INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

5-Channel Microphone Array with Binaural-Head for Multichannel Reproduction

John Klepko

Faculty of Music
McGill University, Montreal
September, 1999

A thesis submitted to the
Faculty of Graduate Studies and Research
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

© John Klepko, 1999

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-55349-3

Canada

ABSTRACT

With the recent emergence of new release formats capable of delivering discrete multichannel surround-sound, there is a need to research unique recording methods to take advantage of the enhanced spatiality compared to conventional 2-channel stereophonic systems. This dissertation proposes a new microphone technique that incorporates head-related spatial cues through use of binaural artificial-head microphone signals sent to the surround channels. Combining this with 3 spaced directional microphones for the front channels shows promising results towards reproducing a 3-dimensional sound field. This dissertation describes a complete investigation of the proposed recording technique including an analysis of the basic concept, performance and suggested applications.

RESUMÉ

À la suite du développement récent de nouveaux formats qui ont la capacité de transmettre des multi-canaux indépendants ambiophoniques, le besoin de faire des recherches est survenu pour trouver de nouvelles méthodes uniques d'enregistrement qui peuvent profiter de la supériorité spatiale sur la stéréophonie à deux canaux. Cette thèse propose une nouvelle technique d'enregistrement qui inclut des directives spatiales qui proviennent de signaux binauraux d'un microphone tête-artificielle pour les canaux ambiophoniques. On peut voir des résultats prometteurs pour la reproduction d'un champ sonore tri-dimensionnel en joignant ces signaux binauraux à trois microphones directionnels espacés pour les canaux de l'avant. Cette thèse décrit une étude en détail de la technique d'enregistrement proposée et inclut une analyse de l'idée de base et du rendement ainsi que des suggestions d'utilisation.

Acknowledgements

I must express my heartfelt gratitude to the following:

Dr. Wieslaw Woszczyk, my supervisor and teacher throughout my graduate studies at McGill; his keen insight always pinpointed the areas of my work that needed strengthening; his high demands always kept me striving for better; in short, I could not imagine a greater mentor.

Prof. George Massenburg; a truly outstanding individual and teacher whose encouragement and example have, and will always be a source of profound inspiration.

Peter Cook for his kind and trustworthy friendship, and for his patience in hearing my ideas through all of this.

My parents, whose great love and support allowed me to pursue my higher goals.

My dearest Lisa, whose unwavering love and strength guided me to overcome the many obstacles and doubts through this work.

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

There has long been a quest towards the recreation of a soundfield in its fullest spatial dimension through the use of audio technology. Since the late 1950's, 2-channel stereophonic systems have been the predominant format for music reproduction through tapes, long-play records, compact discs and radio broadcast. However, the spatial rendering ability of conventional stereo systems is limited to the plane tended between the 2 loudspeakers. Recent advancements in audio technology have been towards the development of multichannel audio display systems which use multiple loudspeakers surrounding the listener.[1] Such surround-sound systems offer a much greater spatial reproduction ability. Of particular interest to the music production community is the emergence of new release formats capable of delivering 5 (or more) discrete channels.

This development opens up a whole new frontier for music creation and representation. With this, there is a definite need to research unique recording methods to take advantage of the enhanced spatial ability. One obvious need is the development of new encoding methods through microphone techniques. However, there should be a certain hierarchy followed before leaping directly towards developing new microphone techniques.

[1] Although the film industry has been using multichannel surround-sound steadily since the late 1970's, applications for music purposes have not been exploited.

Firstly, a comprehensive understanding of our natural spatial hearing process should be garnered. Research (particularly in the area of cognitive psychology) is constantly revealing new understandings about spatial hearing which can be applied towards the development of surround-sound systems.

Secondly, we must attempt to understand how the surround-sound loudspeaker configuration is interpreted by our spatial hearing sense. The loudspeaker arrangement surrounding the listener constructs a new virtual environment where different sonic elements can be placed within. We must seek to understand its inherent limitations - what it can and cannot do (Klepko 1999).

Finally, we can proceed to the last stage of encoding where microphone techniques can be developed that cater to, and support the surround-sound listening environment. Inherent within this research stage is a need to forge new methods of testing microphone techniques in terms of spatial as well as timbral qualities. These test specifications should be able to be applied towards any microphone system designed for surround-sound.

This dissertation proposes a microphone system as one possible solution towards the problem of spatial rendering of a sonic event within a surround-sound environment. The research follows the hierarchy as outlined above ultimately leading towards testing, evaluating and applying a new microphone technique.

Chapter 1 of the dissertation begins with a discussion of the concept of space and spatial perception as an egocentric process. This is followed by a review of the different audio systems (from monophonic to 3-D sound) focussing on their ability to represent spatial properties of a sonic event. It ends with a comprehensive analysis of the requirements of auditory images.

Chapter 2 presents the rationale of the dissertation beginning with a basic description of the proposed system followed by a breakdown analysis of the strengths and weaknesses of its component parts.

Chapter 3 goes on to describe the theory, design and construction of the system. Included within this chapter is a review of some precedent similar research ideas.

Chapter 4 describes the evaluation process of the microphone system with discussion of the test results. Inherent to this chapter is a methodology for evaluation of surround-sound microphone techniques. It should be noted at this point that all the experiments described within this chapter were run with a small sample of test subjects, typically 4 or 5. This was partly due to a practical limitation of the research lab being in a constant state of change and upgrade making it difficult to schedule listening tests within such periods. As well, it was felt that this number of subjects was in agreement with (and influenced by) recognized and authoritative published psychoacoustic research practice (Begault 1992) (Begault 1996) (Begault & Erbe 1994) (Grantham 1986) (Perrot 1974) (Perrot 1984)

(Strybel, Maniglas & Perrot 1989) (Strybel, Maniglas & Perrot 1992) (Trahiotis & Bernstein 1986).

Chapter 5 contains proposals for different applications mostly unique to this microphone system. It discusses the requirements of these applications and, the results of experimental recordings made for this purpose. There is a digital tape which demonstrates these applications. It is not essential to this dissertation and is included only as an option. (Appendix "A" outlines the contents of this demonstration tape).

Chapter 6 summarizes the work with conclusions and suggestions for further related investigations.

CHAPTER 1 - CONCEPT OF SPACE

1.1 Physical vs. Cognitive

Before we can effectively design and implement an audio display system, we must gain an understanding of the mechanisms behind the occurrence of a sound and its subsequent reception by the human auditory system. Blauert (1997, p.2) simplifies this concept by distinguishing between the occurrence of a vibration or wave phenomena (of a frequency between 16 Hz and 20 kHz); "sound event", (From this point on, I suggest an alternate term, "sonic event" to be more precise and consistent than "sound event") and the perceptive result, "auditory event".

This distinction between sonic and auditory event is related to the fact that spatial cognition is essentially an egocentric process. Judgements concerning the spatial characteristics of an object (and its environment) are always in reference to the listener/observer. Marshall McLuhan stated it succinctly;

> "Objects are unobservable. Only relationships among objects are observable". (McLuhan, 1964)

For instance, we may describe the spatial characteristic of an object (or sonic event) as; "behind, to the right", or, "large, wide" etc.. These articulations are all relative to other objects whether they be the observer him(her)self, the observer's *memory* of other objects not present, or surrounding surfaces (walls, ceilings, floors).

Central to this concept is the realization that the auditory event may not correspond correctly to the sonic event. It is a case of cognition (subjective) vs. physical reality (objective). It should be understood that differences (i.e. errors) *do* occur in natural hearing situations - these errors are not a problem unique to interactions involving audio display systems.

Our spatial ability is highly dependent upon both our visual and auditory sense. This bi-modal interaction reinforces our overall perception of space helping to reduce errors of judgement. However, the potent visual cue can influence the position of the auditory event even if the sonic event position does not coincide. This can be a form of trickery known as the "ventriloquism effect" which was observed while creating an artificial discrepancy between visual and auditory localization cues (Pick, Warren & Hay 1969). A common, everyday example of this is found in movie or television presentations where the speaking voice will appear to originate from the mouth of the actor despite the fact that the loudspeaker actually emitting the voice is located elsewhere.

Research supports another peculiar auditory/visual interaction effect referred to as "visual facilitation" (Warren, 1970). It was found that localization variances were greater when the subject was tested in a darkened over a lighted room. In either case, the sound source was not visible. It was postulated that, in a lighted room, the subject could more easily establish a frame of reference with which they could place a sonic event.[1] In other words, the sense of vision served to organize auditory space.

In general, the complete process of spatial perception involves the auditory reception of a sonic event and the assignment of some form of meaning to that event. A flowchart of this process could be as follows;

1. physical - sonic event
2. representation (optional)-audio/visual
3. auditory stimulation - auditory event
4. articulation - meaning, context

1. The sonic event can result from a static object, or one in motion. Most musical instruments are played from one position, however, the subsequent room sound in the form of reflected sound wave energy, travels around and past the listener. This can be perceived as some form of motion as the sound modulates and grows throughout the room environment. Also, the different room modal resonances vary according to position and frequency (pitch). As a result, different pitches will excite resonances at different parts of the room which

[1] An interesting note is that this visual facilitation effect only occurred with adults. Children (under 13 years of age) did not show this effect.

can cause the direction of the auditory event to be anything but static.

2. This stage of representation refers to the reproduction of a recorded sonic event (or transmission of a live sonic event) over loudspeakers or headphones, with or without a corresponding visual display. Discussion of the various forms that this stage can take is central to the next section (1.2). It is labeled as optional here since it is an intermediary step that is not essential to the perceptual process between the sonic and auditory event.

3. This stage refers to the stimulation of the auditory nerve. For optimal performance, it is assumed that both ears are healthily functioning and that neither is occluded in any way.

4. This stage refers to the process of understanding the data collected through our auditory sense. We usually attach some sense of form, meaning and context to the stimuli. In the textbook, *Spatial Cognition*, Olson and Bialystok state:

> "...meaning is usually taken as a higher order interpretation of experience, while perception is a direct consequence of that experience, hence, meaning appears to have a secondary or derived status". (Olson, Bialystok, 1983)

It can be argued that listening to music should be a more phenomenological or gestalt experience. For example, we perhaps should not be so concerned with identifying objects (and their location) within a musical performance. Such an analytical approach could distract from the deeper emotional intent of the music or art form. Phenomenologicalist Don Ihde, interestingly describes an idealistic listening experience as one that unifies the sound sources with the listener:

> "If I hear Beethoven's Ninth Symphony in an acoustically excellent auditorium, I suddenly find myself immersed in sound which surrounds me. The music is even so penetrating that my whole body reverberates, and I find myself absorbed to such a degree that the usual distinction between the sense of inner and outer is virtually obliterated". (Ihde 1976)

This description is from the end-user or listener perspective. The sound engineer/producer must be aware of this perspective, but also take a highly analytical stance when recording/mixing a musical performance. (The power to manipulate spatial images is far greater now that surround-sound technology is becoming more prevalent).

However, in everyday listening experience, we do extract meaning from the sonic events. This "meaning" of the objects could exist in the form of;

1. locatedness of object
2. character (size, shape) of object
3. character (of object's) environment
4. identification (or failure to identify)

1. Determining the location of a sonic event is essentially an egocentric process. We develop a lexicon to help organize the spatial characteristics of the event. This lexicon is of a relative, contrastive binary nature separated into 4 general dimensions;

> a) The vertical axis which is defined by gravity and described by the terms; above/below, up/down, over/under, high/low etc.
>
> b) The medial axis extending through the center of our bodies and described by the terms; front/back, ahead/behind etc.
>
> c) The horizontal axis which is perpendicular to the medial axis and mainly described by the term; left/right.
>
> d) The distance dimension described by; near/far.

By fulfilling all 4 dimension categories, we can completely describe the location of a static auditory event.

2. The character of the object can also be defined by a binary lexicon. Such terms include; large/small, tall/short, wide/narrow, etc..

Judgements about the size of an object could be linked to incidental learning (Perrot & Buell 1982). For example, an auditory image that has much low frequency energy and a relatively high loudness level is often associated with a larger object.

3. The character of the environment where the sonic event/object exists can be ascertained by listening to such acoustical measures as reverberation decay time and, early reflection arrival time. More accurate judgement requires experience in several different room environments.

4. The final stage of *identification* is derived from the previous three. Identification of the object (or environment) is based on the listener's experience, expectations and unique interpretations in conjunction with the quality of the (audio) display. Any of these factors can affect the success of identifying the object, but this thesis is mainly concerned with the accuracy of the audio display system. Distortions in the encode/decode signal path can cause discrepancies in the spatial as well as spectral and temporal domains.

One final related point questions the accountability of the listener in determining the success of an audio display system (Shlein & Soulodre 1990). Are there listeners who are more skilled then others in resolving the signals presented from an audio display system? Theodore Tanner (of Spatializer Audio Labs) asks: "...what do we do about persons who have a pre-conceived notion that virtual imaging systems cannot operate accordingly?" Tanner goes on to suggest:

> "Our psychological predisposition can affect our perceptions. Experience plays a large part in aspects of stream segregation. I contend this is the case with virtual imaging systems and learning how to localize virtual images plays an important role in making perceptual judgements." (Tanner 1997)

Here again, it must be argued (as in pages 6-7) that the enjoyment of music through virtual reality systems should be an immediate experience not conditional on any previous experience with listening through an audio display system. Specialized training and learning with audio display systems is more relevant to critical perceptual tasks required in military applications, aircraft pilots etc..

1.2 AUDIO REPRESENTATIONS

Audio technology is a carrier of information from the sonic event to the auditory event. The following sections (1.2.1 - 1.2.6) discuss the different forms of audio representations.

1.2.1 Monophonic

A monophonic system is the most basic audio representation of a sonic event. It implies a single channel of transmission or storage. The simplest mono system is comprised of a single (encoding) microphone and a single (decoding) loudspeaker. However, audio systems are still considered to be monophonic if they mix several microphones into a single channel, or, distribute a single microphone signal to 2 (or more) loudspeakers.

Despite significant research in the 1930's on stereophonic and multichannel audio[1], monophonic was still used for several decades for records, films, and radio and television broadcasts.

A monophonic system may be sufficient to represent a sonic event (for example, the sound of a musical instrument) in terms of its frequency range (bandwidth), timbre (within that range) and dynamics (soft/loud). But it cannot convey a sense of space, or

[1] See sections 1.2.2 and 1.2.5 for a brief survey of this research on stereo and multichannel audio respectively.

represent the spatial soundfield inherent with any sonic event. The single microphone simply "compresses" all the spatial information (including direct sound, early reflections, and reverberation) into one channel.

There have been ideas presented to *synthesize* a spatial image from a monophonic channel distributed to 2 loudspeakers (Schroeder 1958) (Gardner 1969) (Bauer 1969). The common basic idea behind these schemes is to split a monophonic signal into 2 loudspeakers that are treated with a slight difference between them. These differences may be obtained through imposing time delays, phase shifts, complementary spectral contouring, and adding artificial reverberation.

1.2.2 Stereophonic

The term "stereophonic", is actually a combination of 2 Greek words. "Phonics" is the science of sound, and "stereo" means "solid" implying a 3-dimensional object (or image) possessing attributes of width, height, and depth. However, stereophonics (or stereo, for short) is most commonly thought of as being 2-channel.

With stereo came the ability to present a left-right soundstage and to distribute the direct sounds across it fairly independent of reflected and reverberant energy. The recommended loudspeaker/listener relationship (as shown in figure #1) is that of an equilateral triangle with both loudspeakers at ±30°.

The first demonstration of stereophonic audio is believed to be part of the Paris Exhibition of Electricity held in 1881. (Hertz 1981) The engineer, Clement Ader, designed a system comprising of 10 microphones that were placed in a line across the front of a stage; 5 assigned to the left channel, and 5 to the right. This was not a recording, but a live transmission of singers and orchestra performing on the stage of the Grand Opera of Paris. The 2-channel signal was transmitted over telephone lines (3 km. away) to earphones worn by visitors at the exhibit[2] who could perceive a left-right virtual soundstage.

Significant research in stereophonic audio did not occur again until the early 1930's with concurrent experiments at Bell Labs in New Jersey and by Alan D. Blumlein at E.M.I. in England.

The engineers at Bell Labs explored "auditory perspectives" with an emphasis on spatial reproduction (Steinberg & Snow 1934) (Snow 1953). They focussed on developing techniques that comprised of a line array of spaced omnidirectional microphones for 2 or 3-channel stereo. This system relied mainly on time-of-arrival differences between the left or right-assigned microphones. These techniques are still in wide use today.

[2] Stereophonic *reproduction* usually implies using loudspeakers. In this case, the listeners used earphones where the proper term would be "bi-phonic".

Alan Blumlein's landmark patent in 1931 (Blumlein 1933) was the first of many dealing with improvements in audio through stereophonic 2-channel techniques. His main developments involved crossed coincident pairs of microphones that relied on intensity differences to present a spatial soundstage. This was quite a different approach than the time-difference based techniques developed at Bell Labs. Blumlein's work resulted in many variations of coincident microphone techniques still used today such as; M-S (mid-side stereo), Blumlein (crossed figure-eight microphones), and X-Y (crossed cardioid microphones).

Another important invention of Blumlein's was a method to record 2-channel signals into a spiral groove that would become the basis for stereophonic vinyl phonograph records.

However, the first commercial release of a stereophonic phonograph record would not be until 1957.[3] Furthermore, it was common practice to continue to release monophonic versions alongside a stereo one until about 1968 (Eargle 1980). This was because of the monophonic limitations of AM radio, jukeboxes and portable transistor radios that were prevalent at the time. Recording engineers were (and are still to some extent today) concerned with the monophonic compatibility of stereophonic recordings. This is because mono is actually a summation of the left and right stereo

[3] Recordings on tape format (7.5 ips speed) did precede vinyl phonograph records as a consumer stereophonic release in the early 1950's but did not receive the widespread success that LP records did.

signals. Any considerable phase differences will result in comb-filtering[4] of the mono version.

[4] The term "comb filter" graphically refers to the alternating pattern of cancellations and reinforcements across the frequency range of a signal. In simple terms; combining in-phase signals causes a level doubling (reinforcement) and, combining out-of-phase signals causes a cancellation.

1.2.3 Binaural (with earphones)

Binaural technology is derived from the human spatial hearing ability where our anatomy modifies the impinging soundfield. We can localize the direction of a sound source due to difference of 2 input signals, one at each ear. The presence and shape of the torso, shoulders, head and outer ears (pinnae) reflect and diffract the sound waves in different ways dependent upon the direction of the sound source. These disturbances of the sound field cause shadowing effects and resonances, which alter the spectrum of a sonic event. Directional encoding results from; 1) pinna reflections, 2) concha[1] resonances 3) shoulder reflections 4) torso reflections 5) head diffraction. The physical separation of the ears gives us additional timing difference cues, which are important to localizing sounds. All of this combined is known as the "head-related transfer function" (HRTF).

Binaural technology is an approach to recording that seems rather obvious and primitive in its simplistic nature. It is a 2-channel system where the 2 recorded signals are meant to be reproduced independently to each ear - hence the term, *binaural.* Small microphones may be placed just inside each ear canal entrance of a live human subject. But, usually the more practical solution involves a life-size replica (or manikin) of a human head (and torso) with microphones inserted within ear simulators. In this way, the sound field is modified by the same physical features as an average human

[1] The *concha* is the main (and largest) cavity of the pinna.

head. Such a 2-channel microphone system with manikin has been most commonly referred to as *artificial head, binaural head, dummy-head, head-and-torso-simulator (HATS),* and (in German), *kunstkopf.*

The earliest documented experiments with artificial head microphones were conducted at Bell Laboratories in 1932 (Fletcher, 1933). The experiment involved a tailor's wooden manikin named "Oscar", which had microphones mounted flush within each cheekbone just ahead of the ears. A rather dramatic public demonstration was carried out as described below:

> "Striking evidence of the naturalness which can be secured with such a binaural system was obtained at several formal demonstrations in Philadelphia. When the guests had put to their ears the receivers connected to Oscar, who was in another room, someone would say confidently in Oscar's ear, "Please move over." A surprisingly large number of the guests would start to obey the command before realizing that it came from the receivers. Afterwards someone would whisper in first one and then the other of Oscar's ears, and would tear paper, jingle keys, or thump a tambourine to illustrate the fidelity of the system to sounds of high frequency. Finally Oscar would be brought out and set in the midst of the audience so that they could compare direct with transmitted sounds."

Leopold Stokowski[2], then conductor of The Philadelphia Symphony Orchestra, participated in the musical demonstration recordings that were also made with Oscar.

[2] Stokowski often offered his services towards the advancement of audio technology.

Some artificial head microphones[3] were later developed for purposes other than music recording such as: medical (audiology) applications (Nordlund & Líden 1963); space aeronautical acoustics (Bauer *et al.* 1967) and (Torick et al. 1968).

Burkhard and Sachs (1975), of Knowles Electronics Inc., designed a head-and-torso manikin for hearing aid and acoustic research. This binaural microphone known as KEMAR (Knowles Electronic Manikin for Acoustic Research), has become a standard tool in audiological research[4]. Physical dimensions of KEMAR were average-modeled according to data obtained of over 4000 male American air force pilots. For flexibility, there are 4 interchangeable sets of pinnae, and a hair wig can be an option.

Brüel & Kjaer are a company in Denmark that specialize in the design and manufacture of audio and acoustical measurement equipment. They have designed 2 different HATS models intended for applications such as testing telephones, headsets, hearing aids, and headphones as well as an instrument for evaluating room acoustics, noise levels in automobiles and speech intelligibility. The model 4128[5] features ear simulators which include a model of the ear canal. The model 5930 can be more suited towards music recording applications as there is no simulation of the ear canal - the

[3] These are distinguished here because they were only developed for research interests and are not available commercially.

[4] There appears to be no reports of using KEMAR for music recording.

[5] This model also includes a voice simulator with a small transducer in the mouth position.

microphone capsules (model 4009) are almost flush with the ear canal entrance.

In 1973, the Sennheiser Company introduced a binaural recording system called the MKE-2002. This model features 2 small condenser microphones mounted on the ends of a stethoscope-like headset allowing the user to mount the microphones just outside their own ears for a more individual head-related recording. The system also comes with a soft rubber head bust so that the user can place the microphone headset on the dummy-head as a more practical option.

Neumann GmbH, a German company that specializes in making high-quality studio microphones introduced the KU-80 artificial head in 1973. This model consisted solely of a manikin head with no shoulders or torso. This was later replaced by the KU-81i, which improved on the latter by placing the microphone capsules at the ear canal entrance, rather than inside. As well, it featured diffuse-field[6] rather than the free-field equalization of the former model. The most recent model KU-100, also has diffuse-field equalization and improved symmetry between the ears.

In the late 1980's, Head Acoustics GmbH of Aachen, Germany introduced an artificial head known as the "Aachen Head". This was the result of research by Klaus Genuit (Genuit & Bray 1989) (Gierlich & Genuit 1989). There are actually 2 different models; the HMS II for

[6] This diffuse-field type equalization (as provided by electronic EQ in the supplied pre-amplifier) is meant for to be compatible for loudspeaker reproduction. (This is also discussed further in sections 2.1.2 and 3.1.1)

acoustic measurement applications, and the HRS II for binaural recording. Both models feature a torso simulation and a simplified head and pinna structure that is based on average dimensions. These models do not replicate the fine physical features of a typical pinna and face which were deemed unimportant to the acoustical properties of the system.

1.2.4 Binaural (with loudspeakers)

The ideal situation with binaural encoding (whether via dummy-head microphone or DSP-simulated HRTF) is that each ear hears only its corresponding signal; i.e. the listener's left (or right) ear is exposed to the left (right) signal. This degree of separation can only be accomplished using earphones. However, limiting binaural reproduction to earphones also limits its usefulness as a recording method.

Attempting to reproduce binaural recordings over loudspeakers would cause an undesired crosstalk component to reach each ear. Any side (left or right) signal would reach the intended ear as well as diffract around the listener's head and reach the opposite ear. This would result in serious spatial distortion where sounds that were encoded above, behind and to the sides would all be reproduced within the angle subtended by the loudspeakers. This would of course negate the unique virtue of spatial encoding that binaural methods possess.

The most common solution to this problem is a form of acoustical subtraction where the unwanted crosstalk signal is cancelled by another almost identical signal that is polarity-inverted. This type of crosstalk-cancellation scheme involves cross-feeding a signal that is equalized to match the crosstalk leakage signal at the opposite ear. This is because the acoustical crosstalk would originate from the opposite loudspeaker typically placed 30° off-axis, and would undergo spectral filtering as a result of diffraction around the head. This equalized signal should also undergo a slight delay to compensate for the extra distance it must travel so that the cancellation signal arrives at the same time as the original crosstalk signal. A simple polarity inversion (-180°) will cancel the signal out to a greater degree.

Atal and Schroeder (1962) were the first[1] to propose such a system having to rely on a large mainframe computer for the calculation and generation of the crosstalk canceling signals.

Variations of the Atal/Schroeder crosstalk-canceling scheme have appeared in the research literature under various names; TRADIS (True Reproduction of All Directional Information by Stereophony) system designed by Damaske (1971); Transaural recording (Cooper & Bauck 1989 and 1992); Sonic Holography (Carver 1982)[2].

Other related research includes; (Møller 1989), (Gierlich and Genuit 1989) and the MINT[3] system, which actually uses a 3rd centre loudspeaker to assist in the crosstalk canceling action.

[1] Actually, Benjamin Bauer (1961) presented an idea (one year earlier) to "...convert a binaural program for reproduction over a stereo loudspeaker system". His proposal outlined a simple polarity reversal of crossfed signals but showed no concern or mention of filtering and delaying these crosstalk-canceling signals.

[2] Carver Corp. actually marketed this "Sonic Holography" crosstalk-canceling circuit within their own home stereo preamplifiers. The process was the same as other crosstalk-canceling schemes but was mainly intended to enhance conventional stereophonic listening (by extending the lateral soundstage) rather than be applied to binaural recordings. A similar system was marketed by Polk Audio with the model SDA-1 and SDA-2 loudspeakers.

[3] The MINT system was developed by Miyoshi & Kaneda in 1988 and found reported in (Tohyama, Suzuki, & Ando 1995).

These schemes work best provided that the listening room has a fairly dead acoustic, otherwise, side-wall reflections could increase the complexity of the crosstalk signal making it difficult to cancel. Also, the cancellation process works best for a tightly confined listening area. Any lateral movement of the listener will produce a type of remainder signal caused by the imperfect acoustical subtraction.

Another problem is that the artificial crosstalk-canceling signal will diffract around the head becoming unwanted leakage itself. Some schemes would calculate and produce an additional pair of crosstalk-canceling signals to deal with this.

1.2.5 Multichannel

Stereophonic 2-channel has prevailed as the storage and delivery medium mainly due to limitations of technology. Vinyl discs could only store left/right modulated grooves, FM radio could only transmit sum and difference channels, and digital compact disc had a fixed data transfer rate compromising the number of channels for the sake of longer playing time as well as reasonable sound fidelity limited by quantization and sampling rates.

But even at the beginning of "stereo" audio research, it was realized that 3 channels (or loudspeakers) across the front could be much superior - hence the birth of the multichannel idea.

1.2.5a Three-channel

In their pioneering research on audio spatial reproduction, Bell Labs ran experiments to determine the optimum number of channels[1] (Steinberg & Snow, 1934) (Snow 1953). They used a "screen analogy" to describe a theoretical ideal situation with an extremely large number of microphones placed in a straight line each connected to a correspondingly placed loudspeaker. This "infinite screen" approach was only theorized as its impracticality was realized. Instead, more practical arrangements that used 2 or 3 channels were tried and compared. The 3-channel system involved 3 separate

[1] This was as an alternative approach to the binaural headphone reproduction experiments using "Oscar", the dummy-head microphone.

microphones or, 2 microphones (feeding left and right) with the 3rd centre-channel being derived from a bridged summation (L+R) of the left/right microphone signals.

Among their principal conclusions:

> "The 3-channel system proved definitely superior to the 2-channel by eliminating the recession of the center stage positions and in reducing the differences in localization for various observing positions. For musical reproduction, the center channel can be used for independent control of soloist renditions. Although the bridged systems did not duplicate the performance of the physical third channel, it is believed that with suitably developed technique their use will improve 2-channel reproduction in many cases". (Steinberg & Snow 1934)

It should be pointed out that these experiments were aimed at improving audio spatial reproduction in large room auditoriums and stages as opposed to smaller, more intimate home listening environments. Nonetheless, this work provided a foundation for realizing the psychoacoustic benefits of using a 3rd center channel.

The main benefit of the 3rd channel is for film presentations where the all-important dialogue can be "anchored" in the center no matter how much any viewer is seated off to one side. Even with home or studio listening of music, any sounds designated to the center loudspeaker will not suffer a type of spatial intermodulation distortion as the listener moves left or right.

Two other advantages of a center loudspeaker are related to the interaural crosstalk that occurs with a center phantom image produced by two (left and right) loudspeakers. (Holman 1993 and 1996) points out that the acoustical crosstalk signal is delayed to the opposite ear. This delay is a function of the effective distance between the ears for a sound arriving from ±30 degrees and causes a cancellation at 2 kHz. The result is a timbre mismatch between the left/right loudspeakers, and the center. This timbral inequality would be most noticeable during simulated motion (pans) across the front, or when there are 3 similar sounds meant to be perceived at the left-side, center and right-side positions. This comb-filtered effect would not be present from a single sound source center loudspeaker.

The other advantage, although somewhat abstract in explanation, is that a center loudspeaker sounds more "natural" since each ear only receives one signal. A left/right derived phantom center uses up more "brain power" to make sense of the auditory event that appears to come from front-center despite the fact that the *sonic* event actually occurs from two places at 30° to the left and right of the listener.

An additional advantage is the avoidance of the problem of vocal/speech sibilance "splattering" towards the sides which often occurs with stereo central phantom images. This is due to the inevitable (but varying degrees of) lack of matching between a left/right pair of high-frequency drivers (tweeters).[2] Obviously, a single loudspeaker cannot suffer from this problem (Holman 1991).

Since the 1950's, various schemes were developed making use of a third center loudspeaker. Paul Klipsch (1959) proposed a simple system that derived 3 channels from 3 microphones that were stored on a standard 2-channel tape.[3] The center loudspeaker feed was the bridged output of the left and right channels recorded on tape where the 3rd centrally-placed microphone signal was recorded equally on both tracks.

Both Duane Cooper (1970) and Richard Cabot (1979) proposed a system called "triphonic". Their proposed loudspeaker placement was essentially the same with one center loudspeaker in front, and two flanking loudspeakers to either side, altogether forming an equilateral triangle around the listener. The center loudspeaker derives its input from a summation of left and right channels, and the 2 side loudspeakers would be connected in anti-phase. Cooper recommended mixing a coincident pair of forward-facing cardioid microphones with a spaced pair of bidirectionals with their 90° nulls pointed towards the front center stage area. The bidirectional ambient microphones would be spaced far enough apart to have a

[2] Actually, any component of the (encode/decode) signal chain can contribute to this effect as long as there is some difference in the high frequency phase or frequency response. Therefore, microphones, mixing consoles, and amplifiers could introduce slight differences that would widen the center phantom image.

[3] Early on in the stereophonic era, Paul Klipsch (1960a) also proposed that; "The optimum stereo system in terms of fidelity of tonality and geometry must consist of 3 channels with corner flanking units".

low degree of correlation between themselves and, the front pair. Each pair would be assigned and blended to left and right as recorded on a 2-channel tape. Upon playback, the summed center channel would have a mix of both pairs, and the side-rear loudspeakers would cancel out the forward signals and be left with opposite-polarity, but uncorrelated ambient microphone signals.

Cabot's triphonic encoding involved a coincident setup of 3 microphones. One omnidirectional microphone summed to both channels to feed the front-center loudspeaker and, two perpendicular-crossed bidirectionals fed to both channels in anti-phase which would be routed to the side loudspeakers via an (L-R) wiring.

1.2.5b Surround

The 1940 Walt Disney production of the film "Fantasia" is considered the first attempt at surround sound presentation. Building on the research from Bell Labs, the Disney sound department (in conjunction with the RCA sound engineering staff and Leopold Stokowski) devised a special system called "Fantasound" (Klapholz 1991) (Holman 1998) which recorded 4 separate audio tracks directly onto the film; one screen-left, one screen-center, one screen-right and one "control" track. The system required a technician to mix the show live from within the theatre where at times, the left and right channels would be routed to one of almost 100 rear loudspeakers surrounding the audience.[4] This system was basically a "road-show" that lasted only one year and its development wasn't continued mainly due to the onset of World War II. However, its basic surround ideas would influence film sound in much later years.

Hafler (1970) devised an ingenious yet simple variation on reproducing existing 2-channel recordings to allow for surround sound in a home listening environment.[5] All that it took was an extra loudspeaker (or optionally, two) placed in the rear of the listening room. This "surround" loudspeaker would be driven by the difference signal (L-R) from the stereo power amplifier. It is expected that most stereo recordings' difference signal would contain uncorrelated information in the form of reflected energy and reverberation which are ideal soundfield components for enveloping the listener via the surround loudspeakers. The complete system also features a 4th center loudspeaker signal derived from the summation of left and right.

