Surround Sound Recording:
Is It Worth It?

Robert Anderson

Submitted in partial fulfillment of the requirements for the
Master of Music in Music Technology
in the Department of Music and Performing Arts Professions
in the Steinhardt School
New York University

Advisors: Kenneth J. Peacock, Robert J. Rowe

May 10, 2007

Foreword

I wish to express my sincere thanks to the NYU Viol consort, Professor Margaret Panofsky and William Monical, without whom this project would have been impossible.

I would also like to thank the esteemed faculty of the Music Technology Department for their help and guidance with this endevour.

The support of my family and my wife cannot go unacknowledged. I could not have achieved any of this without them being behind me to encourage me and to give me the stable foundation upon which to build.

Table of Contents

Introduction

The purpose of this paper is to explore a surround sound recording project of classical music from start to finish. The goal is to see, from the perspective of a recording engineer who is versed in on-location recording of two-channel stereo music, what sort of impediments there are to doing such a project (in addition to the usual difficulties of on-location recording work). Aspects such as the cost of additional equipment, added time and complication to the session, as well as the delivery and distribution of the final product will all be explored.

Part of this paper is philosophical as well. First, there is the philosophy behind the recording technique itself: shall we attempt to recreate a feeling of "concert realism" in the living room or should we attempt to use the additional channels more adventurously? Then there are the broader questions: is there really any point to doing a surround recording over a traditional stereo recording? Is the added value worth the extra effort and cost involved? Who is listening to these recordings and what are their expectations? How can we push the envelope in setting new standards for this type of recording?

The current state of the record industry in general and classical music recording in particular is not of consideration in this paper, although given these conditions at the time of this writing, one cannot expect that the already shrinking budgets for such projects will grow in the near future - a concern given the fact that a good stereo recording can be accomplished with a minimum of equipment, but a surround recording adds cost that will be more difficult for the engineer to recoup.

Section 1: Research

Sound and Hearing

In order to understand the science behind stereophony and three-dimensional audio, we must first have some understanding of how we localize sound in the real world, and how we can get an impression not only of the location of a sound source, but also the surrounding environment.

Albert Bregman gives my favorite illustration of the miracle of human hearing:

"Imagine you are on the edge of a lake and a friend challenges you to play a game. The game is this: your friend digs two narrow channels up from the side of a lake. Each is a few feet long and a few inches wide and they are spaced a few feet apart. Halfway up each one, your friend stretches a handkerchief and fastens it to the sides of the channel. As waves reach the side of the lake they travel up the channels and cause the handkerchiefs to go into motion. You are allowed to look only at the handkerchiefs and from their motions to answer a series of questions: How many boats are there on the lake and where are they? Which is the most powerful one? Which one is closer? Is the wind blowing? Has any large object been dropped in the lake?

"Solving this problem seems impossible, but it is a strict analogy to the problem faced by our auditory systems...We are not surprised when our sense of hearing succeeds in answering these questions any more than we are when our eye, looking at the handkerchief, fails." ([18] p.5)

Of course, there is a bit more to it than that. The human auditory system is a highly evolved combination of bio-mechanical sound-pressure transducers which are hard-wired to a psychological analysis stage. One of my previous papers deals with this subject in more depth, but here some of the physical components of hearing are of interest.

Our ears are separated by approximately seventeen centimeters. In between is an acoustic baffle that is approximately spherical in shape (a.k.a. the head). The ear canal itself can be modeled as a quarter-inch diameter pipe 1.25 inches long, terminated at one end by the eardrum. Around the opening of the ear canal is an oddly-shaped flap of cartilage and skin called the pinna. The resonance of the pipe coupled with interference of waves reflecting off of the folds of the pinna result in a distinct frequency response for each ear depending on the direction and elevation of the sound source.

A sound source is often modeled as a theoretical point source that emits spherical waves evenly in all directions. This model is not completely adequate as many sound sources are directional and radiate different frequencies with greater or lesser intensity in different directions. With greater distance, spherical waves flatten out into plane waves, which we localize differently. In addition, most sound recording (and sound reproduction) takes place in an indoor environment that has surfaces that can reflect or absorb sound with differing amounts at different frequencies. The reflections provide us with much information about the size and shape of the space, the materials that make up the surfaces, and in some cases can reinforce or interfere with our ability to localize the source of the sound. Objects that are larger than the wavelength of the sound can obstruct it, creating an acoustical "shadow."

The propagation of sound waves can be modeled by Huygen's principal, which states that "all points on a wave act as sources of circular waves (or spherical waves in three dimensions)... The intensity of the circular wave emanating from any point is the greatest in the direction of the original wave propagation and gradually decreases with angle, becoming zero in the opposite direction." ([19] p. 28) Another way of looking at this is that these individual wavelets could be considered as secondary sound sources. Looked at in this fashion, the wave could either have been emitted by the original point source, or by all the secondary sources along the edge of the wave front [20].

Localization

When a sound wave approaches from directly in front or behind, each ear receives more or less the same signal; that is, the loudness at each ear will be the same and both ears will hear the sound at the same time. In the case of a sound directly to the left of the listener, the wave will reach the left ear first and less than a millisecond later the right ear will receive it. In addition, the right ear signal will be reduced somewhat in intensity, compared to the left ear, mainly due to the acoustic shadowing of the head at higher frequencies. These two cues are referred to as interaural time differences (ITD) and interaural level differences (ILD) repectively. Also, due to the ear canal, pinna and the size and shape of the head, there will be differences in the frequency response of the two ears to any sound that is not directly "on-center."

One of our strongest cues for localization is time of arrival. One part of this is described by the law of the first wavefront (or the precedence effect): "when two coherent sound waves are separated in time by a very short interval (less than 28 ms) the first signal to arrive at both ears will provide dominant directional cues." ([13] p. 2.14) Even when strong reflections of the sound occur with the direct sound, it is rare that our ears are fooled as to where that sound came from, in fact these reflections often reinforce our perception of the source location. A related aspect is described in [21]: "We perceive this [time of arrival] difference at any change in the sound - a transient, a pause or a change in timbre. For this reason, we localize transients more easily than continuous sounds." ([21] p. 91) Lipshitz adds: "The level and time-of-arrival differences reinforce each other: a source to the left is louder in the left ear as well as reaching it first...For impulsive and modulated types of signals it appears that the time-of-arrival difference is the primary cue. " ([16] p. 14)

It is generally accepted that at frequencies above 1500 Hz, we use mainly interaural intensity differences to localize a source whereas below 700 Hz, we use mainly time of arrival differences to localize sound sources. This is mainly because the difference in phase of higher frequencies becomes relative and hard to determine since the wavelengths are small with regards to the size of the head, but where lower frequencies are concerned, we can reference the phase of the same wave cycle on both sides of the head [16]. Also, the head will acoustically shadow the higher frequencies and create an intensity difference between the ears. Between 700 and 1500 Hz, we use a combination of both interaural time differences and interaural level differences which would tend to reinforce each other. While it is widely believed that humans cannot localize sound sources below 100 Hz, this has been disputed and it is the opinion of this author that we can in fact localize low frequencies, but that this ability is somewhat dependent on the listening environment. Malham [22] also makes the case that there are methods of low-frequency perception that cannot easily be tested such as bone-conduction and chest cavity resonances.

Another important source of directional information comes from our ability to turn our heads. As we move our heads, we can minimize - or maximize - the interaural time and level differences (including the frequency dependent differences due to the head related transfer function). This may also be our main mechanism for determining whether a sound is coming from in front or behind [22], [21].

Without getting into the psychological and perceptual aspects, this is essentially how we hear things in a natural soundfield. Most of the time we take it for granted that we know where something is simply by hearing it. We can tell how far away it is, whether or not it is moving, even get an idea of the size of the object.

A bit about stereophony

Stereophony is an aural illusion that plays upon the aural directional cues that humans use everyday to locate sound sources in the real world:

"*Of course, it's our imagination, or more properly, our higher-level cognitive associations with previous memory of how these situations look and sound, that allows us to accept the vibrations of a two-inch transducer as a form of virtual acoustic reality.*" ([23] p. 15)

Even without reading the previous section, it is immediately intuitive that a sound source that is on the left side will appear louder in the left ear than in the right. Likewise, a sound from that source will reach the left ear before it reaches the right. Because of the spacing of the ears, the acoustic baffling provided by the head at higher frequencies, and perhaps most importantly, the learned memory of how the location of a sound source affects the relationship of time, loudness and frequency response between the ears, humans can make use of interaural level and time differences to ascertain the location of a sound source. Loudspeaker stereophony makes use of differences in the signals sent to two speakers to generate a phantom sound source between the speakers. A sound that is played louder or earlier at one speaker will create the impression that the sound source is located closer to that speaker than the others. A sound that is fed to each speaker at the same time and level will be perceived to have originated directly in between both speakers.

Stereophony is distinct from other techniques of three-dimensional sound reproduction, especially binaural reproduction, in that it is dependent on acoustic crosstalk between channels, unlike binaural reproduction which relies on complete separation between channels: "In stereo sound reproduction using two loudspeakers, each ear hears <u>both</u> loudspeakers, and is intended to do so." ([16] p. 4) We can expand this to five, six, seven or more loudspeakers.

The history of stereophony goes back over one hundred years, but it truly came of age in the 1930's. In the United States, Harvey Fletcher and the Bell Labs scientists were performing experiments, while across the pond at the same time, Alan Blumlein was also dabbling in stereophony for EMI. It is interesting to note that two distinct schools of thought arose out of each of these, and the dichotomy continues to this day.

The dichotomy arises out of two different theories on how a wavefront can be reconstructed. Both are based on the previously mentioned concept of Huygen's principal, namely that the propagation of a spherical wavefront generated by a single point source can be modeled as being made up of wavelets, each acting as a secondary point source. Bell Labs took a "macro" approach to reproducing this wavefront, whereas Blumlein took a "micro" approach.

The Bell Labs philosophy was as follows: if a listener could be separated from a sound source by a curtain of microphones, each one infinitely small and linear in its response, and this curtain of microphones were attached to a curtain of speakers (again, each one infinitely small and linear in its response), the wavefront would essentially pass through the curtains unchanged resulting in complete transparency from the sound source to the listener. Of course, such a curtain would be impractical to say the least, so the Bell scientists began to narrow it down a bit. First they eliminated the need for height information - thus the curtain could be reduced to a single horizontal line of microphones and loudspeakers. On this horizontal line, it was decided that a having a microphone/loudspeaker to the left and right of the listener, coupled with one in the center, would be adequate to convey a realistic sense of depth and transparency. In other words, by capturing and reproducing the wavefront across a broad area, one could produce the effect of a three dimensional sonic image with each sound source localised from left to right, as well as give an impression of its relative distance to the listener.

Blumlein had a different take on things. It was his belief that two loudspeakers were sufficient for the stereophonic illusion to work. Ironically, he was working on reproducing sound for film (ironic because the Bell Labs three channel frontal approach became the film standard and Blumlein's work ended up becoming the basis for two-channel home stereo - including the stereo phonograph disc). Blumlein believed that the best way to properly transmit meaningful directional cues to a person would be to capture the soundfield at a single point at the center of their head, and then regenerate those cues over the speakers. Theoretically, the wavefronts generated by each speaker would again intersect at this single point, reproducing the proper directional information at the listener's ears as though they were in the same room as the sound source.

These two different takes have spawned endless (and some might say pointless) debate, but have also generated different microphone techniques. The Bell Labs work inspired the use of two or more spaced microphones - most commonly omnidirectional in pattern - often referred to as AB stereo when two microphones are used. This type of technique essentially uses temporal cues between channels to generate a stereo image. Because of the random phase relationships between channels, the image tends to feel broad and diffuse, giving the psychoacoustic feeling of being in a large space also referred to as "spaciousness."

The famous "Decca Tree" technique is a specific spaced three-microphone technique that was invented at Decca Studios in 1954. Originally, it utilised three cardioid microphones - but eventually the standard became three omni's - with the center mic set somewhat in front of the two flanking microphones, using a natural precedence effect to generate a stable center image. This technique lends itself well to surround recordings of large ensembles since the use of three spaced microphones works quite well with the use of three front speakers.

Blumlein patented the technique that bears his name: a 90-degree crossed pair of bidirectional microphones aimed 45-degrees off-axis to the sound source, which generate discrete left and right channels. This technique captures differences in sound intensity due to the directional pattern of the microphones. In the same patent, Blumelin also made the case for its mathematical equivalent - the Mid-Side or MS technique.

For this technique, Blumlein also used coincident bidirectionals at a 90-degree angle to each other, but with one aimed directly forward and the other aimed directly to the side. The MS technique is an important deviation from all the other techniques in that it derives its stereophonic signal from a sum and difference matrix rather than from discrete channels. This signal can produce a number of equivalent directional patterns for the array, depending on the ratio of mid signal to side signal. Mixing equal amounts of mid and side signal with two bidirectional microphones theoretically yields the same pattern as a traditional Blumlein array. The side microphone must always be a bidirectional microphone, but any pattern may be used for the "mid" microphone, the most common implementation being a cardioid. Like the Decca Tree, this stereophonic technique has found applications in surround recording, but in a completely different way: it is the grandfather of the Ambisonic technique and there have been other implementations of the MS-system for minimalist surround arrays.

Importance of loudspeaker orientation

There are limitations to this stereophonic illusion. It only works if the listener and the loudspeakers are properly positioned. If the loudspeakers and the listener are in the proper positions relative to one another, and the program material has the proper amount of correlation, "phantom images" of sound sources will appear between the two loudspeakers. It is true that each speaker must receive a discrete channel of information. The key is that there must be some correlation between the signals sent to each loudspeaker, else the phantom image collapses or splits and appears to come from one or the other loudspeaker. Of course, if there is too much correlation, there is no sense of depth or localisation: the image will either appear to be very narrow or even "dual-mono."

In its most basic form, loudspeaker stereophony requires at least two speakers. It is the differences in amplitude and phase between the two loudspeakers that can fool our hearing system into believing these phantom directional cues. For example, imagine a listener sitting at the apex of an equilateral triangle formed by the listener and two speakers, with each speaker being the exact same distance away. If the same sound were to be produced in such a way that there were no differences between the two loudspeakers, each ear would receive the exact same sound at the same time at the same level. Our perceptual systems are hardwired to interpret this as a sound coming from directly in front, despite the fact that the sound is the result of two identical sources which are actually a certain distance to either side. A phantom image of the sound source will materialize (sonically at least) directly in between the speakers.

If both loudspeakers were to receive the exact same program material all of the time (total correlation between loudspeakers all the time), there would be no possibility of reproducing localization cues between the loudspeakers. Even with the listener in the proper position, all of the sources would seem to come from the same location between the speakers. This sensation can be experienced by playing a monophonic recording over stereo loudspeakers. There must be some difference between the loudspeaker signals to generate a sense of three-dimensional space,

Consider a case where the sound is slightly delayed to the right speaker. Since the sound from the left speaker will now arrive first, and according to the law of the first wavefront, we will perceive the source of this sound to be positioned more towards the left side. How far to the left depends on how much the signal is delayed. According to Bartlett [21], it takes a delay of approximately 1.5 ms to move a sound completely to one speaker or the other. Correspondingly, if you increase the loudness of the signal in one speaker relative to the other, we will perceive the sound as having come more from the direction of the louder speaker. How much more depends on how much louder it is. However, since the crosstalk between speakers is heard at both ears, this difference in loudness is apparently interpreted by our brains as an interaural time delay [21], [16]. It takes approximately a 15 decibel interchannel difference to move the phantom source all the way to one loudspeaker.

This is all completely dependent on proper angle and spacing between the loudspeakers and the listener. The ideal two-speaker stereo setup involves both speakers being equidistant from each other and the listener. Each speaker should be at a 30-degree angle inclined towards the listener. Different angles and different spacing will affect localization in the reproduced soundstage. This is an important point, since all of our current 5.1 standards involve much wider spacings and angles between the surround speakers, and a smaller distance and narrower angle between the three front speakers.

One problem with two-speaker stereophony is the fact that the phantom images in between the speakers are unstable, especially as the listener position moves towards one speaker or the other, the image will begin to "lean" towards the nearer speaker. Image distortion also results as the listener moves closer to, or farther from the speakers. The result is a fairly narrow "sweet spot" where the imaging and color are ideal. Every other listening position has to contend with less than optimal imaging and coloration of the sound source.

