recommended loudspeaker arrangement for surround monitoring. Front Left and Right speakers have an included angle of 60°. Rear Left and Right speakers are positioned between 100° and 120° from the centre-front (*i.e.* Centre positioned speaker).

Surround sound recording and reproduction has been investigated from several "point-of-views": from head-related and transaural, to single-point (*e.g.* Ambisonic) techniques. Stereo-derived techniques are the main focus of this thesis. These stereo-derived techniques are basically derivations of current stereo formats, making use of both intensity and time-of-arrival differences at the microphones to provide the signals that will create “phantom images” in the reproduction system, as well as de-correlated signals that might provide for a more spacious sound presentation. According to Eargle’s Microphone Book (Rayburn, 2012), stereo-derived systems are based on three acoustical elements which are necessary for convincing surround sound reproduction: 1) accurate pickup of direct sound, 2) pickup of sufficient early reflections from the recorded room which will be generally conveyed by all of the loudspeakers in the surround system, and 3) pickup of uncorrelated reverberation which will be presented over the entire surround system (Rayburn, 2012). These conditions are the same as those required for effective 2-channel stereo. However, the proper reproduction of early and late reflected energy from a primary environment will be improved in the secondary listening room when presented over the 6 channels of a 5.1 surround system (Theile, 2001). For the work undertaken in Chapter 8 of this thesis, the ITU recommendation was adopted.

5.6.1 Examples of microphone techniques for surround sound presentations

As discussed previously, some of the microphone techniques for surround sound recording are approached from coincident and spaced arrays, with most of the surround microphone techniques being designed using the same principles of 2-channel stereo where the output signals from such arrays feed the 5 cardinal points (if a 5.1 system is used) of a surround sound layout.

Surround microphone techniques have been, in more recent years, widely proposed (Rumsey, 2001; Streicher & Everest, 2006; SCHOEPS GmbH, 2006; Williams, 2013; DPA Microphones A/S, 2013) and they can be categorized, like 2-channel stereo techniques, as coincident, near coincident and spaced arrays.

5.6.1.1 Coincident techniques

Going back to the era of quadraphonic surround sound, in the 1970’s, a technique was proposed by Yamamoto (1973) which comprised of four cardioid capsules, each pointing at each quadrant of a circle. This configuration is essentially similar to having 2 coincident pairs of XY cardioids at 90°, in which one pair is forward facing towards the sound stage, and the other pair is rearward facing towards the hall. Georg Neumann GmbH manufactured the QM69 model of a quad cardioid microphone, and today there are manufacturers such as Pearl Mikrofonlaboratorium AB, Line Audio Design and Nevaton Microphones who still make quadraphonic microphones.

The "double MS" technique proposed by Curt Wittig, has become a widely used technique, especially in the movie industry, mostly because of its versatility and ease of configuration (SCHOEPS GmbH, 2006). The array is basically made of 2 MS pairs in a back-to-back configuration (one pair forward facing, and one pair rearward facing) which must be organized in a coincident fashion. Since the two MS pairs provide a Side facing microphone (with their 0° axis "aiming" at the same point) it is possible to "drop" one of these capsules and use just one which will be shared by both fore/aft pairs. By using this arrangement, it is possible to have a microphone array, comprised of 3 microphones, which will provide 5 channels. The decoding of these arrays is similar to the decoding process discussed in Section 5.3.1. The forward facing pair will provide for the frontal sound stage in a 5.0 reproduction layout, while the rearward facing pair will provide the "ambience" (*i.e.* room reflections) of the sound field recorded. The central channel information can be obtained by feeding the centre loudspeaker with the forward facing Mid signal (Wittek, 2009).

The Soundfield microphone, introduced by Gerzon (1975), consists of four sub-cardioids capsules closely spaced together in a tetrahedral arrangement, where the output signals from the microphones are labelled A-Format. The coincidence of the capsules and polar pattern response is further improved by means of elaborate frequency-dependent matrixing. After the appropriate processing of the A-Format signals, a new B-Format is derived: a pressure component (W) and three pressure-gradient components (X, Y, Z) (*i.e.* fore/aft, left/right and up/down figure-of-eights). From these signals, 2-channel stereo or surround sound layouts can be driven. The idea for this technique derives from some of the ideas proposed by Blumlein (1933). Such technique can provide surround sound imaging, including height information, independently of the number of output loudspeakers used. The initial approach used was that the signals from the B-format should feed regular polygon surround arrays, for best performance. Since most of the commercially available surround sound speaker layouts are irregular arrays, such as the ITU-R BS.775-1 (1992-1994) 5.1 surround sound system (Figure 5-12), a decoding processing was later proposed for converting B-Format signals into G-Format signals which feed 5.1 surround sound systems (Gerzon & Barton, 1992). Soundfield microphone techniques are here introduced since they were investigated as a surround technique for 5.1 presentations, but they will not be further discussed since it is beyond the scope of this thesis. Refer to Rumsey (2001), Streicher *et al* (2006) for more technical details, and the Michael Gerzon audio pioneer web page for a list of publications regarding Ambisonics and the Soundfield microphone (Thorton, 2009).

As with "normal" stereo coincident techniques, these ambisonic arrays are capable of providing accurate source localisation, but criticism of them are similar to what was previously presented in Section 5.5 with regard to the lack of a more "airy" or "spacious" sound (Theile, 2001; Streicher & Everest, 2006).

5.6.1.2 Near-coincident techniques

Similar to the proposed data for a useful range of microphone angle and spacing for 2-channel stereo recordings, which has been extended to multichannel surround (Williams, 2013), it is possible to find several approaches for near-coincident surround microphone arrays that have been exploited either for music or film presentations. In all of the proposed approaches, the idea is to provide an accurate sound stage with an increased sense of spaciousness due to inter-capsule separation which leads to time-of-arrival differences in the signals of the microphones (Theile, 1991; 2001). Currently, it is possible to find (commercially available) surround sound microphone techniques that fall into this category. Arrays, such as the Optimized Cardioid Triangle (OCT) Surround, ORTF Surround, IRT Cross (a.k.a. Surround Ambience Microphone SAM), Sound Performance Lab (SPL) Atmos 5.1 and the Schoeps KFM360 – DSP4, have been proposed by several authors and microphone companies (Rumsey, 2001; Rayburn, 2012). For the investigations undertaken in this thesis, the OCT Surround (Theile, 2001) (see Figure 5-13) and the KFM360 – DSP4 (see Figure 5-14), proposed by Jerry Bruck (1997), were used. The OCT Surround, introduced by Gunther Theile, is a discrete microphone array, which was derived from the OCT microphone technique for 2-channel stereo recordings (Theile, 2000), by which the output signal from each microphone discretely feeds each channel of a 5.1 surround sound system. The use of super-cardioids for Left and Right channels increases channel separation (*i.e.* minimized interfering crosstalk between the channels) and optimized stereo imaging across the Left-Centre-Right (*i.e.* frontal speakers) channels. This is of importance because one of the problems with arrays containing three-frontal-microphones is the occurrence of multiple phantom images between Left-Centre, Centre-Right and Left-Right (Theile, 2001). As for the KFM360 – DSP4, proposed by Jerry Bruck, the array comprises two MS pairs, separated by a spherical baffle, which have a pressure microphone for the Mid signal. It is possible to separately obtain forward and rear patterns decoded from each MS pair at each side of the sphere which allows for an excellent fore/aft spatial distinction. The centre-front channel is achieved by using a 2-to-3 matrix proposed by Gerzon (see Section 8.2.3 for more details) (Rayburn, 2012).

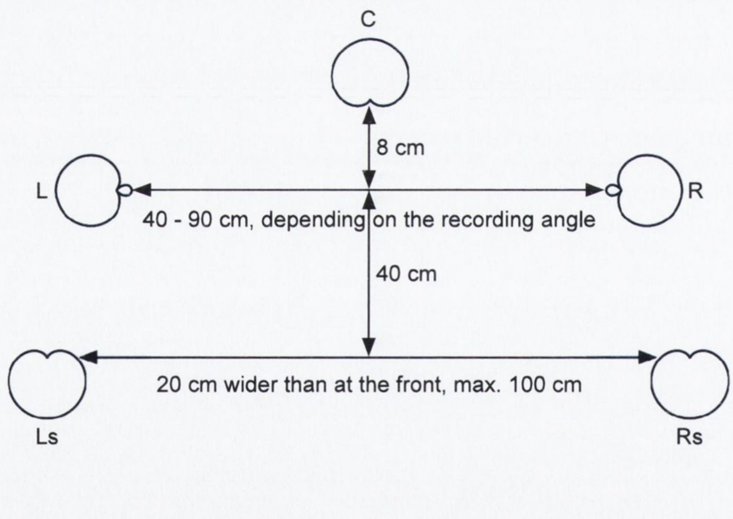


Figure 5-13: Optimized Cardioid Triangle (OCT) Surround. The Centre, Left surround and Right surround microphones use cardioid polar patterns, while the Left and Right microphones use super-cardioid for improved frontal separation.

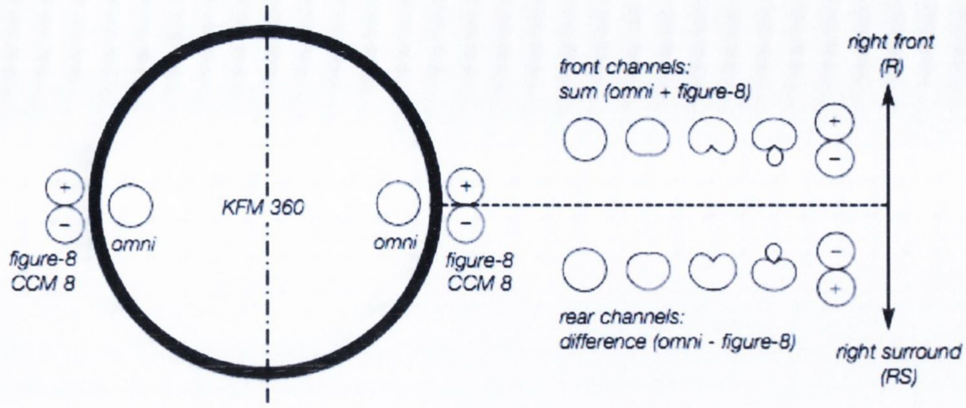


Figure 5-14: Schoeps KFM360 – DSP4 microphone array. The front and rear pattern selection for front and rear channels is represented (SCHOEPS GmbH, 2013a)

Both of these arrays, due to inter-capsule separation either from the microphone spacing or the minimized crosstalk due to the use of more directional polar patterns (OCT) and baffles (Bruck array), claim to provide natural recordings which deliver good directional clarity, while providing the necessary reflections captured in the primary room to the secondary listener environment, allowing for an improvement in the perceived impression of spaciousness when compared to 2-channel stereo recordings.

5.6.1.3 Spaced Arrays

Spaced array microphone techniques are arranged in a manner similar to that of 2-channel AB stereo techniques, where the time-of-arrival of a sound source between microphones is of importance for delivering the cues for sound stage representation. Such arrays are setup with an inter-capsule distance of several metres and are mostly used for capturing the diffuse field of a room. These techniques are designed to work in conjunction with other techniques (coincident or near-coincident), working as a complement to such recordings. An example of such an array is the Hamasaki Square (Hamasaki & Hiyama, 2003). These techniques are mentioned here for the sake of comprehensiveness, but will not be discussed further, as they were not used in the investigation conducted for this thesis.

5.7 Summary

An overview of the technology and techniques involved in *'purist'* stereophonic sound recording was presented. The term *'pure'* here means that the approach to stereophonic recordings is based on the principle of capturing a primary room sound field for later reproduction in a secondary listening environment by utilizing microphones techniques capable of capturing the entire sound field, as opposed to using individual spot microphones for each source which are recorded in multi-track fashion for later mixing.

Since the introduction of stereophonic techniques, there have been investigations addressing their improvement and refinement, either by revising existent techniques or by proposing new ones. However, most of the investigations have been aimed at the physical reconstruction of the primary room sound field in a secondary listener environment, where source localisation has been prioritized. Despite the importance of such approaches, it is also of importance that the concerns of improving stereophonic reproduction should also be aimed toward other sound field perceptual characteristics which ultimately will provide a "better sounding recording". The following chapters will provide more insight into how the perceptual characteristic of spaciousness can be presented and controlled in stereophonic and surround sound recordings and reproduction.

6 INFLUENCE OF DIFFERENT TEST ROOM ENVIROMENTS ON IACC AS AN OBJECTIVE MEASURE OF SPACIOUSNESS

6.1 Introduction

Early lateral reflections have already been discussed in relation to the perception of spaciousness. A strong single early lateral reflection added to a direct sound leads to a change in the perception of spaciousness, as was discussed previously in Chapter 3, Section 3.3. To see the effect of one or multiple early lateral reflection(s), an anechoic chamber is usually used since the lack of acoustical reflections makes it possible to control the sound field electro-acoustically and be sure that only the reflection under study is contributing to the results. However, getting access to an anechoic chamber to perform these tests can be difficult, and a limitation. Because of this, it was decided to explore whether similar IACC trends for changing early reflections parameters could be achieved in echoic rooms in comparison to an anechoic reference, when playing a direct sound and a simulation of a single early lateral reflection. This would constitute an added value in these types of studies.

IACC was chosen from other means of objectively assessing spaciousness (see Chapter 4), because the human capability of processing spatial information is principally a binaural process (Schroeder, Gottlob, & Siebrasse, 1974; Ando, 1977; Blauert, 1997; International Organization for Standardization (ISO), 2009). Monaural cues, such as pinnae effects, are of importance for processing spatial information, but since it was suggested that the degree of the spatial effect was related to the cross correlation between signals at the two ears, attention should be paid to the binaural process (Toole, 2008). Barron proposed the need of early lateral reflections for a desirable spatial effect (1971).

Using loudspeakers to generate a direct sound plus a single lateral reflection which can be digitally controlled in level and time of arrival, it is possible to setup this experimental arrangement in any room of choice. This was undertaken in order to examine if IACC measurements of similar setups of direct sound and a single lateral reflection would give the same IACC trends in different rooms with changing reflection amplitude and delay. The results achieved show that it is possible to obtain similar trends for IACC peak values for similar setups where only the room changed. The rooms used were the Music and Media Technologies studio in Trinity College Dublin Ireland, the *Serviços de Áudio* studio in the School of Music, Arts and Performing Arts of the Polytechnic Institute of Oporto (Portugal), and the Multichannel Anechoic Chamber of the School of Science and Technology, Aalto University, Helsinki (Finland). From now on, these rooms will be here referred to as Dublin, Oporto and Helsinki respectively.

6.2 Experimentation

Apart from the anechoic chamber in Helsinki, which is a non-reflective room, the studios used, although designed specifically for use in sound recording, had reflections which contributed to the reconstructed sound fields for which the IACC measurements were being made. Because of its non-reflective characteristics the results from the anechoic chamber were used as a reference for the results from the other two rooms. The three rooms did not differ much in terms of spatial volume, and were of a generally cubic shape, except for the room in Oporto, which has an irregular geometric shape as can be seen in Figure 6-1.

Figure 6-1: Details of the irregular shape of the room in Oporto.

The anechoic chamber in Helsinki has a free measure of about 4.2m (length of about 4.2m; width of about 4.2m; height of about 4.2m), with an absorption wedge length of 80cm, the room is assumed to be anechoic at frequencies above 100 Hz (Lokki, Patyen, & Pulkki, 2008). The studio in Oporto has a length of about 3m, a width of about 5.2m with a maximum height of about 2.8m. The studio in Dublin has a length of about 3.8m, a width of about 2.8m and a height of about 2.15m.

6.2.1 IACC measures for assessing spaciousness

In all three rooms, a frontal speaker (Genelec 1029 in Dublin and Oporto and a Genelec 8030A in Helsinki) was placed at 0° on-axis to a dummy-head, (a Neumann KU-100 except in Helsinki, which was a Cortex Electronic Manikin MK1). This speaker was intended for the playback of the direct sound. Another same-model speaker was placed at an angle of 60° to the right from the dummy head with the intent of simulating a single lateral reflection which constituted the indirect sound. The choice of angle for the simulated lateral reflection was based on Ando's findings for preferred echo directions which are more effective from about 30° to 90° with the maximum preference ratings around 60° (Ando, 1977). Both speakers were fed the same 15th order Maximum Length Sequence (MLS) signal, generated using the MLS Signal Generator plug-in of the Aurora 4 plug-in bundle (Farina, 2007). The Aurora 4 plug-in bundle makes it possible to retrieve the binaural impulse responses from the generated MLS signal played back through loudspeakers. Using the Deconvolve Multiple MLS module within the Aurora 4 plug-in facilitates the retrieval of the binaural impulse responses; the binaural impulse responses are later used in the Acoustical Parameter module of Aurora 4 to calculate the IACC in compliance with ISO 3382 (Farina, 2007). The use of MLS signal and the Aurora 4 software was carefully chosen. The very experienced audio researcher, Angelo Farina, has suggested (personal communication: Farina, 2014) that IACC measured using MLS test signals can be used to estimate the auditory spaciousness that would be experienced when listening to music if some of the known problems with MLS signals are not invoked:

- Non-linear artifacts arising from audio equipment distortions are avoided.

- The system signal-noise ratio is maintained at an adequate level for all measurements
- The clocking used in the recording and playback systems are matched
- All spurious disturbances (such as people movements) of the acoustical system are avoided.

Provided sources of distortion such as these are avoided, which was the case in experimentation conducted throughout the research for this thesis, then the IACC values measured using MLS test signals will be indicative of the IACC values that would be generated using music signals.

The frontal speaker played the MLS signal always at a consistent level; the laterally placed speaker signal was changed in level and delay relative to the frontally placed speaker signal. Both speakers were placed at a distance of 1m from the centre of the dummy head. A schematic of this setup can be seen in Figure 6-2.

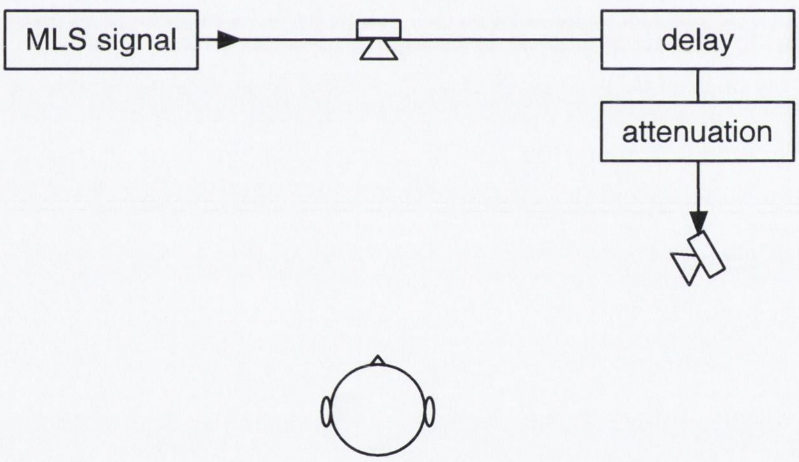
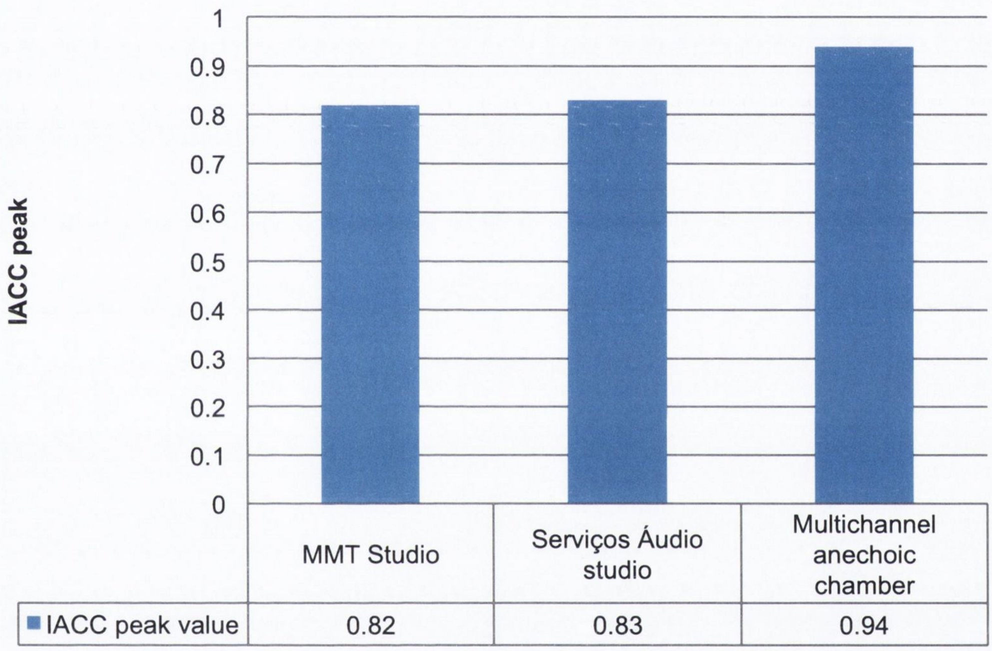


Figure 6-2: Direct sound with a single simulated lateral reflection. This setup is similar to the setup used by Ando and Kageyama (1977)

For the laterally placed speaker, the intent was to recreate a single lateral reflection in which the level changed in steps of 2dBFS from -18dBFS to 0dBFS. This change in level was repeated for different delay times which were introduced for the lateral speaker signal. The resultant direct and indirect composite signals were recorded using the dummy-head for later IACC peak value measurement. All the signals recorded were recorded at a sampling frequency of 44.1 kHz with 24bit resolution, using a MOTU-traveler mk3 sound card and a Digital Audio Workstation (DAW) on a laptop personal computer.

As a control, the IACC peak value was always measured from the frontal (0° on-axis) speaker, without the laterally positioned, indirect sound, speaker. The peak values achieved for this case, for each of the rooms used, can be observed in Table 6-1.

Table 6-1: IACC peak values in each of the rooms used for measurement. Frontal direct sound with no reflection.



6.2.2 Results

The results obtained from the IACC measurements, which were all calculated using the Acoustical Parameter module of the Aurora 4 using IACC full (*i.e.* $t_1 = 0$ and $t_2 = \infty$, see Equation 4.2) (Farina, 2007), indicate that an increase in the amplitude of the reflection resulted in a drop in the IACC peak value. The relative amplitude of the reflection to the direct sound is of major importance for the perceptual impression of spaciousness, while a delay introduced in the reflection does not cause significant changes in the overall IACC peak, which is used as an index for the impression of spaciousness. The results from this experiment are in agreement with the findings of Barron and Marshall (1981). From Figure 6-3 to Figure 6-5, several graphs can be seen with the results obtained for 10, 30 and 50ms delays introduced for a single lateral reflection. When observing the values for IACC (*A* graphs), the horizontal axis corresponds to the relative change in level in dBFS of the indirect sound speaker, and the vertical axis corresponds to peak value of the normalized IACC between 0 and 1. When observing the values for ITD (Inter-aural Time Delay) (which can be used as an indication of Image Shift – *B* graphs following), the vertical axis corresponds to the *tau* value in milliseconds measured with a 1ms window. In this case the *tau* shift, if existent, would be to the right, since the early reflection is positioned to the right of the centre front. From the graphs, it is can be seen that there is a similar drop in the IACC peak as the level is increased for each of the different delays introduced, indicating, as previous work had done (Barron & Marshall, 1981), that the change in time of arrival of an early reflection does not change the overall spatial impression as much as the change in level of the early reflection does. In the graphs for different rooms, it can also be observed that the trend of the drop in peak value with increasing reflection level is consistent for different test environments, and is thus independent of the room in which the measurement was being made. This is of importance for the tests to be undertaken because it indicates that further investigation of the control of the perception of spaciousness can be examined even using echoic rooms.

A

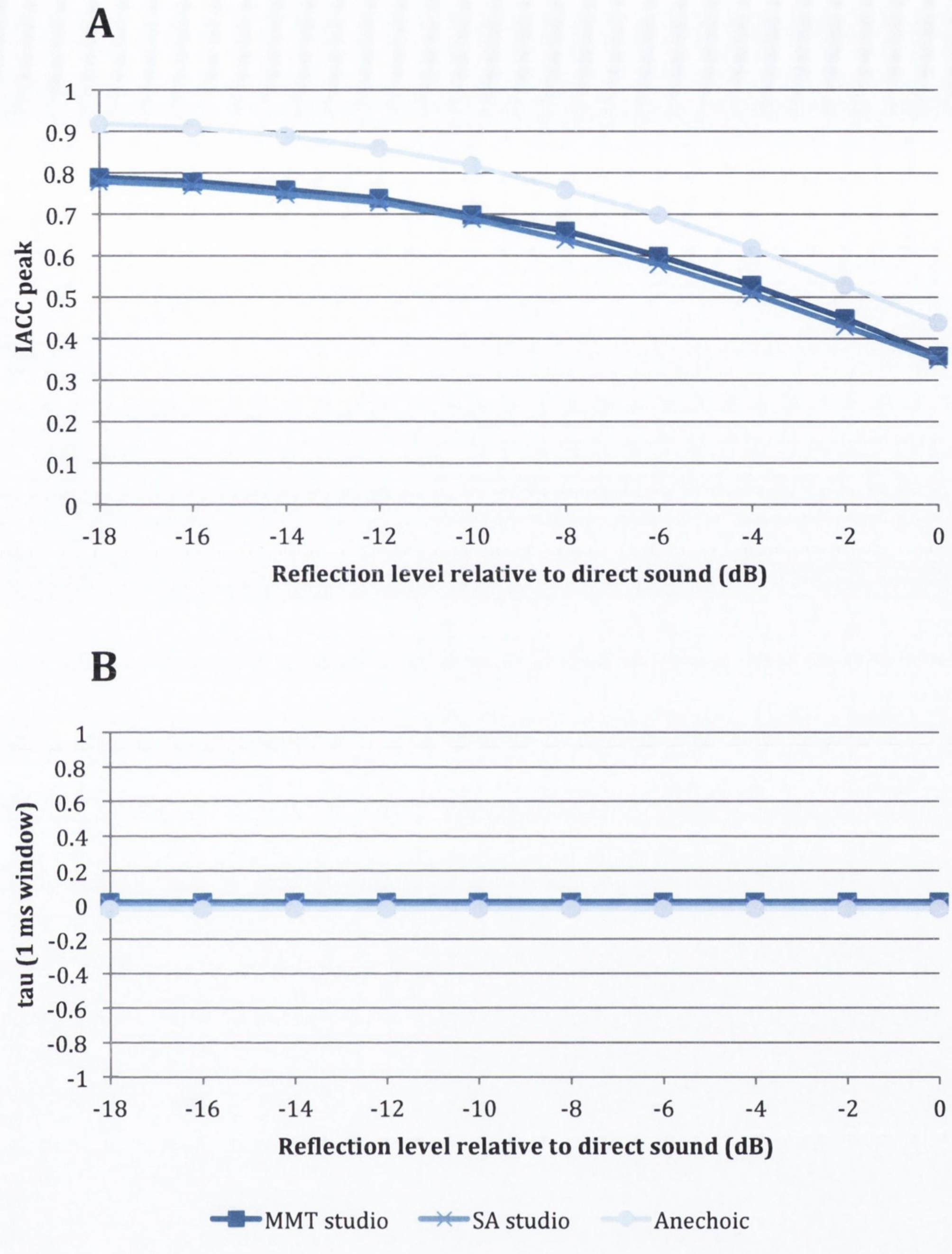


Figure 6-3: IACC (A) and ITD (B) measurements with early reflection at 10ms. In ITD (B) there is no image shift; all the measurements, made in the different rooms, indicate no change of ITD.