Peter Scheiber (1971) went a step further with this idea by devising a matrix approach to encoding 4 channels onto a 2-channel storage medium, then retrieve 4 separate channels again for playback. This form of matrix array is commonly referred to as "4-2-4".

[4] The engineers were asked to develop a method to simulate motion of a sound between the front and back of the theatre. They succeeded by designing the first panoramic potentiometer (nicknamed at the time, "panpot") which was a 2-ganged volume constant-power volume control. The separate volume controls were "ganged" by what looks like (in one photo) crude bicycle chains driven by large sprocket wheels. This panpot idea (minus the chains) is found on every mixing console today.

[5] Hafler was assigned to the Dynaco Corporation at the time and this method of wiring became alternately known as the "Hafler Hookup" or "Dynaco Hookup".

Scheiber's visionary patent called the process, "Quadrasonics" and led to much activity in the audio community towards a new type of spatial audio reproduction more commonly known as "Quadraphonics". This is significant in that it was the first major attempt (which was even embraced by record companies and radio broadcasters) at surround-sound for music reproduction in the home. Reproduction was via 4 loudspeakers placed (roughly) in a square subtending 4 corners around the listener.

Ultimately, by the end of the 1970's, quadraphonics had failed to be accepted in the marketplace. Its demise has been blamed on several different factors most notably, consumer dissatisfaction. The audio equipment manufacturers involved could never settle on a single standard format, which ultimately confused the public over the proliferation of systems that were incompatible.[6] Add to that the economic impracticalities for record stores having to stock/display dual-inventories and, the additional production costs to the record companies for quadraphonic releases (over the stereophonic versions).

Despite the failure of quadraphonics, important audio technology innovations emerged due to the great competitive research effort. These were mainly the development of multiple playback channels

[6] There were at least 3 different matrix (4-2-4) systems; "SQ" from CBS, "QS" from Sansui, and "EV-4" from Electro-Voice. One discrete quad format was the "CD-4" from a joint effort by JVC corp. and RCA Records.

and the idea/implementation of matrixing to store and extract 4 channels out of 2.

This matrix technology spawned a whole new improvement and standard for film sound. In 1976, Dolby Laboratories introduced a (4-2-4) matrix surround system called "Dolby Stereo" for movie theatres. This consisted of 3 channels across the front (Left-Center-Right) and a mono channel for surround. The surround channel would be distributed over many loudspeakers to the sides and rear walls of a cinema for the purpose of even coverage of the audience area. In 1982, Dolby introduced "Dolby Surround" for the home theatre environment that was emerging due to the popularity of the videotape formats (beta and VHS). This system was a 3-2-3 matrix having only Left, Right and mono Surround channels.

The passive matrix decoders used in these systems suffered from channel crosstalk, which would often result in sounds intended for the front appearing in the rear and *vice versa*. In 1987, Dolby improved this situation by introducing the "Dolby Pro-Logic" active-matrix surround system (and Dolby SR for movie theaters) using "intelligent" processing that steered sounds (that were dominant in level) to their intended position. With "Pro-logic" also came a center-channel, which the Dolby Surround format for home did not have.

It should be noted at this point, that these matrix systems were developed as a compromise to make the best out of analog recording and storage, which had a practical limitation of 2 channels. Once digital recording technology began to progress, it was found that data reduction (compression) schemes could be applied to audio signal encoding with little perceived loss in sound quality while extending its storage and delivery capacity.

Research involving these digital data reduction schemes lead to applications for *discrete* multichannel surround sound capabilities. In 1991, Dolby Labs introduced the AC-3 data reduction scheme which was applied to (among other uses) "Dolby Stereo Digital" format for the movie theatres. The following year, "Dolby Surround Digital" for the home environment was introduced. Both systems featured encoding/decoding of 5 independent, full-bandwidth channels plus a sixth channel called "LFE" (Low Frequency Effect) meant for subwoofer reproduction. The popular term for this discrete playback format is "5.1" in reference to the 5 independent channels with the ".1" designating the limited (10%) bandwidth of 20 to 120 Hz.

With discrete surround, the problems of inter-channel crosstalk, which caused gross spatial errors, were overcome allowing proper control over sound placement. Also important was that the surround loudspeakers were no longer fed from a single, monophonic signal, but 2 independent channels. Now, the rear soundfield could be expanded spatially. (This would have greater artistic implications for music applications more than film presentations).

The Dolby AC-3 (Dolby-Digital) encoded surround scheme also became accepted as an option for the audio portion of the DVD-Video format as well as HDTV broadcast. At the time of this writing (April 1999), Dolby Labs introduced a new coding system designed to simplify the distribution and post-production of digital multichannel programs ultimately intended to be AC-3 encoded. This system called Dolby E, can distribute up to 8 channels of audio plus "metadata"[7] all transmitted via AES3 (2-channel digital) or recorded onto 2 audio tracks on a digital VTR.

Other (incompatible) data reduction schemes were developed which were part of competing film presentation systems. In 1993, DTS (Digital Theatre Systems) launched their 6-channel format (L, C, R, with stereo surrounds + LFE channel).[8] In 1994, Sony introduced their SDDS (Sony Dynamic Digital Sound) which is a "7.1" system featuring 5 front channels plus stereo surrounds and one LFE channel.

Also at the time of this writing, Dolby introduced a new 6.1 system called Dolby EX, that features an added rear center-channel. It is not a true discrete 7 channels but a a 7-6-7 scheme where the additional channel is matrixed from the 2 surround channels. This format is intended for movie theater presentations and was developed by Dolby Labs in collaboration with Gary Rydstrom from Skywalker

[7] Metadata is "data about the data" used as control information for the decoding stage. (www.dolby.com)

[8] DTS also introduced a discrete multichannel format capable of 6 channels being encoded (via a lossy perceptual data reduction scheme) onto a standard audio compact disc. A special decoding chip is required inside a preamplifier.

Sound (of Lucasfilm). The first production to use this format is, "Star Wars Episode One: The Phantom Menace".

1.2.5c Other considerations

Currently, there is some debate about what type of loudspeakers to use for the surround channels. The arguments for either side seem to be driven by the intended creative or artistic function of the surround.

Those who consider the surround's function as reproducing a vague ambient soundfield prescribe the use of *dipole* radiator loudspeakers. (Holman 1991a) Dipoles have a figure-8 polar pattern and are recommended to be positioned with their null towards the listening area (or audience). In this way, no direct sound from the loudspeakers reach the listener, only reflected energy from the room surfaces which would be more diffuse avoiding localization of any sounds in their direction. This approach can be more appropriate for film presentations (theatre or home) where distracting the viewer's attention away from the front screen (by clear sounds from the rear) would be best avoided.

For music applications, dipole radiators would seem to be advantageous if the surround channels only represented ambient and reverberant sound energy. However, many music producers feel that having 5 matched loudspeakers would allow more flexibility and power to place direct sounds anywhere in the soundfield. As well, this would allow panning motional effects to be more even and continuous. With matched direct radiator loudspeakers, there would also be the possibility for equality of sound timbres around the listener.

Despite the aforementioned auditory image benefits of using a center loudspeaker in place of a stereo phantom image, there can be some reservations about its usefulness. This is mainly due to the fact that many so-called home theatre surround sound systems have a different center loudspeaker than the left and right. The center loudspeaker is usually of a smaller dimension and different (sideways) orientation to facilitate placement atop a television monitor. So any sounds assigned to the center could be quite compromised in terms of sound quality. As a result, many sound mixers avoid placing important sound elements solely in the center channel.

1.2.6 3-D Synthesis

3-D synthesis is another expedient approach that expands the spatial reproduction capability of audio by using only 2 channels. The design of these systems adapt to the realization that most people listen via 2-loudspeaker/channel stereo.

The general idea behind this approach is based on imposing a binaural (2-channel) HRTF equalization, phase and timing differences on monophonic input signals. This requires DSP (digital signal processing) to store many HRTF "signatures" that correspond to different positions around the listener.[1] The sound engineer decides where to place the sound element, and the corresponding transfer function is applied to the input signal producing a left and right output.

To reproduce this effect over loudspeakers requires an interaural crosstalk cancellation scheme similar to those discussed in section 1.2.4. Because of the reliance on crosstalk-canceling, the effect only works best for a confined central listening area.

This design approach has been realized into many products intended for the production of music, sound effects, multimedia, games, virtual reality, etc..

[1] This HRTF data can be measured by the designers themselves, or taken from various sources made public. For example, researchers at MIT Media Lab have made available comprehensive HRTF measurements from a KEMAR artificial head. This data can be found at; <http://sound.media.mit.edu/KEMAR.html>

There is a multitude of products available now that is too numerous to mention,[2] but below is a list some that have had more of an impact on music/sound production.
(They are all real-time[3] processes).

- **Q Sound** as a stand-alone unit, or software "plug-in" for Digidesign Pro Tools systems. (<www.qsound.ca>)

- **Roland** RSS (Roland Sound Space) with elevation and azimuth control, SRV-330 (digital 3-D reverberation), SDE-330 (digital 3-D delay) (<www.roland.com>)

- **I.R.C.A.M.** "Spatialisateur" software as part of MAX and jMAX options. (<www.ircam.fr/forumnet/>)

- **Desper Products** "Spatializer" and SP-1 "Spatial Sound Processor" stand-alone units.

- **Aureal** "A3D Pro" software "plug-in" for Digidesign Pro Tools. (<www.aureal.com>)

Harman International proposed an interesting system they called "VMAx" (Toole 1996) which is an abbreviation for "virtual multi-axis". The idea uses 2 loudspeakers (placed in front of the listener) that employ interaural crosstalk cancellation[4] with the intention of creating a "virtual" surround-sound reproduction environment. Binaural HRTF's are used to simulate the typical locations of surround loudspeakers (L, R, C, LS, RS). In other words, phantom images of

[2] As well, there are new additions and updates that make it impractical to compile a comprehensive list.
[3] There is a slight, but imperceptible delay due to processing and conversion.
[4] They applied the crosstalk-cancelling "transaural" patent of Cooper and Bauck (1992).

loudspeakers are synthesized to replace real ones. Virtual reality is simulating virtual reality.

1.3 Requirements of auditory images

Before successful design of a multichannel reproduction technique or system, we must gain an awareness of the requirements of auditory images regarding their spatial attributes.

Since we are beginning with a basic limitation of transmission channels equal to 5, we must expect that the system induces auditory images at locations other than the positions of the (5) loudspeakers.[1] Further to this limitation is the idea that we can only expect a coarse reconstruction of the spatial characteristics of the sonic wavefront. Instead, the system must extract and convey only the psychoacoustically pertinent information - a sort of data reduction. For example, our localization ability is much worse for sounds arriving from the sides, and overhead. In this situation, we may not need to encode/decode as fine a detail of spatial information as needed from the frontal direction.[2]

This brings up the question of whether the virtual reality system should cause the listener to perform better localization than in a natural setting. For example, front-to-back confusion often occurs in natural listening (especially for steady-state sounds). A virtual

[1] In addition, we often want to avoid localizing the loudspeakers as sound sources themselves. As Woszczyk (1993) advises: "The presence of the loudspeakers should be transparent and only virtual sources should be perceived".

[2] This is one explanation for the current practice of multichannel loudspeaker layout where 3 of the 5 channels are dedicated to reproduce the front total area of only 60° with the 2 surround channels to cover the remaining 300° total area.

reality system may defeat these localization weaknesses thus departing from its name-saken goal of simulating "reality".

One primary requirement is that the listener be able to externalize the auditory images. Virtual reality auditory display systems that use headphones (or earphones) too often suffer from "in-head" localization. Although it may seem that loudspeakers cannot produce such an effect, there have been experiments reported that do (Toole 1969).

Woszczyk (1993) proposes many interesting requirements of a multichannel system:

- the ability to portray small sound sources to large ensembles
- the ability to portray a variability of sizes and shapes
- the ability to portray a full range of perspectives from near and intimate to distant
- the ability to portray vertical cues for sensations of envelopment and object volume[3]
- the ability to control fusion and diffusion of images

The auditory image requirements can be broken down into 3 separate soundfield component categories:

1- Direct sound (localization)
2- Early reflection zone (sensation)
3- Reverberation (sensation and distribution)

[3] see also; (Furuya *et al*, 1995) for experiments regarding vertical reflections and their effect on image broadening.

The multichannel system must have the fundamental ability to present the direct sound of the sonic event with timbral accuracy and clarity (if it is intended).

The direct sound component is also most important in allowing us to discern the object's location. We rely on the transient component of the direct sound to be intact in order for the precedence effect to operate (in a room enclosure). The transient contains relative timing cues that help us differentiate it from the subsequent conflicting cues added by the room reflections.

Blauert (1996) summarizes a related phenomenon known as the "Franssen Effect". This effect features 2 loudspeakers (roughly) in a stereophonic layout. Loudspeaker #1 is fed a sine tone with a slow rise and fall time. Loudspeaker #2 is fed a sine tone that is amplitude-modulated by a rectangular-shaped pulse - in effect, a switching transient. The result is that only loudspeaker #2 is perceived as operational and the auditory event is located in that direction. It is as if the other loudspeaker never existed. This illustrates the strength of transient timing cues and the precedence effect leaving a powerful aural impression on the listener analogous to a brief flash of light obscuring our vision.

The complex directional radiation pattern of acoustical instruments cannot be captured by a single close microphone (Woszczyk 1979). An array of microphones placed at least 1 meter away would better encode the direct sound radiation pattern and its interaction with surrounding room surfaces. This interaction could either help reinforce or obscure the sound source location.

The presence of other instruments in a musical arrangement may be considered as "noise" signals, which interfere with and hinder localization. In effect, they act as spatial "maskers". Good and Gilkey (1994, 1996) reported experiments on the effect of noise on sound localization. More research needs to be conducted into the effect of multi-instrumental musical textures on the listener's localization ability.

One common current aesthetic regarding music[4] applications for surround sound is to present the direct sound of the instruments mainly via the front channel loudspeakers, and reserve the ambience for the rear surround channels (Yamamoto 1998) (Eargle 1998) (Fukuda et al 1997). However it should be noted here that the proposed system carries the intention to present direct sounds from any direction in the horizontal plane including the sides and rear. This is especially significant to the application of "spatial music" recording. (see section 5.1)

The early reflection(s) component of the total soundfield can be defined as the reflections of the direct sound that emanate from the room enclosure surfaces and arrive within 80 milliseconds. They can be labeled as 1st-order, 2nd-order, 3rd-order (etc.) reflections

[4] Most notably, classical music.

referring to their place in a time-sequence of any given sound travel path.[5]

Begault (1994, p.110) clearly describes the perceptual influence of the early reflection component:

> "The timing and intensity of early reflections can potentially inform a listener about the dimensional aspects of the environment immediately surrounding the sound source....Early reflections are not usually heard as distinct, separable sound sources, but instead are indirectly perceived, in terms of their overall effect on a sound source. A notable exception are reflections that are delayed long enough to be heard as distinct echoes".

To elaborate, if the reflection is delayed relative to the direct sound by less than 2.5 milliseconds, the image will be shifted towards its direction and slightly blur the overall auditory event. If it is delayed between 2.5 and 5 milliseconds, the image position will remain steady but will be somewhat blurred. If the reflection arrives from a lateral position and its delay is greater than 5 milliseconds. an enhanced sense of spaciousness will be the result (Theiss and Hawksford 1997).

An enhanced sense of spaciousness is deemed an important attribute of an auditory image. There is much evidence of listener preference for spaciousness as part of music listening (Barron, 1971) (Schroeder *et al*, 1974) (Ando, 1985) (Blauert and Lindeman, 1986). Therefore, it

[5] A 1st-order reflection refers to the reflected energy from a surface impinged upon by the direct sound. A 2nd-order reflection is a reflection of the 1st-order one, a 3rd-order is a reflection of the 2nd-order one, etc....

is important to have the ability to represent and control this property of spaciousness through the surround-sound medium.

Early reflections contribute to spaciousness through their influence on broadening the auditory image. This effect, commonly referred to as "Apparent Source Width" (from here on abbreviated as: ASW) is the sense that the auditory image is larger than the visual image of a sound source. It has been described as, "the image of an instrument or of the orchestra (will be) directionally softened and widened without losing a precise attack". (Damaske 1997)

(Ueda and Morimoto 1995), (Blauert and Lindemann 1985), (Morimoto and Maekawa 1988), (Morimoto, Iida, Sakagami and Marshall 1994), (Morimoto et al, 1993), (Morimoto et al, 1994) performed significant research into what affects ASW.

From their work, it can be summarized that ASW increases when:

- interaural cross-correlation coefficient (IACC) decreases[6]

[6] IACC is a measure of correlation between the 2 ear signals. A coefficient value of 1 equals maximum correlated binaural signals. Values less than 1 indicate increasing difference between the left and right (Beranek 1996).

Measurements using an artificial head result in maximum IACC for a sound arriving from 0° and slightly less for 180°. IACC falls to a minimum at around 60° (MacCabe & Furlong 1994) (Ando 1985). Listener preference has also been investigated showing that early reflections arriving from ±20° scored the highest (Ando 1985) (Ando and Kurihara, 1985). So it would seem necessary to pay particular attention to encoding sonic events (or, reproducing auditory events) from this area (±20° - ±60°).

- loudness increases
- lateral energy fraction increases
- low-frequency energy increases (especially between 100 - 200 Hz)

Furthermore, it has been suggested that lateral reflections having spectral energy only above 1.6 kHz do not contribute to ASW (Morimoto, Iida, Sakagami and Marshall 1994). This supports the notion that low frequencies are influential on spaciousness.

One final important (but almost overlooked) point is the conclusion (by Morimoto *et al*, 1993) that ASW is widest for a direct sound arriving from 0° than it is for those at 30°, 60° and 90° even if the IACC is kept constant. This is most likely due to the maximization of the interaural-difference (or binaural) effect along the symmetrical axis for a sound arriving from straight-ahead and center.

The potential for surround-sound to reproduce and create these qualities of spaciousness should be further explored.

(Morimoto et al, 1993 & 1994) concluded that ASW is equal despite the number of early reflections and their direction as long as the IACC was constant.

CHAPTER 2 RATIONALE OF DISSERTATION

Unlike with stereophonic recording, there are no clearly established microphone techniques for 5.1 multichannel recording. There is a definite need to introduce and investigate new techniques unique for surround-sound in a systematic way.

2.1 5-Channel Microphone Array with Binaural Head

2.1.1 General description

The proposed solution to improved multichannel representation of the spatial characteristics of a soundfield is basically a static array of 5 microphones. The array can actually be thought of as comprising of 2 distinct systems: one for the front 3 channels, and the other for the 2 surround channels. Figure #2 graphically depicts the microphone setup.

The first system is mainly to encode the front horizontal 60° (±30°). 3 directional pressure-gradient microphones are used here. The left and right channels use super-cardioid pattern microphones while the center channel uses a cardioid pattern.[1] The lateralization is achieved through both intensity and timing differences between the independent microphone channels. (see section 3.2) Intensity differences are provided by the directionality of the microphones.

[1] The microphones used are the Sennheiser model MKH-50 (super-cardioid), and the MKH-40 (cardioid). See figure #3 for specifications.

Timing differences are provided by the physical separation of the microphones.

The 2nd system uses a custom-designed binaural head fitted with a pressure omnidirectional microphone in each "ear". (see section 3.1) Through binaural encoding of the sonic event, the system can represent the spatial horizontal territory from 30° all the way around to 330°. It is this system that allows representation of the most demanding and difficult area to reproduce which is the side areas between ±30° and ±90°.[2] It is well known that intensity difference techniques cannot provide stable and credible auditory images in this region. Spaced-apart microphones allow for additional time-difference cues that can improve the illusion of sounds from that region. It is the spectral head and pinna-related cues provided by the binaural head in conjunction with the spacing between the head and the frontal trio of microphones that allow images to be reproduced here. These same cues allow imaging of sounds to the rear of the listener. As well, auditory images can be resolved from overhead positions with a fair degree of control. (see section 4.1.2)

The inclusion of binaural technology into this surround-sound system grew out of initial research by the author into music recording applications using artificial head microphones. It was found that various binaural demonstration sound effect recordings (Audiostax,

[2] This particular region carries special importance since reflected sound energy from here produces a much sought-after sensation of spaciousness through increasing the apparent source width (ASW) and listener envelopment (LEV). Section 1.3 discusses the relevant factors of this attribute.

1987) exhibited extraordinary spatial impressions over conventional 2-channel stereo. It was felt that there should be more research into how to take advantage of this spatial-rendering ability for music reproduction despite its apparent lack of acceptability in this application.

Distance and depth attributes of a sonic event can also be encoded with the proposed system (see section 4.1.3).

Sound objects in motion, whether actual or simulated (by distributed loudspeakers) can be captured and reproduced (see section 4.1.4).

Although, as expressed above, the array can be bisected into 2 separate sub-systems, they work together to form unified and whole spatial images.

2.1.2 Problems with binaural technology

The basic difficulty with incorporating binaural technology into a virtual reality system is that, in effect, the designer must select one particular head and torso-related characteristic. This is in conflict with the significant individual variation of physical dimensions inherent with every end-user of such a system.[3] This variation

[3] The size, shape and angle of the pinnae usually vary with age so that

equates with the directional sensitivity of the listener resulting in inaccurate localization cues.

Recent research by MØller *et al.* (1997) compared the localization performance of 8 different models of artificial head microphones with that of real life. The subjects listened to recordings made by the 8 different binaural microphones through headphones. The results show a higher percentage of localization errors for all head models than for real life. The most significant differences being for sound sources along the median plane.

2.1.2a Front/back reversals

The compromise of having to use a non-individualized HRTF is believed responsible for many of the shortcomings of binaural reproduction. It can lead to equalization (spectral) errors as well as commonly reported localization errors such as front/back reversals. However, many researchers (Wenzel et al. 1993), (MØller et al. 1996), (Begault & Wenzel 1993), (Butler & Belendiuk 1977) have concluded that quite useable directional information can still be translated through non-individualized HRTF's with the major focus of localization discrepancies occurring along the median plane with front/back ambiguity.

an individualized HRTF can become obsolete even for a particular person.

The simplified spherical model (Kuhn 1977) of the human head in a sense, predicts the so-called "cone-of-confusion" related errors of front/back ambiguity. This is based on the model's assumption of IADs (interaural amplitude differences) and ITDs (interaural time differences) derived from a symmetrical sphere-shaped head. But, an actual human head is asymmetrical implying an effectively changing diameter for every angle of sound incidence. So in real life, this asymmetry coupled with spectral "shadowing" of the pinnae, and, slight head movements can often resolve ambiguous front/back sound locations due to confusing IAD/ITD cues.

With binaural reproduction, front-to-back reversal is a more common occurrence than back-to-front. This may be the result of a missing visual correlate with the auditory image - if the object can be heard but not seen, it is reasoned that it must be behind. As an example of another manifestation of this effect; even a visual cue as "artificial" as a pair of loudspeakers positioned in front of the listener can result in frontal auditory images with ease, and rear auditory images being difficult to imagine (Toole 1991, p.7).

Mori *et al.* (1979) suggest that while reproducing binaural signals over loudspeakers:

> "...in the case where the reproducing loudspeakers are positioned at the rear (270°), the localization of a rear sound source (Ø=180°) is accurate....when the loudspeakers are in front (30°), most listeners assume (incorrectly) that the actual sound source was also

> recorded up front. Thus it is clear that unambiguous determination of front and rear sound sources depends greatly on the relative positioning of the reproducing loudspeakers. In the case where only the forward direction is considered to be important, 2-channel reproduction with two front loudspeakers is sufficient. However, when rearward sound sources must also be given consideration, 4-channel reproduction with four loudspeakers becomes necessary". [4]

As in any sound localization task, the spectral content of the signals can influence the success rate of frontal vs. rear. Signals with a limited bandwidth (particularly ones lacking high frequencies) are more likely to lead to front/back confusion (Begault 1994, p. 79).

It should also be noted that many of the reported front/back reversal experiments were observed under artificial, controlled conditions where the listening room was anechoic, and, symmetrical about the front and rear axis. Front/back ambiguity can be reduced by recordings made and/or reproduced in an asymmetrical environment which is the usual situation anyhow.

[4] From this, it is clear that actual sound sources (i.e. loudspeakers) helped to coerce an auditory image at its intended front (or rear) location. But it is not clear whether a visual "awareness" (by the subjects) of the loudspeaker positions played a part in this.

2.1.2b In-Head-Localization

Another common ailment with binaural reproduction is the effect that the auditory image tends to be perceived "inside" the head of the listener. This is a gross misinterpretation of the direction of a sonic event since the notion of localization should also include distance as well as direction. The desired effect is usually the opposite; the image should be "externalized". In-head-localization (IHL) is mainly due to the listener's ears receiving an implausible combination of sound pressure levels and relative signal delays. Consider a typical situation with centrally pan-potted signals (presented via headphones) where both ears hear a sound with the exact same level and time of arrival across all frequencies. This situation cannot occur in real life listening since the head and outer ears modify the oncoming sound wave causing differences at each ear. The senselessness of these artificial cues leads to a defaulted auditory image inside the head. In nature, it makes sense that only one's own voice (with closed mouth) can be localized within the head.

Localization over headphones is more properly referred to as *lateralization*, (Plenge 1974) since auditory images are not externalized but appear at positions just outside each ear and in a line between the ears. Hence, lateralization better describes our spatial sense being confined to a lateral left-right axis.

This problem is the result of diotic reproduction over headphones. The effect is difficult to induce via loudspeaker reproduction although Toole (1969) reports some experiments that do.

Toole reproduced coherent signals (octave-band-filtered white noise) over 5 different variations of loudspeaker layouts within an anechoic room:

A- one loudspeaker at 180°
B- one loudspeaker at 0°
C- two loudspeakers at 90° and 270°
D- two loudspeakers at 0° and 180°
E- four loudspeakers at 0°, 90°, 180°, and 270°

Actually, in-head localization (IHL) was reported in all configurations. The single rear position (A) was the least effective at producing IHL while position E resulted in the highest percentage of IHL with in most cases being 100%. Position C fell somewhere in the middle of these results. One interesting comment (that warrants further investigation) by some subjects was that "the low frequency sounds were too large to be contained within the head". (Toole, 1969 p. 945) This might explain the subtle trend in progressively lower IHL percentages as the f_c octave-band decreased as reported by Toole. This may also be do to the increased influence of bass frequencies on bone-conduction which could cause IHL, much in the same way as one's own speaking voice. Also, the anechoic test room used in Toole's invesigation (coupled with loudspeaker sound sources) produced an unnatural listening environment where IHL auditory images can more easily be formed. This is in effect, similar to headphone reproduction.

However, with typical surround-sound loudspeaker configurations, IHL is not really a concern unless reproducing "artificial" signals such as sine tones, or filtered noise. Music is a dynamically changing signal that doesn't easily allow the establishment of in-head auditory images. The reason that IHL can occur so readily with headphone reproduction is the resultant static soundfield reproduction despite listener head movement. The use of non-individualized HRTFs can also play a part in eliciting IHL.

2.1.2c Binaural signals over loudspeakers

Loudspeaker playback presents other complications regarding the reproduction of binaural signals. It has been proposed (Pompetzki 1990) that the difficulties involved in loudspeaker reproduction are to blame for the lack of acceptance of binaural (dummy-head) recordings for music applications.

First, there is the fundamental problem that loudspeaker playback will result in significant interaural crosstalk signals. Many schemes have been developed to counteract the unwanted crosstalk leakage. Generally speaking, this is a complicated solution. (These were briefly discussed in section 1.2.4).

Pompetzki (1990) lists 4 reasons why crosstalk-canceling schemes "fail":

1- They require a very precisely defined listening area.

2- This tight listening area necessitates only one listener.

3- The listening room must be highly absorbent in order to minimize crosstalk.

4- Special equipment is required at the playback-end of the signal chain in order to process the crosstalk cancellation.[5]

As well, crosstalk-canceling signals need to be removed themselves so that each generation of canceling signals will need their own attendant cancellers. This obviously complicates the processing required and can introduce unwanted phase distortions in the signal.

2.1.2d Binaural microphones - characteristics

The designers of typical microphones for music recording applications strive for a transfer function independent of direction of sound incidence. An ideal omnidirectional microphone will pick up sound from all directions at equal level across its entire frequency range. An *ideal* directional microphone will attenuate sounds from different angles, but this attenuation will be independent of frequency. These are termed "free-field"[6] microphones since their primary frequency response is calibrated to be as flat as possible for sound incidence at 0°. Of course in practice, it is difficult to realize this ideal since the physical size and shape of the microphone body

[5] Pompetzki (1990) proposes a solution that the processing (crosstalk cancellation and equalization) be made at the originating end of the chain.

[6] The term "free-field" also implies the exclusion of any reflected sound energy (i.e. anechoic).

and capsule element obstruct and diffract off-axis sounds. This results in a non-linear attenuation of frequency as a function of angle.

In contrast, a binaural artificial head microphone is intended to have a different transfer function for different angles of incidence. This is the basic idea of the "Head-Related Transfer Function" (HRTF) where, the physical geometry of the head, shoulders, and pinna influence and impose a unique frequency response for every possible direction of sound incidence.

If we equalize the left and right signals from an artificial head microphone to be flat for frontal sounds, then we are applying the same "free-field" equalization principle as regular studio microphones. An argument in support of this approach might be that most sound sources are recorded (and reproduced via loudspeakers) from the front, so the artificial head should act in the same manner as a regular stereo microphone. However, the timbre of sounds from other directions will be altered. (In general, the practical result of this is a high-frequency roll-off for off-axis sounds from 30° - 330°).

Often, a binaural microphone is placed further away from the sound sources - outside of the room radius and into the diffuse field of the hall. A free-field equalized artificial-head will then result in an overall colored timbre of diffuse sound reflections and reverberation that make up the majority of the total sound.

A solution to this problem is the so-called "diffuse-field"[7] equalization. The goal of this approach is to have an average equalization for all possible directions (including that of the sound source) so that tonal colorations are kept to a minimum (Theile 1986) (Møller 1992) (Larcher et al. 1998). In the end, the difference between the average frontal response and the average diffuse response is not that great.

For most angles of incidence, there is a broad resonance centered around 2500 Hertz. This peak could be between 10 to 15 dB and is caused by the primary resonance of the largest external ear cavity known as the "concha". An artificial head microphone would add this resonance to the incoming sound as well. However, reproducing these signals over loudspeakers (from any azimuth) will cause a doubling of this resonance as the sound passes through the external ear twice - once via the artificial head, and the second time through the ear of the listener. This principle resonance needs to be eliminated through electronic equalization (Killion 1979). Otherwise, a marked tonal coloration of the sound will occur especially since the human auditory system is most sensitive to this mid-frequency region of the spectrum.

Griesinger (1989) proposes an additional form of equalization for loudspeaker reproduction of artificial-head recordings. In natural hearing, there is little difference in level between the 2 ear signals at

[7] The diffuse field is defined as "a sound field in which the sound pressure level is the same everywhere and the flow of energy is equally probable in all directions" (Brüel & Kjaer Pocket Handbook - Noise, Vibration, Light, 1986).

low frequencies due to the long wavelengths diffracting around the head. So, at low frequencies (below 800 Hz) we rely upon slight phase differences to locate a sound. An artificial-head microphone would also encode little level difference (at low frequencies) so that when its signals are replayed over 2 loudspeakers, we would have difficulty localizing the sound. Any phase differences would be obscured by the crosstalk signals reaching the contralateral ear. Griesinger proposes a method to increase (level) differences or, separation between the 2 reproduction channels by adding a left minus right (L - R) signal at low frequencies. This is in effect, adding a reverse polarity (-180°) boosted low-frequency left channel signal to the right channel, and vice versa. The result is increased separation at low frequencies which helps resolve the location of sounds. Griesinger calls this "spatial equalization" but the basic idea is often referred to as "shuffling" as derived from Alan Blumlein's patent inventions (Blumlein 1933).

2.1.2 Free-field Microphone Techniques

The basic premise for free-field microphones is that their frequency response is expected to be flat (or flat as possible) only at 0° on-axis.[1] As the direction deviates from 0° (in any plane), the frequency response changes. These changes are merely influenced by the shape and dimensions of the microphone capsule and body. On the other hand, an artificial-head binaural microphone also has deviations in frequency response dependent upon angle of sound incidence - but these intentional deviations are related to natural spatial hearing. Free-field microphones have no such natural relation, so their directional response becomes randomized depending only on the microphone design.

Directional microphones follow the pressure-gradient principle and as such, suffer from inconsistencies in low-frequency response dependent upon the distance from the sound source (Eargle 1981). Close mic'ing within (typically) 0.5 meters will exhibit an unnatural bass boost (known as the "proximity effect"). More distant microphone placement will result in a weakened bass response tending to sound thin at great distances more than a few meters. Artificial-head microphones use pressure-sensitive microphones[2]

[1] There are many microphones that are by design, meant to be non-flat at 0° on-axis. Microphones intended for close vocal recording or, bass instruments are 2 examples where the frequency response is meant to "flatter" or enhance the pickup of certain sound types.