Another limitation of two-speaker stereophony is the lack of a true sense of depth and space. Much of our perception of depth and location in a concert hall setting comes from early and late reflections within the space itself [12]. According to Lipshitz:

"*it is not possible, with two channels alone, to fully place the listener into another imaginary environment... The best one can do is to provide the illusion that the end of the listening room* *between and beyond the loudspeakers has been removed, enabling the listener to "listen into" the original recording venue as if it were situated behind the loudspeakers....If more than two transmission channels are available one can do much better. Using three or more transmission channels in a properly designed surround sound system one can hope to place the listener within the ambience of the recording locale." ([16] p.6)*

We shall see in the next section whether this hope can actually be realised.

Surround Sound
So-called “surround sound” is an extension of the concept of loudspeaker stereophony. Instead of being limited to the front quadrant, surround stereophony positions loudspeakers around the listener in an attempt to achieve a more enveloping and immersive sonic experience.

Our current 5.1 "standard" was hardly the first attempt at enveloping or surrounding the listener with reproduced sound. Attempts at surround sound go back nearly as far as stereophony itself. Almost at the same time that Blumlein and the Bell Labs team were inventing stereo, Walt Disney was inventing Fantasound, expressly for his movie *Fantasia*. Fantasound was close to the current 5.1 standard in that it used three front channels and two surrounds. As with the contemporary Bell Labs and EMI developments, the work of the Disney engineers was way ahead of its time. Not only did they invent the concept of surround sound for film, they also invented the concepts of panning and multitrack recording. It was a long time before anything comparable was even attempted.

It wasn't until the 1950's and 1960's that two-channel stereo caught on with the public. In the 1970's a new consumer surround sound format emerged. Known as Quad, it involved four speakers surrounding the listener at 90-degree angles. Quad failed for a number of reasons but its weaknesses were clear: the angles between the speakers are too large for good stereo imaging, there was not yet a feasible delivery system, and the cost to the consumer of additional speakers and amplifiers.

Ironically, the first Quad format was 4-track reel-to-reel tape which was able to generate four discrete channels of audio information and it worked quite well. With a proper quadrophonic speaker setup, the results were quite good – for pop music recordings, it could have been a great listening format. Even for some classical applications, the Quad format could have added much to the listening experience. Perhaps the downfall of Quad was the fact that it was forced onto vinyl. This required matrixing four channels down to two channels, and then back to four at the listening end. One of the most confusing elements of Quad were the many competing matrixing standards. Few of these actually worked well in practice – in many cases the image was so vague and distorted that it ceased to be pleasant.

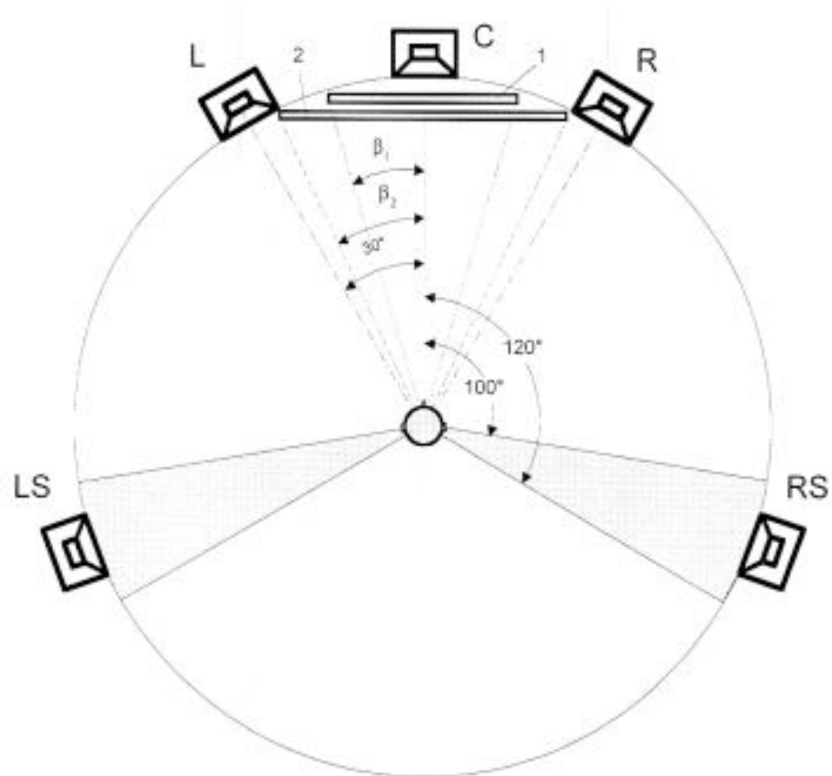
Quad could live on in the 5.1 standard. There are many applications in surround recording where the center channel is unnecessary – and could even be detrimental to the recording. There are many situations where a surround mix could simply use the front left – right and surround left – right speakers without utilizing the center channel.

Ambisonics also developed around the same time. Ambisonics is different from other surround systems in that it is matrixed rather than discrete and rather than relying on a specific speaker configuration, it can be decoded for different numbers of speakers in different layouts. Ambisonics lives on to this day, but in relative obscurity. It will be covered in greater detail later in this paper.

It wasn't until digital delivery systems came into being that surround sound could be brought into the home reliably. Once again, film led the way. In the theatre, three (or more) front channels were already the norm for film mixers, and experiments had been done using speakers in or around the audience for effect, but there was not yet a standard way of doing business for surround sound. In the 1970's, Dolby pioneered a matrixed delivery system on optical film where information for a center channel as well as a single surround channel could be derived from a stereo audio track. It was also not uncommon to have a dedicated speaker for low frequency effects. Eventually the 5.1 standard for film emerged, where there would be three channels for the front, two channels for the surrounds and a bandlimited low frequency effects channel. It was not until the introduction of the DVD that this format began to take hold for the common household. As the new HDTV standards begin to catch on, popularity of surround sound with picture - be it film or television - will continue to grow. So why has the audio industry lagged so far behind this trend?

Unfortunately, this paper cannot encompass all of the possible answers that this question arouses. However, as with everything else in audio lately, the possibility that we have two existing standards could be partly to blame. I am not only referring to the two conflicting delivery formats (SACD and DVD-Audio), but I am here referring to the two conflicting standards for surround sound speaker placement.

The generally accepted standard is ITU-R BS.775-1, which stipulates that the loudspeakers should be placed in a circle, each one equidistant to the listener. The left and right loudspeakers are the same as with two-channel stereo: angled at +/- 30-degrees, with a center loudspeaker in between. This makes the total front angle 60-degrees. The surround speakers are each 100 - 120 degrees off center, slightly above and behind the listener. This standard was developed with film in mind, where the surround speakers are rarely used for important sonic elements. The angle between the right front and the right surround is 70 - 90 degrees. The angle between the two surround speakers is 120 - 160 degrees. Both of these angles are too great for accurate stereo imaging, which means essentially that the surround speakers are not particularly useful for placing phantom sound sources around the listener, either between the front and surround speakers or between the two surround speakers themselves.



ITU-R BS.775-1 speaker placement

While this placement may not be ideal for stereo imaging to the sides and rear, it may be the best for giving the listener a sense of envelopment with a minimum number of loudspeakers. According to research done by Hamasaki, he found that it was possible to create a “minimally impaired diffuse field with only four uncorrelated channels and that the angles of the loudspeakers concerned corresponded closely to the ITU standard positions” excluding the center channel. ([31] p.77) The AES Staff Writer calls this a “convenient conclusion that appears to validate the choice originally made in that standard.” ([31] p. 77) However, there is an advantage to adopting this format which has nothing to do with audio quality.

According to Theile [12], "the 3/2 stereo format... provides easy programme exchange with film sound... [and] enables the sound engineer to exploit binaural cues more effectively than is possible with two-channel stereo, and thus to create a new dimension of spatial depth, spatial impression, and enveloping atmosphere." Williams admits to the limits of this format: "the slight increase in stability of the front sound stage is compromised by the addition of an inherently unstable rear segment" ([25] p. 2) but later writes that this is not so significant because it plays into the fact that humans have more trouble pinpointing sound sources in the regions to the sides and behind the head.

The National Academy of Recording Arts and Sciences has released an alternate recommendation for the placement of the surround speakers: "the P&E Wing recommends surround placement between 110 and 150 degrees, with the optimum range being 135 to 150 degrees." ([24] p. 2-3)

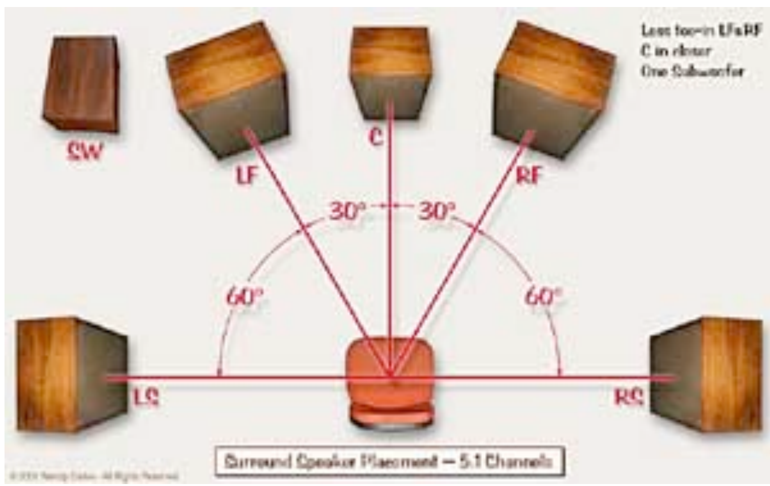
NARAS Recommendation

This wider placement of the surrounds almost completely sacrifices the ability to place phantom sources to either side of the listener, but this places the surrounds closer to the optimum 60 degree stereophonic angle behind the listener. In a presentation to the New York Section of the AES, Robert Auld made the point that localization of phantom images between two speakers to the side of a listener is nearly impossible since, due to the principals of stereophony, two ears have to be involved in the localization. In other words, the only way for a person to perceive a phantom source between the left front and left surround speaker is to turn their head to the left. To me, this indicates that the NARAS standard would seem more conducive to mixing and presenting music in surround due to the fact that there are now two viable stereophonic imaging sectors: one directly in front and the other directly to the rear, and that sacrificing the side imaging sectors is not a big loss. However, as with many things in audio, a good visual or geometric position does not always translate to a good aural result. Enter Wendy Carlos to the debate.

Wendy Carlos has done a great deal of experimentation with surround recording and mixing, and has her own recommendation for the placement of the surround speakers. One experiment she recommends is to play some music facing a pair of speakers as you would in stereo. As the music is playing, swivel your chair around so that the speakers are directly behind you. When you compare the imaging of the two, the image behind you becomes narrower and individual sources that were clear when you were facing the speakers become hard to localize. In my case, I tried this on a recording of a saxophone quartet that I had done and noticed that the low frequency instruments were most affected. Wendy Carlos' comes to the conclusion that: "we're interested in an optimum plan…for human surround sound. Since the back of our heads is not nearly as sensitive to sound directionality and nuance…we ought not to "waste" too much effort trying to obtain what we can't: a uniform sound field. That's where so many surround concepts fall down, assuming we humans can hear 360 degrees and follow it all accurately." [33]

Wendy Carlos prescribes the opposite of the NARAS standard; rather than rotating the surrounds further around to the back, they should actually be further forward. She actually prefers the surrounds to be directly to the side at +/- 90 degrees.

The largest angle she recommends is +/- 95 degrees from center: "The side channels still sound reasonably to the sides…but you'll find that the effects that you want to locate rearwards remain rearwards a little better…you don't want to go too far with this variation. More than 10-12 degrees rearward bias on LS and RS, and you create the same old problems." [33]



Source: http://www.wendy carlos.com/surround.html

I purchased a 5.1 home theatre system recently and I noted with interest the recommendation for speaker placement. Unlike the ITU or NARAS standards, the diagram shows the front speakers in nearly a straight line, with the center speaker set slightly forward of the left and right speakers.

With all of these conflicting recommendations, it becomes confusing for the engineer: how should one monitor in surround? One could argue that most consumers are oblivious to proper speaker placement and that this has no impact on the ability of the audio industry to deliver meaningful content in a surround sound format. I disagree because I feel that the placement of the surround speakers in the mixing or recording studio will effect the way an engineer treats these speakers: how important they are to the overall mix and what sort of content will be placed in them. If the engineer does not place anything of real importance in the surround speakers, or places things in the speakers that cannot be properly reproduced at the listening end, why should we be surprised that the consumer does not find surround sound recordings to have much added value over a CD?

I give the last word to Wendy: "Don't worry about some theoretical ideals of 'completely reconstructing the listening space.' You can't. Not even with 7.2 channels…Go back to the basics. Get the overall balances between the sounds right, that's not going to change. Set the equalization and wet-dry mods where they sound best, the same as usual. Let the reverb come mostly from the side channels, along with at least some of the instruments (don't waste LS & RS just on reverb/echo)…The transition to full surround is simply a move to a superset of everything else we already know, from creating, engineering and producing, to the "all-enveloping" new playback systems at home. We don't have to begin all over again. This is NOT rocket science!" [33]

Section 2: Surround Microphone Techniques

With the emergence of 5.1 as the de facto standard for surround sound, numerous microphone arrays and techniques have emerged. Some are descended directly from two-channel stereophonic techniques as well as tried-and-true film scoring techniques, others evolved with the surround specification itself. Even the most revolutionary approaches still have their roots in stereophonic techniques.

In a previous section the two most basic approaches to stereophonic recording were outlined, but almost all stereophonic microphone systems can be divided into one of three categories: spaced, near-coincident, or coincident. In addition, there is also the recording technique that evolved with multi-track recording: using close-mic's on each instrument and constructing an artificial stereo image using the pan-pot control. We find these same categories in the surround-sound field. Many are variations on the same theme.

Commonly, using a single stereophonic technique proves inadequate for a satisfying recording, especially when dealing with a large ensemble such as an orchestra. Often spot microphones are combined with a main stereo array to highlight soloists or lend presence or clarity to particular instruments, and additional stereo arrays may be used to capture sections of the ensemble that are too distant from the main array to be reproduced with enough clarity or balance. In addition, more microphones may be employed solely to capture ambience. We see similar approaches in many surround recording techniques.

Coincident Arrays

Coincident microphone techniques make use of inter-channel level differences to generate a stereo image. With any coincident technique, the aim is place the diaphragms of the microphones as close as physically possible, so that any time and phase differences between the channels are minimised. Directional microphones must be used to capture intensity differences based on the angle and distance of the sound source. In stereophonic applications, this leads to a clear, precise and stable image upon reproduction, and also has the advantage of being mono-compatible, often a concern in broadcast applications such as radio or television. The disadvantage is often a coloration of the sound due to the use of directional microphones (especially in the bass frequencies) sometimes characterized as a lack of "warmth" as well as a lack of "spaciousness" because there are no phase differences between the channels. Sometimes the reproduced image can seem narrow; i.e. it does not stretch all the way to both speakers. While there are not many examples of coincident surround techniques the few that do exist are based on the MS-technique of Alan Blumlein, namely Ambisonics and Double-MS arrays.

Ambisonics

Ambisonics is the only single point surround sound technique that has the potential to reproduce the entire sound field, including elevation and height information. As mentioned before, Ambisonics evolved around the same time as Quad but it never quite got off the ground. Why Ambisonics never caught on with the mainstream is a matter of debate, but as surround sound recordings become more common, Ambisonics is beginning to catch on with more engineers. The concept itself is actually quite simple: at its core, it merely expands Blumlein's Mid-Side technique to three dimensions.

As mentioned previously, the MS technique uses a bidirectional microphone for the "side" signal but can use any type of directional pattern for the "mid" signal. It is worth mentioning here that a cardioid directional pattern is derived mathematically by adding equal parts omni and figure-eight. Therefore, if we were to use an omni as our mid capsule and add equal parts mid and side, we would have the mathematical equivalent of two coincident cardioids at 180-degrees - in other words, we would have all of the information in a 360-soundfield encoded into left and right channels.

Now imagine a typical MS array with an omnidirectional "mid" microphone. The omni captures the absolute pressure component of the sound field at a single point while the bidirectional "side" microphone captures the pressure gradient from right to left. Now add another bidrectional microphone facing front to back. We now have two separate MS arrays each using the omni as the mid capsule - one can encode side to side and the other front to back. Then add a third bidirectional facing up and down. If you combine the three MS arrays, you have all of the possible directions captured by four microphones.

To better understand how the soundfield is encoded, it helps to know a bit of the math behind it. An ideal omni pickup pattern is equal to a unit circle where the magnitude at any angle is 1. An ideal bidirectional microphone has the pattern of cos(). In terms of Ambisonics, it's as easy as WXYZ. Thus, the position of any sound source within the three dimensional soundfield is encoded in these four signals which make up what is known as B format Ambisonics:

X = cosA.cosB (front-back)

Y = sinA.cosB (left-right)

$Z = \sin B$ (up-down)

$W = 0.707$ (pressure signal)

where A is the counter-clockwise angle from center front and B represents the elevation. [27] Recreating the soundfield is as simple as properly decoding these four signals, similar to recreating a stereo signal from an MS recording.

