

**Perceived differences in recordings produced with  
four 5.0 surround microphone techniques**

Markku Hietala  
Master's Thesis  
Music, Mind and Technology  
University of Jyväskylä  
April 2007

Tiedekunta – Faculty Humanities	Laitos – Department Music
Tekijä – Author Markku Hietala	
Työn nimi – Title Perceived differences in recordings produced with four 5.0 surround microphone techniques.	
Oppiaine – Subject Music, Mind and Technology	Työn laji – Level Master's Thesis
Aika – Month and year April, 2007	Sivumäärä – Number of pages 42+3
<p>Tiivistelmä – Abstract</p> <p>Recently, the surround sound has enjoyed an increased popularity among audio professionals and consumers. The increase in surround music recording has raised quite a few questions concerning the best possible equipment, microphone choice and microphone setup. Several theories have been established to optimize the microphone setup in recording situation. Different microphone setups or microphone arrays are based on various psychoacoustical theories and aesthetic preferences. This research compared four different 5.0 surround microphone techniques in listening test. The evaluated variables were sound source depth and width, the amount of ambience and the naturalness in the recording.</p>	
Asiasanat – Keywords surround sound, microphone techniques, recording, listening test	
Säilytyspaikka – Depository	
Muita tietoja – Additional information	

## Table of contents

1	INTRODUCTION .....	4
2	THE SURROUND SOUND .....	6
	2.1 Early attempts in surround sound .....	6
	2.2 The Quad and its successors .....	7
	2.3 Toward standardization .....	7
	2.4 The surround systems today .....	8
	2.4.1 The matrixed and the discrete systems..	8
	2.4.2 The 5.1 surround system .....	9
	2.5. Spatial hearing .....	9
	2.6 Previous studies .....	11
3	RESEARCH METHODS .....	20
	3.1 The microphone setups .....	20
	3.2 Recording environment .....	23
	3.3 Recording the source material .....	23
	3.4 Listening setup and environment .....	23
	3.5 The questionnaire .....	24
	3.6 Participants .....	24
4	RESULTS .....	25
	4.1 Width .....	25
	4.2 Depth .....	26
	4.3 Ambience .....	27
	4.4 Naturalness .....	29
	4.5 The influence of the room .....	31
5	DISCUSSION .....	34
6	CONCLUSION .....	38
7	REFERENCES .....	40
8	APPENDIX .....	43
	8.1 The questionnaire .....	43

## 1 INTRODUCTION

We live in a world of surround sound. Our natural environment is a surround sound environment. Our hearing system has evolved to perceive sounds arriving from all around us. It is rather surprising that only relatively recently the surround sound has enjoyed an increased popularity among audio professionals and consumers. Admittedly, the surround sound has been around for decades, but only in the 1980's the 5.1 surround was developed to a movie sound standard (Holman, 2000). Later the same standard was adopted for digital television and music as well. The invent of the DVD in 1990's brought the surround sound to consumers. The increase in surround music recording has raised quite a few questions concerning the best possible equipment, microphone choice and microphone setup. Different setups are based on various psychoacoustical theories and aesthetic preferences.

From the available literature on concert hall design (Essert, 1997), psychoacoustics (Blauert, 1983; Campbell & Greated, 1987; Griesinger, 2000; Holman, 2000; Theile, 2001) and the application of psychoacoustic theories to surround sound recording (Griesinger, 2000; Holman, 2000; Theile, 2001; Williams & Le Du, 2000) one can conclude that the early lateral reflections are essential in creating a natural impression of depth, width and ambience in a recording. However, not all the research supports this view. On the contrary, it has been argued that the importance of early reflections has been overestimated (Neher, Rumsey & Brooks, 2002). Also, a claim has been made that the early reflections actually deteriorate the feeling of the captured spaciousness in the recording (Griesinger, 1998).

If uncorrelated early lateral reflections are indeed essential for creating a natural sounding imaging and ambience, and if these reflections will provide the important cues for perceiving the width and depth of the sound source, one would expect the microphone setups that use side pointing cardioids being superior to other setups. However, the

omnidirectional microphones are often described having more “open” or “airy” sound due to their 360 degree polar pattern. Would the use of omnidirectional microphones in surround microphone setups result a better ambience reproduction?

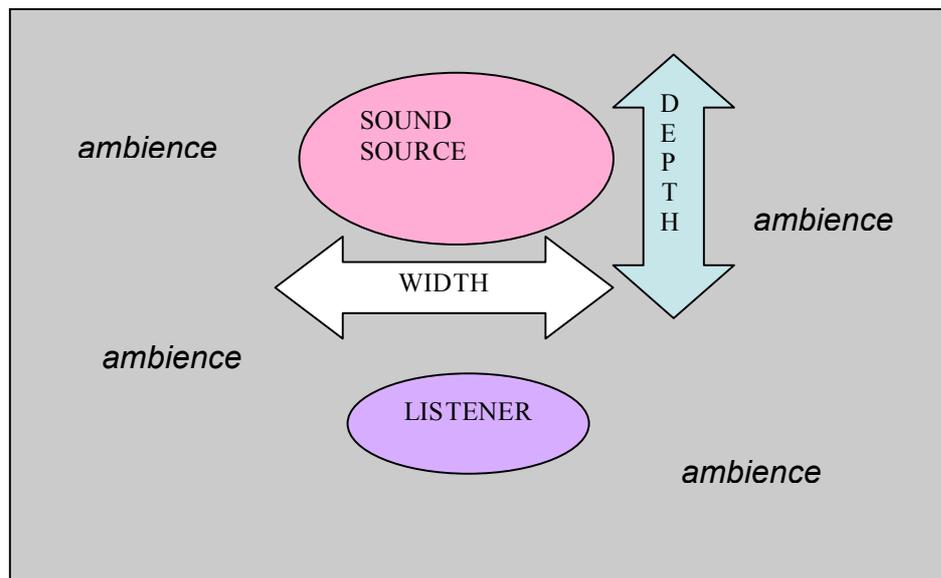
This research compared four different 5.0 surround microphone techniques in a listening test. The evaluated variables were sound source width and depth and the amount of ambience in the recording. Also, the naturalness of the overall sound was evaluated. Because the terminology can be problematic and the verbal taxonomies confusing these concepts were further defined in a following way:

**Width** - the perceived lateral width of the sound source.

**Depth** - the perceived distance inside the sound source.

**Ambience** - the perceived spaciousness around the sound source.

**Naturalness** - the perceived naturalness of the sound source



## **2 THE SURROUND SOUND**

The surround sound has been around for decades. The developments and the requirements in the movie sound have resulted the evolution of surround sound systems. The sound and the music in the movie mainly serve to enhance the viewing experience. For music reproduction first the mono and later the stereo sound were considered sufficient. The technical limitations of the home stereo systems and delivery formats long restricted the use of more than two channels. For example, the phonograph record can carry only two signals. And because of the engraving technique, the length of material on a side is restricted. When other formats were developed - the cassette tape and the CD - they followed the two-channel format, mainly because people were accustomed to listen music in stereo and most homes had systems with two loudspeakers.

### **2.1 Early attempts in surround sound**

The scientist at Bell laboratories experimented with surround at 1930's. They developed a three channel stereophonic system with left, center and right loudspeakers. The system did not use rear surround speakers, since the sound was reproduced in a large, reverberant room. The room acoustics provided enough reflected energy and hence created the feeling of envelopment (Holman, 2000). In 1938 Walt Disney wanted to add surround sound to upcoming animation movie, Fantasia. The Disney engineers developed a technology called Fantasound, a five-channel surround system that had three front speakers and two rear surround speakers. This resembled very closely the currently popular 5.1 surround system. The difference was that the Disney system used only three channels that were reproduced through five loudspeakers (Holman, 2000). The Disney system was only in use a short time with the movie it was designed for.

In the early 1950's the 20<sup>th</sup> Century Fox began using a four-channel system that had three front channels and a single surround channel. This system was used with many released movies at the time. Simultaneously, six-channel systems were used with the musicals *Oklahoma* and *South Pacific*. These systems used five front speakers behind the screen and a single surround speaker behind the audience. At the same time in Germany a system with several loudspeakers around the listener was introduced. Each loudspeaker received the same signal that was delayed in various ways. This created an illusion of an enveloping sound that surrounded the listener (Steinke, 1996).

## **2.2 The Quad and its successors**

In the late 1960's and early 1970's several matrix –based surround systems were developed. The Quad system was designed for music use and it derived four signals from an ordinary LP using phase and amplitude manipulation. The Quad was short-lived. The consumers were not ready to invest for four loudspeakers. The Dolby Stereo for movie sound was introduced in 1976. Like Quad, it could store four channels worth of sound in the space of two tracks. It used three front speakers and a single surround speaker. The Dolby Stereo had both wider frequency and dynamic range than the preceding systems. It was used with several blockbuster movies of the time, for example, *Star Wars* and *Close Encounters of the Third Kind*. It was soon realized that the low frequency reproduction in movie theaters was inadequate and a low frequency-only channel was introduced (Holman, 2000).

An improved version of the Dolby Stereo, The Dolby Pro Logic, was introduced in 1980's.

## **2.3 Towards standardization**

In 1987 the 5.1 surround was developed to a movie sound standard (Holman, 2000). Later the same standard was adopted for digital television and music as well. The invent of the DVD in 1990's brought the surround sound accompanying movies to consumers. The DVD was designed as a delivery format for home movies, but two competing

formats were developed later for music-only use. The Super Audio CD (SACD) was introduced in 1999 by Philips and Sony. SACD stores data in DSD (Direct Stream Digital) format. It uses single-bit delta-sigma modulation coding method which uses sample rate of 2,8224 MHz (Holman, 2000). A dedicated SACD player is needed to play the multi-channel SACD. However, it is possible to make a two-layer disc where the other layer is an ordinary CD, the other layer the stereo or multi-channel SACD. The competing format, DVD-Audio was introduced in 2000. DVD-Audio uses linear PCM coding method. It supports 24 bit word lengths and 192 KHz sample rate. A dedicated DVD-Audio player is needed to play the multi-channel audio. Recently, so-called “universal” players have been manufactured. These players have the capability to play back DVD-video, DVD-audio and SACD and CD disks as well (The Recording Academy’s Producers & Engineers Wing, 2004).

## **2.4 The surround systems today**

### **2.4.1 The matrixed and the discrete systems**

The surround sound can be derived in two ways. The matrix-based systems encode the multi-channel information in two tracks. In sound reproduction this information is recovered to multi-channel format with the use of specific matrixes, mirror image decoding processes. (The Recording Academy’s Producers & Engineers Wing, 2004).

For example, in Dolby Stereo the center channel signal was created by summing the left and right channels and then attenuating the resulting signal three decibels. The surround channel was the filtered and delayed sum of the left and right channels. The surround channel signal phase was also shifted 180 degrees (Holman, 2000).