[2] The human ear-drum is considered to be a pressure-sensitive, so in yet another way, the artificial-head microphone can simulate natural hearing.

that do not suffer from these distance-dependent low-frequency variations (Wuttke 1985).

Free-field microphones have been known to provide quite satisfactory results in terms of spatial encoding of a soundfield to be reproduced over a standard stereo loudspeaker arrangement. The spatial performance of stereo microphone techniques in this region (±30°) is of significance since this is the region where our horizontal localization acuity is at its highest (Mills 1958). This is one domain where free-field microphones are superior to artificial-head microphones both in terms of spatial and timbral imagery (in the frontal ±30° region). With surround-sound, the additional center-loudspeaker channel can allow greater spatial resolution along with more stable auditory images.

Because stereo microphone techniques rely on interaural amplitude and/or timing differences, the reproduction of their signals requires a symmetrical arrangement with respect to the 2 loudspeakers and both ears of the listener. Due to the placement of our ears, we cannot extract an auditory image to the sides from routing signals to (for example) the front-right and surround-right loudspeakers. The interaural differential cannot operate in the same symmetrical (or equal) fashion as frontal stereo. Even if we route a signal to the sides, any slight forward/reverse movement of the listener will cause the image to fall into the nearer (and louder) speaker. The summing-localization breaks down in this situation. On the other hand, by providing the proper head-related direction cues, an artificial-head microphone can deliver the illusion of an image towards the sides. This is because it can reconstruct the "correct" signals at both ears which correspond to that particular direction.

Elevated or overhead auditory images are generally not possible with regular microphones as they are with an artificial-head microphone. However there has been some studies (Blauert 1969) (Bloom 1977) that suggest it may be possible to create elevated auditory images by means of electronic modification of a sound's spectrum. This is through manipulation of "directional bands" (at high frequencies) which have been reported to incite an illusion of elevation.

3.1 DESIGN AND CONSTRUCTION OF BINAURAL MICROPHONE

The body of the head and torso simulator (HATS) was cast of hard fiberglass material.[1] Actually, it does not include a complete torso, but there is a neck and shoulder simulation. The presence of shoulders will cause a small interference effect due to the effective path length difference between the direct sound and shoulder reflection reaching the ear canal entrance (Burkhard and Sachs 1975). The average path length difference here is about 14 cm. resulting in a half-wavelength cancellation effect around 1228 Hz of about 3 dB.[2]

A (photocopy version of a) photograph of the artificial head is shown in figure #4.

The surfaces are smooth and continuous with mild contours that form facial features such as cheekbones, nose, lips, eyes, chin etc.. Although some researchers have gone to extremes to model the physical characteristics of soft human flesh (Bauer *et al.* 1967) and (as reported in; Sunier 1989), there is no evidence that it has any

[1] A small company specializing in the manufacture of mannequins was contracted to build the HATS body to specification. The company's name and location is: Mannequin Shop, 766 Notre Dame Ave., Saint-Lambert, Quebec.

[2] This also corresponds with the findings of Burkhard & Sachs (1975, p.219) when testing the KEMAR mannequin.

significant effect on the response of the head. Burkhard and Sachs (1975) recast a duplicate KEMAR dummy-head with flesh-like rubber to specifically investigate this feature and concluded that (it):

> "...shows no more than 1-dB difference at a few frequencies over the range extending to over 8000 Hz....Being hard headed does not affect significantly the sound that will be incident..." (p. 220)

It is more critical that the material composition of the external-ear replicas match the acoustic impedance characteristics of the human ear since the direction-dependent reflections occur here (so close to the transducer element). A special pair of "ear moulds" was ordered through an audiometric design company.[3] These were made of a soft, malleable rubber material simulating the texture of a real outer ear.

A hole was drilled through the rear block of the ear moulds in order to fit a microphone into each. The microphone was inserted through the back until the capsule was almost flush (within 2 mm.) of the ear canal entrance. This is the optimal position for collection of sound with the directional properties of the sonic event intact. Positions further into the ear canal cause resonances that are non-directional related.

Many designers of custom artificial-heads opt to use miniature microphones since they are easy to insert into the ear-replica.

[3] The ear moulds are called: "soft plastic ear", item number 2317 ordered from, Dahlberg Hearing Systems, P.O. Box 9022, Kitchener, Ontario, N2G 4J3.

However, they suffer from inferior quality compared to regular-size recording microphones. Problems include very low sensitivity that would result in a low signal-to-noise ratio. The resultant high pre-amplifier noise level could mask the fine detail of the spectral direction cues as well as produce a low-quality recording. Also, many miniature microphones have an overall erratic frequency response due to compromises in their size. One other point concerns the recording of very high frequencies (above 10 kHz) where the miniature microphone would sample the sound pressure at one specific point. As Møller (1992) states; "It would probably be better to measure the sound pressure integrated over a cross-section of the ear canal - just like the hearing does itself at the eardrum".

For this design, regular studio microphones were chosen for the in-ear transducers. Despite their size, the head was designed to be hollow enough so that the microphone bodies could pass through to the opposite side. Special right-angle XLR connectors were fitted onto microphone cables to minimize the physical obstruction that regular connectors would otherwise make outside (and behind) the pinnae.

The microphones themselves were of the highest quality of pressure omnidirectional type.[4] They have an exceptionally high signal-to-noise ratio with a sensitivity rating of 25 mV/p and a noise specification of 10 dBA. Another reason for this choice was that they belong to the same "family" type of microphones used for the frontal array with similar capsules (that only differ because of their

[4] The microphones are the Sennheiser model MKH-20. (see figure #3)

intended polar pattern differences). Otherwise, they have the same free-field dynamic specifications (i.e. sensitivity and noise) allowing optimal matching across all 5 channels. The nominal impedance (150 ohms) is also the same allowing more predictable and equal performance when coupled with identical microphone preamplifier channels.

The inclusion of pressure-type transducers as part of the total system is advantageous since (unlike pressure-gradient directional types) they maintain a flat low-frequency response independent of the distance from the sound source (Wuttke 1985).

3.1.1 Measurement and equalization

Initial trials with the artificial-head microphone showed promising results in terms of spatial rendering, but there was an obvious tonal coloration for all sound directions.

Maximum-Length Sequence measurements were taken to verify the frequency response of the head at selected angles; 0°, 30°, 45°, 60°, 90° and 180°. (The graphical results are shown in figures #5 - 10). The frequency range was limited to a bandwidth of 300Hz to 20 kHz since data acquired below 300 Hz was unreliable. The reference calibration frequency response was derived from a Brüel & Kjaer model 4007 microphone.

Taking 1 kHz as the reference, there is clearly a broad peak at approximately 2.5 kHz. This is due to the concha cavity resonance as mentioned previously. Next in importance is the overall downward trend in level above 5 kHz. Within the range of 5 and 20 kHz, there are a number of peaks and dips that tend to move depending on the direction of sound incidence. There is one notable dip at around 8.8 kHz that is present for most directions.

An approach similar to the "diffuse-field" equalization method was chosen to help correct the overall response of the head for loudspeaker playback. Since the above-mentioned anomalies occur to a relatively greater degree at all 6 directions, it was reasoned that these should be equalized. In a sense, it is an average of the 6 samples taken, however, only 1 angle (90°) was chosen to apply a corrective EQ to. In "free-field" equalization methods, the frontal region within ±30° is chosen as the reference. But, this proposed system will rely on another set of (free-field) microphones to represent the frontal region[5]. The artificial-head is meant to represent the spatial regions beyond ±30°, and, the loudspeakers feeding these binaural signals will be positioned to the sides of the listening area.

[5] as presented in section 3.2

A digital equalizer[6] was used employing 6 separate bands. Figure #11 shows the unequalized vs. the equalized signal for the right ear. Figure #12 shows the settings used to achieve the equalized curve.

Various recordings made using the total system were played back using this equalization. It was found to be a large improvement on the overall timbre of the source material. However, it seemed to be too bright overall, as well, the noise floor was lifted to an audible level by some of the extreme settings. As a consequence, it is believed that only 3 bands (or areas) need to be attended to:

1- the large resonance at 2.5 kHz
2- the overall drop at high frequencies
3- the dip at 8.8 kHz

A slight shelving boost (depending on the nature of the recorded source) will compensate for #2. The simple shelving filter characteristic will maintain the individual direction-coded spectral "signatures" intact.

3.1.2 Crosstalk

It would seem that loudspeaker playback would violate a fundamental rule of binaural signals needing to be exclusive in terms

[6] The 90°-reference signal was run through the equalizer section of a digital mixing console, the Yamaha 03D.

of left and right. There is unavoidable contralateral crosstalk with loudspeaker playback in the typical stereo positions of ±30°. However, within this proposal, the binaural signals are fed to the rear surround speakers that are typically placed around ±110-120°.

Referring to Blauert (1996, p.73) the maximum IAD (interaural-level difference) varies between 100° and 110°.[7] McCabe and Furlong (1994) also show lower IACC (inter-aural cross-correlation) values at 100° and 110° compared to 30°.[8] There could be as much as 20-dB difference in level between the 2 ears at 110° for frequencies above 2 kHz. In effect, placing the loudspeakers at the sides results in using the listener's own head as an acoustical barrier that behaves as a crosstalk "minimizer". This contralateral crosstalk reduction is most effective at high frequencies, which is where the fine location-dependent spectral differences reside.

For low frequencies, there is definitely contralateral crosstalk signals occurring from the 2 surround loudspeakers. This is because the listener's own head is not a sizeable enough acoustic barrier to this range of wavelengths.

One way to increase separation at low-frequencies is to use electronic equalization as proposed by Griesinger (1989) as "spatial equalization", and Blumlein (1933) as "shuffling". (an as discussed in section 2.1.2d)

[7] Blauert assumed a spherical model of the head with a 17.5 cm. diameter and ears positioned at 100° and 260°.
[8] The lowest IACC value occurs at 60°.

This was attempted with some of the pre-recorded material where the binaural surround signals were duplicated and cross-fed to the opposite side. They were low-pass filtered (around 500 Hz), delayed (about 700 µs.), and polarity-inverted using a digital mixing console.[9]

The results were not very successful. It is believed that this was mainly due to interference between common signals picked up by both the artificial-head and the frontal microphones. The polarity-inversion altered the timbre of the sound too noticeably. When the binaural surround signals (with the crosstalk EQ signals) were isolated (i.e. the front L,C,R channels turned off), the results were much better. This is in accordance with the intent of Griesinger's and Blumlein's proposal where it is meant to work only with 2-channel source material. With more channels (and loudspeakers) another method of low-frequency separation would have to be researched.

In addition, both Griesinger and Blumlein's proposals were designed for the geometry of typical 2-channel stereo loudspeakers (as in figure #1). Again, perhaps a better solution tailored specifically for surround loudspeakers could be researched.

[9] The Yamaha 02R (software version-2).

3.2 Design of the Frontal 3-Channel Microphone Array

The second (of two) components comprising the total proposed system is specifically designed to spatially codify the frontal area of ±30°. This is the same azimuth range covered by 2-channel stereophonic systems, but now has an additional (third) center channel.

This particular area carries the most stringent demands on imaging since localization acuity is at it highest level here. Mills (1958)[1] conducted localization experiments around the horizontal plane to test for Just Noticeable Differences (JND) in sonic event position which he called; minimum audible angle (MAA). His experiments showed that the minimum audible angle was as fine as 1° in the frontal region. Beyond 30°, the MAA steadily rose up to 60° (MAA = 2°) then, rapidly rose until reaching its largest value (MAA = 10°-40°) near the 90° point. The main explanation for this decrement in acuity is the variance in interaural time differences (ITD) from the front to the side. A 5° change in azimuth from 0° produces a change in ITD of 43 microseconds compared to a 2 microsecond ITD change from 85° to 90° - a factor of 21. This physical fact makes the ITD cue more significant (therefore improving localization acuity) in the frontal area as compared to the sides.

[1] Mills experiments tested the localization acuity of serially or sequentially presented displaced sounds that were spectrally similar. More recent research by Perrot (1984) employed a different approach where the 2 sounds were presented concurrently instead of sequentially. However, the results in both cases were similar.

This inferior localization performance from 30° towards 90° can perhaps reveal some important imaging tendencies with LRC reproduction. Whenever the centrally located listener turns their head to face towards (for example) the right loudspeaker, this results in their left ear facing the left loudspeaker. Therefore, sound from the centre and left loudspeakers arrives at angles between 30° and 60° where the MAA acuity begins to worsen. This situation begins to resemble (the deficiencies) of natural hearing and we should not be too quickly discouraged by poorer localization performance in directions other than the focus of the head movement.

To encode the total 60° frontal region, it was decided to use 3 separate microphones dedicated to the 3 reproduction channels (L, R, C). The intent here was not to record and reconstruct the spatial wavefront, but to attempt to provide the proper IAD and ITD cues at the ears of the listener.

Informal listening tests were performed by the author to help determine ITD and IAD values for localization within the frontal (L-C-R) area of a standard surround sound setup. Only 2 loudspeaker channels were activated: the left and center channel.

Sending a repeating periodic pulse-like signal (of equal amplitude) to both loudspeakers produced an auditory image roughly at midpoint, around 13° left of center. To perceive an image at the more accurate midpoint of 15°, the center channel required a delay of about 204 microseconds, or, a gain of -1.4 dB. This was considered the equilibrium benchmark position equivalent to a center phantom image with 2-channel stereo. So, any intended auditory image at ±15° should have one or a combination of these interaural difference values.

The next position of interest was the most extreme point of ±30° - essentially the position of the right loudspeaker. It was found that a delay of around 412 microseconds (applied to the center channel) was the threshold interchannel delay value that firmly produced an auditory image at 30°. Any increase in delay did not cause the image to shift laterally any further. An interchannel level difference of 12.5 dB produced the same results.

An analysis of these results led to the following comparison. In natural localization, the ITD value for a sound at 30° azimuth is approximated at 247 microseconds. This is derived from Kuhn (1987) using the formula:

$$\mathrm{ITD} = \frac{2a}{c} \sin\emptyset$$

where; a = average radius of head (.085 meters)

c = velocity of sound (344 m/sec.)

$\sin\emptyset$ = angle of sound incidence

The interchannel time difference value of 412 microseconds (for a sound image at 30°) observed in the listening experiment is a factor of 1.67 times the natural ITD of 247 microseconds. The natural value (of 247) corresponds to an ear spacing of 17 centimeters. Therefore, multiplying this dimension by a factor of 1.67 results in a separation of 28.5 centimeters.

In summary:
412/247 = 1.67
(since 247 microseconds corresponds to an ear-spacing = 17cm.)
17*1.67 = 28.5 cm.

The greater interchannel time difference value of 412 microseconds over the natural ITD of 247 is due to the interaural crosstalk that occurs when 2 loudspeakers are delivering signals to the listener. The crosstalk tends to offset the auditory image towards the center somewhat. This crosstalk does not occur for a natural sonic event where only one signal is received at each ear.

It was decided that 28.5 cm. would be a good starting point for the separation distance between the left and center channel microphones. This would result in the observed 412-microsecond time difference for a sound object at the extreme end of the frontal area (i.e. 30°).

In the early stages of this investigation, a coincident positioning of the 3 microphones was attempted. The results were not satisfactory due to several reasons:

1) Only intensity difference cues could be presented, timing difference cues would be absent. When there exists only intensity difference cues as presented by spaced-loudspeakers, any movement of the listener towards a loudspeaker will distort the spatial image. Additional time-differentials are needed to reinforce the intended spatial positions (Snow 1953, Long 1972).

2) First-order gradient microphones do not have enough directivity to avoid buildup of a center image when used in conjunction with a center channel. The included angle of the L and R channel microphones would have to be extremely wide resulting in the optimal 0°-axis of the microphones to be directed away from the frontal soundstage. This situation would have much of the direct sound picked up severely off-axis of the microphones.

Since intensity differences cannot be pronounced enough (for LRC reproduction) with a coincident assembly of microphones, the interaural differences would have to be dominated by timing differences between the microphone signals. However, directional microphones were chosen for the frontal array mostly to allow sufficient separation from sounds intended for the rear channel signals accounted for by the binaural head.

The super-cardioid pattern was chosen for the left and right channels for 3 reasons:

1. "Of all the patterns, it accepts the smallest amount of energy from the rear half-room" (Dickreiter 1989) also, (Boré 1978, Woram 1989). This would help avoid the typical problems of front-back imaging confusion.

2. The extra "focus" or "reach" (over cardioids) of the super-cardioid pattern would allow the left and right microphones to better pick-up the typically further sound sources to the sides of the soundstage. The supercardioid pattern is known to have a "Distance Factor" of 1.9 compared to the cardioid's 1.7 (Woram 1989). This is a factor of roughly 1.11. This factor would compensate somewhat for sound sources that are off to the extreme left or right of the frontal sound stage which can be typically more distant than central sounds.

3. The increased directivity of the super-cardioid pattern coupled with the spacing between the (L and R) would yield a lower inter-microphone cross correlation. This low correlation would discourage the establishment of a strong central phantom image (between the left and right channels) which is desirable to leave room for the separate centre channel signal. A strongly established phantom image in addition to a centre channel would result in a central image buildup.

A cardioid pattern was chosen for the center channel microphone. The reasons are twofold:

1. The maximum rejection of sound from 180° would avoid picking up sounds that are intended to be reproduced at the rear. This would lessen the chances of front-back imaging confusion.

2. The cardioid pattern also allows the widest acceptance angle from the front. This is necessary for the typically most important front-center images. Both in theory and practice, a cardioid can obtain a fuller frequency response across a wider frontal area than with directional microphones of a higher degree such as super-cardioid and hyper-cardioid.

For the frontal area represented by the left, right and center channel (loudspeakers), the directional information is mainly encoded as timing rather than level differences. For the example of a sound source arriving at 30° to the array, the level difference between the right and center microphones is slight.

To illustrate: the polar pattern equation (Woram 1989) for a cardioid is:

$$p = 0.5 + 0.5cos\emptyset$$

for a super-cardioid:

$$p = .366 + .634cos\emptyset$$

so, for a sound incidence at 30° to the array, this will arrive at 0° on-axis for the super-cardioid and will equal unity (or 1) resulting in no attenuation.

The cardioid microphone would pickup the sound from 30° off axis, therefore:

$$p = 0.5 + 0.5\ cos(30)$$
$$p = 0.5 + 0.5\ (0.866)$$
$$p = 0.933$$

the level difference would be calculated from,

$$\begin{aligned} 20 \log (0.933) &= \text{level attenuation (in dB)} \\ &= -\ 0.6 \text{ dB} \end{aligned}$$

From this simple analysis, it is clear that there is little level difference between the right and center microphones forcing the directionality to rely upon spacing-induced timing difference.

Figure #2 shows the final configuration of the 3 front channel microphones and their relation to the binaural head.

A custom-designed microphone stand bar was built to support the 3 microphones on a single stand for ease of setup and movement.

3.3 Convergence of technologies

It appears that the convergence of binaural technology and surround-sound as proposed here can help solve many of the common individual problems of each technique.

For binaural techniques:

- The in-head localization problem is eliminated by use of external transducers (i.e. loudspeakers) which can rarely cause IHL. This is because, unlike headphones typically used for binaural reproduction, they are truly external sources forcing external auditory images. As well, listener head movement will cause corresponding changes in IAD, ITD and spectral cues which would be absent in an artificial listening environment like headphones. The loudspeakers would also "activate" reflected energy in the listening room promoting a static reference environment.

- The problem of front/back reversal is greatly reduced due to the loudspeakers being positioned both in front (-30°, 0°, +30°) and in the rear (110° and 250°). This allows the establishment of more solid auditory images in both quadrants. As well, using the free-field trio of microphones for the frontal region reinforces the clarity of those intended sounds as compared to the rear.

- The hindrance of complicated and difficult-to-implement crosstalk-cancellation schemes deemed necessary for binaural reproduction over loudspeakers is avoided. This is due to the placement of the 2 surround loudspeakers assigned the binaural signals. This placement results in a simple acoustical crosstalk reducer (relying on the listener's own head acting as a barrier) sufficient to transmit the necessary spectral differences to each ear at high frequencies.

- The problem of finding a suitable equalization for headphone-reproduced binaural signals is avoided. (Wearing headphones deforms the listener's outer ear in an unpredictable manner causing a wide variation of tonal colorations). A simple equalization is applied that removes the global timbre colorations evident at all angles of incidence and allows improved reproduction over loudspeakers.

- The noise and inferior frequency response of miniature microphones often used in artificial-heads is avoided by incorporating regular-size, high-quality studio microphones instead.

For multichannel reproduction:

- Although surround-sound reproduction is a great improvement in expanding the spatial ability over stereophonic systems, it still has great difficulty in resolving auditory images in the regions between 30° and 100° and, in the rear. This is because the 5-loudspeaker layout must rely on IAD and/or ITD cues to resolve so-called phantom images. Incorporating binaural head and pinna-related signals into surround-sound can resolve images in these regions by introducing spectrally-coded directional cues. These spectral cues then work in collaboration with the ITD and IAD differences between each pair of left/left-surround, and right/right-surround channels. This is evident with the proposed system by alternately muting and unmuting the front channels (L,C,R) while listening to images intended for the side areas. There is some significant semblance of an image there, but adding the front 3 channels reinforces the image.

- The typical horizontal layout of the 5 loudspeakers in surround-sound cannot really deliver a vertical impression or overhead images. It was found with this system that the binaural signals could resolve some overhead sounds especially those that are not lacking in high frequency content. (see section 4.1.2)

3.3.1 Similar ideas

There have been found some similar research ideas to those presented in this proposal. Burkhard, Bray *et al.* (1991) and Xiang *et al.* (1993) proposed a production technique using an artificial head recording replayed over 4 loudspeakers, one pair in the front and one in the rear. This idea came before the realization of discrete (5.1) multichannel reproduction as we know it today. The same left/right binaural signals were sent to both pairs of loudspeakers with the rear, "at slightly reduced level". The artificial-head was the Aachen

Head[1] which also has neck and shoulder simulation. As part of their summary they concluded, "This procedure allows binaural reproduction with approximately comparable results when using headphones with regard to localization and spaciousness". This method should be questioned in terms of phase-shift comb-filtering and image-shifting that should be audible during slight forward/backward motion of the listener due to each side front and rear loudspeakers reproducing exactly the same signal.

Mori *et al.* (1979)[2] also proposed the reproduction of artificial-head signal over 2 pairs of loudspeakers, front and rear. One difference was that they used 2 separate heads spaced close together, one behind the other and acoustically-separated by a baffle.[3] The front head fed its signals to the front loudspeakers and the rear to the second pair behind the listener. This was a discrete 4-channel system which they called "Q-Biphonic", in reference to the development of quadraphonic equipment around that time period. They concluded:

> "For full azimuthal localization it is necessary to install reproducing system loudspeakers also to the rear thus forming a quadraphonic system....a stable sideward sound-image localization can be obtained since the pairs of forward and rearward loudspeakers function substantially equally with respect to the sideward image."

[1] This artificial head model was briefly discussed in section 1.2.3.

[2] The authors were assigned to JVC Corporation of Japan.

[3] No more details were given of the spacing between heads or size and composition of the baffle separator.

Don and Carolyn Davis (1989) also proposed binaural signal playback over 4 loudspeakers. Similar to the Burkhard idea discussed above, they split each ear signal to 2 same-side loudspeakers simply wired in parallel. The first pair was placed at ±45° and the second pair, to the sides at ±90°. The side loudspeakers were placed (low) directly on the floor to help avoid some strong, delayed reflections from that boundary. The authors seem to imply that the side loudspeakers are considered the main transmitters of the binaural signals by their statement; "The two front loudspeakers are there merely to provide help in hearing signals forward of the listener". They also seem to be relying on the listener's head to provide contralateral crosstalk cancellation although they do not discuss it any further than the following statement: "Many experimenters have tried to duplicate electronically with 2 loudspeakers the exact encoding and decoding that this system provides acoustically, namely cancellation of the contralateral crosstalk components." The other significant difference is their recommendation that the binaural source recordings not be made from a dummy-head, but via miniature probe microphones inserted into the ears. (Hence the name of their system, "In-the Ear Recording and Pinna Acoustic Response Playback").

3.4 Monitoring Equipment and Environment

Since the beginning of this research, there have been a few variations of monitoring environments. Eventually, the listening (control) room has evolved into the setup as described below.

3.4.1 Description of the control room

The room itself is rectangular-shaped with the following dimensions:
Length = 5.1 meters
Width = 3.25 meters
Height = 3.05 meters.

The side walls are completely covered with sound absorbing material composed of (7.5 cm.) semi-rigid fiberglass sheets with fabric on top. This absorbing material is mostly effective for mid and high-frequencies. The floor is covered with carpet. It is important that the listening environment not be too reverberant as this can lead to localization imaging blur. (Although some producers may rely on the "liveness" of a listening room to give added spaciousness due to the decrease of interaural cross-correlation - especially for the surrounds).

3.4.2 Description of the equipment and layout

The power amplifier (model AV-6000) is made by B&K Components. It is a 6-channel amplifier with individual non-discrete continuous level control for each channel.

All 5 monitors were identical, 2-way loudspeakers made by Focus Audio called, Signature Series Model 68. The woofer diameter is 11.5 cm. and the tweeter is 3.2 cm. with the cross-over frequency at 2.7 kHz. (For more detailed specifications, see figure #13).

The placement of the loudspeakers follows the recommendations specified by the ITU-IRT[1] and IAMM (International Alliance for Multichannel Music). The 5 loudspeakers are essentially placed on a circle with the center being the reference. All 5 loudspeakers are equidistant from the central listening position at 1.7 meters distance. The front (left and right) loudspeakers are placed at ±30° - as in a stereo reference layout. The rear surrounds are at ±110°.[2]

The 1.15-metre[3] height of the loudspeakers are about average ear-level (when seated).

[1] See the document, ITU-R REC.775 at the website <www.ebu.ch>.

[2] As explained in section 3.1.2, this position relative to the listener happens to provide the greatest interaural differences - which benefit this project by allowing greater crosstalk cancellation.

[3] as measured at the center of the tweeter.

CHAPTER 4 IMPLEMENTATION AND PERFORMANCE

4.1 IMAGING OF SONIC OBJECTS

Within this section is a discussion of the experimental recordings made using the 5-channel microphone array in order to test its spatial imaging qualities. The spatial soundfield is isolated into 3 dimensions; horizontal, medial and distance; plus a 4th being the motion rendering capability.

4.1.1 Horizontal imaging

Horizontal imaging capability is the 1st-order requirement of any spatial audio display system. With music and film applications, virtually all of the main "action" takes place in front of the observer.[1] A common aesthetic is to relax the demands on imaging in the areas surrounding the listener where ambience and reverberation is presented as an extension/enhancement of the direct sound. However, it is felt here that in order to faithfully encode a sonic event, attention to spatial resolution is equally important in all directions, not only the frontal area. In this way, if the spatial resolution is precise in the sides and rear, then we can believe that the reverberant and ambient sounds are encoded well.

[1] "Spatial music" is a notable exception (which is discussed in section 5.1).

To test for horizontal imaging resolution, direct sounds were used because they are a more precise reference to identify than reflections and reverberation.

A test recording of horizontal positions dividing a complete circle was made through the 5-channel microphone array.

The horizontal area surrounding a supposed listener was mapped out corresponding to 12 clock-points. Each point was assigned an alias letter designation ("A" to "Z"). The more difficult positions were repeated 3 times for redundancy check assigning a different letter each time. These points corresponded to clock positions: 2, 3, 6, 8, 10 and 11 o'clock. The 5 and 7 o'clock positions were repeated twice; 12, 1, 4 and 9 o'clock were only sampled one time. The sound stimulus was male voice speaking in a average-loud voice, "point A,...point A" , (for example) before moving on to another position.

3 different rooms were used to make separate recordings; a large concert hall (average Rt60 = 2.3 seconds), a large shoebox-shaped performance hall (average Rt60 = 3 seconds), and an acoustically dead, small studio room. This variance of room types was chosen to check for localization acuity under different reverberant conditions.

There were 3 test subjects who participated in this experiment, none of whom reported any hearing defects. The playback test was conducted in the control room as described in section 3.4. The subjects were asked to write down the position of the alias letter as they perceived it. The perceived positions were notated on a template sheet as shown in figure #14. The subjects were not restricted from movement but were informed that it was best to keep forward facing for consistency.

The results are shown in scatter-plot form of figures #15-17.

These tests show promising results. There is no significant variance in imaging performance between the 3 room types as the error deviations are small and show no real pattern.

Noteworthy was the fact that there were no front-back reversals which are a common problem with binaural methods. The 3 frontal free-field microphones provide a distinguishing component that locks the front image in place.

Imaging beyond ±30° showed only minor errors. The maximum error margin resulted in an auditory image shift towards an adjacent point on the circle. For a finer and more accurate resolution, it may be required to present sound samples of longer duration. (In music, typically the sound elements exist longer than a few seconds so that the listener can learn and "appreciate" the intended position). Usually, localization performance improves with an increase in signal duration.

In summary, it is evident that auditory images were formed corresponding to locations other than the loudspeakers and, that these images showed small deviations from their intended positions.

4.1.2 Medial Images - test of vertical localization

As covered in the previous section (4.1.1), spatial imaging in the horizontal plane is of primary importance in any audio display system. In music for example, virtually all sounds are emitted from instruments on a plane (almost) equal level to the listener.[1] But to approach a more complete representation of a performance space, the vertical dimension should also be considered.

In many performance spaces, most of the reverberant sound energy is perceived overhead. It should be realized that the ceiling is the largest of reflecting room boundaries and is often the most easily accessible surface to the direct sound making it the most effective reflector. It is mostly recognized as contributing to loudness and distinctiveness of the sound sources (Cremer & Müller 1978).

It has also been proposed that it plays a significant role in overall spatial impression, and envelopment. In their study on the effect of "upside" reflections on envelopment, Furuya, Fujimoto, Takeshima and Nakamura (1995) concluded that ceiling reflections were necessary to "feel a three-dimensional flood of sound in a room" by contributing to a "vertical spread" of the sound image. Through this, they are implying that the "utmost acoustical sensation" of a space requires a uniformity of directional distribution - not just lateral reflections.

[1] Of course there may always be the exception of electro-acoustic music performances where it is possible to suspend loudspeakers overhead of the audience.

The perception of overhead reflections and reverberation is more of a sensation than an accurate localization of sonic events. Our natural hearing ability for overhead median-plane sounds does not rely on binaural cues as such. This is because (for median-plane sounds) there are no interaural differences. So it is believed that the spectral-filtering of the pinnae is the cue that allows us to localize overhead sounds - it is essentially a monaural process.

In their research on median-plane localization, (Roffler & Butler 1968), (Gardner 1973) and (Gardner & Gardner 1973) all supported the importance of the pinna cues. A summary of their key observations is below.

To localize a sound overhead:

1. The sound itself must have a complex spectrum.

2. The sound must include frequencies above 7-kHz; the presence of lower frequencies neither assisted or hindered localization.

3. It is better if the sound spectrum is broadband rather than narrow-band.

4. The ability of localizing sounds centered around the 8-kHz and 10-kHz band approaches that of broadband sounds.

5. Localization is improved as center-frequency of noise band increases.

(Roffler & Butler 1968) and (Blauert 1996) also point out that vertical localization ability increases with familiarity of the source signal - a type of learned localization.

4.1.2a Objectives

Certain expectations about the vertical imaging capability of this system can be derived from the above research on median plane localization. It has been well observed in psychoacoustic research that human localization performance is generally much worse in the median plane (and overhead) than in the horizontal plane. It should also be noted that the above-mentioned research in median plane localization used various types of noise as test signals.

Complete and accurate vertical imaging is not expected from this system since this is not even possible in natural hearing. Unlike the research in vertical localization, this investigation is based on creating an "illusion" of a sound source position. This is due to the fact that the *intended* vertical positions do not coincide with the *actual* sound sources as there are no loudspeakers above the listener in the standard 5.1 surround-sound arrangement. As well, it is realized that the difficulty and errors which occur naturally in vertical localization can obscure the test results of this system.

The main objective of this section is to ascertain whether this system is able to resolve recorded sound events at these difficult median plane positions. A sub-objective is to discover whether certain sound stimulus types are easier to localize correctly than others (Klepko 1998).

4.1.2b Procedure

The test positions were limited to 5 positions along the median plane as shown in Fig. #18. The sound samples included 5 second samples of: male speech, 2 separate percussion instruments (shaker, sleigh bells) and 1/3-octave band filtered noise[2] played through a small 2-way loudspeaker. The center frequencies were; 250, 500, 1000, 2500, 4000, 6000, 8000, 10,000, and 12,500 Hz. Thus, the samples can be classified as natural sounds (speech, percussion), and synthetic sounds (filtered-noise). All 5-second samples were divided into 3 short, repeated phrases (or bursts). They were recorded with the 5-point microphone array from a distance of 1.5 meters. In an attempt to simulate a practical application of this system, the recording of these test stimuli was performed on a large concert hall stage.