The Soundfield microphone is the only single-point surround sound microphone in existence. Realistically speaking, there is no way to get the omni and three figure-eight microphones to be truly coincident, so the designers of this mic took a slightly different approach. They used four cardioid capsules arranged in the form of a tetrahedron, each on the surface of an imaginary sphere about the center of the array. The signals from these four capsules are then electronically processed to duplicate the results of four truly coincident capsules.

There is no predefined speaker layout for Ambisonics, meaning that it can be dematrixed for any number of speakers in any layout (within reason). That being said, many recommendations call for no less than six speakers laid out around the listener and others recommend one or two more above for elevation. Of course, this also requires an Ambisonic decoder in the listening room, making it expensive and impractical for many end users. As a result, G-Format Ambisonics was proposed.

G-Format Ambisonics is basically Ambisonics that is pre-decoded for the standard 5.1 speaker layout. It is also encoded in such a way that the original Ambisonic information can be retrieved and decoded for other speaker layouts if the listener has the proper equipment.

Ambisonics is intriguing to me because of its versatility. If one were to pony up the necessary $10,000 for the basic system, you would have a microphone that could be used as a main mic in stereo recordings, yet have the capability to use it for surround recordings to fit all current and future standards since you could theoretically use it for both 5.1 and 7.1 situations. The ease of use is quite attractive as well: set up one microphone on one stand to do a surround recording.

There are a few downsides however. On the subject of flexibility, the same money would buy 5 or more Schoeps, Neumann or DPA microphones that could be used for surround arrays or spot mic's or any other purpose you could envision. While single-point surround and stereo is an attractive proposition in many cases, it may not be the way to go in every situation. If you have the Soundfield microphone, there is no other way to use it. Either it is your main mic or it sits in the case. One other complaint brought up by Jerry Bruck at an AES demonstration was the "lack of spaciousness" that is experienced with coincident techniques.

Dual-MS

The good news is that one can make a “poor-man’s” Ambisonic microphone array by using the principals of the Mid-Side technique. An Ambisonic microphone would capture the pressure component and three pressure gradients which, in microphone terms, translates to an omnidirectional microphone and three bidirectional microphones. Keeping in mind that very few playback systems can reproduce height information (a.k.a elevation), we can eliminate one of the pressure gradients. This leaves us with a single omni and two figure-eight capsules. In theory, one could use a Blumlein configuration with the omni capsule in between the two figure-eights and have two M-S systems – right-front and left-surround, and left-front and right-surround, with the omni being fed to the center channel if so desired.

An even better arrangement would be to break this down even further. If we were to replace the omnidirectional microphone with two cardioids – one facing forwards and the other aft, we could eliminate the front-back bidirectional mic and the remaining array would have two cardioids and a single figure-eight. This could then be matrixed very simply – one M-S system for front left and right and another for left-surround and right-surround, the front Mid capsule could be fed to the center channel. Both systems are completely separate and completely coincident.

The advantages to this approach are obvious: a single point array that is mono, stereo and surround compatible, and a 5-channnel surround recording can be made with just three microphones. Not surprisingly. Schoeps makes such an array with miniature capsules, which is quite small and easy to rig. This set (WSR-DMS LU Double M/S) will set you back a cool $8000.00. [29]

Chances are, most engineers will have these mic’s available to them already. It is just a matter of having the patience and ingenuity to come up with a way of setting up the array in a manner that is not visually obtuse and in such a way that it would remain stable while hanging in a concert hall.

Spaced Arrays

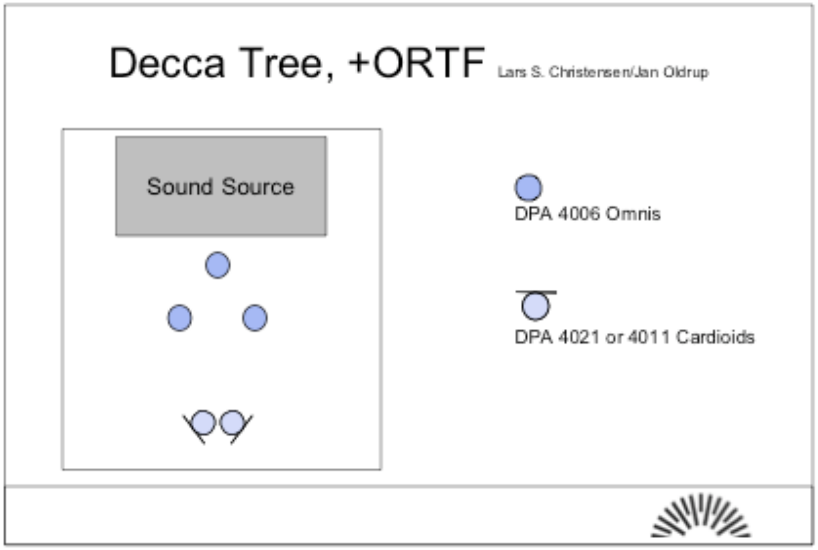
In two-channel stereophony, spaced microphone techniques mostly make use of inter-channel time differences to generate a stereo image. Generally, the image is somewhat vague, diffuse and sometimes even unstable, depending on the relative distance between the mic's and the overall stereo width of the ensemble. However, spaced techniques have the advantage that omnidirectional microphones can be used. Omni's offer perhaps the best "true color" representation of the recorded sound source, since they lack the coloration of (better said “lack of”) the bass frequencies that is often encountered with directional microphones. Also, due to random phase relationships between channels in these recordings, listeners often describe a feeling of ambience or "spaciousness.” For these reasons, spaced techniques have been the bread and butter of tonmeisters for decades in the orchestral recording world. They give an impressive feeling of size and space at the expense of a precisely articulated stereo image.

The use of spaced microphone arrays is quite well documented in the surround world, most of these taking their inspiration from the Decca Tree. This technique has been used in the orchestral and film scoring worlds for many years to give an impressive feeling of breadth while maintaining a rock-solid central image. For surround recording, it makes a great deal of sense to use this technique – the three microphones in the tree correspond well to three front loudspeakers. The easiest way to adapt this technique to surround is simply to hang two more microphones on the tree for the left and right surround channels, and this is essentially what Ron Streicher suggests in [30].

The Streicher paper goes a bit farther than that of course. Rather than the traditional three omni's, the author suggests substituting a Soundfield mic or an M/S array for the center microphone, and two Schoeps wide cardioids (MK21) for the left and right. On each end of the tree, a rearward-angled hyper-cardioid is used to record the surround channel. This technique is both flexible and pragmatic; flexible in that there are a number of options for the surround mix – especially if a Soundfield microphone is used as the center mic, and pragmatic in that it requires only one bar for all five microphones. This bar could be flown from a rigging point or mounted on a stand.

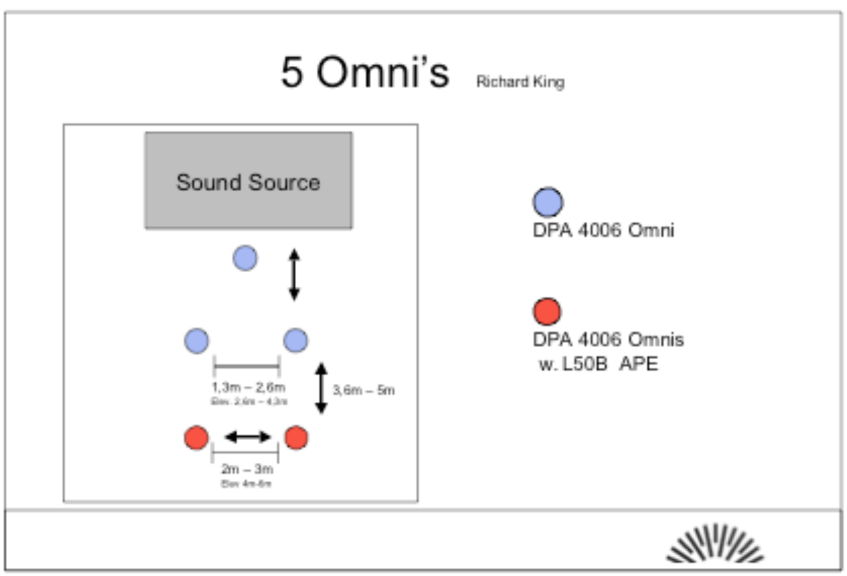
One concern would be the use of five different capsules to create a single image. In a "minimalist" technique, conventional wisdom dictates that matching capsules are preferred. When using three spaced microphones in two-channel stereo, one can often get away with using a different microphone for the center since it will be mixed equally to both channels, but in the case of discrete feeds to each loudspeaker, this could potentially be an issue. However, the author makes no mention of any such problems.

Other engineers have used the Decca-Tree successfully in surround applications. Lars Christensen advocates the use of a Decca-Tree with omni's for the front three speakers and an ORTF or widely-spaced A-B pair of cardioids for the surround channels. He states, "it is necessary to have a big physical distance between the front set and the surround set. If this is not the case you will get the impression [of] not having the orchestra in front of you due to the relatively strong content of direct sound in the surround mic's." ([9] p. 8) He recommends placing the surround mic's 8 – 10 meters from the Decca-Tree, which would normally be located at or about the critical distance. Christensen even advocates adapting the use of the Decca-Tree technique for small ensembles, by simply adjusting the size of the triangle to suit the situation: "[for] small ensembles like quartets 60 cm is adequate…"([9] p. 8). Jason Corey recommends a similar configuration, but suggests pointing the LS and RS cardioids towards the ceiling: "aiming the surround cardioid microphones to the ceiling has two advantages. First the direct sound from the ensemble is attenuated because it is arriving near the null of the polar pattern…The level of front-back coherence can be adjusted by changing the angle of the microphones and thereby controlling the amount of direct sound in the surround channels. Second, the often ignored vertical dimension of an acoustic space provides diffuse signals that are ideal for the surround channels.” ([9] p. 12)



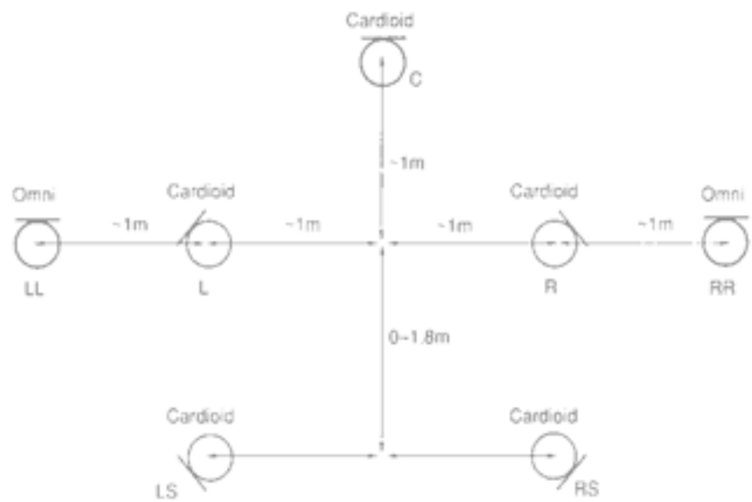
Source DPA [9] p.4

Richard King advocates the use of five omni’s: a Decca Tree in the front and two omni’s in the rear: “The five microphone configuration utilizes low frequency correlation that is offset by the natural time delay which relates to the distance from the front L/R to the rear L/R. This front to back relationship is however, de-correlated in the mid and high frequencies, which keeps the image of the ensemble in the front speakers, even if the listener moves out of the sweet spot towards the back of the listening environment.” ([9] p.13)



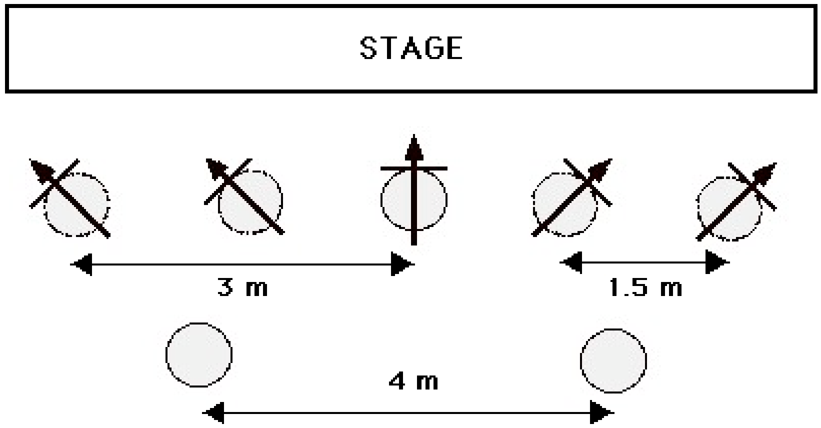
Source DPA [9] p. 5

Another spaced-microphone technique is the Fukada Tree, which is based on the Decca Tree, but takes the concept a step further. Firstly, the Omni’s in the tree are replaced with cardioids. A pair of flanking omni's is added to this array, and two rearward angled cardioids generate the surround feeds. Unlike the Streicher array, these surround mic's are usually mounted separate from the main array. The use of five microphones of the same type is another difference, however, using cardioids for both the front and rear could theoretically lead to some image instability. In the Fukada Tree, we encounter a technique that uses seven microphones to achieve its effect, instead of the minimalist five. Ostensibly, the flanking omni's would capture more of the low frequencies which would be lacking with the cardioids, and might capture additional ambience, giving the engineer some ability to blend the front and rear parts of the array.



Fukada Tree

One spaced-microphone array that is not based on the Decca-Tree idea, was proposed by Hamasaki. It involves five spaced supercardioids across the front of the orchestra in a line, about 1.5 meters apart. Two omnidirectional microphones are placed slightly behind the line of supercardioids about four meters apart. These omni's are run through a lowpass filter and are used "to compensate for the poor LF response of the supercardioids." ([31] p. 78) This array is then combined with another array placed 2 to 10 meters back in the hall to capture ambience.



Source [31] p. 78 – Hamasaki proposal for recording orchestra in a concert hall

Near-Coincident Arrays

Widely spaced microphones generate a robust image, perhaps not the most accurate, but impressive in size and rich in ambience. These techniques work well on large ensembles. However, there are situations where a widely spaced technique would not be effective, such as a recording of a small group or any scenario where image accuracy, rather than trying to impress the listener with a wall of sound, is more important. The aim with a near-coincident technique is produce a more natural sonic picture.

In two-channel stereophony, near-coincident techniques are a compromise between the intensity stereo and spaced-microphone approaches. A small spacing between the microphones is used, often imitating the spacing between the ears of a listener (about 8 – 12 inches). Directional microphones are usually employed in these techniques (ORTF and its many variants), although omni's can also be used with success (for instance, a small AB or the OSS technique). Often, the microphones are placed at an angle to take advantage of inter-channel level cues resulting from the directional pattern of the microphones. This "middle-of-the-road" approach seeks to preserve some of the "spaciousness" of the spaced techniques while preserving the clarity and stability of coincident approaches. There are a few examples of near-coincident surround arrays, notably the Jerry Bruck adaptation of the Schoeps Sphere microphone, the OCT developed by Gunther Theile and the many arrays that were presented by Michael Williams. The latter two are of most interest to me because these techniques do not require much in the way of additional equipment for an engineer who is used to working in stereo.

In the realm of two-channel stereophony, near-coincident techniques represent the “safe and quick” category since they almost always produce a decent sounding result, even in less than optimal placements. They also represent a single-point pickup that can be applied under almost any recording conditions. The image is usually accurate and stable and these techniques can be used with an ensemble of almost any size with good results.

When it comes to the world of surround recording, it seems that near-coincident techniques have proven to be the most difficult. Near-coincident microphone arrays expose many of the weaknesses of five-speaker stereophony, since the very aim of using a near-coincident technique is to achieve a natural representation of the original sound source. The holy grail of surround recording would be one technique that could be rigged at a single point and would be reliable in almost any recording situation, reproducing the sound field around the listener with an accurate, clearly-articulated and stable image, while still giving a sense of ambience and spaciousness. Unfortunately, even some of the best minds have had trouble coming up with a way to rig five microphones in such close proximity without side effects.