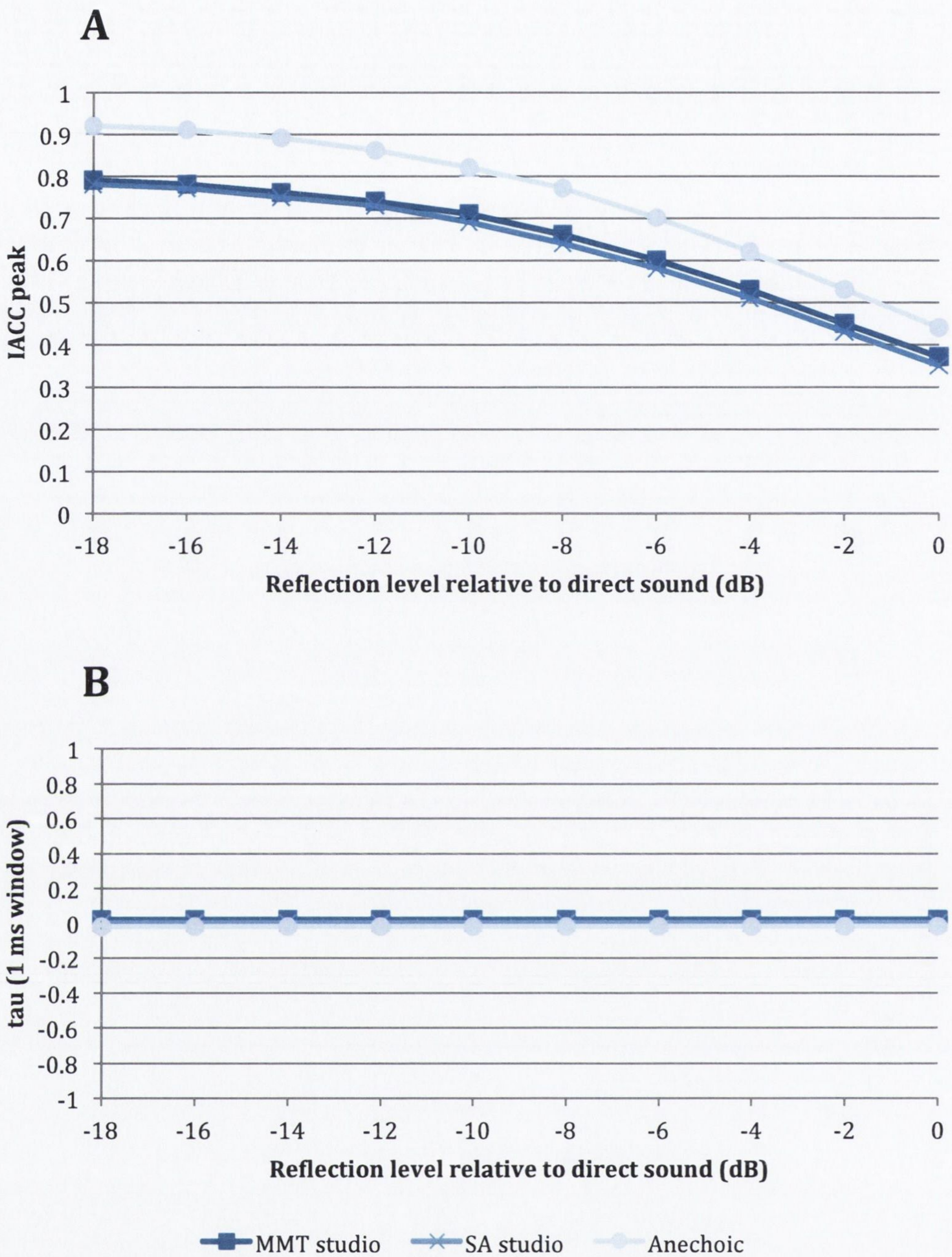


Figure 6-4: IACC (A) and ITD (B) measurements with early reflection at 30ms. In ITD (B) there is no image shift; all the measurements, made in the different rooms, indicate no change of ITD.

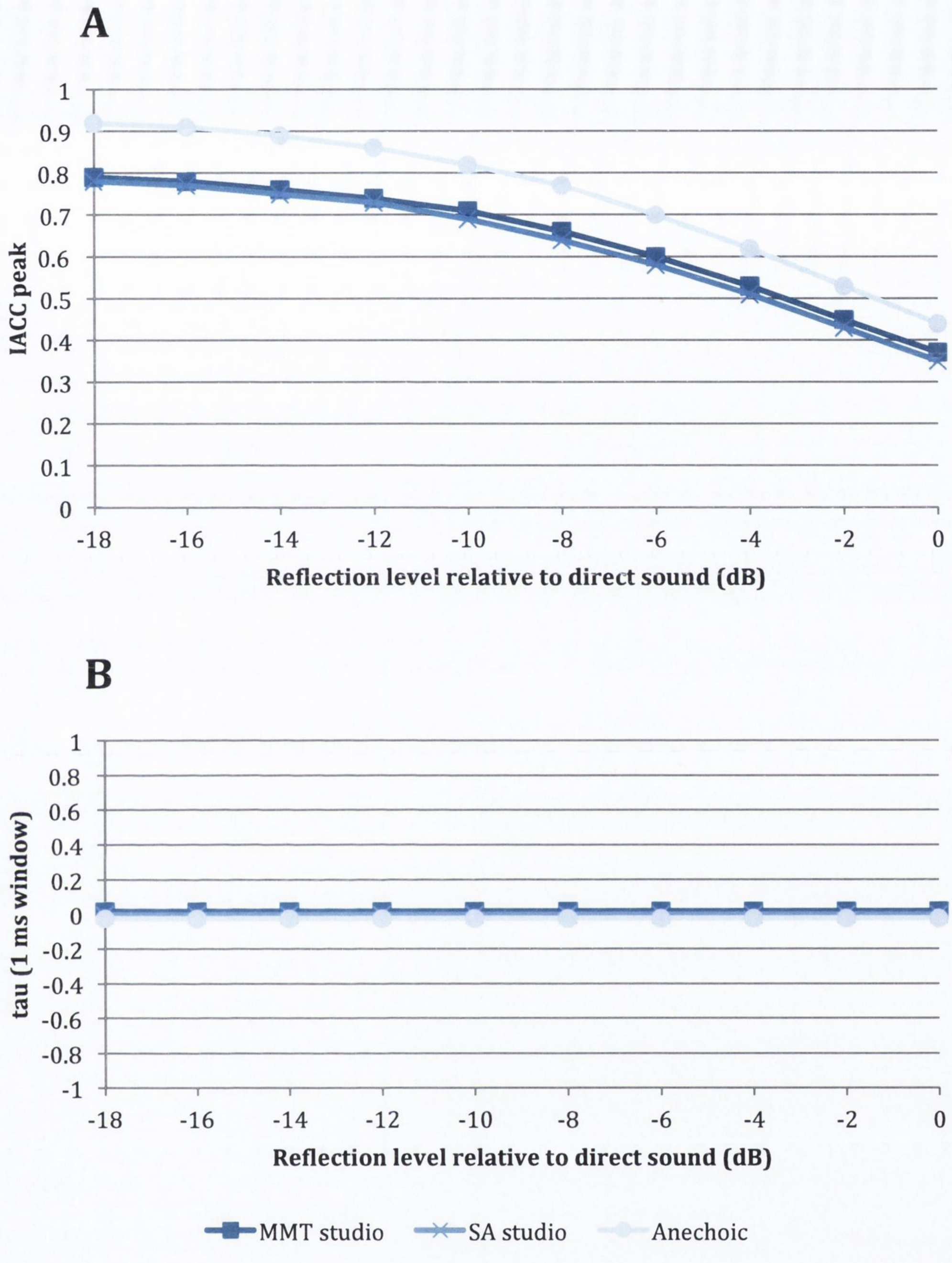


Figure 6-5: IACC (A) and ITD (B) measurements with early reflection at 50ms. In ITD (B) there is no image shift; all the measurements, made in the different rooms, indicate no change of ITD.

In order to see if the variations in IACC were indicative of changing correlation and not due to changes in the normalising power, an experiment reported in APPENDIX II – IACC Results for Centre Front Direct and Indirect Components was conducted in the Oporto studio. Repeating the experiment described in Section 6.2.1 but having the simulated reflection coming from front centre (0°) (*i.e.* positioned from the same place as the direct sound), shows no changes of IACC peak value which maintained practically the same figure (IACC = 0.83) for all simulated delayed reflections and relative levels used. From the results it is concluded that the IACC changes are indeed related to inter-aural differences introduced from the lateralization of the early reflection.

6.3 Summary

The studies undertaken have verified that the measurement trends observed for IACC for changing reflection amplitude are location independent. This was achieved through the repetition of the IACC generation and recording process in different environments, followed by a comparison of the results achieved. It was verified that the IACC peak value drop was indeed related to the lateralization of the early reflection (see APPENDIX II – IACC Results for Centre Front Direct and Indirect Components). All the IACC measurements were made using the full bandwidth of the signal (see APPENDIX III – IACC Results for Octave Bands, Full Bandwidth, and Full Bandwidth with A-Weighting Filtering, where it is shown that spectral limitation of the test signals does not affect IACC alteration trend). Only one simulated reflection was used in this setup since it has been established by Ando & Gottlob (1979) and re-stated by Blauert (1997) that spaciousness is not significantly affected whether one or multiple lateral reflections are used (see also APPENDIX IV – IACC Results for Single and Multiple Reflections, which demonstrates that the IACC alteration trend is not affected by the number of synthetic reflections used).

Auditory scenes with different IACC values were generated and recorded in particular test environments in order to explore the effect of room acoustic details on IACC, and therefore on perceived spaciousness. Given the results achieved, it has been concluded that further studies of perceived spaciousness can be examined in echoic rooms. Therefore, investigation of the perceived spaciousness of different recording formats and reproduction systems can be undertaken in echoic environments without loss of IACC trend accuracy. By being able to objectively assess the perceived spaciousness of a sound field, a better understanding, and therefore a greater control potential, of the overall spaciousness experienced in a listening environment can be obtained. This invites the question: What are the factors which influence the spatial impression created by a reconstructed sound field using different recording and reconstruction formats?

Using MLS signal to obtain the binaural impulse signals which are used for IACC measurement can be used to estimate the auditory spaciousness that would be experienced when listening to music.

Part of the work presented in this chapter has been published as a conference proceeding, further details can be found in APPENDIX VI – Resulting Publications and Presentations.

7 INFLUENCE OF DIFFERENT MICROPHONE ARRAYS ON IACC AS AN OBJECTIVE MEASURE OF SPACIOUSNESS

7.1 Introduction

In the previous chapter it was established that similar IACC trends were evident for the addition of a single lateral reflection to a direct sound, independent of the rooms used for the experimental setups. These findings were crucial for the following work, since access to an anechoic chamber was limited. After establishing this, experiments were conducted to compare the spaciousness reconstruction effectiveness of microphone arrays for stereo and surround formats, using IACC as an index of perception of spaciousness.

Thus, using loudspeakers to simulate direct sound and a single lateral reflection, which could be digitally controlled in level and time of arrival, it was possible to use the setup discussed in Section 6.2 in any room of choice in order to study spaciousness control. This was undertaken in order to compare IACC measurements for similar direct/indirect sound field configurations in a primary recording setup with the IACC measurements for a reconstructed sound field in a secondary reproduction environment, where stereo and 5.1 were used as default sound field reconstruction systems. By setting up different microphone arrays, while changing parametric details in their setups, it was possible to record the simulated direct sound and single lateral reflection sound fields. These recorded sound fields were then reproduced over stereo and 5.1 systems where a dummy-head microphone system placed at the sweet spot recorded the reconstructed sound field ear signals. The IACC measurements obtained from the binaural signals of the dummy-head revealed changes which depended on the different recording microphone arrays, and on changes in details made to these arrays. By analysing the so-achieved results, it is possible to develop an appreciation of how the ultimate listener perception of spaciousness might be controlled using either the choice of microphone array, or the parametric choices made for such arrays.

7.2 Experimentation

The experimental setup used for investigating the influence of choice of recording microphone array on reconstruction sound field spaciousness involved two loudspeakers, one for frontal, direct sound, and one for a simulated single lateral reflection angled at 60°, which together define a primary sound field which was to be recorded suing various microphone arrays. Control in level and delay for the simulated single lateral reflection was made possible using a digital audio workstation which fed the loudspeakers with an MLS signal, in a manner similar to that described in Section 6.2.1. A schematic of the organization of the experimental setup is represented in Figure 7-1. All the microphone arrays were placed at a distance of 1 meter (measured from the centre of the array) from the speakers.

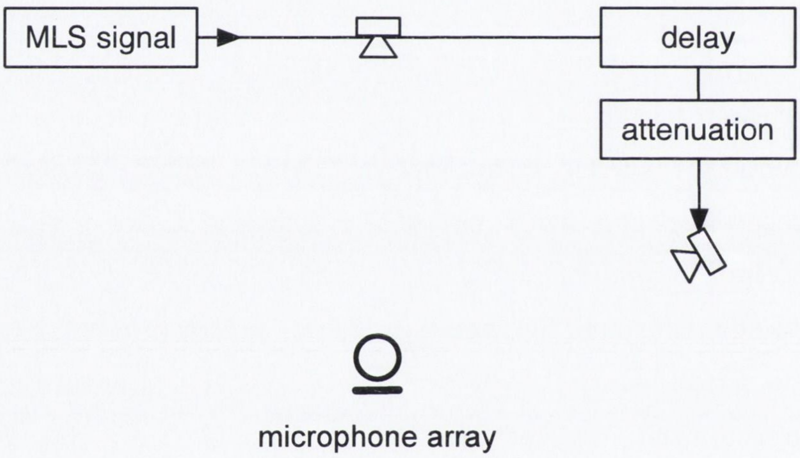
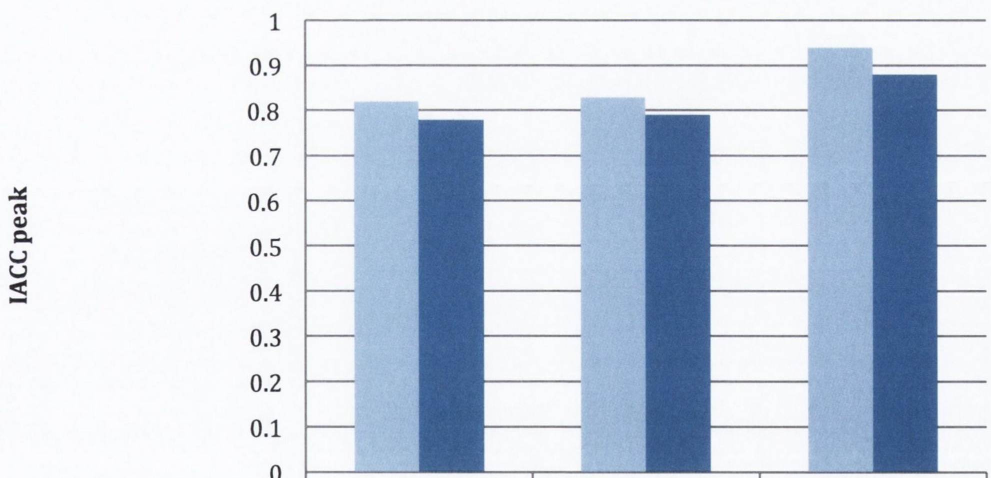


Figure 7-1: Primary (recording) environment configuration.

Starting with XY and ORTF stereo microphone techniques, the sound field which resulted from the frontal and lateral loudspeakers was recorded. The reflected sound loudspeaker signal was varied in steps of 2dBFS from -18dBFS to 0dBFS, which was repeated for different signal delays. These recordings were then played back through a 2-channel stereo reproduction system with the Left and Right speakers positioned at ±30° from centre front, and a dummy-head microphone system placed at the listening 'sweet spot'. A control measurement was first made for the stereo reconstruction configuration in which an MLS signal was fed to both loudspeakers of the stereo playback system. This was done in order to ensure that the peak of IACC was at *tau* = 0, demonstrating that the dummy-head was indeed at the sweet spot (*i.e.* signals from both speakers were arriving at both ears with no time difference), and also in order to evaluate if there was a difference between the IACC peak value measured for direct sound (played back by an actual centre stage loudspeaker), and the IACC peak value for a phantom centre image created by the stereo playback system. In the first stage of these experiments, XY (with an included angle of 90°) and ORTF stereo recording formats (see Section 5.3.1) were used employing a matched pair of microphones (Rode NT-5); the experiment was repeated in each of the rooms used and is described in Chapter 6.

The IACC peak value in each room for just the direct sound is shown in Table 6-1. Table 7-1 shows a comparison of IACC peak values in each room for frontal direct sound and centre front phantom image. This comparison indicates that stereo stage phantom images demonstrate an increased spaciousness impression as compared to the spaciousness impression for physical sources at the same centre front locations. The reason for this might be to do with the cross-talk which occurs at the ears of a listener and which is inherent to stereophonic reproduction (see Section 5.3).

Table 7-1: IACC peak values in each of the rooms used for measurement. Comparison between frontal direct sound with no reflection and centre phantom image, with no reflection.



	MMT Studio	Serviços Áudio studio	Multichannel anechoic chamber
Frontal direct sound	0.82	0.83	0.94
Centre phantom image	0.78	0.79	0.88

7.2.1 Results

IACC results were measured for the reconstructed experimental sound fields using a stereo recording made with an XY microphone configuration and are shown in Figure 7-2, Figure 7-3 and Figure 7-4 for 10ms, 30ms and 50ms 'reflection' delays, respectively. It can be seen that for this recording technique, the IACC peak value drops as the amplitude of the reflection is increased. However, the drop in the IACC peak for XY stereo reconstructed sound fields is not as significant as the results shown in Figure 7-5, Figure 7-6 and Figure 7-7 which were achieved using an ORTF microphone configuration for recording.

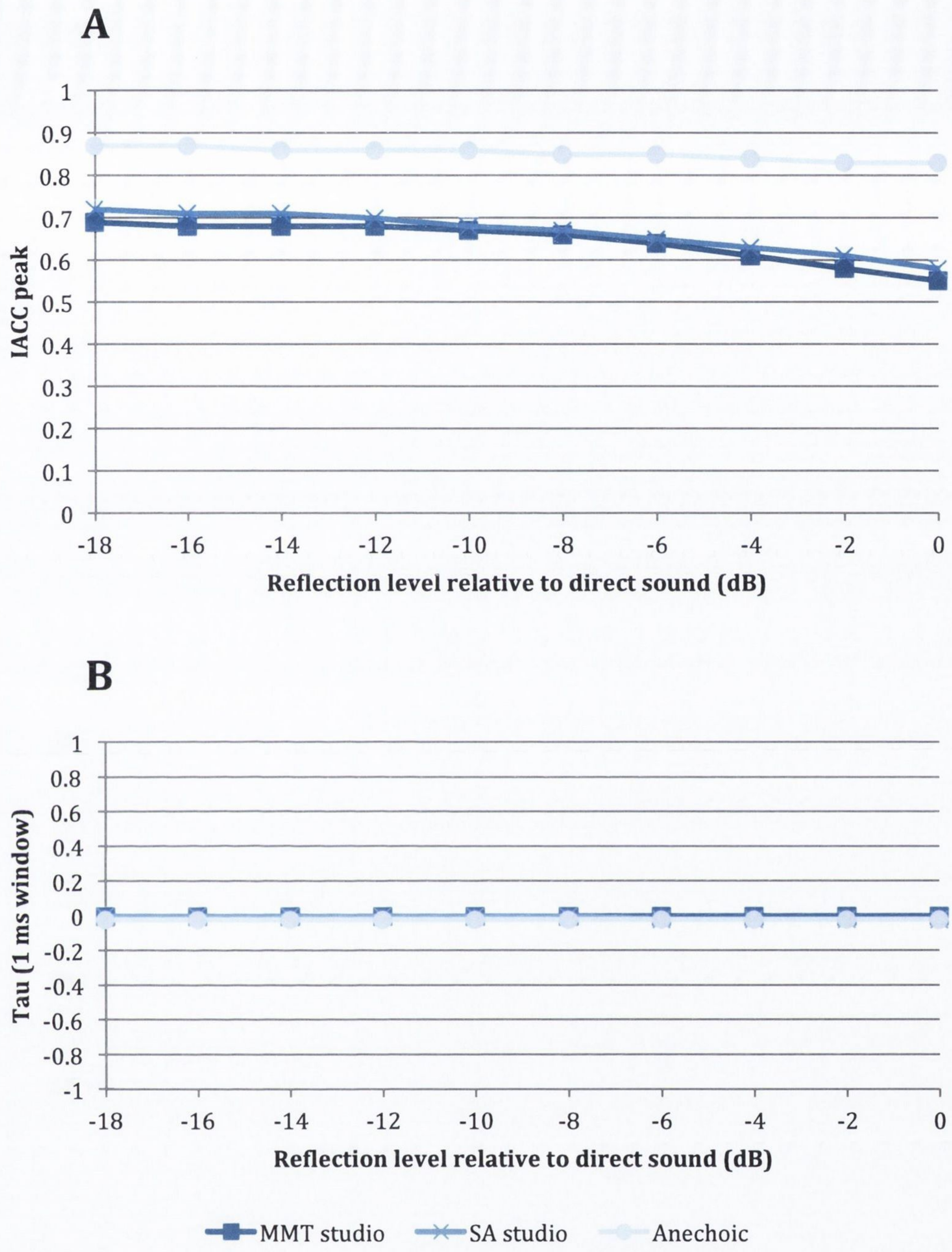


Figure 7-2: IACC (A) and ITD (B) measurements of XY stereo recording technique with an early reflection at 10ms, using stereo reconstruction. In ITD (B) there is no image shift; all the recordings, made in the different rooms, indicate no change of ITD.

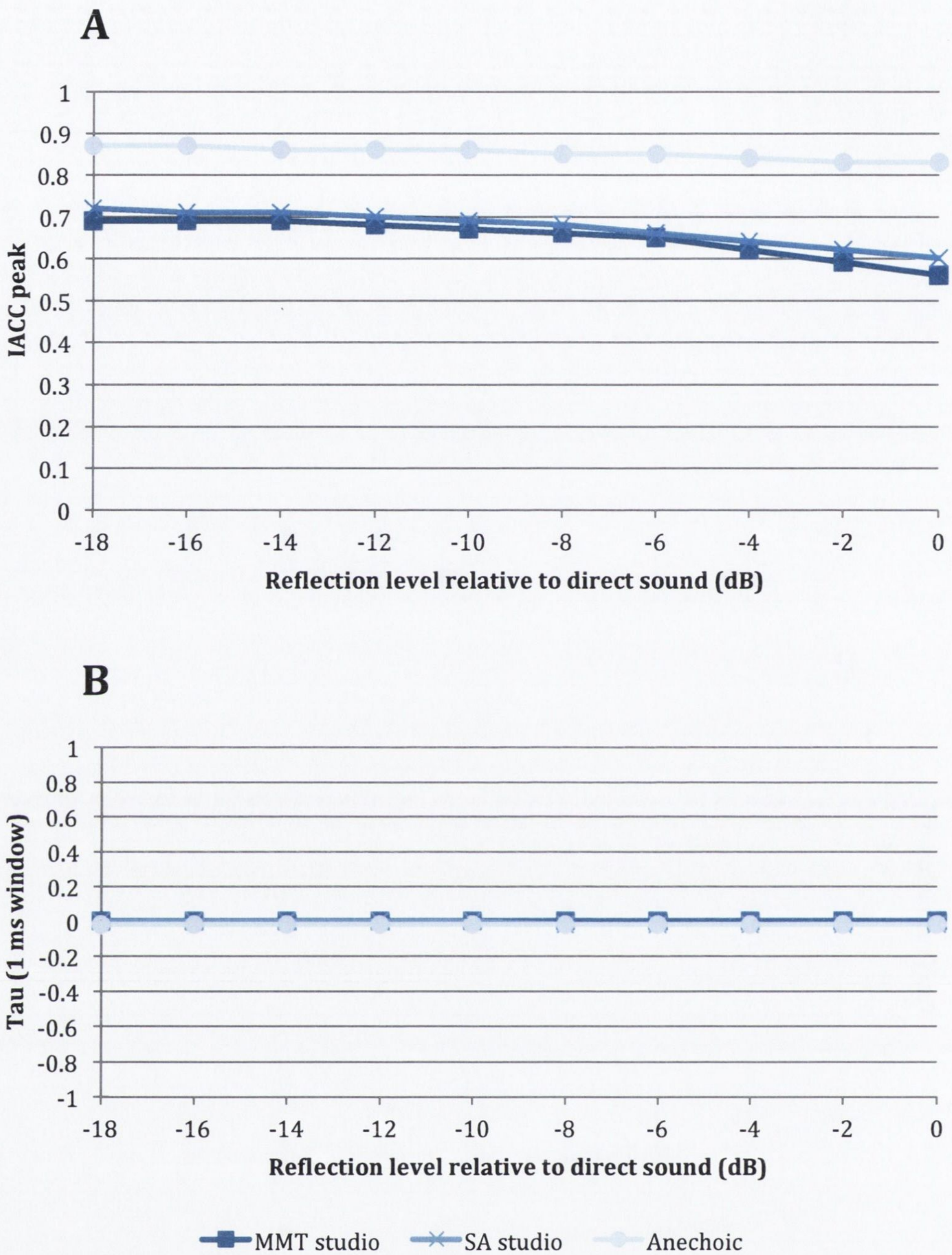


Figure 7-3: IACC (A) and ITD (B) measurements of XY stereo recording technique with an early reflection at 30ms, using stereo reconstruction. In ITD (B) there is no image shift; all the recordings, made in the different rooms, indicate no change of ITD.

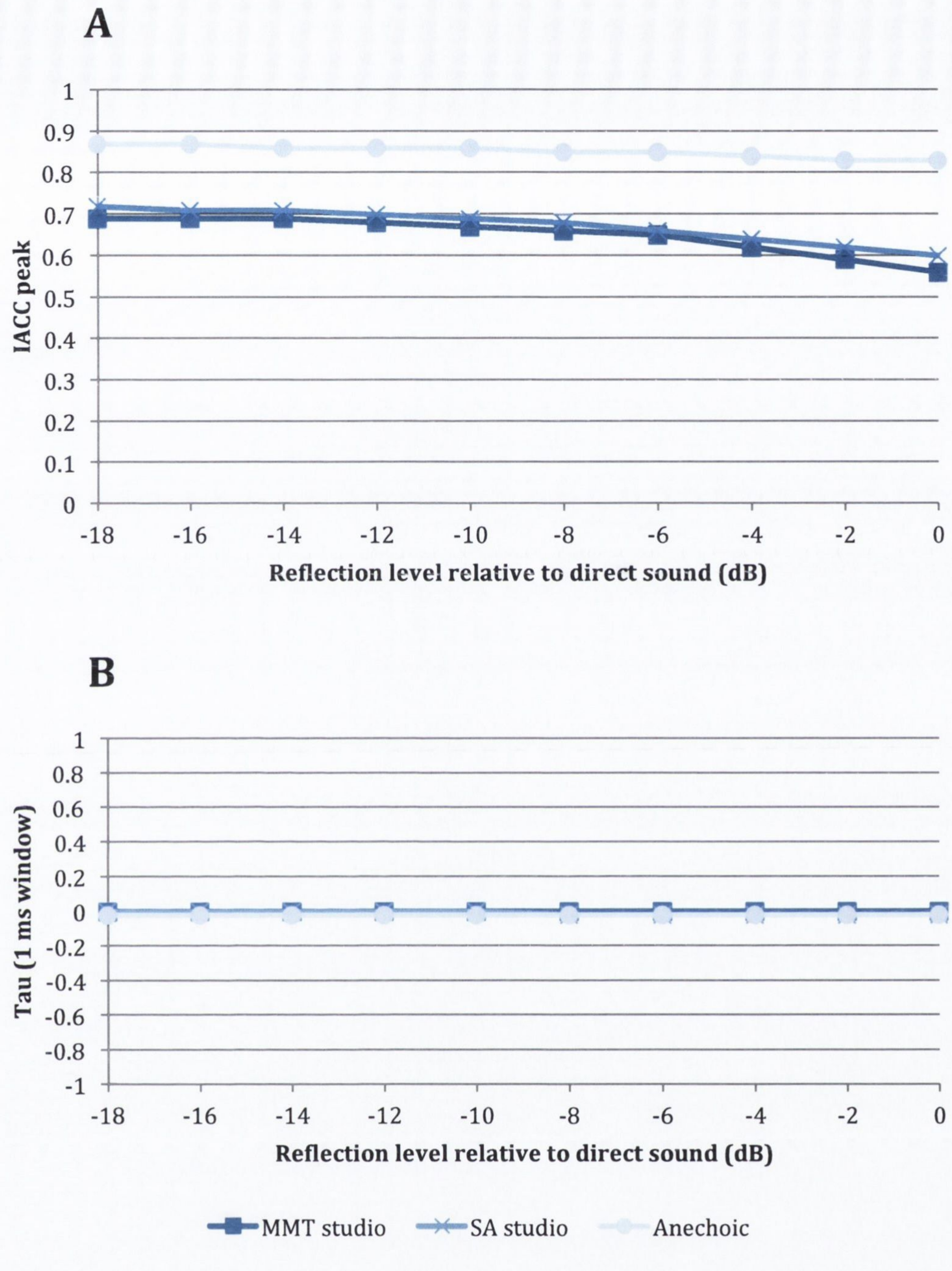


Figure 7-4: IACC (A) and ITD (B) measurements of XY stereo recording technique with an early reflection at 50ms, using stereo reconstruction. In ITD (B) there is no image shift; all the recordings, made in the different rooms, indicate no change of ITD.

