RECORDING SOUND FOR 3/2

INTRODUCTION
&
CRITICAL REVIEW

How humans perceive sound, how people are trying to record and recreate sound, and why there still are a lot of things to improve upon in recording and recreation of sound.

Arjan van Asselt, MA
June 2002

INDEX

1 – Introduction

Before really getting into the subject, some assumptions and definitions have to be made. There are many terms and expressions in the field of this research, many often wrongly used or incomplete in what they intend to describe.

The word “sound” can mean a lot of things. The dictionary states:

sound[1] (sound)
n.

1.
    a. Vibrations transmitted through an elastic solid or a liquid or gas, with frequencies in the approximate range of 20 to 20,000 hertz, capable of being detected by human organs of hearing.
    b. Transmitted vibrations of any frequency.
    c. The sensation stimulated in the organs of hearing by such vibrations in the air or other medium.
    d. Such sensations considered as a group.
2. A distinctive noise: *a hollow sound.*
3. The distance over which something can be heard: *within sound of my voice.*
4. *Linguistics.*
    a. An articulation made by the vocal apparatus: *a vowel sound.*
    b. The distinctive character of such an articulation: *The words bear and bare have the same sound.*
5. A mental impression; an implication: *didn't like the sound of the invitation.*
6. Auditory material that is recorded, as for a movie.
7. Meaningless noise.
8. *Music.* A distinctive style, as of an orchestra or a singer.
9. *Archaic.* Rumor; report.

In this writing however, the word “sound” or “sound event” will be used as a means to describe a certain event or rather, series of events in “vibration of the air”, in a certain place, room, or venue.

Theoretically, the simplest way sound can be generated, is from a single point in space, where it radiates from, equally in any direction. It pretty much would look like when a stone is thrown into a pond. Although in practice this can never be exactly correct, since this would mean the object generating the sound would be infinitesimally small, it serves good for mathematical purposes.

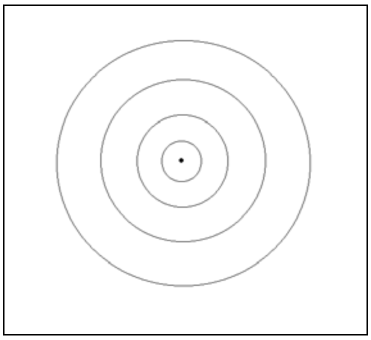


Figure 1 - sound waves radiating from a single point in space

Now, also for mathematical purposes, it can be stated that when being far enough from the source, the spherical waves can be treated as plane waves.

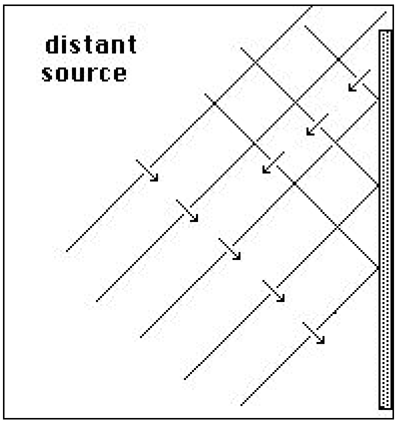


Figure 2 - when coming from a large distance, soundwaves and their reflections can be regarded as plane waves, rather than co-centrical

The complete human hearing mechanism, existing of several “devices”, mostly located in the human head, will be named the “earbrain”. The first and probably most important part of the human hearing mechanism, which in itself can be divided into several sections, for instance outer-ear, inner-ear, and so forth, will be referred to as “ear”. The second, also very important part, is the human brain, which interprets the impulses it receives from the ear. Together they act as a rather sophisticated system, as will be shown in the next chapter.

Currently, the most common way to recreate sound, is by means of so called “loudspeakers”. Loudspeakers can generate vibration of the air through an electrically driven cone. As has been known for quite some time, more loudspeakers can act together and enhance each other in recreating a soundfield.

There are several possibilities in the number of speakers that can be used, and how they can be positioned. The most commonly known is the 2-speaker stereo setup, which will be explained later. There are many other possibilities, the majority of them only being used in experimental circumstances

For this writing there will be a focus on the so-called “3/2 speaker setup”, which one will also be explained later. By no means this is a “perfect” number of speakers, nor a perfect layout. But since this is the setup that is being used in the majority of larger cinemas, and is advancing rapidly into (modern) living rooms (“hometheatre”- courtesy of DVD-Video), it is the setup most likely to be popular for the years to come.

2 – Human perception of sound

2.1 – Introduction

In fact being a rather more biological and physical matter, the human perception of sounds needs to be understood to a great extent, in order to form and understand a proper theory of the probably more artistical matter of “recording and recreation of sound”.

Not too much can be found on the scientific knowledge of human perception of sound. Although there have been written numerous chapters about it, mostly in books about recording and microphone techniques, these tend to go no further then a rather shallow description of some basic principles.

Yet there is one exception, and it is this work that will be the guideline for this part of the review. The book is called “Spatial Hearing”, and has been written by Jens Blauert. (see “resources”) It would be meaningless to aspire the same thoroughness and level of completeness as can be found in this book. Therefore, here will only be another summary, focusing only on meaningful aspects for 3/2 recording.

2.2 – The human ears

The human ears, theoretically, could be regarded as simple transducers, that convert vibration of air (“sound”) into mechanical vibration, and then into neurotransmitters that can be interpreted by the brain. So, theoretically, the ears act like microphones: they pick up vibrations in the best possible way (meaning: without any unwanted alteration) and convert it into some other kind of vibration. In the case of a microphone it is vibration of the air that gets converted into mechanical vibration (the membrane), and then into electrical vibration. (AC, Alternating Current)

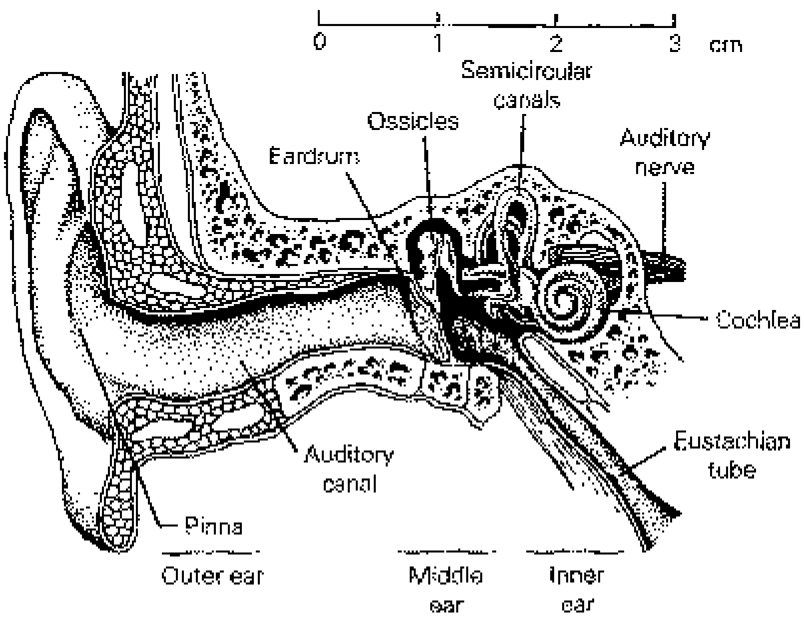


Figure 3 - the human ear

As mentioned in the general introduction, it is the human brain that interprets the data the ears receive. Many aspects of the sounds received by the ear, like pitch, location, distance, duration, etc., are being “tagged” by the brain and used for several purposes.

Humans have two ears, located on either side of the head. Not only can this be seen as a safety-precaution, in case one gets disabled one still has another one to rely on, it also has its use for some other applications. Just like how the two human eyes interact strongly in order to see “depth” and hence be more accurate in interpreting the distance of objects, the human ears make use of the same principle to locate sound sources. The difference of the sound arriving at the two ears can be used in order to interpret where exactly the origin of the sound-event took or is taking place. There are two aspects of this difference that help the brain interpret the location: the time difference between the two receivers, and the level difference between the two receivers.

2.3 – Interaural time difference (ITD)

Because of the distance between the two ears, differing per human being, but typically about 17,5 centimeters, the average diameter of the human head, a sound hardly ever arrives at the two ears at exactly the same point in time. Taking only the horizontal plane into account, then only when the origin of the sound is exactly right in front, or right behind the listener, the time of arrival is exactly the same.

So in any other case, there will be a time difference between the two ears, and the brain can learn how to interpret these difference and make them correspond to an interpretation of where the sound source must be.