These sound sample types and positions were randomly re-organized and re-recorded onto a test tape. The test took place in the room as described previously in section 3.4.

[2] These filtered-noise samples were taken from an audio test CD called, "Sound Check" assembled by Alan Parsons and Stephen Court.

5 subjects (with no reported hearing defects) were involved in the test. They were instructed to note down the perceived position (along the median plane) of the sound event. A sample copy of the template they used to notate the positions is shown in figure #19. The subjects were seated but not restricted from moving.

4.1.2c Results and discussion

Figures #20-23 show in graphic form a compilation of the results.

From these results it can be shown that it is possible to perceive overhead despite the absence of an actual sound source (loudspeaker) above the listener.

The high value (figure #21) for the direct overhead position was due to the noise samples which were more often localized (correctly) above. The natural sounds (i.e. male voice, shaker, bells) did much worse for the overhead sounds but scored excellent for the front and rear positions. An explanation for this might be that the natural sounds are more familiar and are "expected" to be on ground-level rather than above. The narrow-band noise bursts are synthetic sounds which are unfamiliar. These intangible sounds carry no expectations possibly allowing the listener to believe the illusion of an overhead position. But it should also be noted that the natural sound types scored the highest overall percentages over the synthetic (noise) types which is due to the good localization for the front and back positions.

As expected, the low-frequency centered noise bands were difficult to localize showing a higher percentage of errors especially for overhead. This may be partly explained by the natural hearing's difficulty with such sounds (in a room enclosure).

Also, as evident in figure #23, there was no trend towards better overhead localization with increasing center-frequency as reported by Gardner (1973).

In conclusion, this investigation helps show that the proposed system has the ability to represent or stimulate auditory events overhead the listener. This is entirely due to the pinna cues provided by the dummy-head microphone and the acoustical crosstalk minimization inherent in the surround loudspeaker placement.

4.1.3 Distance

A sense of a sonic object's *distance* is another dimension in spatial hearing which contributes to the complete localization process.

Distance can be separated into 2 different categories: *absolute* and *relative*. A sense of absolute distance implies that the listener can identify the range of a sonic object upon first hearing it, and without any other references. (For example, upon hearing an isolated sound, the listener would be able to state the precise distance). Relative distance implies comparing the sound to other sounds (or to itself) at varying distances (Mershon & Bowers 1979) (Nielsen 1993).

Distance (or depth) is a spatial attribute often overlooked in audio display systems. This is partly because a true sense of depth and distance is difficult to achieve over loudspeakers. In many recordings, we may ask whether the object *really* sounds further away, or are we just too easily accepting of the intended audio illusion simply because (for example) there is more reverberation attached to that sound.[1]

In stereophonic reproduction, the depth of field is limited by the distance of the loudspeaker pair to the listener. That is, auditory

[1] Imagine this in the context of a recording reproduced through a single, tiny loudspeaker placed off to one side.

images cannot be perceived closer than the loudspeaker positions[2] (Wöhr, Thiele, Goere & Persterer 1991) (Nielsen 1993) (Michelsen & Rubak 1997). This may also be related to the so-called "proximity-image effect" reported by Gardner (1968a) where test subjects consistently (and wrongly) perceived the sounds as coming from the nearest of a line of 5 loudspeakers .[3]

It should also be considered that sounds which are assigned to one loudspeaker seem slightly more forward-projected than the same sound as a phantom-image between the 2 loudspeakers. (This effect is even more pronounced with rear "stereo" surrounds and side images in multichannel reproduction).

Whereas cues for azimuth localization are predominantly binaural, distance cues are more often monaural. That is, we can judge the distance of a sonic event without necessarily relying on interaural differences.

A brief review of the various distance cues is necessary at this point towards an understanding of the performance of the proposed system in this respect.

[2] An illusion of sounds being closer than the loudspeaker pair can be somewhat obtained through manipulation of the signal polarity, but this method is unreliable.

[3] Perhaps this is also related to the "ventriloquism effect" (Pick, Warren & Hay 1969) and reported here in section 1.1 as well as by Begault (1994, p. 84).

Distance cues can be separated into 5 different categories:

1. Loudness differences
2. Spectral differences
3. Direct-to-reverb ratio differences
4. Binaural differences

1. A basic acoustics principle states that the sound pressure level (as received by a stationary observer) will fall 6 dB for every doubling of distance. This predictable rate of attenuation is valid only for anechoic environments and is often referred to as the "inverse-square law". This change in intensity that manifests itself as a loudness cue is the most relied upon for distance judgements (Coleman 1963) (Ashmead, LeRoy & Odom 1990) even though this reliance may result in error.

It is a relative cue otherwise it would require a high level of familiarity with the sound source (Mershon & King 1975).

It is also a monaural cue since we can perceive loudness differences with one ear (or one loudspeaker).

2. Changes in distance cause 2 types of spectral changes. At great distances there is a high frequency attenuation due to absorption properties of air. But, as a sound source moves closer, its loudness increases which causes us to perceive an accentuation of bass

frequencies as is evident by the equal-loudness contours (a.k.a. Fletcher-Munson curves). This low-frequency boost can be further accentuated when using directional pressure-gradient microphones which cause an increase in bass as distance decreases (known as the "proximity effect").

This is a relative cue otherwise, simple alterations/distortions in the frequency response of the audio signal path can result in misjudgement of distance. It is also a monaural cue.

3. A change in distance causes a change in the relationship between the direct and its reflected sounds. With increasing distance, the intensity of the direct sound diminishes as the reverberant energy remains essentially constant (Mershon & King 1975) (Wöhr, Thiele, Goere & Persterer 1991).

As a listener (or microphone) is moved increasingly outside the room-radius, there is a corresponding sense of increasing distance (or separation) as the reflected energy begins to dominate the direct sound.

Changes in the relationship between the 1st-order reflections and direct sound occur with distance. That is, the level and time[4] both become more integrated and fused between the reflection and its

[4] The (ITDG) Initial Time Delay Gap (Beranek 1996) becomes smaller as distance increases.

direct sound source (Kendall, Martins & Wilde 1990). For example, the 1st-order floor reflection appears at a relatively lower amplitude and time delay when the receiver is up close to the direct sound source. As the source is moved away, the level of the floor reflection and direct sound become more equal and, the time of arrival difference is shortened causing a more fused image of the 2 soundfield components. This is related to the so-called precedence effect. As the time delay gap between the direct sound and first reflections widens, it becomes easier to segregate the 2 soundfield components.

Since most people spend a greater proportion of time indoors, it stands to reason that they are better accustomed to localization judgements within room enclosures. It can be equally considered that the inclusion of room reflections both aids and hinders distance judgements. It aids by providing additional cues (of a different dimension) to loudness. It hinders by complicating and obscuring the direct wavefront's otherwise predictable minus-6-dB per distance doubling inverse-square law.

Room-related cues to distance are monaural and mostly relative. They can be absolute if the listener is highly familiar with the room environment.

4. It is still unclear whether binaural differences are used by the psychoacoustic system to discern distance - however there is

objective evidence that changes in distance, cause changes in interaural signals (Hirsch 1968) (Greene 1968) (Molino 1973) (Cochran, Throop & Simpson 1967) (Mershon & Bowers 1979) (Nielsen 1993).

Binaural differences are particularly significant when the sound source is at very close proximity where the separation between the ears and the size of the head become significant in relation to the source distance. At greater distances, the relative interaural-time (ITD) and interaural amplitude (IAD) differences are negligible. It might be assumed that since sounds that arrive from ±90° result in the greatest ITD and IAD values, distance judgement would have the strongest possible cue available from binaural processing. However, the research shows conflicting reports on whether this factor of localization ability is enhanced dependent on direction angle (Cochran, Throop & Simpson 1968) (Gardner 1969) (Holt & Thurlow 1969).

However for sources arriving from around 0°, another form of binaural difference comes into play with changes in the Inter-Aural Cross Correlation (IACC). As the sound source recedes into the distance, the binaural differences are less and the direct sound signals arriving at both ears becomes more correlated. At very close proximity, the head itself will impose appreciable differences resulting in a lower IACC value. As well, IACC would lessen as the room reflections would seem to come from a wider angle in relation to the sound source and receiver positions.[5]

4.1.3a Procedure

To test the microphone array's ability to encode distance, sample recordings were made on the stage floor area of a concert hall.

3 different sound types were recorded including male speech, sleigh bells and a large orchestral bass drum. There were 3 sets of samples corresponding to 0°, 270° and 180°. Per set, each sound type was recorded at 3 equally-spaced distances (1.8 meters) apart. The samples were about 10 seconds duration, repeated twice with a short gap (2 seconds) between. The voice sample counted from 1 to 10.[6] The sleigh bells were continuously shook for 10 seconds, and the bass drum was struck by hand. Care was taken to ensure even dynamics throughout (mezzo-forte).

Actual instruments (and voice) were chosen over loudspeaker reproduction so that the complex directional patterns of the acoustical sounds would be preserved. This is important especially since this localization "test" is not taking place in an anechoic room, but will rely on room-related cues. The benefit of loudspeaker reproduction would be to maintain repeatability of dynamic

[5] This factor should be designed into the DSP algorithms of artificial reverberation software to enhance the engineer's ability to control any sound element's distance.

[6] I decided to "perform" the speech samples myself so that one of the sound types would be reasonably familiar to the test subjects.

expression so that it is less of a variable. However, it was reasoned that this would be somewhat unnatural where microphones would record sounds played by loudspeakers, again recorded by microphones (5-channel array) and reproduced over loudspeakers before reaching the test subject's ear.

Each 3-sample set was organized onto a test tape according to azimuth - 3 voice sample positions at 0°, 3 sleigh bell positions at 0° and 3 bass drum positions at 0°. Then the same for 270° and then 180°.

The playback of the test tape was held in the control room described in section 3.4.

4 subjects were used with no reported hearing defects. The subjects were informed that the test tape always presented the closest (#1) position first as a reference, then followed by the further 2 positions in any order. They were asked to identify the position of the 2nd and 3rd sample only (in reference to the 1st) and indicate on a paper template their choices. (The template is shown in figure #24). Each sound sample type set was repeated a second time only. In the event of any uncertainty, they were asked to avoid marking a guessed choice.

Once an azimuth set was complete, the test proceeded to the next azimuth. This sequence was: 0°, 180°, 270°. There was no level compensation for the different azimuths on playback.

4.1.3b Results and discussion

The results are as follows:

Out of an individual maximum possible correct score of 9:

subject "C" scored 7

subject "D" scored 5

subject "K" scored 8

subject "L" scored 8

total correct = 28 out of 36 (or 77.8 %)

These overall subject scores (77.8%) show that it may be possible to perceive changes in distance (or relative distance) to a more favorable degree.

Subject "D" misinterpreted the test procedure somewhat by anticipating that some of the samples may have been repeated. As a result, the 4 errors out of 9 of his responses were the result of his believing that the sound samples were (or could be) at the same distance position.

Organizing the number of correct responses from the 4 test subjects differently:

	0°	270°	180°		
	FRONT	SIDE	REAR		
sleigh bells	4	3	4	=	11 correct out of 12
bass drum	3	3	4	=	10 correct out of 12
male speech	4	3	0	=	7 correct out of 12
total =	11	9	8	*correct out of 12*	

The results show that the distance of the sleigh bells sound-type was the most correctly identified with the speech sample being the last. It is not surprising for the sleigh bells to score highest due to the predominance of high frequency energy in its spectral content. (Generally for any type of localization task, the ability increases as the sound's high frequency content increases).

Oddly, speech ranked below the bass drum for correct responses. This is due to an isolated factor of the total failure (0%) for this sound type at the rear positions. The other 2 angles (0° and 270°) scored almost perfectly.

Looking at the results in terms of azimuth shows that 0° (in front) provided an almost perfect performance. This is despite the expectation that distance localization for the side directions provide maximum binaural differences - hence a stronger set of distance cues than any other direction.

The rear position scored the lowest mostly due again to the isolated poor performance for the speech sample from that direction.

The subjects were asked for any comments on the test experience. All of them pointed out (in some way) that they relied on the change in room response as a cue for distance more than loudness level. It may be significant in that the samples were actually recorded in a concert hall with considerable reverberant energy (with an average Rt60 of 2.3 seconds). This may have caused the direct-to-reverb cue to supercede the loudness cue. This may not have been the case in a room with a much drier acoustic.

4.1.4 Moving sound sources

One obvious advantage of surround-sound audio presentation is the ability to simulate movement of sounds around the listener. This ability is certainly exploited in film presentations with "fly-bys" of aircraft and spacecraft.

Chowning (1971) developed a computer program to simulate motion over 4 loudspeakers in a sort of quadraphonic (4-corner) configuration. The program manipulated spatial cues of intensity differences, doppler effect and reverberation changes. This system was unique (at the time) in that it was intended for new music applications allowing the composer the option of motional gestures.

Despite Chowning's work, methods to simulate movement via audio displays have not really evolved much past the practice of active intensity panning of the signals in the intended directions.

In addition, the study of sound sources in motion has not been researched nearly as much as stationary sound source localization. (This is despite the fact that we exist in a constantly mobile environment where sound source objects *and* listeners are often in motion. Of course, the traditional performance practice of music rarely finds the musicians (or audience) in motion.[1])

[1] This is discussed further in section 5.1.

The understanding of moving sound source localization can be adapted from what is known about static sound source localization. A sonic object in motion results in a continuous set of changing binaural cues through intensity, phase and time differences. Motional localization ability is the result of sequential processing of these modulating cues.

The cues that are relevant to the localization of moving sounds are:

1. interaural temporal differences (binaural cue)

2. interaural intensity differences (binaural cue)

3. pitch modulation (Doppler effect)[2] (monaural cue)

4. continuous spectral changes due to pinna filtering (this can be a monaural cue)

5. intensity change as the sound changes in relative distance (monaural cue)

6. change in reverberant properties of the auditory image (binaural or monaural cue)

The most notable references on moving sound source research are (Harris & Sergeant 1971) (Perrot & Musicant 1977, 1981) (Grantham 1986) (Rosenblum, Carello & Pastore 1987) (Perrot & Tucker 1988) (Strybel, Manligas & Perrot 1989, 1992).

[2] As a sound source moves towards a single point of observation (i.e. a listener or microphone), there is a shortening of the wavelength which causes a gradual rise in pitch. As the sound moves away, there is an accompanying fall in pitch as the wavelength increases. Typical manifestations of this effect are found in a moving train's whistle blow, or a car horn sounding as it passes by.

The keynote work by Mills (1958) produced the concept of the "Minimum Audible Angle" (commonly called MAA) which the above research evolved from. The MAA is an index for a JND (Just Noticeable Difference) in sequential stationary source position. This concept was adapted into a new spatial index called the "Minimum Audible Movement Angle" (MAMA) which measures the minimum angle of travel required for detection of the direction of sound movement.

A general summary of the salient conclusions from the above research activity is as follows:

- MAMA has a direct relationship with the velocity of sound object travel. The faster the object travels, the greater the minimum audible movement angle.

- For sound stimuli which is impulsive in nature (i.e. non-continuous), shorter durations result in an increased value of MAMA.

- MAMA was the least at 0° azimuth. It gradually increased towards ±60° and grew sharply beyond that towards the highest values at ±80° to ±90°. In other words, localization performance deteriorated as the sonic object moved to the sides of the listener. This is the same result as Mill's MAA findings for stationary sonic objects.

- There is no specialized motion detection mechanism. The auditory system must rely on successive spatial changes and apply the same processing of cues as for stationary sound objects.

Klipsch (1960b) points out an additional consideration when recording sonic objects in rapid motion. A spatial distortion problem involving the doppler effect can arise when using 2 (or more) spaced-apart microphones to record the movement. It was observed while recording a passing train with 2 microphones spaced 15 meters apart that there were 2 doppler effect events. It is as if each microphone is a separate "observer" perceiving the doppler effect at 2 different times. (The extra-wide microphone spacing was used to exaggerate the effect - but the same principle applies to a relatively lesser degree with closer separations). What this means is that the recording engineer should be cautioned against using widely-spaced microphones when trying to capture the sound of objects in rapid motion.[3]

There are some examples included on the (optional) audio tape which demonstrate the effect of moving sound sources. These are:

1. pilot-study walkaround (speech while walking around the array)
2. backyard sound effects (flying birds and airplanes)

[3] The doppler effect may be a more subtle pitch-shifting of the overtone components of a sound source.

3. coin rolling (across the concert hall stage floor)
4. audience anticipation (members walking about, talking etc.)
5. pre-choir rehearsal (members walking about, talking etc.)

4.1.4a Procedure

In addition to those audio examples that feature sonic objects in motion, some tests were recorded to specifically exercise the motional rendering of this 5-channel microphone array. The intentions behind this test were to verify whether a motion path could be perceived without hesitation, and whether certain paths were more difficult to perceive.

On the stage area of a concert hall, 3 different sound types were recorded while in motion. These were male speech, percussion shaker, and sleigh bells. The shaker and sleigh bells were both shook continuously, while the speech counted out aloud discrete consecutive numbers. Each segment lasted approximately 10 seconds. 3 different trajectory paths were recorded in either direction: across the front, along the right side and, across the rear. The distance of the "line" from the array was about 2.5 meters. Diagonal paths were not attempted due to the awkwardness of passing through the microphone setup.

It was realized that the actual performance could result in a somewhat inconsistent sound. The alternate (and most common hearing research) approach is to use sounds emitted from a loudspeaker which can be controlled for consistency. However, there are serious mechanical difficulties in putting a loudspeaker in motion. As well, any such mechanical system could introduce extraneous noises which would contaminate the purity of the intended signal.[4] Further reasoning towards using actual sounds instead of loudspeaker-reproduced sounds is that this is the type of sound source that the microphone-array is intended to record naturally. (Hearing research that uses loudspeakers usually interface the sounds directly to the participating subject - rather than having a microphone encoding stage in between).

The 9 different (3 sounds x 3 paths) motion samples were randomly reorganized onto a test tape. The total test duration was 4 minutes. Each sample was only heard once (with a 10-second gap between them) so that the subject's response would be their immediate, absolute judgement.

There were 4 test subjects of which none had any known hearing defects. They were asked to simply draw an arrow-line in the path of their perceived sound motion. The template they used is shown in figure #25. They were instructed to skip to the next segment in the event of uncertainty.

[4] While performing these sounds, shoes were removed and every effort was made to reduce extraneous walking noises.

4.1.4b Results and discussion

The results show that all 4 subjects were able to perceive the motion paths with 100% success. Perhaps this perfect rate does not indicate enough about the continuity of motion rendering. That is, the subjects may simply be applying a series of static localization cues to make a guess at the direction and azimuth - perhaps even only being able to localize the start and end points as a minimum.

Comments by the participants indicate that the frontal path was easiest to perceive, with the rear as second and the side-path being the least clear.

This observation is in agreement with research on MAMA which shows that it is greatest at the sides - that is, less localization resolution occurs naturally at the sides (Harris & Sergaent 1971) (Grantham 1986) (Strybel, Manligas & Perrot 1992).

The explanation for the superiority of the frontal region results is due to our greater natural localization acuity there (Mills 1958), and the fact that there are 3 loudspeakers to render this area instead of 2 as is the case for the sides and rear.

The more lucid motion path to the rear over the sides might be due to the fact that the surround loudspeakers are positioned symmetrically about the listener's head allowing binaural cues to be maintained. The loudspeakers rendering the side images cannot fully utilize such interaural differences.

Due to the prescribed spacing of the microphones, it is also expected that there would be slight distortion of the doppler effect (Klipsch 1960b) especially for the side images. This could be perceived at some conscious level to degrade the sensation of natural motion.

CHAPTER 5 - Applications

5.1 Music - spatial compositions

One of the more potentially compelling applications for any surround-sound system is the recording of music composed with a specific, pre-designed spatial organization. Although the performance realization of many of these works was designed to be a unique, live experience, there remains the need to document them through recording. Multichannel surround-sound has the potential (where stereophonic sound fails) to capture and reproduce these works with the intended spatial element intact.

Probably the earliest examples of spatialized music can be found in liturgical music with its processional and recessional movement of Gregorian monks (while singing) that mark the start and end of a mass celebration. Elements of this custom are still practiced today.

Existing stereophonic recordings of this event certainly fail in providing a sense of the chant entering from and departing to the rear of a church. The sound would appear to come and go from some distance in front of the listener.

Surround-sound practices which use many microphones distributed far apart from each other would also fail here. Engineers often place ambient microphones in the rear of a hall which are fed principally to the surround channels. With this method, reproduction of the processional (for example) would result in the voices entering from the back, but eventually coming near, then far again as they advance towards the front. This would obviously produce the wrong spatial perspective. A more unified position of all the microphones such as the system presented in this research, would stand a better chance of imitating the movement past the listener similar to the actual event.

The Renaissance period also introduced the idea of *coro spezzato* (broken choir) through its church music. Here, the choir would be divided into 2 to 4 distinct groups dispersed throughout the cathedral. The separate choirs could be positioned on balconies above and behind the listener resulting in an "other-worldly" or "heavenly" sensation. This spatial separation of the choir termed "polychoral", was most notably developed and exploited by Giovanni Gabrieli (c. 1557 - 1612) to further promote his four, five and six-part polyphonic writing. [1]

As music entered the so-called Classical and Romantic periods, it gradually became more of a concert event. As such, music was composed with the background intent that the performer(s) would be positioned on a separate stage area in front of an audience, much like theatrical performances. Commercial interests favored this spatial arrangement where the audience could clearly see (and hear) the performers.

[1] Polychoral music can also refer to the responsorial and antiphonal singing of Gregorian chant where 2 distinct groups (or soloist and choir) alternately sing in dialogue. However, there is no extraordinary separation or placement of the singers here that could benefit from a surround-sound rendering.

It was not until the late nineteenth century that composers began again to explore the spatial distribution of musicians. Gustav Mahler's use of off-stage instruments not only had a dramatic effect, but a symbolic one as well where brass and percussion instruments formed an off-stage "military" band playing fanfares over the lyrical string orchestra. Mahler maximized the differentiation of the two groups by introducing spatial separation as well as the more conventional means of contrasting timbres and rhythmic patterns.

Charles Ives also contributed to spatialized music inspired by his youth experiences where he observed 2 separate marching bands converging during town parades. He was struck by the odd effect produced by the gradual convergence (and divergence) of the two streams of music that had no relation in tempo or tonality. Although he composed many works based on this idea, *The Unanswered Question* (1908) was the only piece that fully explored the spatial dimension. There are three distinct bodies of sound; a string orchestra, a group of four flutes, and a solo trumpet which are all intended to be separated in the performance hall. Although the piece requires two conductors, there is no intended rhythmic coordination except for approximate entrance points. The following quote from Ives himself characterizes his sense of awareness of the possibility of using spatial separation in music;

> "Experiments, even on a limited scale, as when a conductor separates a chorus from the orchestra or places a choir off-stage or in a remote part of the hall, seem to indicate that there are possibilities in this matter that may benefit the presentation of the music, not only from

> the standpoint of clarifying the harmonic, rhythmic, thematic material, etc., but of bringing the inner content to a deeper realization (assuming for arguments' sake, that there is an inner content)." (Cowell 1933)

Ives' work and ideas had a profound effect on composer Henry Brant. Brant composed music as if he were conducting experiments focussing specifically on spatial elements.[2] He composed pieces that purposely explored various modes of space and the problems associated with it such as: vertical height, "walls of sound", rhythmic coordination, distance, sound movement, performers' movement, harmonic blend ("spill"[3]), and "outdoor music".

Brant, like Ives, observed the possible benefits of spatial separation on harmonic and rhythmic clarity:

> "The total impression of spatially distributed music, in its clarity of effect and in the special kind of relationships produced, is to some extent equivalent to setting up the performers close together on stage, as usual, but writing the music in such a way that each texture remains in its own octave range, with no collision or crossing of textures permitted. The spatial procedure, however, permits a greatly expanding overall complexity, since separated and contrasting textures may be superimposed freely over the same octave range, irrespective of passing unisons thus formed, with no loss of clarity." (Brant 1967, p. 225)

[2] His , "Space as an Essential Aspect of Music Composition" (Brant 1967), is a culminative report of his findings.

[3] Brant's use of the term, "spill" is also often used by British recording engineer/producers when referring to the sound of an instrument being picked up by a microphone intended for another instrument thereby reducing separation. Also known as "leakage".

This same idea can be observed from a perceptual point-of-view where the separation of sound sources (instruments) can relieve dissonance (if so desired). Albert Bregman notes that:

> "Since we know that spatial separation is a basis for segregating signals we should be able to suppress the dissonance between two tones by placing their sources in quite different spatial locations". (Bregman 1990, p.522)

In brief, it is our auditory perceptual biases[4] that can cause spatial distributions to be more effective through harmonic delineation and signal segregation. But these same perceptual biases can conflict with certain expectations of instrumental, contrapuntal, and textural balance. For instance, assuming equal distances from a listener, a certain sound source (or instrument) will produce different loudness sensations according to its position. In general, sound sources placed above, behind or to the sides of the listener will appear less loud than those positions from the front. So the composer must understand and anticipate the auditory perspective when arranging the music both contrapuntally and spatially, so as to avoid being misled by an idealized vision which can exist in the imagination and physical realm, but not in the perceptual realm.

A typical surround-sound playback environment has the ability to translate these particular balance perspectives since the loudspeaker

[4] Our visual perception is an entirely different modality that, when exposed to spatially distributed performers, would evoke more a sense of theatrical enhancement and/or symbolic associations than our auditory perception would. (Schornick 1984, pp. 2-3)

arrangement provides sources (real and virtual) around the lateral plane of the listener. In effect, it is a type of *spatial quantization* down to 5 positions which represent the recorded sound field.[5]

Increased spatial separation leads to increased temporal separation due to the relatively slow speed that sound travels. This may result in another performance practice problem where rhythmic accuracy can be compromised due to extreme physical separation of musicians (often with separate conductors).

Brant chose to look at this problem as an opportunity to free composers from the constraints of strict rhythmic ensemble when he proposed an "...idea of non-coordinated, but in-its-essential-parts-controlled, rhythm..." as a new, alternate means of expression.

> "Whether the spatial separation is small or great, it will cause a deterioration in the rhythmic coordination (keeping together) of the separated groups, and hence in their harmonic connection....
> However, it can be turned into a unique advantage if the composer will go one step further, and plan his music in such a way that no exact rhythmic correspondence is intended between the separate groups". (Brant 1967, pp. 233-4)

Composer Harry Somers also arrived at the same idea:

[5] To date, there have been other proposals for more than 5 channels (plus one LFE channel) that allow more surround and overhead loudspeakers.

> asymmetry that will result from the time lags and which would be impossible to notate accurately. The impossibility, or at least extreme difficulty, of synchronization, is one of the strongest considerations with regard to the nature of the musical materials and their application". (Somers 1963, p.27)

If a feature of the spatial music is to have such a rhythmic freedom and aleatoric synchronization then, the recording engineer/producers should be cautioned against close multi-micing since this could destroy the special timing relationships. A more unified microphone grouping could provide better results allowing something closer to a singular point-of-view (as an ideally located listener or conductor would hear) with the different sound source transit times arriving relatively intact.

Another category of spatial music is that which is performed in an outdoor environment. Henry Brant's proposal for "outdoor music" was finally realized through many compositions (or "soundscapes") by Canadian, R. Murray Schaefer beginning in the late 1970's. This is an extreme form of spatial music where musicians perform in an outdoor environment such as a wooded lake, a park or a city hall square. A side benefit of such an open environment is the allowance for extreme separation of the performers involved, which would not be possible in most indoor performance halls. Imagine a successful surround sound recording of an outdoor performance of spatial music; it would be an interesting challenge to try and transport the listener (in their home environment) to a performance out on a wilderness lake for example.

Composers, performers, and recording engineers should be aware of the possible variants in sound production unique to an outdoor environment (Embleton 1996). The propagation of sound is strongly affected by temperature, temperature gradients, humidity and wind. Shadow zones, upwards and downward refraction (bending) and erratic absorption are some of the unique sound altering phenomena that occur outdoors. These effects are mostly unpredictable making the performance of music outdoors a risky venture. Add to that, the common problems that face the sound engineer when using microphones outdoors; wind turbulence can cause extreme rumble noise, and humidity can cause condenser microphones to produce spurious noises or even fail temporarily.

Incorporating motion of sound sources might seem to be more the domain of electro-acoustic music composers since they can easily mobilize sound around a listener by distributing the sound around an array of loudspeakers[6] (Chowning 1978).

With the exception of the processional/recessional sung parts of a liturgical mass, actually having active mobile performers would usually have a strong theatrical connotation, which could be a distraction from the music.[7] Composer, Roger Reynolds puts it another

[6] Pierre Schaeffer, the French composer largely responsible for "musique concrete", coined the term "trajectoires sonores" (sonic trajectories) to denote the spatial movement of sound. Pierre Boulez referred to the circular movement of sound as "space glissandi".

[7] Of course, opera music features movement of the singers, but a theatrical element is definitely intended.

way: "Our experience with sound sources in motion around us is too easily associated with insects, airplanes, and wind" (Reynolds 1978).

An interesting and subtle way of creating a sense of motion with actual performers is through dynamic and timbral shifting. One way to accomplish this is to have stationary performers distributed around an audience stagger the playing of identical pitches and chords. The resulting effect simulates motion as the notes and timbres gradually displace from one instrument (group) to another. Karlheinz Stockhausen used this effect with the three displaced orchestras in *Gruppen für Drei Orchester* (1955-57).

"Répons" (1991-1988), by Pierre Boulez is considered among the masterpieces of spatialized music (Harley 1994). The title refers to the responsorial singing part of a Christian mass where a soloist is "answered" by a choir. In Répons, there are 6 instrumental soloists that alternate with an ensemble of 24 musicians. The spatial aspect is realized in live performance where the audience surrounds the ensemble, and both are surrounded by the 6 soloists and 4 loudspeakers (1988 version) that amplify and redistribute the sound of the solo instruments. Circular motion of sounds is articulated through amplitude modulation techniques controlled by computer where the speed of motion is dependent upon the amplitude envelope of any particular note event (Boulez & Gerzso 1988).

The idea of spatial movement of sounds can be taken further through the intent to evoke or outline geometric shapes. The outline of a circle surrounding an audience would be an obvious choice as well as a square, triangle, or star shape. Somewhat more ambitious applications of this idea are those which intend to outline a specific, more complex architectural shape. Two such examples are John Tavener's *Ultimos Rios* (1972) which outline the shape of a cross, and Charles Hoag's *Trombonehenge* (1980) featuring thirty trombones surrounding the audience with the shape of the ruins at Stonehenge. However, it would seem that only a very specific limited listening area would allow the proper appreciation of this effect.

5.1a Tape Examples

The optional demonstration tape contains many examples of music recorded through the 5-point microphone array.

Excerpt #15 is a recording of a John Cage composition, "4^3", for 4 musicians playing rainsticks, 2 pianos, and a sine tone oscillator. This is a spatial composition intended to have the musicians separated from each other. Here, one musician playing the piano and rainstick is located several meters behind the microphone array.

Excerpts # 4 - 8 feature a percussion ensemble, organ 6-voice choir, SATB choir with organ, and a jazz quintet. The percussion ensemble (#4), the 6-voice choir (#6), and the jazz quintet (#8) were spatially distributed around the microphone array.

5.2 Sound effects

Surround-sound in general, offers a much expanded spatial "canvas" with which to portray sound effects. With the proposed microphone system, the sound designer is able to compose and capture more natural soundscapes, and effects. These may be considered "extra-musical" effects within a music composition or, applied within a dramatic production/performance.

The proposed 5-channel microphone array has an advantage through its ability to render overhead sounds as well as motional (both horizontally and vertically) sounds. Music is usually performed from the front "stage" area by immobile performers/instruments, but motional and vertical sound effects can be produced if desired.

Using this system to record an environment as a background can be quite effective. This could be a fully spatial-textured ambient background where music or drama can be superimposed upon (using simple mono microphones that would give contrast for a more clarified and intelligible result). The background would be reproduced as more of a continuous field of sound whereas other microphone techniques might produce a collection of distinct images (at the loudspeaker positions). The intent would be to avoid having the 5 loudspeakers "call attention" to themselves in the same way that proponents of using diffuse dipole loudspeakers (for the surrounds) claim. But using dipoles would discourage a continuous, equal field of sounds since they would differ from the front (LCR) loudspeakers in terms of clarity and timbre.

5.2a Tape examples

The tape examples feature an outdoor as well as some indoor ambiences.