A

B

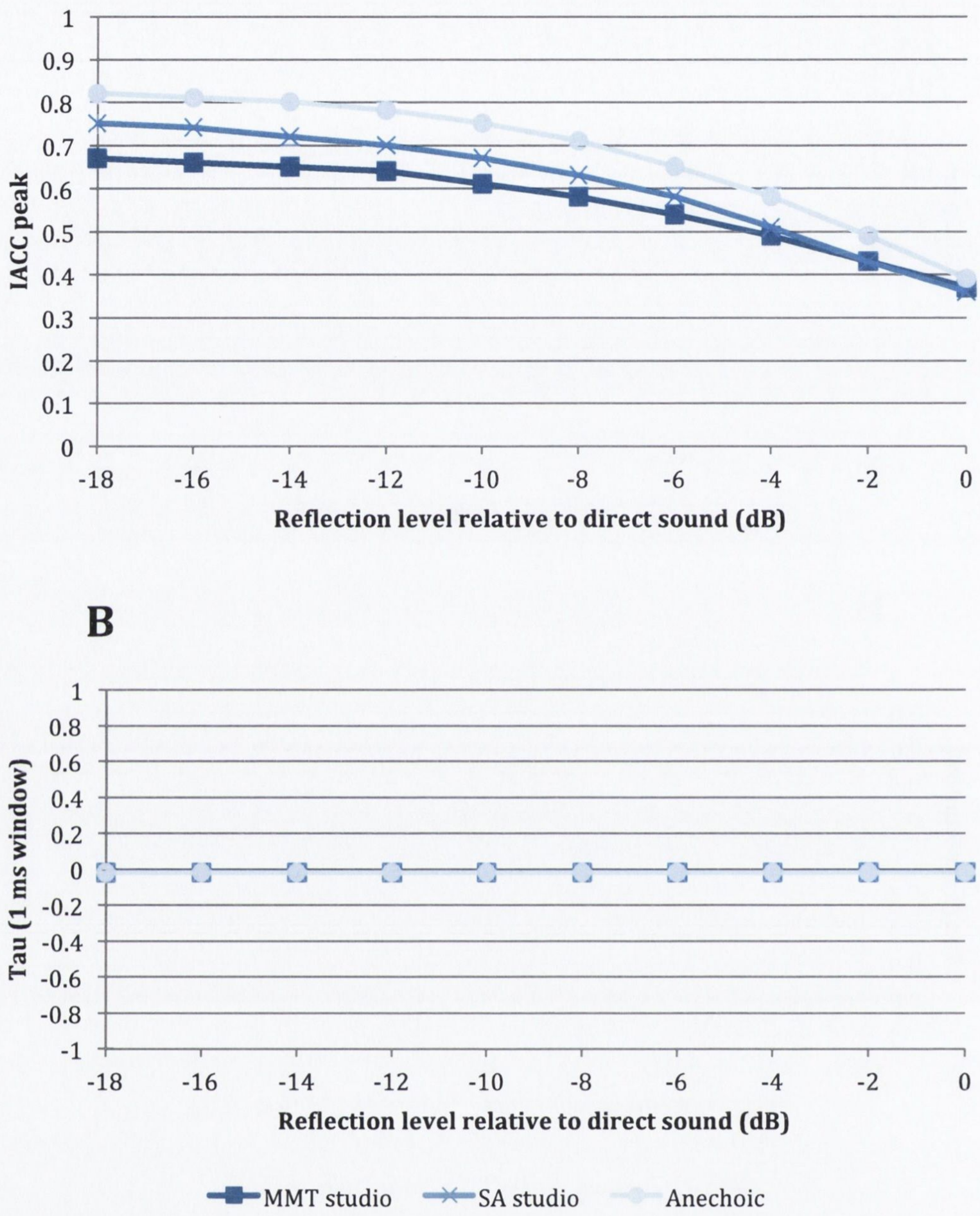


Figure 7-5: IACC (A) and ITD (B) measurements of ORTF stereo recording technique with an early reflection at 10ms, using stereo reconstruction. In ITD (B) there is no image shift; all the recordings, made in the different rooms, indicate no change of ITD.

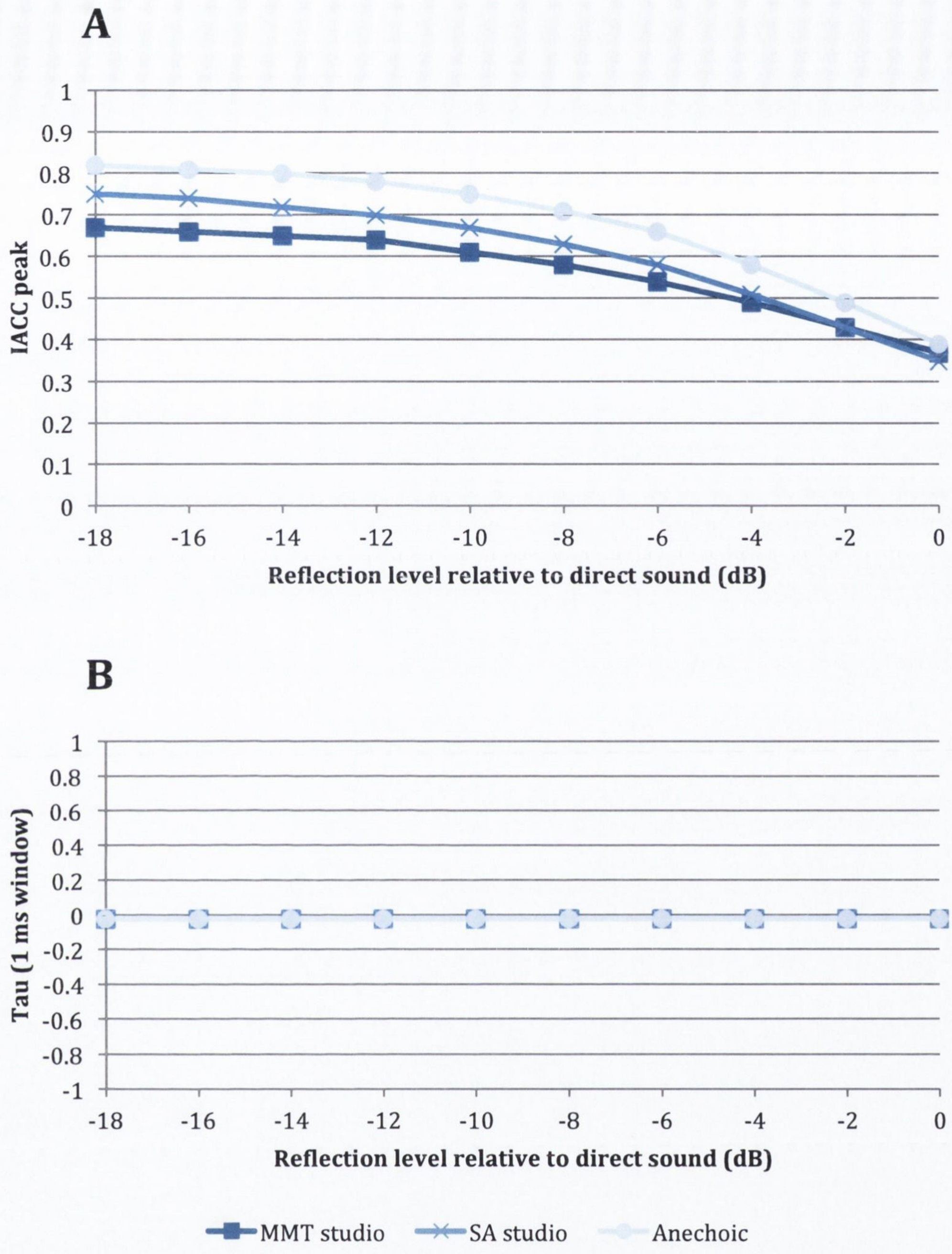


Figure 7-6: IACC (A) and ITD (B) measurements of ORTF stereo recording technique with an early reflection at 30ms, using stereo reconstruction. In ITD (B) there is no image shift; all the recordings, made in the different rooms, indicate no change of ITD.

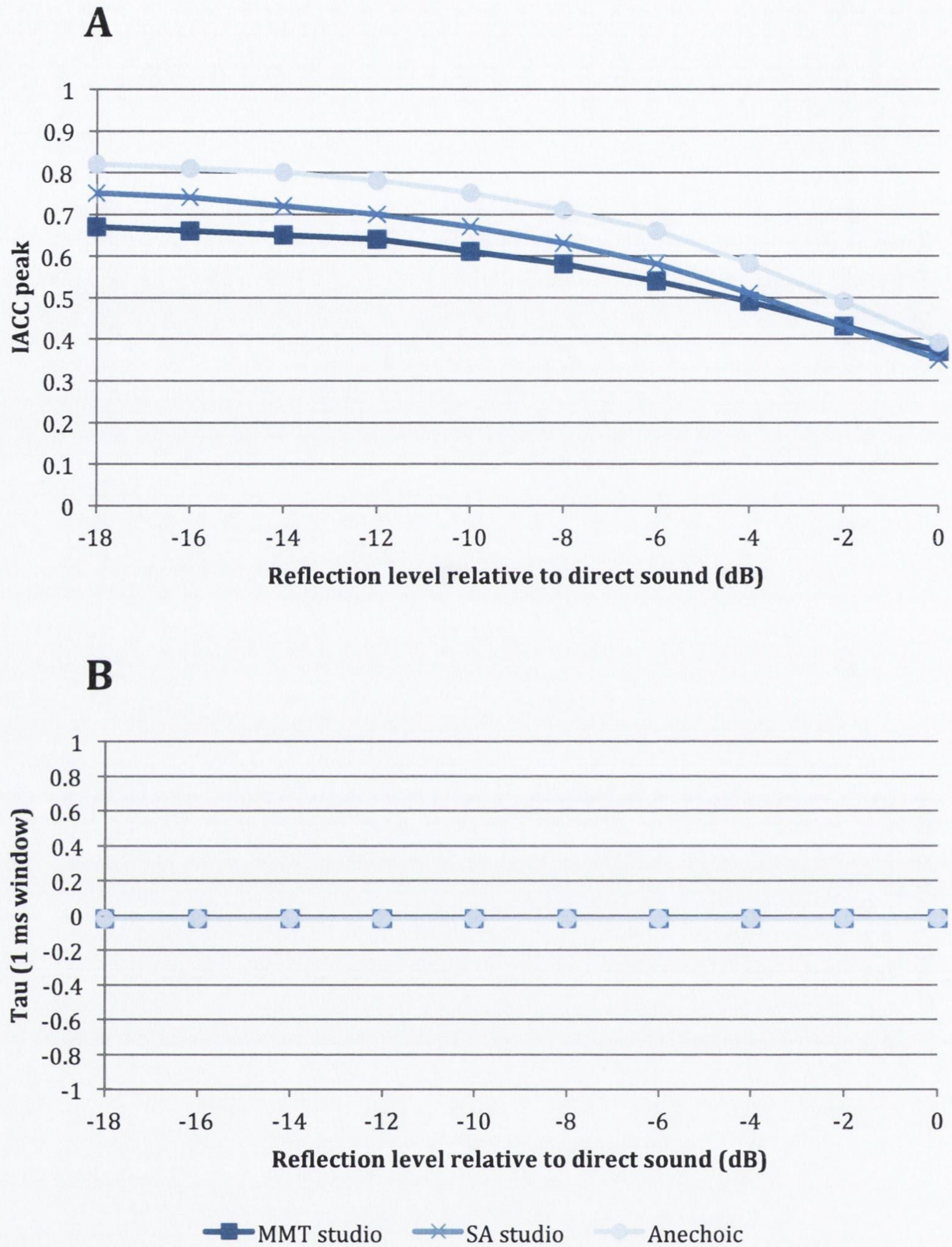


Figure 7-7: IACC (A) and ITD (B) measurements of ORTF stereo recording technique with an early reflection at 50ms, using stereo reconstruction. In ITD (B) there is no image shift; all the recordings, made in the different rooms, indicate no change of ITD.

The IACC measurement method used allowed an indirect assessment of the spaciousness impression experienced by stereo reconstructed sound field listeners. There are clear differences between the reconstructed spaciousness impressions using coincident and spaced stereo recording techniques, as demonstrated by comparing XY and ORTF reconstructed sound field IACC values. The introduction of inter-microphone delay for ORTF recording leads to reduced IACC peak values, with an attendant increase in the sense of spaciousness, which is commonly expected, but the results achieved do confirm the identification of the necessary conditions for auditory spaciousness perception, that being that if listeners are presented with significantly dissimilar ear signals there will be an increased sense of 'space' or 'roominess' experienced.

7.3 Other microphone array experimentation

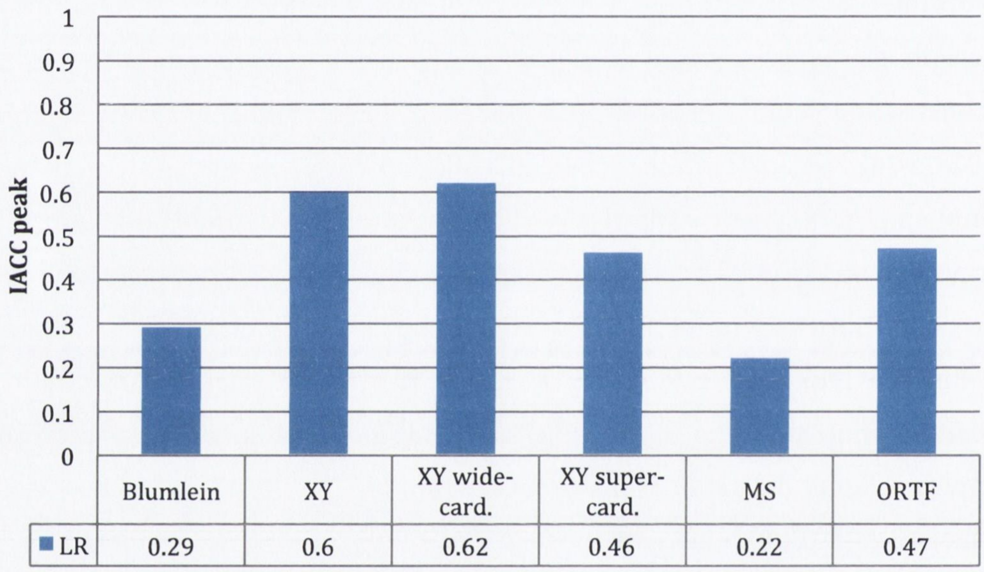
Following the experimentation carried out using XY and ORTF stereo recording reconstructions, further work was conducted using different microphone arrays in order to see how IACC peak values varied between the reconstructed sound fields for both 2-channel stereo and 5.1 (ITU-R BS.775-1, 1992-1994) reconstruction. The different microphone arrays used were Blumlein, XY, XY with wide cardioid, XY with super-cardioid, (all of the XY arrays had an included angle of 90°), MS, AB, Double MS, Bruck Array a.k.a. KFM Surround (SCHOEPS GmbH, 2013a), Soundfield and OCT-Surround (see Chapter 5). Since it had been established that a drop in IACC peak values generally occurs as the amplitude of the reflection rises in relation to a direct sound, a simplification of the experimentation was adopted by just using an early lateral reflection at -4dB in level with a 50ms delay time, relative to the direct sound. Also, since it had been established that the rooms used did not change the trends in IACC peak value decrease, the experimentation was conducted, for practical reasons, only in Oporto.

7.3.1 Results

The results obtained are shown in Table 7-2. Given that all other parameters are the same, it can be appreciated that a change in microphone configuration allows for a change in IACC peak value, which indicates a difference in the perceived spaciousness of listener experience. Table 7-3 presents the results obtained from the variation of microphone spacing in AB and OCT microphone arrays. For the case of AB stereo format, two different microphone capsules were used: pressure (omnidirectional) and velocity (figure-of-eight). It can be seen that, for AB stereo, as the spacing between capsules is increased the IACC peak value decreases, indicating an increased spaciousness listener experience. However, while this is true for AB microphone arrangements, in the case of OCT microphones the capsule separation between the left and right pointing hyper-cardioids does not introduce a significant change to the IACC peak value. The results achieved also show that the choice of microphone capsules for AB arrays significantly affects the IACC peak value.

Table 7-2: IACC results for different microphone arrays when played back through 2 channel stereo (A) and 5.1 (B).

A



B

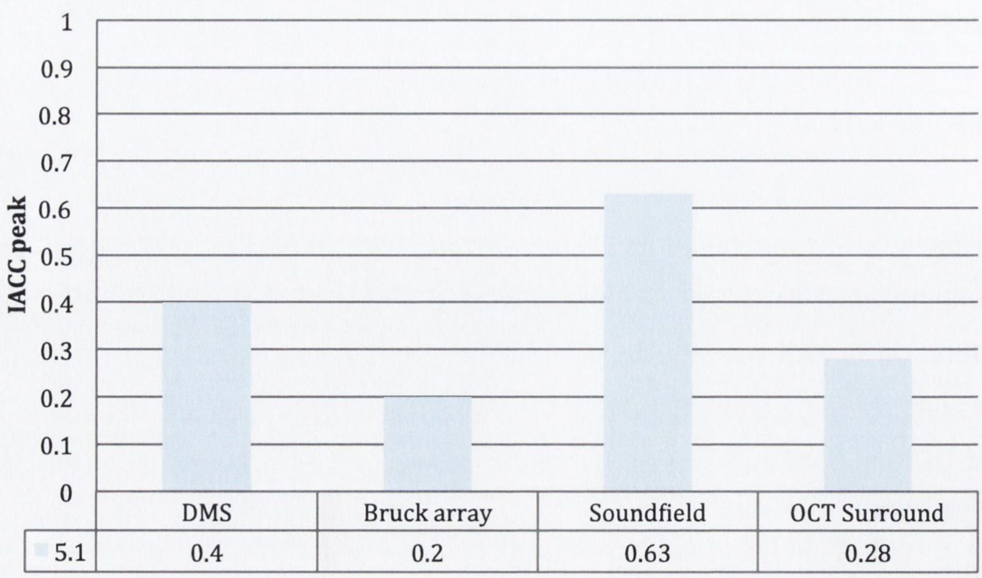
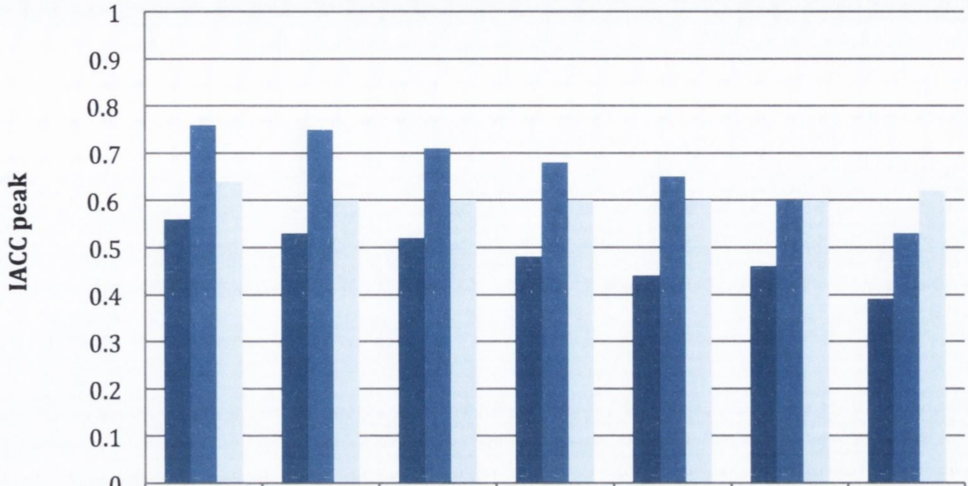


Table 7-3: IACC results for different AB spacings using pressure and velocity microphones, and for OCT, when played back through 2 channel stereo.



	40 cm	50 cm	60 cm	70 cm	80 cm	90 cm	100 cm
AB - Omni	0.56	0.53	0.52	0.48	0.44	0.46	0.39
AB - Fig 8	0.76	0.75	0.71	0.68	0.65	0.6	0.53
OCT	0.64	0.6	0.6	0.6	0.6	0.6	0.62

7.4 Summary

The effect of recording format details on perceived spaciousness has been investigated. A physical measure, IACC, which correlates with perceived spaciousness, was used as a spaciousness index thereby allowing quantitative assessment of qualitative impression generated in a variational study of spaciousness.

It was deemed to be important to verify that the measurement technique used was location independent by repeating the IACC generation and recording process in different environments and comparing the results arrived at. The similar measurement trends achieved for XY and ORTF in all measurement environments suggested that the comparative tests to be performed were 'environment neutral'.

Audio scenes with different IACC values were then generated and recorded in particular test environments using different stereo and surround microphone techniques in order to explore the effect of recording technique details on IACC, and therefore on the perceived spaciousness of the reconstructed sound fields.

From the results achieved, it can be concluded that microphone parametric variation can be employed to facilitate spaciousness control for reconstructed sound fields by sound engineers and producers. Of particular note is the spaciousness control potential suggested by the IACC results, indicating that spaciousness impressions can be varied by producers in order to enhance the spatial audio impression created by stereo, and also by 5.1 surround sound system reconstruction. The understanding of how to create sound field reconstruction with controllable spaciousness can help promote the use of surround sound systems, in particular 5.1, for recorded music presentation which will make engineers and producers more confident about the use of such systems, since it can be demonstrated that a controllable impression of spaciousness permits enhancement of content presentation.

Part of the work presented in this chapter has been published as a conference proceeding and presented as a poster session, further details of which can be found in APPENDIX VI – Resulting Publications and Presentations.

8 SPACIOUSNESS CONTROL IN STEREO AND 5.1

8.1 Introduction

Stereo techniques, as reviewed in Chapter 5, became the accepted norm for audio recording and reconstruction approximately 30 years after their invention. During this gestation phase various refinements were introduced until eventually the technological developments were such as to allow commercial exploitation. Stereo became accepted as an improvement in terms of sonic reconstruction due to technological developments that facilitated a considerable advance in aesthetic control. Stereo recordings were not regarded as being a reconstruction of a physical sound field, but rather as an adaptable medium for artistic communication (as discussed in Chapter 2). The created sound was recognised as potentially having a perceptual impact which was under the control of the original artists, sound engineers, and producers.

Spaciousness is regarded as an important perceptual aspect of a sound field reconstruction (see Chapter 3, Section 3.2.1) and has been identified as one of the most significant factors in distinguishing good spaces for music performances. The audible effect of spaciousness, or at least the impression of a space, has been used to some extent in movies. The fact that cinema has been using multiple channels and loudspeakers to deliver such audible effects has contributed to making it easier to engage audiences in the visual images. Gary Rydstrom (Sonnenschein, 2001, p. 170) explains that the synergy between sound and image, in films, might, for example, have visuals telling one thing and the sound telling something completely different, like the off-screen world (*i.e.* ambience, space, etc.). The use of surround sound channels in cinema will help create a variety of aural movements and (re)-creation of acoustical spaces which will help break through the 2-D film image to create a 3-D visual space (Sonnenschein, 2001).

The use of 2-channel stereo can deliver some impression of the acoustic properties of a primary recorded sound field. With more channels, as have become commercially available in formats such as 5.1, the control of cues that provide a perceptual impression of spaciousness can potentially be further manipulated and controlled. Despite the added ability which surround sound formats might provide in delivering impressions of the acoustic properties of a primary recorded sound field in comparison to 2-channel stereo, it is not necessary to represent all the physical details of a real sound field in a listening room; it is sufficient only to provide key cues to elicit a recollection or an emotion (Toole, 2008).

In a simplistic form, spaciousness can be presumed to be dependent on the ear signal difference. Sound energy that reaches the listeners by paths other than those of the direct sound will contribute to the perceptual impression of spaciousness, but so also will choices in microphone arrays and details. In Chapter 6 the influence of a room environment with a setup comprised of a controlled direct sound and a single lateral reflection was measured using IACC as an objective parameter related to perceived spaciousness. From the results achieved it was confirmed that a strong lateral reflection influenced the IACC peak value independently of the listening room type, indicating a change in the sound field spaciousness that would be perceived. In Chapter 7 the influence of different microphone arrays and choices in microphone details on IACC was investigated. Here, the results indicated that changes in microphone array configuration details accounted for changes in perceived spaciousness impression.

8.2 Control of spaciousness

Is it possible to control the extent of spaciousness in recording reconstructions? What can be done to the stereo system L and R, and the surround system L-C-R-Ls-Rs, signals to cause ear signal differences?

Research undertaken in concert hall design has investigated the optimization of halls in terms of shape and construction materials, such that early reflections are produced at such a level at the listeners' ears so as to create the effect of a more spacious sound. Equally, reflection delay and level can be controlled by means of electronic devices. It is possible then to use room acoustics and signal processing, along with careful microphone selection and placement, to achieve and control the impression of spaciousness in recordings (Streicher & Everest, 2006). This is one way of controlling IACC, and therefore perceived spaciousness. As investigated by Marshall, Barron and others, lateral reflections contribute to perceived auditory spaciousness. Lateral reflections contribute significantly to ear signal differences, which is why there is a drop in IACC when these are presented at a listener position. However, for this thesis, and as is presented in this chapter, the primary concern is with listener perception when presented with two-channel stereo, three-channel stereo and 5.1 reconstructions at the listener position, and therefore with the manipulation of the recorded signals for the creation of the perceptual impression of spaciousness.

8.2.1 Stereo shuffling

Alan Blumlein first introduced the use of shuffling in the 1930's (Blumlein, 1933) when he invented stereo recording. The fact is that Blumlein thought of stereo not just as an L and R system, but rather as a Sum (Mid) and Difference (Side) signal, where L=M+S and R=M-S. Such a process is also known as Mid and Side, or MS, stereo (see Chapter 5). The summation or Mid signal (M) emphasizes all that is common on the L and R channels, while the difference or Side signal (S) includes all that is different between the L and R channels.

Blumlein's original idea included the generation of M and S signals from any L and R signals, which can then be processed independently and converted back into L and R for stereo listening. This encoding and decoding process allows for control of the level of M and S, and it was discovered that an increase in the S level resulted in an increase in the perceived width of reconstructed recordings' auditory perspective, even so far as to "put the stereo images beyond the left and right speakers". This increase in width is not without problems (Gerzon, 1986), and Blumlein, realizing this, devised a more sophisticated process for stereo widening – *i.e.* shuffling.

Shuffling is a frequency dependent width control, where equalization of the M and S signals can be implemented independently as a means to increase the perceived impressions of auditory width and spaciousness.

More recent work carried out by recording engineers and researchers such as David Griesinger (1985) and Michael Gerzon (1986) revived shuffling as a technique to enhance perceived spaciousness in stereo reproductions. Both these authors have suggested that adding a 6dB boost to the difference channel of a stereo recording below 700Hz with a low shelf filter will cause a perceived increase in the impression of spaciousness. Such an approach, as described, was exploited in this thesis and should not be confused with the Blumlein shuffler, since the use of the described shelf-filter is designed to increase stereo width and not to convert phase differences into amplitude differences (Gerzon, 1994)

8.2.2 Use of up-mixing techniques

Producers and audio engineers have frequently used up-mixing techniques as a means of 'transforming' stereo recordings for multichannel system playback. Such up-mixing is therefore not entirely a novelty; Gerzon introduced a technique, reported on in the 70's, by which he devised a form to translate stereo recordings for 4-channel reproduction (Gerzon, 1970). More recently, it has been possible to use commercially available up-mixing processors from companies such as SoundField (2013) and Waves Audio (2013), that make possible the up-mix of stereo to 5.1 or to any in-between reproduction setup (LRC, LRLsRs, etc.).

Using the approach outlined by Gerzon as a point-of-departure for the work reported here, an up-mixing process was developed to convert stereo recordings for 5.0 and 5.1 playback. The initial approach devised by Gerzon up-mixed stereo to a 4-channel setup, where the speakers were placed in a room in a "diamond" shaped configuration. A centre speaker was placed centre front to the listeners, and Left and Right speakers were placed at ± 90°. A surround speaker was also used, and placed behind the listening position, although Gerzon *et al.* reported that the use of such a speaker could be discarded in their setup (Gerzon, 1970). The centre speaker was dedicated to play the sum of the stereo channels (M signal), the Left and Right speakers would play the stereo channels as they were, while the surround (rear) speaker would play the difference of the stereo channels (S signal).

Gerzon's implementation was here adapted to up-mix stereo to 5.1 in a related manner. Using the ITU-R BS.775 1 (1992-1994) as the standard for the 5.1 playback system, the sum (M signal) of the stereo channels was fed to the centre channel and the difference (S signal) of the stereo channels was fed to the surround channels. The Left and Right speakers were fed with the stereo channels as they were. Also, the sum (M signal) was Low Pass Filtered at 80Hz and fed to the LFE channel. This added more low frequency information to the reconstructed sound field, but did not change IACC measurements when comparing 5.0 to 5.1, as will be seen later.

8.2.3 Development of a spaciousness processor

The development of a Virtual Studio Technology (VST) plugin (Rumsey, 2004) (see also (Wikipedia Foundation, Inc., 2014)) that allows for the control of the parameters discussed in the previous chapters was crucial for the continuing work related to this thesis. Having a processor that could produce both *shuffling* and up-mixing techniques allowed for an easier control of all parameters in one 'box'.

The plugin was built using the SonicBirth (Missout, 2007) programming environment which allows for object based programming of VST and AU plugins. An overall schematic of the patch can be seen in Figure 8-1.