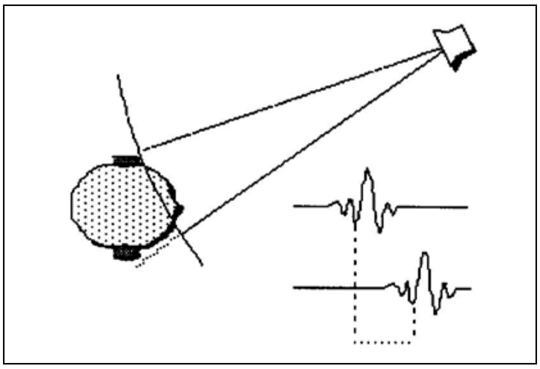


Figure 4 - Interaural Time difference (ITD)

How it is interpreted whether a sound source is coming from the front or the rear (for every time difference, there are two possibilities for the location of the source, one being the exact mirror-image of the other), is covered in the paragraph about the “pinna”.

In many books and writings it is stated that this way of localizing sound sources only works at lower frequencies, because, as will be explained in the next paragraph, interaural level differences only work at higher frequencies. Yet it can be easily shown that interaural time differences does work at any frequency (if only because there is no reason to assume that it doesn't!).

The main reason why interaural time differences generally aren't considered as an important cue at higher frequencies, is because interaural level differences give a much stronger and clearer cue in that frequency range. And since interaural level difference doesn't work at lower frequencies, interaural time difference simply takes over the task, because in that range it's providing the strongest cues.

2.4 – Interaural level difference (ILD)

Also because of the distance between the two ears, but mainly because the separation ("shadowing") and diffraction that the head itself causes, there is a difference in level between the two ears when a sound arrives at the head. Once again, the only exception is when a source is exactly in front or in the rear of the listener. It is also referred to as interaural intensity difference (IID).

This kind of "hearing differences" only works for somewhat higher frequencies, because at lower frequencies the head doesn't provide for an "obstacle". Lower frequencies have longer wavelengths, so do not get affected by smaller objects. If an object is smaller than one half a wavelength in size, it is "invisible" to the sound, therefore it doesn't affect it.

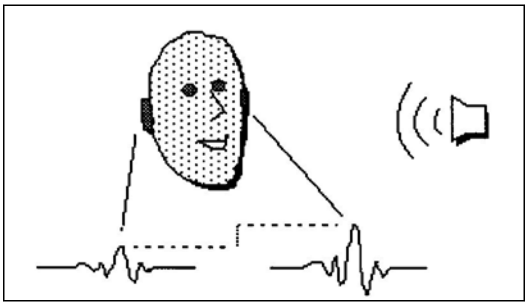


Figure 5 - Interaural Level difference (ILD)

The "crossover frequency" lies at about 1700 to 2000 Hz., depending in the size of the head. So frequencies below 1700 to 2000 Hz. are being located by means of interaural time difference, and frequencies above 1700 to 2000 Hz. by means of interaural level difference.

Frequencies below about 120 Hz cannot, or hardly, be localized by the human earbrain. The wavelength of these frequencies is so large, that the distance between the two ears should be much larger than the average 17,5 centimeters in order to notice any difference in time of arrival. Our heads simply aren't large enough.

Human beings are very good in localizing sharp transient sounds, sounds with a sharp attack and a relatively short decay. These are sounds like crackling of bushes, trees or rocks falling, gunshots, etc. From nature's point of view, this makes total sense, because those are typically the sounds to be aware of. (even though it could be argued nature didn't anticipate for gunshots and the likes)

More extended types of sound-events, like waterfalls, the beep of a TV-channel-when-not-broadcasting (a sine-wave), etc., are much harder to localize. Once again, from nature's point of view, this also makes sense: those sounds typically aren't caused by something that could be a threat or prey.

The reason why this is the way it is, can be explained rather easily. When a sharp and short impulse occurs, the time difference at the two ears can be noticed very easily. When a sound-event lasts longer, the sound (that still is part of the same sound event!) that comes later, can obscure the initial time difference, the earbrain might loose track of what is happening.

Also, when a sound event lasts longer, it is bound to "fill the room", and therefore obscuring the level difference at the two ears. With reflections coming from all around, the level of the reflected sound might get equally loud or even louder than the initial sound. The earbrain is likely to loose track completely of where the sound originally came from.

2.5 – Head movement

There is a third, rather important cue the human earbrain relies upon when localizing sound, in addition to interaural time and level differences. Moving the head (slightly) can provide for additional cues in interaural time and level differences.

It can be easily explained by the analogy of how pigeons (and other birds) can see depth. Because those creatures have eyes on the side of their heads, rather than on the front, like humans and most (all?) other mammals, what they see with one eye doesn't correspond by any means with what they see with the other. Therefore, they cannot see depth the way humans do (as explained earlier in this writing).

But nature has found a way: when holding their head –and eye- in one position a "screenshot" is taken. Then the head –and eye- is moved slightly and quickly, and another "screenshot" is taken. The difference between the two pictures provides cues for depth and distance of the pigeon's surroundings.

Now humans can do the same trick with their hearing mechanism. When the head is in a certain –fixed- position, the brain can memorize the cues (ITD, ILD) that are given by the two ears. These cues can sometimes be inconclusive or even contradictive. But when moving the head slightly, another set of cues can be given, and compared to or "superimposed upon" the cues that are in memory. By collecting cues this way, hypothesis of where the sound event is taking place can be verified or proved to be wrong. A stronger and more conclusive sense of where a certain sound-event is taking place can be obtained this way.

2.6 – Pinna

The pinna (outer ear) is largely in charge of separating sounds coming from the front from sounds coming from the rear, and giving cues about where in the vertical plane a sound event is taking place. The latter, although being very interesting and challenging, will be left alone in this writing, since the focusing is on the 3/2 speaker setup, one of the many that doesn't take sound anywhere outside of the horizontal plane into account.

The pinna acts as a mechanical "equalizer", or rather "audio filter". Sound entering the ear canal will come across the complex structure of the earflap. Part of the sound will just, unaffectedly, enter the ear canal, but another part will be reflected back and forth between the (unique) curves of the earflap, before entering the ear canal. This "bouncing back and forth" of the sound, will cause phase-differences at certain frequencies compared to the direct sound that is entering the ear canal. This causes certain frequencies to become stronger, when both direct and indirect (the sound being "processed" by the pinna) are in phase, or weaker, when both sounds are out of phase.

It can also happen that a certain frequency gets "trapped" in a certain curve of the pinna, when that wavelength is exactly the size of the curve. This by itself will act as a (Helmholtz-) resonator, and will cause amplification of this frequency.

This filtering does happen to any sound that enters the ear, and it only depends on the angle in which the sound enters the ear, how the sound will be filtered. The brain simply learns how to interpret the coloration caused by the pinna. Because the pinna are different for every single human being, everybody has learned how to interpret ones own pinnae.

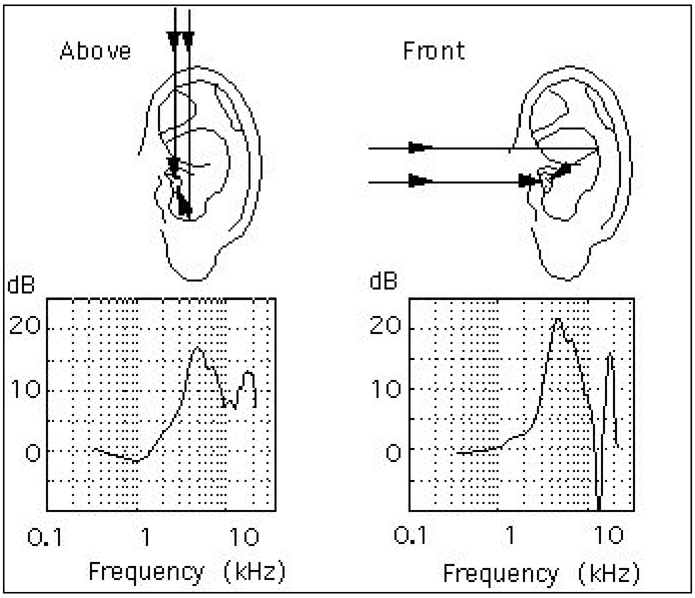


Figure 6 – filtering curves caused by the pinna

Perhaps the strongest filtering takes place when a sound event takes place at the rear of the listener. Then the entire pinna will filter the sound, and because the pinna will act as a boundary for (very) high frequencies, the sound will loose some of its brightness in general.

This implies the listener has to be familiar with the sound in the first place, in order to determine whether the sound is “less bright” than usual, hence it must be coming from the rear. The artifact commonly known as “front-back confusion”, which happens frequently to many listeners, when listening to binaural recordings (with either a dummy head or miniature microphones in somebody’s ears).

All this explains why new sounds, sounds that are unfamiliar to the listener, are harder to localize. It also explains why people have a hard time listening with “somebody else’s ears” (possible when using miniature microphones at “somebody else’s” ears, and listen back through headphones).

Yet it has been proven that people can learn how to listen to “somebody else’s ears”, simply by practicing a lot, hence providing evidence for the assumption that the brain only relies on learning and practicing, in order to interpret the cues provided for by the pinna.

Another reason why human beings aren’t good at localizing sounds with lower frequencies, is because of the size of the pinna. It only affects sounds with higher frequencies, to lower frequencies it is “invisible”.