The outdoor ambience (excerpt #12) is a recording made in the backyard of a house in the early morning. This captures an early rush hour din reverberating from a distance as a general backdrop. The sound of some individual cars starting and leaving from a closer distance can be heard. Most striking is the sound of the birds chirping above in the trees with the occasional overhead fly-by where the rapid flutter of the wings can almost be felt. A few small-engine airplanes can also be heard passing above.

Indoor ambiences include an audience in anticipation of a concert (excerpt #9); and a choir preparing for rehearsal (excerpt #13). Both of these were recorded in the same large performance hall space[1] and feature effective motion effects as people walk around and pass the artificial head. The conversation that takes place nearby and in the distance, demonstrates the systems capability of rendering depth.

[1] Redpath Hall at McGill University.

A pilot-study recording (excerpt #1) effectively displays the capability of the system for complete surround spatial rendering. Here, a "walkaround" is featured while speaking and playing different percussion instruments (in typically difficult to perceive locations). Other sounds to notice are footsteps, a door slamming shut and overhead sounds. The whispered speech phrases made close to the ear of the artificial-head microphone are not as effective through loudspeakers - they sound at the distance of the loudspeakers themselves and not the intended close proximity.

A more stark indoor ambience is presented in (excerpt #10). This is a “tapping-in” of a vocal performance preparation featuring a solo voice and vocal coach giving instruction while playing piano. The images are very solid especially the vocal coach who is positioned at around 60°. (The singer is at around -10°. The room is a small-medium performance hall where the strong reflections off of the rear windows can be perceived quite effectively.

(Excerpt #2) shows 2 passes of a coin rolling across an empty concert hall stage floor. The first trajectory outlines a large circle around the head; the second rolls across the back.

The final sound effects example is (excerpt #11). This sample was made in an apartment with wooden floors and a typical ceiling height of 2.5 meters. The interesting effect here is a sense of the room boundaries surrounding the listener. The early reflections strongly reinforce the acoustical identity as the room is activated by the sound of a small radio turning on and off briefly.

5.3 Multitracking-overdubbing

The 5-point microphone array can be viewed as a *minimalist* approach using one microphone for every channel of a surround-sound system - in effect, a live-to-5-track recording. This is not unlike the common practice of using a single pair of microphones to capture a sonic event for 2-channel stereo reproduction. If applied well, the results can be striking in terms of the overall purity of sound and, realistic acoustic perspective. However, problems can arise when relying on these minimalist approaches to record a multi-instrumental performance.

1. The inter-instrument balance may be difficult to achieve properly in one unified performance or setting.

2. The desired spatial layout of the instruments in the surround-sound representation may be compromised due to physical difficulties of situating the musicians without their getting in each other's way.

3. A poor performance by any of the musicians would ruin the combined take in spite of a potentially excellent performance by the others.

4. Independence of the different parts could not be maintained in the event that at a later production stage, certain instrumental parts may be omitted temporarily or completely.

Separate mic'ing through overdubs and/or close mic'ing is the typical method chosen to overcome these problems. However, this approach lacks the purity and sense of integral acoustic space that minimalist approaches can offer.

A novel solution to this conflict is to use overdubbing techniques to build the total musical arrangement through layering of individual instruments, 5-tracks at a time as recorded through the microphone array. This hybrid approach can solve all of the performance problems as well as the difficulties in achieving a controlled audio perspective as mentioned above, while maintaining the sonic virtues afforded by a minimalist technique.

A room with the appropriate acoustic qualities should be chosen where the microphone array would be set up and never moved throughout the entire overdubbing process. In this way, the qualities of the room will be preserved at every layer and the relationship of the room surfaces to the microphone array will be constant throughout. With such an approach, balances, perspectives, spatial positions and depth can be controlled by simple placement of each musician with respect to the array. As well, individual instrumental performances can be optimized through re-takes and insert recording (i.e. "punch-ins").

Each layer would consist of a set of 5 channels as recorded by the microphone array. Any layer could consist of more than one instrument if desired. For example, a common format of 24-tracks would allow 4 layers (4x5) as maximum.

In the end, all of the sets are mixed down with the fader levels equal.[1] All left channels will be combined to the left-total track on the mixdown recorder, all right channels to the right-total track, and so on.

One limiting factor with this method is that there would be only one room ambience. Although this is the element that binds and unifies the musical performance as if it occurred at the same time.

5.3a Tape experiment

A short, pilot-study recording was made using this idea and is included in the sample tape (excerpt #3).

The first step in the process was to record an electric bass-guitar part onto one track of a 24-track recorder.[2] This served as the foundation of the arrangement which was then monitored through headphones while overdubbing the subsequent instrumental layers. The layers were as follows: layer 1, Steinway 9-foot grand piano;

[1] Minor level deviations may be made for the sake of a desired balance.
[2] A Sony DASH-S 24-track digital open-reel recorder.

layer 2, drum kit; layer 3, shaker and congas; layer 4, bass part reproduced through a powered loudspeaker, melodica , and triangle.

The instruments were spatially distributed around the array in order to test the imaging quality and stability at different areas of the surround-sound reproduction field.

The overall effect is that of being engulfed by the instruments as if in the center of the activity. The results show good imaging of all instruments especially in the difficult zones at around 65° and 290° where the congas and piano were positioned respectively. The drum kit image (directly to the rear) is solid with a realistic sense of size corresponding to that proximity. The sensation of the separate performances taking placing in a single room environment is obtained.

5.4 Reverb Chamber

The proposed 5-channel microphone array is intended to be used to capture the complete sonic event with the spatial properties of the direct, early and late soundfields intact. However, it can also be used to add reverberation to dry signals that (may) have been recorded by close microphones. This can be accomplished simply by placing the microphone array in a reverberant room which also has a loudspeaker feeding the direct dry signal of an instrument or voice into the room. The microphones will pick up the reflected sounds and feed these signals back to a mixing console where it is mixed together with the original dry sound.[1] It is important that the microphones pick up as little direct sound from the loudspeaker as possible. This can be achieved by loudspeaker placement, or by using a dipole loudspeaker with its null oriented towards the microphone array. If the intention is to have more reflected energy from the rear, then the loudspeaker should be placed behind the array to allow it to more strongly excite the room from that location.

The novelty of this approach is to have a multichannel surround-sound representation of actual room reflections that are separately controlled from the direct sound. The proposed array has its strength

[1] This idea has its origins in the late 1950's and is commonly referred to as a reverb or echo "chamber". Usually the reverberated sound is picked up by 1 or 2 (stereo) microphones. This idea is rarely found today since it is considered impractical to devote a whole room for such a purpose. As well, there is a limited amount of control over the reverb effect unless the trouble is taken to alter the acoustical surfaces of the chamber.

in delivering a natural-sounding acoustic sensation and is well-suited in applications where a natural perspective is desired.

Another advantage of such a system is the ability to postpone decisions related to the ambience component of a mix to a later stage in the production process. One disadvantage is that it takes up 5 channels of the mixing console to return the reverberated signals.

The demonstration tape has an example (excerpt #14) where a dry close-mic'd jazz piano performance is recorded onto tracks 7 and 8. These tracks were routed to a loudspeaker in a concert hall (with Rt60=2.3 seconds) where the 5-channel microphone array was positioned on stage to pick up the reflections from the hall. To hear this effect requires 7 channels returned at a mixing console with 2 different grouped elements: the original piano sound in stereo, and the 5- channel reverb pickup from the microphone array. The 2 elements can be mixed to taste.

5.5 Dramatic productions

Surround-sound representation in general can provide a compelling audio version of a dramatic production or play (Haines and Hooker 1997). The expanded spatial territory permitted (over 2-channel stereo) through multichannel renderings can allow the listener to be more engaged in the action of the story.

Use of the 5-channel microphone array can provide a more natural, unbiased perspective for the listener. The actors would simply balance their performance around the microphone array. The performance of dialogue and sound effects would be spatially "mapped" around the array.

One limitation of this approach is that the environment (indoors or outdoors) would be static so that the performance would have to be moved to a different location in order to achieve the desired variation of acoustic perspectives.

There are no tape demonstration examples of this application. However, listening to the examples under the Sound Effects section (5.2) could provide an idea of how this might work.

5.6 Audience reactions

An audio representation of a live concert performance can be made more interesting if the sound of the audience is captured as well. Including audience reactions such as applause, cheering, and laughter can allow the listener at home to feel more involved in the event. Unlike stereo, surround-sound can spatially separate the audience from the on-stage performance and envelop the listener with audience sounds. (Stereo television productions of live talk-shows often use 3-D spatializing and reverse-phase processing to force the sound image of the audience past the boundaries of the front loudspeakers).

It was hoped that the 5-channel microphone array would faithfully capture this component of a live performance. However several trials proved unsatisfactory.

The applause seemed too overpowering so that the signal levels from the binaural head microphone would have to be reduced drastically. This would depart from the intention of equal levels throughout a recording. Spatially, the applause sounds had too much contrast of depth with nearby audience members seeming too close and clear while the rest of the audience clapping sounds like distant noise. It was also hoped that there would be a more continuous curtain of sound surrounding the listener. However, there was not a strong sensation of rear sounds, but mostly direct clapping sounds coming from the surround loudspeakers.

Also, during the performance, audience noise (movements, coughing etc.) seemed too distracting. Such noises in stereo seem less disturbing even though they come from the front. This is probably due to their being spatially masked by the sound of the music.

CHAPTER 6 SUMMARY AND CONCLUSION

6.1 Summary

This work began with an analysis of the requirements for fulfilling complete auditory images with a concentration on spatial attributes. To this end, a general review of the different forms of audio display systems available to the sound engineer/producer for rendering the spatial characteristics of a sonic event was presented. Special attention was focussed on an analysis of the strengths and weaknesses of 2 different audio display technologies: binaural and multichannel surround-sound.

Through this analysis, a unique hybrid approach was proposed that was derived from a convergence of 2 technologies, binaural and surround-sound. This approach resulted in a microphone technique primarily aimed at capturing and displaying the spatial qualities of a sonic event.

Details of the design and construction of the artificial-head microphone were discussed along with the rationale for the use of the 3-microphone technique for the frontal sounds.

After some pilot studies to arrive at the final specifications, experiments were conducted to test the spatial fidelity of the system under different demands. Several experimental recordings were made to test the usefulness and limitations of the system. From these recordings, many different potential applications were presented and discussed.

6.2 Conclusion

Audio systems have evolved to a high degree of fidelity when it comes to recording/reproducing the subtle timbral qualities of a sonic event. However, the spatial component of the sonic event places a much higher demand on audio systems which they have not been able to meet to the same degree.

Two totally different, independent approaches show promising results when representing these spatial characteristics - these are binaural, and surround-sound methods. But they are not without their weaknesses.

This work proposed, implemented and investigated a hybrid approach in the form of a 5-channel microphone array. The proposed technique overcame some of those weaknesses through the simple convergence of their strengths.

The common in-head localization (IHL) problem with binaural (headphone) reproduction is eliminated by use of external transducers (i.e. loudspeakers) which can rarely cause IHL. The surround loudspeakers (LS, RS) are truly external sources forcing external auditory images. As well, listener head movement will cause corresponding natural changes in interaural and spectral cues which would be absent in a more artificial listening environment such as headphones. The loudspeakers would also "activate" reflected energy in the listening room itself promoting a more solid and integrated reference environment.

There have been solutions to binaural recording using loudspeakers, (see section 1.2.4) but these involve complicated and difficult-to-implement crosstalk-cancellation schemes. The typical surround-loudspeaker placement results in a simple acoustical crosstalk reducer (relying on the listener's own head acting as a barrier). This is sufficient to maintain the necessary left-right (independent) spectral differences at high frequencies.

Finding a suitable equalization for headphone-reproduced binaural signals can be difficult and prohibitive. A simple equalization is applied that removes the global timbre colorations evident at all angles of incidence and allows improved reproduction over loudspeakers. This equalization combines both a free-field and diffuse-field approach. The free-field approach is inherent in the equalization to correct the gross timbral errors especially due to double-pinna filtering. The diffuse-field approach imposes a high-frequency shelf-boost to compensate for this loss over the course of random sound incidence.

The problem of front/back reversal is greatly reduced due to the loudspeakers being positioned both in front (-30°, 0°, +30°) and in the rear (+110° and -110°). This allows the establishment of more solid auditory images in both front and back quadrants. As well, using the free-field microphones for the frontal region maintains the clarity of those intended sounds as compared to the rear.

The typical high noise and relatively poor frequency response of miniature microphones often used in artificial-heads is avoided by incorporating regular-size, high-quality studio microphones in their place.

Although surround-sound reproduction is a great improvement in expanding the spatial ability over stereophonic systems, it still has great difficulty in resolving auditory images in the regions between 30° and 100° and, in the rear. This is because the 5-loudspeaker layout must rely on IAD and/or ITD cues to resolve so-called phantom images. The spatial images often lack any sense of integration to the sides and rear. In effect, surround-sound can often suffer from sounding too "discrete" with separate images at each loudspeaker position. Introducing binaural head and pinna-related signals into surround-sound can resolve images in these regions by involving spectrally-coded directional cues. These spectral cues then work in collaboration with the ITD and IAD differences between each pair of left/left-surround, and right/right-surround channels.

The typical horizontal layout of the 5 loudspeakers in surround-sound cannot really deliver a vertical impression or overhead images. It was found with this system that the binaural signals could resolve some overhead sounds especially those that are not lacking in high-frequency content.

In general, the intended use of such a microphone array is to capture the acoustical space from one point-of-view position. In many situations, the musicians must be carefully positioned around the array to achieve an effective balance and acoustical perspective. The placement is crucial as this is in essence, a "live-to-5-track" recording, not unlike trying to capture a performance to stereo with only 2 microphones. Compensating levels at a later production stage will compromise the integrity and balance of the room sound foundation. However, it would be possible to supplement the array with subtle use of close spot microphones to highlight certain instruments, especially those assigned to L,C and R channels since they would not adversely affect the binaural cues which are only assigned to the surround channels (LS, RS).

The system can be used as a subset of a larger multitrack project where only selected elements are recorded with the 5-point microphone array. Creative application of this idea could build a production where different elements co-exist but have separate acoustical settings so that the listener can hear into several spaces at once.

A complete piece can be built by overdubbing separately one (or more) musical parts at a time in place around the microphone array. Of course, this involves using 5 tracks at a time which may not be practical.

Recording live concerts may be unacceptable considering the distracting presence of the dummy-head within the microphone array. Audience reaction sounds (i.e. applause) did not reproduce very successfully in terms of spatial imaging.

The extended spatial ability of this system can enhance dramatic productions where the actors can actively move and place themselves around the microphone array where the studio floor is mapped out with position markers. This is not unlike early live radio productions or even quadraphonic experiments with this genre.

There exists a substantial body of spatial compositions (especially modern works) by composers such as Ives, Cage and Xenakis where the instruments are intended to be displaced around the audience. Conventional recordings of these works entirely miss out on the essential spatial element. Through use of this proposed technique, these compositions can be realized in the recorded medium with the intended spatial element intact.

A further refinement of the system is achieved through the intentionally designed use of similar transducers (microphones, loudspeakers) and amplifiers. In practice, the simplicity of the system would almost allow the user to set the levels, and then not really have to monitor the recording in surround sound since the matched (equal) levels of all channels should ensure proper encoding as intended.

The proposed approach is downward compatible to stereo although there will be no surround effect. Informal tests of this downsizing ability shows that the level of the binaural (surround-channel) signals should be reduced by about 3 dB, otherwise, in many cases there will be an overabundance of reverberant sound. However, stereo headphone reproduction will resolve a full surround effect due to the included binaural head-related signals.

Downsizing to Dolby matrix multichannel (5-2-4 in this case) is feasible except that it will not properly reproduce the rear binaural signal pair because of the mono surrounds. A matrix system would down-mix the 2 artificial-head signals into one mono surround channel therefore losing its binaural effect. As well, some of the fine spectral detail would be lost due to the usual bandpass filtering scheme (100 Hz - 7 kHz) of the surround channel in such matrix systems.

6.3 Further investigations

The initial research here could prompt more detailed and refined testing of the spatial imaging capabilities.

Further refinements of the equalization could lead to improved timbral characteristics while maintaining the spatial qualities.

More field recordings need to be made in order to "practice" and optimize placement of the array relative to performers and room surfaces. (As it stands now, it is still at an undeveloped stage in this regard).

The idea of using binaural head-related signals for the surrounds can be adapted for other work in spatial audio displays. The binaural signals do not have to be derived from an artificial-head microphone, but can be synthesized (through DSP) from existing HRTF databases.

More research can be conducted on integrating additional close microphones with the array. These can be to increase relative loudness of an instrument within a mix, or to add presence and detail. These signals should be subtly mixed in with the front 3 channels (L, C, R) and delayed to be in approximate synchronization with the array.

In general, this dissertation described one method of rendering spatial qualities of sound through the surround-sound medium. Many other techniques need to be developed in order to give the sound engineer/producer a broader choice for any situation.

standard stereophonic loudspeaker/listener arrangement

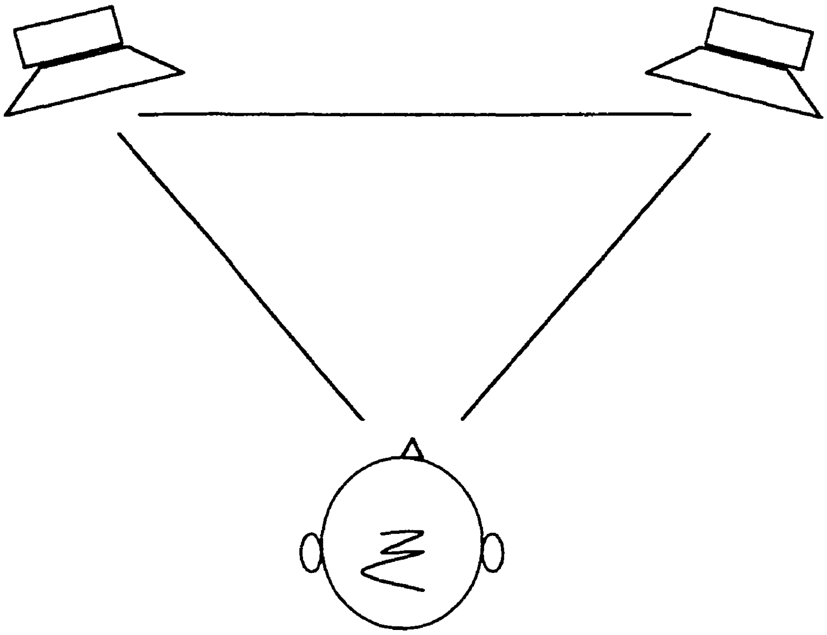
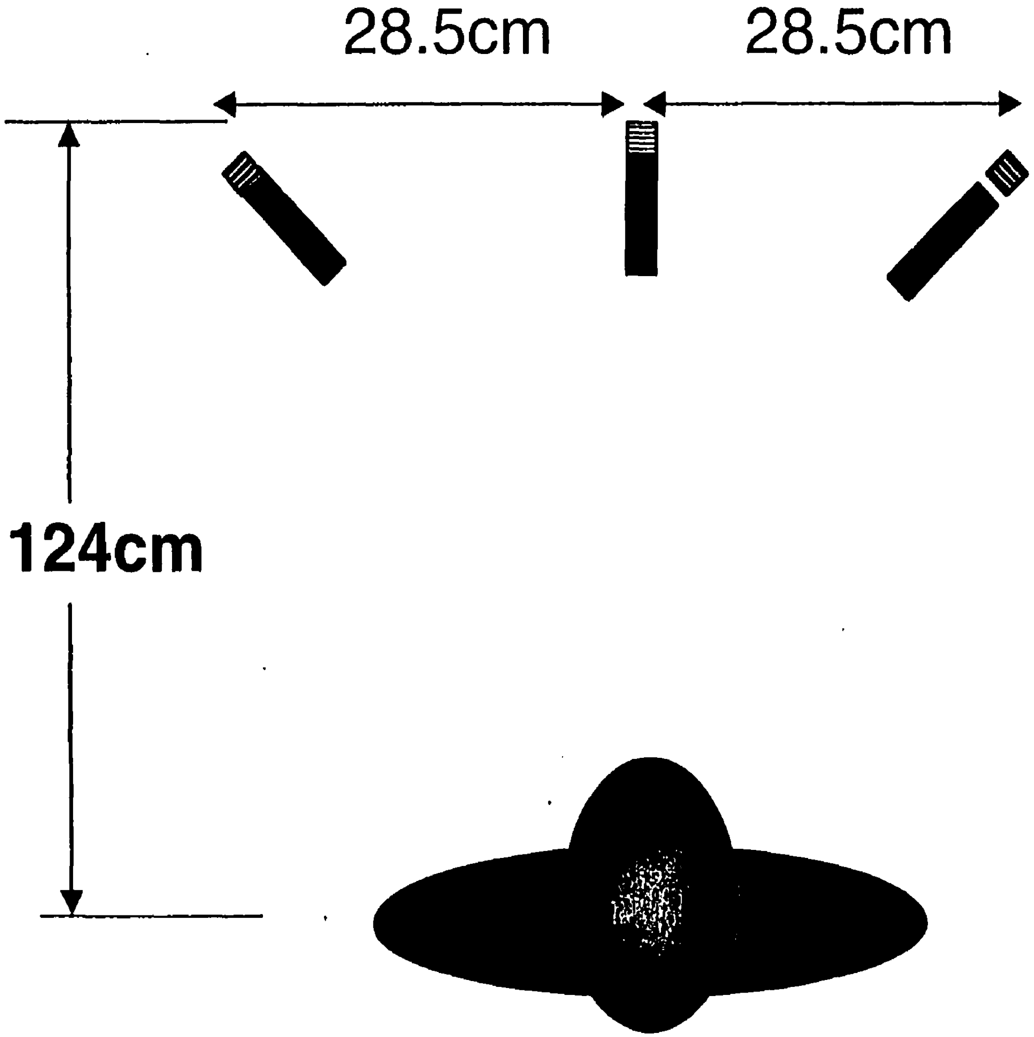


figure #1

Microphone Setup

John Klepko • Faculty of Music – McGill University

figure #2

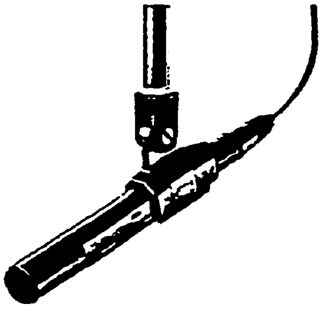


5-channel microphone array dimensions

RECORDING MICROPHONES

figure #3

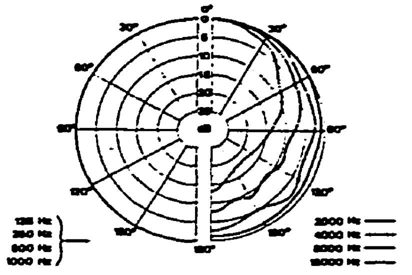
MKH 20 P48U3



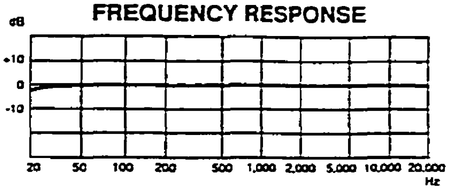
Omnidirectional

Low distortion push-pull element, transformerless RF condenser, flat frequency response, diffuse/near-field response switch (6 dB boost at 10 kHz), switchable 10 dB pad to prevent overmodulation. Handles 142 dB/ SPL. High output level. Ideal for concert, Mid-Side (M-S) acoustic strings, brass and wind instrument recording.

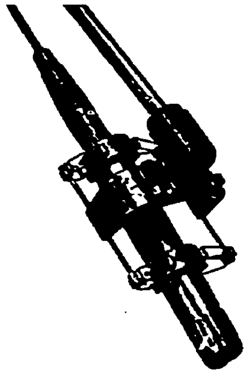
Frequency Response	20 Hz-20 kHz
Sensitivity	25 mV/Pa (-32 dBV)
Rated Impedance	150 Ohms
Minimum Load Impedance	1000 Ohms
Equivalent Noise Level	10 dBA/20 dB CCIR468
Signal-to-Noise Ratio	84 dBA
Maximum SPL	134 dB (142 dB)
Maximum Output Voltage	2.5 V
Phantom Power Supply	48 V
Plug	3-pin XLR
Dimensions	1 inch (25 mm) x 6.12" (153 mm)
Weight	3.5 oz. (100 g)



FREQUENCY RESPONSE



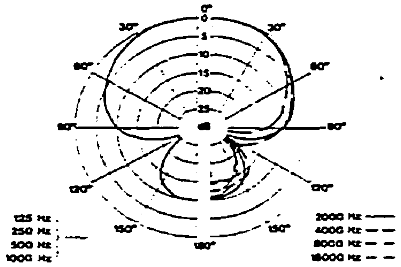
MKH 50 P48U3



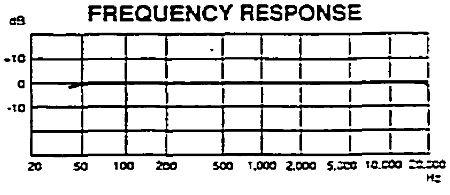
Supercardioid

RF condenser with narrower pick-up pattern than cardioid making it highly immune to feedback. Low distortion push-pull element, transformerless, high output level. Switchable proximity equalization (-4 dB at 50 Hz) and 10 dB pad. Off-axis attenuation for better isolation makes it ideal for multi-track recording, live performances, stage overheads, and orchestra pick-up.

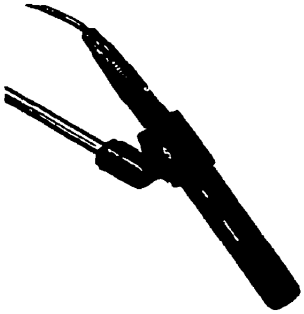
Frequency Response	40 Hz-20 kHz
Sensitivity	25 mV/Pa (-32 dBV)
Rated Impedance	150 Ohms
Minimum Load Impedance	1000 Ohms
Equivalent Noise Level	12 dBA/22 dB CCIR468
Signal-to-Noise Ratio	82 dBA
Maximum SPL	134 dB (142 dB)
Maximum Output Voltage	2.5 V
Phantom Power Supply	48 V
Plug	3-pin XLR
Dimensions	1 inch (25 mm) X 6.12" (153 mm)
Weight	3.5 oz. (100 g)



FREQUENCY RESPONSE



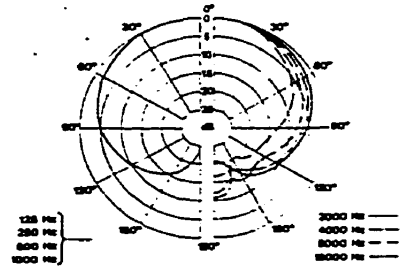
MKH 40 P48U3



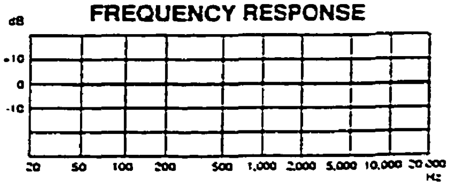
Cardioid

Highly versatile, low distortion push-pull element, transformerless RF condenser, high output level, transparent response, switchable proximity equalization (-4 dB at 50 Hz) and pre-attenuation of 10 dB to prevent overmodulation. In vocal applications excellent results have been achieved with the use of a pop screen. Recommended for most situations, digital recording, overdubbing vocals, percussive sound, acoustic guitars, piano, brass and string instruments, Mid-Side (M-S) stereo, and conventional X-Y stereo.

Frequency Response	40 Hz-20 kHz
Sensitivity	25 mV/Pa (-32 dBV)
Rated Impedance	150 Ohms
Minimum Load Impedance	1000 Ohms
Equivalent Noise Level	12 dBA/21 dB CCIR468
Signal-to-Noise Ratio	82 dBA
Maximum SPL	134 dB (142 dB)
Maximum Output Voltage	2.5 V
Phantom Power Supply	48 V
Plug	3-pin XLR
Dimensions	1 inch (25 mm) x 6.12" (153 mm)
Weight	3.5 oz. (100 g)



FREQUENCY RESPONSE



front view of binaural head

figure #4a

oblique view of binaural head

figure #4b

side view of binaural head

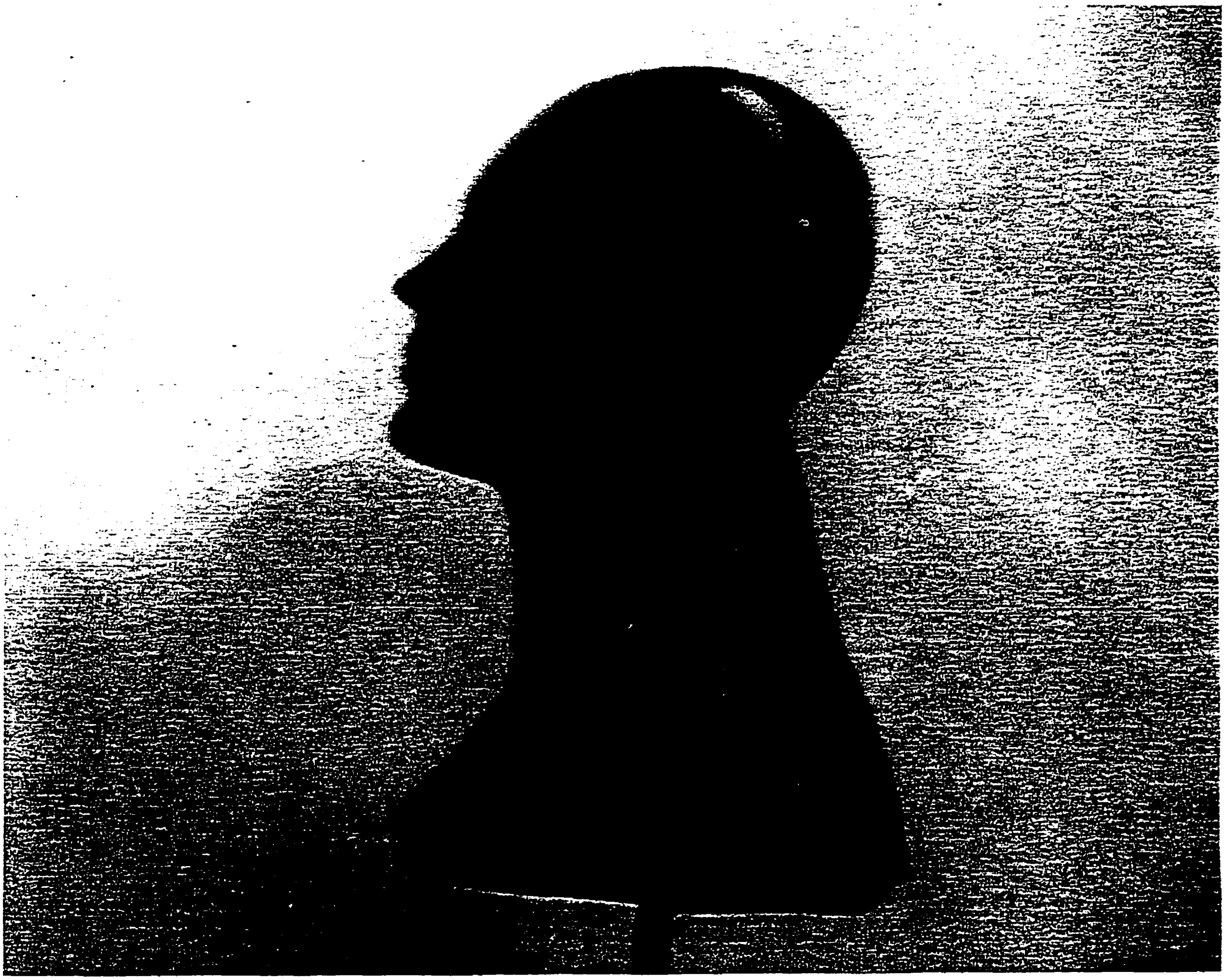
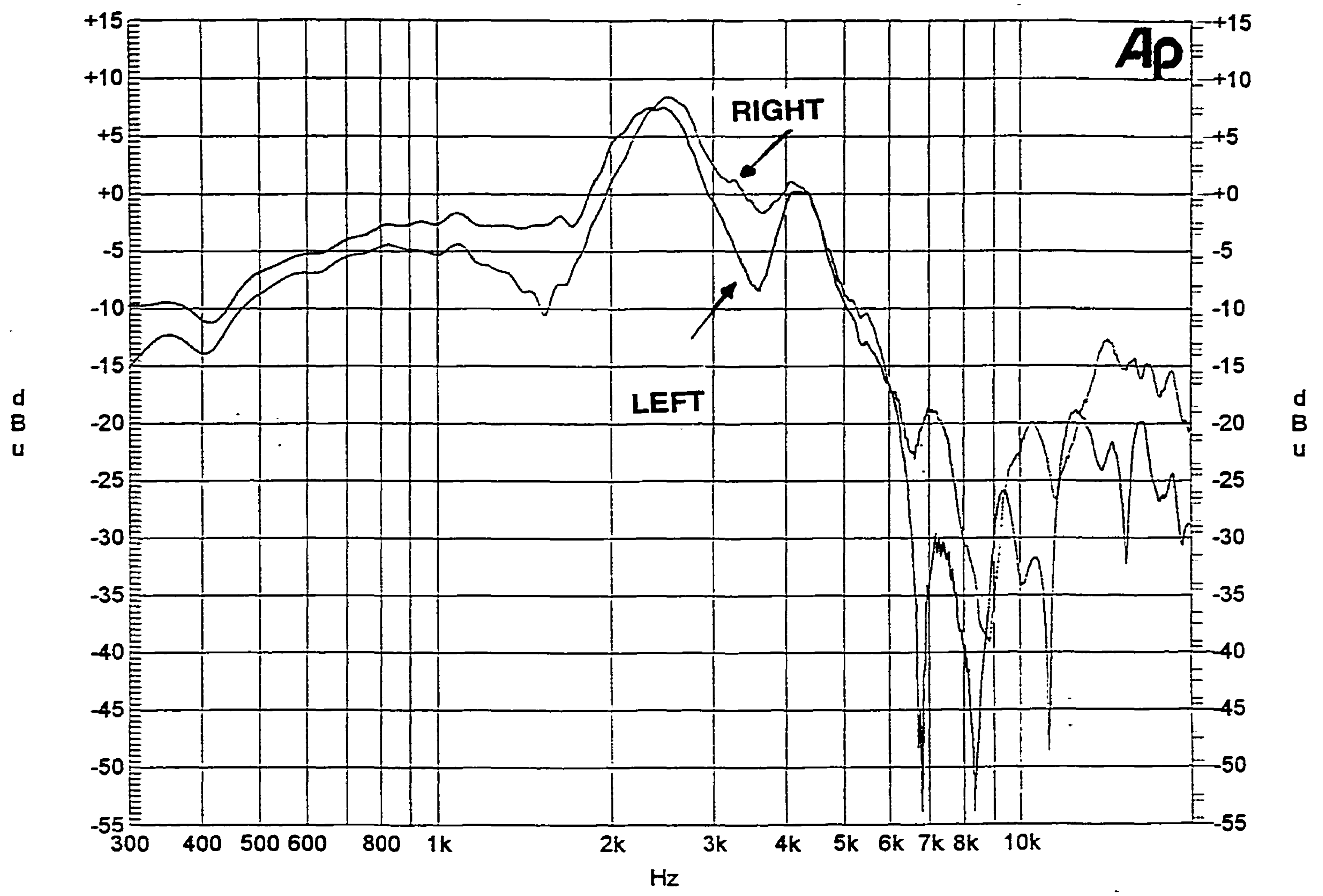