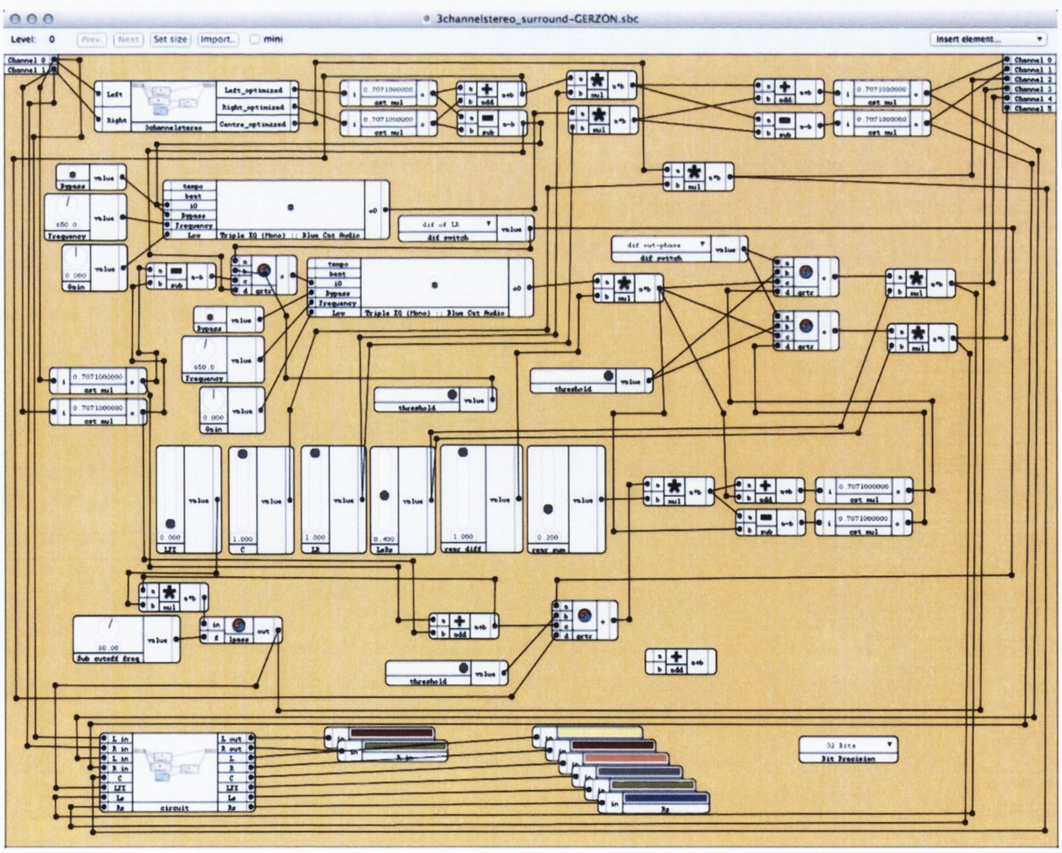


Figure 8-1: Overall schematic of the spaciousness processor patch.

The patch allows for the input of stereo signals (L and R) which are then encoded to Sum (M) and Difference (S) channels (Figure 8-2). Separating these signals allows therefore for the *shuffling* of the S signal where a low-shelf filter with variable f_c and gain can be applied (Figure 8-3). Since this patch also allows for stereo up-mixing to 5.1, a copy of the S signal can be independently *shuffled*, if desired, and routed to the surround channels. The routing of the S signal to the surround channels is based on Gerzon's approach described in Section 8.2.2, and can be sent either in-phase or out-of-phase (*i.e.* the Left surround and Right surround channels can be fed the same S signal or it can have a polarity reversal between them). Also, the surround channels can be fed with the M signal, which in small amounts helps surround channels not being to much out-of-phase (Gerzon, 1971). The levels of M and S can also be changed independently (Figure 8-4 and Figure 8-5).

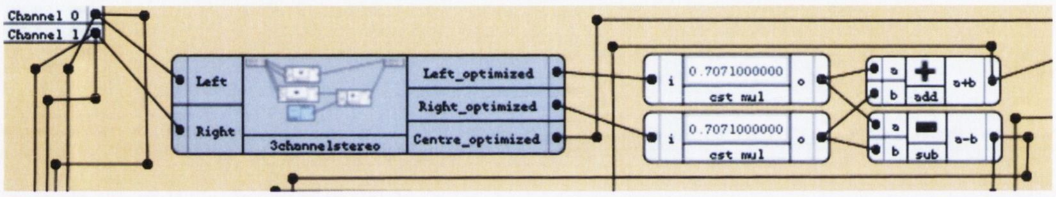


Figure 8-2: Encoding of Left and Right signals to Sum (M) and Difference (S) signals. The "add" (+) and "sub" (-) objects allow for the addition and subtraction of the Left and Right signals which results in an output of M and S signals. Channel 0 and Channel 1, represent the Left and Right signal input, respectively.

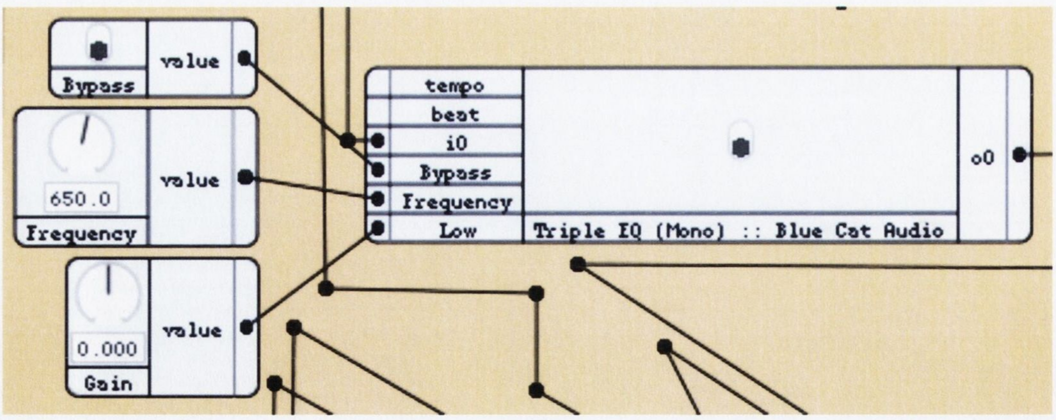


Figure 8-3: *Shuffling* of the S signal which was obtained from subtracting the Left and Right signals. The gain of a low shelf filter can be control with variable f_c.

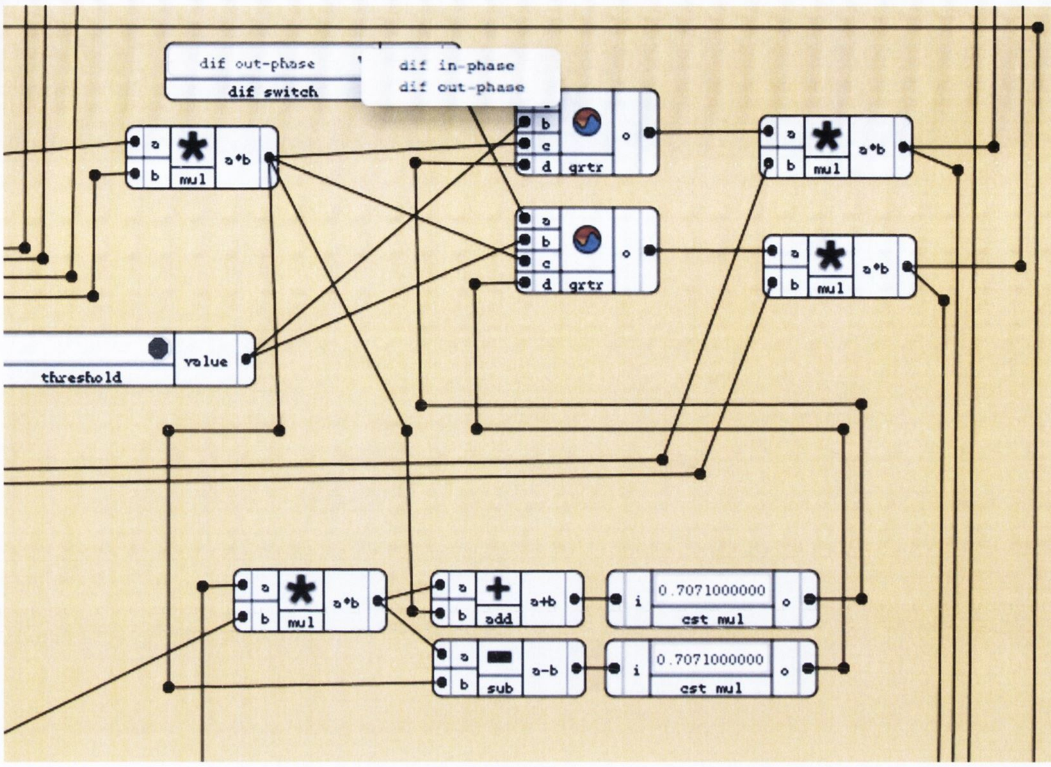


Figure 8-4: Feeding the surround channels with either an in-phase or out-of-phase S signal. The "add" and "sub" object allows for the decoding of M and S signals which will allow for the control of the amount of sum and difference signals to be fed to the surround channels.

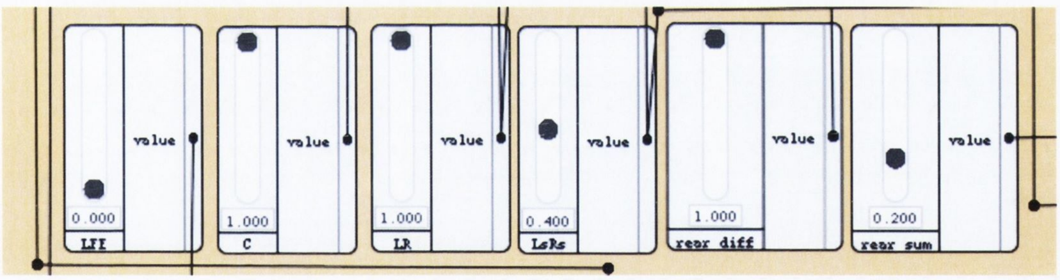


Figure 8-5: Level control section for LFE, Centre, LR, LsRs and surround (rear) sum and difference signals.

After *shuffling*, the M and S signals are decoded to stereo signals (L and R) which are then routed to the front left and right channels (Figure 8-6). The centre channel will play the M signal and the rear channels will play the S signal. All channel pairs (*i.e.* L-R and Ls-Rs) can be controlled in level, as well as the centre channel (Figure 8-5). This makes it possible to control the level of surround channels vs. front channels vs. centre image, which allows for the control of how the sound is presented to the listener potentially leading to a more or less spacious impression of the reconstructed sound field.

It is also possible to feed the LFE channel with a signal. This is done by low pass filtering the M signal and routing it to this channel. The cutoff frequency, f_c, can be varied, and the level of the signal sent to the LFE channel can also be controlled (Figure 8-7).

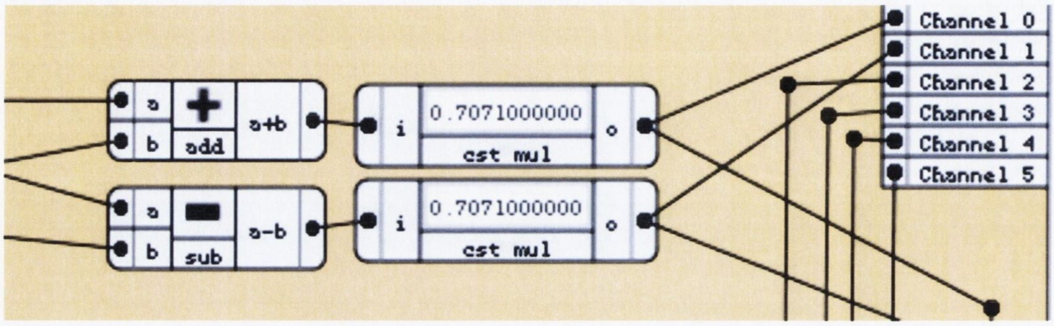


Figure 8-6: Decoding of the M and S signals to Left and Right signals. Channel 0, Channel 1, Channel 2, Channel 3, Channel 4 and Channel 5 are the outputs for Left, Right, Centre, LFE, Left Surround and Right Surround, respectively.

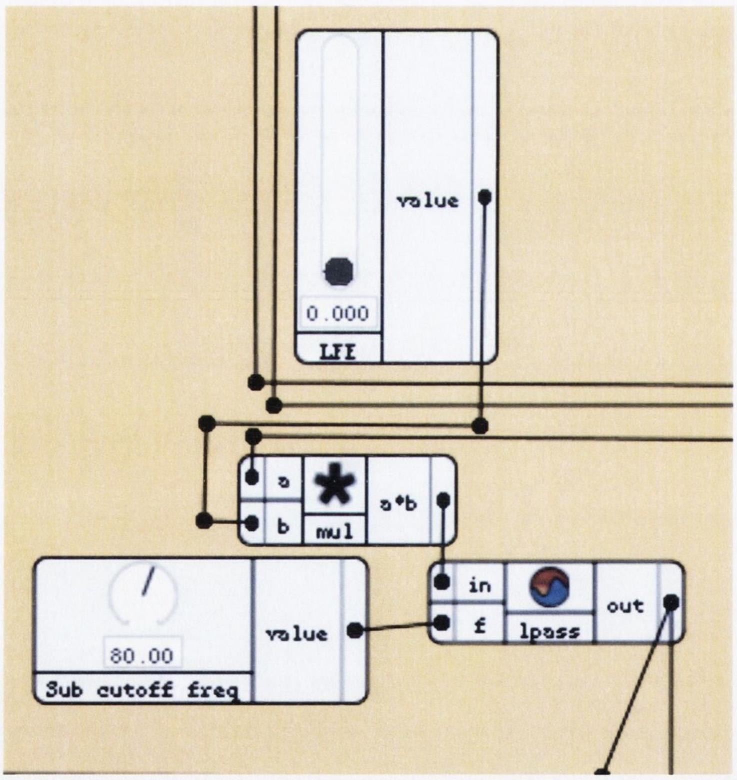


Figure 8-7: Low pass filter applied to the M signal which is then routed to the LFE channel. The cutoff frequency can be varied.

The details described here are for the most recent version of the implementation of this plugin. Equation 8.1 and Equation 8.2 give the MS encoding and decoding matrix for the previous version of this processor:

$$M = 0.5(L + R)$$
$$S = 0.5(L - R)$$

Equation 8.1: MS encoding from L and R channels; used in previous versions of the VST plugin.

$$L = M + S$$
$$R = M - S$$

Equation 8.2: MS decoding to L and R channels; used in previous versions of the VST plugin.

In previous versions, the S signal sent to the rear channels was always in-phase, and routing the M signal to these channels was not made possible. The most recent version of the plugin has a 2 to 3 conversion matrix (see Figure 8-8) where the input of stereo (L and R) is converted into M and S signals as represented by Equation 8.3, and converted back to L and R by the inverse matrix represented in Equation 8.4. The implementation of these encoding and decoding equations in the processor is demonstrated in Figure 8-2 and Figure 8-6, respectively.

$$M = 0.7071(L + R)$$
$$S = 0.7071(L - R)$$

Equation 8.3: Sum (M) and Difference (S) signal encoding from L and R.

$$L = 0.7071(M + S)$$
$$R = 0.7071(M - S)$$

Equation 8.4: Sum (M) and Difference (S) signal decoding to L and R.

Following Gerzon, the M signal is then split via a constant-power sin/cosine pair of gains where an M component with gain $cos\emptyset$ is sent to the Centre channel, and an M component with gain $sin\emptyset$ is sent to an MS decoder for Left and Right channel outputs (Gerzon, 1992b). It has been found (Gerzon, 1992a) that the optimum value of $\emptyset$ for good imaging depends on frequency, with $\emptyset = 35.5°$ for frequencies below 5kHz and $\emptyset = 54.7°$ for frequencies above 5kHz. The crossover point, between the two values of $\emptyset$, should involve use of a smooth transition near 5kHz. Figure 8-8 shows a block diagram of the 2 to 3 matrix which was implemented in the latest version of the VST plugin. Figure 8-9 shows the implementation of the constant-power sin/cosine pair of gains using the optimum value of $\emptyset$.

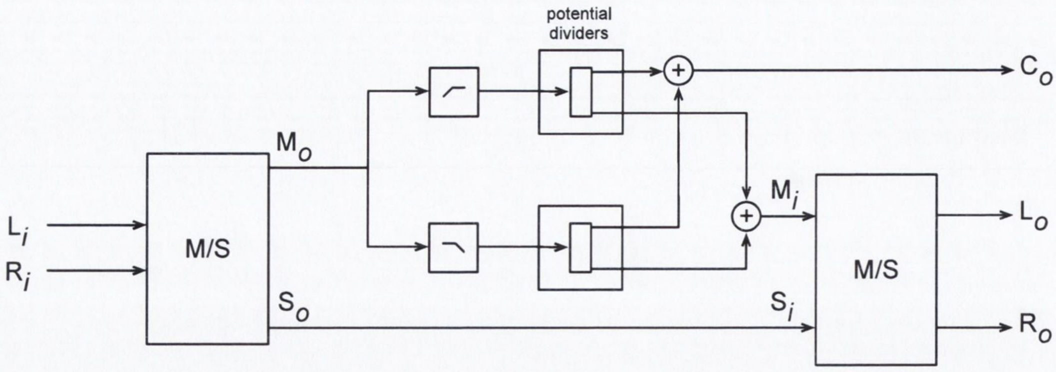


Figure 8-8: Frequency dependent version of Gerzon's 2 to 3 decoder (adapted from (Gerzon, 1992b).

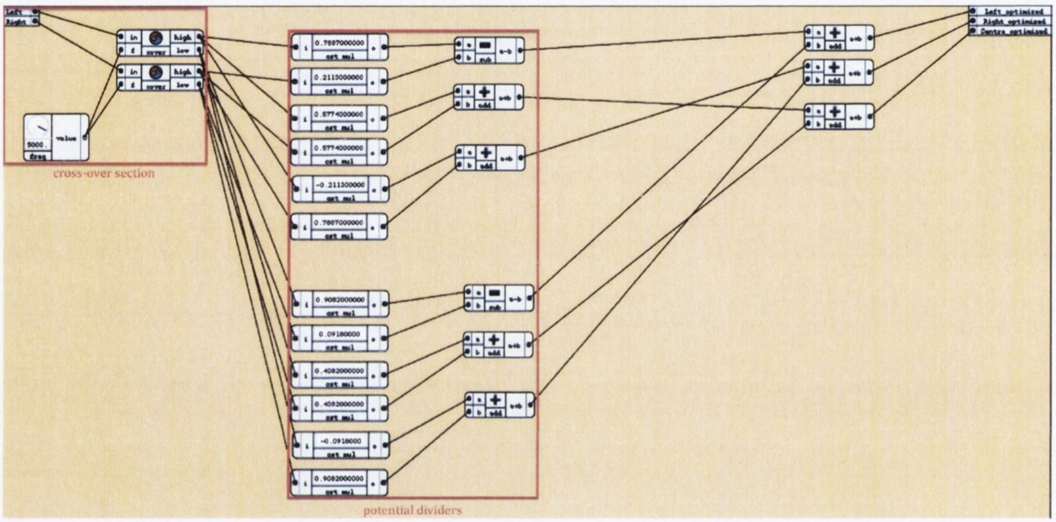


Figure 8-9: Details of the implementation of Gerzon's 2 to 3 decoder (Gerzon, 1992b) in the VST plug-in developed for this thesis; the cross-over section and the potential dividers are highlighted.

In the next sections, results will be presented indicating the effectiveness of the control of listener spaciousness impression using the processor described here. All the results were obtained using the VST plugin with the surround channels, which was fed with the S signal, in-phase, and with the 2 to 3 matrix described previously. The implementation of the latest version is an improvement over the previous version performance.

8.3 IACC measurement

In order to control IACC, a similar test setup to that used for the listening room effect test was employed for the reconstructed sound field spaciousness study, in order to investigate the effect of recording and reconstruction system detail on IACC (see Figure 6-2).

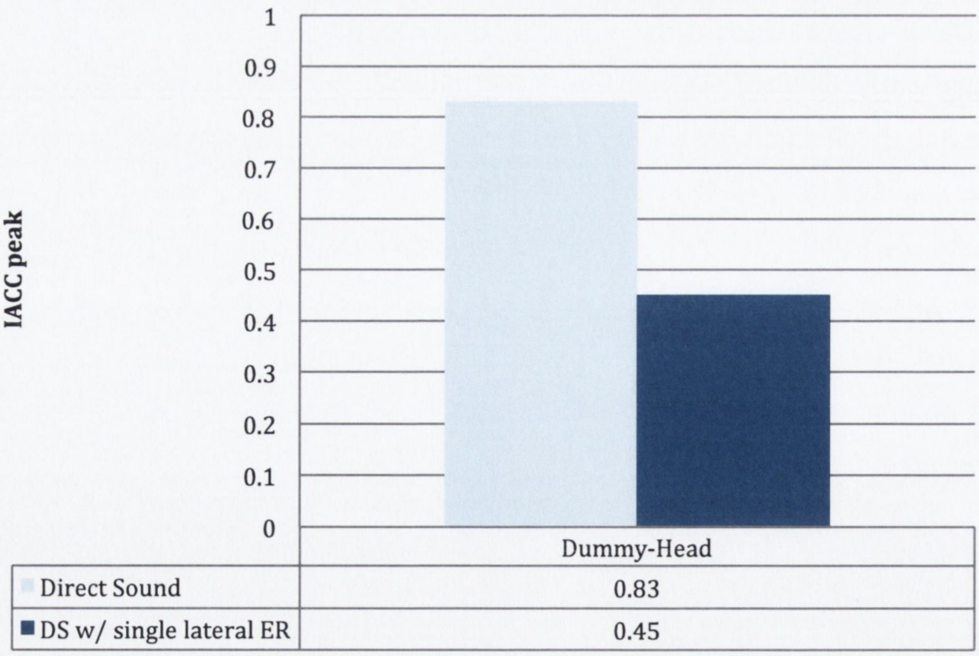
As a control, a dummy-head microphone system captured the direct sound and early lateral reflection, where the indirect signal could be varied in level and delay. The control experiments results are discussed in Chapter 6, Section 6.2.1. Later the dummy head microphone was replaced with different micing techniques which recorded the same sound field as previously captured using the dummy-head (see Figure 7-1)

The recorded MLS noise signals were then played back through stereo and 5.1 systems to see how each of the microphone recording techniques behaved in terms of the reproduced spaciousness. To assess this, measured IACC was again used, with a dummy-head microphone system being placed at the sweet spot of the playback system (see Figure 8-10), so that the IACC peak value of each reconstructed recording could be measured.

Table 8-1 shows the IACC values for the Direct Sound only, and for the Direct Sound with an added single lateral Early Reflection at 50ms and -4dBFS relative to Direct Sound. The results shown here, which are for reference, relate to the work presented in Chapter 6, and demonstrate that the added reflection produces a significant drop in IACC value, which is an indication of an increased sense of spaciousness. Previously, results were presented for the effect of a single early lateral reflection on IACC, for different delays and levels of the simulated reflection. However, here only the effect of an early reflection with 50ms and -4dBFS level relative to direct sound was used, since it had been established that a strong early reflection is capable of affecting an IACC result. The question that was addressed was whether different recording and reconstruction system details when used with signal processing and matrixing can influence these measured IACC results. This is thus an extension of the work presented in Chapter 7.

Figure 8-10: IACC measurements of reconstructed sound fields. *Serviços de Áudio* studio at ESMAE-IPP.

Table 8-1: IACC values for Direct Sound and Direct Sound with lateral Early Reflection. Early Reflection delayed by 50ms and at a -4dB level relative to the Direct Sound.



8.4 IACC measurement of different recording techniques

As mentioned previously, all stereo and surround recordings captured the same sound field as the dummy-head – a Direct Sound with an added single lateral Early Reflection at 50ms with -4dB relative amplitude to the direct sound.

The microphone techniques used for this experiment were: Blumlein, XY wide cardioid, XY super-cardioid, MS, AB, Double MS, Bruck Array a.k.a. KFM Surround (SCHOEPS GmbH, 2013a) Ambisonic and OCT-Surround. All the stereo and surround formats are well known and for further details on the setup and characteristics refer to Chapter 5 of this thesis (Rayburn, 2012), (Streicher & Everest, 2006) and (Rumsey, 2001). For all micing techniques, the microphones were placed in the spot where the dummy-head was placed (Figure 6-2 and Figure 7-1) and captured the signals produced by the direct sound and early reflection speaker. In the case of the OCT surround setup, the 3 main mics (L-C-R) are placed in the same "spot" as was the dummy-head, thereby positioning the surround mics further away from the sound source.

All the recording techniques were shuffled, and the stereo techniques were also up-mixed using the VST Spaciousness Processor plugin developed by the author as detailed previously (see Section 8.2.3). Stereo *shuffling* was implemented using a low-shelf filter that increased the difference channel amplitude of the stereo signal by 6dB, leaving the sum channel amplitude of the stereo signal unaltered. Native surround recording techniques were played in 5.1 using the decoders that are used with the systems. In the case of the Ambisonic system, the Soundfield MKV system was used with the SP451 for converting B-Format to G-Format (5.1) (see Chapter 5, Section 5.6.1.1). The Double MS technique was decoded using the Schoeps VST decoder (2013b), which allows for stereo and surround matrixing. The MS technique was decoded within the digital audio workstation by duplicating the S channel. The M signal was fed to both Left and Right channels, while the original S signal was fed to the Left channel, and a copy of the S signal with polarity reversed was fed to the Right channel. A schematic representation of the MS decoding is shown in Figure 8-11.

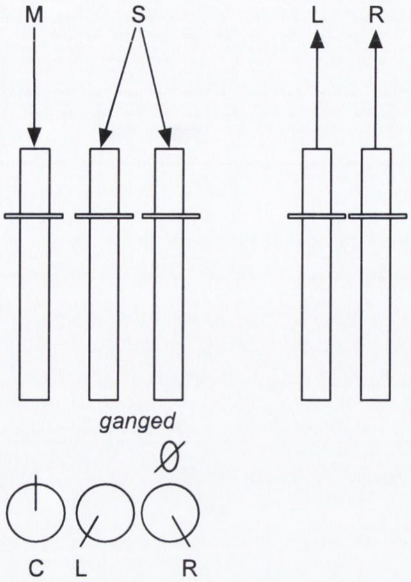
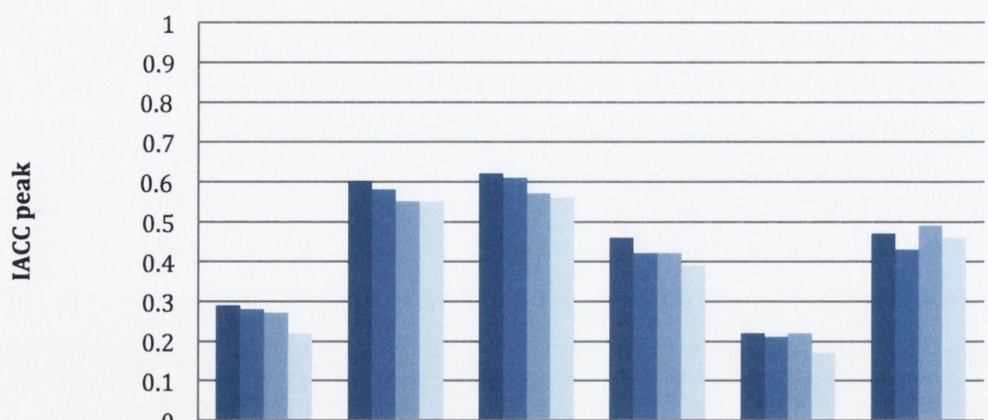


Figure 8-11: MS decoding to LR using a mixer. This can be either implmented with outboard equipment or internal to any DAW.

8.5 Results

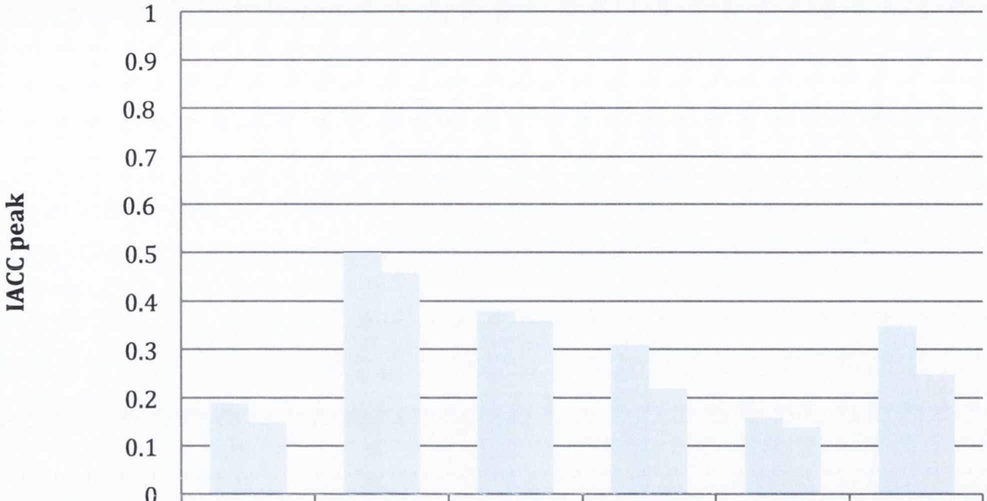
The following tables show the IACC results achieved for different micing techniques using stereo, LCR and 5.1 playback systems. All the measurements where conducted in Oporto (see Chapters 6 and 7).