The matrixed surround systems are still widely at use, for example in analog television broadcasts and computer and video games. However, the matrixed surround systems have inherent restrictions due to limited separation between channels and signal shifting problems that are caused by the complex phase correlations in the stereo signal.

The discrete systems use individual channels that each carry unique audio information. Each channel is assigned to a dedicated speaker, sometimes an array of speakers. The DVD-video, DVD-audio and SACD are all discrete systems. The Dolby Digital, DTS (Digital Theatre Systems) and MLP (Meridian Lossless Packing) encoding formats are discrete as well (The Recording Academy's Producers & Engineers Wing, 2004). Discrete surround sound is the best method for delivering music in multi-channel audio. However, one needs to keep in mind that there are sound quality differences between the discrete systems, because of the different data compression schemes used. For example, the Dolby Digital uses 11:1 compression, DTS uses 3:1 compression and the MLP used on DVD-A uses 1.85:1 data compression.

#### **2.4.2 The 5.1 surround set up**

The 5.1 surround system is made up of six channels – left, center, right, left surround, right surround and a LFE (low frequency effects) channel. The LFE is the .1 channel since it carries only low frequencies, approximately one tenth the bandwidth of the other channels. There are many other discrete surround systems today - 6.1, 7.1 and 10.2 - but the 5.1 system is a current industry standard specified in 1994 by International Telecommunications Union (ITU specification, ITU-R BS .775). Also, the 5.1 system is the most widely used music-only surround set up today. Music released in SACD and DVD-audio formats is designed for the 5.1 listening setup. Recently, the ITU specification has been criticized because it was done almost entirely with classical music in mind. Also, many technological and aesthetical developments in surround mixing during recent years have demonstrated that the ITU specifications may not be ideal for all music applications. (The Recording Academy's Producers & Engineers Wing, 2004)

#### **2.5 Spatial hearing**

The human sound localization ability is at best very accurate and depends on several auditory cues. The differences caused by the distance

between the ears and the filtering and the reflecting effect of the upper torso are called the Head Related Transfer Functions (HRTFs). Each individual has unique HRTF due to differences in head and ear shape and size (Blauert, 1983). The most important localization cues that can be extracted from the HRTFs are the Interaural Time Difference (ITD) and Interaural Level Difference (ILD) (Holman, 2000). For example, the human head is relatively small object compared to a sound having a wavelength of 80 Hz. This corresponds to the low E string on guitar. The wavelength is approximately 4,2 meters long and thus, is not shadowed or blocked by the head and the sound arrives both ears at the same level but in slightly different time. With low frequency sounds ( $f < 1500\text{Hz}$ ), the ITD cues dominate.

With high frequency sounds ( $f > 1500\text{Hz}$ ) the ILD cues - sometimes called Interaural Intensity Difference (IID) or Interaural Amplitude Difference (IAD) - are more important in sound localization. For the shorter wavelengths the head appears as a relatively large object that blocks the high frequency sound waves. Because of this, the ear that is more distant from the sound source will receive less sound energy (Holman, 2000). At higher frequencies the delay between the ears becomes ambiguous. It is important to bear in mind that the given 1500 Hz is really an approximation, a transition region, rather than a precise value. It is derived from the measured delay between the average distance of the two ears.

In a reverberant space the ITD cues have been found to dominate through the Haas effect (Wightman & Kistler, 1992). The hearing tends to concentrate on the angle of the arriving sound if the onset of the sound is clear. This is sometimes called the precedence effect. Also closely related to this phenomenon is The Law of the First Wavefront (Holman, 2000). For example, if there is a short delay between the signals of the two sound sources the signal that arrives at the ear first is perceived as the sound source location. The later arriving signal is suppressed by the hearing in the localizing process. This explains why a listener can easily localize sound in reverberant space.

Not only the ITD and ILD have effect on the perceived arrival direction of the sound. There are other mechanisms at work. The sounds arriving from the front have a different timbre than sound arriving from the side or from the back. The pinna of the ear reflects and filters sound differently depending on the arrival direction and the arrival height. The timbre of a sound source will also serve as a distance cue. The high frequencies attenuate faster than the low frequencies. Distant sounds are perceived as darker or duller than the sounds closer to the ear.

The ratio of direct to reflected sound is also an important range cue (Blauert, 1983). The direct sound usually arrives at the ear first, while the reflected sounds are delayed and attenuated. However, if the reflected sound energy is increased enough a summing localization comes in to effect. A phantom image is created somewhere between the direct and the reflected sounds and this is perceived as the sound source location. The stereo speaker system uses this phenomenon in creating the phantom center image (Holman, 2000). The importance of the early reflections in the sound localization was discussed in detail in the literature review.

## **2.6 Previous studies**

In *"Hearing of Music in Three Spatial Dimensions"*, Blauert (1983) describes the importance of decorrelation in lateral and rear reflections. This decorrelated reflected sound energy contributes to the transparency of a sound source, i.e. all the instruments in the orchestra are heard as separate entities in the sound field. Several uncorrelated reflections will enhance the perception of the width and depth of the sound source, the localization of the sound source and spaciousness around the sound source. Blauert mentions that several listening tests have confirmed that people prefer sound fields in which the ambience can be experienced. According to Blauert, the early lateral reflections have a strong impact on the perceived ambience around the sound source. The reflections from the median plane will not have same importance as the lateral reflections.

In fact, if the sound source is at the median plane, the simultaneous reflections at median plane will deteriorate the sound quality by adding coloration due to comb filtering. The early lateral reflections between 10 ms and 50 ms are most important for perceiving the spaciousness and enhancing the imaging. The shorter delay times are not effective since they can cause phasing coloration on the timbre of the sound source. Longer than 50 ms delay times are often perceived as a separate echo. Naturally, all these early reflections need to be lower in sound level than the direct sound to be perceived as natural sounding. Reflections 20 dB below the level of the direct sound can still create the feeling of spaciousness.

The timbre will also affect the perceived width and depth of the source. The increase in energy above 3000 Hz in the early lateral reflections will increase the perceived source width. The lower frequency increase in the early reflections will increase the perceived depth of the sound source. The theory of decorrelated early reflections has been used in concert hall designs and sound recordings in order to create natural sounding ambience and imaging for the listeners (Essert, 1997). Blauert explains that the spaciousness of the concert hall is dependant on the level of the sound source. The louder sound will create louder reflections and reverberation. This makes the acoustic designing of a concert hall very difficult.

In *"Spaciousness and Envelopment in Musical Acoustics"* Griesinger (1998) argues that the definition of spaciousness is problematic. For him the spaciousness equals to a sound field that gives an impression of a large and enveloping space. Griesinger uses the definition in a very different way than Blauert above. Griesinger explains that the early lateral reflections have a strong impact on the source width. The source width increases if the lateral reflections arrive during the rise time of a sound event. This affects the sharpness of the image. The image of the sound source is sharp if there are no reflections during the rise time of a sound event. Hence, the early lateral reflections actually decrease the impression

of imaging and spaciousness, because these reflections are perceived as being a part of the direct sound. The early reflections will color and lengthen the direct sound. If the length of the sound source is increased this way, less space is left between the repeated sound events and this results fewer opportunities for listener to perceive spaciousness. The sound events actually mask the spaciousness they have created. This contrasts the arguments made by Blauert (1983). Griesinger adds that when the early reflections arrive 50 ms after the end of the sound event an impression of a small room ambience is created. He argues that this impression is not spacious. The spaciousness is determined by the late arriving reflections, reflections that are delayed more than 160 ms.

In *"The Theory and Practice of Perceptual Modeling – How to Use Electronic Reverberation to Add Depth and Envelopment Without Reducing Clarity"* Griesinger (2000) criticizes the most surround microphone setups, because they rely on a relatively narrow sweet spot, i.e. the location of a listener is critical for perceiving the intentions of a recording engineer. Also, if people are listening in a large group, some people are seated outside the sweet spot - the optimal listening area - and will experience deteriorated version of the recording.

Griesinger argues that the microphone techniques that use main microphones in the hall and spot microphones on the stage, close to musicians, will create a wrong impression of the sound event. For example, in a typical orchestra recording the main microphones are placed 3-4 meters high behind the conductor. If the orchestra is a large symphony orchestra of approximately 100 musicians, those in the rear of stage (percussion, harp, brass) are too far way from the main microphones. The main microphones will pick only the reflected energy of those instruments. To overcome this problem the recording engineers usually place several spot-microphones on the stage, close to the instrument groups to get more direct sound on the recording. According to Griesinger, this creates several problems. The spot-microphones will pick sound earlier than the main microphones. For some instruments the

distance from the main microphones can be up to eight meters. This creates approximately a delay of 24 ms. When the spot microphones are mixed in with the main microphones this delay will create unsatisfactory coloration and will make the localization worse. It is, therefore, essential to come up with a microphone set-up or create other solutions to take all abovementioned facts in to account.

Griesinger proposes a setup that uses spot microphones that are decorrelated by use of carefully selected recording angles. Also the rear microphones should be separated from the main microphones by a distance of a hall radius. Hall radius is sometimes called a critical distance. A critical distance of a hall or a room is a measured distance from the sound source where direct and reflected sounds are equal in power. Griesinger suggests the use of digital reverberation during mixing to further decorrelate the surround channels. He recommends at adding both the early lateral reflections and late reverberation digitally to improve the sound. The early reflections should have a length between 20 ms and 50 ms. The late reverberation should start at 150 ms. Surprisingly, here Griesinger clearly supports the views portrayed by Blauert. He also contradicts his previous research by adding: “ Our studies of the perception of reflections show that early reflections that do not come from the same direction of the sound source can create a perception of distance and depth. If these reflections fall between 20 ms and 50 ms, they do not affect the intelligibility or clarity of the direct sound.”

In “*Multichannel Microphone Array Design*” Williams and Le Du (2000) have created an elaborate theory for obtaining full 360 degrees coverage with five microphones and simultaneously minimizing the cross-talk between the microphones. This theory includes the concept of critical linking. Critical linking enables the reproduction of the continuous sound field around the listener, despite the listener seating position within the surround speaker setup.

Williams and Le Du acknowledge the importance of uncorrelated early lateral reflections in creating a natural sounding ambience. The recording angles of the front, left and right microphones are hence critical in capturing the early reflections in an optimal way. The distance and the angle between the three front microphones must be carefully adjusted to obtain smooth, full coverage of the sound source. At the same time the cross-talk between the microphones need to be minimized. If these conditions are met the result will be a natural reproduction of the sound event with superior imaging qualities. Williams and Le Du use mainly cardioid and supercardioid -patterned microphones to obtain the directivity and to eliminate the cross-talk. They introduce seven table groups that present 220 possible surround microphone array combinations. All the arrays are based on following theoretical concepts:

*The Coverage Angle* – the total coverage of the stereophonic recording angle of a single pair of microphones making up a part of the total multichannel array.

*Negative Angular Offset* – negative angular offset is achieved when the coverage angle is rotated counter clockwise by moving the microphones.