figure #4c

MLSSA measurement of binaural head at 0°

McGill University 03/29/98 17:29:36

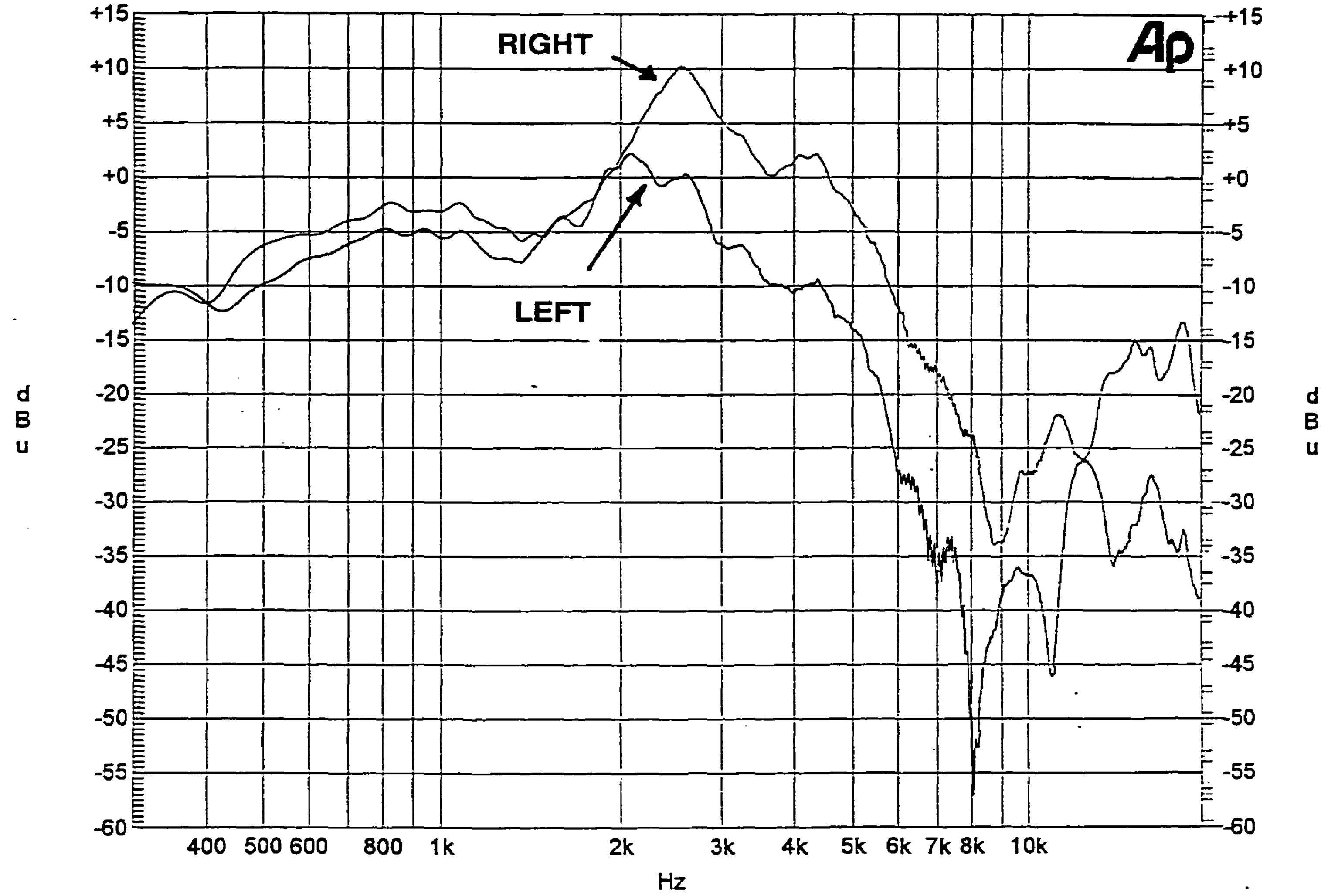


AZIMUTH = 0°

figure #5

MLSSA measurement of binaural head at 30°

McGill University 03/29/98 17:31:51



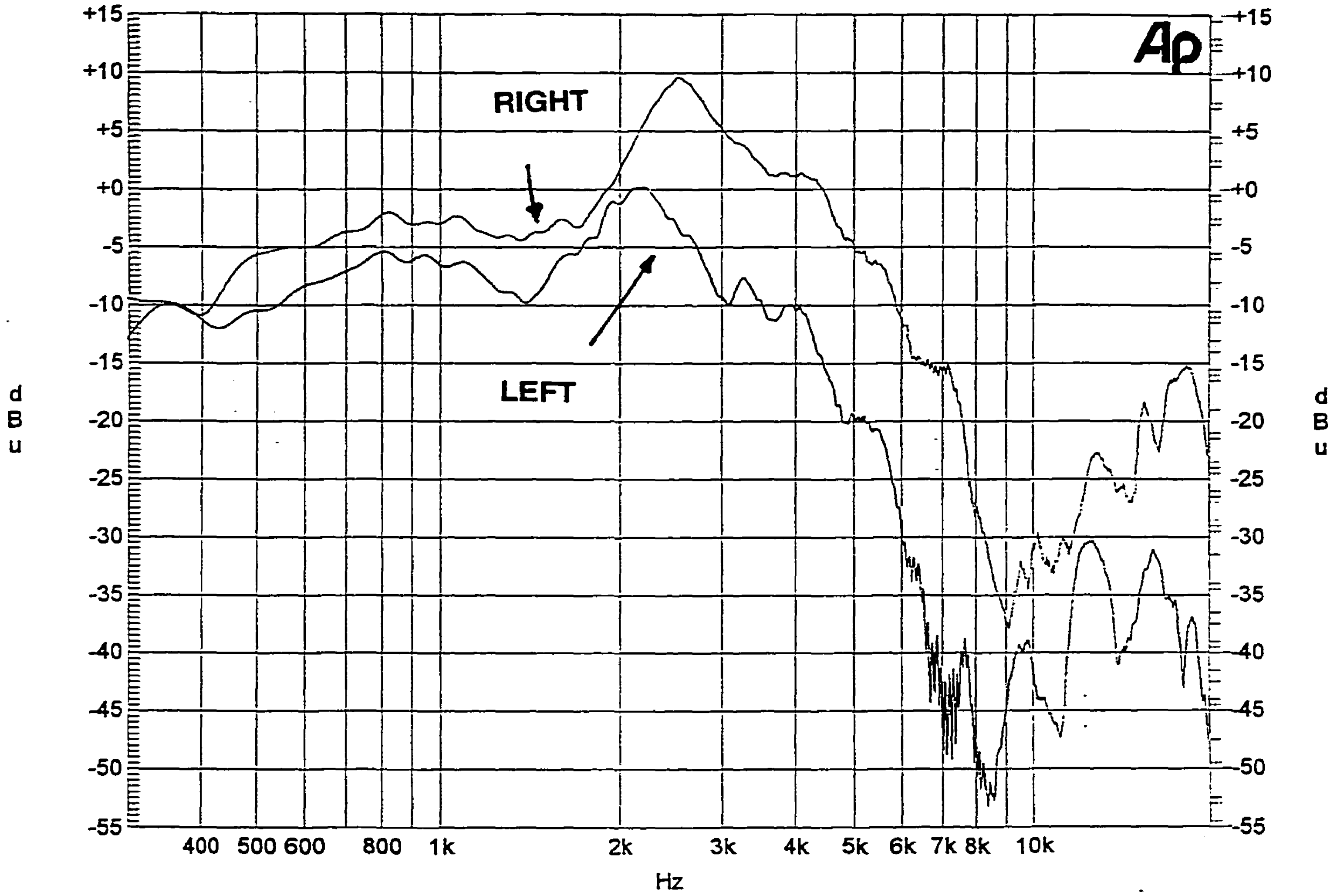
AZIMUTH = 30°

figure #6

MLSSA measurement of binaural head at 45°

McGill University

03/29/98 17:32:17

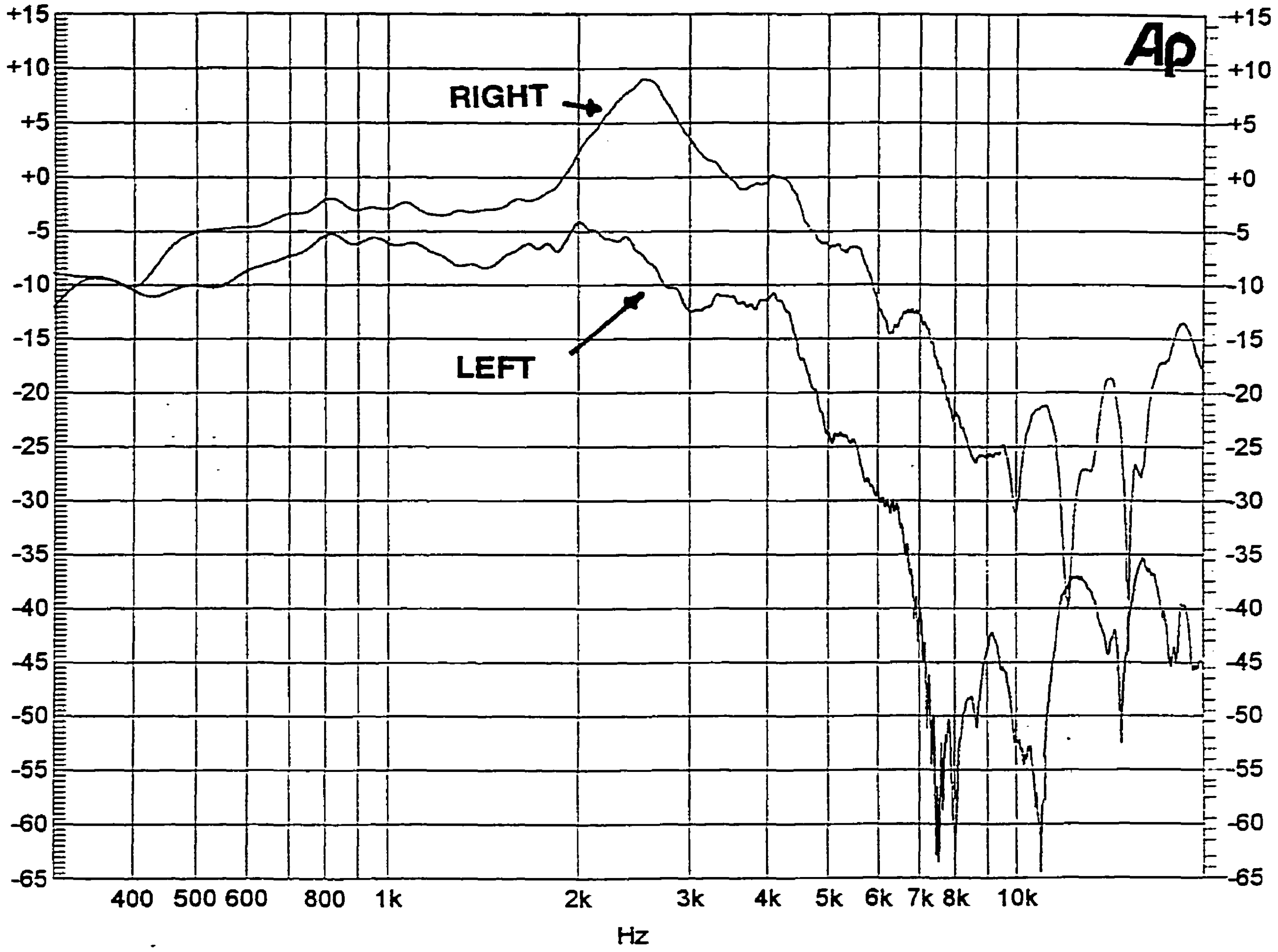


AZIMUTH = 45°

figure #7

MLSSA measurement of binaural head at 60°

McGill University 03/29/98 17:32:45

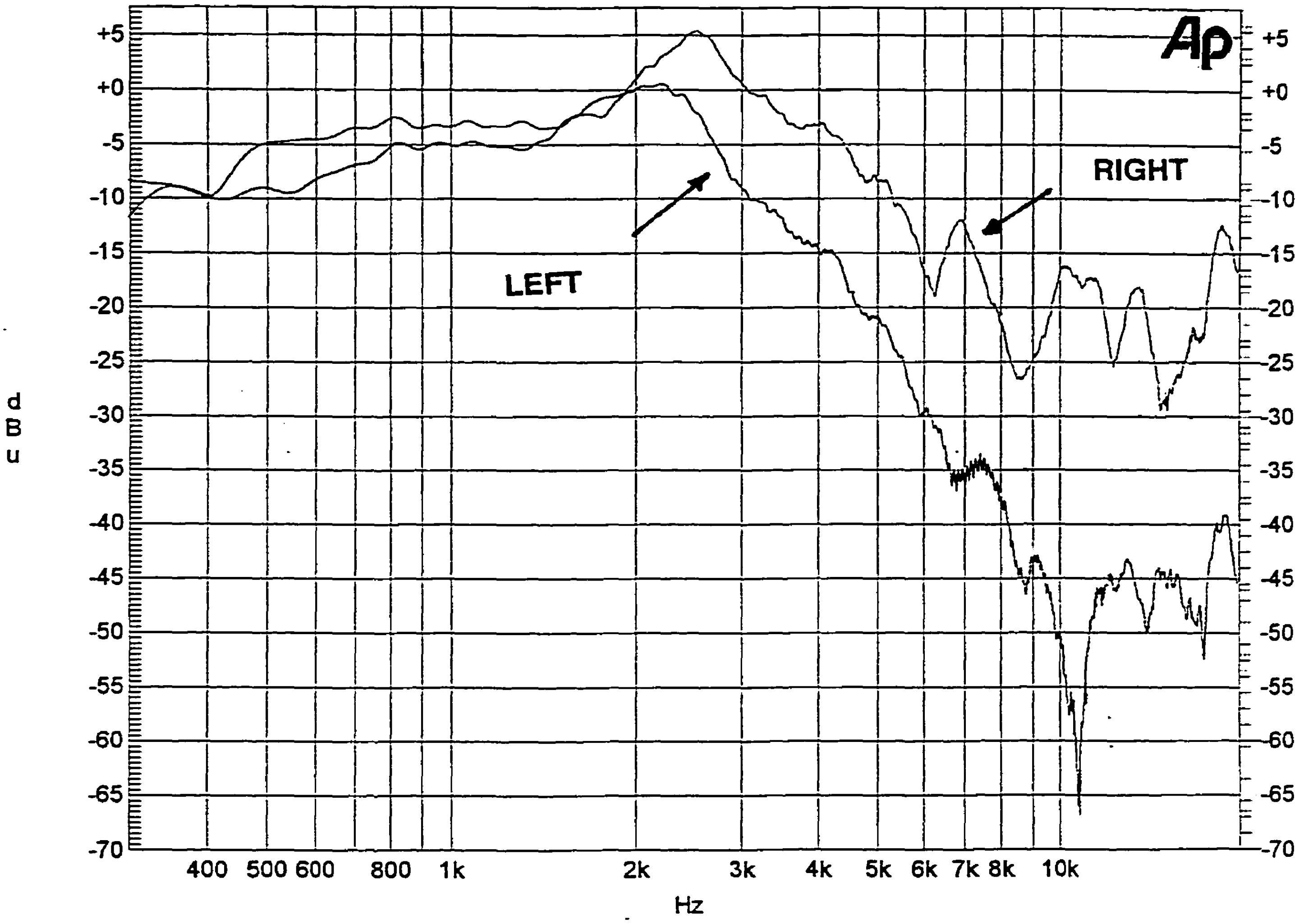


AZIMUTH = 60°

<u>**figure #8**</u>

MLSSA measurement of binaural head at 90°

McGill University 03/29/98 17:33:43

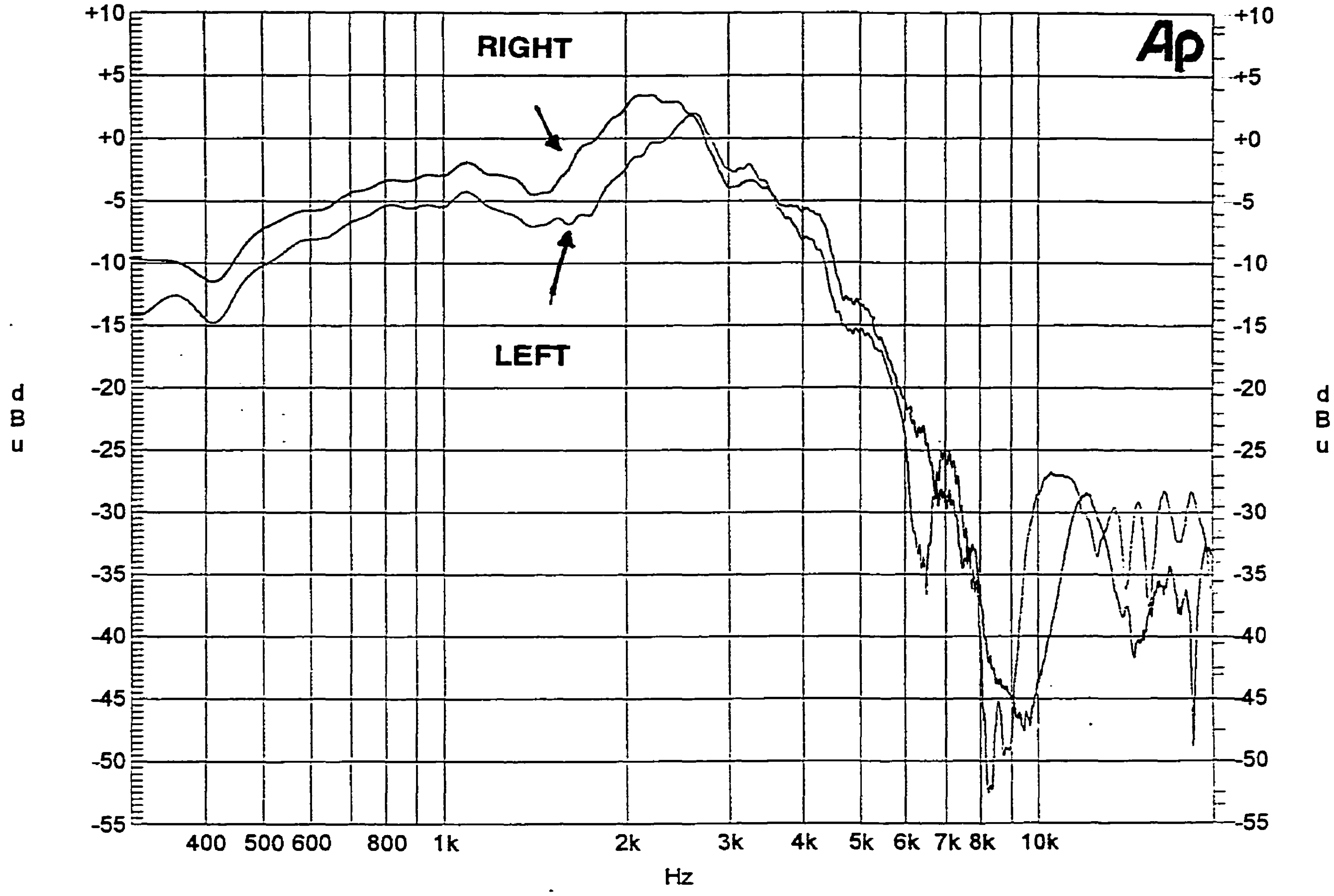


AZIMUTH = 90°

figure #9

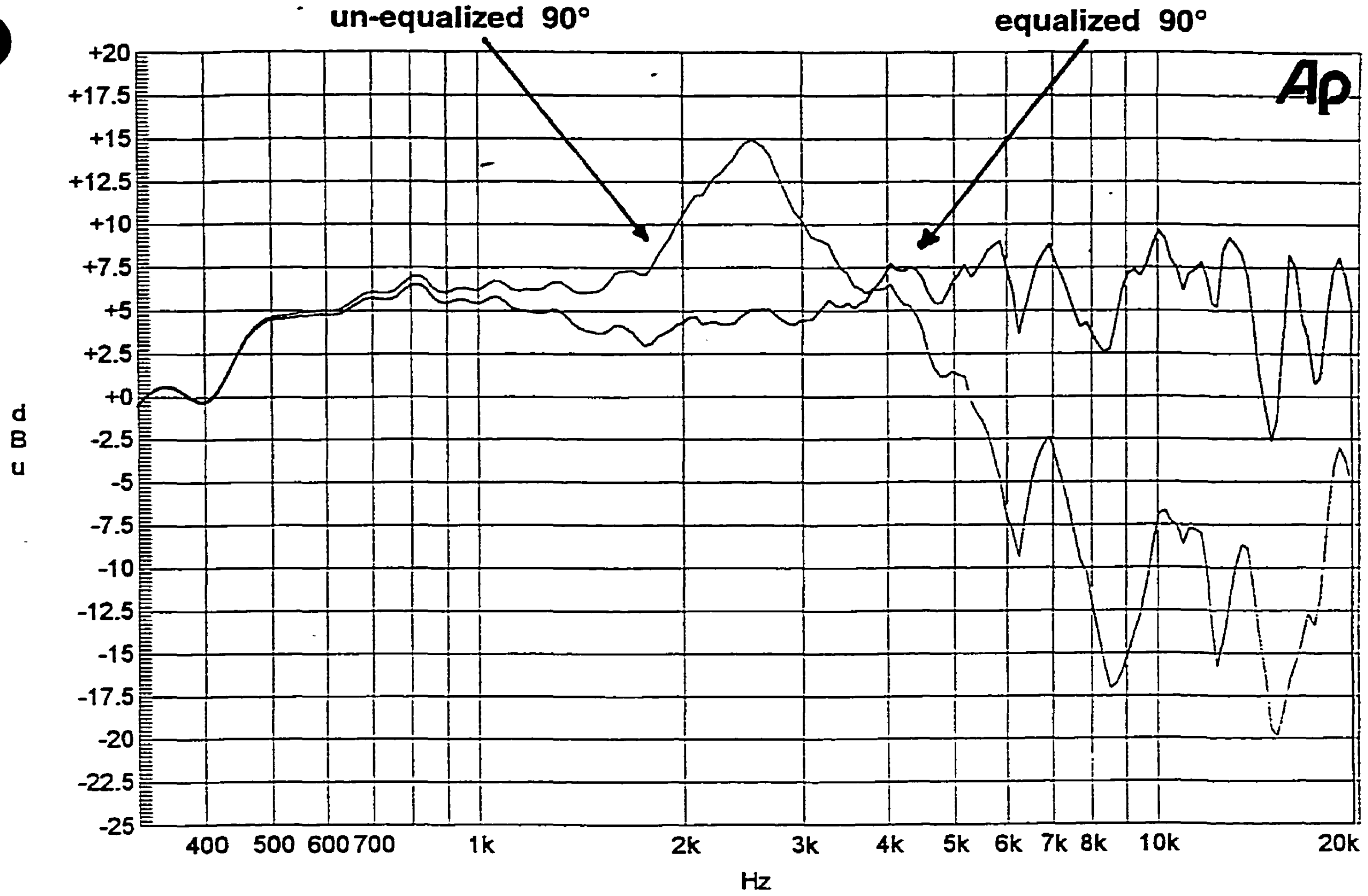
MLSSA measurement of binaural head at 180°

McGill University 03/29/98 17:34:07



AZIMUTH = 180°

figure #10



un-equalized vs. equalized measurement of binaural head (right ear)

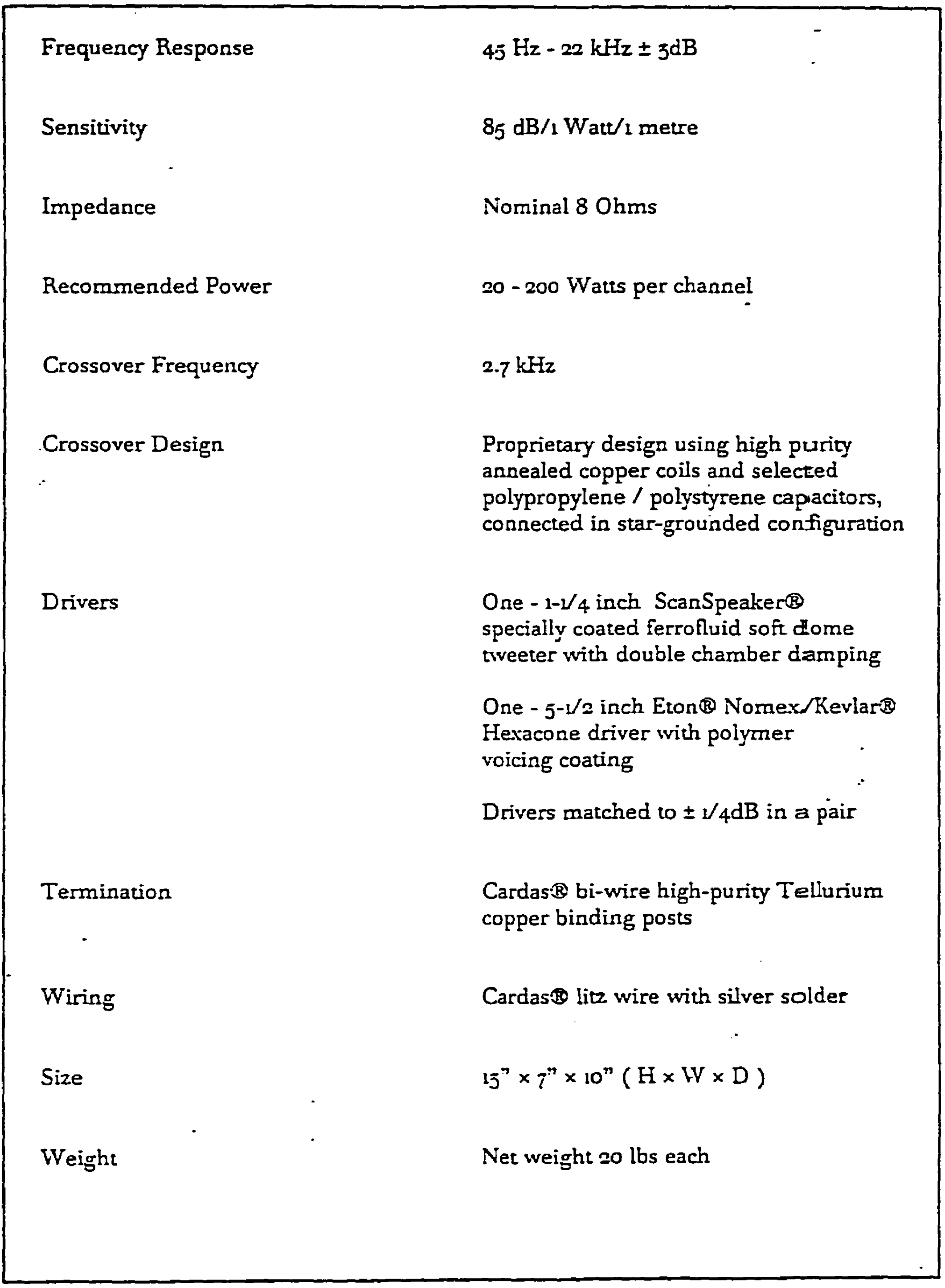
figure #11

	BAND 1	BAND 2	BAND 3
LEVEL	-12 dB	+12 dB	+8 dB
Fc	2.52 kHz	8.98 kHz	6.72 kHz
Q	3.2	7	hi-shelf