Table 8-2: Comparison between stereo, shuffled stereo, 3-channel stereo and shuffled 3-channel stereo playbacks.



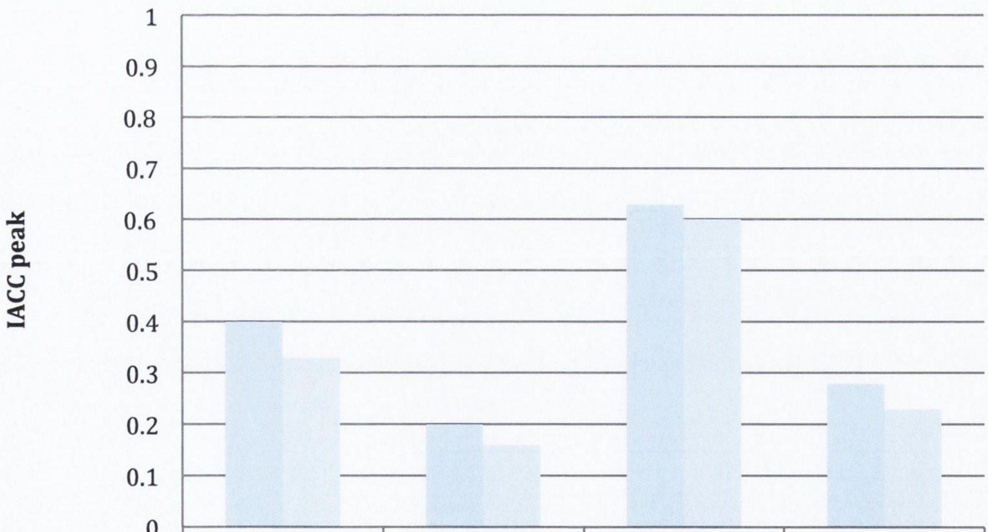
	Blumlein	XY	XY wide-card.	XY super-card.	MS	ORTF
LR	0.29	0.6	0.62	0.46	0.22	0.47
LR shuflled	0.28	0.58	0.61	0.42	0.21	0.43
LRC	0.27	0.55	0.57	0.42	0.22	0.49
LRC shuffled	0.22	0.55	0.56	0.39	0.17	0.46

Table 8-3: Comparison of different stereo microphone arrays when up-mixed from stereo to 5.1 and to 5.1 with *shuffling*.



	Blumlein	XY	XY wide-card.	XY super-card.	MS	ORTF
5.1	0.19	0.5	0.38	0.31	0.16	0.35
5.1 shuflled	0.15	0.46	0.36	0.22	0.14	0.25

Table 8-4: Comparison between different surround microphone arrays. 5.1 and *shuffled* 5.1 playbacks.



	DMS	Bruck array	Soundfield	OCT Surround
5.1	0.4	0.2	0.63	0.28
5.1 shuflled	0.33	0.16	0.6	0.23

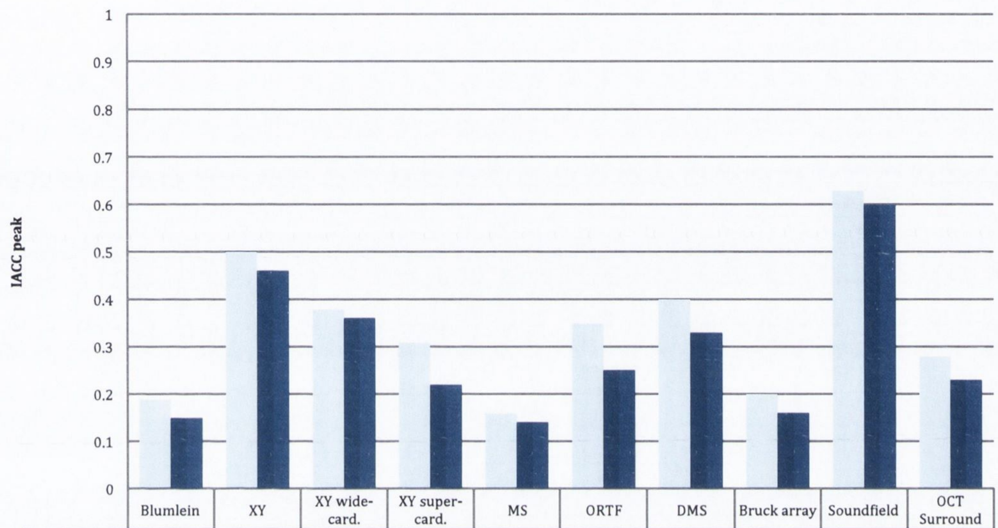
From an overall analysis of the above tables, it can be seen that shuffling and up-mix techniques produce a decrease of IACC peak values. However, it is curious to see that certain coincident stereo techniques, such as Blumlein, XY wide-cardioid and MS, do not produce a change in IACC peak value when reproduced with shuffled stereo (as defined by Gerzon (1986)), as compared to "normal" stereo reproductions. The shuffling technique here was implemented using techniques derived from both Griesinger (1985) and Gerzon (1986), which were aimed at making coincident recordings sound more spacious. The reason for this is still in need of further research and investigation.

The results confirm what was found in Chapter 7, that microphone parametric variation can be employed to facilitate spaciousness control for reconstructed sound fields. Also, appropriate matrixing of stereo recordings to 5.1 playback systems using the VST spaciousness processor will introduce reductions in IACC peak values, indicating an increase of perceived spaciousness impression. Not only do up-mixing techniques provide for a sense of an increased spaciousness, but also appropriate filtering of the difference channels in stereo recordings will produce changes to the sonic results. These findings are of particular interest to sound engineers and producers, as referred to previously, since the understanding of how to create sound field reconstructions with controllable spaciousness can help promote the use of surround sound systems for music presentation, providing more pleasant, spacious recordings.

In Table 8-2, where an LRC reproduction system was analysed, it can be seen that the LRC *shuffled* version produces an IACC peak drop in all cases except for XY wide cardioid. From Table 8-2 it can also be observed that stereo recordings, when presented over 3-channel stereo, show lower IACC peak values. These results indicate that LRC playback systems increase the perceived impression of spaciousness. However, LRC playback of MS and ORTF recordings produce higher IACC peak when compared to LR playback. Since the Centre channel is being fed with the sum of L and R signals, this higher IACC peak could be the result of presenting a more correlated sum signal (see Chapter 5).

The introduction of surround channels (Table 8-3 and Table 8-4) seems to make a big difference to IACC. All the stereo microphone techniques produce lower IACC values when used with surround channels as compared to stereo playback. Native surround recording techniques also produce low IACC peak values, except for the Soundfield array which produces an even higher IACC peak value than the stereo microphone recordings presented in Table 8-2 when played back through stereo loudspeakers. Also, the LFE channel does not change the IACC peak values obtained for 5.0 and 5.1 systems, as can be observed in Table 8-5.

Table 8-5: Comparison between 5.0 (without LFE channel) and 5.1 (with LFE channel) presentations; *shuffled* presentations results are also shown.



	Blumlein	XY	XY wide-card.	XY super-card.	MS	ORTF	DMS	Bruck array	Soundfield	OCT Surround
5,0	0.19	0.5	0.38	0.31	0.16	0.35	0.4	0.2	0.63	0.28
5.1	0.19	0.5	0.38	0.31	0.16	0.35	0.4	0.2	0.63	0.28
5.0 shuffled	0.15	0.46	0.36	0.22	0.14	0.25	0.33	0.16	0.6	0.23
5.1 shuflled	0.15	0.46	0.36	0.22	0.14	0.25	0.33	0.16	0.6	0.23

8.6 Subjective listening tests

In order to confirm the significance of the IACC results from the objective measurements outlined in this and previous chapters, a subjective listening test was conducted. The purpose of the test was to align the objective measures of spaciousness for the reconstructed sound fields with subjective preference assessments. The differences in spaciousness between microphone technique recordings, *shuffled* and up-mixed versions, as indicated by the IACC measurements, were expected to be reflected in the results of a subjective test.

As the objective measures compared several different microphone technique recordings with their *shuffled* and up-mixed versions and also with native surround microphone techniques (see Chapter 7 and Chapter 8), ideally the subjective test should also compare recordings from all of these microphone techniques playing back through LR, LRC and also 5.1. Since this would prove to be an impracticable test method, the subjects were presented with some selected examples of un-processed stereo recordings and their spaciousness processed versions and were asked to score the spaciousness attribute of the processed version in comparison to the un-processed version. Also subjects were asked to compare some selected stereo microphone recording techniques with surround microphone recording techniques and different stereo recordings.

In brief, the subjective test entailed the evaluation of the spaciousness of several different recording techniques (*e.g.* single point "micing" and "multi-micing), which were processed with the spaciousness processor developed during the research undertaken for this thesis, and also a comparison of different recording techniques. The results of the test produced an average score for each example. The scores were then correlated with the delta IACC values obtained from the comparison of the objective measurements for the paired examples presented.

8.6.1 Examples

Eight music examples were used in the test:

i. Blumlein, MS, OCT, DMS and Ambisonic recordings of a direct front signal and a single, 50ms -4dB relative to direct sound, early reflection, as previously described in Section 7.2, of a anechoic sample of Mozart's overture, "Le Nozze di Figaro" (Denon, 1995);
ii. MS and Double MS recordings of a sample of Shostakowitsch 5th Symphony, beginning of the 4th movement (SCHOEPS GmbH, 2014);
iii. AB recording of a sample of Bach's "Herz und Mund und Tat und Leben" cantata, recorded live by the author in a chapel in Portugal (Figure 8-20);
iv. Sample of Bach's Prelude in E-flat major (2000), which according to the record label is a single spaced AB recording;
v. Sample of Será Una Noche's Malena (1999), which according to the record label is a single spaced AB recording;
vi. Sample of Cooder and Bhatt's Meeting By The River (2008), which according to the record label is a single point Blumlein recording;

vii. Sample of Sting's Perfect Love...Gone Wrong (2001);

viii. Sample of Tchaikovsky's Swan Lake Scene Act II (1999).

Each sample was edited to be approximately 10 seconds in duration using short fade ins and outs where necessary. All the examples were adjusted so that their loudness was constant from example to example. This was made possible with the use of the loudness meter included in Nuendo 6 DAW (STEINBERG MEDIA TECHNOLOGIES GmbH, 2014).

Table 8-6: List of examples presented to the subjects, where they were asked to judge B in comparison to A in terms of perceived spaciousness

	A	*B*
Example 1	*Mozart;* *Blumelein recording*	*Mozart;* *Blumlein recording, spaciousness processor applied*
Example 2	*Mozart;* *original anechoic recording*	*Mozart;* *Blumlein recording, spaciousness processor applied*
Example 3	*Mozart;* *MS recording, spaciousness processor applied*	*Mozart;* *MS recording*
Example 4	*Mozart;* *OCT recording, spaciousness processor applied*	*Mozart;* *OCT recording*
Example 5	*Mozart;* *Blumlein recording*	*Mozart;* *OCT recording*
Example 6	*Shostakowitsch;* *MS recording*	*Shostakowitsch;* *MS recording, spaciousness processor applied*
Example 7	*Shostakowitsch;* *Double MS recording*	*Shostakowitsch;* *MS recording*
Example 8	*Bach;* *AB recording*	*Bach;* *AB recording, spaciousness processor applied*
Example 9	*Shostakowitsch;* *MS recording*	*Shostakowitsch;* *MS recording*
Example 10	*Mozart;* *Blumlein recording*	*Mozart;* *Blumlein recording*
Example 11	*Mozart;* *Double MS recording*	*Mozart;* *Ambisonic recording (presented in G-format)*
Example 12	*Shostakowitsch;* *MS recording, spaciousness processor applied*	*Shostakowitsch;* *MS recording, spaciousness processor applied*
Example 13	*Bach – lute music;* *AB recording*	*Bach – lute music:* *AB recording, spaciousness processor applied*

Example 14	*Cooder & Bhatt;* *Blumlein recording, spaciousness processor applied*	*Cooder & Bhatt;* *Blumlein recording*
Example 15	*Sting;* *spaciousness processor applied*	*Sting;*
Example 16	*Sting;*	*Sting;*
Example 17	*Tchaikovsky;*	*Tchaikovsky;* *spaciousness processor applied*
	A	***B***
Example 18	*Será Una Noche;* *AB recording*	*Será Una Noche;* *AB recording, spaciousness processor applied*
Example 19	*Será Una Noche;* *AB recording*	*Será Una Noche;* *AB recording*

The examples were presented in a pair-comparison where original recordings, either by the author or by the record labels, were compared to processed versions. Some examples were pair-comparisons of different microphone techniques, and also there were examples where A and B were exactly the same samples acting as controls (*e.g.* example 9, 10, 12, 16 and 19). The list of examples presented is shown in Table 8-6

8.6.2 Subjects

Twenty one listeners took part in the test, all of whom were staff or students of the School of Music, Arts and Performing Arts of the Polytechnic Institute of Oporto. The majority of the subjects had previously taken part in other listening tests and had an interest in either audio and music technology or music recordings in general. The subjects could be considered as "selected assessors" (Bech & Zacharov, 2006) and therefore be expected to produce reliable judgments. None of the subjects reported any known hearing defects. The tests were held over a three-day period.

8.6.3 Physical setup and test preparation

The Oporto studio described in Section 6.2 was used with a 5.1 playback system in accordance with the ITU-R BS.775.1 standard (1992-1994) and can be seen in Figure 8-12. This studio was chosen because of its low background noise and controlled environment, and also as it was the studio used for the objective measures of IACC described in Chapters 6, 7 and 8.

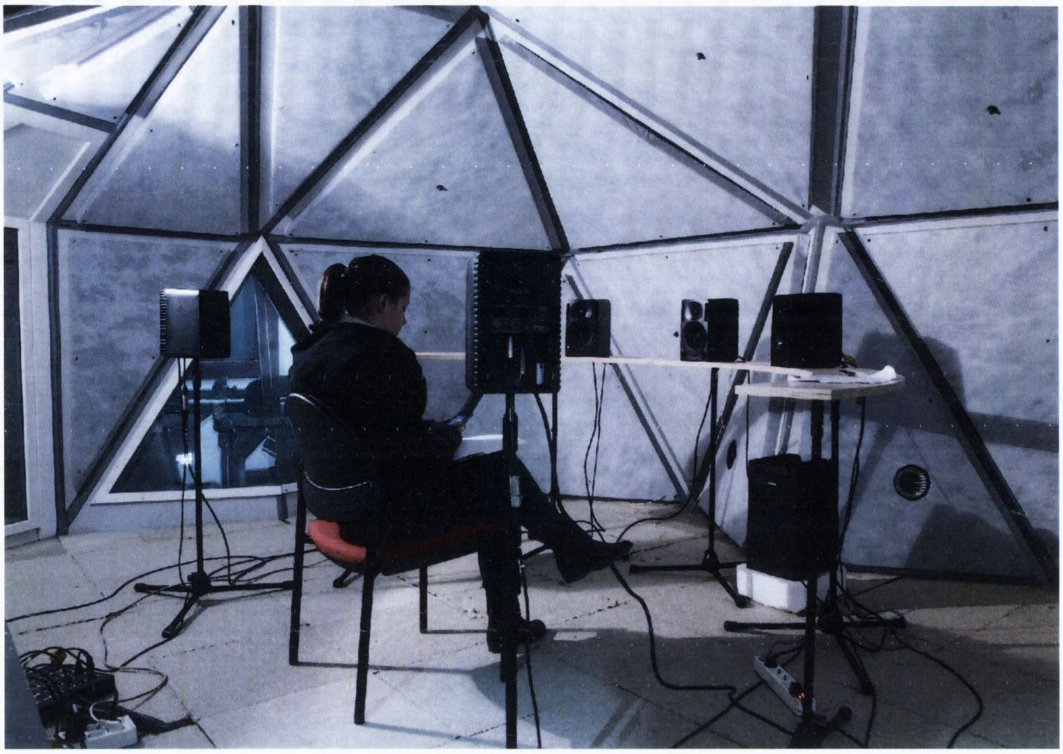


Figure 8-12: Detail of the setup for the listening tests. *Serviços de Áudio* studio at ESMAE-IPP.

Five random lists of the examples were created, so that the order of playback of the examples could be presented differently to the subjects. This was done to avoid subjects from influencing other subjects by revealing their answers. The examples were organized on a computer using Nuendo 6 which is capable of programing different playback orders of the examples, and also of playing stereo and 5.1 formats. The computer was then connected to the loudspeakers via a multichannel audio interface. The equipment used was the same as described in Section 6.2.1. To allow subjects to swap between A and B of each example a tablet interface was used with TouchOSC (HEXLER.NET, 2014) which permits sending MIDI information over Wi-Fi, therefore allowing communication with Nuendo 6. For this listening test, an interface for TouchOSC was specifically designed, using the TouchOSC Editor (Figure 8-13).

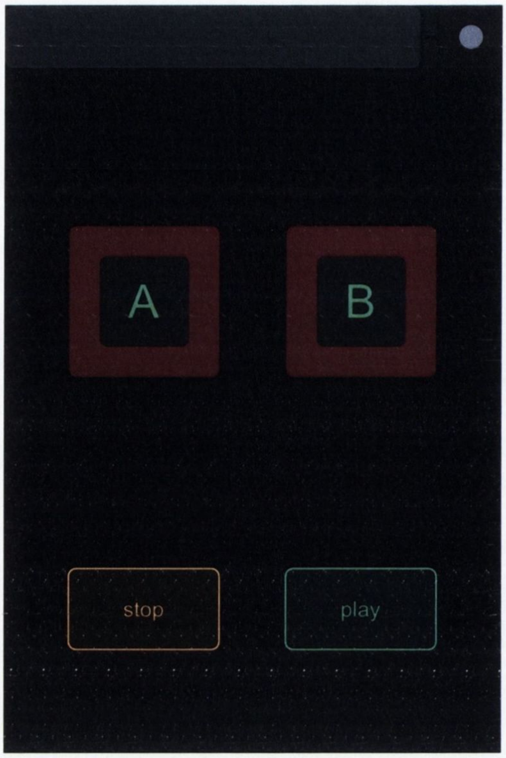


Figure 8-13: TouchOSC interface for the listening test. Subjects could swap seamlessly between A and B of each example and also start and stop each example.

8.6.4 Test procedure

The subjects were presented with the 19 pair-comparison examples (Table 8-6), one after the other. They could swap seamlessly between A and B as they wished and, since the examples were presented in a loop, they could listen to them until they were comfortable with an answer. Before engaging in the listening test the subjects were allowed to familiarize themselves with the interface and the environment and also with an example where A and B were made to be extremes was presented (*e.g.* anechoic recording vs full spaciousness sound).

The subjects were asked to rate each example, using a Comparison Category Rating (CCR) scale (Bech & Zacharov, 2006), always comparing B to A, in terms of spaciousness (see Table 8-7). When a subject had finish rating a particular example, they would let it be known that he or she was ready to compare the next example. The subjects took approximately between 10 to 30 minutes to complete the test. Instruction can be seen in APPENDIX V - Spaciousness Assessment Questionnaire.

Table 8-7: Comparison Category Rating Scale used for the listening test.

-2	-1	0	1	2
much less Spacious	less Spacious	equal	more Spacious	much more Spacious

8.6.5 Results of the subjective listening test

From the results obtained, regarding the comparison between the spaciousness processed version and the original version examples (see examples 1, 2, 3, 4, 6, 8, 13, 14, 15, 17 and 18 in Table 8-6), it was possible to draw a frequency distribution chart based on the total number of answers (n=231) provided by the subjects. This chart shows that the majority of subjects (86.58%) were able to answer positively in relation to an increased spaciousness, where 43.72% felt a "more spacious" and 42.86% felt a "much more spacious" sound when presented with the spaciousness processed version of the examples (see Figure 8-14). These results show what previous IACC measures have indicated: an increased sense of spaciousness when up-mixing and *shuffling* is applied to the original recordings (see Section 8.5), confirming the efficiency of the spaciousness processor as a tool to provide an increased sense of spaciousness.

When comparing the native Double MS surround sound recording with its counterpart stereo MS recording technique (*i.e.* example 7 on Table 8-6), the total of answers provided show that the majority of subjects (80.87%) indicate an increased sense of spaciousness, where 52.38% felt a "more spacious" sound and 28.57% felt a "much more spacious" sound in favour of the Double MS technique (see Figure 8-15).

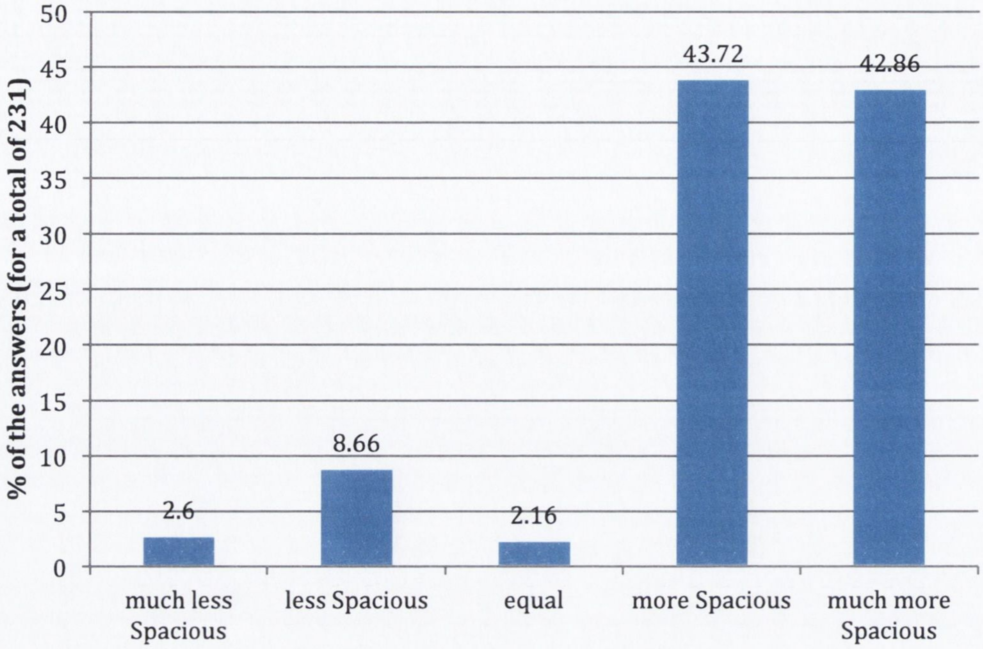


Figure 8-14: Percentage of answers given when comparing the spaciousness processed version and original version of the examples listened.

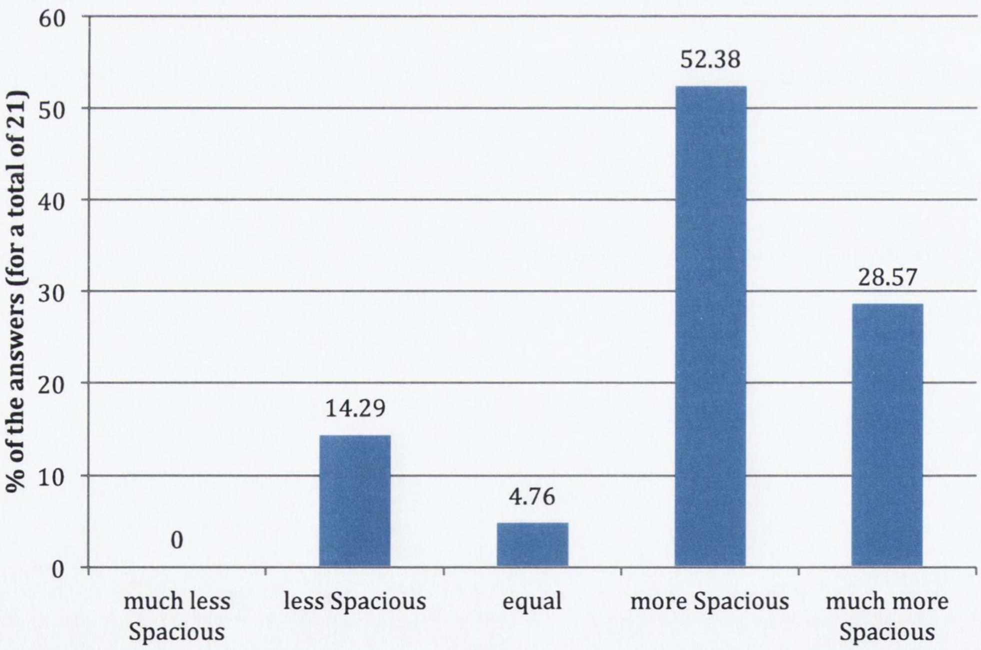


Figure 8-15: Percentage of the answers given when comparing Double MS with MS recording technique.

Comparing IACC results of Double MS with MS recording techniques (see Table 8-2 and Table 8-4) it is possible to observe that MS has a lower IACC peak value, indicating an increased sense of spaciousness. These results are not in complete accordance with the subjective results, but it should be notice that the IACC results obtained for either MS or Double MS were from microphone configurations, experimental conditions and decoders used for this thesis, while the program material used for the subjective assessment of MS and Double MS recordings were obtained from Schoeps (2014) and decoded using their decoders (2013b).

The comparison of subjective results between Blumlein and OCT recording technique (*i.e.* example 5 on Table 8-6) reveals that the impression of spaciousness is most obvious with the Blumlein technique, where 76.19% assessed Blumlein has having a "more spacious" sound (see Figure 8-16). The subjective results here are in accordance with the objective measures of IACC, where Blumlein recording technique provided a lower IACC value in comparison to OCT, indicating an increased sense of spaciousness for the Blumlein technique.

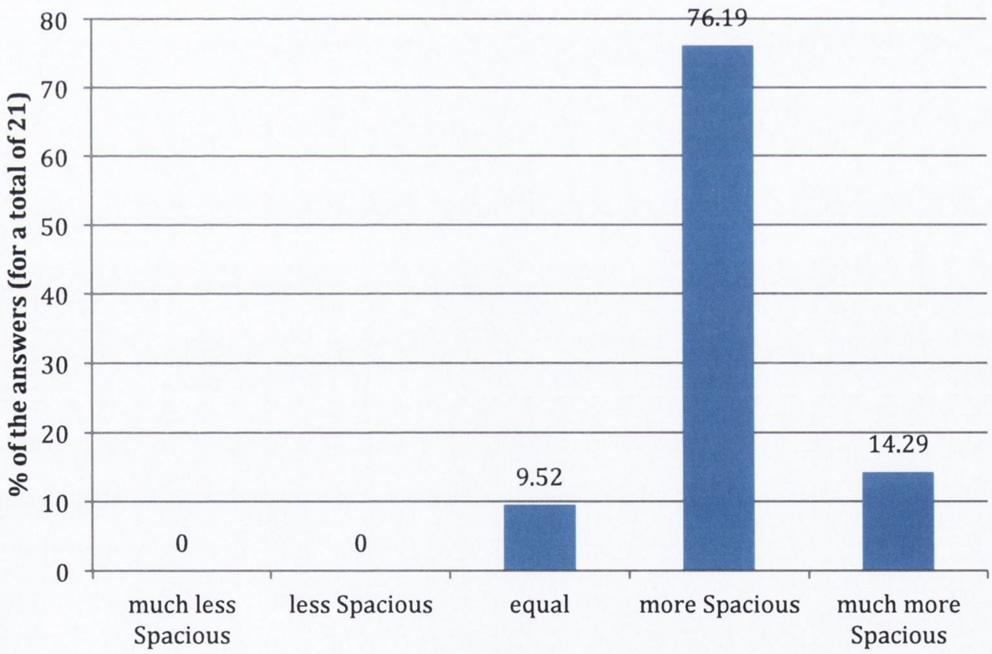


Figure 8-16: Percentage for total answers provided when comparing Blumlein with OCT recording technique.

Finally, when comparing the native surround recording techniques of Double MS with Ambisonics decoded to G-Format (*i.e.* example 11 on Table 8-6) it is shown that 69.9% find Double MS to have a "more spacious" feel while 33.33% indicate that both techniques have an equal impression of spaciousness (see Figure 8-17). The IACC results previously obtained for both these techniques indicate that Double MS has a lower IACC when compared with the Soundfield recording (*i.e.* Ambisonics decoded to G-Format) which indicates an increased sense of spaciousness for Double MS (see Table 8-4).