*Positive Angular Offset* – positive angular offset is achieved when the coverage angle is rotated clockwise by moving the microphones

*Critical Linking* – critical linking is achieved when the stereo recording angles of both front left segment and front right segment are linked to produce continuous coverage of the frontal sound events.

Front left/right segment is the coverage angle created with left and center and right and center microphones.

In “*Multichannel natural Music recording based on Psychoacoustic Principles*” Theile (2001) covers the use of several surround microphone setups and argues that cardioid or super-cardioid microphones will give better results than setups using omnidirectional microphones. With the use of cardioids a minimized cross-talk interference is achieved. The traditional Decca Tree setup that uses omnidirectional microphones will capture too much ambience and reverberation in the recording. This will result a

diffused sound field with poor imaging qualities. Theile supports the views presented by Williams and Le Du (2000) and uses that theoretic background in designing his own microphone arrays.

Theile compares the localization curves achieved with three different front channel microphone setups – the INA-3 (Ideale Nieren-Anordnung) setup, Decca Tree and Optimized Cardioid Triangle (OCT). He presumes that if the microphone setup corresponds with the optimum localization curve the resulting recording will portray a natural and well-balanced distribution of the frontal sound sources. In closer analysis, both the INA-3 and Decca Tree reveal unwanted cross-talk properties and because of this, an unsatisfactory localization curve. This creates unwanted sound coloration and inadequate localization focus. The OCT front microphone setup uses a cardioid-patterned center microphone and super-cardioid side microphones that are faced to left and right at 90-degree angles off center. This setup results less cross-talk between all front microphones and a more satisfactory localization curve and more natural sounding recording. Theile argues that both the pickup patterns and the pickup angles of OCT setup will better capture the early lateral reflections than the INA-3 or Decca Tree setups, and that the reflections will be naturally uncorrelated. The conducted listening test supports the theory.

#### *The ORF Surround Listening test 2001*

The Austrian Broadcasting Corporation conducted a listening test in Vienna in 2001. Austrian Radio Symphony Orchestra performed pieces by Mozart and L. Berio. These performances were captured with seven different surround microphone setups. The recordings were then compared in two blind tests. First test had 18 participants, all experienced recording engineers. The second listening test had 14 students as test subjects. The participants were asked to evaluate the spatial presentation, the timbral characteristics and the spatial imaging of the recordings. These three main categories were further divided in more detailed attributes that were evaluated in a scale of 1 to 5.

This test took into account the several theoretical backgrounds in surround microphone techniques. At least one microphone setup from each following theoretical group was included in the test.

1. *The No Sweet-Spot Group*

This group uses a spaced omnidirectional main microphone array in front of the orchestra. A selection of spot microphones is also used. Often artificial digital delays and reverberation is added to the recording. The theories and practices of David Griesinger (2000) belong to this group. A Decca Tree microphone setup belongs to this group.

2. *The Sweet-Spot Group*

For example, the “Critical Linking” –theories of Michael Williams (2000) and microphone arrays derived from those theories belong to this group.

3. *The Natural-Illusion Group*

Microphone techniques that create clear and precise imaging and clearly captured lateral reflected energy belong to this group. For example, the theories and microphone setups introduced by Theile (2001) belong to this group.

4. *The Verisimilitude Group*

For this group, the goal is to duplicate the concert experience as accurately as possible. The enveloping quality of the microphone setup is considered very important. The use of dedicated surround microphones, dummy heads, convolution and wave field synthesis is preferred. Most of the dedicated surround microphones depend on some kind of digital encoding/decoding algorithm. The B-format is probably the most common.

The results of the listening tests showed that the OCT surround and OCT front with the Hamasaki Square (named after its inventor, Kimio Hamasaki) produced the best overall spatial impression. Also, the depth and width of the sound source were best reproduced with these

microphone setups. They were among the best when judging the quality of the timbre as well.

Surprisingly, the rather expensive special surround microphones did not fare that well. It was argued that their reliance on the specific digital matrixing procedures contribute to the results. Maybe the digital encoding/decoding parameters are not fully developed yet?

The fact that the special surround microphones did not do well is actually good news for those who cannot afford the special equipment. Satisfying results can be achieved with the microphones that are used on a daily basis in most recording studios.

In *“Effect of Rear Microphone Spacing on Spatial Impression for Omnidirectional Surround Sound Microphone Arrays”* Rumsey and Lewis (2002) tested a rear pair microphones of a 5.0 surround microphone setup. A pair of rear facing omnidirectional microphones were used to capture the ambience. The distance between rear and front microphones was kept constant, but the distance between rear pair was varied from two to five meters. Five different solo instrument extracts were recorded and the results compared in a listening test of fourteen participants.

The rear microphone pair was evaluated in terms of envelopment, naturalness and spaciousness. For each attribute, an evaluation scale from 0 to 10 was used. A multifactor ANOVA (MANOVA) test was performed on the obtained data to determine significant effects. The results show a clear preference for the spacing of three and four meters between the rear microphone pair. The envelopment was perceived best at the spacing of five meters. The spaciousness was perceived most open when the spacing was three meters. The four meters spacing produced the most natural sound. Hence, it was argued that the most enveloping sound is not perceived as being the most natural.

Another research to evaluate just the rear microphones of a surround microphone setup was by Mason (1999). In *“Microphone Techniques for*

*Multichannel surround Sound*” the author explained that previous research (Fukada, Tsujimoto & Akita, 1997) found a preference for rear facing cardioid-patterned microphones that were close to the front microphones over the distantly placed omnidirectional microphones. The author argued that this preference was due to more direct sound captured by the cardioid pair. When this microphone pair was reproduced at the rear channels the sound was more full than with the omnidirectional microphones. This timbral difference was due to amount of captured direct sound.

Mason compared three rear microphone setups - a rear facing ORTF pair that was placed close to the front microphones, same ORTF pair that was delayed 29 ms and a rear facing omnidirectional microphone pair that was placed ten meters behind the front microphones. Four different sound extracts were recorded with each setup. Then a blind A-B paired comparison type listening test was conducted. The participants were asked to rate the auditory events in terms of stereo image and spaciousness. The rating was done on a scale of 1-10. A rating of 10 meant that the auditory event could not possibly be any better.

The results suggest that a surround sound microphone setup which uses digital delay between the front and rear microphones will less likely suffer from stereo imaging problems. Also, the results of Fukada et al. were further confirmed. The rear facing cardioid pair that is placed close to the front microphones produced better spatial impression than distantly placed omnidirectional microphones.

### 3 RESEARCH METHODS

#### 3.1 The microphone setups

A piece of music was recorded with four different surround microphone setups in two different spaces - in the Jyväskylä University recording studio and a large lecture hall. All microphone positions and distances between the capsules were measured in centimeters according the figures 1 - 4. The microphones were at the height of 215 centimeters. The same height was used for all the four surround arrays. The microphone setups were:

1. *Decca Tree* in surround setup (Figure1). This setup has three omnidirectional microphones for the front channels and two rear pointing cardioid microphones angled at 110 degrees to capture the surround channels. Three AKG 414 microphones were used for the front array and Neumann KM 184 microphones were used for the rear pair.

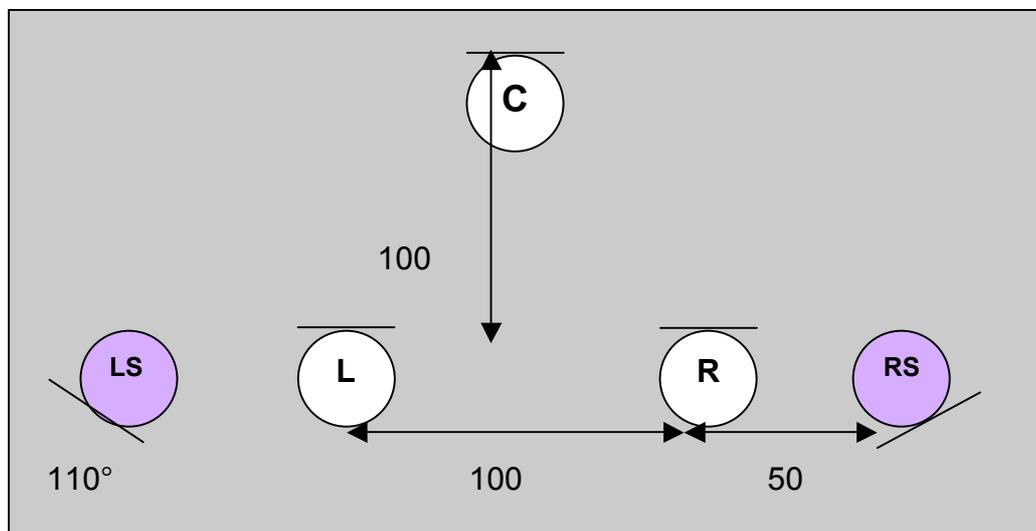


Figure 1. The Decca Tree surround microphone setup

2. *Williams 72*. A setup based on Williams recording angles (Figure 2). The array was named this way because of the three front microphones have a coverage angle of 72 degrees. The INA 5 array introduced in the literature part is actually a variation of this setup.

The microphones used were the same as in Decca Tree setup.

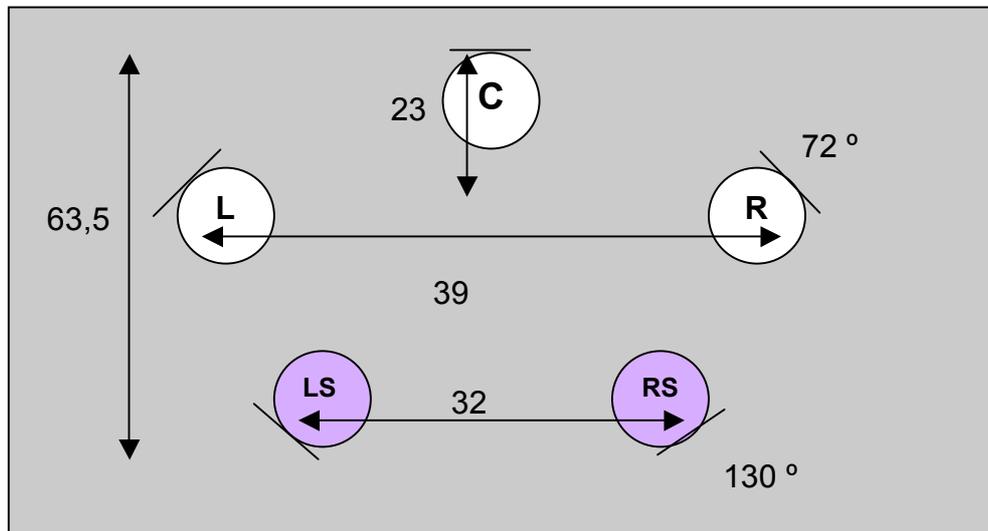


Figure 2. The Michael Williams 72 surround microphone setup

3. *Griesinger Surround Array* (Figure 3). This array has all five microphones in a row in front of the source. The center microphone was a cardioid AT4030 (an omni pattern can be used as well) and all the other microphones were omni patterned AKG 414 microphones.