EQUALIZATION SETTINGS

figure #12

specifications of loudspeakers used in experiments



SPECIFICATIONS

Frequency Response	45 Hz - 22 kHz ± 3dB
Sensitivity	85 dB/1 Watt/1 metre
Impedance	Nominal 8 Ohms
Recommended Power	20 - 200 Watts per channel
Crossover Frequency	2.7 kHz
Crossover Design	Proprietary design using high purity annealed copper coils and selected polypropylene / polystyrene capacitors, connected in star-grounded configuration
Drivers	One - 1-1/4 inch ScanSpeaker® specially coated ferrofluid soft dome tweeter with double chamber damping One - 5-1/2 inch Eton® Nomex/Kevlar® Hexacone driver with polymer voicing coating Drivers matched to ± 1/4dB in a pair
Termination	Cardas® bi-wire high-purity Tellurium copper binding posts
Wiring	Cardas® litz wire with silver solder
Size	15" × 7" × 10" (H × W × D)
Weight	Net weight 20 lbs each

figure #13

template used in horizontal imaging tests

McGill University
Faculty of Music

INITIALS ________________ ROOM ______________________

DATE __________________

TIME ________

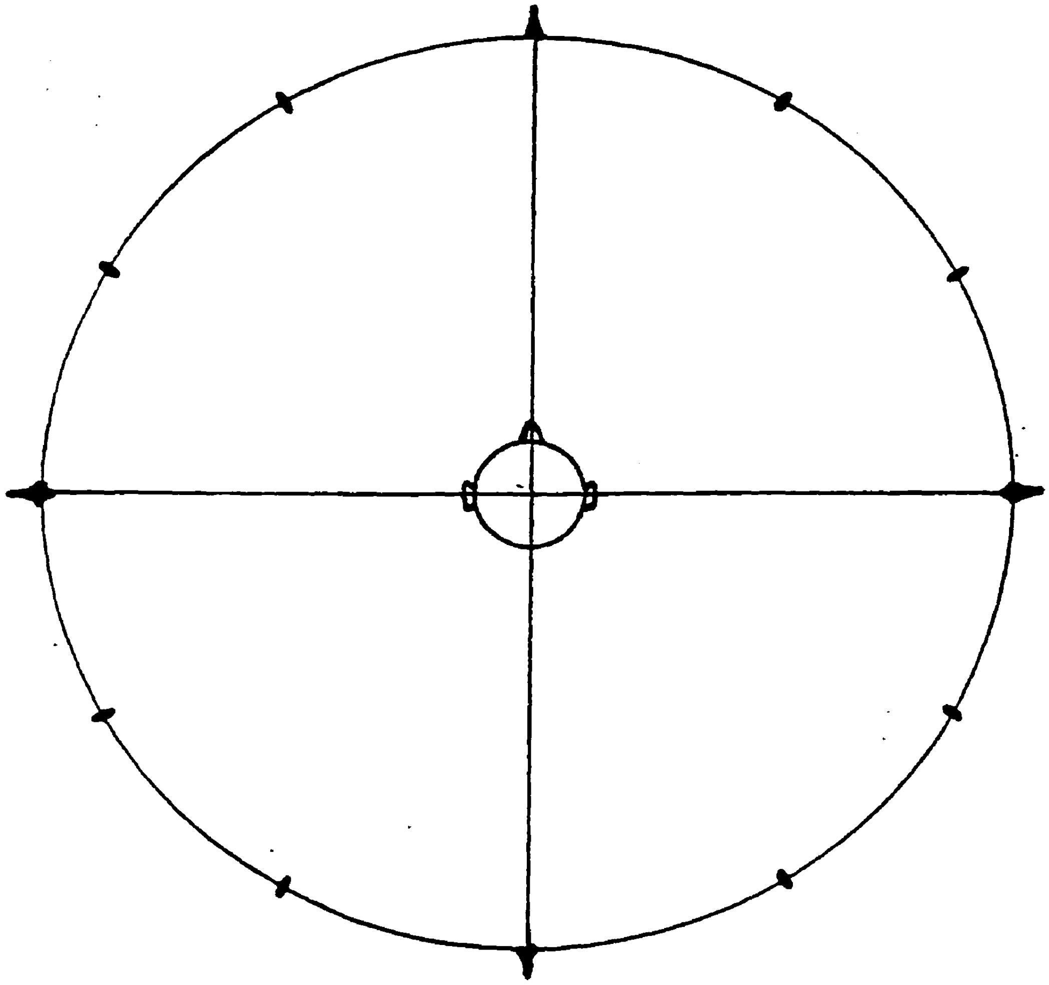


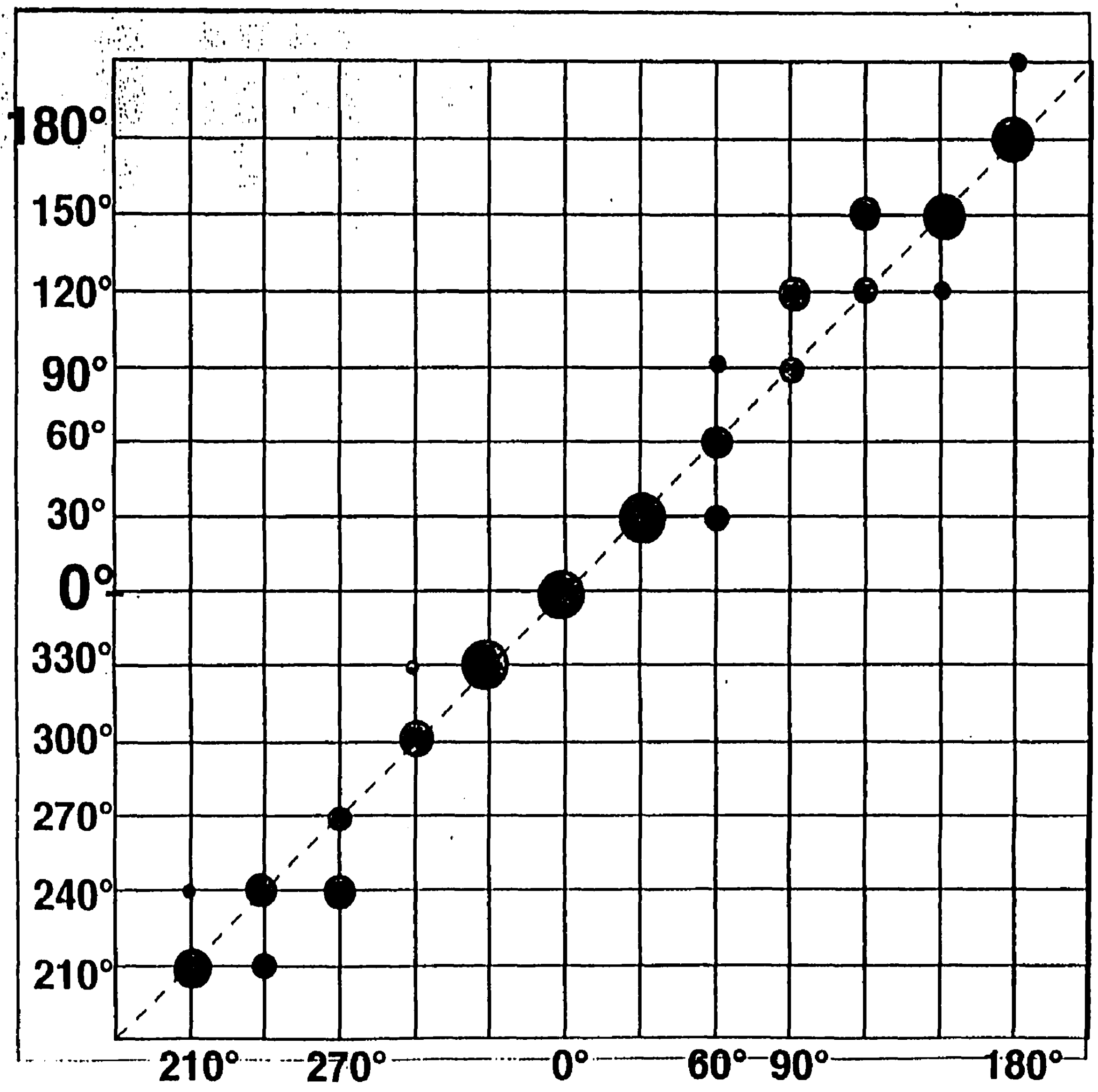
figure #14

Horizontal Imaging Tests - 3 subjects

John Klepko • McGill University

ROOM #1

perceived position



Scatter-plot of perceived vs. actual position (larger bubble implies higher occurrence)

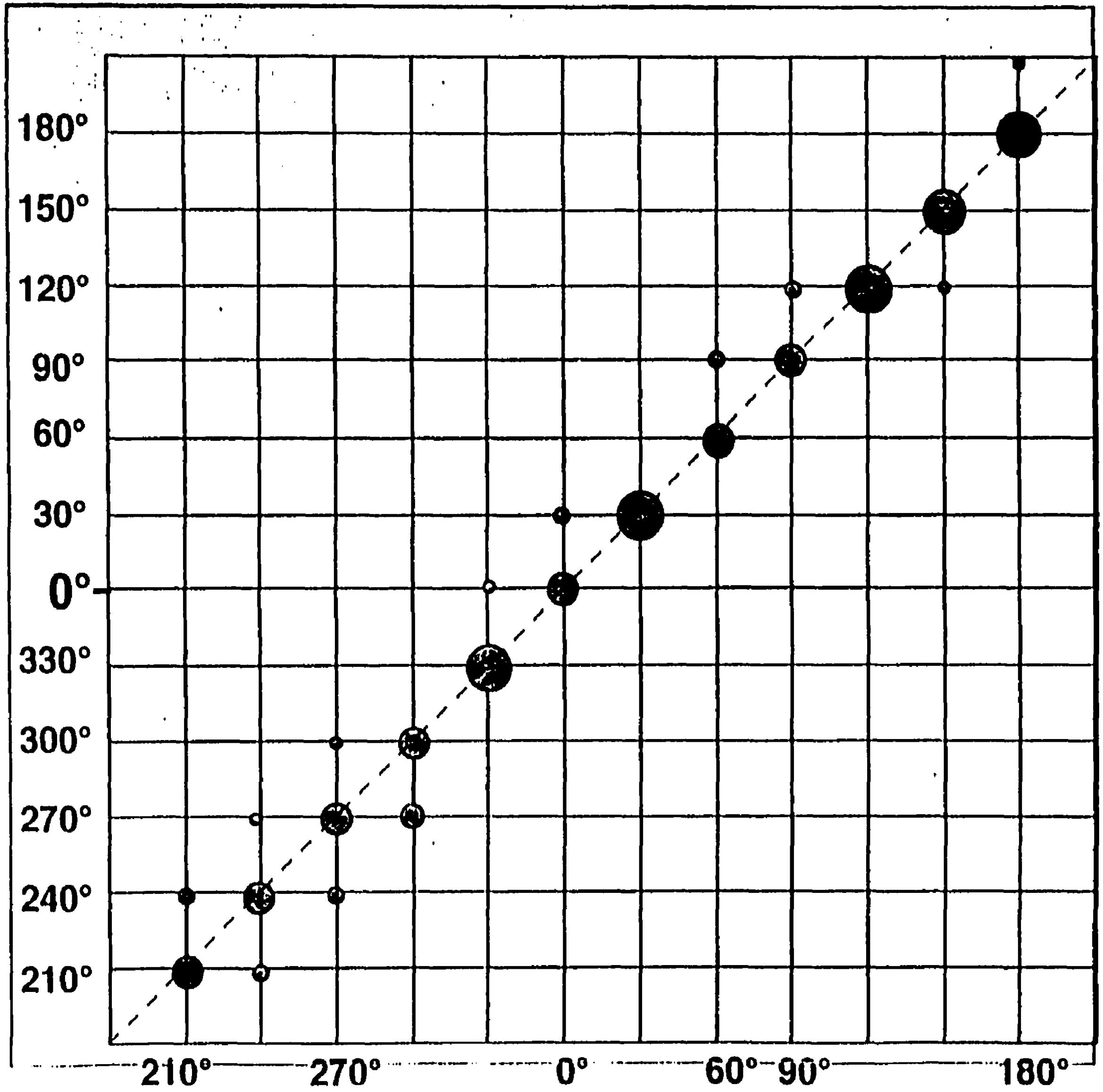
figure #15

Horizontal Imaging Tests - 3 subjects

John Klepko • McGill University

ROOM #2

perceived position



Scatter-plot of perceived vs. actual position (larger bubble implies higher occurrence)

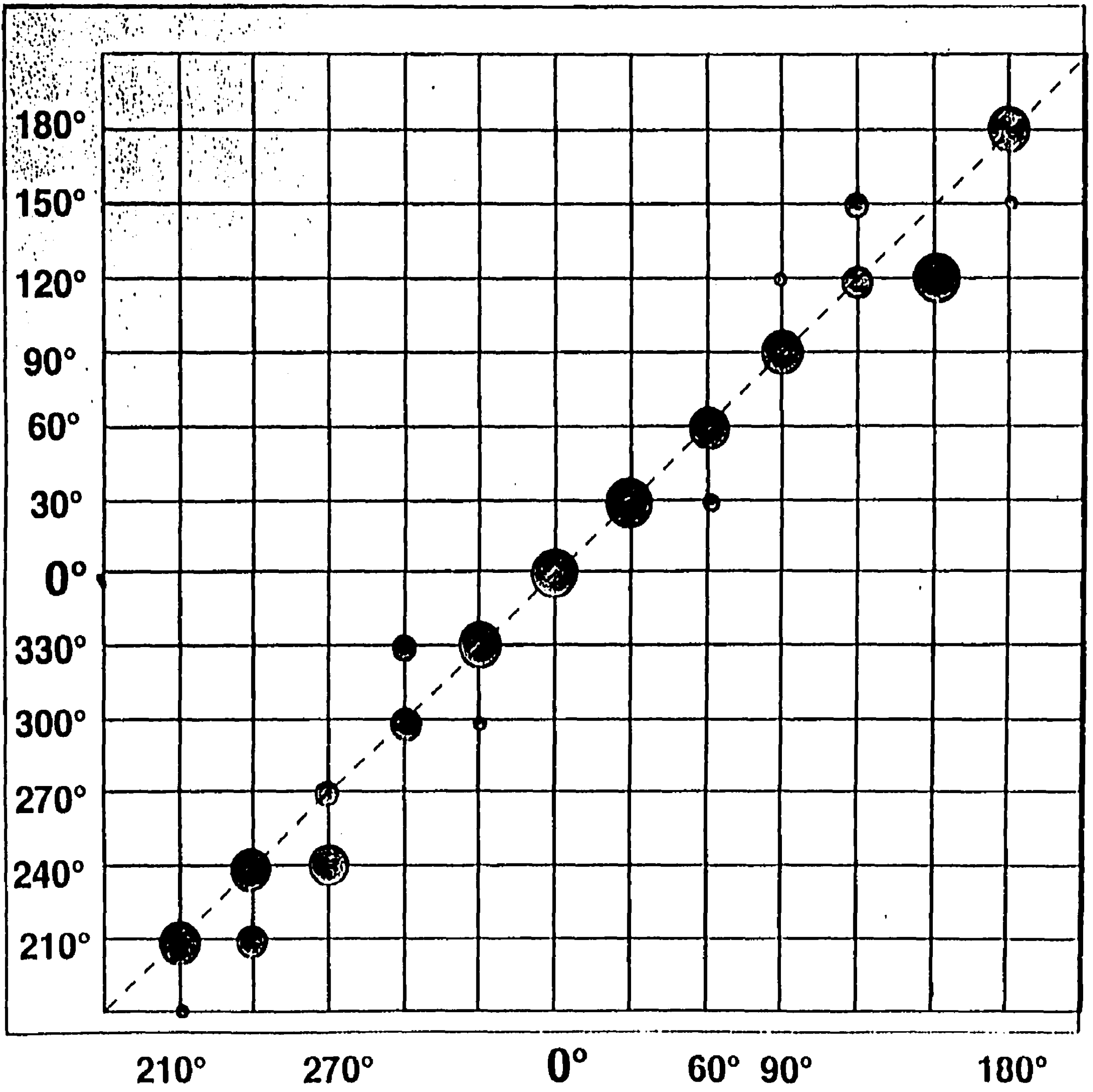
figure #16

Horizontal Imaging Tests - 3 subjects

John Klepko • McGill University

ROOM #3

perceived position

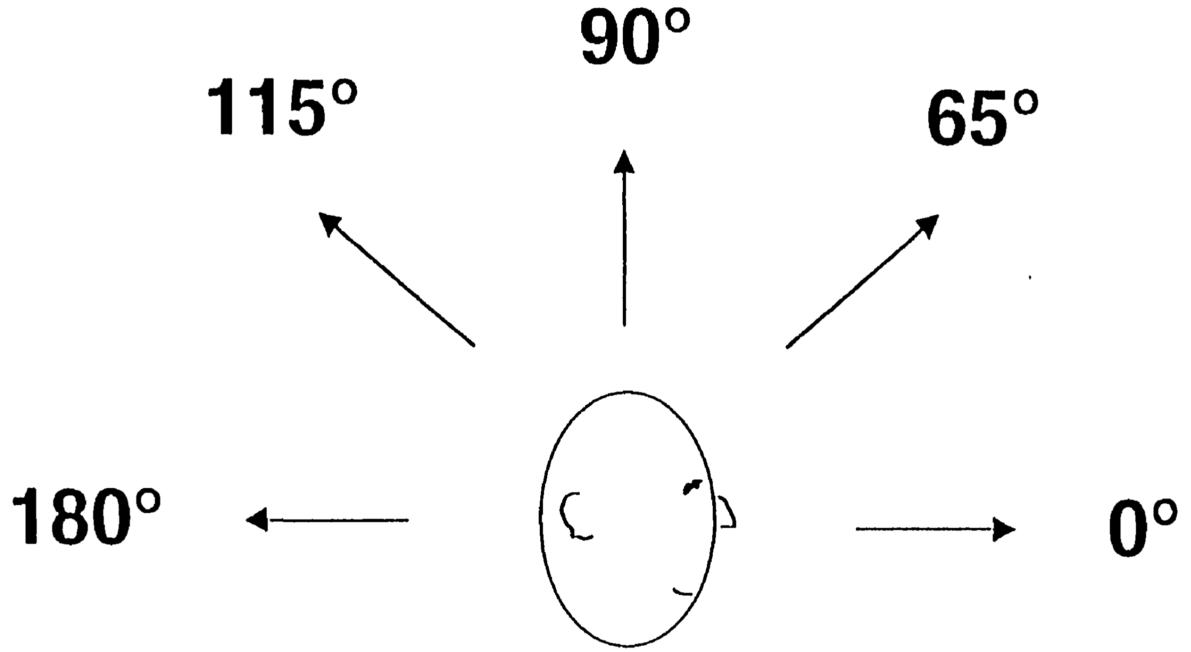


Scatter-plot of perceived vs. actual position
(larger bubble implies higher occurrence)

figure #17

Test positions

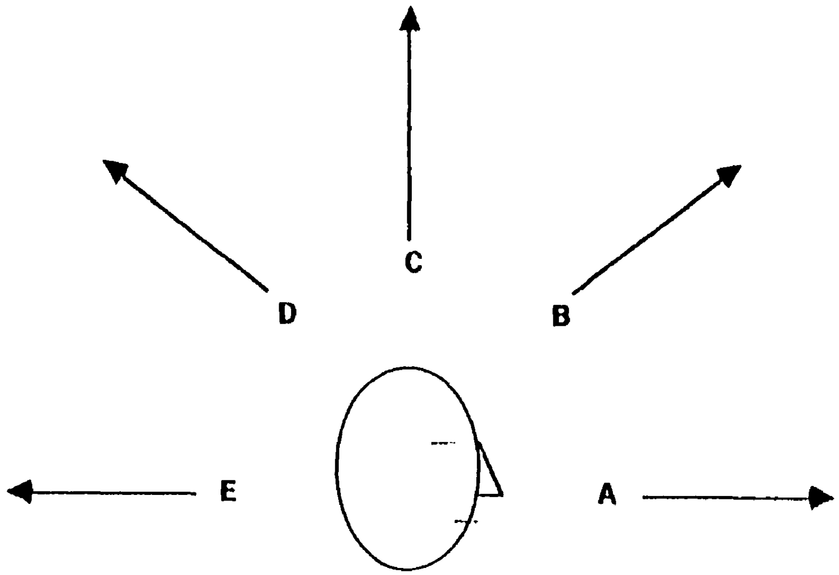
John Klepko • Faculty of Music – McGill University



5 positions tested in vertical imaging experiments

figure #18

Template used in vertical imaging experiments



23 24 25 26 27 28 29 30 31 32

33 34 35 36 37 38 39 40 41 42 43 44

INITIALS............

DATE...................

figure #19

(results)

John Klepko • Faculty of Music – McGill University

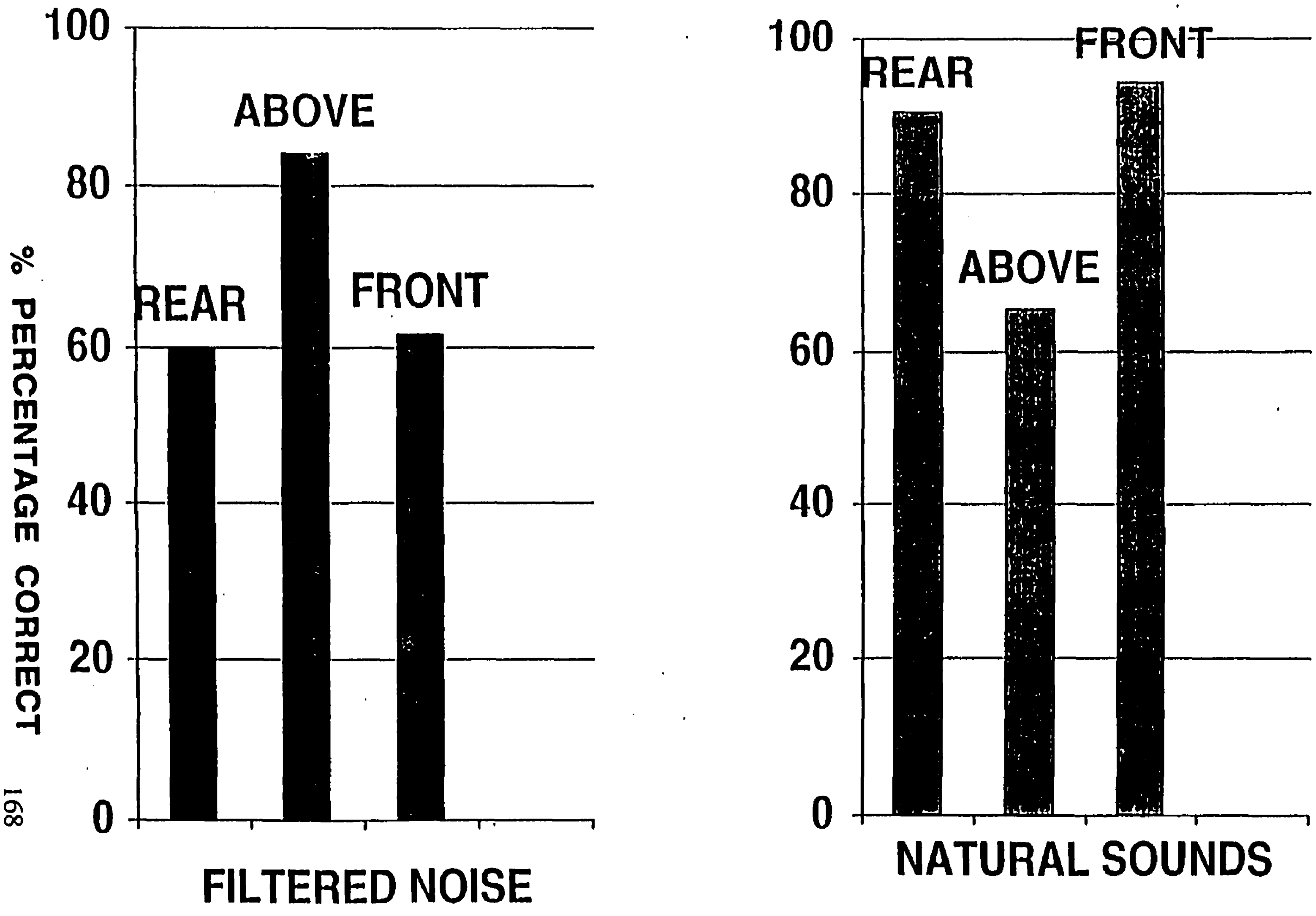


figure #20

vertical imaging test results

5 Positions – all sound types

John Klepko • Faculty of Music – McGill University

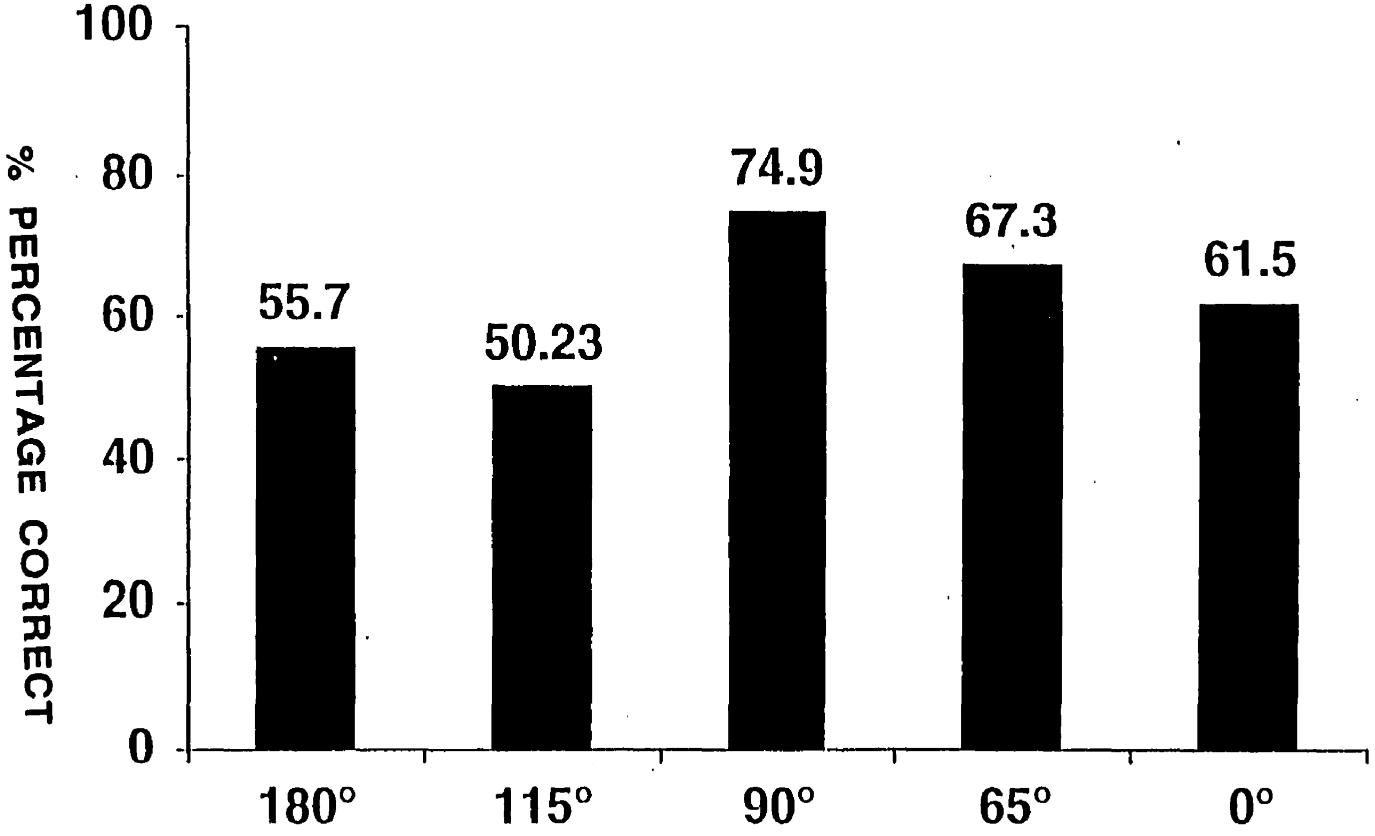


figure #21

vertical imaging test results

Natural sounds – all positions

John Klepko • Faculty of Music – McGill University

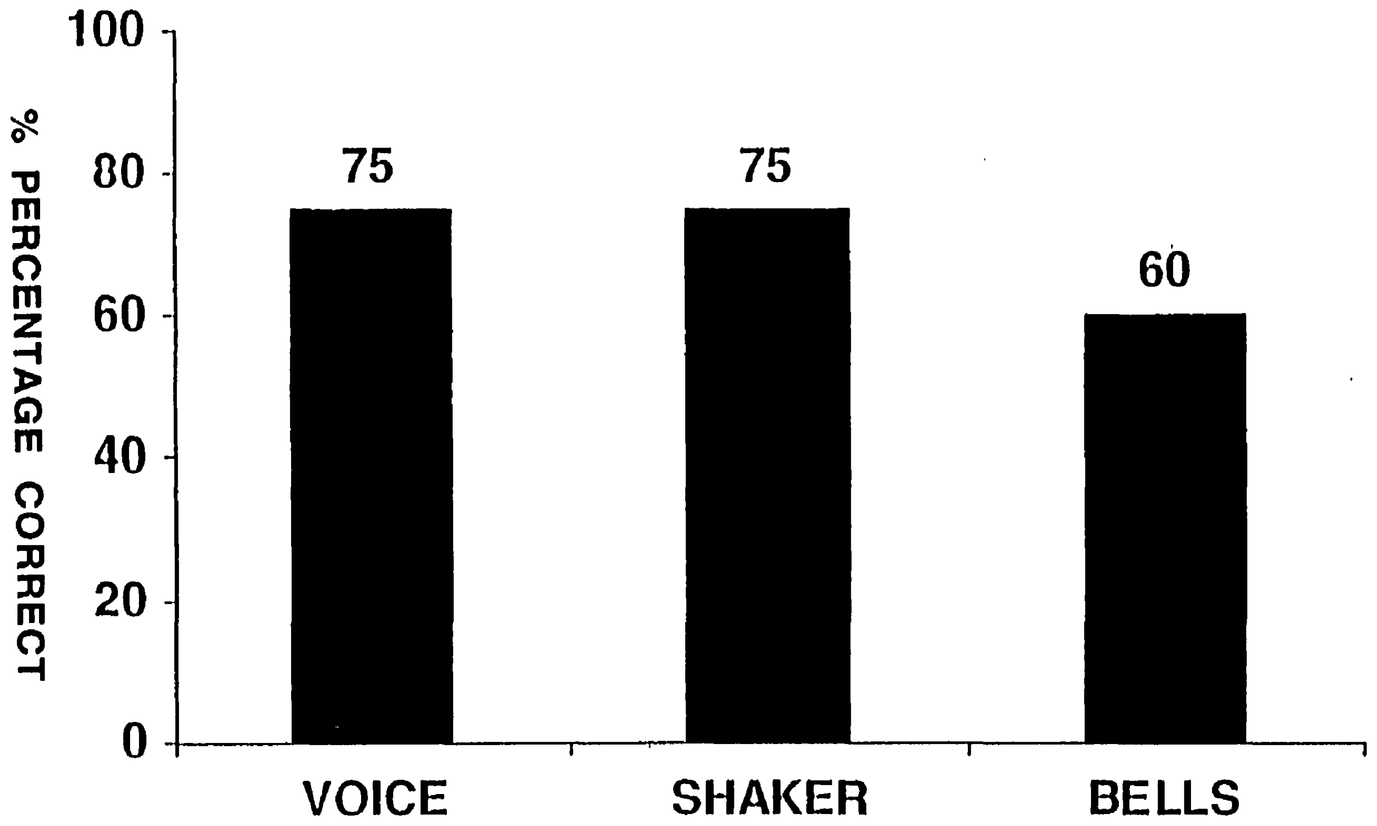
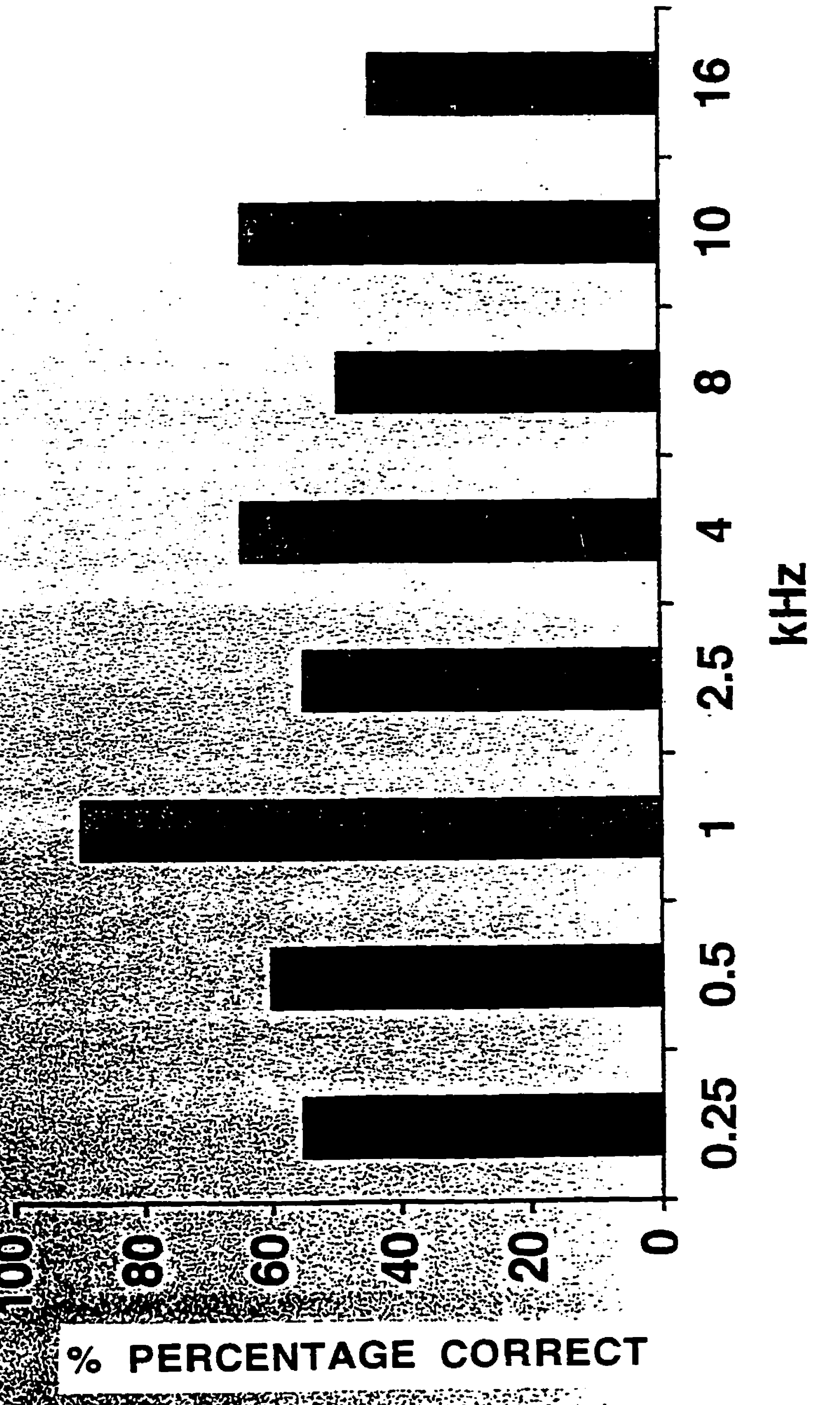