Michael Williams, perhaps best known for his “Stereo Zoom” concept, where he approaches microphone angling and spacing based on a desired Stereo Recording Angle (SRA). Williams defines the SRA as “that sector of the sound field in front of the microphone system which will produce a virtual sound image between the loudspeakers.” ([32] p.5) The SRA of a recording could be considered analogous to the framing of a photograph: the angle between the extreme left and right of the stereophonic image is considered to be the SRA. Sound sources that are at or beyond this angle will be reproduced at the extreme left or right speaker. His paper demonstrated that there are a number of combinations of angle and spacing between microphones for a given SRA, depending on the pickup pattern of the mic’s being used. He does not take into account image distortion of sound sources between the speakers, but shows angle and spacing combinations for microphones of a given pattern that would generate equivalent SRA’s.

Williams takes this concept to the next level in the field of surround sound with a concept he calls Multichannnel Microphones Array Design (MMA). Williams states that “the basis of the the process of Mulitchannel Microphones Array Design is the division of the sound field into individual segments, where each segment is treated as a separate entity…” ([25] p. 2) and each segment is defined by the loudspeakers on either side of the segment (e.g. one segment would be between front left and surround left, another would be defined by surround left and surround right). Essentially he breaks up the 5-speaker surround field into 5 stereophonic pairs of speakers and then considers each segment as a separate imaging sector. In other words, we are concerned with generating phantom images only between each pair of adjacent loudspeakers.

The main factor to consider in each segment is the linearity of reproduction within that segment, a concept that Williams calls Segment Reproduction Linearity [25]. Since the angle between each pair of speakers is different, the linearity of reproduction in each segment will be different: “The linearity of reproduction in each segment…is mainly determined by the angle of reproduction within each segment.” ([25] p. 2) As an illustration of this, he shows how, in two-channel stereophony, the relative distance between the listener and the loudspeakers changes the angle of reproduction. As the angle increases from 30-degrees, the sound sources that are reproduced in the center of the speaker pair are “crushed” towards the left and right speakers, essentially stretching the center. However, the SRA does not change: the sources that were located on the extreme left and right will still be located at the extremes, regardless of how large the angle becomes. Likewise, anything located “dead-center” will still localize directly between the speakers.

Applying this concept to our 5-speaker layout as prescribed by ITU 775, the linearity of reproduction in the front triad will be very good, since the angle between speakers is small. The side segments have a reproduction angle of 80-degrees, so we can say that the linearity in these segments will be poor – the image will be stretched between the speakers. Likewise, the rear segment will exhibit poor linearity since the angle of reproduction is 140-degrees. Williams makes the point that this poor linearity should not be confused with the ability of the listener to localize the sound sources. By virtue of our hearing apparatus, humans are not particularly good at localizing sound sources directly to the sides, but are quite adept at pinpointing sound sources directly in front or behind. This also applies in 5-channel stereophony. While the image itself may be stretched or distorted in a particular segment, the listener’s ability to discern where the phantom images are appearing, or how badly distorted the image is depends on which particular segment is being used to generate the image.

The next trick is to knit all of the contiguous segments together into a continuous whole. This concept is referred to as Critical Linking, defined by Williams as “continuous… (sometimes called “seemless”) reproduction of the complete sound field without overlap or holes.” ([25] p. 9) Critical linking can be obtained by some kind of offset between the front triplet and the back pair of microphones. This offset can be electronic (i.e. the use of delay or a level difference between the front and rear) or can be obtained through the position of the microphones in the array (Microphone Position Offset).

While the strength of near-coincident arrays lies in the fact that they use a combination of both interchannel level and time differences, Williams recommends that intensity differences should provide the dominant cues in the front triplet. He also provides the following advice: “A good ‘rule of thumb’ is that the sound sources should be no nearer than about 5 times the distance between the microphones.” ([25] p.9)

Multichannel Recording according to Theile

Dr. Gunther Theile of IRT has conducted his own experiments in near-coincident surround microphone arrays, and his findings are quite interesting. His paper *Multichannel Natural Music Recording Based on Psychoacoustic Principals* is a comprehensive guide to many of the issues that arise in natural “surround” recording. This paper is so important to my research that it merits its own subsection of this thesis.

His approach is somewhat different to that of Williams in that he is not separating the stereo field into five segments and then trying to link these segments together, rather he is treating the surround field as a frontal segment and a rear segment. This is illustrated by his use of the description "3/2 stereo." His stated goal is to develop a technique for "optimum 'naturalness'" of the stereophonic imaging. He initially defines optimum naturalness as the reproduced sound being "nearly identical to the original sound image" ([12] p.3) a goal which he acknowledges is beyond the capabilities of the 3/2 stereo format. This goal of "identity" must give way to a more realistic goal of naturalness. The criteria he lays out for a "natural" presentation are as follows: "[the image] should be satisfying aesthetically and at the same time it should match the tonal and spatial properties of the original sound." ([12] p. 3)

Lack of naturalness is not necessarily a defect. Thiele acknowledges that there are sometimes aesthetic goals which cause the stereo image to differ from "naturalness," for instance the use of a Decca Tree to record an orchestra would result in a "dense and open sound picture perceptible over a huge listening area" ([12] p.5), but where the goal is "a precisely dispersed, deeply staged, stable and wide image of the orchestra in the room, where the foreground distance is not fixed at the loudspeaker distance and the spatial impression and acoustical environment allow a convincing illusion of 'being in the hall' over a reasonable listening area" ([12] p. 5), this requires a recording technique which relies on a more naturalistic approach.

A accompanying table outlines specific imaging parameters for the sound engineers' condsideration, on which he displays various acoustic elements present in the concert hall experience, and the relative importance of each in terms of different aspects of recording aesthetics:

	Orchestra elements	Soloists	Room	Audience
Horizontal direction	●●●	●●●	●	●●
Elevation	○	○		○
Near-head distance				○○○
Distance, spatial depth	●●●	●●●		●●
Spatial impression			●●●	●
Envelopment			●●	●●●
Sound colour	●●●	●●●	●●●	●●

Consideration of specific imaging parameters for natural sound design

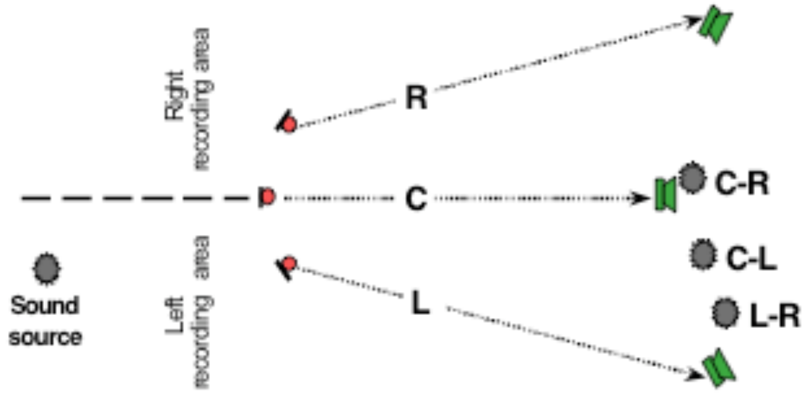
Imaging of orchestra elements, soloists, room acoustics, or ambient atmosphere (audience) requires the application of specific phenomena of spatial hearing, each of them governed by particular laws and needing adequate microphone configurations and mixing methods. However, in the case of 3/2-stereo there are constraints regarding perception of elevation and near-head distance.

Source: [12] G. Thiele 10/2001

Theile dismisses the use of coincident microphone techniques: "It is well known that pure coincidence microphone concepts are not able to produce a satisfying natural spatial impression, due to the lack of adequate interchannel temporal relations" ([12] p. 9). He also advocates the use of directional microphones "to minimize interfering crosstalk [and] also to attenuate the indirect lateral and rear sound." ([12] p. 9) According to Theile, in the bad old days of two-channel stereo, omnidirectional microphones were used primarily to "effectively record affluent reverberatory components" (p.9) This becomes problematic in a natural surround recording, since it is assumed that the indirect sound will be reserved largely for the surround channels.

Thiele raises one extremely important issue that is also recognized by Williams/LeDeu: one must take the center channel into special account when planning a frontal microphone array, however Theile gives a number of methods by which to treat the center channel. In recording a large orchestra, the common technique that has been utilized by film mixers is to treat the left and right speakers as would normally be done in a two-channel stereophonic recording and rely on phantom imaging. Important sources, such as soloists could be mixed exclusively to the center channel to ensure solid imaging. One could also deploy two stereo arrays, one for the left side and one for the right. The two "inside" microphones are summed to the center channel and the overall level of the center is reduced by 3 dB. The "outside" mic's receive a corresponding delay. This treats the center-left as one separate stereo imaging area and the center-right as a completely separate area, not unlike Williams' consideration of each sector as a separate area with a corresponding SRA. Another method for recording a large orchestra could include five microphones spread across the front, with the mid-left and mid-right microphones fed to the left and right speakers and also to the center with their levels reduced by 3 decibels. The center microphone is fed exclusively to the center channel, and the left and right flanks are fed to the corresponding speakers. This results in stable phantom images half-left and half-right as well as stable images at the speakers themselves.

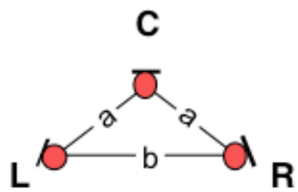
In my case, the main issue is not recording an orchestra, but rather how to approach recording a chamber ensemble. Theile acknowledges that a completely different approach must be taken in such a case: "here we need a real three-channel main microphone concept to achieve directional stability from the center loudspeaker and – at the same time – ensure…maximum localization focus and minimum coloration due to combing effects. ([12] p.8) This is more difficult in practice than one would think.



Source: [12] G. Thiele 10/2001

The main issue that arises with a near-coincident placement for three microphones is that, in reality, you are actually dealing with three stereophonic arrays rolled into one: there will be directional cues from center-left and center-right, as well as from the left-right pair of microphones. Since these are all close together, there will be conflicting directional cues generated by each of these arrays, all of which will be reproduced with nearly equal intensity and with very little delay. With a widely spaced array, there will be interchannel time differences (ITD's) on the order of 3 to 6 ms (or more) when a single sound source is picked up by adjacent microphones. The time cue is great enough that the ear will not be confused, because the delay is great enough to place the source firmly in one channel or the other; in stereophony, it requires only a delay of 2 ms to localize a source completely at one speaker. In a near-coincident array, we may be dealing with ITD's of less than 1 ms and so this "precedence effect" will not be helpful. Likewise, interchannel level differences for a source that is near the centerline of the array may be "too close to call." Thus, rather than achieving the goal of greater stability and enhanced image clarity, the center channel becomes a liability if not treated properly.

This calls into question the viability of using a small Decca-Tree for a chamber ensemble as suggested by Christensen in [9]. It would seem that there would not be adequate interchannel separation between the front three microphones, especially if omni's were employed. Here, Theile also debunks Williams' notion that one can simply decide on an SRA for each separate segment and then simply link those segments together – especially in the front triad - using the "INA 3" arrangement as an example. The INA 3 is a near-coincident technique developed using the Williams' curves:



The triangle arrangement has been designed in line with the so-called "*Williams-Curves*" aiming optimum attachment of the recording areas for L-C and C-R. the distances a and b are calculated for cardioid capsules dependent on the resulting recording angle φ :

φ = 100°:	a = 69 cm	b = 126 cm
φ = 120°:	a = 53 cm	b = 92 cm
φ = 140°:	a = 41 cm	b = 68 cm
φ = 160°:	a = 32 cm	b = 49 cm
φ = 180°:	a = 25 cm	b = 35 cm

The off-center angles of the microphones are always ε = ½ φ

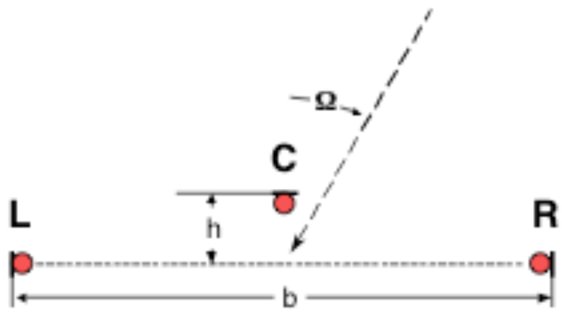
Source: [12] G. Thiele 10/2001

This array attempts to achieve an even distribution of sources in both the L-C and R-C segments. According to the Williams/LeDeu paper, the next step is simply a matter of combining the two segments. The problem is that the acoustic interference between the two stereophonic pairs is not really taken into account.

Theile makes the following statement: *"the optimisation regarding the attachment of adjacent recording areas does not imply a minimisation regarding artefacts due to interchannel crosstalk. Minimum impairments of localisation focus, clarity, and timbre may not be achievable with this configuration, because the acoustical channel separation is not sufficient."* ([12] p. 9) He goes on to point out that part of the reason that there is insufficient separation is due to the directional characteristics of the microphones employed. Were one to employ super- or hyper-cardioid microphones, the interchannel separation might be better, but this may require a different set of angles and spacing for the microphone array to achieve the same linearity within each segment.

He also points out the weaknesses inherent in an array that resembles a so-called ORTF-triplet. This array is made up of three near-coincident microphones in a line – each 17.5 cm from the next. It does employ super-cardioid microphones at the left and right, each of which is angled +/- 30-degrees, with a cardioid center microphone. This will result in greater separation between the left and right channels, but does not solve the problem of a source in the center being picked up by both L-C and R-C arrays, resulting in conflicting directional cues and loss of focus in the center. Also, because there will be mic's with differing polar patterns making up each pair, the distribution of sources in the L-C and R-C sectors will be less than linear.

Thiele presents his own array, the Optimised Cardioid Triangle or OCT array. This array is based on using the directional characteristics of microphones to reduce inter-channel crosstalk in the frontal triplet, as well as to create inter-channel time differences that will reinforce the image rather than confuse it.



OCT - an optimised triangle configuration

The microphone characteristics of capsules L and R are super-cardioids. They are faced side wards (off-center angle $\varepsilon = 90°$), in order to ensure maximum channel separation. Preferably the freefield equalis ation of capsules L and R is based on $\Omega = 30°$, and the center microphone is a cardioid.

Distance b depends on the recording angle φ
Distance h = 8 cm. - If cardioids are used instead of super-cardioids, it is h = 12 cm.

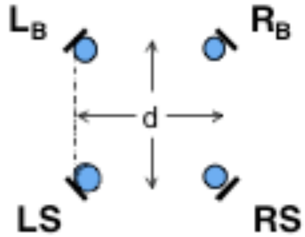
source: Theile [12] p. 13

In a near-coincident array, the inter-channel level and time differences work together. In other words, the angular shifts for a phantom sound source results from the *sum* of the inter-channel level (ΔL) and time differences (Δt). In the case of the above OCT, if the C microphone is a cardioid and the L and R mic's are supercardioid, the interchannel level differences will be as much as 9 dB. Since the center mic is forward of the left and right microphones, it will generate an inter-channel time difference for sources in the center. As shown in the above diagram, varying the distance between the L and R microphones will change the recording angle and varying the distance *h* will vary the time difference of C-L or C-R. The interchannel intensity difference or the interchannel time difference by itself may not be quite enough to localize a source completely at one speaker, the perceived location of the phantom image will be the sum of both Δt and ΔL, resulting in a clear directional cue.

One of the most interesting conclusions that Theile draws is that there really is no single array that would be best to capture both the frontal direct sound and the ambient "surround" sound – in other words, no array arranged around a single point as we do in stereo. For the OCT array, he writes that if one wanted a minimalist approach to the surround recording (meaning no more than five microphones), one could add two cardioids facing the rear of the hall, about 50 cm apart and about 40 cm behind the OCT array. He suggests that ideally two arrays should be used, one to capture the direct frontal sound, which would ideally be located near the critical distance and the other in the diffuse field to capture the ambience: "in many recording situations a compact main microphone causes an unfavourable compromise." ([12] p.33)

There are many techniques that are designed specifically to capture ambience. Some of these are two microphone techniques such as the Woszcyk Technique which employs two coincident cardioids at 180-degrees. The polarity of one mic is reversed. According to Wieslaw himself, the purpose of this is to generate the equivalent of a bidirectional capsule, but with control over each side of the capsule.

Another technique is the IRT Cross, which employs four cardioids at 90 degree angles, about 25 cm apart. This technique has the disadvantage of picking up more of the direct sound, since the sound source is not in the null of any of the microphones in the array.



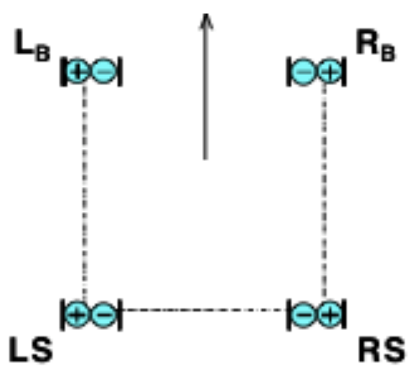
"*IRT-Cross*" for atmo recording

A square arrangement of cardioids (d = 20…25 cm) is known as "Atmo-cross" and offers realistic stereophonic images of ambient sound such as for example applause.