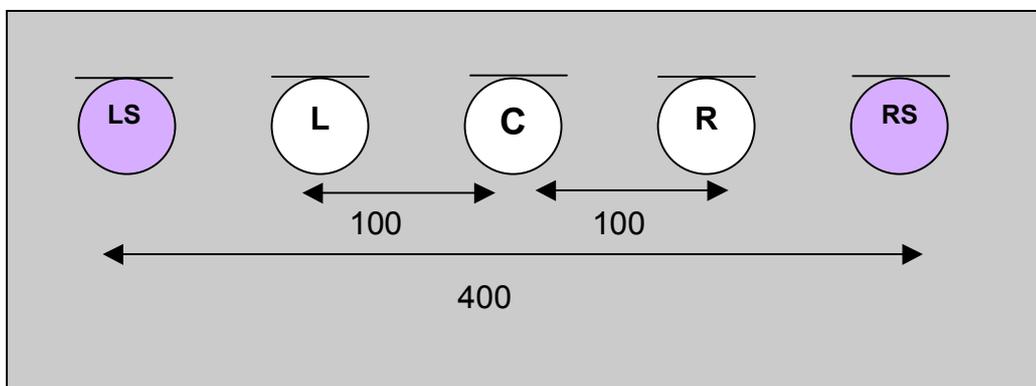


Figure 3. The Griesinger Surround Array

4. OCT with the Hamasaki Square (Figure 4).

All the front microphones are cardioids. The center microphone was an Audio-Technica 4033. Neumann KM 184 pair was used for the left and right microphones and they were angled  $90^\circ$  to the side. The four rear microphones of the Hamasaki Square are figure-8 patterned and angled  $90^\circ$  to the side. Four AKG 414 microphones were used for this. In the hall the Hamasaki Square was 250 centimeters behind the front microphones at the height of 260 cm. Very often when an orchestra is recorded in a hall or a church with good acoustics the microphones are placed higher, around 3-4 meters. In the studio the Hamasaki Square was altered due to the room size. The distance from the center microphone was reduced to 2 meters and the square size was 60 cm

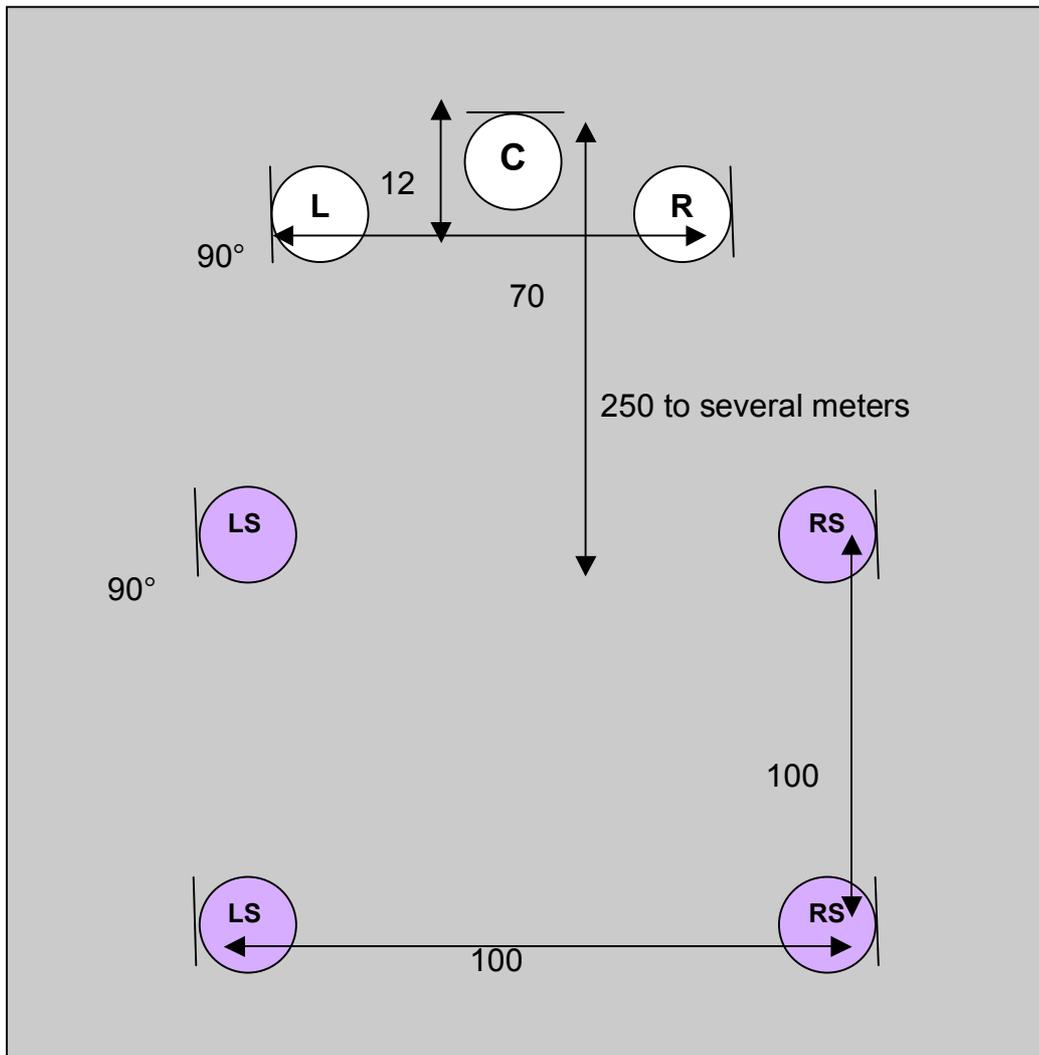


Figure 4. The OCT+Hamasaki Square

### **3.2 Recording environment**

The recordings were done at the Jyväskylä University studio recording room and music department lecture hall on 26th and 30th of October 2006. The studio recording room measures 50 square meters. The hall is a medium size lecture hall that has seating capacity of 350. It may not be optimal for recording music, but it provided a good comparison for the considerably smaller and drier acoustics of the studio recording room.

### **3.3 Recording the source material**

The four examples were recorded to Pro Tools 6.9 at 24 bit/44.1 kHz. All the recordings were done using Pro Tools remote preamplifier. The microphones were calibrated by playing pink noise from a test CD through a pair of Genelec 8030 loudspeakers at the sound source position. The tweeter distance between the speakers was 135 cm and the tweeter height was 110 cm from the floor. The preamplifier gain to the Pro Tools was set so that all the microphones were at the equal level, as close to zero as possible. The test signal was also recorded at the beginning of each example to help to calibrate the examples for the listening test. In all of the microphone setups the sound source was at the distance of 180 cm from the center microphone.

An excerpt of a previously recorded American traditional song *I Am a Pilgrim* played by David Grisman (mandolin) and Tony Rice (acoustic guitar) was used as a source material. This example was from CD David Grisman and Tony Rice: *Tone Poems - The Sounds of Great Vintage Guitars and Mandolins* (Acoustic Disc ACD-10, San Rafael, CA). The example was played through the same Genelec speakers used in microphone calibration. For the listening test the example was edited 2 minutes long.

### **3.4 Listening setup and environment**

The listening tests were done at the Jyväskylä University studio control room during 13th and 14th November and 13th December 2006. The control room measures 40 square meters and was built on the ITU-R BS.1116 specifications. It has Genelec 1037 surround monitoring system,

laid out according the ITU-R BS.775. The sound examples were measured with Minilyzer ML 1 SPL meter and calibrated to produce a 75 dBA peak level at the listening position. All the music examples were setup in Pro Tools in such a way that the participant could freely compare the examples by soloing each one at his own will. The participants were encouraged to spend as much time as needed with the examples.

The participants sat at the center of the room, one at the time at the optimal listening position. The lights of the room were dimmed and the computer monitors removed to help the participants rely solely on the aural cues.

### **3.5 The questionnaire**

The participants were asked to read the questionnaire (Appendix) prior to the test and then the test procedure was explained verbally. The participants evaluated the perceived width and depth of the sound source and the amount of ambience around the sound source. Also, the naturalness of the each example was evaluated. A discrete scale was used, values ranging from 1-5. For example, in the case of sound source width, a small value was given if the sound source was perceived narrow, a large value if the sound source was perceived wide. The recorded microphone setups were given a letter name in the following way: **A** (Decca Tree in the studio), **B** (Williams setup in the studio), **C** (Griesinger Surround Array in the studio), **D** (OCT with the Hamasaki Square in the studio), **E** (Decca Tree in the hall), **F** (Williams setup in the hall), **G** (Griesinger Surround Array in the hall), **H** (OCT with the Hamasaki Square in the hall).

At last, the participants were taught to use the Digidesign Control 24 control surface in order to audition the examples on their own.

### **3.6 Participants**

Sixteen volunteers participated in the listening test. Most of the participants were students from the Jyväskylä University music department. Also, some members of the staff took part in the listening test.

## 4 RESULTS

The data obtained was analyzed using SPSS program. Two-way repeated measures ANOVA test was performed on the collected data.

If significant main effects were found, further post hoc tests were done.

### 4.1 Width

Figure 5 displays the average perceived width values with 95% confidence intervals for each microphone technique and recording room. No significant effects of the used microphone technique were found. ( $df=3$ ;  $F=2.489$ ;  $p=.072$ ). The recording room had a significant main effect ( $df=1$ ;  $F=6.622$ ;  $p=.021$ ). The musical examples recorded in the hall were perceived wider sounding than those recorded in the studio. There was no significant interaction effect.