It is possible to “record” how exactly, for a particular human being, the filtering caused by the pinna and the filtering and diffraction caused by the head takes place. The person under test is equipped with miniature microphones at the entrance of the ear canals. Then, in an anechoic room and at a fixed distance from the head, but with an altering location around the head, impulse responses are recorded.

The impulses are short bursts of white (or pink) noise, that –theoretically- contain all frequencies at an equal level. (for pink noise this is somewhat different, but for measuring purposes it doesn’t matter, as long as the differences are taken into account). Now, by comparing the original impulse, and its responses (the signal recorded by the microphone in the ear), the amount of filtering can be determined.

For every position of the source-location, a graph can be drawn, and a function derived from it. This is commonly known as “Head Related Transfer Function”, or rather “HRTF”. By applying the filtering of a given HRTF at a certain location to a (monaurally recorded) sound, it should –theoretically- be possible to give the human being that was under test, the impression of this sound occurring at the exact same position, when listening through headphones, that is.

2.7 – Torso

It appears that actually the entire human body, with the body parts closest to the ears being most importantly, play a role in localizing sounds. A humans upper-body, particularly the shoulders, act as an obstacle and/or refractor for sound. There hasn't been too much research on this issue, but it is most likely that, although it certainly does play a roll, it isn't a very important one.

2.8 – Cone of confusion

There are several situations in which the earbrain can get confused, is inadequate or can be tricked. A few –the most important- will be mentioned here.

One is the so called "cone of confusion". When a sound source is located at either the exact left or right side of the head, the source is hardly localizable. This because the sound gets "blocked" by the head, it doesn't, or hardly does, reach the ear at the other side of the head. The sound that possibly does arrive, is typically so attenuated, that it is disregarded by the brain anyway.

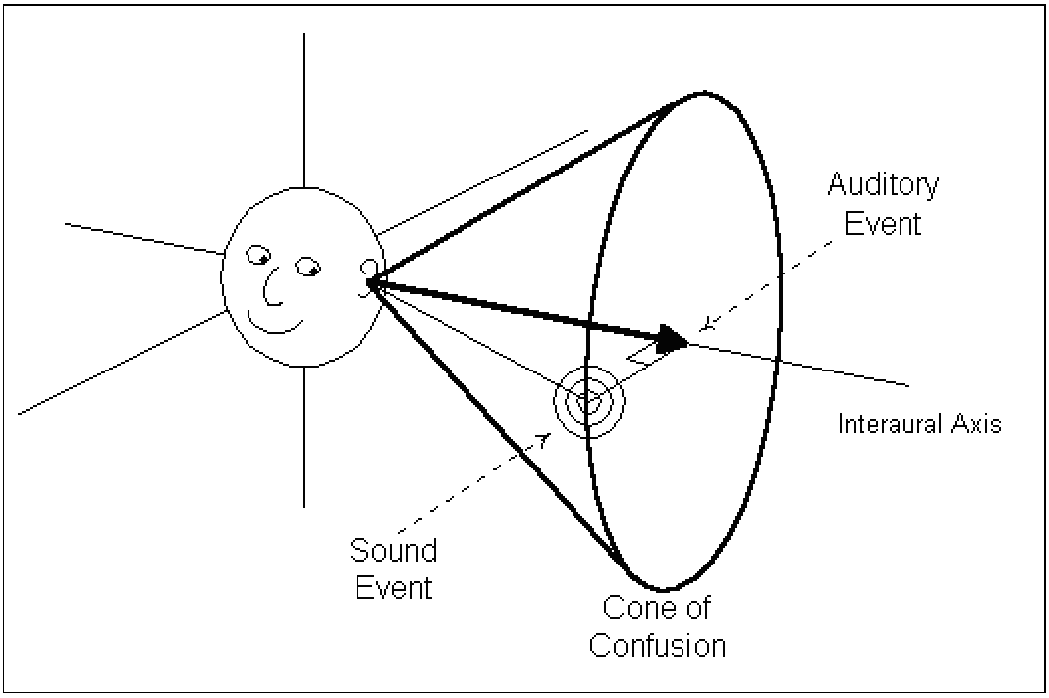


Figure 7 - Cone of confusion

The only way to localize a source located at the side, is by head movement, and so, in effect, still isn't "localizing sound at one side of the head". This because intrinsically, by moving the head, the source isn't located there anymore, and the data collected by the ear on the opposed side comes into play again.

2.9 – Haas or precedence effect

The earbrain bases the determination of the location of a sound source in the very first part of the sound event. It already has been explained that the ear is best in localizing short, short transient sounds. In fact the system is entirely based on this. A sound event typically starts with a rather short attack, sounds that do not, are hard to localize.

After the brain has been “triggered” by such an event, it becomes “immune” to any other trigger for the next 10 to 30 milliseconds. Any sound occurring in this period of time, gets “fused into” the original event. So if a sound event occurs on the right side of the head, and another, slightly later, on the left side, the brain makes it like there is only an event on the right side of the head, the one on the left gets disregarded.

3 – Recording and playback of sound

3.1 – basic stereo microphone techniques

The ways of recording sound that are described in this chapter, are typically meant for playback over a typical 2-speaker stereo system. Because this is the way sound has been recorded and recreated for many, many years, a great deal is known, and therefore makes a good study.

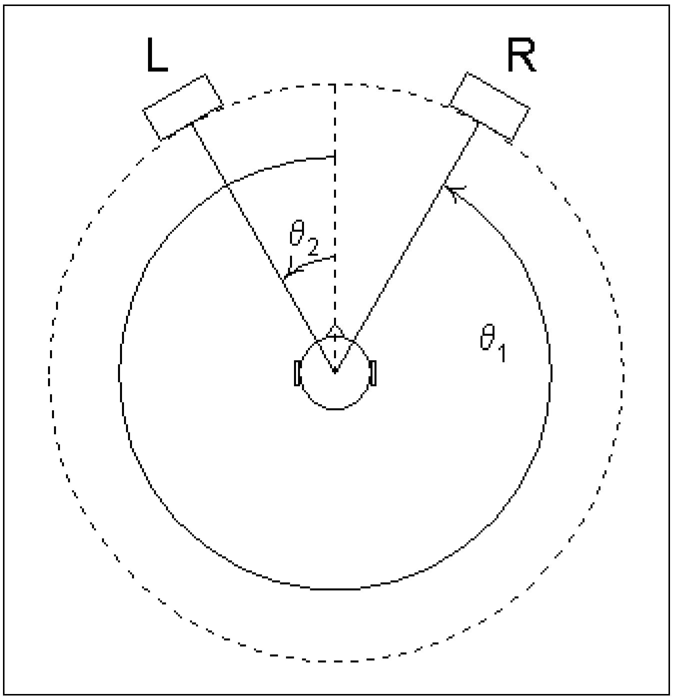


Figure 8 - typical 2-speaker stereo setup

In this setup, the two speakers are setup equidistant from both the listener and each other. The angles of this triangle are 60 degrees each, both speakers being plus and minus 30 degrees off-center.

3.1.1 – XY

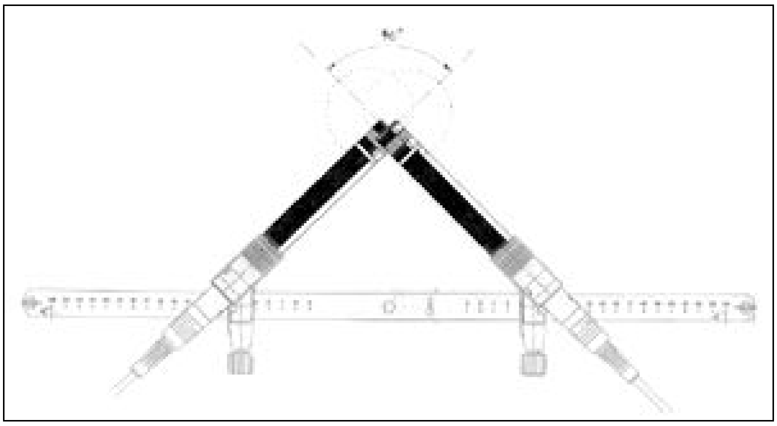


Figure 9 - XY-microphone setup (picture courtesy of DPA)

This setup is a very basic setup, hardly ever used in classical recordings, rather often in pop music. The two microphones are cardioid, and setup with an angle of 90 degrees. This technique only provides only for amplitude-stereo (ILD), since there cannot be any time delay-stereo (ITD), because the microphones are at the same spot.

It provides a signal that is for 50% mono, because of it's inherent coincident nature. The two cardioid microphones aren't separated in any way, and therefore they'll pick up a lot of the same sound.

Although giving a very clear stereo-image because of the good separation in the level between the two microphones, it is part of the microphone's nature to "color" the sound in a rather unwanted way.