However, this configuration can have disadvantages in room microphone applications if the R/D-ratio level picked with LB and RB is not large enough (distance from the L-C-R main microphone to short

Source: [12] p. 32

My personal favorite is the Hamasaki Square, which employs four bidirectional microphones with their nulls aimed at the ensemble, the positive polarity side of the capsule aimed at the walls.



"*Hamasaki-Square*" for spatial imaging

the square arrangement is proposed to be configured with figure of eight microphones (spacing d = 1 m). The null point of the directional characteristic of each microphone is facing the stage in order to reduce the energy of direct sound as much as possible. For lateral sound a pure A/B-microphone is effective. The array has been placed "at a very high position in the concert hall where the sound is very diffused and energy from direct sound is diminished" The microphones L_R and R_R are routed to channels L and R or panned between L-LS and R-RS, the microphones LS and RS are routed to channels LS and RS.

Source: [12] p. 32

One characteristic of a good "atmo" technique is that it is able to reject most of the direct sound while picking up as much of the ambience and reflections as possible. Another favorable characteristic is that there is adequate interchannel coherence between the loudspeakers on playback. It has been shown that if the mic's are too far apart, there is not enough cohesion between the channels and the ambience begins to collapse around the individual speakers rather than giving a feeling of immersion. If there is too much interchannel coherence, "a phantom sound source becomes audible above [the] listener's head, and subjective envelopment disappears accordingly." ([12] p. 32)

Sphere Microphone

The Schoeps Sphere microphone system differs from the previous near-coincident concepts quite radically. Rather than trying to rig five directional microphones in some sort of configuration, it utilises a sphere with two omnidirectional microphones on either side. Jerry Bruck of Posthorn Recording came up with the idea of adding a bidirectional capsule just under each omni. This yields two M/S arrays separated by a sphere that is about eight inches in diameter. One M/S array is de-matrixed for front left and surround left, the other gives the signals for front right and surround right. In theory, one can “zoom” in or out by manipulating the left/right balance on each side. This is a novel approach that is easy to carry and rig. The whole system, including processor, is offered as the KFM 360 5.1 Surround Sound System for about $15,000.

Section 3: Planning the Session

It was my intention from the outset to record an early music ensemble since I tend to gravitate towards this type of music and performance. Recording early music ensembles usually involves recording instruments that the typical recording engineer does not encounter on a daily basis. Also, since much of the music from this period is intricately contrapuntal, it is important to make sure that all of the voices are properly represented. In addition, there is a certain aesthetic and often certain temperaments associated with certain composers and certain time periods. While all of these factors contribute to my interest in the music, they also must be a concern during the recording.

One of the things that initially attracted me to NYU was the fact that there was an early music program at the school, which possessed a number of period instruments. I contacted Dr. Stanley Boorman, the coordinator of the program, and expressed my interest in doing a surround sound recording of an early music ensemble, and he graciously put me in contact with Margaret Panofsky, who is in charge of the viol consort at NYU. The consort is made up of five players: two trebles, two tenors and a bass.

I met with Margaret to discuss some of the possibilities for the project, and to find out more about the ensemble and the music. She was kind enough to furnish me with scores and we took time to discuss the different approaches that could be used in this project. Several possible venues were discussed and a timetable established. Fortunately, she is no stranger to the recording process so she was able to bring up important issues and address many of my questions.

It was decided that ensemble would need time to rehearse adequately, so a date in early December was proposed for the initial recording, and additional work could be done over the break (ostensibly after this paper is completed).

The Instruments

The viol, also known as viola da gamba (Italian: *viol of the leg*), refers to a family of bowed stringed instruments, similar in some respects to the modern violin family. The viol emerged in 11th century Europe, possibly as the result of applying a bow to a plucked string instrument such as a lute. After waning in popularity for a time, the instrument reemerged in the 15th Century and enjoyed a period of immense popularity among amateurs and professionals alike. During the Baroque period, as music moved into the concert hall, the more powerfully-toned violin family became the dominant stringed instruments and the viol family moved into obscurity.

The *da gamba* in the name indicates that the viol was held between the knees, similar to the cello - not under the chin like a viola or violin. Unlike the cello, there is no support to rest the instrument upon while playing. The viols are fretted like a guitar, and these frets

can be moved to attain different temperaments. Commonly, there are six strings "tuned in a sequence of a 4th, 4th, M3rd, 4th, 4th (d-g-c'-e'-a'-d" for the treble). [1] The bow is held underhand and in some cases it is arched outwards from the hair, not inwards like a violin, and the player can adjust the tension of the hair with his/or hand. The viols that I recorded used a more modern Italian style bow.

While the viol came in many shapes and sizes, the most common were the treble, tenor and bass viols. The bass viol is approximately the size of a modern cello and the tenor and treble are successively smaller. The tone of a viol "is quiet but with a rather distinctly nasal quality " [1] whose clarity lends itself rather well to contrapuntal music. To my ear, its sound is a bit more "glassy" than a violin's. In some ways, the viol consort could be considered the forerunner of the modern string quartet. However, unlike the modern string quartet, the quality of the instruments is more homogenous, the overtones of the instruments seem to blend together, and the dynamics of the instruments are evenly balanced. They are well suited to ensemble playing, and were often used to accompany solo voice. A disadvantage to this "nasal" quality is the fact that close-microphones will not flatter the instruments, a concern in this case. Also, since the timbre of the instruments lends itself to the blending of the ensemble, contrapuntal lines between two of the same instrument can easily be obscured.

The music

The literature for the viol family ranges from arrangements of choral music to solo and ensemble pieces, culminating perhaps in the Fantazias of Henry Purcell. The consort in this case will be playing five-part fantasias by three English composers: William Byrd, John Jenkins, and John Ward.

The fantasias of the 16th century (which the English sometimes called *fancies*) are a type of composition that was supposed to resemble improvisations, perhaps a bit like the later *impromptu* of the Romantic period. According to [10], they fall into the same category as the *ricercare* which was a type of prelude which developed into a fugue-like form. These were not dance pieces, nor were they supposed to resemble dance forms of any sort. These were polyphonic art-pieces written both for solo and ensemble playing, demonstrating the skill of both the composer and musician.

William Byrd (1543-1623) was the premier English composer during the Elizabethan period. A Catholic, he nevertheless was directly in the employ of the protestant Queen as a composer, singer and musician. Byrd was prolific, writing masterpieces in virtually every genre sacred and secular. Known particularly for his vocal and keyboard writing, he also composed many consort pieces. One of the master composers of Renaissance music, his work and name are held in the same regard reserved for his contemporaries Palestrina and Di Lasso.

As a composer, John Jenkins (1592-1678) seems to have been a bit of a "chamber-music" specialist. The majority of his surviving work is for consort, much of it seemingly geared towards semi-skilled and amateur players: "A great deal of his consort music, apart from the fantasia-suites, is geared towards amateur performance...ideally suited to the talents of the less able performers he was likely to muster in consorts at the country households of the nobility..." [6] who were his primary employers as a freelance musician. It may be inferred from the above that the fantasias were in a different category, perhaps written for consorts of professional musicians as might be employed at the courts, more complicated musically and requiring more virtuosity.

John Ward (1571-1638) is somewhat more obscure than the other two composers mentioned here. It seems that little is known about his life. It does seem, however, that his consort music was well regarded and quite popular during his lifetime.

Aesthetic and Practical Concerns

Typically, in most classical music recording situations, whether two-channel or surround, the engineer strives to capture a certain amount of "concert realism." Streicher and Everest [13] break it down into two major categories: either put the listener in the concert hall ("you are there") or put the ensemble in the listener's living room ("they are here"). There are several factors that must be kept in mind: the stereophonic image, the mixture of ambience and direct sound, the feeling of blend between various elements of the ensemble, and the color and quality of the tone.

There are always tradeoffs and compromises to be made. Even in the world of two-channel stereo, there is no silver bullet - no one technique that fits the bill in every situation. In a multichannel surround configuration, there is even more to consider. Firstly, in addition to the traditional phantom image between the left and right speakers, there are now two smaller stereophonic imaging areas in the front of the listener (L-C and R-C). Also possible are an area behind the listener (Ls and Rs) as well as on either side (Rf-Rs and Lf-Ls), although these speakers are too far away from each other and the angles not optimal for true stereophonic imaging. It is important to note that any pair of speakers in the system could theoretically interact.

I had to decide exactly how I would approach this session. Should I take the “safe” route and attempt to recreate the experience of being in the concert hall or should I try something different, perhaps a bit more “aurally exciting”? Much research has been done on how one should go about recreating an event in surround sound, and many theses have been done comparing one surround recording technique to another. This paper is not so much concerned with these questions, and there was no money on the line in this situation, so I felt free to attempt an experiment of my own.

It occurred to me that one of the challenges that we face when we record classical music on-location is trying to capture as accurately as possible the positioning of the instruments that make up the ensemble. We are almost preconditioned to accept that the instruments should appear in certain positions – the seating is arranged and I simply try to figure out how best to place my mic's so that the image will reflect what is before me. I was curious to know how it would work out if I took the opposite approach – to create the desired image in my head and then arrange the musicians in the positions that correspond to this image.

This ensemble was almost too easy: five instruments correspond perfectly to a five loudspeaker setup. Of all the music ever written, five-part Renaissance viol music could possibly stand to benefit the most from 5-speaker surround sound. Traditionally, a viol consort is seated in such a way that the two trebles are next to each other, the bass is in the center and the two tenors are seated next to each other on the other side. In such an arrangement, it is easy to blur and confuse the lines of the instruments in a traditional stereo recording. By placing each instrument in its own speaker, the contrapuntal lines could unfold with unprecedented clarity.

Another benefit of this approach is that it entirely avoids the use of phantom images by placing a static image in each speaker. By taking this road, I can sidestep the tricky question of accurate localization between loudspeakers – something which is subject to change in any listening environment where the speaker layout differs even slightly from whatever standard I choose for my monitoring setup, or where the acoustics of the listening environment interfere with loudspeaker imaging. It also presents an intimate "they are here" feeling where the loudspeaker becomes the proxy of the live player in the listener's living room. In two channel stereo, this would be annoying; in 5-channel surround it is immersive and engaging.

If I were recording an orchestra, I may not have taken this approach, though this philosophy could theoretically be applied to a larger ensemble. However, I believe that the projecting the illusion of size and depth, while maintaining image stability may be a bit more difficult to attain using this approach. In my opinion, smaller ensembles have more to gain from an approach such as this, especially where intricate contrapuntal textures can be masked or confused in a smaller image.

I am a fan of placing the low frequency instruments right down the center, so it made sense for me to place the bass in the center speaker. Also, since there are two of everything else, it would be most balanced to have one of each in a left or right speaker. Not only would this make for a pleasant surround image, but a two-channel stereo mixdown of this image would also preserve the countrapuntal clarity. From there it was a question of whether I should place the trebles in the front or in the surrounds. I felt that the high frequency instruments would most grab the listeners' attention, so it would feel more natural to have the trebles in the front and the less obtrusive tenors behind.

At this point it comes to a decision: shall I use a specific free-field microphone technique and position the instruments in such a way around the array that they will give the desired image? Firstly, the use of a widely-spaced technique seemed to be out of the question, at least for the main array. Since this was essentially a chamber music ensemble, the approach needed to be smaller and more intimate, with some concern for image clarity since we cannot really hide behind the impression of size in this case. That would imply that a near-coincident or coincident array might be the way to go. This left me with the dilemma of using a more "traditional" approach in an untraditional manner, or simply abandoning the idea of using a prescribed arrangement of microphones and attempting something different.

One idea was to close mic the instruments and use an artificial panning, in which case they could sit more or less however they like. Close mic'ing a viol might not give a pleasant sound, and according to [14], there should be a healthy amount of bleed on all the mic's, and none of the mic's should be directly on-axis to any source. There is the issue of speaker layout to contend with: I am adamant that the material must work in any 5.1 speaker layout, which for me eliminates the possibility of using phantom images between the speakers. In this case, the weakness of the "minimalist" arrays is their lack of flexibility on playback. Audio Research Labs makes a plug-in that would enable me to take the close mic'ed instruments and pan them around the surround field using their own proprietary algorithms, which employ a combination of delay and EQ, and have a natural-sounding result, which was demonstrated by Paul Geluso at a meeting of the NYU AES chapter. However, after hearing the plug-in personally, I realized that this could possibly color the sound of the viol in an undesirable way.

In the end, I decided to strike a compromise: the musicians would sit in a circle, more or less corresponding to the speaker layout. I wanted to seat them in such a way that a cardioid microphone pointed on-axis to one instrument would have maximum rejection of the other instruments in the ensemble. Each instrument would have a spot mic, but the mic would not be too close, so I would avoid the undesirable close-mic'ed sound and retain a sense of blend of the ensemble on playback, yet still have maximum control over the balance and placement of each instrument. All of the microphones should be the same make and model to keep the tone of the viols homogenous. In theory, this technique is also completely stereo compatible: if I were able to pan each instrument in two speakers, keeping a clear stereo image and not have any audible phase issues between the microphones, the recording should translate well to surround. Since most monitoring on location is done in headphones, this ability to monitor in stereo and mix in surround would be important. Using this technique, isolation and phase issues between the mic's should be apparent even in stereo. One issue I did not consider was front to rear imaging,

A problem results from such close microphone placement, especially with cardioids: there might be a feeling that the musicians are too close – the room would not be well represented and the sound of the ensemble might not blend together as with a simple stereo array. Also, as mentioned previously, I am not the only recording engineer who prefers the sound of an omnidirectional microphone to a cardioid when recording classical music. To address these two issues, I opted to set up an “ambience” array made up of omnidirectional mic’s in addition to the close mic’s. One question that arises is the number of microphones to use for this array: I could use two, four, or five microphones. In any case, as with the spots, it would be preferable to have mic’s of the same make and model.

Two mic’s would be the easiest – I could set them up inside the circle, high above the heads of the musicians. These signals could be sent equally to the front and rear speakers. However, the worst-case scenario with this would be the room sound from the ambient mic appearing as a point source between the front and rear speakers! The ambience should surround the listener, not become a distinct localizable source. I could use four microphones and position them in between each of the musicians, again high overhead and pan the signals to each of the corner speakers. This is a better solution, but the question remains as to what goes in the center channel. The bass might appear very dry and “in-your-face” compared with the rest of the ensemble. As a result, I decided that this array would be made up of five omni’s, and now the question was a matter of placement. Should I place a single omni behind each player (outside the circle)? Perhaps I should place the mic’s in the hall to capture the diffuse sound field where ensemble would be blended together by the room. Both of these placements come with their own set of problems.

Placing omni’s behind each player would reinforce the image well. However, because of the close proximity, audible comb filtering could occur between the “spot” mic and this ambient mic. The ambient microphones might end up being too far apart, which could mean that there would not be the desired amount of correlation between them. These mic’s are supposed to blend the spot mic’s together, not add phase problems of their own.

On the other hand, a more distant array might end up being too far away and an audible “echo” could occur unless the close mic’s were delayed. I prefer to avoid processing whenever possible. In a live recording situation, I have found that it is best to have the microphones placed in such a way that they do not need artificial “fixing” after the fact. Also, the more distant placement requires a room that is both pleasant sounding and quiet. This is not an easy thing to find – especially in New York City. For the first recording session, I had managed to book the Frederick Loewe Theatre at NYU, which is not a bad-sounding room, but it can be quite noisy – especially with the traffic on West 4th Street.

The week before the session, I was able to attend one of the group's rehearsals. The group had been practicing sitting more or less in a circle, with the two trebles on either side of the bass and the tenors sitting opposite, a request I made several weeks before, so that it would not be strange to the group to be in this formation on the day of the recording. I stood in the middle of the circle facing the bass to see how this arrangement would sound in surround and I was pleasantly surprised at the result. Most interesting to me was the difference in sound when I changed the height of my listening position. When I was standing at my normal height (about 5' 9"), the sound was a bit thin, but when I crouched down to beneath the music stands, the sound was very rich and full - all of the sparkly overtones that I associate with the viol, but with the fullness of the body of the instrument. This planted the thought that rather than follow the normal convention of mic placement high above the musicians, I might place the mic's down low to capture this aspect of the viols' sound.

Limitations of the Media

The medium dictates certain elements of the recording process, most notably sampling rate and bit depth. For example it would make sense to record a project for CD at 44.1 kHz sampling rate. Likewise, it would make sense to record a project for SACD using DSD. Since my target medium for the full-quality product is DVD-Audio, there wer certain things I needed to take into account to make my life simpler (better?) in the end.