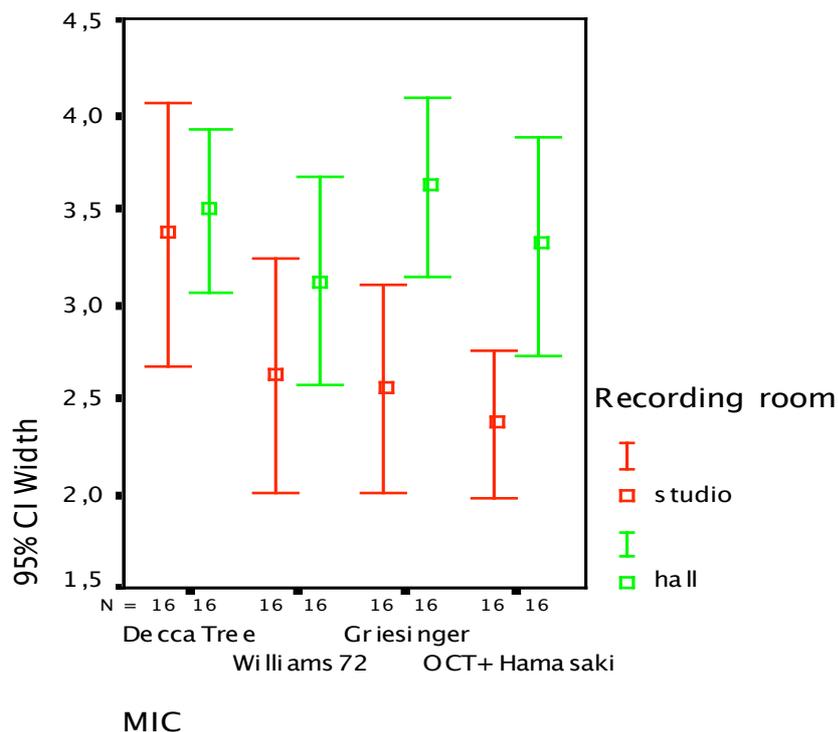


Figure 5. The Width values

Table 1. Tests of Within-Subjects Effects

Measure: WIDTH		Type	III	Mean	F	Sig.
Source		Sum	df	Square		
		Squares				
MIC	Sphericity Assumed	7.187	3	2.396	2.489	.072
Error(MIC )	Sphericity Assumed	43.313	45	.963		
ROOM	Sphericity Assumed	13.781	1	13.781	6.622	.021
Error(ROOM)	Sphericity Assumed	31.219	15	2.081		
MIC *ROOM	*Sphericity Assumed	4.406	3	1.469	2.233	.097
Error(MIC *ROOM)	Sphericity Assumed	29.594	45	.658		

## 4.2 Depth

Figure 6 displays the average perceived width values with 95% confidence intervals for each microphone technique and recording room. No significant effects of the used microphone technique were found. ( $df=3$ ;  $F=.288$ ;  $p=.834$ ). The recording room had a significant main effect ( $df=1$ ;  $F=8.206$ ;  $p=.012$ ). The musical examples recorded in the hall were perceived deeper sounding. There was also a significant interaction effect ( $df=3$ ;  $F=4.397$ ;  $p=.009$ ).

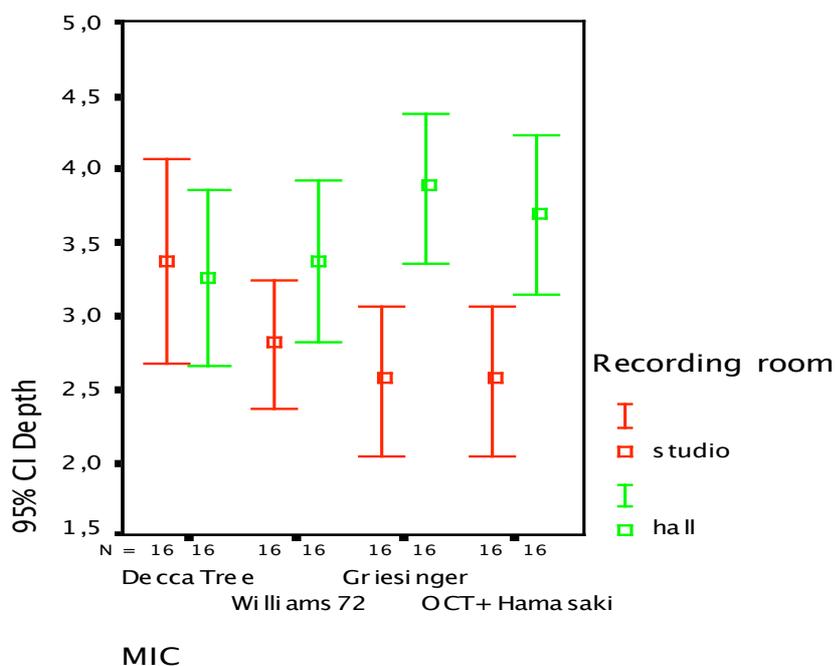


Figure 6. The Depth values.

Table 2. Tests of Within-Subjects Effects

Measure: DEPTH		Type	III	Mean	F	Sig.
Source		Sum	df	Square		
		Squares				
MIC	Sphericity Assumed	.937	3	.312	.288	.834
Error(MIC )	Sphericity Assumed	48.813	45	1.085		
ROOM	Sphericity Assumed	16.531	1	16.531	8.206	.012
Error(ROOM)	Sphericity Assumed	30.219	15	2.015		
MIC * ROOM	*Sphericity Assumed	10.031	3	3.344	4.397	.009
Error(MIC * ROOM)	Sphericity Assumed	34.219	45	.760		

### 4.3 Ambience

Figure 7 displays the average perceived width values with 95% confidence intervals for each microphone technique and recording room. A significant main effect was found on the used microphone technique (df=3; F=5.462; p=.003). The recording room had a significant main effect (df=1; F=9.558; p=.007). The musical examples recorded in the hall were perceived more ambient sounding than those recorded in the studio. There was no significant interaction effect.

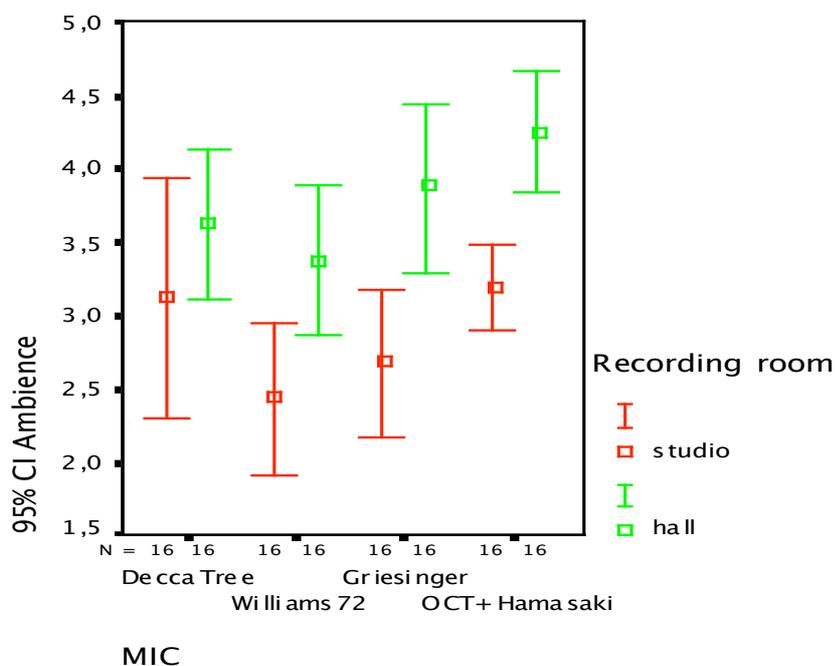


Figure 7. The Ambience values

Table 3. Tests of Within-Subjects Effects

Measure: AMBIENCE		Type	III	Mean	F	Sig.
Source		Sum	df	Square		
		Squares				
MIC	Sphericity Assumed	10.711	3	3.570	5.462	.003
Error(MIC )	Sphericity Assumed	29.414	45	.654		
ROOM	Sphericity Assumed	27.195	1	27.195	9.558	.007
Error(ROOM)	Sphericity Assumed	42.680	15	2.845		
MIC *ROOM	*Sphericity Assumed	2.148	3	.716	.759	.523
Error(MIC *ROOM)	Sphericity Assumed	42.477	45	.944		

A post hoc test revealed significant differences between Williams 72 array and OCT with the Hamasaki Square (p=.009).

Table 4. Tukey HSD post hoc test

(I) MIC	(J) MIC	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Decca Tree 2	Williams7	.4688	.25188	.250	-.1875	1.1250
	Griesinger	.0938	.25188	.982	-.5625	.7500
	OCT+Hamasaki	-.3438	.25188	.524	-1.0000	.3125
Williams7 2	Decca Tree	-.4688	.25188	.250	-1.1250	.1875
	Griesinger	-.3750	.25188	.447	-1.0312	.2812
	OCT+Hamasaki	-.8125	.25188	.009	-1.4687	-.1563
Griesinger	Decca Tree	-.0938	.25188	.982	-.7500	.5625
	Williams7 2	.3750	.25188	.447	-.2812	1.0312
	OCT+Hamasaki	-.4375	.25188	.309	-1.0937	.2187
OCT+Hamasaki	Decca Tree	.3438	.25188	.524	-.3125	1.0000
	Williams7 2	.8125	.25188	.009	.1563	1.4687
	Griesinger	.4375	.25188	.309	-.2187	1.0937

Based on observed means.

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

The Williams 72 array had the “driest” sound of all the evaluated microphone setups, whereas the OCT with the Hamasaki Square was the most ambient sounding.

#### 4.4 Naturalness

Figure 8 displays the average perceived width values with 95% confidence intervals for each microphone technique and recording room. A significant main effect was found on the used microphone technique (df=3; F=4.5; p=.008). Also, the recording room had a significant main effect (df=1; F=4.697; p=.047). The musical examples recorded in the hall were perceived more natural sounding than those recorded in the studio. There was no significant interaction effect.

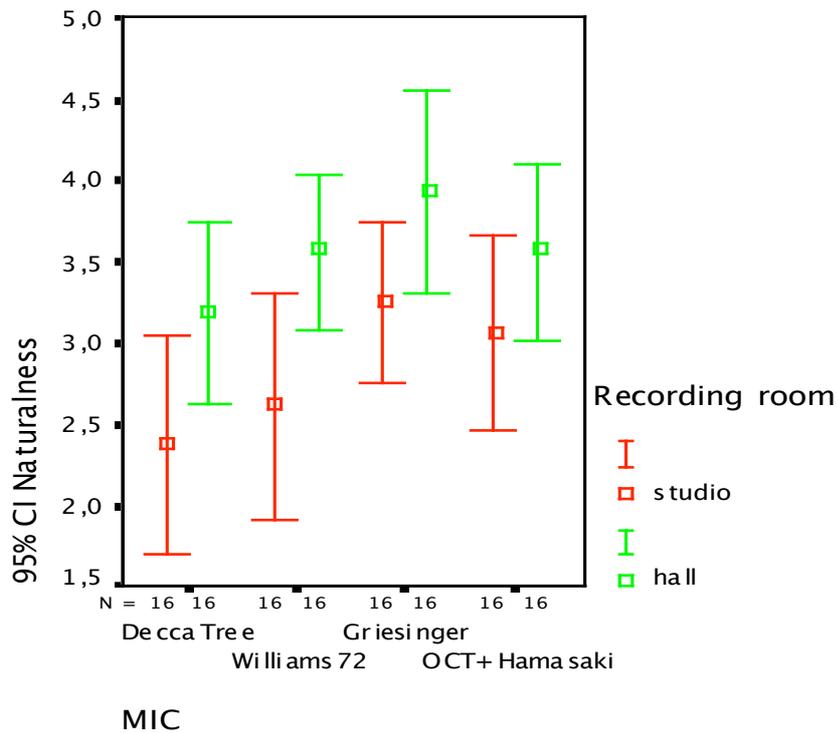


Figure 8. The Naturalness values

Table 5. Tests of Within-Subjects Effects

Measure: NATURALNESS		Type	III	Mean	F	Sig.
Source		Sum	df	Square		
		Squares				
MIC	Sphericity Assumed	11.336	3	3.779	4.500	.008
Error(MIC )	Sphericity Assumed	37.789	45	.840		
ROOM	Sphericity Assumed	17.258	1	17.258	4.697	.047
Error(ROOM)	Sphericity Assumed	55.117	15	3.674		
MIC *ROOM	Sphericity Assumed	.836	3	.279	.443	.723
Error(MIC *ROOM)	Sphericity Assumed	28.289	45	.629		

A post hoc test found significant differences between the Decca Tree and the Griesinger Surround Array (p=.020).