3.1.2 – AB

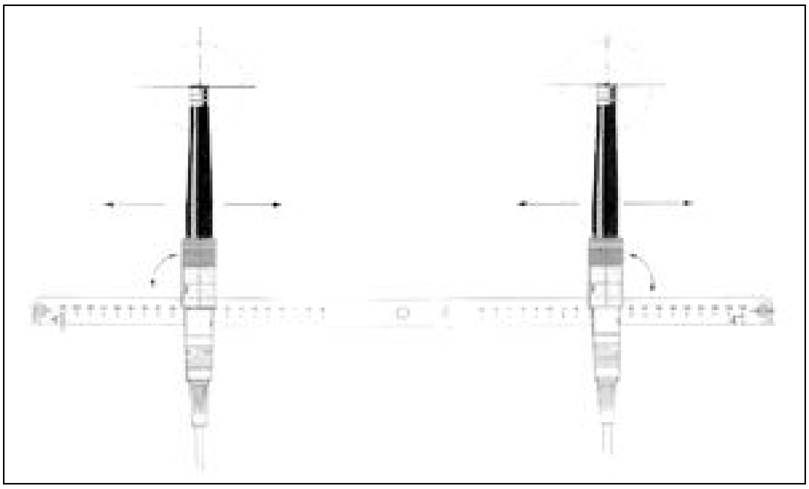


Figure 10 - AB-microphone setup (picture courtesy of DPA)

This microphone technique records both amplitude-stereo and time delay-stereo, making it a physically and mathematically more correct way of recording. It can be argued that for a sufficient amplitude difference the distance has to be too great in order for the time-difference to be adequate. This is because of the omnidirectional nature of the microphones: they pick up a lot of the same sound coming from any direction.

The level difference can be enhanced by placing a baffle (officially known as "Jecklyn-disk) in between the two microphones. This provides also for a kind of "shadowing" effect, somewhat similar as in natural human hearing.

3.1.3 – ORTF

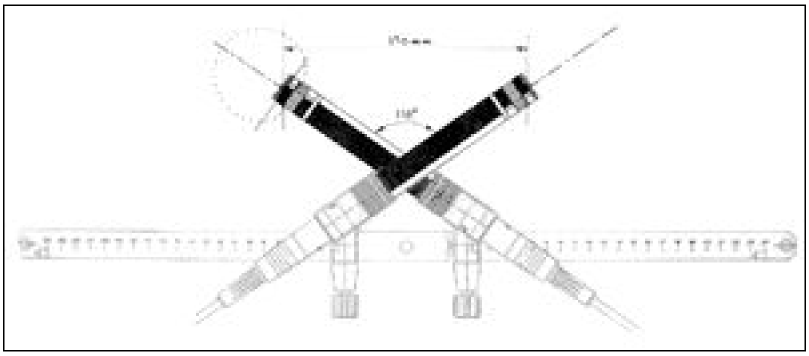


Figure 11 – ORTF-microphone setup (picture courtesy of DPA)

In theory, this technique should provide the “best of both worlds” when assumed the two worlds are XY and AB. Both amplitude and time delay are recorded, and since the distance is usually at about 17,5 centimeters, a recording close to natural human hearing can be provided for. The microphones are angled at about 110 degrees.

3.1.4 – MS

In this method, a cardioid microphone is pointed straightforward, and a figure-of-8 microphone is pointed sideways. The cardioid should provide for a stable center-image, panned equally to both speakers. The figure-of-8 microphone-signal is splitted into two signals, of which one is phase-flipped. One is sent to the left speaker, the other to the right, according to which side of the soundfield they’re supposed to pick up.

3.1.5 – Blumlein

This technique, though hardly being used anymore because of its spaciousness, by most people considered as being “too much”. Two figure-of-8 microphones are being used, angled just like in the XY-setup as explained before. Because of its nature, the signal at the rear of the microphone is “mixed in” equally loud on the opposite front-side.

3.1.6 – DECCA tree

This technique can be seen as an expansion of the commonly known AB-technique, explained earlier. The two AB-microphones are panned to either side of the stereo-setup. An additional “center-microphone”, typically located right in-between, somewhat in front of the AB-pair, can be mixed in equally in both speakers. The amount of which according to taste.

3.1.7 – OHNO (double AB)

This system is known for its richness in depth and accurate localization. It is a combination of two AB-pairs. One being spaced rather close, the other being spaced rather wide. Although originating only from practical experimentation, by someone bearing little or no theoretical, but al the more practical experience, theoretically it makes perfect sense. Only recently it was discovered that Mr. Blumlein already invented such a setup in the 1920's.

The "small AB" would take care of recording the correct ITD and provide for a rather stable center, the "wide AB" would take care of the proper ILD. When additional shelving-equalizing is applied, the picture could be cleared up and made even more consistent to human hearing (see chapter about human hearing)

3.2 – LCR-microphone techniques

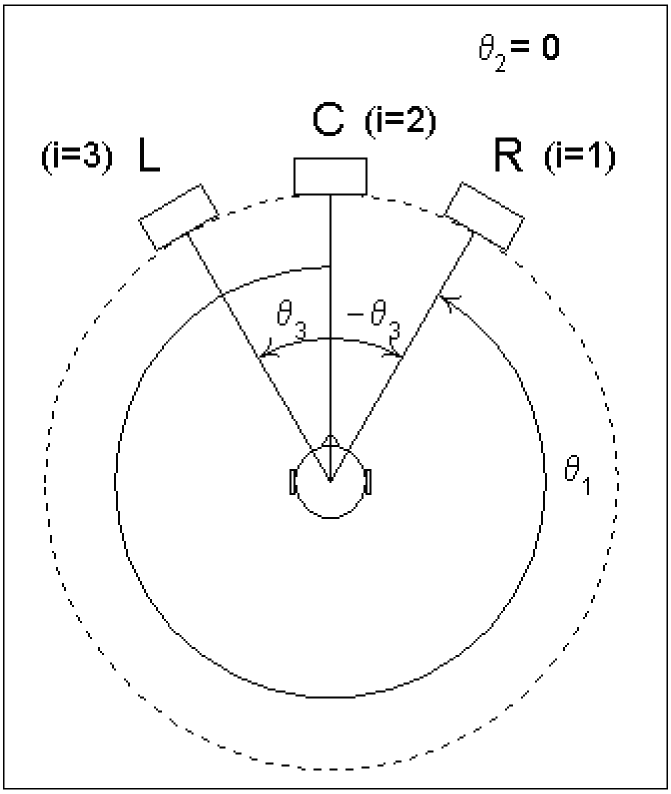


Figure 12 - typical LCR speaker setup

The LCR speaker setup consistes of the regular stereo-pair, situated plus and minus 30 degrees from the centerline, with an aditional centrerspeaker, right on the centerline.

3.2.1 – Stereo+C

This is a simple expansion of the AB-system. An additional omni directional microphone is added to the AB-system, several feet higher. This should provide for enough channel separation without sounding “too close”.

3.2.2 – OCT

This microphone-setup makes extensive use of the profound characteristics of a certain microphone manufactured by Schoeps, the MK41. The microphone is hyper-cardioid, but still has a very good off-axis frequency-response, something very unique for this type of microphone.

Two microphones are angled back-to-back, facing the front sideways. This gives an excellent channel separation. The distance between the two microphones can be altered in order to provide for correct ITD.

A similar microphone faced forward, right in-between and somewhat in front provides the feed for the center channel. OCT stands for “Optimized cardioid Triangle”.

3.2.3 – INA5

This microphone-setup is especially designed for playback through a 3/2 speaker setup, including the surround speakers. 5 Microphones are employed in exactly the same positions as where the speakers are supposed to be. Only the radius, hence the distance between the microphones, is smaller. The microphones of use are cardioids. Their faced “outward”, in effect exactly the opposite direction the speakers are facing.

3.2.4 – Fukada tree

This is an expansion of the DECCA tree, which has been described earlier. The three omni-microphones in front are supplemented with two widely-spaced omni-microphones some distance behind the front-system.

3.2.5 – Soundfield microphone

This very special microphone actually contains “4 microphones-in-one”. In theory this microphone records “all” necessary data of a soundfield in the position where it’s located. The converter that comes with the microphone can provide for 4 channels of audio signals. The first three contain information about the three mathematical axis (x, y and z), the fourth contains “w”, information about the sound pressure at the location. This technique only makes use of level-differences, because of its inherent coincidental nature.

Optimally, this information should be decoded into a symmetrical speaker-array, the more speakers, the better. This system, part of what is called “ambisonics”, can be adapted, although not optimally because of its non-symmetrical layout, to a 3/2 speaker-setup. People are still working on a more optimal decoding-schema, with so far the best result being the “Vienna-decoder”.

3.2.6 – Schoeps-sphere

Derived from the principle of the dummy-head, Schoeps developed a less personalized sphere microphone. The idea is still that the recording of the soundfield is as close to how humans perceive sound, yet generalizations are being made in order to be more compatible. Like with the soundfield microphone, a special decoding scheme is used to provide for the proper speaker-feeds.