It doesn't seem as thought there would be too many limitations to DVD-Audio, but there are a few. DVD-Audio is, in theory, quite flexible in its implementation; one can mix sample rates and bit depths between channels, use various forms of data compression and so forth. I would not be spending the money necessary to get the "professional" DVD-Audio authoring software, so I would be limited to using LPCM formats.

If I were in a situation where most of the important information was in the front three channels and the rear channels contained only ambience, I might decide to mix sample rates and/or bit depths. For example, one could opt for 2 channels of 96k at 24-bits and 4 channels of 48k at 24-bit. Also, one could have 3 channels each of 96k-24-bit and 48k-24-bit. The channels are separated into two hierarchical groups. According to [35]: "The sampling frequencies and word lengths of [Channel Group] 1 must be greater than or equal to [Channel Group] 2." ([35] p. 392) It is possible to have as many as four channels in Channel Group 1. However, in this case, since all the channels will be of equal importance, I would want to keep the sample rate and bit-depth uniform for all channels. This leaves me with two options: a higher sampling rate (96 kHz) at a lower bit depth (16-bits) or a lower sampling rate (48 kHz) and a higher bit-depth (24-bit). I chose the latter for three reasons.

Firstly, smarter people than I make the argument that the greater bit depth is more important than the higher sampling rate. While I have no tests, subjective or otherwise, to corroborate this theory, I will at this time accept it in blind faith. Secondly, if I were to record at 96 kHz with my own equipment, the higher sampling rate would cut my track count in half. For this project in particular, this would not be the end off the world – I only need 10 tracks - however, with my rig, it eliminates the possibility of a redundant backup. In the world of live recording, this is not an option. Thirdly, and perhaps most practically, Dolby Digital uses a sampling rate of 48kHz. Rather than running my finished product through a sample rate converter, for me it just seems better to start off where you plan on ending up.

Section 4: Execution
Session 1: Triskaidekaphobia at NYU
The first recording session took place on December 13th in the Loewe Theatre at NYU. Rather than cart all of my own gear in and restrict myself to monitoring in stereo, I opted to use NYU's Studio H, which is located next to the lighting booth in the rear of the theatre and is connected to the stage via a 16 + 4 channel snake. Studio H is equipped with 5 Meyer HD1 speakers and a Tascam DSM 7.1 to monitor in surround.

My time was limited: I had four hours to setup and record without the benefit of an assistant, and my plan was ambitious: I wanted to record with five spot mic's, five ambience mic's and the Soundfield microphone. I brought an eight-channel Millennia preamp with me, hoping that I could take advantage of using the API preamps installed in H to fatten up the spot mic's, and use the Millennia to amplify the ambience mic's. I was also under the impression that I might need the additional channels if I were planning to use the Soundfield microphone. I would need to take the necessary time to hook this into the setup in H.

I brought some mic stands with me, hoping to eliminate the time that it takes to find a decent stand in Loewe. I also brought eight 50-foot mic cables that I knew were in good condition to avoid at least some of the cabling issues that might arise such as bad cables or not having enough 50-footers. Knowing I would need additional microphones for my planned setup, I brought my own matched pair of Schoeps cardioids and a pair of Neumann KM130's with diffraction spheres. I knew that channel 1 on the snake was not working, so I planned my inputs from channel 16 backwards. Despite what I thought was adequate foresight, there were several unanticipated factors that nearly made the session a disaster.

Perhaps the most important factor that I did not foresee was the amount of time it takes to get from the control room to the stage and back. With my own rig, it is not uncommon for me to set up and wire up my gear, run a snake, and place up to 16 mic's in about an hour and a half, all by myself. I am normally within a 50-foot snake run of my microphones, so troubleshooting a bad cable or adjusting mic placement does not normally pose a significant time issue for me. I usually plan for about two hours of setup time so that I have a bit of padding in case something goes awry. My goal was to be fully up and running before 8:30.

This session was a simple matter of setting up ten microphones in an existing installation, and then figuring out the Soundfield mic, all of which should have been a walk in the park. However, in Loewe, to get from the control room to the stage, one has to walk through the lighting booth, which spans the entire width of the theatre, after which there are 20 steps on a narrow spiral staircase to navigate. Then one must backtrack a bit to the entrance of the theatre and cross from the very rear of the theatre to the front. In other words, each round trip is equivalent to more than a lap around the entire theatre, plus a trip up and down the stairs. Even at a good sprint, this takes up valuable time.

Another factor which I did not take into account was the fact that the last time I had done a surround session in H was more than a year and a half ago, and I had never done so on my own. Although I have recorded several concerts using this studio, they were all in two-channel stereo. I consider myself to be a competent ProTools operator, but had not used it in a surround situation. Unfortunately, I was unable to drop by for a walk-through prior to the session, and I was unfamiliar with the signal routing in this particular studio. I was under the mistaken assumption that the surround monitoring setup was normalled in H, and there was also the fact that I was unaware that I would be unable to get the center speaker working at all.

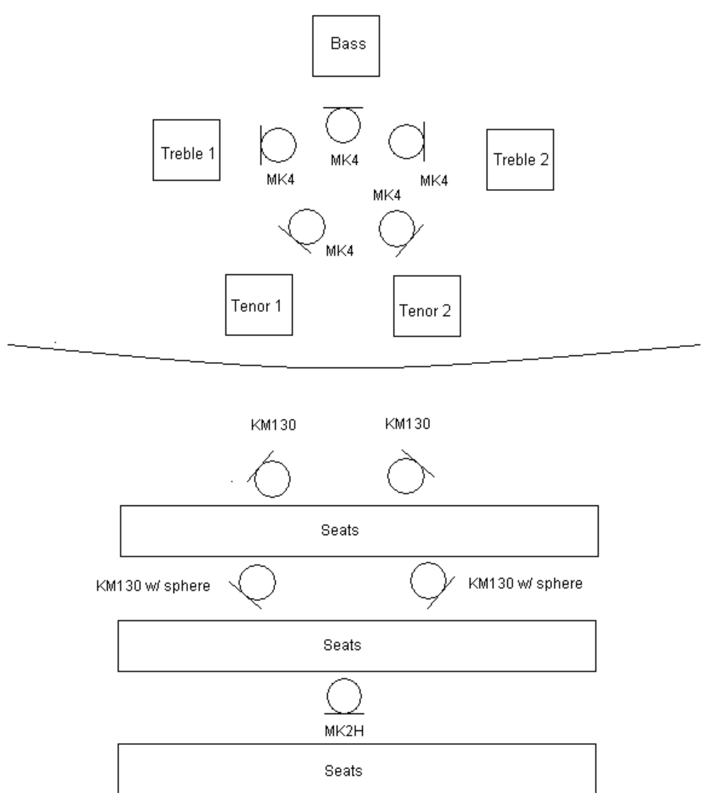
I had reserved the hall starting at 6:30 and the musicians were due to arrive around 8:00, so I arrived at 6:00 to get the keys and load in. A small orchestra was rehearsing in the hall, so I started lugging some of my gear up to the studio. As the orchestra began to disband, I deposited my stands and cables near the stage and brought all of the mic's and cables, followed by the snake down from H which required three round trips to the control room. I readied the chairs and stands for the musicians, and lined up five tripod stands which I would need on stage. Next, I set up my five 15-foot stands in the hall, three forming a Decca Tree facing the rear of the hall with the center mic in the third row, the two flankers in the second row, and then another pair in the first row. The idea was to have five channels of hall sound to bind the sound of the spot mic's together. Since I did not want the front three ambience mic's to interfere with the clarity of the imaging, I felt that orienting the array facing the rear of the hall would be more advantageous. I would simply reverse the right/left perspective so that the image in the ambience array would agree with that of the close array.

This use of five omnidirectional microphones is similar to the Richard King-Sony/Phillips technique, but my real inspiration was the Decca Tree. I needed to have a discrete feed for each speaker that would have some decorrelation, i.e. I could not simply double the right and left front channels in the right and left surrounds, and so I used two more omni's for the surrounds. I decided to use NYU's KM130's for the surround pair and my own 130's with the diffraction spheres for the left and right. These spheres provide greater directionality in the high frequencies changing the omni to a supercardioid at around 1 kHz, giving these channels even more decorrelation and focus in the front speakers. Since there were no more KM130's at my disposal, I chose a Schoeps CMC6/MK3 as the center mic since it has a similar frequency response to the KM130.

I first set up the two surround ambience microphones and plugged them into channels 15 and 16 on the snake. I then made a trip back up to the studio, just so I could make sure that I was getting signal to the control room. I put the rack case containing my Millennia in position to be connected, but decided to save the patching for last since it was a largely unnecessary luxury. Similarly, I decided to wait with the Soundfield mic until I had my main setup ready to go since it would essentially be redundant.

I set up a ProTools session, using the 5.1 SMPTE mix format, with the sampling rate set to 48 kHz and the bit depth at 24-bit. I verified that I was in fact getting signal to the preamps, but I could not get any signal to show up in ProTools at first, until I realised that channels 23 and 24 on the mixer were somehow in the signal chain between the API's and ProTools. Once I raised these faders to unity, I was satisfied that I could at least get signal into the computer. However, I could only monitor the signal if I was listening to the board mix. In other words, I could not hear what was coming out of ProTools. I decided to deal with this later. I went back down to the hall to get my mic's set up and wire up the talkback speaker.

I finished setting up the rest of the ambience array and plugged them into channels 12 through 14 following the convention L, C, R, LS, RS. I placed the tripod stands in the midst of the music stands. I placed a mic clip on each stand and ran the cables, deciding to keep the mic's out of harm's way by letting the musicians arrive and setup before I put them in place.



Having done everything I could in the hall at that point, I went back to the control room to try and figure out the routing problem. As I expected, the other surround mic's also showed up on channels 20 through 22 on the board and I raised the faders to get the signal into ProTools. I still could not hear any return from ProTools and could only monitor the board mix in stereo. Nevertheless, I powered up the surround speakers and dragged them into position, following (to some extent) the NARAS recommendation for the angle and placement. At this point the ensemble arrived, so I went back down to the theatre to greet them. As they got settled, I put the close mic's on the stands and plugged them in, using channels 7 though 11 on the snake. The arrangement of players and microphones was as follows:

I chose Schoeps cardioids (MK4) for the close mic's, each focused on a particular instrument hoping to get a certain amount of bleed between the mic's to blend the array together, while still keeping enough isolation so that each viol could stand alone in its own speaker. This thinking was partly based on advice I received in a demonstration [15] that each mic should be in the diffuse field and so able to "hear" all of the instruments. I was trying to strike a compromise between a dry spot mic (which especially on a viol could sound obnoxious) and a true minimalist array such as the Williams/LeDeu or the OCT arrays.

However, the recommendations from [15] called for all mic's in an array to be off-axis to and high above the instruments. I began my career in recording studios where close mic'ing was the rule, and I have always felt that when close mic's are done correctly, they give an element of control and stability that most stereo arrays do not. Simply placing isolated spot mic's in each speaker with no correlation does not sound good, but neither does an unstable or unclear image. I had the idea that if one can combine both philosophies successfully in surround, the result should be quite pleasing to all ears, regardless of the speaker position on the user end.

I hard patched API's 7 through 11, which were amplifying the close mic's, to the Mytek inputs 1 through 5 on the analog patchbay. I also hard patched API's 12 through 16, which were carrying the ambience mic's, to the analog inputs of the 192 HD. Still, I could only get signal on the ambience mic's through the mixer channels. I resigned myself to this and set these faders at unity. Unfortunately, one of the spot mic's was not finding its way into ProTools, but I figured that this must be a bad cable so I would address it later. In the meantime, I set up a session for a stereo mix (which is something I had done recently in H) so that if push came to shove, I would at least be able to record the session, even if I had to monitor in stereo. This worked and I was able to monitor the output of ProTools as I expected. I was also happy to find that the talkback speaker was functioning properly.

Source	**Mic**	**Preamp**	**Converter/Input**
Treble 1	CMC6/MK4	API 7	Mytek 1
Bass	CMC6/MK4	API 8	Mytek 2
Treble 2	CMC6/MK4	API 9	Mytek 3
Tenor 1	CMC6/MK4	API 10	Mytek 4
Tenor 2	CMC6/MK4	API 6	Mytek 5
Amb L	KM130	API 12	192 1
Amb C	CMC6/MK3	API 13	192 2
Amb R	KM130	API 14	192 3
Amb Ls	KM130	API 15	192 4
Amb Rs	KM130	API 16	192 5

I tried hard patching the first 5 analog outputs on the HD 192 to some patch points labeled L, C, R, LS, and RS on the analog patchbay, expecting that I would be able to simply route the audio directly from ProTools. This was unsuccessful, so I closed this session and created a new session (48k/24bit) using what appeared to be a custom surround setup called "Tonmeister Surround," expecting that the monitoring connections would be somehow be normalled. At first I heard nothing, so I checked the I/O setup and found that the outputs for the surround feeds were digital. I then took a look at the AES patchbay and found that I needed to hard patch the 192 to the DSM 7.1. I was now able to hear sound coming from the rear speakers. I went back to the theatre to finish placing my mic's and swap out the bad cable.

The musicians had swapped out the heavy metal music stands that I set up, finding that they were able to hear each other much better when they used their own wire stands. I wanted to try aiming the mic's under the music stand, remembering how impressed I was with the sound at the rehearsal when I knelt down in the middle of the ensemble. I swapped out what I thought was the bad mic cable, adjusted my mic's and went back to the control room to have a listen.

At this point I realised that the center speaker was not powered on, so I walked around the back of it to correct this. I threw the power switch twice, but there was no change in the speaker's condition. I traced the power cable to a power strip at the base of the speaker stand. There was no LED to indicate the status of the power strip, so I flipped the switch to what I thought was the "ON" position and powered off the computer, mixing board, and ProTools interfaces. Realizing my mistake, I said a few choice words and turned the power strip back on. Of course, I had to start the computer and wait for everything to come back online again. I opened the session and got back to work. At this point, I abandoned the idea of using the center speaker and decided to use aux sends to feed the bass and center ambience mic's equally to the right and left speaker, generating a phantom image instead of using a discrete feed. This proved to be an adequate solution to the problem.

The low mic placement was wonderful for the bass and on the one tenor that I could hear, but the treble viols sounded a bit thin and screechy. I was still not getting any signal to the preamp from the other tenor. Thinking that I may have discovered another bad channel on the snake, I made a return trip to the theatre. I changed the placement of the two treble mic's so that they were above the music stand pointing down at the instrument and changed the snake feed on the tenor from 11 to 4. Returning to the control room, I found that the trebles were sounding much better, but the tenor was still not making its way up to the control room.

I swapped out the mic and cable to the second tenor and plugged it into snake channel 6, thus ensuring that whatever was wrong would now be fixed. Seeing that my preamp was now receiving signal, I hard patched API 6 to the proper Mytek input and the second tenor emerged from the left surround speaker. Finally, I had signal from all of my mic's, I could monitor in surround, I was happy with the sound, my levels were good and I was ready to roll. All of this had taken nearly half an hour longer than I anticipated.

The musicians had chosen Jenkins' Fantasy No. 3 for the session and were more or less ready for a take at this point. They did a complete take of the piece and I invited them up to the control room for a listen to ensure that they were happy with the sound before I went too far into the session. Fortunately, they were extremely pleased with the quality and the blend. The separation of the instruments in the speakers lent a bit of clarity to each of the contrapuntal lines as I had hoped, but there was a real sense of blend and ensemble between the bleed of the spot mic's and their mixture with the hall mic's. At this point, the hardest part of my job was past and we could get down to work. The musicians spent some time discussing how they could improve the performance and returned to the stage.

I was disappointed to find that I could not monitor the stereo downmix in my trusty headphones while I listened in surround. The control room in H is quite noisy, the speaker placement is less than optimal and the room is quite active acoustically, so I hoped to check the sound in my familiar Sony 7506's. I was unable to figure out a quick and easy way to do this without disturbing the surround mix and I wanted to devote my attention to the score, so I abandoned the idea and focused on the performance instead.

Occasionally throughout the evening, a player's seating position shifted slightly, for instance after sitting down from a stretch, and this unexpectedly had a drastic effect on the stability of the image. I had to adjust the mic placement, or have the player shift their position accordingly. Most notably, I felt that there was an isolation issue with the tenor and treble viols towards the end of the evening. Since we were doing takes in sections at that point in the evening, I was reluctant to move the mic or the player since this might affect my ability to edit the sections together. Overall, the session seemed to be a success. The musicians did about twenty takes and called it a night around 10:30.