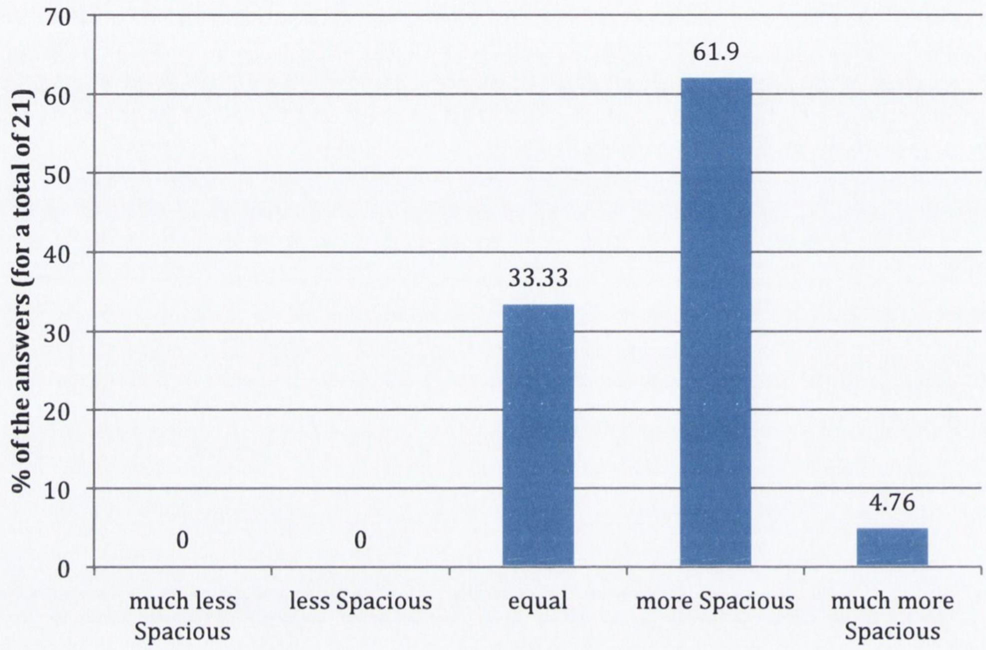


Figure 8-17: Percentage of answers given when comparing Double MS with Ambisonic (decoded to G-Format) recording technique.

Since the data from the type of questionnaire used is categorical data also referred as *non-parametric* (Bech & Zacharov, 2006), a "sign test" was used for the statistical analysis of the data (Svensson, 2001). Each example was evaluated to see if a classification above 0 was obtained, meaning an increased sense of spaciousness, which was revealed to be greater than the significance level ($p < 0.05$) for all the processed examples with the exception of example 14 (see Table 8-6). All of the controls (*i.e.* example 9, 10, 12, 16, 19; see Table 8-6) did not veer significantly from 0 ($p > 0.05$). This demonstrates that the subjective perception was considered positive for all but one of the processed examples and neutral (= 0) for the control examples. Following this, for the cases where the perception was positive it was evaluated if it was "much more" positive (*i.e.* classification > 1). This was only statistically significant ($p < 0.05$) for examples 2, 3 and 13. The output of the results can be seen in Table 8-8.

Table 8-8: Statistical results for all examples used.

		(>0)			Probability >0	(>1)			Probability >1	Statistics		
Example	**N**	**<0**	**0**	**>0**	**P***	**<1**	**1**	**>1**	**P***	**Median**	**Average**	**delta_IACC**
1	21	1	2	18	0.000	3	12	6	0.254	1	1.095	0.14
2	21	0	0	21	0.000	0	2	19	0.000	2	1.905	0.64
3	21	1	0	20	0.000	1	1	19	0.000	2	1.762	0.65
4	21	1	0	20	0.000	1	15	5	1.109	1	1.095	0.41
5	21	0	2	19	0.000	2	16	3	0.500	1	1.048	0.35
6	21	5	2	14	0.032	7	7	7	0.605	1	0.714	n/a
7	21	3	1	17	0.001	4	11	6	0.377	1	0.952	n/a
8	21	3	0	18	0.007	3	14	4	0.500	1	0.905	n/a
9	21	1	19	1	0.750					0	0.000	0
10	21	2	18	1	0.875					0	-0.048	0
11	21	0	7	14	0.000	7	13	1	0.996	1	0.714	0.23
12	21	0	21	0	1.000					0	0.000	0
13	21	0	0	21	0.000	0	12	9	0.002	1	1.429	n/a
14	21	7	0	14	0.095					1	0.571	n/a
15	21	5	0	16	0.013	5	8	8	0.291	1	0.857	n/a
16	21	4	14	3	0.773					0	-0.095	0
17	21	2	0	19	0.000	2	12	7	0.090	1	1.143	n/a
18	21	1	1	19	0.000	2	11	8	0.055	1	1.238	n/a
19	21	1	15	5	0.109					0	0.190	0

** if P=0, must indicate P<0.001*

A comparison between the objective measurements and subjective results was undertaken using the IACC values from the results presented in Section 7.3.1 and 8.5, which are assumed to be representative of each of the paired-example since the recording technique and recording conditions were exactly the same. The delta IACC could only be calculated for examples 1, 2, 3, 4, 5, and 11 (see Table 8-8). The controls were assumed to have a delta IACC of null, since the examples presented in A and B were exactly the same. In this comparison it can be observed that there is a linear relation (R^2=0.909) between the average of subjective preference and the delta IACC of each paired-example. There is, however, one example (*i.e.* example 1, see Table 8-6) which is considered to be an outlier, since the high subjective appreciation is not in correlation with the delta IACC (see Figure 8-18.) From all the results obtained from the subjective listening test it is possible to state that the objective measurements explain 90.9% of the average subjective perceptual variation.

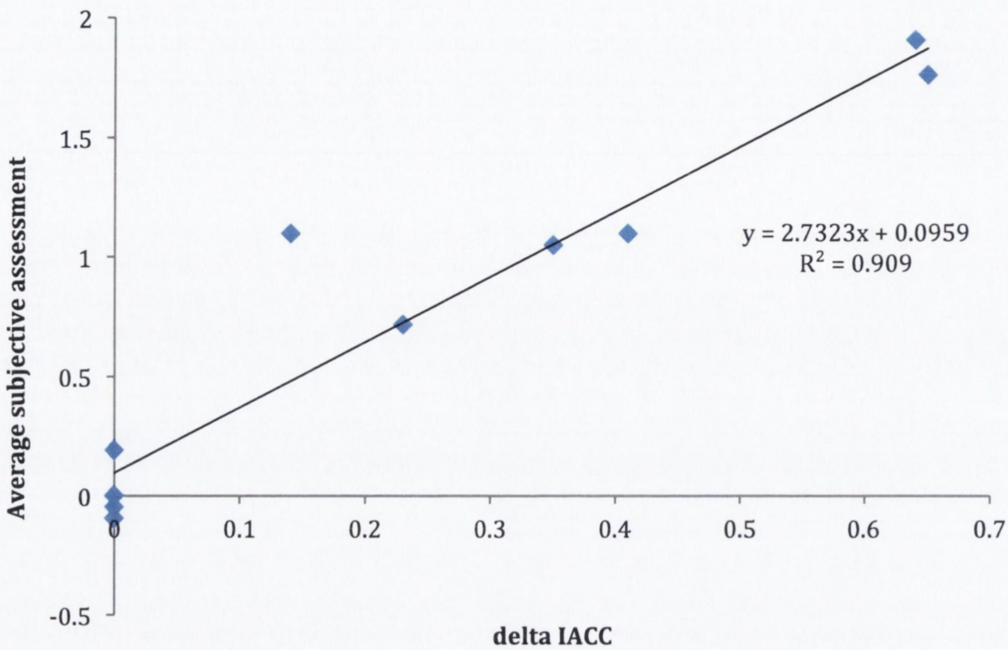


Figure 8-18: Delta IACC quantiles against average subjective assessment quantiles.

8.6.6 Casual listening comments

The author, supervisor and some colleagues also undertook some casual listening tasks which are not to be considered as listening tests for subjective evaluation, which was already dealt with in the previous sections. The material used for these listening tasks was previously recorded using all the techniques detailed, except for the Bruck array. The recordings were of a classical guitar duo and of an orchestra that performed in two different chapels in Portugal. The microphone techniques used were setup so as to be placed as closely together as possible, as can be seen in Figure 8-19 and Figure 8-20.

It was suggested previously that the overall spaciousness could be improved with 5.1 playback as compared to stereo. The stereo recordings that were up-mixed to surround using the spaciousness processor were judged to deliver an increase in perceived spaciousness. The surround channels needed to be carefully balanced so that the stereo image was not disturbed, and so that there was not too much de-correlation leading to an out-of-phase reproduction. It was found that using the up-mix technique proposed by Gerzon, and adapted by the author, delivered a pleasant, spacious sound as was reported on originally by Gerzon (1970). The comments also revealed that LRC playback (*i.e.* without the surround channels) was capable of producing a better sounding sound field reconstruction in terms of spaciousness, especially when shuffling was applied. It was also to be noticed that there was an increase in the sweet spot, with listeners outside the 5.1 system loudspeaker arrays even claiming that they could still perceive a stereo image and a spacious sound. Why this increased sense of sweet spot is noticed was not investigated, since all the experimental methodology was aimed for sweet-spot position listening. Further investigation should be conducted to assess this reported increase sense of sweet spot. Listeners who were seated close to the surround loudspeakers also claimed they did not feel the surround loudspeakers were perturbing their listening experience or the stereo image.

The surround recording techniques by themselves delivered a pleasant and spacious sound in comparison to stereo recording techniques. The Double MS and OCT surround produced the best impression and sensation. There was a general opinion that Ambisonics, when reproduced in 5.1, did not deliver what was expected in terms of spaciousness. The decoding of B-Format to G-Format was considered to be quite precise in terms of stereo imaging (*i.e.* localisation), but was lacking 'space'. The best ambisonic subjective results obtained were by using the figure-of-eight preset on the decoder Surround Zone VST (TSL Professional Products Ltd., 2013), but the listener spaciousness experience was still not close to what the other surround microphone techniques delivered. Different decoder designs might produce different perceptual impacts, but these were not experimented with.

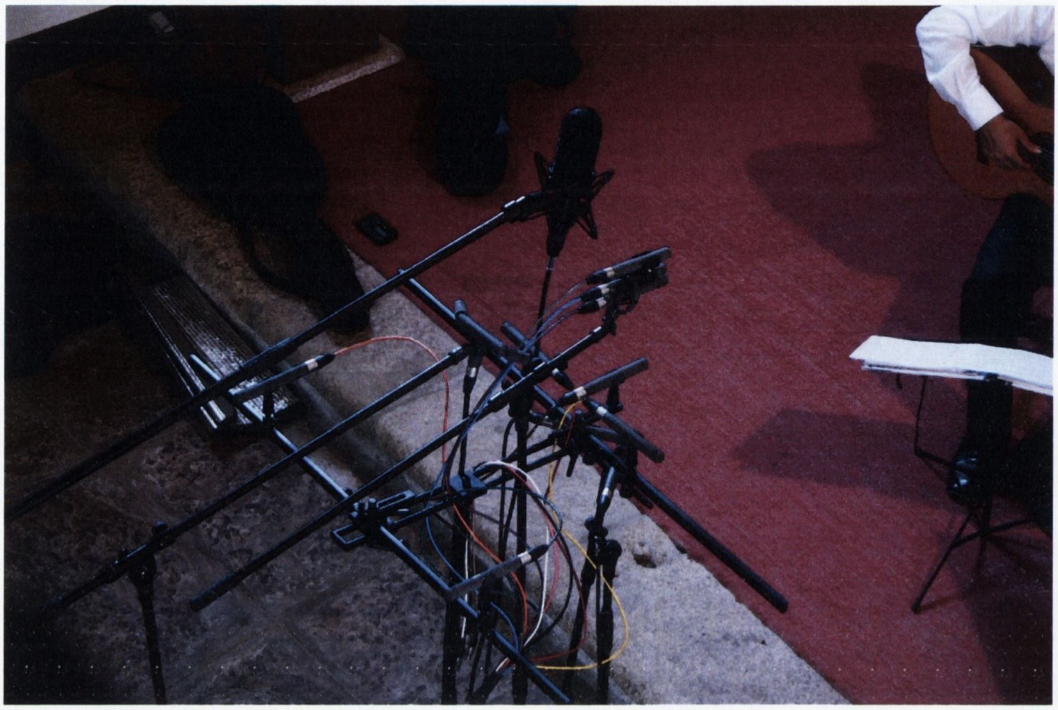


Figure 8-19: Comparative setup for stereo and surround recordings of a classical guitar duo in a chapel in Vila Nova de Cerveira, Portugal.

Figure 8-20: Comparative setup for stereo and surround recordings of an orchestra in Aveiro's cathedral, Portugal (front view).

Other listening demos were undertaken with commercially available stereo recordings, including both purist stereo recordings, and pair-wise panning stereo mixes. These recordings were up-mixed using the method discussed previously. Listening to these results was revealing, as the stereo recordings were improved dramatically regarding perceived spaciousness, without any compromise of stereo imaging. However, the details of the program and music played do appear to condition subjects' preferences, since musical preference might be taken into account by subjects. It was the listeners' opinion that there were not any unpleasant artefacts. Also, it was agreed that the up-mix technique gave better results for purist stereo recordings. However, pair-wise panning mixes seemed to gain a lot more, especially if artificially generated reverb had been used, making the sound of such effects more natural and pleasant.

8.7 Summary

A series of stereo and surround microphone techniques were investigated with respect to perceived spaciousness. Stereo shuffling and 5.1 up-mixing techniques were used in order to investigate the possible control of spaciousness in stereo and 5.1 reproduction systems. By using IACC as an objective measurement that correlates with the perception of spaciousness, several comparisons were presented in which a decrease in IACC peak value was shown, indicating a spaciousness increase for the reconstructed sound fields.

It was demonstrated that different microphone techniques contributed differently when reproduced over stereo and 5.1, and that the added LFE channel did not change the results obtained, in terms of IACC.

From the experimentation carried out, it can be concluded that spaciousness control is possible for stereo and 5.1 playback, either by controlling details of the microphone techniques used for recording, or by using shuffling and up-mixing techniques as was implemented in a custom VST spaciousness processor

A subjective listening test was devised in order to establish the effectiveness of the spaciousness processor, and also to establish a parallel between IACC measures of several examples and the subjective judgment of such examples. From all the results obtained from the subjective listening test it can be stated that the objective measurements explain 90.9% of the average subjective perceptual variation.

It has been confirmed that IACC is correlated with the perception of spaciousness for reconstructed sound fields, and that the perception of spaciousness is dependent on the differences between the left and right ear signals that reach the brain. Such a process for objectively assessing spatial sound reconstruction characteristics using different microphone techniques and reconstruction formats has not, to the author's knowledge, been previously presented.

Part of the work presented in this chapter has been published as a conference proceeding, and further details can be found in APPENDIX VI – Resulting Publications and Presentations.

9 DISCUSSION AND CONCLUSIONS

9.1 Introduction

Sound recording can be a process by which engineers, artists and producers use technology in order to achieve artistic communication. The simple press of a button introduces an input in the creation of *art*, and the choice of microphone techniques, the use of processing tools, can all contribute. In fact, they are all intentional dimensions in the creation of any sound recording which will lead to a music production, sound art installation, or sound design of a film.

The study presented in this thesis contributes to the creative process of sound recording by analysing the perceptual feature of the impression of spaciousness for the reconstructed sound fields of recordings. The understanding of how spaciousness can be controlled, either by choice of microphone technique or by signal processing, has been the core of the experimental work which ultimately led to the development of a spaciousness processor in the form of a VST and AU plugin. All of the work conducted for this thesis was motivated by the notion that sound recordings can communicate emotions and sensations. That is, sound recording is an *art*. Therefore, the research undertaken for this thesis was on the control of the perceptual effect of the reconstructed sound field, with a focus on the impression of spaciousness.

Spaciousness is appreciated as a strong positive and desirable component of concert hall acoustics, which fact has been confirmed by extensive architectural acoustics studies. However, spaciousness in sound reproduction, which is parallel to that experienced in concert halls, has not been comprehensively dealt with previously. Surround sound systems with laterally placed speakers reproducing uncorrelated signals can create this desirable feature. But, if such a feature is desired why have surround sound formats, such as 5.1, not become the norm for music production? Despite the fact that sound for films, and to a lesser degree sound installations and contemporary electroacoustic music, have been delivered in surround sound formats for many years now, it is the case that musical presentations of recorded music are still mostly presented in 2-channel stereo despite the increasing domestic availability of 5.1 reconstruction systems. There seem to be questions that need to be answered before the effort of moving to more channels for recorded music reconstruction can be considered. How signals can be controlled, and what choices can be made in the recording process, in order to influence the resulting spaciousness experienced by listeners, are questions that were raised for this study.

Stereophonic reproduction using only Left and Right channels is appreciated as being a reliable format for delivering good localized sonic imaging and other perceptual features. However, there are perceptual characteristics which ultimately will provide a "better sounding recording" that should be addressed if surround sound formats (*e.g.* 5.1) are to be used. It was the intent of the research study undertaken for this thesis to provide answers, and also to raise other questions regarding the improvement of spaciousness in sound recording reconstruction.

9.2 IACC as an index for spaciousness

The adoption of IACC as an index for measuring spaciousness was considered because of the fact that it is the differences in the signals presented at the Left and Right ears which will influence the perceptual impression of auditory spaciousness. IACC is a binaural measure that represents these differences, and was presented here along with another measure for spaciousness known as Lateral Energy Fraction. There has been extensive investigation related to the use of these techniques to investigate spaciousness, since there still exists issues and confusions regarding the results provided by these techniques (Ando, 1985; Potter, 1993; Okano, Beranek, & Hidaka, 1998; Barron, 2000; Okano, 2002; Mason, 2002). IACC is considered a standard for measuring spaciousness and envelopment (International Organization for Standardization (ISO), 2009) and given that the IACC measure relates directly to listener ear signals, as compared to the more indirect Lateral Energy Fraction, it was decided to use IACC as a spaciousness index.

It is emphasized that IACC was here used as a measure which correlates with the impression of spaciousness, and is not the cause of it. By noting the results produced from the measurements undertaken, it is possible to compare how IACC values changed for each of the variations made to the microphone array parametric details, and also due to the signal processing introduced in the recordings. This comparative observation indicates that the listener impression of spaciousness is being affected by such alteration of the recorded signals.

For some cases in the experimental work, the IACC differences, when compared from one setup to the other, did not produce significant IACC differences according to the research results available for IACC Just Noticeable Difference (JND). However, it has been established that an IACC reduction is correlated with judgments of spaciousness preference, even if a definitive value of JND for IACC is still not entirely clear. In fact, IACC JND values seem to vary with different signal stimuli, and with respect to frequency. IACC is strongly related to spaciousness, and the processor control developed allows production control of listener IACC, and of the listener spaciousness experience. This is not to say that there are no other parameters which contribute to spaciousness, beyond inter-aural differences. It is certain, however, that differences between the ear signals will contribute to perceived spaciousness, and this was the focus here.

9.3 Stereo recording techniques and spaciousness

Stereo, and more recently surround system, microphone techniques have been investigated over the years. The improvements and refinement of these techniques have been mainly concerned with the localisation aspects of the reconstructed sound fields. However, as presented in this thesis, published results exist which show that producers and also engineers have been of the opinion that stereo recording and reconstruction is capable of more than just merely enabling phantom image localisation between the loudspeakers.

Better sounding recordings will always be a concern for recordists, and it is this concern that is investigated by the author in this thesis. Source localisation is an important aspect of a reconstructed sound field, but other perceptual characteristics are also of concern. Subjective statements concerning the perception of the perceived attributes of different microphone techniques often involve loose comments such as their being able to produce a more “airy” sound, or an increased “roominess”, or “lacking in space”. These do not help in understanding how a microphone technique can behave in delivering signals that will affect the perceptual feature of spaciousness for listeners but they do indicate the perceptual impression experienced with different stereo microphone techniques. It is worthwhile to provide studies that can help producers and artists to more fully understand and control how a recording can be made to sound better. It is suggested here that not only is an increased understanding of the perceptual significance of the physical details of microphone techniques and their operation important, but that also to have the knowledge so as to be able to use these techniques to enhance the perceptually artistry of sound recording is of considerable note. The experienced spaciousness of reconstructed recordings is a feature which needs to be better understood.

9.4 Measuring using IACC

The first objective of the work undertaken for this thesis was to establish *perceptually significant* metrics, based on those used in room acoustics, by which spatial hearing criteria, and the associated "auditory impression", could be assessed for a 2-channel stereo, 3-channel stereo, or a surround sound recording and reconstruction context. As a contribution to the field, this objective was achieved by firstly establishing if the measurement setups used were room independent, *i.e.* could they be used independent of the rooms in which these sound field reconstruction setups were installed. Most research regarding perceptual features relating to spatial impression is conducted in anechoic chambers, which environments are difficult to access for many researchers. By conducting experiments and comparing the measured IACC trends for controlled stimulus situations in three different test environments, one of which was an anechoic chamber and was used as a reference environment, it was established that the measured reconstructed sound field IACC trends for changing microphone and reconstruction format details were similar for all environments. This experiment was of great importance, since the author did not have an anechoic chamber at his disposal, and it meant that further experiments could be undertaken in the other available rooms used throughout the experimental period of this thesis work. Also, it is worthwhile commenting that the results achieved from the IACC trend measurements in these different rooms are a contribution to researchers who are interested in conducting IACC measurements using artificially created reflections, but do not have easy accesses to anechoic chambers.

Further research was conducted, in the rooms available to the author, in order to comparatively assess the reconstruction effectiveness of a number of surround microphone arrays, by comparing spatial hearing measurements for a primary (*i.e.* auditorium) space with those of a secondary (*i.e.* listening room) space. For this comparison, audio scenes with different IACC values were first generated and recorded in particular test environments. The IACC results achieved in the secondary listening environment showed that microphone parametric variation can be employed to facilitate spaciousness control for reconstructed sound fields. Parametric variation, such as the introduction of inter-microphone delay, leads to a reduced IACC peak value, and an increased sense of spaciousness, none of which is news, but which does confirm the identification of the necessary conditions for auditory spaciousness perception. That is, if listeners are presented with significantly dissimilar ear signals there will be an increased sense of "space" or "roominess" experienced. The importance of such experiments, and the results obtained, is that they will help understand which details can be changed in microphone arrays in order to influence a listener's perceptual impression of spaciousness. Producers and sound engineers will ultimately better understand how to create sound field reconstructions with an enhanced spatial audio impression, and how to apply this control in reproduction formats such as 5.1, which although by now widely available, is still not a norm for most of the music produced or reconstructed today.

It was established that changes in IACC peak values could occur in reconstructed sound fields based on the choice of microphone technique, and on changes in setup details of the arrays (*e.g.* inter-capsule distance, polar pattern characteristics and angle of microphones) in recording primary environment sound fields. Experimental work was undertaken in order to see how changes in reproduction format (*e.g.* 2-channel stereo, 3-channel stereo and 5.1), or in the introduction of signal processing of the recorded signals, such as shuffling and up-mixing techniques, could also provide IACC variation possibilities. The results achieved were also conclusive, since they showed that the application of shuffling, and up-mixing techniques for stereo recordings could provide signals to the ears of listeners which by their controlled dissimilarity will present an enhancement to the impression of spaciousness experienced. These findings are important since it has been shown that spaciousness can be controlled for both stereo and 5.1 reconstruction systems. Also, such a process for objectively assessing the spatial sound reconstruction characteristics using different microphone techniques and reconstruction formats has not, to the author's knowledge, been previously presented.

9.5 Spaciousness processor

The last objective proposed for the work undertaken was that of the development of a spaciousness processor for 5.1 system use which derives from standard 2-channel recording methods, and is based on measured perceptually significant parameter changes. Such a processor was based on the inspiring work of Michael Gerzon who, in the 1970's, detailed one method of presenting surround sound derived from 2-channel stereo recordings. However, the processor developed here differs from the method proposed by Gerzon in that for this study the illusion of spaciousness was the main concern, while that of the setup devised by Gerzon was aimed at the achievement of sound source localisation from all around using conventional 2-channel stereo recordings over a quad array of speakers presented in a "diamond" shape configuration. The concern of this thesis was to examine the parameter changes which could be made in the recording process and in post-processing of the recorded left and right signals so as to enhance the experienced listener spaciousness using a well-established surround sound format such as the ITU-R BS.775-1 (*i.e.* 5.1).

In the development of a processor which producers, musicians, and artists alike can employ to alter the spatial experience of the recorded sound listener, lies most of the originality of the work undertaken. The processor facilitates control over different signals derived from 2-channel stereo recordings, how they are subsequently fed to the loudspeakers, and also how they are to be filtered. The contribution to the field is that with such a processor (which ultimately can be used not just by sound engineers, but also home listeners) the degree of spaciousness impression to be experienced is left to individual taste and artistic expression for any reconstructed sound field. There are a number of good aspects about "standard" recording techniques, but here it is identified that there are features of auditory spaciousness which are not well recorded or reproduced, since certain recording techniques fail to present signals at the ears of listeners' ears which are necessary to create the dissimilarities needed for perceived spaciousness. This thesis has been concerned with identifying a technological means of changing that.

9.6 Subjective listening test

A subjective listening test was conducted in order to reinforce the findings from the IACC measurements, and to allow "closure" of the research outlined in this thesis permitting the effective validation of the objective measurements undertaken.

The subjects were presented with selected examples of un-processed stereo recordings and their spaciousness processed versions and were asked to score the spaciousness attribute of the processed version in comparison to the un-processed version. Also subjects were asked to compare some selected stereo microphone recording techniques with surround microphone recording techniques and different stereo recordings.

From all the results obtained from the subjective listening test it has been established that the objective measurements explain 90.9% of the average subjective perceptual variation.

9.7 Future work

This thesis has presented several answers relating to the field of perceived "auditory spaciousness" for reconstructed sound fields using 2-channel stereo, 3-channel stereo, and 5.1. Nevertheless, there are also several questions that have been raised by the work undertaken which should be addressed in future work.

Some issues to be considered are:

- Conduct studies in order to determine whether the spaciousness processor is introducing artefacts to the spatial quality of the recorded material. It is required that the spaciousness processor should only enhance auditory spaciousness, and while doing so it should not introduce unpleasant artefacts to the recordings (*e.g.* phasiness, locatable changes in sound sources, timbre changes)

- Although 5.1 surround systems are widely available for home listening, there has been a market tendency towards an increased use of headphone listening, mostly initiated by the audio industry investment in portable audio systems (*e.g.* smartphones, mp3 and iPod players, computer laptops). In light of this, investigating how the spaciousness processor could be optimised for headphone presentation is something that should be paid attention to.
- The effectiveness of the spaciousness processor in conjunction with motion pictures and audio for gaming is a feature that is to be investigated. Providing a spacious control of the reconstructed sound field to listeners, while viewing films and playing video games, might be a desirable feature in the overall experience.
- In the case of MS recording techniques, how might different M signals, *i.e.* different directivities in the microphones used for the M signal, influence IACC peak values?
- Can Just Noticeable Difference for IACC be investigated in order to make it more consistent and definite?
- There is a general agreement that IACC is related to spaciousness. Are there other dimensions of perceived spaciousness, such as sound level at lower frequencies (Beranek, 1996), which should be investigated?

Answers to these questions and the proposed research topics could contribute to the ultimate quest of: "how can we make it sound better?"

GLOSSARY

AB	Spaced stereo recording technique, generally uses two omni-directional microphones spaced apart.
ACF	Autocorrelation Function.
ASW	Apparent Source Width.
AU	Audio Units are a system-level plug-in architecture provided by Core Audio in Mac OS X developed by Apple Computer.
BIR	Binaural Impulse Response.
DAW	Digital Audio Workstation.
Double MS	Coincident surround recording technique which uses double mid side configuration.
IACC	Inter-aural Cross Correlation.
IRT cross	Near-coincident surround recording technique, also known as "Theile" or "Atmos" cross, proposed by Gunther Theile.
ITD	Inter-aural Time Delay
LEV	Listener Envelopment.
LF	Lateral Energy Fraction.
LG	Lateral Sound Level.
MLS	Maximum Length Sequence: is a type of pseudorandom binary sequence.
MS	Coincident stereo recording technique arranged in a mid side configuration.
NOS	*Nederlandsche Omroep Stiching*: Near-coincident stereo recording technique.
OCT	Optimized Cardioid Triangle: near-coincident stereo recording technique, proposed by Gunther Theile.
OCT Surround	Optimized Cardioid Triangle Surround: near-coincident

	surround recording technique, proposed by Gunther Theile.
ORTF	*Office de Radiodiffusion-Télévision Française*: Near-coincident stereo recording technique.
ORTF Surround	*Office de Radiodiffusion-Télévision Française Surround*: Near-coincident surround recording technique.
RT	Reverberation time.
VST	Virtual Studio Technology: makes use of digital signal processing to simulate traditional recording studio hardware in software.
XY	Coincident stereo recording technique arranged in a crossed configuration.