3.3 – rear-channel microphone techniques

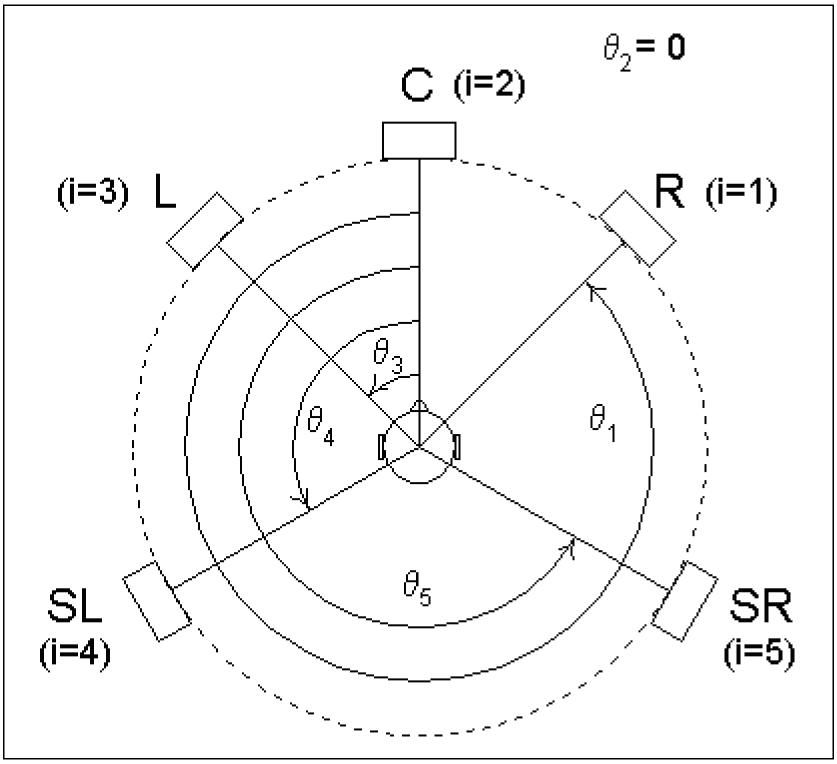


Figure 13 - 3/2 speaker setup

3.3.1 – OCT

In addition to the LCR-array, described before, two hyper cardioids can be applied facing backwards, to give feeds for the rear speakers. They typically are positioned near the left and right microphones.

3.3.2 – INA5

As already described, the INA5-array employs two, rather widely-spaced, cardioid microphones facing the rear.

3.3.3 – Spaced omni directional microphones

As used with the Fukada tree, two rather widely spaced omni-microphones can be used in any situation. It can be argued what kind of signal is actually desired in the rear speakers, but it is to the author's believe that a strongly decorrelated, diffuse, indirect and pleasingly sounding signal is optimal.

As long as the spacing between the two microphones is adequate, this setup can deliver exactly that.

3.3.4 – Spaced cardioid microphones

Like with the spaced omnis, this setup can come pretty close to the same results. Care has to be taken not to pick up any unwanted direct signal, as well as any strong reflections coming from any hard surface at the sides or the rear of the venue. Typically the (frequency-) response of this type of microphones isn't considered to be as "pleasingly sounding" and "correct" as omni directional microphones. A trade-off has to be considered with respect to achievable channel separation (hence decorrelation) and frequency-response.

It can be very usefull in a live-recording situation, where noise coming from the audience can be somewhat attenmuated from the rest of the signal, by pointing the microphones cleverly.

3.3.5 – Soundfield microphone

As described, the soundfield microphone is supposed to record "all" data about a certain soundfield, hence a feed for the rear speakers should be achievable. So far, the author hasn't heard any convincing results with that, assumingly partly due to incorrectness of the decoding scheme. It is also arguable that it is impossible to achieve sufficient channel separation and diffuseness due to the coincident nature of this setup.

3.3.6 – Namasaki-square / NHK-square

A rather original way to record signals for the rear speakers is the so called Hamasaki-square. In this, a square of 4 figure-of-8 microphones is employed, with a variable spacing, but typically a rather wide one. Th routing of the outputs of these four microphones is somewhat unusual: the two rear ones go straight to the rear speakers, the two in front go to both rear speakers and front-left and front-right speakers. Theoretically this should provide for a more "enveloping" reproduction, although it can be argued, partly due to the "cone of confusion" that this is impossible with a 3/2 speaker setup.

4 – Why more research is needed

As can be concluded from the foregoing, there are many, many factors that play a role in making a good and accurate sound recording. It is to the authors believe that, probably even with a playback system much more expanded than the 3/2 speaker setup, there will always be a lot of compromises.

In the most ideal situation, all aspects that apply to recording and recreating sound have to be true to nature. All cues for localization have to be present and non-contradictive. The recreation of the frequency-spectrum has to be exaclty similar to the original, just as with any other factor one could think of.

One part of the story is the human hearing-mechanism, the “earbrain”. It is highly sensitive to certain aspects of sounds, yet sometimes almost completely insensitive to others. Also, the earbrain can easily be confused or fooled with respect to certain aspects. It is very important to know these factors when attempting to record and recreate sound. This in order to prevent certain faults from happening or use these aspects to ones benefit.

Another part of the story is a rather technical one. In order to record a soundfield, a certain event a certain space, room or environment, as faithfully as possible, one would, theoretically, need thousands, if not millions of “perfect” microphones. In order to recreate the same soundfield, one would need some thousands, if not millions of speakers. This already has been hypothesized in the early 1920’s, when experiments for broadcasting sound were conducted at Bell Labs.

This approach has rather recently been adopted by the Technical University of Delft, the Netherlands, though up until recently focusing only on the recreational-side of the chain. A couple of hundreds of speakers (around 300) are possitioned in a large square, or rectangle, around the listener. All these speakers together reproduce the soundfield as calculated by a very powerfull computer. Although theoretically very well supported and understood, in practice this is a very costly and impractical approach that, in this form, never will be commercially viable.

An approach aiming for the same goal, but approaching matters from exactly the opposite side, is the aforementioned “Ambisonics”. Trying to be as flexible as possible, therefore being commercially more viable, it is also tried to reconstruct a certain wavefield as faithfully as possible.

In Ambisonics, sound is recorded carrying specific information about its position in space. In sound-synthesis this can be done simply by filling out the spatial coordinates. In recording this can be seen as the specific loudness of a certain sound in each of the dedicated tracks of the recorder, attached to the axis’ x, y, z and w.

Upon playback, the decoder is given the exact location of the speakers, and it calculates how loud each sound should be played back through every speaker, in order to position the sound at the exact location of its origin. It can be seen by this that more speakers would automatically result in a better performance.

Because this 4-track-ambisonics is only theoretically correct, and holds quite a bit of drawbacks and flaws in practice, an improved version, using 9 tracks has been developped: “2nd order ambisonics”. By storing more information of the original, a better picture can be drawn in recreation. But this is still in its rather early stages, and a lot more research has to be done in order for things to live up to their capabilities and expectations.

A lot is known about recording sound for stereo (2-speaker), since it has been practiced for so many years. In the early days of recording, people where considered with recording the complete soundfield, and recreating it as accurately as possible. Due to technical limits, many ideas and theories had to be compromised, and using only two speakers seemed to be adequate.

After settling for the 2-speaker setup, research and progression seemed to have dozed of, since nothing changed for about 50 years. People practicing recording and recreation of sound seemed to have forgotten what it exactly was they where after. They regarded the two speaker-setup as a given, neglecting al the artifacts and discrepancies it posed. They simply wanted to “deliver the message”, not taking care of true physical or mathematical correctness.

Now, with the rising of “cinema-sound” and “home-theatre” in peoples living rooms, a lot of possibilities are opened to improve upon the recreation of sound. Was it already known to many insiders that more speakers evidently are probably always better than less, this expansion of the playback mechanism should be used in order to prove this to anyone.

With the 3/2 speaker-setup, it is possible to improve upon the physical and mathematical correctness of the recreation of sound. Even though it is not perfect by any means, for the moment this is the system that has the best chances in making it as a standard, if it isn’t one already.

So know the bottleneck is at the recording-side, if it is assumed that the 3/2 setup is the one to go with. Developments in recording sound have been restricted to the old-fashioned 2-speaker setup for nearly 50 years, so little is commonly known about recording for a “more-than-2-speaker-setup”.

Only the very few people working on ambisonics since the late 1970’s have gained quite a bit of knowledge about this matter. At the moment ambisonics is having a revival, so hopefully a lot of people will take knowledge about the system and learn whatever can be learned from it.

It can be argued that, also in the 1970’s, quadrophonics also had to do with improving upon the 2-speaker-setup. Yet this format (if it ever were a “fixed standard” in the first place) has little or nothing to do with recreating true physically and mathematically correct sound and soundfields. It was mainly used for pop music, and a lot of “unnatural ping pong-stereo” was going on.

It is to the authors believe that anyone trying to record sound in the most natural way, should be aware of all the physics that applies. This would mean the physics of sound, the physics of playback-systems, the physics of the human hearing mechanism, etc.