Clocking Problems

I copied the session to my hard drive, packed everything up and headed home, thinking that my troubles were behind me. When I got home I did a quick listen over headphones on my Mbox. The music sounded a little thin and when there were sections of people speaking, there was a slight chipmunk effect. It was at this point that I realised ProTools in Studio H must have been slaved to a master clock. I set the session for 48 kHz, but the master clock must have been set to 44.1 kHz. Now everything was sped up and pitch shifted. How could I recover? I could not simply export and convert the files - they would be permanently enshrined in their chipmunk-like state, regardless of the sample rate. The only solution would be to somehow digitally re-record the material while it was played back at the proper sample rate. I would have to abandon my hope to keep the session at 48 kHz, but at least I would have a usable result.

I attempted an experiment to see if this would work. I ran a SPDIF cable from a Digi 002 rack to the input of the Mbox and set the clock to slave to SPIDIF. The clock in the 002 was internally set to 44.1 kHz. This produced the desired result - the chipmunk effect was corrected. When I reverted to the internal clock on the MBox, the chipmunk effect returned. This proved that it was in fact possible to correct the playback speed by using an external clock, and ProTools would behave even though the external clock signal was different from that of the session. Of course, since this was a 10-track session, I needed a multitrack solution. The tools I had at my disposal were a Digi 002 rack, an Alesis HD24 and an Apogee AD16.

The situation was resolved thusly: I ran a word clock cable from the AD16 to the HD24 and set the Apogee at a 44.1 kHz sampling rate and set the HD24 to slave to the external word clock signal. I then ran an ADAT cable from the AD16 to the optical input on the 002, thereby ensuring that both devices would be receiving the exact same clock signal. I ran another ADAT cable from the 002's optical output to the ADAT input on the HD24. I opened the session in ProTools using the 002 as my interface and set the clock to slave to ADAT. The 44.1 kHz sampling rate LED on the 002 blinked at me constantly throughout the operation, but otherwise ProTools seemed happy with the arrangement. The pitch and playback speed were now correct. I created a new song on the HD24 and set its sampling rate to 44.1 kHz. I then created 8 Aux inputs in ProTools and wired the analog outputs of the HD24 to the analog inputs of the 002 so I could monitor the results of the transfer. A test run affirmed that there would be no clock errors in my signal and the sampling rate problem would be corrected. Since I would only be able to digitally transfer 8 tracks at a time, this would have to be done in two passes. The last step was to record an impulse onto every track in the session so that I would be able to line up the last two tracks with the first 8. This was accomplished by setting all of the track inputs in ProTools to Mic 1 on the Digi 002. I plugged in a condenser mic and turned on the phantom power while simultaneously recording this signal to all tracks. I then copied this impulse to the end of the session on every track to ensure that there was no drift during the transfers.

I recorded the first 8 and the last two tracks to the HD24 at 44.1 kHz/24-bit. I then set up a new session in ProTools for 44.1 kHz and wired up the optical output of the HD24 to the optical input of the 002. On the first pass, I re-recorded the first 8 tracks back into ProTools and then the last two (which were the ambient surround mic's) on a second pass. Using the impulse at the beginnings of these tracks, I was able to line up the audio to sample accuracy. The ordeal was finally over - the session was safe in the bag and the mixing and editing could begin.

Even without taking into account the fact that I would not conceive of using ten microphones to record an ensemble like this for a stereo project. one can see that the PITA factor was unusually high for this session. While much of the aggravation can be attributed to the fact that I was working in an unfamiliar studio with no assistance, some of the complication was a direct result of the fact that I was doing a surround recording. The studio was ready to roll for a stereo session with a CD project in mind – no patching required – just sit down, turn it on and you can begin tracking, which is the case in almost any recording studio on the planet. Trying to get the surround monitoring system to work took a bit of time and work that would otherwise not have entered into the equation. If this recording were for a CD release, I would have left my clock set for 44.1 kHz, but since my target medium is DVD-A, I opted to use the 48 kHz sampling rate – which proved to be a problem after the fact.

I found that editing Renaissance consort music is a tricky affair. Because of the contrapuntal texture, there are not many unison cadence points. Because each line is largely independent, each take is slightly different for each of the five instruments, making edits between cadential points difficult at best. To compound the problem, this ensemble was not playing with the aid of a reference such as a metronome or click (even to start off), and so there was not a reliable way to ensure that there was no tempo drift between takes. I was glad to find that I was able to do my edits working over headphones – a plus for me since I live in a small basement apartment and I don't wish to drive my neighbors (or my wife) crazy with viol music at 2:00 in the morning.

As I suspected when planning the session, some image issues in surround were audible when listening in stereo. The image instability between the two tenor viols, which I had discerned in the surround speakers in Studio H, was also apparent when monitoring in headphones. At the time, I did not have the option to listen in surround in my home studio, so I had to work in two-channel stereo. To monitor the recording while editing, I had each tenor panned half left and half right. At those points where I had noticed instability during the recording session, one tenor would "pull" to the center or just suffer from a bit of smear in the stereo image. This implied to me that it might be possible to do a surround recording of this nature without actually having to monitor in surround at all times. The implications of this could be a boon to situations that are common to engineers who often work outside the studio. When working on-location, one often has to set up in the same room as the musicians. At the very least, even when working with loudspeakers, monitoring conditions are less than ideal in most circumstances. If the engineer can still rely on trusted headphones to monitor while working, one could imagine that a surround recording could, at least part of the time, be monitored in stereo.

Session 2: On Location on Staten Island

We decided to do a second recording session so that the consort would have more time to prepare, and also because we would be able to accomplish more in an all-day recording session. The venue was William Monical's violin shop and showroom down in Staten Island. This type of session was more what I had in mind to do from the outset: rather than using one of NYU's studios, I wanted to see what it was like to run a surround session on-location. At least I knew what to expect gear-wise since I would be using my own rig. One thing I lacked was the ability to monitor in surround. Despite my find from the previous session about monitoring in stereo, I felt that it would be wise to at least have the option available.

To solve this problem, I purchased five KRK RP5 speakers. These are powered monitors (which eliminates the need for separate amplifiers) that generally run about $150 each. I managed to get all five for $600.00 even. I later performed some experiments on them to see how well these speakers actually work and I must say that they fall far short of what I would describe as a reference monitor. My stereo monitors for my home studio are Event TR6 speakers, which perform reasonably well for the price that I paid, but these are discontinued, not to mention a bit larger and heavier than the KRK monitors – also a factor in my purchasing decision. While I would have preferred to purchase the KRK V4 model – quite a bit smaller and perhaps better quality – these would have blown my budget at $200 apiece, $1000 for the set. Forget about buying five Genelecs – at $400 each for the smallest monitor, I would need to invest $2000.

Of course, the five monitors add a bit of PITA factor, since these speakers must also be lugged around in addition to the rest of my gear. For this session in particular, the extra lugging is actually a major factor since there are a number of stairs that one must navigate to enter the shop, and the speakers can only be carried two at a time (at best). This resulted in more than three additional trips on each load-in or load-out. However, other than the additional baggage and schlepping, as well as the need for additional power and the extra step of wiring the speakers to my interface, this session was really no more difficult than any other multi-track recording session I have done in the past.

Not enough can be said for William Monical. He is the caretaker and technician for the instruments of the NYU viol consort and being so was an excellent person to have around for the session. Not only was he kind enough to open his home and shop to us on a Sunday, but he proved to be an invaluable asset: he was able to give helpful pointers to the musicians, and he ended up being a great set of “extra ears,” following the score on his own and providing an expert opinion on each take. His insight due to his intimate knowledge of the music and the instruments proved to be priceless.

This room is much smaller than Loewe Theatre and so the ambient microphones would probably receive more direct sound from the ensemble, not to mention there would be a much higher incidence of strong first and second reflections and less ambient reverberation due to the smaller dimensions. As a result, I decided that a five-omni array would probably not be the best way to go. Instead, I opted for a modified Hamasaki Square.

This technique calls for four microphones, but in keeping with my original philosophy that I would be using a mixture of close mic’s and room mic’s to achieve the end result, and I might need to be able to add ambience to the bass viol in the center channel, I decided I would put up a fifth “center” microphone as an addition to the four mic’s of the square. In the worst-case scenario, I simply could choose not to use it. The most common implementation of the Hamasaki Square calls for four figure-of-eight microphones with their nulls pointed at the ensemble, forming a square with each microphone at the corner, each approximately 2 meters apart. Another implementation calls for the rear two microphones to be cardioids, also with their nulls facing the ensemble.

Not having enough bidirectional microphones to pull off this technique on my own, I was able to borrow a pair of C414's from NYU, hoping to use them with my own pair of C414B-ULS's to form the archetypical Hamasaki Square. At the very least, I would have four microphones with very similar characteristics, should I opt to use the figure-eight and cardioid square. Here we see another barrier to surround recording for the smaller on-location outfit. It is common to purchase microphones in pairs, but not as common to purchase four or five of the same model at a time. I have many good microphones as two-of-a-kind, but I do not own four or five of any one model. If I were not a student at NYU, I would have had to borrow from a friend, rent, or worse: it might have been necessary to go out and purchase additional microphones. At $800 to over $1000 apiece, this could be an expensive proposition, each mic costing almost a full day's earnings for even a well-established outfit. It is an investment, but a costly barrier to entry.

But even my borrowing scheme went slightly awry: I failed to specify that I needed the ULS model of C414, and found that I had been loaned the higher-end TL-II model instead. These microphones have more of a high-frequency boost than their ULS counterparts since they are designed for recording in the diffuse field. This was not such a big deal to overcome: I opted to use a pair of TLM-103's, which have a similar boost to the TL-II's, as the rear pair of microphones in the square. For the extra "center" mic in between the 103's, I employed a U87, set slightly closer to the wall than the two flanking 103's in a manner similar to the Decca-Tree.

As before, I had the ensemble sit in a five-pointed star configuration. The room has a hardwood floor that is mostly bare. William suggested that the ensemble should be seated on a large area rug, as this had been used in the past when recordings had been done in his shop to dampen reflections from the floor. I wanted to seat the musicians a little further apart than in the NYU session to try to get a bit more isolation, which would hopefully translate to better image stability. I also decided to place the microphones somewhat closer to the instruments, risking a less desirable sound to obtain better image clarity. As before, I was using Schoeps MK4 cardioids as the close mic's. Fortunately, the rug proved to be just large enough to accommodate the ensemble.

At my disposal, I had eight channels of Millennia HV-3 preamps, four channels of Hardy M-2 preamps, and four channels of Sytek MP4-Aii. I felt that it would be important to try to match the close mic's as much as possible. Towards this end, I used the first five channels of Millennia for these microphones. I used the remaining three channels of

Millennia for the left, center, and right ambient microphones, and used two channels of the Hardy for the surround ambient mic's. I needed only ten inputs, so I could utilize my Apogee AD16 and still have channels left over. Since I am recording at 48 kHz, the first 8 digital outputs could be split from the Apogee – one cable feeding the HD24 and the other feeding the Digi 002 rack. I used two XLR Y-cables to split the output of t channels 1 and 2 on the Hardy – the Ambient LS and RS mic's - between the Apogee and the analog inputs on my 002.

Source	Mic	Preamp	Converter/Input
Treble 1	CMC64	Millennia 1	Apogee 1
Treble 2	CMC64	Millennia 2	Apogee 2
Bass	CMC64	Millennia 3	Apogee 3
Tenor 1	CMC64	Millennia 4	Apogee 4
Tenor 2	CMC64	Millennia 5	Apogee 5
Ambient L	TLM103	Millennia 6	Apogee 6
Ambient R	TLM103	Millennia 7	Apogee 7
Ambient C	U87	Millennia 8	Apogee 8
Ambient LS	C414-TLII	Hardy 1	Apogee 9/Digi 5
Ambient RS	C414-TLII	Hardy 2	Apogee 10/Digi 6

The monitoring setup was another story. Because of my relatively close proximity to the musicians, I did not want the speakers to be active while they were tracking. Also, I wanted to check my hypothesis that it would indeed be possible to monitor in stereo and predict how stable the image would be in surround. I would be monitoring through my Digi 002 rack using ProTools LE, which does not support surround sound. I connected the speakers to Analog outputs 3 through 7, which would enable me to reserve outputs 1 and 2 for a stereo headphone mix. In this way, I could easily mute the speakers while tracking and still monitor with my headphones.

I assigned each track's main output to Analog 1-2 and utilizing an auxiliary send on each track, I sent each of the mic's to its corresponding speaker:

Source	Speaker
Treble 1	Left
Treble 2	Right
Tenor 1	Left Surround
Tenor 2	Right Surround
Bass	Center
Ambient Left	Left
Ambient Right	Right
Ambient Left Surround	Left Surround
Ambient Right Surround	Right Surround
Ambient Center	Center

I was able to position myself between two workbenches in Bill Monical's workshop, which is adjacent to the main showroom where the ensemble was. The front speakers were laid out on the bench in front of me and the two surrounds were positioned behind. It more or less corresponded to the NARAS specifications, with a wider stereo pair behind me, and the left, right and center channels set up more or less as they should be.

Post production

I wished to be able to finish as much of this project as I could in my home studio. Not only is this more convenient for me, but it would be good for me to be able to repeat as much of this in the future so that I can offer the option of surround sound recording to future clients. I now was the proud owner of five cheap monitors, but my work environment is only set up for stereo monitoring. At the very least, I had to purchase a pair of monitor stands so that I could set up my two surround speakers.

I purchased a pair of cheap Samson monitor stands for $65. The adage "you get what you pay for" stands here; I brought the first pair home and found that one of the stands was completely defective. I returned to the store and exchanged the stands for a new pair, brought them home and set them up. As I was tightening the collar on one of the stands after making the height adjustment, the collar (which is made of plastic) snapped. Unwilling at this point to make a third trip on the same day, I got out my trusty electrical tape and made and improvised collar, which more or less did the job. I replaced my beloved Events with the KRK's in the stereo left and right positions, placed the center speaker on top of my 19" Sony CRT monitor (a rather risky, and less than ideal, placement, but I had no other option), and positioned my surround speakers in the recommended ITU placement. I ran cables and power for the additional three speakers, using outputs 1 – 5 on the Digi 002 rack for L, R, Ls, Rs

In addition to ProTools LE, there is an elderly version of Nuendo (v 1.5.1) on my editing PC, which could do the job of true surround mixing if I so desired. This would involve learning my way around a DAW with which I have little or no experience at a point where time is of the essence. I did go through the steps of importing the audio for one of the pieces into Nuendo – a rather time consuming process since there were 26 takes involved in the first piece, each of which has 10 corresponding audio files. Nonetheless, I set each of the tracks up properly on the timeline and made sure that my I/O was setup correctly. After I realized how much time it had taken to do this for a single piece, and realizing I had to repeat this entire process for four songs, I decided that I would do all of my editing in ProTools, and perhaps import the edited audio into Nuendo for the final mix and bounce.

Since ProTools LE does not support surround sound, I had to improvise. For the monitoring it was not difficult since I was not planning on interchannel panning or any such effects; I simply set up each track to feed a bus corresponding to the proper speaker. Since I do not have proper surround monitoring controls, I set up 5 Auxiliary tracks, one each to serve as the master volume control for each speaker, and grouped these faders so I could mute and adjust the monitoring levels simultaneously.

My next problem was not quite so simple: the room in which I recorded was small and there was no ring to it; I needed to add some reverb to the recording. I do not own a surround reverb, but even if I did I could not create a surround track on which to put it. I needed to improvise some sort of surround reverb in this stereo editing environment. I resolved the situation in this way: I created two stereo tracks, placing the same type of stereo reverb on each, one representing the front half of the listening area, and the other representing the rear; each of these tracks was fed by an independent stereo bus. The front instruments were fed to the front reverb, and also to the rear, but the level to the rear reverb was reduced by 6 dB. Likewise, the instruments that were located in the rear speakers were fed to the rear reverb, but also to the front reverb with lower levels. When listened to by itself, the reverb did not exactly sound convincing, but it seemed to give the illusion of a bit more space around and behind the instruments. Interestingly enough, the reverb would actually localize between the front and rear speakers. I thought of adding a bit of delay to the sends, but decided to come back to this idea if there was time.

The closer mic spacing in combination with the slightly more distant seating between players seemed to work well – the surround image was pretty solid even though there was still plenty of bleed between the microphones. Perhaps the homogeneity of the sound suffered just a little, but it is difficult to say if that were solely attributable to the microphone placement. The closer spacing did capture a bit more of the bow noise and other artifacts, and also made the dry spot mic's a little claustrophobic sounding, but when mixed with the ambient array and the "pseudo-surround" reverb, an acceptable sound could be obtained.