Table 6. Tukey HSD post hoc test

(I) MIC	(J) MIC	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Decca Tree	Williams72	-.3125	.27641	.671	-1.0326	.4076
	Griesinger	-.8125	.27641	.020	-1.5326	-.0924
	OCT+Hamasaki	-.5312	.27641	.224	-1.2514	.1889
Williams72	Decca Tree	.3125	.27641	.671	-.4076	1.0326
	Griesinger	-.5000	.27641	.274	-1.2201	.2201
	OCT+Hamasaki	-.2187	.27641	.858	-.9389	.5014
Griesinger	Decca Tree	.8125	.27641	.020	.0924	1.5326
	Williams72	.5000	.27641	.274	-.2201	1.2201
	OCT+Hamasaki	.2813	.27641	.740	-.4389	1.0014
OCT+Hamasaki	Decca Tree	.5312	.27641	.224	-.1889	1.2514
	Williams72	.2187	.27641	.858	-.5014	.9389
	Griesinger	-.2813	.27641	.740	-1.0014	.4389

Based on observed means.

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

The Decca Tree was perceived having the most unnatural sound, while the Griesinger Surround Array was perceived being the most natural sounding microphone setup.

## 4.5 The influence of the room

Since the statistical analysis revealed that the room had significant effect on all of the measured variables, it was interesting to look a bit closer at the room effect on each individual variable and microphone technique. The mean values are shown in the following figures. The results confirm the observations made during the recording.

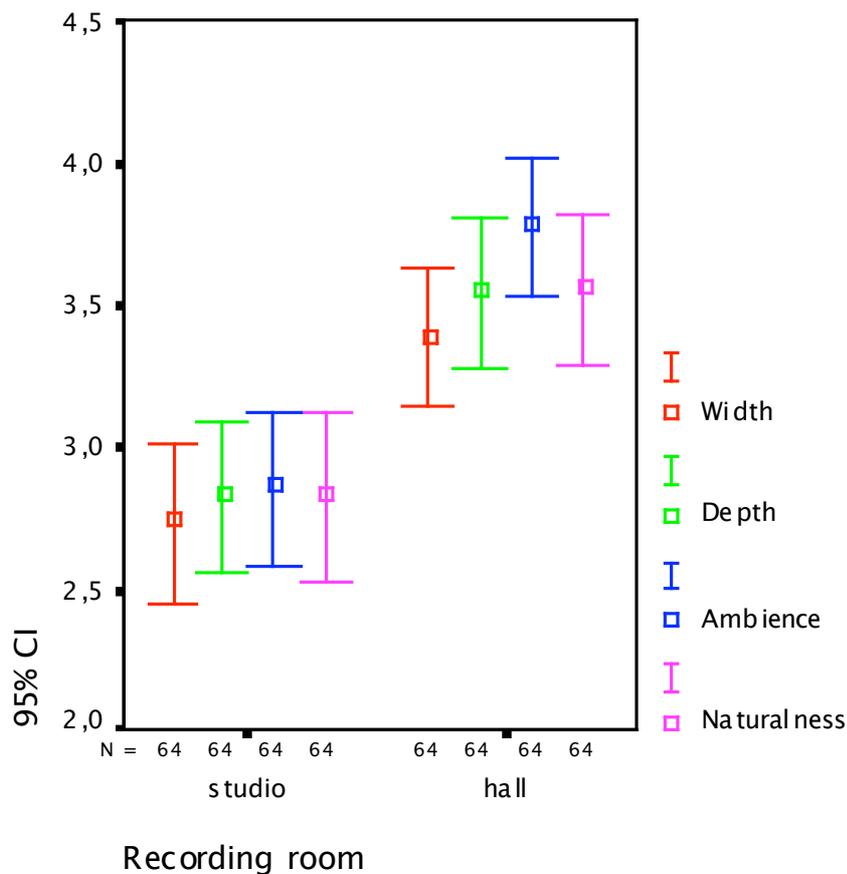


Figure 9. The mean values of each perceived variable for each room

Following figures (10 – 13) show more closely the recording room influence on each microphone technique. The hall recordings were perceived wider, deeper and more natural sounding than the recordings done in the studio. Also the hall recordings had more natural sound. The hall was preferred over the studio in almost every case. Only the Decca Tree was perceived having deeper sound in the studio than in the hall.