APPENDIX I – Magnitude and Phase Response of a Blumlein Pair

Figure I to Figure IV show the results of a Blumlein pair recording a single, frontally located, loudspeaker (Fostex 6301B) which was fed with pink noise. Using Smaart (Rational Acoustics, LLC, 2013) to plot either magnitude or phase against frequency for each channel of the Blumlein array, it is possible to see how the phase relationship between channels changes as each of the microphone quadrants (*i.e.* 0°, 90°, 180° and 270°) is pointed at the sound source. The black line represents to the Left channel output of the array, while the red line represents to the Right channel output of the array. It can be seen that when the array is at 0° (Figure I), the magnitude and phase relationship are similar in both channels, and when the array has its 180° (Figure II) quadrant pointing at the source, the magnitude is maintained but now with a polarity reversal. Rotating the microphone so that either the 90° or 270° quadrants are facing the sound source, alternately, will produce the same magnitude but with an out-of-phase relationship between Left and Right channels.

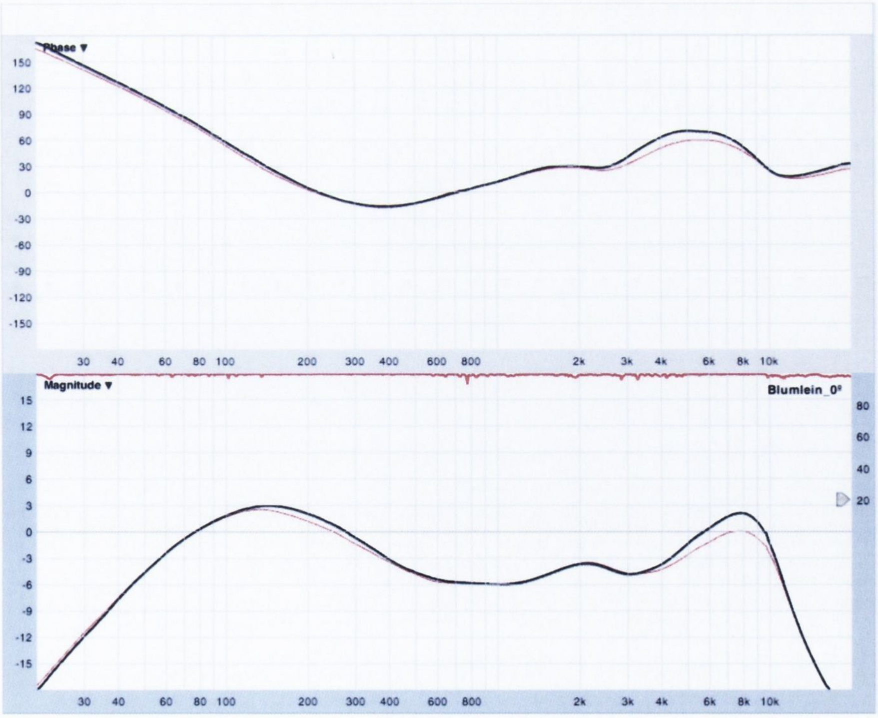


Figure I: Magnitude and Phase relationship between Left and Right Channel of a Blumlein pair. Black line is Left channel and red line is Right channel. The array is facing the sound source at 0°. Phase is presented from 0° to ±180°; Magnitude is presented in a range of ±15 dB; Frequency is presented in 1/3 octave bands.

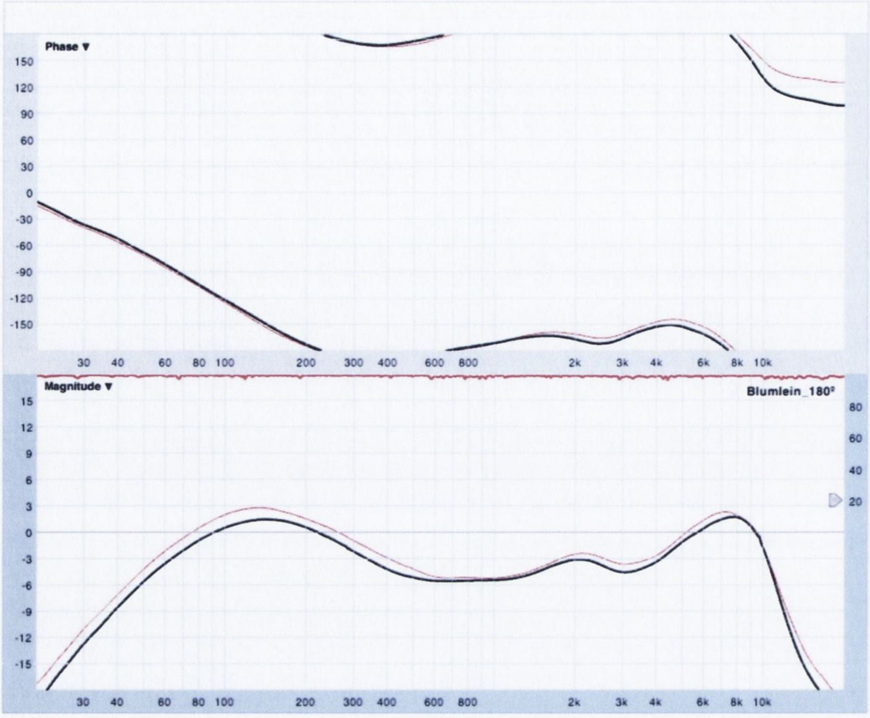


Figure II: Magnitude and Phase relationship between Left and Right Channel of a Blumlein pair. Black line is Left channel and red line is Right channel. The array is facing the sound source at 180°. Phase is presented from 0° to ±180°; Magnitude is presented in a range of ±15 dB; Frequency is presented in 1/3 octave bands.

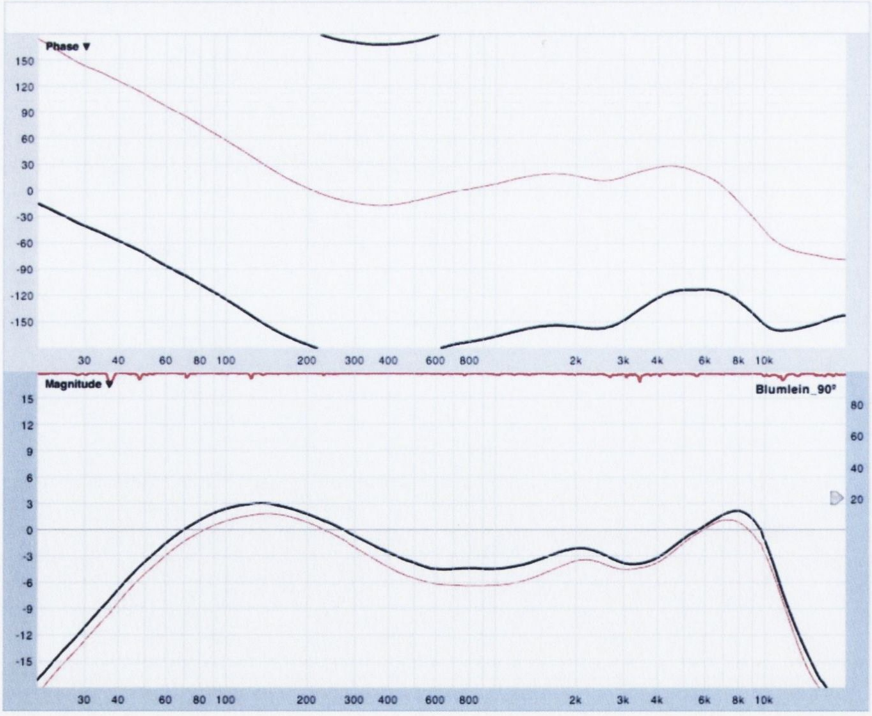


Figure III: Magnitude and Phase relationship between Left and Right Channel of a Blumlein pair. Black line is Left channel and red line is Right channel. The array is facing the sound source at 90°. Phase is presented from 0° to ±180°; Magnitude is presented in a range of ±15 dB; Frequency is presented in 1/3 octave bands.

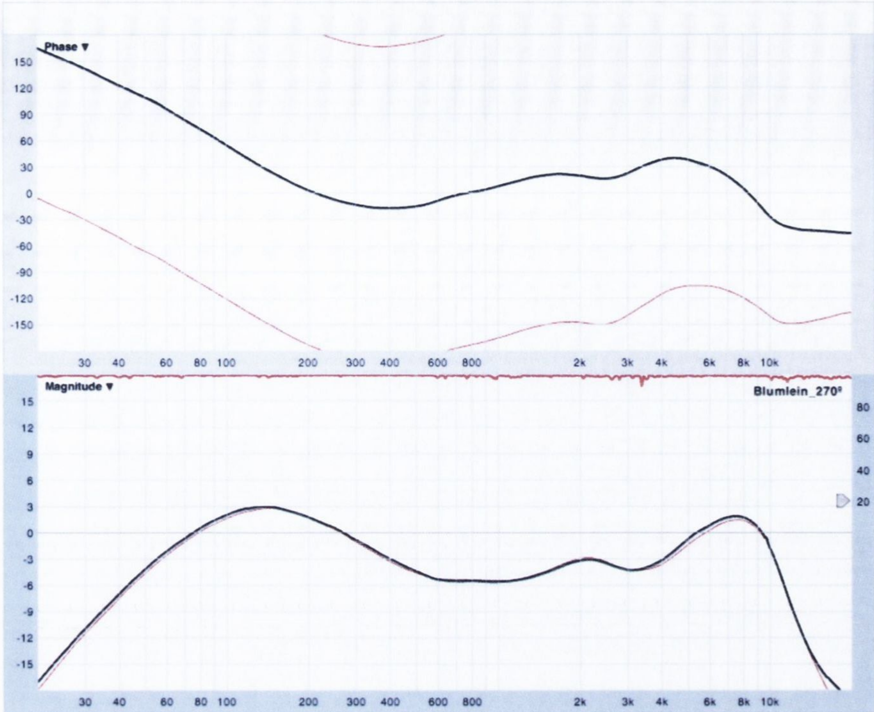


Figure IV: Magnitude and Phase relationship between Left and Right Channel of a Blumlein pair. Black line is Left channel and red line is Right channel. The array is facing the sound source at 270°. Phase is presented from 0° to ±180°; Magnitude is presented in a range of ±15 dB; Frequency is presented in 1/3 octave bands.

For the 90° (Figure III) side of the array the Right channel will be in-phase, when compared to the 0° quadrant, and the Left channel will have a polarity reversal. The same happens at 270° (Figure IV) of the array, but now the Right channel is in polarity reversal and the Left channel is in-phase. When Left and Right channel of a Blumlein pair are added together into mono, the out-of-phase regions will cancel which is good for monophonic compatibility.

APPENDIX II – IACC Results for Centre Front Direct and Indirect Components

Could it be possible that the IACC peak value was varying because of the normalization power of the measurement used and not because of inter-aural differences? To answer this question an experiment conducted in Oporto studio was undertaken similar to that described in Section 6.2.1, but instead of having a laterally placed early reflection, this component was made centre front as the direct sound component. The dummy head recorded the resulting signals played back from the centre front direct and indirect sound and IACC measures were calculated from the binaural impulse responses. The delays used for the early reflection were the same (10, 30 and 50ms) and the amplitude of the reflection varied in a similar fashion, as described in Section 6.2.2. From Figure V it can be observed that there are no changes of IACC peak value which maintained practically the same (IACC = 0.83) for all simulated delayed reflections and relative levels used. From the results it is concluded that the IACC changes are indeed related to inter-aural differences introduced from the single lateral early reflection.

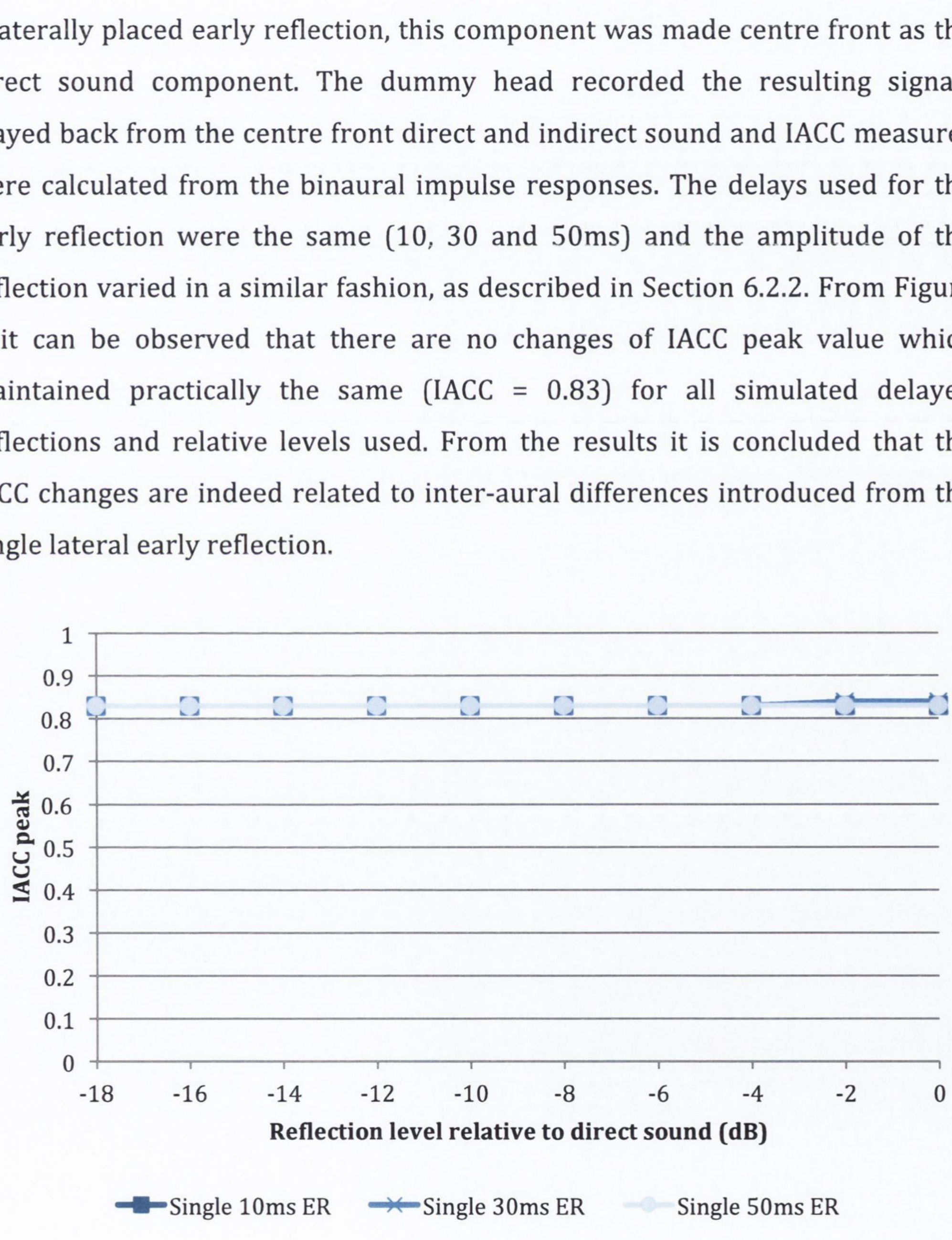


Figure V: IACC measurements with a single centre front (*i.e.* originated from where the direct sound was originated) early reflection at 10ms, 30ms, 50ms.

APPENDIX III – IACC Results for Octave Bands, Full Bandwidth, and Full Bandwidth with A-Weighting Filtering

Figure VI to Figure VIII show the results of different IACC measurements for several octave bands, full bandwidth, and full bandwidth with A-weighting filtering. These results were obtained from the binaural recordings made in Helsinki, where a direct sound and a single lateral reflection positioned at 60° from the left, with 10ms, 50ms and 80ms delay, were presented to the ears of a dummy head (see Figure 6-2) in an anechoic environment. It can be seen that the IACC trends are broadly similar and that as the reflection level is increased the IACC peak drops accordingly.

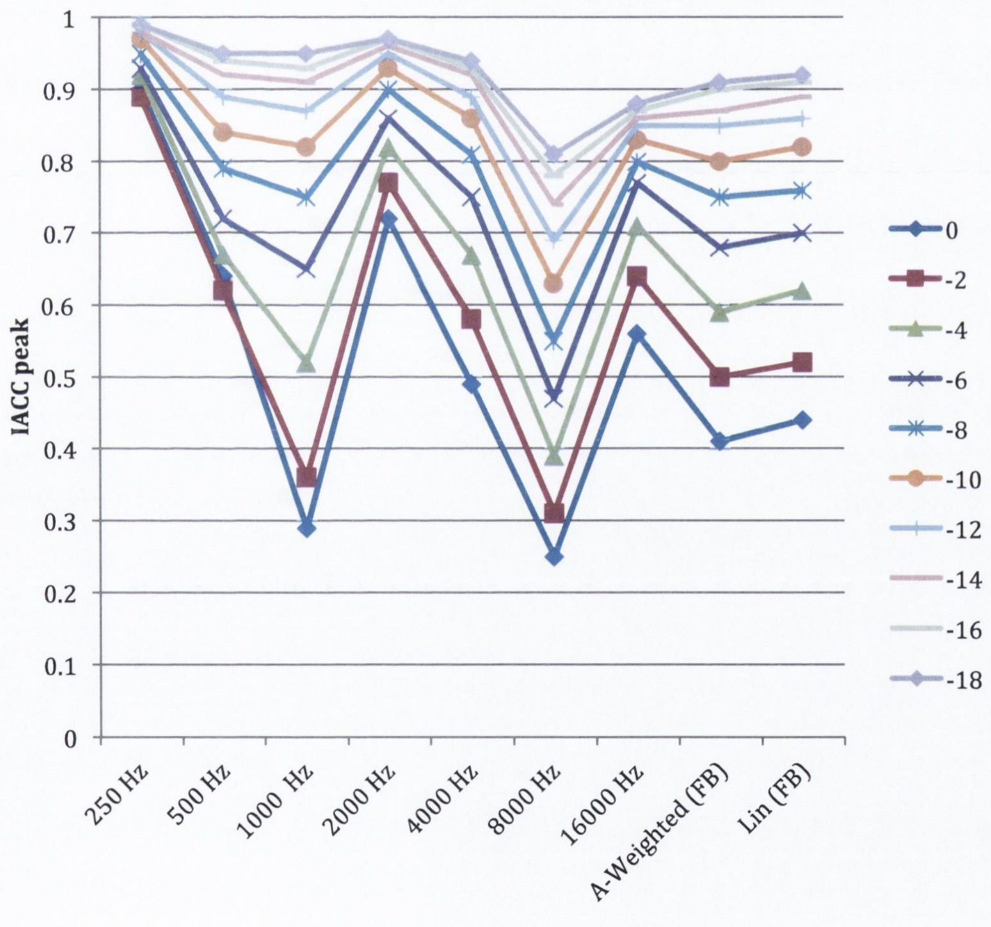


Figure VI: IACC measurements with early reflection at 10ms, for 7 octave bands, A-weighted filtered full bandwidth, and full bandwidth with no filtering applied. Each colour-coded curve represents the relative gain, in dBFS, of the simulated early lateral reflection in relation to the direct sound, from -18 to 0dB. The IACC measurements were obtained using the Aurora Acoustical Analysis plugin (Farina, 2007). Results are from Helsinki.

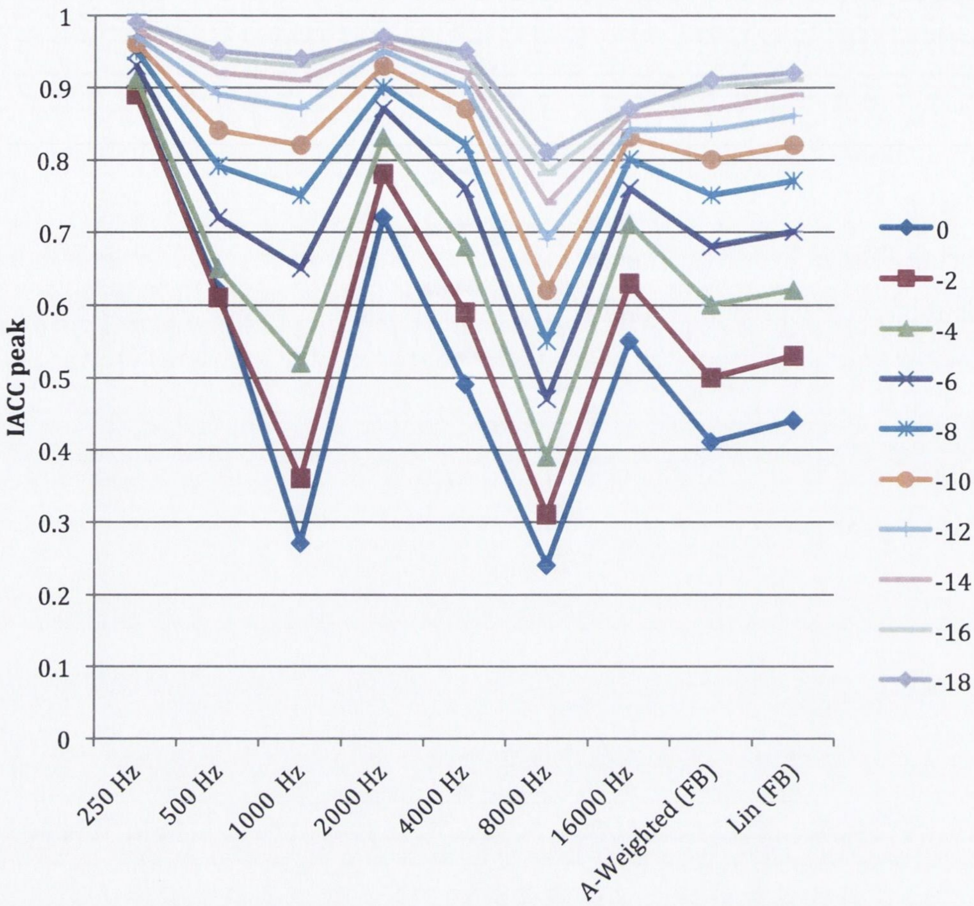


Figure VII: IACC measurements with early reflection at 30ms, for 7 octave bands, A-weighted filtered full bandwidth, and full bandwidth with no filtering applied. Each colour-coded curve represents the relative gain, in dBFS, of the simulated early lateral reflection in relation to the direct sound, from -18 to 0dB. The IACC measurements were obtained using the Aurora Acoustical Analysis plugin (Farina, 2007). Results are from Helsinki.

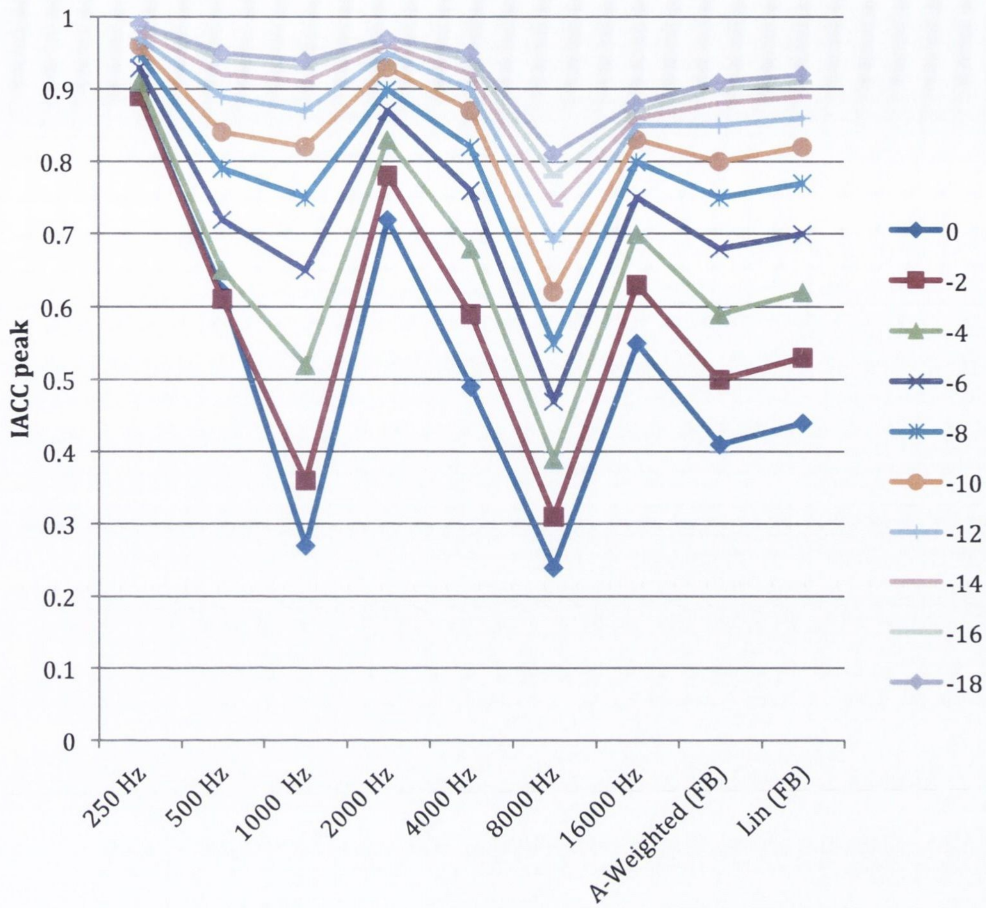


Figure VIII: IACC measurements with early reflection at 50ms, for 7 octave bands, A-weighted filtered full bandwidth, and full bandwidth with no filtering applied. Each colour-coded curve represents the relative gain, in dBFS, of the simulated early lateral reflection in relation to the direct sound, from -18 to 0dB. The IACC measurements were obtained using the Aurora Acoustical Analysis plugin (Farina, 2007). Results are from Helsinki.

In light of these results, IACC measurements using a full bandwidth test signal can be used, since the trend of IACC decrease, as a function of amplitude of the reflection, is preserved irrespective of any spectral weighting. All of these measurements were conducted in Helsinki, using an anechoic chamber.

APPENDIX IV – IACC Results for Single and Multiple Reflections

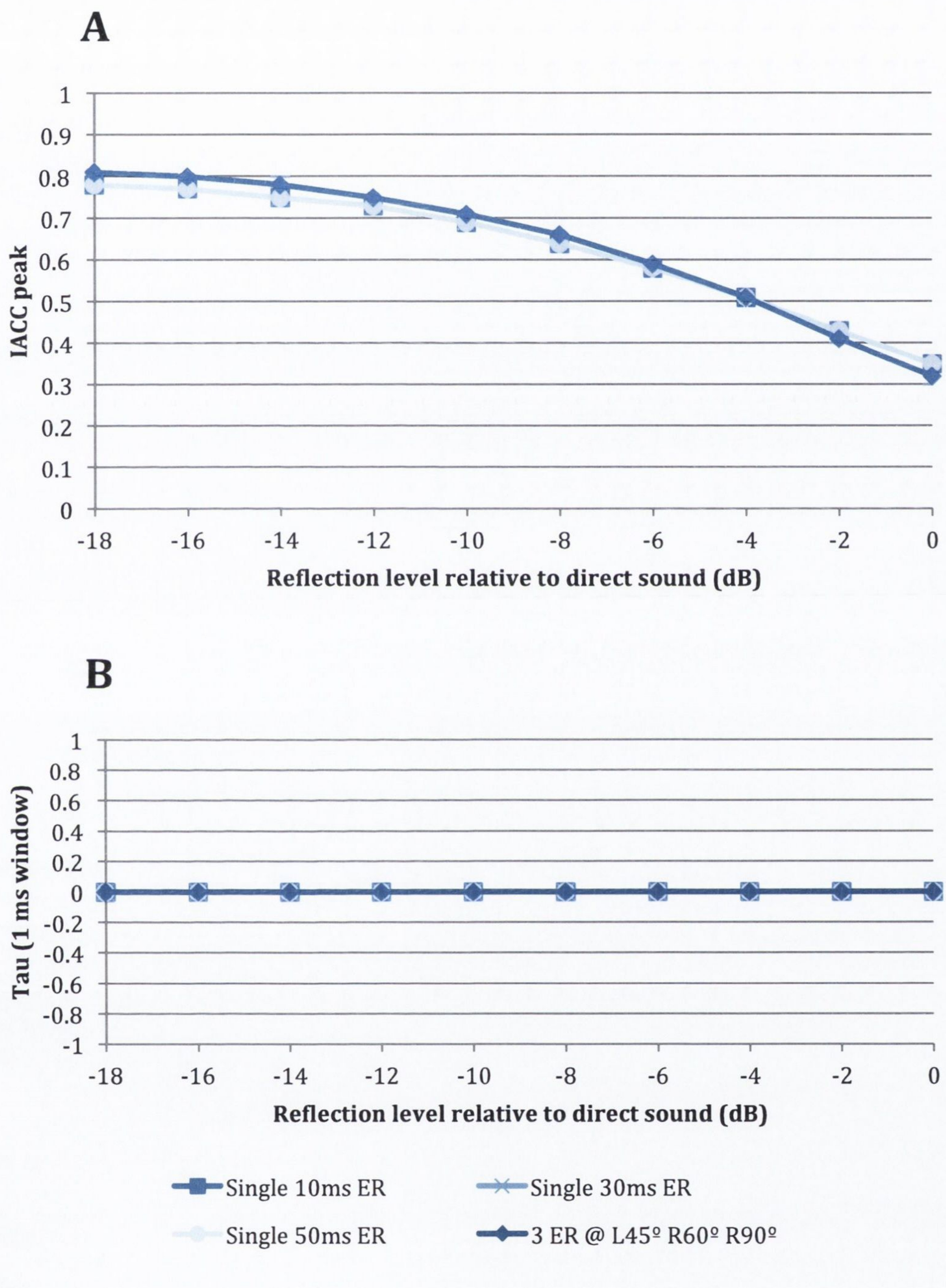


Figure IX: IACC (A) and ITD (B) measurements with early reflections at 10ms, 30ms, 50ms and with multiple (3) reflections. In ITD (B) a consistent Tau value indicates that there is no image shift; all the measurements, made in Oporto, indicate no change of ITD.