I was surprised to find that the surround effect was a bit distracting when I first began attempting to do edits. For some reason I felt a little unsure of the edit points –not something I normally would experience when working in stereo. This may have just been psychological, or perhaps it was because I chose the most difficult piece to edit starting off – Byrd's *Fantasia of Five Parts*. This five-and-a-half minute piece has only one unison cadence point, other than the ending (and even that cadence stretches the definition of unison). Once I got used to it, this did not seem to be much of an impediment. The clarity of the counterpoint in surround did serve as a bit of a barrier to easy editing – it was no longer possible to attempt to "bury" mistakes in the mix – flubbed notes in the inner voices that were not so prominent when monitoring in stereo were now painfully clear.

On the other hand, the use of close microphones allowed a freedom with editing that I would not have had otherwise. It was possible to overlap edits on the entrances in such a way that would have been impossible with a typical stereo recording. To make sure that the surround mixing was not masking some of the edits. I checked the first edit on headphones and was satisfied that the edit was not discernable in stereo either. I later did much of my editing with headphones on, as I normally would working in stereo.

At first, mixing also seemed a bit more difficult than in stereo. Part of it was the closer mic'ing, but I had a feeling that I was giving too much prominence to the trebles. As the high voice, it is natural for them to stick out a bit more, but being in the front speakers, it seemed as though I might have been featuring them even more. The thought occurred to me that since we check stereo mixes in mono, perhaps it would be just as useful to check a surround mix in stereo. I tried this experiment, and much to my surprise, I had mixed the tenors forward of the trebles. Also surprisingly, the bass was more prominent in the stereo mix than when mixed to its own speaker in surround.

After I had completed the edits, I decided to try a different way of publishing the mix. Since ProTools LE does not allow a surround bounce to disk, I could either export my edits to Nuendo or I could simply record the separate stems in real time and export these as my surround files. I often use a similar technique in stereo, where I simply bus all of the tracks to a single stereo track and mix in real time saving the need for complicated automation and such. I did this, and mixed the viols in real time to five separate audio tracks – one for each speaker. After I finished the pass, I exported the five files as 48k/24-bit AIFF files to a subfolder of the session.

I repeated this process for the edited audio for the NYU session that was recorded in December. Interestingly enough, there was image instability for the first treble that I haed not realized. Occasionally it would jump from the front speaker to the rear where the first tenor was, or spread itself out between the two speakers. This instability would have been nearly impossible to detect in stereo since the two instruments were panned to the same side. It also was virtually undetectable in Studio H where the session was recorded. As you may recall, I had the surround speakers in the NARAS format, which calls for the surround speakers to be farther away from the front speakers. In my small home listening environment, these problems were immediately apparent.

Delivery

It was time to take DiscWelder Bronze for a test drive. DiscWelder is a line of DVD-Audio authoring software distributed by Minnetonka Audio Software. The Bronze version is the bottom-of-the-line version and I purchased it from Full Compass for about $75. Unlike its far more expensive brethren, It will not allow any compression schemes such as MLP or Dolby, not does it have any video capability or support for DSD. It will only write LPCM files to a DVD-R. The disc it generates would only play back in players that supported the DVD-Audio spec AND could read DVD-R. This was just fine for my purposes, since I had no thoughts of complicated authoring schemes. If I had such visions, my price tag could have been as high as $3000.00

I opened the program and decided to make a test disc consisting of the first NYU session as well as one of the songs from the Staten Island session. DiscWelder has a relatively simple interface, though it was not immediately intuitive to me how I was to proceed. I figured it out in short order (there was not much to figure out) and burned a

DVD-R in my ancient DVD burner at its maximum speed of 2x. For a bit over eight minutes of material, this took a total of about 7 minutes.

Unfortunately, I have no idea if the process was successful. I attempted to play the disc in my own home theatre system, but sadly my own DVD player does not support DVD-Audio. Although I suspected this at the outset, I was still disappointed to see "Disc Error" on the display. In all fairness, many DVD players do support DVD Audio, and it is becoming more and more common, but here is yet another reason why surround sound recordings have not caught on with the listening public.

In addition to delivering the final product as a DVD-A with its five channels of full 48 kHz 24-bit LPCM quality, I would also like to produce some sort of compressed surround sound format so that the mix could ostensibly be distributed over the web or even streamed from the client's website (someday). Since classical music seems to have fared better on the web and the trend for many ensembles (such as the New York Philharmonic) has been to post live recordings of concerts on the web, this may eventually be the place where surround sound comes into its own. Many PC users (especially gamers) are already set up for 5.1 and Windows Media Player versions 9 and up support surround sound formats. It was in attempting to find a compressed surround sound format that I experienced the most difficulty.

In my first forays into this part of the project I was attracted to the MPEG family. Since MP3 is the most ubiquitous form of compressed audio, it made sense that this might be the way to go. I began with some white papers on MPEG surround formats on the Fraunhofer IIS website. I was very attracted to the idea that it was possible to have something that was encoded for surround sound, but could be handled by people who are set up only for two-channel stereo. The Fraunhofer website seemed to suggest that this was possible with one of their products. Despite all of the promise of surround sound for the MPEG family, there is not yet a product that can accomplish all of these wonderful things, since there are several companies involved and, as of this writing, the licensing has not yet been worked out.

Next I was attracted to the Windows Media Player 9 encoder. I was not able to find sufficient documentation on this encoder, although for me the product would be free of charge. That being said, Windows Media Player is universal and easy to obtain for both Mac and PC users, making it an attractive format for web distribution. It was unclear to me at first as to whether or not it would perform compression on full-blown PCM audio files, and whether or not the material had to be streamed from a server or if it could be downloaded like an MP3.

Lastly, I explored the possibility of Dolby Digital or DTS as the compressed format. A little bit of digging led me to the SurCode product line of Minnetonka Audio Software. Surely one of these formats would at least be compatible with consumer DVD players, which would be the most likely device to be set up for 5.1 surround sound in most households. Once again, the audio community shoots itself in the foot by having two competing formats. Minnetonka has a CD-DTS encoder available for an MSRP of $99.00. Theoretically, I could author a CD in this format and play it in a typical DVD player (although some DVD players cannot read CD-R's). Fortunately, most DVD players support both Dolby and DTS encoding.

Technically, Dolby Digital (AC-3 file format) would be the best of the two for web distribution since it has a higher compression ration (8:1 as opposed to DTS's 3:1), not to mention the fact that AC-3 is a file format and could theoretically be played back on computer players that support the Dolby Digital format. This format is also backwards-compatible with two-channel stereo systems. However, the Dolby Digital Encoder is much more expensive (the 5.1 version has an MSRP of $995.00, street price is about $400). Fortunately, NYU has this software at the school, so issuing the files in this format costs me nothing.

Yet the distribution of this format remains a problem. To my knowledge, the only way to distribute the audio in Dolby AC-3 is to put it in a DVD. This DVD would probably be a video disc, since that would no doubt be the most playable format. The question for this becomes: what do you do with the video part? It is also a bit of a pain for the consumer. Unlike an MP3 sort of setup where you can preview the material and then it is simply a matter of downloading the file and it is instantly available for playback, this would involve downloading the the file, creating some sort of video or audio DVD (which many consumers will not have access to the necessary tools to encode surround sound on either format), and then being able to hear it in your home theatre system. It seems that the greatest weakness of this format is at the very last point: the point of delivery. Until this becomes a simpler and less costly process, I cannot imagine the consumer being very interested in pursuing it.

That being said, authoring for DVD-A is a relatively simple affair. The software is readily available and not particularly expensive with a street price of about $75.00. Since many PC's and Mac's have DVD burners on board, this seems to be the easiest way to go about distributing surround sound recordings. However, the average consumer does not have a DVD-Audio player, or even a universal player, and there does not seem to be a dramatic shift in that direction. As mentioned previously, some DVD players don't support the DVD-Audio format. It certainly does not help the popularity of the format if the most easily authored and created format is not the easiest to play back, as is the case with CD and CD-R.

Section 5: Conclusion

So is it worth it? For this project I have to say it was. The excitement it generated among the musicians brought a freshness to the recording sessions that is not normally the case. The music itself benefits from this type of recording – the contrapuntal lines are easier to follow and there is no masking of any of the voices. The conversations and duets between parts become immediately apparent. Professor Panofsky feels that there may some educational value to this type of recording. Viol players would be able to hear and learn their parts with greater ease if more recordings were made in this manner. I could see a similar application in the choral world where some parts, especially inner voices, get lost in the mix.

Another benefit that chamber music gets from this sort of recording is the restoration of intimacy between player and listener. One of the players even remarked that it allows the listener to experience the feeling of being a part of the ensemble instead of just a passive listener. The immersion in the music, the feeling of being completely enveloped in the sound of the music, and the ability to take part in the musical conversation are all parts of the experience that keeps many ensemble players coming back for more. This sort of recording can bring more of that experience to life for the listener. My wife, not really a connoisseur of western classical music, sat in the midst of my little surround setup with a smile and gave her verbal stamp of approval, "I like!"

On the down side, this type of presentation can be a bit distracting until you become accustomed to it. Also, since I insisted on putting the trebles up front the focus is still on the front pair – the trebles call more for your attention due to the fact that they are higher in pitch as well as the fact that they are in front of you. Thus, even in surround they can overshadow the tenors at times.

On the cost side, I managed to get off relatively cheaply – less than $1000. However, I would not recommend the monitors that I bought for high-end studio mixing. Also, while I only spent $100 on additional software, the price could have been much higher. If I wanted more professional authoring features and/or the option for Dolby AC-3 encoding, at least another $500 would be necessary. Schoeps just raised their prices again, bringing the cost of a pair of cardioids up to nearly $3000. If I did not have access to NYU's mic closet, I probably would have needed to invest at least another $5000 in microphones alone – a high price to pay for an entry-level recording. For well-established Tonmeisters who already own 10 Schoeps, DPA's or other appropriate microphones, this might not be an issue.

Perhaps the highest cost was in time. While some of that time was spent resolving first-time surround recording issues, doing a surround project involves much more time than a comparable multitrack stereo recording, both in pre- and post-production.

Perhaps the greatest surprise to me was the relative ease with which I was able to do a surround recording. Another surprise was the amount of time I did not need to be working in surround sound, at least for this project. I could edit working in stereo (if need be), and I could even track while listening in stereo, although I would not recommend this without having a way to monitor in surround – there is no substitute for really being able to hear what is going on, especially with the situation of front-to-back image instability. Whether you like it or not, the monitor investment cannot be avoided because at the end of the day, no matter what, one has to mix the project in surround.

Other than the authoring software, the only additional equipment that is a real necessity is the monitoring equipment. Cheap monitors simply won't do for this, and never mind if you already have money invested in a pair of older, discontinued monitors (Genelec owners take note). There is no way around buying the additional monitors and whatever hardware you need to deploy them. Included here is a table outlining just some of the additional costs one might encounter when moving to surround recording.

Stuff	Stereo	Surround	Price Difference
Mic's (minimum)	2	5	$4000 +
"T-Bar"	SabraSom $40 Schoeps $400.00	SabraSom $750 OCT Bar $2050	$700.00 $1600.00
Recording Software	Almost Any	Nuendo ProTools TDM Pyramix, etc.	$2000 - $10,000
Authoring	Free!	Diskwelder	$70.00
Data Compression	Free!	SurCode	$400.00
Monitors	2 or Headphones	5	$600.00 +
PITA	Low to High	High	Priceless!

The pitfalls of image stability when recording in surround are a bit disturbing. It is not simply a matter of setting up five microphones and pressing record. Much attention must be given to the crosstalk between channels – especially front to back. Unlike in stereo, where an image may wander a bit in between two speakers, it is more disturbing when an image jumps behind you, or gets smeared in between two speakers that are a large distance apart. The fact that you can never be assured of how the speakers will be laid out at the listening end is a huge disadvantage. In two-channel stereo, the worst that can happen is that the left and right channels get reversed, or that the two speakers are not really in the optimal position. With surround, all bets are off and the effect that the recording can have on the listener will be impacted directly.

I hope to get the opportunity to record in surround again. It would be nice to be confronted with a scenario calling for a larger scale such as an orchestra, choir or combination of the two, though I would no doubt have to change my approach for a different ensemble. I do believe that more surround recording should be done in general. The results are worth it, if it is not too much of a stretch timewise and financially, and the outcome generally seem to give good two-channel recordings as well. The more that is done in this format, the better the techniques and the tools will become. Perhaps one day the media, distribution, and consumers will catch up.

Sources

[1] http://www.music.iastate.edu/antiqua/t_viol.htm

[2] http://en.wikipedia.org/wiki/Viol

[3] http://vdgsa.org/pgs/stuff.html

[4] http://www.classical.net/music/comp.lst/byrd.html

[5] http://www.classical.net/music/comp.lst/jenkins.html

[6] Ashbee, Andrew. *The Four-Part Consort Music of John Jenkins.* Proceedings of the Royal Music Association, Vol. 96. (1969 - 1970). pp. 29-42.

[7] Edwards, Warwick A. *The Performance of Ensemble Music in Elizabethan England.* Proceedings of the Royal Music Association, Vol. 97. (1970-1971), pp. 113-123

[8] Robinson, Lucy. *John Ward, 'to satisfie both quickness of heart and hand'*. Early Music. February 2006 pp. 157-158

[9] Nymand, Mikkel; Brixen, Eddy Bogh; Stove, Morten. Miking with DPA Microphones, A workshop in Surround Miking. Banff - June 2003 http://www.dpamicrophones.com/

[10] Hanning, Barbara Russano. Concise History of Western Music, 2nd Ed. pp. 146-149. W.W. Norton & Company 2002

[11] AES staff writer. Stereophonic Recording Techniques: Old Challenges, New Approaches. J. Audio Eng. Soc., Vol. 54, No. 3, 2006 March

[12] Theile, Günther. Multichannel Natural Music Recording Based on Psychoacoustic Principles. www.irt.de/wittek/hauptmikrofon/theile/Multich_Recording_30.Oct.2001_.PDF

[13] Streicher, R. and Everest, F.A. The New Stereo Soundbook. 2nd Edition. Audio Engineering Associates. 1998

[14] Williams, Michael and Le Dû, Guillaume. *The Quick Reference Guide to Multichannel Microphone Arrays Part 1 : using Cardioid Microphones*. Audio Engineering Society Convention Paper 5336. 110th Convention 2001 May

[15] Notes from presentation by Paul Geluso. AES meeting November 3, 2006 at the Clive Davis School at NYU

[16] Lipshitz, Stanley P. *Stereo Microphone Techniques: Are the Purists Wrong?* Audio Engineering Society Convention Paper 2261. 78th Convention 1985 May

[18] Bregman, Albert: Auditory Scene Analysis.
The MIT Press, 1990.

[19] Berg, Richard E., Stork, David G.: The Physics of Sound 2nd Ed.
Prentice Hall 1995

[20] Daniel, J., Nicol, R., Moreau, S.: Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging
AES Convention Paper 5788, 114th Convention. March 2003

[21] Bartlett, Bruce, Bartlett, Jenny: On-Location Recording Techniques
Focal Press, 1999

[22] Malham, D.G.: Approaches to Spatialisation
Cambridge University Press. 1998

[23] Begault, Durand R.: 3-D Sound For Virtual Reality and Multimedia
Academic Press 1994

[24] Massey, Howard. Editor. The Recording Academy Producers and Engineers Wing: Recommendations for Surround Sound Production. National Academy of Recording Arts and Sciences 2004

[25] Williams, Michael. Multichannel Sound Recording Practice Using Microphone Arrays. AES 24th International Conference on Multichannel Audio

[26] AES Technical Council. Multichannel Surround Sound Systems and Operations. Document AESTD1001.1.01-10

[27] Malham, D.G. SPATIAL HEARING MECHANISMS and SOUND REPRODUCTION. University of York, England 1998

[28] Phone conversation with Craig of Minnetonka audio 12/5/06 at 6:00 pm EST
Craig@MinnetonkaAudio.com

[29] Schoeps Mikrofone U.S. Pro Net Price List. February 2005

[30] Streicher, Ron. *The Decca Tree – It's Not Just for Stereo Any More*. February 2003

[31] AES Staff Writer. *Novel Surround Sound Microphone and Panning Techniques*. J. Audio Eng. Soc., Vol. 52, No. 1/2, 2004 January/February

[32] Williams, Michael. *The Stereophonic Zoom*. Rycote 2002

[33] *http://www.wendy carlos.com/surround.html* and *surround2.html*

[34] Notes from a Meeting of the New York Section of the Audio Engineering Society:
"Surround 4.0: The History of Quad from the '70's to the present day"
Presenters: Robert Auld and Jerry Bruck. April 10, 2007

[35] Pohlmann, Ken C.Principals of Digital Audio 4th Ed. McGraw-Hill 2000