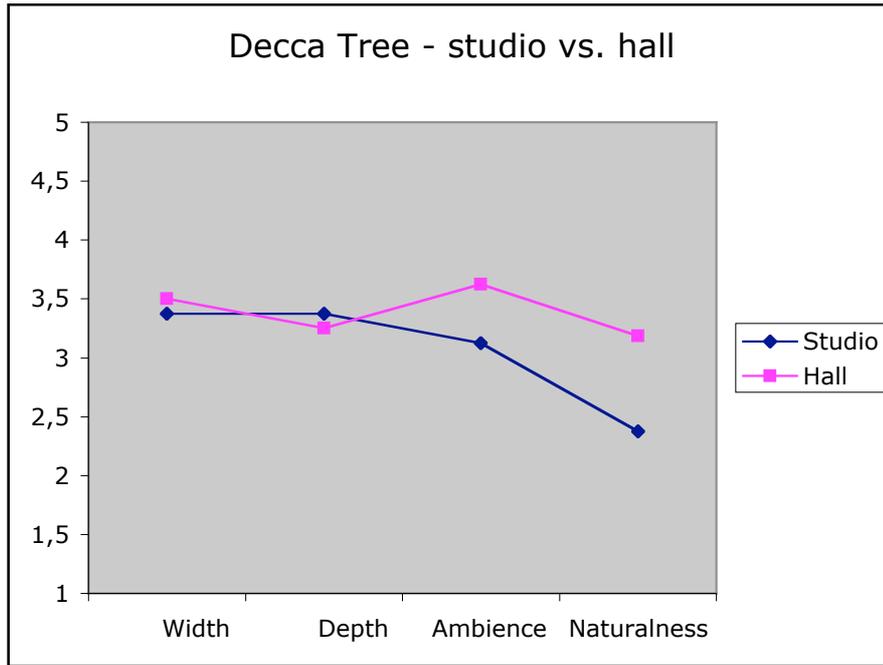


Figure 10. The influence of the recording room on the Decca Tree

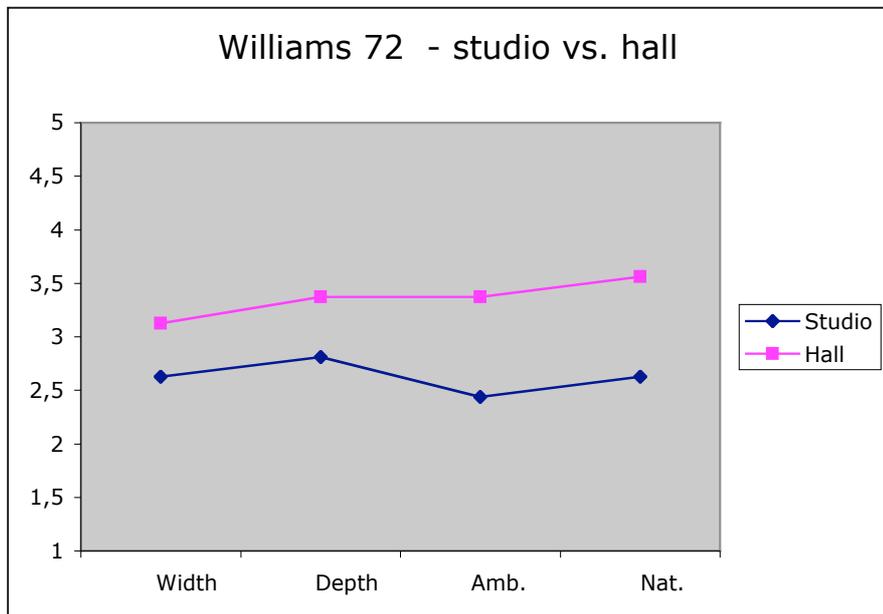


Figure 11. The influence of the recording room on the Williams 72

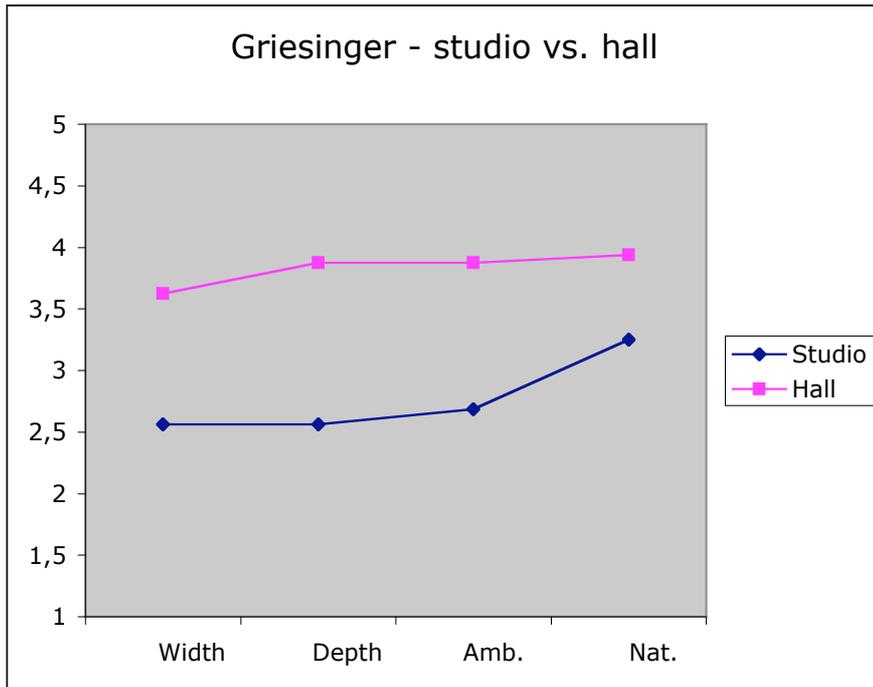


Figure 12. The influence of the recording room on the Griesinger Array

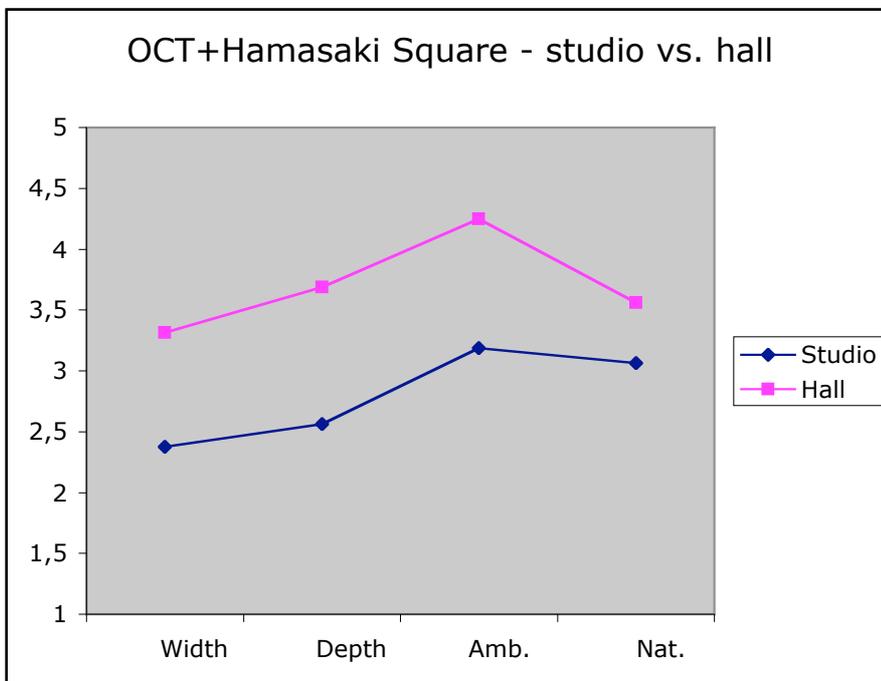


Figure 13. The influence of the recording room on the OCT+Hamasaki Square

## 5 DISCUSSION

The statistical data confirmed the observations made during the recording. The used recording space had a significant effect to the obtained results. The microphone setups recorded in the hall were perceived having wider, deeper, more ambient and more natural sound. Only the Decca Tree setup that was recorded in the studio was perceived having more depth than Decca Tree recorded in the hall.

The studio recording room was perhaps too small for these multichannel recordings. Since the walls are relatively close to each other there is too much reflected energy in the room. This causes smearing of the image and overall deteriorating of the sound. The studio ambience had a colored, "boxy" quality that was clearly revealed by the surround microphones. This coloration deteriorated the sound and maybe partly explains why almost all the excerpts recorded in the studio were ranked lower than those recorded in the hall. However, the studio room produced more consistent results in all the evaluated variables than those recorded in the hall (see figure 9). Again, this is what one expects from a room that was specially designed for recording music.

Overall, the hall space was better suited for these microphone arrays, simply because it had more volume. The hall had longer reverberation time than the studio and the timbre of the reverb was darker and warmer. The decay of the reverb was more pleasing to the ear. It is interesting that only Decca Tree got somewhat similar ratings on width, depth and ambience despite the recording space used. This may suggest that it is well adapted to varying recording environments and produces relatively consistent results.

Although it was not the purpose of this study to find the best surround microphone setup, some general characteristics of each microphone array were revealed. These are discussed briefly in the following.

### 1. The Decca Tree in the studio

The Decca Tree was perceived having the widest and deepest sound of all the microphone setups recorded in the studio space. It was ranked second when the amount of ambience was evaluated. However, it was ranked last when the naturalness of the sound was evaluated.

### 2. The Decca Tree in the hall

When the width variable was evaluated the Decca Tree was perceived having narrower sound than the widest set up, the Griesinger array. The Decca Tree recorded in the hall was perceived having the flattest and least natural sound. What explains this result? Why the Decca Tree was perceived being the least natural of all the evaluated setups? A post hoc test revealed significant differences between it and the Griesinger array. Did the timbre of the sound contribute to the results? Actually, during recording the examples it was observed that all the microphone setup had a different timbre character. A timbre analysis done on Matlab MIR Toolbox revealed that the Decca Tree in the studio was the brightest sounding microphone array. It had the highest Spectral Centroid and Spectral Roll-off values (figure14).

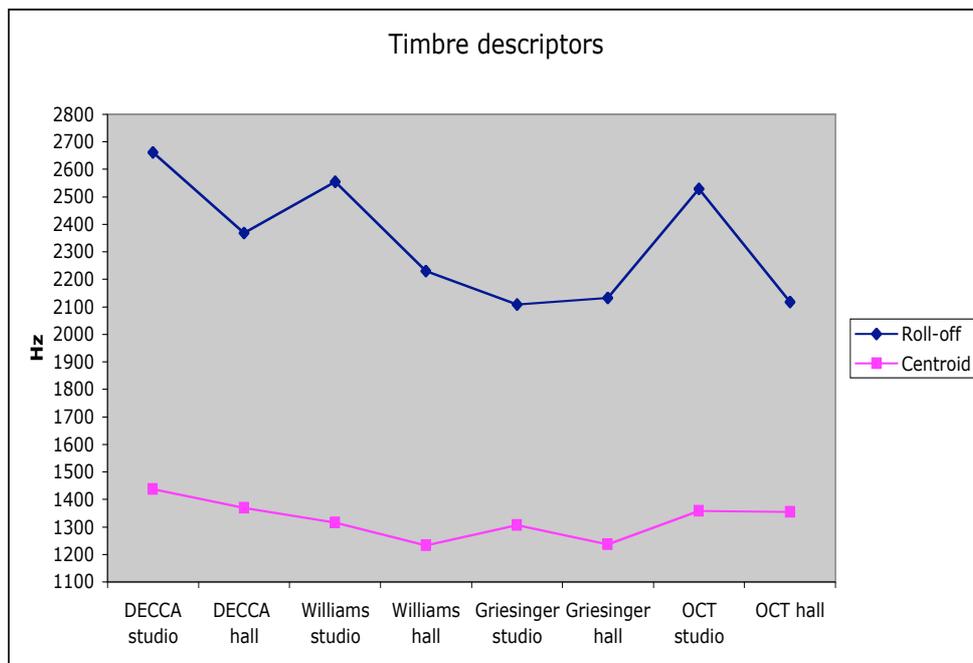


Figure 14. Mean Spectral Centroid and Spectral Roll-off values of the evaluated microphone techniques

Perhaps the Decca Tree sound was perceived being overly bright and hence least natural? In a future study it would be interesting to evaluate the timbre differences more closely and perhaps search for correlations between the timbre descriptors and the evaluated variables.

### *3. The Williams 72 in the studio*

The Williams 72 setup was based on the carefully calculated recording angles to obtain the best possible critical listening and least amount of cross talk between the microphones (Williams & Le Du, 2000). In the studio recordings it was perceived having the second widest and deepest sound after the Decca Tree. The Williams 72 array was ranked last in the evaluated ambience and the performed post hoc test revealed significant differences between Williams 72 and the OCT with the Hamasaki Square. These microphone setups clearly captured ambience differently. The Williams 72 array had the “driest” sound of all the evaluated microphone setups. This is not necessarily a negative thing and may suggest that it can be used in different situations and for recording different type of music than the other setups. It was perceived having not so natural sound, slightly more natural than the Decca Tree.

### *4. The Williams 72 in the hall*

The Williams 72 array performed better in the hall ambience. The obtained data revealed that it was clearly more sensitive to the space it was used in than, for example, the Decca Tree.

The sound of the Williams 72 recorded in the hall was perceived being wider, deeper and more ambient than the studio version, but it was still the driest sounding microphone setup. Also, it was perceived as being more natural sounding in the hall.

### *5. The Griesinger Surround Array in the studio*

The Griesinger Surround Array was perceived having rather narrow and flat sound. It and the OCT with the Hamasaki Square were perceived having the least depth of all the setups recorded in the studio space. The Griesinger Surround Array was perceived having a slightly wider sound

than the OCT with the Hamasaki Square. Also, it was not that ambient sounding either, but it was clearly the most natural sounding setup.

#### *6. The Griesinger Surround Array in the hall*

The Griesinger Surround Array clearly performed better in the hall space than in the studio. It was ranked first in width, depth and naturalness. The results may suggest that the “open” sound of omni-patterned microphones please the ear and contribute to the perceived naturalness of the sound. Perhaps the spacing of the microphones has also something to do with it? Maybe the width of the array creates longer level and arrival time differences between the microphones capsules despite the used omni pattern. Perhaps the ambience of the hall enhanced this character. A post hoc test revealed significant differences between it and the Decca Tree on the evaluated naturalness of the sound. Again, the timbre analysis revealed that the Griesinger Surround Array was the darkest sounding setup. When one thinks about the practicalities of each microphone array, the Griesinger array was the easiest and fastest to setup.

#### *7. The OCT with the Hamasaki Square in the studio*

The OCT with the Hamasaki Square was perceived having the narrowest and the flattest sound. It was perceived having the most ambient sound and when the naturalness was evaluated it was ranked second, after the Griesinger Surround Array.

#### *8. The OCT with the Hamasaki Square in the hall*

Like most of the microphone setups, the OCT with the Hamasaki Square clearly performed better in the hall than in the studio. Because of the sheer size of the setup, it needs relatively large space around it. It was ranked first when the amount of ambience was evaluated. This was expected from what was perceived during the recording. The design of the array explains the results. It uses four microphones for capturing the surround channels whereas other arrays use only two. The side pointing figure-eight -patterned microphones pick up efficiently the lateral reflecting energy thus enhancing the amount of ambience perceived in

the recording. And the distance from the front microphones helps to reduce the cross talk. The sound of the OCT with the Hamasaki Square was perceived rather natural. It ranked second when naturalness was evaluated. It suffered in the “tight” studio space and the results between the studio and hall recordings are considerable.

It is interesting to investigate the results keeping the used microphone polar-patterns in mind. The Decca Tree and Griesinger Surround Array used a combination of omni- and cardioid-patterned microphones while the Williams 72 array was all cardioid setup. The OCT with the Hamasaki Square used a combination of cardioid-patterned and figure eight microphones. Microphone frequency responses change with the chosen polar pattern. Generally, the omni-patterned microphones capture the low frequency energy more efficiently (Starck, 2002). This needs to be taken into calculation when speculating with the timbre differences between the setups in possible future studies.

## **6 CONCLUSION**

Above microphone techniques clearly require sufficient space around them. The sound of the room will affect the character of the sound source. This is nothing new, however. Any recording engineer knows through his or her experience that in order to get the best possible sound one needs to use the best possible source in best possible space with the best possible routing and processing. All the variables in a signal chain are interactive. In surround recording the space has even more importance than in the regular multi-track recording. It is advisable to find a best possible sounding space when recording is done in surround.