It has been established that preference tests of sound fields with multiple reflections give similar results as those obtained for a one reflection sound field (Ando & Gottlob, 1979). The above Figure IX presents the IACC results obtained using the author's test arrangement using one single reflection with 10ms, 30ms and 50ms delay (see Section 6.2.2), and also the results obtained using 3 reflections, one positioned 45° to the right of the dummy head and two positioned to the left of the dummy head at 60° and 90°. These 3 reflections had 10ms, 30ms and 50ms delays, respectively (Figure X). The total amplitude of the 3 reflections was normalized with that of just one single reflection. It can be seen that the IACC trends remain similar as a function of reflection amplitude irrespective of reflection number.

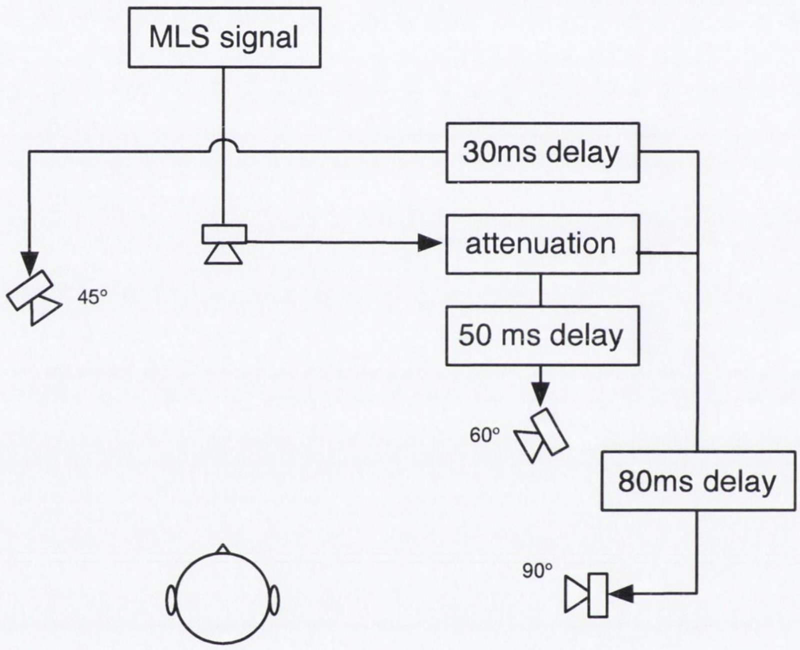


Figure X: Direct sound with multiple lateral reflections.

APPENDIX V - Spaciousness Assessment Questionnaire

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

You will be presented with audio examples for pair wise comparison between example A and example B. A tactile tablet device will act as a switch allowing you listen to each of the examples (A and B) as you wish. You can listen to the examples as many times as you wish and can take as long as is needed to make your judgments.

You are asked to judge each example **solely in terms of perceived spaciousness.** How much more spacious is example **B** when compared to **A**?

When answering, you may wish to consider some attributes that have been used to describe perceived spaciousness:

> ***Source width***. Does the sound source (instrument/voice) appear to be broad?
>
> ***Performance environment.*** Does the environment appear to be broad and spatially extended?

To grade this you are asked to mark the following scale:

-2	-1	0	1	2
much less Spacious	less Spacious	equal	more Spacious	much more Spacious

For each example, please mark your preference in the sheet provided.

Once you have recorded your answers please say "Finished" or words to that effect so that the next example may be presented. Please allow about 10 seconds before pressing the A or B key again.

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

APPENDIX VI – Resulting Publications and Presentations

Following are the titles and abstracts of papers and poster presentations which are outcomes of the research undertaken for this thesis.

- **Influence of Different Microphone Arrays on IACC as an Objective Measure of Spaciousness** (Conceição & Furlong, 2013). – *Presented at the 134th AES convention in Rome.*

  Abstract

  Inter-Aural Cross Correlation measurements are used as physical measures which relate to listener spaciousness experience in a comparative study of the influence on spaciousness of different microphone arrays, thus allowing an objective approach to be adopted in the exploration of how microphone arrays affect the perceived spaciousness for stereo and surround sound reconstructions. The different microphone arrays recorded simulated direct and indirect sound components. The recorded signals were played back in three different rooms and IACC measurements were made for the reconstructed sound fields using a dummy head microphone system. The results achieved show how microphone array details influence the IACC peak, and lead to a better understanding of how spaciousness can be controlled for 2 channel stereo and 5.1 presentations. Parametric variation of microphone arrays can therefore be employed to facilitate spaciousness control for reconstructed sound fields.

- **Spaciousness Control in Stereo and 5.0** (Conceição & Furlong, 2011). – *Presented at the International Conference on Spatial Audio in Detmold*

  Abstract

  Stereo and surround microphone techniques are investigated in relation to perceived spaciousness. This paper examines what factors can influence the perception of spaciousness in stereo and surround recordings, and examines if 5.1 system reproductions can improve perceived spaciousness, and if they allow for a greater control of it. By using IACC to objectively assess the perceived spaciousness of controlled stereo and surround recordings, a greater understanding of the factors that influence overall spaciousness perception in reconstructed sound fields was achieved. Comparative recordings of classical music, for subjective assessment, were presented using coincident and non-coincident techniques, which demonstrated different spatial reconstruction possibilities. By using *shuffling* techniques and MS processing for up-mixing stereo to 5.0, it is demonstrated that spaciousness perception can be controlled for stereo recording using 5.0 audio reconstruction.

- **Influence of Different Test Room Environments on IACC as an Objective Measure of Spatial Impression or Spaciousness** (Conceição & Furlong, 2011). – *Presented at 131st AES conventionin New York.*

  Abstract

  To investigate the perceptual impression of spaciousness, a physical measure, which relates to listener spaciousness experience, was used. A variable setup was introduced that made possible the control of spaciousness in different rooms. IACC measurements were made using a frontal loudspeaker for the direct sound, with a second loudspeaker for an angled single early reflection positioned in the horizontal plane. Measurements were performed under controlled conditions in which a dummy-head measured Inter Aural Cross Correlation was captured for variable sound fields. It was possible to conclude that there is a similar trend in IACC results from the repetition of the experiments in different rooms. That is, measurement room acoustic details are not crucial to observed trends in IACC variation.

- **Influence of Different Microphone Arrays on IACC as an Objective Measure of Spatial Impression or Spaciousness** (Conceição & Furlong, 2012). – *Poster session of the 25th AES UK Conference: Spatial Audio in Today's 3D World*. York, UK.

BIBLIOGRAPHY

Ando, Y. (1985). *Concert Hall Acoustics.* New York: Springer-Verlag.

Ando, Y. (1977). Subjective preference in relation to objective parameters of music sound fields with a single echo. *Journal of the Acoustical Society of America , 62* (6), 1436-1441.

Ando, Y., & Gottlob, D. (1979). Effects of Early Multiple Reflections on Subjective Preference Judgments of Music Sound Fields. *Journal of the Acoustical Society of merica* , 524-527.

Ando, Y., & Kageyama, K. (1977). Subjective Preference of Sound with a Single Early Reflection. *Acustica , 37*, 111-117.

Audio Enginering Society. (1986). *Stereophonic Techniques - An Anthology of reprinted articles on stereophonc techniques.* New York: Audio Enginering Society.

Bach, J. (2000). Eduardo Eguez - The Lute Music of Johann Sebastian Bach, volume one. MA Recordings - audio CD.

Bachelard, G. (1986). *The New Scientific Spirit.* Beacon Press.

Barron, M. (2000). Measured Early Lateral Energy Fraction in Concert Halls and Opera Houses. *Journal of Sound and Vibration , 232* (1), 79-100.

Barron, M. (1971). The subjective effects of first reflections in concert halls - the need for lateral reflections. *Journal of Sound and Vibration* , 475-494.

Barron, M., & Marshall, A. H. (1981). Spatial Impression due to early lateral reflections in concert halls: the derivation of physical measure. *Journal of Sound and Vibration* , 211-232.

Bartlett, B., & Bartlett, J. (1999). *On-Location Recording Techniques.* Focal Press.

Bartlett, B., & Bartlett, J. (1999). *On-Location Recording Techniques.* Woburn: Focal Press.

Bech, S., & Zacharov, N. (2006). *Perceptual Audio Evaluation - Theory Method and Application.* Chichester, UK: Willey and Sons.

Beranek, L. (1996). Acoustics and Musical Qualities. *Journal of the Aocustical Society of America , 99* (5), 2647-2652.

Beranek, L. (1996). *Concert and Opera Halls: How They Sound.* Acoustical Society of America.

Blauert, J. (1997). *Spatial hearing: The Psychophysics of Humann Sound Localization.* Cambridge, Mass.: MIT Press.

Blauert, J., & Lindemann, W. (1986). Auditory Spaciousness: Some further psychoacoustic analyses. *Journal of the Acoustical Society of America , 80* (2), 533-542.

Blesser, B., & Salter, L.-R. (2007). *Spaces Spaek, Are You Listening. Experiencing Aural Architecture.* Cambridge: The MIT Press.

Blumlein, A. D. (1933). *Patent No. 394,325.* UK.

Bradley, J. S., & Soulodre, G. A. (1996). Listener envelopment: An essential part of good concert hall acoustics. *Journal of the Acoustical Society of America , 99* (1), 22-22.

Bradley, J. S., & Soulodre, G. A. (1995b). Objective measures of listener envelopment. *Journal of the Acoustical Society of America , 98* (5), 2590-2597.

Bradley, J. S., & Soulodre, G. A. (1995a). The Influence of Late Arriving Energy on Spatial Impression. *Journal of the Acoustical Society of America , 97* (4), 2263-2271.

Bregman, A. S. (1990). *Auditory Scene Analysis - The Perceptual Organization of Sound.* The MIT Press.

Bregman, A., & Ahad, P. A. (1995). Audio-CD, Track 5. *Demonstrations of Auditory Scene Analysis [Sound Recording]* . Auditory Perception Laboratory, Psychology Department, McGill University.

Brown, H. I. (1979). *Perception, Theory and Commitment, The New Philosophy of Science.* Chicago: University of Chicago Press.

Bruck, J. (1997). The KFM 360 Surround Sound: A Purist Approach . *103rd AES Convention* (p. 4637). New York: Audio Engineering Society.

Chernyak, R. I., & and Dubrovsky, N. A. (1968). Pattern of the Noise Images and the Binaural Summation of Loudness for the Different Interaural Correlation of Noise. *Procedings of the 6th International Congress on Acoustics*, (pp. A53-A56).

Clark, H. A., Dutton, G. F., & Vanderlyn, P. B. (1958). The 'Stereosonic' Recording and Reproduction System. *Journal of the Audio Engineering Society , 6* (2), 102-117.

Conceição, M., & Furlong, D. (2013). Influence of Different Microphone Arrays on IACC as an Objective Measure of Spaciousness. *134th AES Convention* (p. 8885). Rome: Audio Engineering Society.

Conceição, M., & Furlong, D. (2012, March 25-27). Influence of Different Microphone Arrays on IACC as an Objective Measure of Spatial Impression or Spaciousness. *Poster session of the 25th AES UK Conference: Spatial Audio in Today's 3D World* . York, UK.

Conceição, M., & Furlong, D. (2011). Influence of Different Test Room Environments on IACC as an Objective Measure of Spatial Impression or Spaciousness. *131st AES Convention* (p. 8555). New York: Audio Engineering Society.

Conceição, M., & Furlong, D. (2011). Spaciousness Control in Stereo and 5.0. *International Conference on Spatial Audio* (pp. 27-32). Detmold: Verband Deutscher Tonmeister and ETI, University of Music Detmold.

Cooder, R., & Bhatt, V. M. (2008, September 1). A Meeting by the River. *A Meeting by the River* . Water Liy Acoustics - audio CD.

Cox, T., Davies, W., & Lam, Y. (1993). The sensitivity of listeners to early sound field changes in auditoria. *Acustica , 79*, 27-41.

Denon. (1995). Anechoic Orchestral Music Recording. Performed by the Osaka Philarmonic Orchestra and conducted by Enkoji, Tsubonou and Kawaguchi. *Denon CD PG-6006* . Japan: Denon.

Dooley, W., & Streicher, R. (1982). M-S Stereo: A Powerfull Technique for Working in Stereo. *Journal of the Audio Engineering Society , 30* (10), 707-718.

DPA Microphones A/S. (2015). *5100 Mobile Surround Microphone.* Retrieved June 03, 2015 from DPA Microphones: http://www.dpamicrophones.com/en/products.aspx?c=Item&category=117&item=24312

DPA Microphones A/S. (2013). *Surround Techniques.* Retrieved August 19, 2013 from DPA Microphones: http://www.dpamicrophones.com/en/Mic-University/Surround-Techniques.aspx

Everest, F. A., & Pohlmann, K. C. (2009). *Master Handbook of Acoustics - Fifth Edition.* New York: McGraw-Hill.

Farina, A. (2007). *AURORA Plug-ins.* Retrieved July 16, 2013 from Aurora download: http://www.aurora-plugins.com/Aurora/download/

Fauconnier, G., & Turner, M. (2003). *The Way We Think: Conceptual blending and the mind's hidden complexities.* Basic Books.

Forsyth, M. (1985). *Buldings for Music: The Architect, the Musician, and the Listener from the Seventeenth Century to the Present Day.* Cambridge: The MIT Press.

Furlong, D. J. (1989). Comparative Study of Effective Soundfiled Reconstruction. *87th AES Convention* (p. 2842). New York: Audio Engineering Society.

Gardner, B., & Martin, K. (1994). *HRTF Measurements of a Kemar Dummy-Head Microphone.* MIT Media Lab Perceptual Computing Technical Report #280. MIT Media Lab.

Gerzon, M. A. (1971, August). A year of surround-sound. *Hi-Fi News* .

Gerzon, M. A. (1994). Applications of Blumelein Shuffling to Stereo Microphone Techniques. *Journal of the Audio Engineering Society , 42* (6), 435-453.

Gerzon, M. A. (1976). Blumlein Stereo Microphone Technique. *Journal of the Audiio Engineering Society , 24* (1), 36.

Gerzon, M. A. (1992b). Microphone Techniques for 3-Channel Stereo. *93rd AES Concention* (p. 3450). San Francisco: Audio Engineering Society.

Gerzon, M. A. (1992a). Optimum Reproduction Matrices for Multispeaker Stereo. *Journal of the Audio Engineering Society, 40* (7/8), 571-589.

Gerzon, M. A. (1973). Periphony: With Height Sound Reproduction. *Journal of the Audio Engineering Society, 21*, 2-10.

Gerzon, M. A. (1971, July). Recording Techniques for Multichannel Stereo. *British Kinematography Sound and Television*, pp. 274-279.

Gerzon, M. A. (1974b, December). Stabilising stereo images. *Studio Sound*.

Gerzon, M. A. (1986, July). Stereo Shuffling: New Approach - Old Technique. *Studio Sound*.

Gerzon, M. A. (1970, August). Surround sound from 2-channel stereo. *Hi-Fi News*.

Gerzon, M. A. (1974a, December). Surround-Sound Psychoacoustics. *Wireless World*.

Gerzon, M. A. (1975). The Design of Precisely Coincident MIcrophone Arrays for Stereo and Surround Sound . *50th AES Convention* (pp. L-20). London: Audio Engineering Society.

Gerzon, M. A. (1990, June). Three Channels: The Future of Stereo? *Studio Sound*.

Gerzon, M. A. (1971, March). Why Coincident Microphones? *Studio Sound, 13*, pp. 117, 119, 140.

Gerzon, M. A., & Barton, G. J. (1992). Ambisonic Decoders for HDTV. *92nd AES Convention* (p. 3345). Viena: Audio Engineering Society.

Griesinger, D. (1999). Objective measures of spaciousness and envelopment. *AES 16th International Conference: Spatial Sound Reproduction* (pp. 16-003). Rovaniemi: Audio Engineering Society.

Griesinger, D. (1996). Spaciousness and Envelopment in Musical Acoustics. *101st AES Convention* (p. 4401). Los Angeles: Audio Engineering Society.

Griesinger, D. (1985). Spaciousness and Localization in Listening Rooms - How To Make Coincident Recordings Sound As Spacious As Spaced Microphone Arrays. *79th Convention* (p. 2294). New York: Audio Engineering Society.

Haas, H. (1972). The Influence of a Single Echo on the Audibility of Speech, Doctoral dissertation, University of Gottingen. *Reprinted in Journal of the Audio Engineering Society , 20,* 146-159.

Hamasaki, K., & Hiyama, K. (2003). Reproducing spatial impression with multichannel audio. *24th International Conference: Multichannel Audio, The New Reality* (p. 19). Banff: Audio Engineering Society.

HEXLER.NET. (2014). *Software - Touch OSC.* Retrieved 12 31, 2014 from Hexler.net: http://hexler.net/software/touchosc

Hirst, J. M. (2006). Spatial Impression in Multichannel Surround Systems. *PhD Thesis* . Salford, UK: University of Salford UK.

Holman, T. (2008). *Surround Sound Up and Running.* Burlington: Focal Press.

Ihde, D. (2009). *Postphenomenology and Technoscience.* Albany: Sunny Press.

Ihde, D., & Selinger, E. (2003). *Chasing Technoscience.* Bloomington: Indiana University Press.

International Organization for Standardization (ISO). (2009). *ISO 3382-1:2009 - Acoustics -- Measurment of room acoustic parameters -- Part 1: Performance spaces.* Geneva: ISO.

ITU-R BS.775-1. (1992-1994). *Multichannel stereophonic sound system with and without accompanying picture.* Geneva, Switzerland: International Telecomunications Union.

Jeffress, L. A. (1948). A Place Theory of Sound Localization. *Journal of Comparative Physiology and Psychology , 61,* 523-527.

Kendall, G. S. (1995). The Decorrelation of Audio Signals and Its Impact on Spatial Imagery. *Computer Music Journa , 19* (4), 71-87.

Kuhl, W. (1978). Raumlichkeit eine Komponente des Horeindrucks. *Acustica , 40,* 167-181.

Kurozumi, K., & Ohgushi, K. (1983). The relationship between the cross-corelation coeficient of two-channel acoustic signals and sound image quality. *Journal of the Acoustic Society of America , 74* (6), 1726-1733.

Lipshitz, S. P. (1986). Stereo microphone techniques...Are the Purists wrong? *Journal of the Audio Engineering Society , 34* (9), 716-744.

Lokki, T., Patyen, J., & Pulkki, V. (2008). Recording of Anechoic Symphony Music. *Acoustics 08 Paris* (pp. 6433-6438). Paris: Société Française d'Acoustique.

Marshall, A. H. (1967). A note on the importance of room cross-section in concert halls. *Journal of Sound and Vibration , 5* (1), 100-112.

Mason, R. (2002, February). Elicitation and measurement of auditory spatial attributes in reproduced sound. *Thesis submitted in fulfilment of the requirement of the degree of Doctor of Philosophy* . Surrey, UK: University of Surrey.

Missout, A. (2007). *SONICBIRTH.* Retrieved July 16, 2013 from SonicBirth v1.3.0: http://sonicbirth.sourceforge.net/

Morimoto, M., & Iida, K. (1993). A New Physical Measure for Psychological Evaluation of a Sound FIeld; Front/Back Energy Ratio as a Measure of Envelopment. *Journal of the Acoustical Society of America , 93* (4), 2282-2282.

Morimoto, M., & Iida, K. (1995). A pratical evaluation method of auditory source width in concert halls. *Journal of the Acoustical Society of America , 16*, 59-69.

Morimoto, M., & Maekawa, Z. (1989). Auditory Spaciousness and Envelopment. *Proceedings of the 13th ICA*, (pp. 215-218). Belgrade.

Nakayama, T., Miura, T., Kosaka, O., Okamoto, M., & Shiga, T. (1971). Subjective Assessment of Multichannel Reproduction. *Journal of the Audio Engineering Society , 19*, 744-751.

Noche, S. U. (1999). Será Una Noche. MA Recordings - audio CD.

Okano, T. (2002). Judgments of noticeble differences in sound fields of concert halls caused by intensity variation in early reflections. *Journal of the Acoustical Society of America , 111* (1), 217-229.

Okano, T., Beranek, L. L., & Hidaka, T. (1998). Relations Among Interaural Cross-Corelation Coeficeint (IACCe), Lateral Fraction (LFe) and Apparent Source Width in Concert Halls. *Journal of the Acoustical Society of America , 140,* 255-25.

Peltonen, T. (2000, October 23). A Multichannel Measurement System for Room Acoustics Analysis. *This Thesis has been submitted for official examination for the degree of Master of Science.* Espoo, Finland: Helsinki University of Technology.

Pollack, I., & Trittipoe, W. (1959). Binaural listening and interaural noise cross correlation. *Journal of the Acoustical Society of America , 31* (9), 1250-1252.

Potter, J. M. (1993). On the binaural modeling of spaciousness in room acoustics. *PhD thesis* . Delft: Technische Universiteit Delft.

Pulkki, V. (2002). Microphone techniques and directional quality of sound reproduction. *Convention Paper 5500.* Munich: AES 112th Convention.

Rational Acoustics, LLC. (2013). *Smaart Sofware.* Retrieved August 06, 2103 from Rational Acoustics: http://www.rationalacoustics.com/store/smaart.html

Rayburn, R. A. (2012). *Eargle's Microphone Book.* Oxford: Focal Press.

Read, O., & Welsh, W. L. (1959). *From Tin Foil to Stereo.* Indianapolis, Indiana: Howard Sams.

Reichardt, W., & Lehmann, U. (1978). Raumeindruck als Oberberiff von Raumlichkeit und Halligkeit. Erlauterung des Raumeindrucksmasses R. *Acustica , 40,* 277-289.

Rumsey, F. (2004). *Desktop Audio Technology; digital audio and MIDI principles.* Oxford: Focal Press.

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

Sabine, W. C. (1923). *Collected Papers on Acoustics. Cambridge: Harvard University Press.* Retrieved August 24, 2012 from The Virtual Laboratory: Essays and Resources on the Experimentation of Life: http://vlp.mpiwg-berlin.mpg.de/library/data/lit39364

SCHOEPS GmbH. (2013a). *KFM Surround.* Retrieved July 16, 2013 from Schoeps Mikrofone: http://www.schoeps.de/en/products/categories/KFM-Surround

SCHOEPS Gmbh. (2015). *ORTF Surround.* Retrieved June 03, 2015 from Schoeps Mikrofone: http://www.schoeps.de/en/products/categories/ORTF-Surround

SCHOEPS GmbH. (2006, November). Schoeps Surround Brochure. *Surround Recording Techniques* . Karlsruhe, Germany: Shoeps Mikrofone.

SCHOEPS GmbH. (2014). *Sound Samples.* Retrieved December 31, 2014 from Schoeps Mikrofone: http://www.schoeps.de/en/downloads/sound_samples

SCHOEPS GmbH. (2013b). *Surround Plug-in for RTAS, VST Souble MS Tool.* Retrieved July 16, 2013 from Schoeps Mikrofone: http://schoeps.de/en/products/dms_plugin

Schroeder, M. R. (1980). Acoustics in human communications: Room acoustics, music and speech. *Journal of the Acoustical Society of America , 68* (1), 22-28.

Schroeder, M. R., Gottlob, D., & Siebrasse, K. F. (1974). Comparative study of European concert halls: correlation of subjective preference with geometric and acoustic parameters. *Journal Acoustical Society of America , 56* (4), 1195-1201.

Snow, W. B. (1953). Basic Principles of Stereophonic Sound. *Journal of the SMPTE , 61,* 567-589.

Solvang, A. (2008). Spectral Impairment of Two-Dimensional Higher Order Ambisonics. *Journal of the Audio Engineering Society* (56), 267-279.

Sonnenschein, D. (2001). *Sound Design: The Expresive Power of Music, Voice, and Sound Effects in Cinema* . Studio City: Michael Wiese Productions.

STEINBERG MEDIA TECHNOLOGIES GmbH. (2014). *Products - Nuendo.* Retrieved 12 31, 2014 from Steinberg Creativity First: https://www.steinberg.net/en/products/nuendo_range/nuendo/start.html

Steinberg, J., & Snow, W. (1934). Auditory perspectives - physical factors. In *Stereophonic Techniques* (pp. 3-7). Audio Engineering Society.

Sterne, J. (2003). *The Audible Past.* Durham & London: Duke University Press.

Sting. (2001, June 11). Perfect Love...Gone Wrong. *Brand New Day* . A&M - audio CD.

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

Streicher, R., & Everest, F. A. (2006). *The New Stereo Soundbook* (3rd ed.). Pasadena, CA: Audio Engineering Associates.

Svensson, E. (2001). Guidelines to Statiscal Evaluation of Data from Rating Sacles and Questionnaires. *Journal of Rehabilitation Medicine* , 47-48.

Swedien, B. (2009). *Make Mine Music.* Milwaukee: Hal Leonard Books.

Tchaikovsky, P. I. (1999). Tchaikovsky Ballet Suites - performed by Wiener Philarmoniker, conducted by Karajan. *DECCA Records, Legends 1965 - Legendary Performances* . London: The Decca Record Company Limited - audio CD.

Theile, G. (2000). Multichannel natural music recording based on psychoacoustic principles. *108th AES Convention* (p. 5156). Paris: Audio Engineering Society.

Theile, G. (2001). Natural 5.1 Music Recording Based on Psychoacoustic Principles. *19th International Conference: Surround Sound - Techniques, Technology, and Perception* (p. 1904). Bavaria: Audio Engineering Society.

Theile, G. (1991). On the Naturalness of Two-Channel Stereo Sound. *Journal of the Audio Engineering Society*, *39* (10), 761-767.

Thorton, S. (2009, February). *Michael Gerzon Audio Pioneer 1945-1996.* Retrieved August 19, 2013 from Michaelgerzonphotos.org.uk: http://www.michaelgerzonphotos.org.uk/index.html

Tobias, J. V. (1972). *Foundations of Modern Auditory Theory.* Academic Press.

Tohyama, M., & Suzuki, A. (1989). Interaural cross-correlation coeficients in stereo-reproduced sound fields. *Journal of the Acoustical Society of America*, *85* (2), 780-786.

Tohyama, M., Suzuki, H., & Ando, Y. (1995). *The Nature and Technology of Acoustic Space.* London: Academic Press Limited.

Toole, F. E. (2008). *Sound Reproduction: The Acoustics and Psychoacoustics of Loudspeakers and Rooms.* Burlington: Focal Press.

TSL Professional Products Ltd. (2013). *The Surround Zone Software.* Retrieved October 07, 2013 from SoundField: http://www.soundfield.com/products/s_zone.php

TSL Professional Products Ltd. (2013). *UPM-1 Stereo To 5.1 Upmixer.* Retrieved July 16, 2013 from SoundField: http://soundfield.com/products/upm1.php

Vries, D., Hulsebos, E. M., & Baan, J. (2001). Spatial fluctuations in measures for spaciousness. *Journal of the Acoustical Society of America*, *110* (2), 947-954.

Waves Audio Ltd. (2013). *UM225 / UM226.* Retrieved July 16, 2013 from Waves: http://www.waves.com/plugins/um225-um226

Wikipedia Foundation, Inc. (2014, January 30). *Wikipedia, The Free Encyclopedia.* Retrieved March 6, 2014 from Virtual Studio Technology: http://en.wikipedia.org/wiki/Virtual_Studio_Technology

Williams, M. (2013). *Microphone Arrays for Stereo and Multichannel Sound Recording - Volume II* (Vol. II). Milano: Editrice Il Rostro.

Wittek, H. (2011, March 23). *Image Assistant.* Retrieved August 13, 2013 from Hauptmikrofon.de: http://www.hauptmikrofon.de/index.php?option=com_content&view=article&id=1:image-assistant&catid=29:stereo&Itemid=40

Wittek, H. (2009). M/S Technique for stereo and surround. *14th AES Regional Convention*. Tokyo: Audio Engineering Society.

Yamamoto, T. (1973). Quadraphonic One-Point Pickup Microphone. *Journal of the Audio Engineering Society*, *21* (4), 256-261.