figure #22

vertical imaging test results

figure #23 vertical imaging test results



template used in distance imaging tests

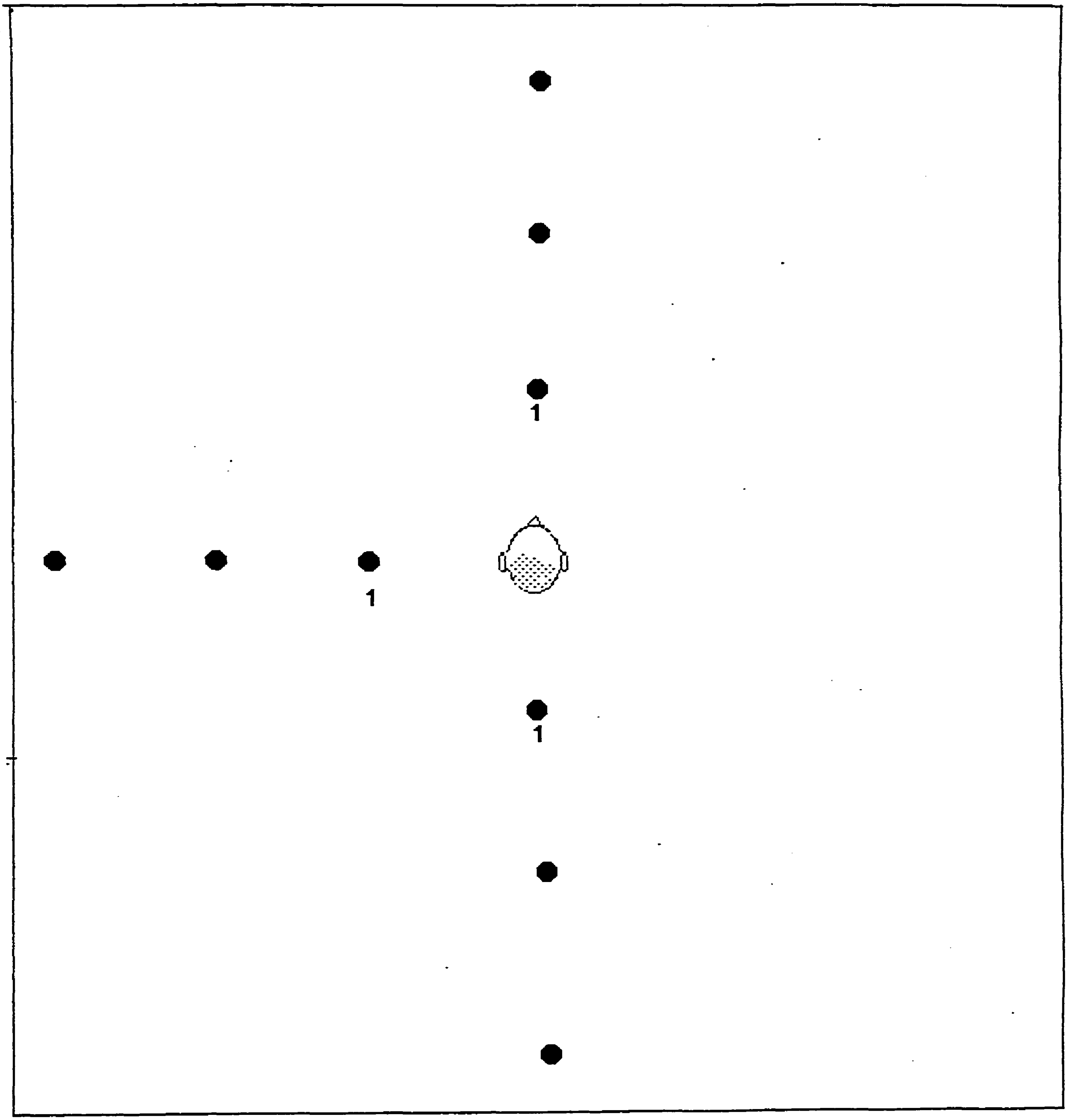


figure #24

template used in motion imaging tests

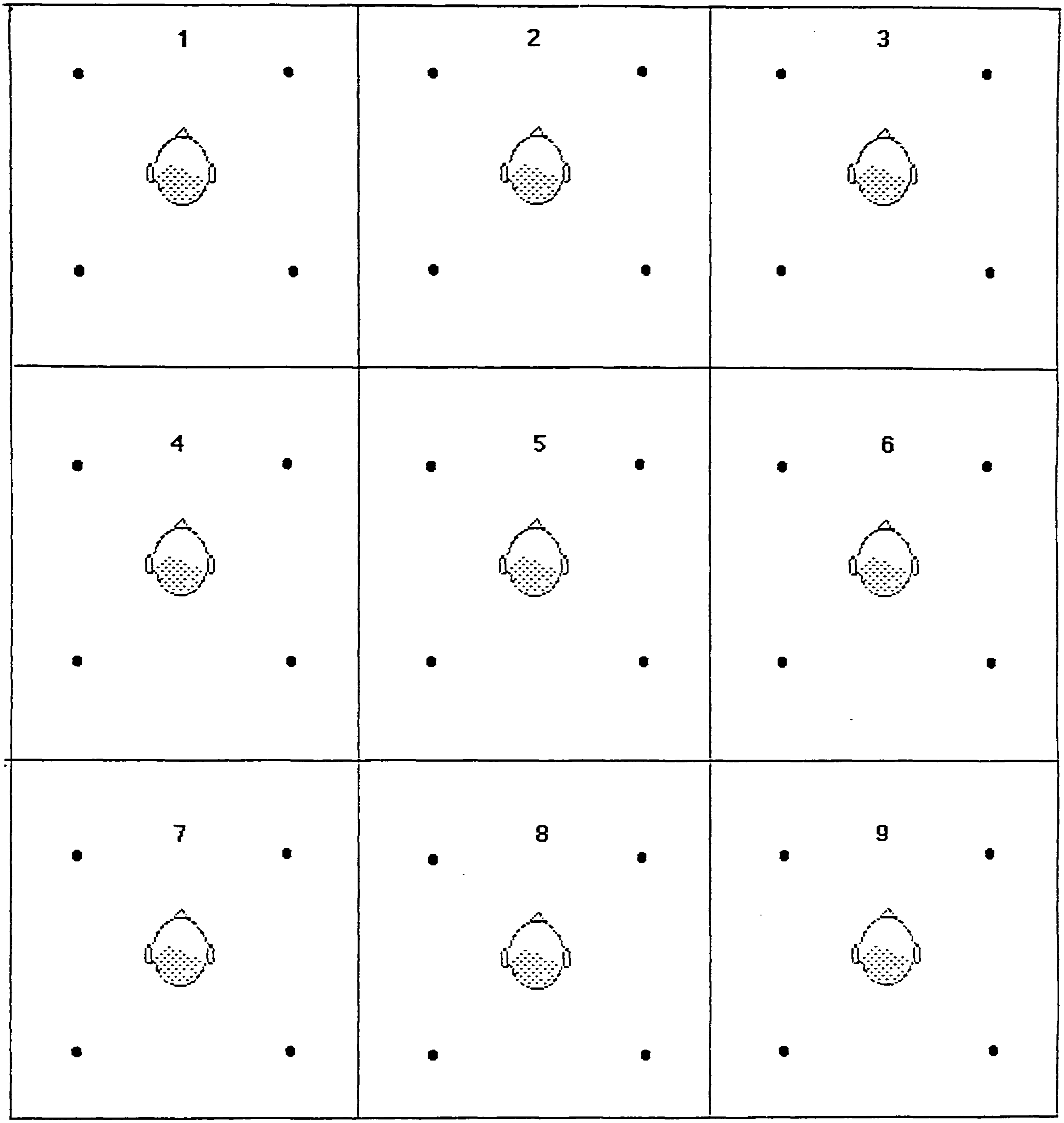


figure #25

APPENDIX A

DESCRIPTION OF OPTIONAL DEMONSTRATION TAPE

Format: DA-88, 8-track digital
Channel 1 = Left front (L)
Channel 2 = Right front (R)
Channel 3 = center front (C)
Channel 4 = blank
Channel 5 = Left surround (LS)
Channel 6 = Right surround (RS)
Channel 7 = direct sound (excerpt #14 only)
Channel 8 = “ “ “ “ “

Sampling frequency = 48 kHz

CONTENTS

1. pilot study walkaround (in 3 parts)..................1:00–9:40
2. coin rolling (in 2 parts)..9:50–10:30
3. jazz multitrack/overdub (re: section 5.3)......10:40–13:10
4. percussion ensemble...13:20–15:10
5. organ...15:20–20:33
6. 6-voice choir..20:50-24:10
7. SATB choir w/organ 24:18–27:45
8. jazz quintet..27:55–30:30
9. audience anticipation..................................30:45-37:42
10. vocal practice...37:48-38:54
11. apartment radio...39:05-39:35
12. backyard ambience...39:40-45:45
13. choir prep. noise...45:55-51:04
14. reverb chamber – jazz piano....................51:15-53:20
15. John Cage “4^{3}”..53:40-1:08:47

References

Ando, Y. (1985). *Concert Hall Acoustics.* Springer-Verlag, Berlin.

Ando, Y., and Kurihara, Y., (1985). "Nonlinear response in evaluating the subjective diffuseness of sound fields", *Journal of the Acoustical Society of America,* vol. 80, no.3, p. 833.

Ashmead, D., LeRoy, D., and Odom. R., (1990). "Perception of the relative distances of nearby sound sources", *Perception & Psychophysics.* vol. 47, no. 4, pp. 326-331.

AudioStax (1987). *The Space Sound CD: Dummy-head recording.* Düsseldorf. Germany, AX CD 91 101.

Barron, M., (1971). "The Subjective Effects of First Reflections in Concert Halls - The Need for Lateral Reflections", *J. Sound Vib.* 15, p. 475-494.

Bauer, B., (1961). "Stereophonic earphones and binaural loudspeakers", *Journal of the Audio Engineering Society,* vol. 9, no. 2.

Bauer, B., Rosenbeck, A. J., and Abbagnaro, L., (1967). "External-Ear replica for acoustical testing", *Journal of the Acoustical Society of America.* vol. 42.

Bauer, B., (1969). "Some techniques toward better stereophonic perspective", *Journal of the Audio Engineering Society,* vol. 17, no. 4.

Begault, Durand R., (1992). "Perceptual effects of synthetic reverberation on three-dimensional audio systems", *Journal of the Audio Engineering Society,* vol. 40, no. 11.

Begault, D. R., and Wenzel, E. M., (1993). "Headphone Localization of Speech", *Human Factors,* vol. 35, no. 2.

Begault, Durand R., (1994). *3D Sound for Virtual Reality and Multimedia.* Academic Press Inc., Chestnut Hill, Mass..

Begault, Durand R., Erbe, T., (1994). "Multichannel spatial auditory display for speech communications", *Journal of the Audio Engineering Society,* vol. 42, no. 10.

Begault, Durand R., (1996). "Audible and inaudible early reflections: Thresholds for auralization system design", *presented at the 100th Convention of the Audio Engineering Society* , Copenhagen, May, preprint #4244.

Beranek, Leo. (1996). *Concert and opera halls: how they sound.* Acoustical Society of America, Woodbury, New York.

Blauert, J., (1969). "Sound localization in the median plane", *Acustica,* vol. 22.

Blauert, J., and Lindemann, W., (1985). "Auditory spaciousness: Some further psychoacoustic analyses", *Journal of the Acoustical Society of America*. vol. 80, no. 2, p. 533.

Blauert, J., (1996). *Spatial Hearing: the Psychophysics of human sound localization*, MIT Press, Cambridge, Massachusetts.

Bloom, P. J., (1977). "Creating source elevation illusions by spectral manipulation", *Journal of Audio Engineering Society*. vol. 25. no. 9.

Blumlein, A., (1933) "Improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems", British patent no. 394,325 reprinted in the *Journal of Audio Engineering Society*, vol. 6, no. 3, (April).

Boré, G. (1978). *Microphones: for professional and semi-professional applications*. Georg Neumann, GmbH Berlin. (p. 19)

Boulez, P., Gerszo, A., (1988). "Computers in music", Scientific American, April vol. 258, no. 4.

Brant, Henry. (1967). "Space as an Essential Aspect of Musical Composition" in *Contemporary Composers on Contemporary Music*. edited by Schwartz, Elliot, Childs Barney. Holt, Rinehart and Wilson. New York.

Bregman, Albert S. (1990). *Auditory Scene Analysis: The perceptual organization of sound*. MIT Press. Cambridge, Mass.

Burkhard, M., and Sachs, R., (1975). "Anthropometric manikin for acoustic research", *Journal of the Acoustical Society of America*, vol. 58, no. 1.

Burkhard, M., Bray, W., Genuit, K., and Gierlich, H., (1991). "Binaural Sound for Television" from the proceedings of the AES 9th International Conference, Detroit, Michigan, February, p. 119

Butler, R.A., and Belendiuk, K., (1977). "Spectral cues used in the localization of sound in the median sagittal plane". *Journal of the Acoustical Society of America*, vol. 61., pp.1264-1269.

Cabot, R., (1979). "A triphonic sound reproduction system using coincident microphones", *Journal of the Audio Engineering Society*, vol. 27, no. 12.

Carver, R., (1982). "Sonic Holography", *Audio*, (March).

Chowning, J. (1971). "The simulation of moving sound sources", *Journal of the Audio Engineering Society*, vol. 19, no. 1.

Cochran, P., Throop, J., and Simpson, W., (1968). "Estimation of distance of a source of sound", *American Journal of Psychology*, vol. 81, p. 198.

Coleman, P., (1961). "Failure to localize the source distance of an unfamiliar sound", *Journal of the Acoustical Society of America*, vol. 34, no. 3, p. 345.

Coleman, P., (1963). "An analysis of cues to auditory depth perception in free space", *Psychological Bulletin*, vol. 60, no.3.

Cook, P., (1991). "Toward a microphone technique for Dolby Surround encoding", Master's thesis, Faculty of Music, McGill University, Montreal.

Cooper, D., (1970). "How many channels?", *Audio*, (November).

Cooper, D., and Bauck J., (1989). "The Prospect for Transaural Recording", *Journal of the Audio Engineering Society*. vol. 37, no. 1/2.

Cooper, D., and Bauck, J., (1992). "Head Diffraction Compensated Stereo System" United States Patent #5,136,651

Chowning, John M. (1971) "The Simulation of Moving Sound Sources". *Journal of the Audio Engineering Society*, vol. 26, no. 1, pp. 2-6.

Cremer, L., and Müller, H. (1978), *Principles and Applications of Room Acoustics - Volume I*, Applied Science Publishers Ltd., Essex, England. translated from German by Schultz, T..

Damaske, P., (1971). "Head-related two-channel stereophony with loudspeaker reproduction", *Journal of the Acoustical Society of America*, vol. 50, # 4.

Damaske, P. (1997). "Diffuse onset of reverberation". in *Music and Concert Hall Acoustics*. ed. Ando, Y., Nason, D.. Academic Press, San Diego.

Davis, Don & Carolyn, (1989). *"In-the Ear Recording and Pinna Acoustic Response Playback"* presented at the 87th Audio Engineering Society convention, New York, October.

Dickreiter, M. (1989). *Tonmeister Technology*. Temmer Enterprises Inc. New York, N.Y. (p. 69)

Doll, T., Hanna, T., and Russotti, J., (1992). "Masking in three-dimensional auditory displays", *Human Factors*, vol. 34, no. 3.

Eargle, John, (1980). *Sound Recording*. Van Nostrand Reinhold Company Inc., New York, N.Y., 2nd Edition.

Eargle, John, (1981). *The Microphone Handbook*. Elar Publishing Co., Plainview, New York.

Eargle, John, (1998) from *"Mixing Music for Multichannel Recording"*, presented at the Los Angeles section of the Audio Engineering Society, June 23.

Embleton, Tony F W. (1996). "Tutorial on sound propagation outdoors". *Journal of the Acoustical Society of America.* vol. 100, no.1. pp. 31-48.

Fletcher, H. (1933). "An Acoustic Illusion Telephonically Received", *Bell Laboratories Record*, June, p.28, (reprinted in *Audio*, July 1957).

Fukuda, A., Tsujimoto, K., Akita, S.. (1997) *"Microphone Techniques for Ambient Sound on Music Recording"*, presented at the 103rd Convention of the Audio Engineering Society, New York, N.Y., November. preprint #4540.

Furuya, H., Fujimoto, K., Takeshima, Y., Nakamura, H. (1995). "Effect of early reflections from upside on auditory envelopment". *Journal of the Acoustical Society of Japan*, vol. 16, no. 2.

Gardner, M. B., (1968a). "Proximity image effect in sound localization", *Journal of the Acoustical Society of America*, vol. 43, p163.

Gardner, M. B., (1968b). "Distance estimation of 0° or apparent 0°-oriented speech signals in anechoic space", *Journal of the Acoustical Society of America*, vol[illegible], no. 1, p. 47.

Gardner, M. B., (1969). "Image fusion, broadening and displacement in sound localization", *Journal of the Acoustical Society of America*, vol. 46, pp. 339-349.

Gardner, M. B., (1973). "Some monaural and binaural facets of median plane localization", *Journal of the Acoustical Society of America*, vol. 54, no. 6.

Gardner, M. B., and Gardner, R. S., (1973). "Problems of localization in the median plane: effect of pinnae occlusion", *Journal of the Acoustical Society of America*, vol. 53, no. 2.

Genuit, K., and Bray, W., (1989). "The Aachen Head System", *Audio*, December.

Gierlich, H., and Genuit, K. (1989). "Processing artificial head recordings", *Journal of the Audio Engineering Society*, vol. 37, no. 1/2. (January/February)

Good, M.D., and Gilkey, R.H. (1994). "Auditory localization in noise:I. The effects of signal-to-noise ratio", and "Auditory localization in noise:II. The effects of masker location". *Journal of the Acoustical Society of America*, vol. 95., p. 2896.

Good, M.D., and Gilkey, R.H. (1996). "Sound localization in noise: The effect of signal-to-noise ratio". *Journal of the Acoustical Society of America*, vol. 99., p. 1108.

Grantham, W., (1986). "Detection and discrimination of simulated motion of auditory targets in the horizontal plane", *Journal of the Acoustical Society of America*, vol. 79., p. 1939.

Greene, D., (1968). Comments on "Perception of the range of a sound source of unknown strength", *Journal of the Acoustical Society of America.* vol. 44, no. 2, p. 634.

Griesinger, D., (1989). "Equalization and Spatial Equalization of Dummy-Head Recordings for Loudspeaker Playback", *Journal of the Audio Engineering Society*, vol. 37, no. 1/2. (January/February)

Hafler, D., (1970). "A New Quadraphonic System", *Audio*, vol. 54, (July).

Haines, T., and Hooker M. (1997). "Multichannel audio dramas: a proposal" *as presented at the 102nd Audio Engineering Society Convention*, Munch, Germany, March. preprint no. 4428.

Harley, Maria A. (1994). *Space and Spatialization in Music: History and Analysis, Ideas and Implementations*. Ph. D. Dissertation. McGill University. Montreal, Quebec.

Harris, J., and Sergeant, R. (1971). "Monaural/binaural minimum audible angles for a moving sound source", *Journal of Speech and Hearing Research*, vol. 14.

Hertz, B., (1981). "100 years with stereo: the beginning", *Journal of the Audio Engineering Society*, vol. 29, no. 5. Reprinted from *Scientific American*. no. 3, 1881.

Hirsch, H., (1968). "Perception of the range of a sound source of unknown strength", *Journal of the Acoustical Society of America*. vol. 43, no. 2, p. 373.

Holman, T., (1991). "New factors in sound for cinema and television", *Journal of the Audio Engineering Society*, vol. 39, no. 7/8, (July/August).

Holman, T., (1991a). "Sound system with source material and surround timbre response correction, specified front and surround loudspeaker directionality, and multi-loudspeaker surround". *United States Patent*, #5,043,970.

Holman, T., (1993) . "The Center Channel", *Audio*, (April).

Holman, T., (1996). "Channel Crossing", *Studio Sound*, (parts 1-3), (February, March , April).

Holman, T., (1998). "In the beginning there was...Fantasound", *Surround Professional*, vol. 1, no. 1, (October).

Holt, R., and Thurlow, W., (1969). "Subject orientation and judgement of distance of a sound source", *Journal of the Acoustical Society of America*. vol. 46, pp. 1584-1585.

Kendall, G., Martens, W., and Wilde, M., (1990). "A spatial sound processor for loudspeaker and headphone reproduction", *Proceedings of the AES 8th International Conference*, Washington, D.C., May.

Killion, Mead, (1979). "Equalization Filter for Eardrum-Pressure Recording Using a KEMAR Manikin", *Journal of the Audio Engineering Society*, vol. 27, no. 1/2. (January/February)

Klapholz, J., (1991). "Fantasia: Innovations in Sound", *Journal of the Audio Engineering Society*, vol. 39, no. 1/2. (January/February), p.66.

Klepko, J., (1997). "5-channel microphone array with binaural head for multichannel reproduction", *presented at the Audio Engineering Society Convention, September*, New York, preprint # 4541.

Klepko, J., (1998). "Vertical imaging capability of surround-sound systems through binaural techniques", *presented at the Canadian Acoustical Association conference,* October, London, Ontario. (printed in the proceedings).

Klepko, J., (1999). "The psycho-acoustics of surround sound", *a presentation delivered at the "Audio Production for Discrete Surround-Sound" conference of the Toronto section of the Audio Engineering Society*, Toronto, May, 1999.

Klipsch, P., (1959). "Three-channel stereo playback of two tracks derived from three microphone", *I.R.E. Transactions on Audio*, (March-April).

Klipsch, P., (1960a). "Experiments and experiences in stereo". *I.R.E. Transactions on Audio*, (May-June), p. 91.

Klipsch, P., (1960b). "Double Doppler Effect in stereophonic recording and playback of a rapidly moving object", *I.R.E. Transactions on Audio*, (May-June), p. 105.

Kuhn, George F. (1987). "Physical Acoustics and Measurements Pertaining to Directional Hearing", in *Directional Hearing*. edited by Yost,W. A., and Gourevitch G.. Springer-Verlag, New York.

Larcher, V., Vandernoot, G., Jot, J.M., (1998). *"Equalization Methods in Binaural Technology"*, presented at the 105th AES convention, San Francisco, September, preprint #4858.

Long, Edward M. (1972). "The Effects of Transit Time and Intensity Differentials Upon Recreating Acoustic Fields". *Journal of the Audio Engineering Society,* vol. 20, no. 2.

McCabe, C.J., and Furlong, D.J., (1994). "Virtual Imaging Capabilities of Surround Sound Systems", *Journal of the Audio Engineering Society*, vol. 42, no. 1/2, p. 38.

McLuhan, M. (1964). *Understanding Media: The extensions of man.* New York, McGraw-Hill.

Mershon, D., and Bowers, J., (1979). "Absolute and relative cues for the auditory perception of egocentric distance", *Perception*, vol. 8, pp. 311-322.

Mershon, D., and King, E., (1975). "Intensity and reverberation as factors in the auditory perception of distance", *Perception and Psychophysics*, vol. 18, no. 6. pp. 409-415.

Michelsen, J., and Rubak, P., (1997). "Parameters of distance perception in stereo loudspeaker scenario", *presented at the 102nd Audio Engineering Society Convention*, Munich, Germany, (March), preprint # 4472.

Mills, A. W. (1958). "On the Minimum Audible Angle". *Journal of the Acoustical Society of America.* vol. 30, no. 4.

Molino, J., (1973). "Perceiving the range of a sound source when the direction is known", *Journal of the Acoustical Society of America.* vol. 53, no. 5.

Møller, H. (1989). "Reproduction of Artificial-Head Recordings through Loudspeakers", *Journal of the Audio Engineering Society,* vol. 37, no. 1/2.

Møller, H. (1992). "Fundamentals of Binaural Technology", *Applied Acoustics* , vol. 36, pp. 171-217.

Møller, H., Jensen, C.J., Hammershøi, D., Sørensen, M.F.. (1996). "Binaural Technique: Do We Need Individual Recordings?". *Journal of the Audio Engineering Society,* vol. 44, no. 6.

Møller, H., Jensen, C.J., Hammershøi, D., Sørensen, M.F.. (1997). *"Evaluation of Artificial Heads in Listening Tests"*, presented at the 102nd Convention of the Ausio Engineering Society, Munich, Germany, March. preprint #4404 .

Mori, T., Fujiki, G., Takahashi, N., and Maruyama F., (1979). "Precision Sound Image Localization Technique Utilizing Multitrack Tape Masters", *Journal of the Audio Engineering Society*, Vol. 27, no. 1/2.

Morimoto, M., and Maekawa, Z., (1988). "Effects of Low Frequency Components on Auditory Spaciousness", *Acustica*, vol.66, p.190.

Morimoto, M., Iida, K., and Furue, Y., (1993). "Relation Between Auditory Source Width in Various Sound Fields and Degree of Interaural Cross-Correlation", *Applied Acoustics*, vol. 38, p. 291.

Morimoto, M., Iida, K., Sakagami, K., and Marshall, A.H., (1994). *"Physical measures for auditory source width (ASW): Part 2. Comparison between various physical measures and ASW (Auditory Source Width)"*, from the proceedings of the Wallace Clement Sabine Centennial Symposium, Cambridge, Mass..

Morimoto, M., Sugiura, S., and Iida, K., (1994). "Relation between auditory source width in various sound fields and degree of interaural cross-correlation: Confirmation by constant method", *Applied Acoustics*, vol. 42, p. 233.

Nielsen, S., (1993). "Auditory distance perception in different rooms", *Journal of the Audio Engineering Society*, vol. 41, no. 10, (October).

Nordlund, B., and Líden, G., (1963). "An artificial head", *Acta Oto-laryngol.*, vol. 56, p. 493.

Olson, David.R., Bialystok, Ellen (1983). *Spatial Cognition: The structure and development of mental representations of spatial relations.* Hillsdale, NewJersey, Lawrence Erlbaum Associates.

Perrot, David, R., (1974). "Auditory apparent motion", *Journal of Auditory Research*, vol. 3, p. 163-169.

Perrot, David, R. (1984). "Concurrent minimum audible angle: A re-examination of the concept of spatial acuity". *Journal of the Acoustical Society of America*. vol. 75, no. 4.

Perrot, David, R. (1984). "Discrimination of the spatial distribution of concurrently active sound sources: some experiments with stereophonic arrays", *Journal of the Acoustical Society of America*. vol. 76, no. 6.

Perrot, David, R., and Musicant, A., (1977). "Minimum auditory movement angle: Binaural localization of moving sound sources". *Journal of the Acoustical Society of America*. vol. 62, no. 6.

Perrot, David, R., and Musicant, A., (1981). "Dynamic minimum audible angle: Binaural spatial acuity with moving sound sources", *Journal of Auditory Research*, vol. 21.

Perrot, D., and Tucker, J., (1988). "Minimum audible movement angle as a function of signal frequency and the velocity of the source", *Journal of the Acoustical Society of America*. vol. 83, no. 4.

Perrot, David R., Buell, Thomas N., (1992) "Judgements of sound volume: Effects of signal duration, level, and interaural characteristics on the perceived extensity of broadband noise". *Journal of the Acoustical Society of America*, 72 (5). pp. 1413-1417.

Pick, H.L., Warren, D.H., & Hay, J.C. (1969). "Sensory conflict in judgements of spatial direction". *Perception & Psychophysics, 6,* pp. 203-205.

Plenge, G., (1974). "On the difference between localization and lateralization", *Journal of the Acoustical Society of America*, vol. 56, pp. 944-951

Pompetzki, Wulf. (1990). "Binaural Recording and Reproduction for Documentation and Evaluation". in *The Sound of Audio: the proceedings of the AES 8th international conference*, Washington, D.C. pp.225-229.

Reynolds, Roger. (1978). "Thoughts on Sound Movement and Meaning". *Perspectives of New Music*. vol. 16, no. 2.

Roffler, S. K., and Butler, R.A., (1968). "Factors that influence the localization of sound in the vertical plane", *Journal of the Acoustical Society of America*, vol. 43, no. 6.

Rosenblum, L., Carello, C., Pastore, R., (1987). "Relative effectiveness of three stimulus variables for locating a moving sound source", *Perception*, vol. 16.

Scheiber, P., (1971). "Four Channels and Compatibility", *Journal of the Audio Engineering Society*, vol. 19, no. 4.

Schroeder, M.R., (1958). "An artificial stereophonic effect obtained from a single audio channel", *Journal of the Acoustical Society of America*, vol. 6, no. 2.

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

Scornick, Lynn Barry. (1984). *Composer Prescribed Seating Arrangements in Twentieth Century Music*. dissertation for D.M.A., University of Texas at Austin. University Microfilms International. Ann Arbor , Michigan.

Shlien, Seymour, and Soulodre, Gilbert. "Measuring the characteristics of expert listeners", presented at the 101st AES convention, Los Angeles, California, November 1990. preprint no. 4339.

Somers, Harry. (1963). "Stereophony for Orchestra", *Music Across Canada I*. March, pp. 27-8.

Snow, William B. (1953). "Basic Principles of Stereophonic Sound". in *Stereophonic Techniques: An anthology of reprinted articles on stereophonic techniques*. Audio Engineering Society Inc., New York. p. 19.

Steinberg, J.C., and Snow, W.B., (1934) "Auditory Perspectives - Physical Factors", *Electrical Engineering*, vol. 53, no. 1.

Strybel, T., Manligas, C., and Perrot, D., (1989). "Auditory apparent motion under binaural and monaural listening conditions", *Perception & Psychophysics*, vol. 45, no. 4.

Strybel, T., Manligas, C., and Perrot, D., (1992). "Minimum audible movement angle as a function of the azimuth and elevation of the source", *Human Factors*, vol. 34, no. 3.

Sunier, John, (1989). "Ears Where the Mikes Are: part 1", *Audio*, (November).

Tanner, Theodore. (1997) *Psychoacoustic criteria for auditioning virtual imaging systems.* unpublished manuscript.

Theile, Günther, (1986). "On the Standardization of the Frequency Response of High-Quality Studio Headphones", *Journal of the Audio Engineering Society*, vol. 34, no. 12. (December).

Theiss, Bernd, and Hawksford, Malcolm. (1997). *"Objective Assessment of Phantom Images in a 3-Dimensional Sound Field Using a Virtual Listener"*, presented at the 102nd Audio Engineering Convention. Munich, Germany, March. preprint #4462

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

Toole, F. E., (1969) "In-Head Localization of Acoustic Images". *Journal of the Acoustical Society of America,* vol. 48, no. 4.

Toole, F. E., (1991) "Binaural Record/Reproduction Systems and Their Use in Psychoacoustic Investigations" presented at the 91st convention of the Audio Engineering Society, New York, October, preprint # 3179.

Toole, F. E., (1996) "VMAx - the Harman approach to 3-D audio", Harman International product release notes. (March 8)

Torick, E., Di Mattia, A., Rosenheck, A., Abbagnaro, L., and Bauer, B., (1968) "An electronic dummy for acoustical testing", *Journal of the Audio Engineering Society*, October, vol. 16, no. 4.

Trahiotis, C., and Bernstein, L.R., (1986). "Lateralization of noise", *Journal of the Acoustical Society of America,* vol. 79, no. 6.

Ueda, K., and Morimoto, M., (1995). "Estimation of Auditory Source Width (ASW): I. ASW for two adjacent 1/3 octave band noises with equal band level", *Journal of the Acoustical Society of Japan*, vol. 16, no. 2, p. 77.

Warren, D. H. (1970). "Intermodality interactions in spatial localization", in W. Reitman (Ed.), *Cognitive Psychology* . New York, Academic Press.

Wenzel, E., Arruda, M., Kistler, D., and Wightman, F., (1993). "Localization using non-individualized head-related transfer functions", *Journal of the Acoustical Society of America*, vol. 94, no. 1 .

Wöhr, M., Thiele, G., and Goeres, H.-J., (1991). "Room-related balancing technique: A method for optimizing recording quality", *Journal of the Audio Engineering Society*, September, vol. 39, no. 9.

Woram, John M. (1989). *Sound Recording Handbook.* Howard W. Sams & Co., Indianapolis, Indiana. (p. 99).

Woszczyk, Wieslaw R., (1979). "Improved Instrument Timbre Through Microphone Placement", *Recording Engineer/Producer*, vol. 10, (May).

Woszczyk, Wieslaw R., (1993). "Quality Assessment of Multichannel Sound Recordings" in *AES 12th International Conference*, pp. 197-218.

Wuttke, J., (1985). *"Conventional and New Viewpoints on Electrostatic Pressure Transducers"*, presented at the 79th AES convention, New York, October, preprint #2305.

Xiang, N., Genuit, K., Gierlich, H., (1993). "Investigations on a new reproduction procedure for binaural recordings", presented at the 95th AES convention, New York, October, preprint #3732.

Yamamoto, Takeo. (1997) from lecture notes given at the Faculty of Music, McGill University, Montreal.