A church, a concert hall, a relatively large studio room or a decent sounding club environment would be ideal for recording multichannel audio. Naturally, the style of music and aesthetic goals of the production suggest the appropriate space.

It is important to notice that all the recorded examples were in their raw, unmixed form. One can argue whether this was the best way to evaluate the recorded examples. It is very rare that people get to hear and evaluate sound at this early in the record producing stage. Normally, the balance of the channels would be adjusted according to the taste of the mixing engineer. Often EQ, compression, delay and reverb is added to the channels to obtain the desired results. After the mixing the recording goes through even further subtle processing in mastering studio. The consumer gets to hear the finished product only after it has arrived at the store. Perhaps it would be a good idea to have a listening test done to the final, mixed and mastered version of the recording. This would surely relate more to the real life situations.

In the end one needs to keep in mind that the results obtained by a listening test are not conclusive. The results obtained here may give a reference to work with when applying the tested microphone techniques in real recording situation. Whatever the results are, it must be the ear and the experience of an engineer that dictate the technical choices in order to arrive at the desired aesthetic and artistic goals.

## 7 REFERENCES

Berg, J. & Rumsey, F. (2002). Validity of selected spatial attributes on the evaluation of 5-channel microphone techniques. 112<sup>th</sup> AES Convention in Munich. Convention paper 5593.

Blauert, J. (1983). Spatial hearing – the psychoacoustics of human sound localization. Cambridge, MA. MIT Press.

Campbell, M. & Greated, C. (1987). The Musician's Guide to Acoustics. London: J.M.Dent & Sons LTD

DPA Microphones (2003). A Surround Microphone Set up for Classical Music. DPA Microphones workshop on mic techniques for multichannel audio. AES 24<sup>th</sup> International Conference in Banff, Canada, 2003. Retrieved 4/11/05 [www.dpamicrophones.com](http://www.dpamicrophones.com)

Ekman, H. & Berg, J. (2005). The three-dimensional acoustic environment as depth cue in sound recordings. 118<sup>th</sup> AES Convention in Barcelona, Spain. Convention Paper 6513.

Essert, R. (1997). Progress in concert hall design – developing an awareness of spatial sound and learning how to control it. EBU Technical review, 1997.

Fukada, A., Tsujimoto, K. & Akita, S. (1997). Microphone Techniques for Ambient Sound on Music Recording. 103<sup>rd</sup> AES Convention, 1997. Preprint 4540.

Griesinger, D. (2000). The Theory and Practice of Perceptual Modeling – How to Use Electronic Reverberation to Add Depth and Envelopment Without Reducing Clarity. Tonmeister Conference. Hannover, Germany, Nov. 2000, Preprint.

Griesinger, D. (2000). Spaciousness and Envelopment in Musical Acoustics. Retrieved 3/11/05 from [www.world.std.com/griesngr](http://www.world.std.com/griesngr)

Holman, T. (2000). 5.1 Surround Sound – Up and Running. Boston: Focal Press

Hunter Starck, S. (2002). Live Sound Reinforcement. Boston: MixBooks

ITU-R, Recommendation BS.775-1. 1992-1994. Multichannel stereophonic sound systems with and without accompanying picture. International Telecommunications Union. Retrieved 5/5/06 from [www.itu.int](http://www.itu.int)

ITU-R, Recommendation BS.116-1. 1997. Methods for the subjective assesment of small impairments in audio systems including multichannel sound systems. International Telecommunications Union Radiocommunication Assembly.  
Retrieved 5/5/06 from [www.itu.int](http://www.itu.int)

Katz, B. (2002). *Mastering Audio – The Art and The Science*. Boston: Focal Press  
Mason, R. (1999). *Microphone techniques for multichannel surround sound*. University of Surrey. Retrieved 4/2/06 from <http://epubs.surrey.ac.uk/recording/3>

Neher, T. , Rumsey, F. & Brooks, T. (2002). Training of listeners for the evaluation of spatial sound reproduction. 112<sup>th</sup> AES Convention in Munich, Germany. Convention Paper 5584.

ORF Seminar (2001). ORF Multichannel seminar. Retrieved 2/23/06 from [www.hauptmikrofon.de](http://www.hauptmikrofon.de)

The Recording Academy's Producers & engineers Wing (2004). Recommendations for surround sound production. Retrieved 21/10/05 from [http://www.grammy.com/pe\\_wing/guidelines/index.aspx](http://www.grammy.com/pe_wing/guidelines/index.aspx)

Rumsey, F. & Lewis, W. (2002). Effect of rear microphone spacing on spatial impression for omnidirectional surround sound microphone arrays. 112<sup>th</sup> AES Convention in Munich, Germany. Convention Paper 5563.

Streicher, R (2003). The Decca Tree – It is not just for stereo anymore. Retrieved 21/10/05 from [www.wesdooley.com/pdf/surround\\_sound/decca\\_tree-urtext.pdf](http://www.wesdooley.com/pdf/surround_sound/decca_tree-urtext.pdf)

Steinke, G. (1996). Surround sound: The new phase, an overview. 100th AES Convention in Copenhagen, Denmark. Preprint 4286.

Swedien, B. (2003). *Make Mine Music*. Norway: Mia Musikk

Theile, G. (2001). Multichannel natural recording based on psychoacoustic principles. Retrieved 21/10/05 from [www.irt.de/wittek/hauptmikrofon/theile/multich\\_recording\\_30.oct.2001\\_pdf](http://www.irt.de/wittek/hauptmikrofon/theile/multich_recording_30.oct.2001_pdf)

Wightman, F. & Kistler, D. (1992). The dominant role of low-frequency interaural time differences in sound localization. *Journal of the Acoustical Society of America*, 91(3).

Williams, M. & Le Du, G. (2000). Multichannel Microphone Array Design. 108<sup>th</sup> AES Convention in Paris. Preprint 5157

Williams, M. (1987). Unified Theory of Microphone Systems for Stereophonic Sound Recording. 82<sup>nd</sup> AES Convention in London. Preprint 2466

## 8 APPENDIX

### 8.1 The questionnaire

#### MMT LISTENING TEST

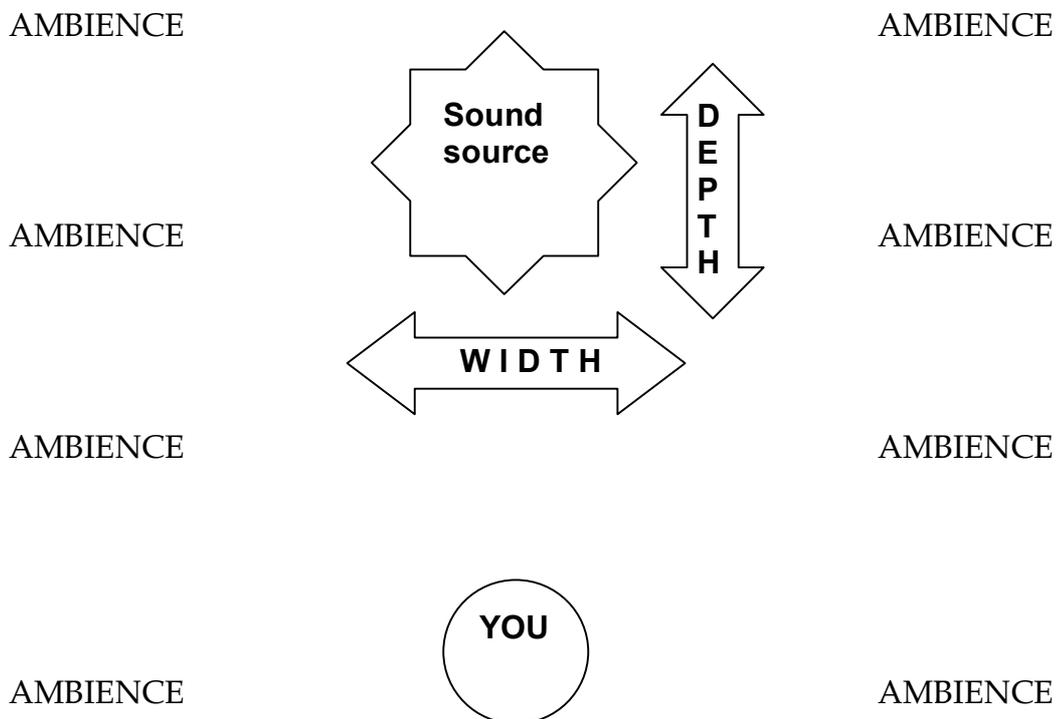
You will hear eight examples of a same piece of music. The examples are labeled as A, B, C, D, E, F, G and H. Your task is to evaluate the each example in terms of :

1. Sound source width
2. Sound source depth
3. The amount of ambience in the recording
4. The naturalness of the sound

You are asked to first listen and then rate the examples on the scale of 1 – 5. You may proceed at your own speed and listen as long as you please. You can compare the examples at your own will by pressing the A, B, C and D labeled buttons at the control desk. By pressing the space bar you may stop and start the music.

Remember, there is no right or wrong answer. Take your time, trust your ears and rate the examples how you hear them.

The illustration below helps to clarify what to listen for.



### 1. The sound source width

		Narrow				Wide
Example	<b>A</b>	1	2	3	4	5
	<b>B</b>	1	2	3	4	5
	<b>C</b>	1	2	3	4	5
	<b>D</b>	1	2	3	4	5
	<b>E</b>	1	2	3	4	5
	<b>F</b>	1	2	3	4	5
	<b>G</b>	1	2	3	4	5
	<b>H</b>	1	2	3	4	5

### 2. The sound source depth

		Flat				Deep
Example	<b>A</b>	1	2	3	4	5
	<b>B</b>	1	2	3	4	5
	<b>C</b>	1	2	3	4	5
	<b>D</b>	1	2	3	4	5
	<b>E</b>	1	2	3	4	5
	<b>F</b>	1	2	3	4	5
	<b>G</b>	1	2	3	4	5
	<b>H</b>	1	2	3	4	5

### 3. The amount of ambience

		Dry			Very ambient	
Example	<b>A</b>	1	2	3	4	5
	<b>B</b>	1	2	3	4	5
	<b>C</b>	1	2	3	4	5
	<b>D</b>	1	2	3	4	5
	<b>E</b>	1	2	3	4	5
	<b>F</b>	1	2	3	4	5
	<b>G</b>	1	2	3	4	5
	<b>H</b>	1	2	3	4	5

#### 4. The naturalness of the sound

		Unnatural			Natural	
Example	<b>A</b>	1	2	3	4	5
	<b>B</b>	1	2	3	4	5
	<b>C</b>	1	2	3	4	5
	<b>D</b>	1	2	3	4	5
	<b>E</b>	1	2	3	4	5
	<b>F</b>	1	2	3	4	5
	<b>G</b>	1	2	3	4	5
	<b>H</b>	1	2	3	4	5