Only when fully comprehending all of these systems (if possible at all; to as large an extent as possible, anyway) the most physically and mathematically correct recording can be made. Also, with this knowledge, any problems that may arise during recording, can be dealt with in the most appropriate way.

Coming to the conclusion of why more research is so much needed in the field of “recording for 3/2”, several tendencies can be seen in all foregoing:

-The 3/2 speaker setup is at this point in time the most viable and most likely to succeed approach in recording and recreating sound.

-People have been loosing focus of what exactly they’re after when recording sound for nearly 50 years due to the technical restrictions of the 2 speaker stereo setup.

-Not to much is commonly known about how the human hearing system works and how sound can be recorded. It is to the author’s belief that the only way to become “perfect” in recording and recreating sound, is to know “everything possibly” about all the factors that do or may play a role.

5 – Resources

A listing of only just some of the resources that were used. In the final version of the thesis, a complete list will be provided for.

[1] - Spatial Hearing
The Psychophysics of Human Sound Localisation
Jens Blauert
MIT Press Cambridge Mass., 1996
ISBN 0-262-02413-6

[2] - Spatial Audio
Francis Rumsey
Focal Press, Boston, 2001
ISBN 0-240-51623-0

[3] - The Proceedings of the AES 19th international conference;
Surround Sound, techniques, technology and perception
Chair: Günther Theile
AES New York, 2001
ISBN 0-937803-43-X

ADDENDUM 1 - Microphone spectrum and impulse response test

Every microphones sounds different. Not only every different type or brand sounds different, also every single microphone of the same brand, same type, and so on. The latter differences are minimal, and usually can be neglected.

Differences between types of microphones (omni, cardioid, figure-of-8, etc.) and brands are all the more noticeable and interesting. High-end microphones can be compared without simply listening, by the graphs provided for by the manufacturer. Problem is that they not always are as honest as one would like, and also test-circumstances can, and probably are influenced, at least they're not exactly the same for every brand.

An omni directional typically has a "flatter" frequency-response than any other type, and is more sensitive. Of all the aspects responsible for how a microphone sounds, some can be measured more easily than others. For this test two of those are used.

The frequency-response of a microphone can be measured simply be recording some pink (or white) noise with the microphone, put in an analyzer and see how flat the response is.

The response-time of a microphone can be measured by applying a known impulse, typically an as-short-as-possible "click", and measuring the response time and distortion of the impulse.

In the anechoic room a test setup was built: a KRK V8 speaker was set up to produce pink noise and an impulse click, with a level of 80 dB(A) at 3 meters distance, up to 22.050 kHz. The microphone was connected to a pre-amp, so that the level of its output was exactly 0 dB. This then was recorded into a computer, where a frequency analyzer was used to draw the graphs and produce the useful numbers.

There is no doubt the speaker, converter, console and cables colored the sound heavily, but for the purpose of this test this wasn't a problem at all. All microphones were measured with exactly the same setup, so for comparison this test is completely valid. Also, most tested microphones are capable of taking care of frequencies much higher than 20 kHz., but restrictions of the AD and DA converters make the use of a brickwall filter at 20 kHz. necessary (hence the roll-off starting at about 15 kHz.)

The following microphones where tested:

2 Schoeps MK2H – omni directional, small diaphragm condenser microphone
2 Schoeps MK4 – cardioid, small diaphragm condenser microphone
1 Stagg OCM 7B – switch. omni/cardioid large diaphragm condenser microphone, switch. low-cut filter
1 Sennheiser MD 421 – cardioid dynamic microphone, switchable low-cut filter
1 Stage Line ECM-20 - omni directional, small diaphragm electret microphone
1 dB-meter (Radioshack)

Examples of what the graphs look like:

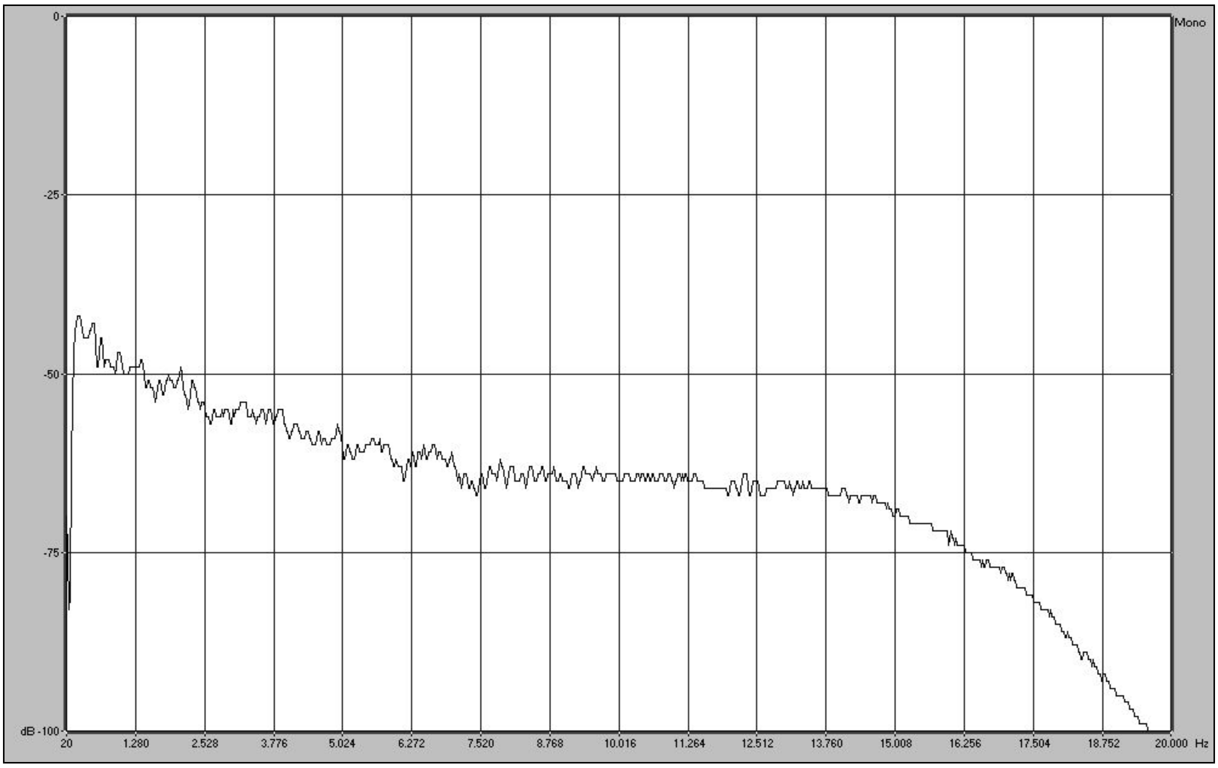


Figure 14 - Frequency spectrum of a Schoeps MK2H microphone (omni directional)

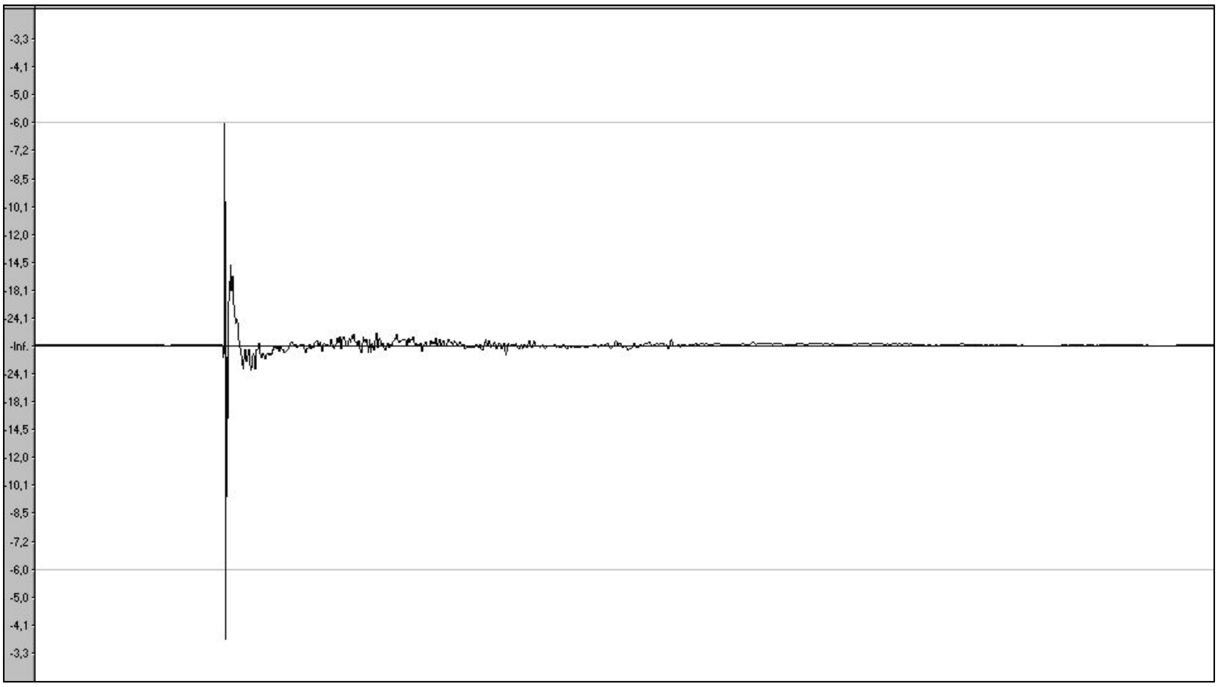


Figure 15 - Impulse Response of a Schoeps MK2H microphone

This test can be very useful when comparing microphones and in practical work. The impulse response tells you something about the sensitivity of the microphone. By knowing the sensitivity of a specific microphone or microphone-type a prediction can be made on how the sound, when applied in a specific situation, or for a specific instrument, will be.

Knowledge about the spectrum can be used in exactly the same manner, and in another, very useful one. It is possible to, by means of an equalizer, adjust the sound of a microphone to a more flat response without any listening (thus canceling any possibility of misjudgment because of listening room or monitoring offsets). Also, it can be tried to make a certain microphone sound more like another one, in cases where it is impossible or unwise to use this other microphone, or when there’s a simple lack of one.

ADDENDUM 2 - Phantom-image coloring

For some applications undoubtedly the biggest advantage of “surround” opposed to ‘stereo”, is the addition of the center speaker. When evolving from mono to stereo, one of the first propositions was to complement the (centered) mono speaker with two speakers, located at about 30 degrees on either side. This would widen the image and make more “room” for the music.

Technical restrictions made it easier and more logic to only store 2 speaker feeds, also because “3” is a somewhat awkward umber, “2” feels more “correct”, in a sense. Hence the center speaker was dropped, and it seemed perfectly possible to do without.

When attempting to place a certain source in the center, when mixing to “regular” stereo, the sound is fed equally into both the left and right speaker. This creates a “phantom image” in the center. Yet there are some problems and drawbacks to this. One is that this trick only really works when the listener is located exactly in the center. When located off center, the image gets drawn into the nearest speaker.

Another drawback is that the sound gets colored rather heavily, because of interference and interaction between the two speakers. Both speakers create sound, and the waves interact when “colliding”, making certain frequencies to gain power, and others to cancel out, according to their wavelength and phase. This kind of comb filtering can be measured in an anechoic room, which is what the author did. Some of the results are presented here.

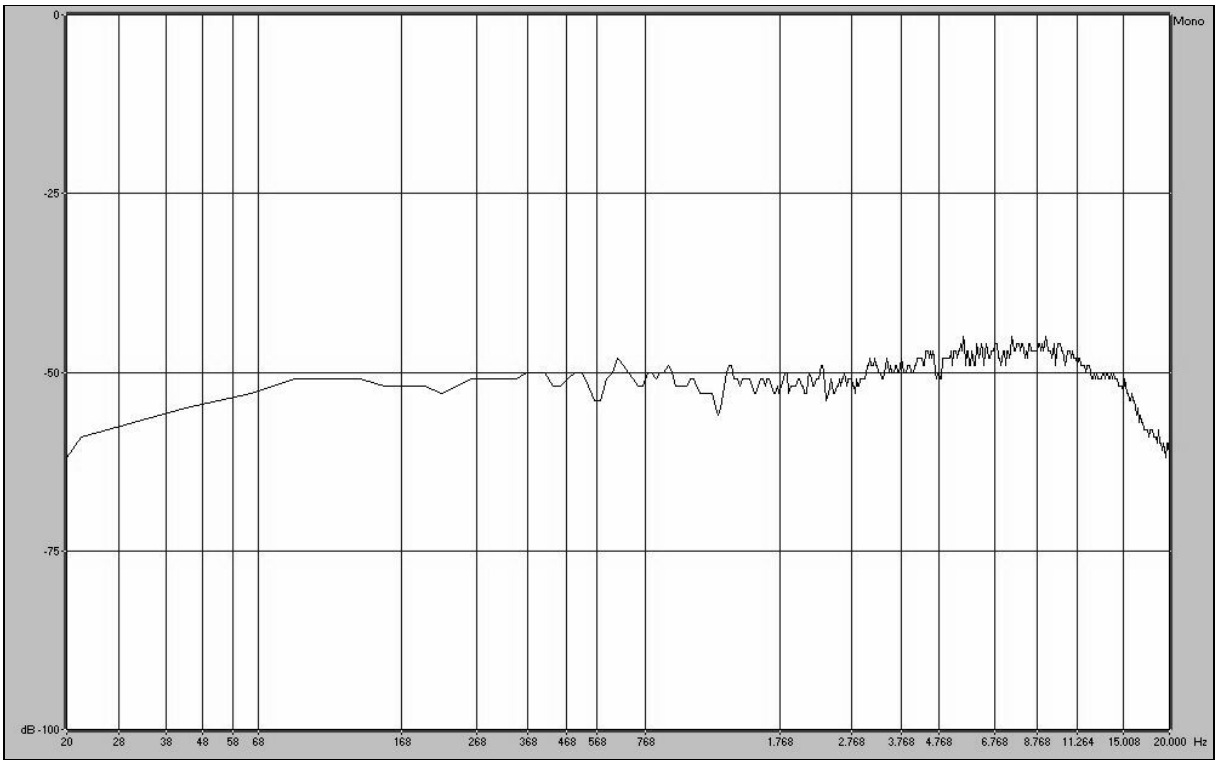


Figure 16 - response from one speaker playing back pink noise

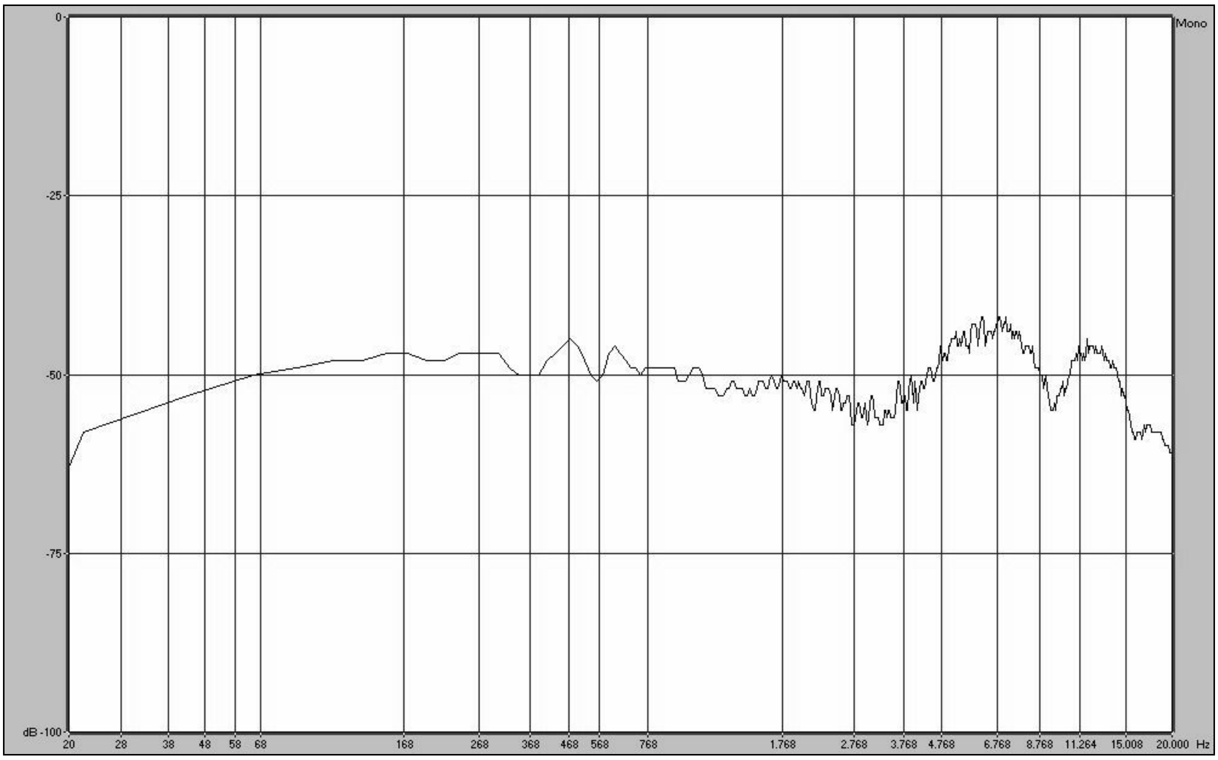


Figure 17 - response from two speakers (plus and minus 30 degrees) playing back pink noise simultaneously

It can be seen that when figure 1 is taken to be “flat”. Just like in addendum one, several circumstances distort the accuracy of the measurement, yet this isn’t a problem for the purposes of this measurement, as long as their taken into account. What’s interesting is the difference between the two graphs, it can be seen that the coloring of the two speakers interfering is rather heavy.

ADDENDUM 3 - Side-localization in a 3/2 setup

Not too much really scientific can be said here about these experiments, since they mostly existed of “playing around”. In the chapter about “Human hearing”, it is mentioned that human beings have a hard time localizing sounds coming from either side of their heads. This is the so-called “cone of confusion”. Only head movements would result in some clues about the exact location of a sound source in that direction.

Now with the 3/2 speaker setup, it turns out to be almost completely impossible to place a sound source at either side of the listener in the first place. The typical “phantom-centering-trick”, when putting a signal in, say, both left-front speaker and rear-left speaker doesn’t work here, because the listener can only rely in one ear for localization. It turns out the earbrain simply picks one of the two speakers, and labels the sound source as coming from that speaker’s location.

When turning the head, off course, the source will be located “at the side”, but also very unstable, because of the rather large and uneven angle between the two speakers.

Moving sounds can give somewhat better results, using psycho acoustics and some imagination. When a sound is panned from the left to the right speaker for instance, it can be rather logical to assume the sound keeps on moving toward the left-rear speaker. And having an image-projection of film confirming this movement, this illusion can become even stronger.

Very recent anechoic tests in panning for surround (in cooperation with Eelco Grimm) showed some very interesting results in this matter. The author will only suffice in stating some brief temporary results, also since this is actually part of a different research-project.

When using simple and common intensity panning (ILD) between the front-left and rear-left speaker, no real movement is noticed, but rather a “fading out” of the front speaker, and a “fading in” of the rear speaker. But when using far less common and somewhat more complicated “time-delay-panning” (ITD), some rather astonishing and way more convincing results can be accomplished.

There are still quite some flaws in the system as was used, and quite a bit unexplained and/or unexplored, but the expectations are rather high…

ADDENDUM 4 – Attempting to recreate the acoustics of a room in 3/2

When a sound source produces sound waves to travel in a room, most of them bounce several times off walls, the floor and the ceiling, before arriving at a listener or microphone, for that matter. This, and exactly how this “bouncing” occurs, is what gives us a notion of “the acoustics” of a room.

In a typical recording or listening situation, one that can be regarded as “good”, or “pleasing”, first the direct sound arrives at the listener of microphone. Then reflections that only get bounced once or a rather few times arrive, they’re called “early reflections”. Then the reflections that are bounced numerous times arrive, they’re typically called “reverb”. It has to be noted that some (e.g. David Griesinger), make an additional separation, early reflections get divided into “early early reflections” and “late early reflections”, but for the purpose addressed here, this separation is unnecessary to make.

Early reflections mainly give us information about the size and shape of a room and provide additional cues for the location of the source. Reverb mainly gives us information about the material of which the room is built.

When trying to recreate a certain sound-event in a certain room, the following setup was made, with rather promising and pleasing results. An anechoically recorded source (in this case female voice, acoustic guitar, violin and viola were tested) was placed in the center speaker. Early reflections were created in the left and right speaker. Reverb was created in the rear-speakers. For this a computer running the Waves TrueVerb-plugin was mainly used.

To make things even more accurate, the reverb can be put in actually every speaker, since those reflections tend to come from “everywhere”. It was found that putting it into the center speaker wasn’t really necessary, and when put there too loud, was capable of obscuring the dry signal.

The early reflections coming from the left and right speaker made the source sound much more “natural sounding” without a sense of hearing echoes or reverb. The source seemed actually even easier to localize. Additional early reflections can be put in the rear speakers, yet care has to be taken that they’re different ones than the ones in front, and also a somewhat softer level is advised. Putting early reflections in the center speaker would be somewhat unnatural and could obscure the image.

ADDENDUM 5 - Anechoic recordings

For the purpose of having some material to test some theories on, to get some hands-on experience and to find out how several instruments “work”, some instruments were recorded in the anechoic room. A side-result was a test-CD, of which the booklet is printed here.

ACOUSTIC GUITAR

This recording is of Harry Saksioni’s “Elixer” (a rather early work of his). Many recordings where made. The first version is a (slightly) edited version with the microphone placed near the sound hole (*fig. 1, number 1*). The next versions are none-edited version with different microphone-placements, in order to get an idea of the frequency-diffraction of the instrument.

The guitar being used was built by Otto Vowinkel in 1990. The kind of wood used for the top was spruce (in Dutch: fichte), as Indian rosewood (Dutch: palissander) was used for the back and rear sides. The strings being used were: “d'Addario Pro Arté hard tension, number EJ46, silver plated wound - clear nylon”. Figure 1. shows the different microphone-locations.

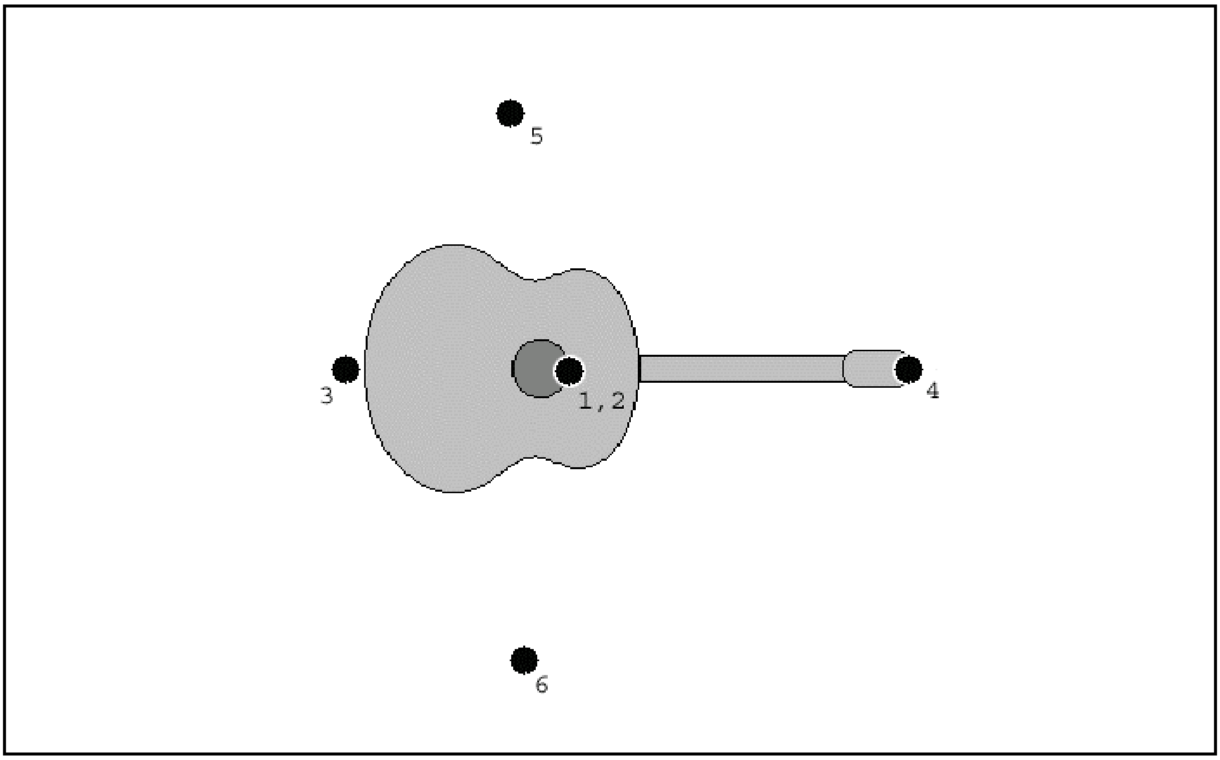


Figure 18

- “1”: track 9 on CD, file: “center” (appr. 20 centimeters from guitar)
- “2”: track 10 on CD, file: “far away” (appr. 2 meters from guitar)
- “3”: track 11 on CD, file: “body”
- “4”: track 12 on CD, file: “headstock”
- “5”: track 13 on CD, file: “high” (right in front of the player’s eyes)
- “6”: track 14 on CD, file: “low” (“shinbone-height)

VIOLIN

The recording is of “Cadenza” for Brahms’ Violin Concerto, transcribed by Joseph Joachim. More than several recordings where made, and the version on the CD is an (heavily) edited version of those. The violin being used was built by Platner, in 1742. The microphone was placed overhead, at a distance somewhere between 50 and 100 centimeters from the violin.

VOCALS

Several different parts and pieces were performed approximately 20 cm. in front of the microphone. The first is “My Funny Valentine”, a jazz-standard. The next is an excerpt of Tracy Chapman’s “Last Night”, a piece meant to be a capella in the first place. The third is a scat-improvisation inspired by the solo of “Not For Me”, another jazz-standard. The last one, called “Maserati”, is an excerpt of a comedy/theatre-performance, in Dutch. Spoken word evolves into singing. The singing is originally meant to be accompanied by piano.

VIOLA

This recording wasn’t part of the plans originally, but since there was time, an instrument and somebody willing to play it, chances were taken. The viola isn’t of top quality. The part being played is “Berceuse”, composed by Gabriel Fauré, transcribed for viola by René Pollain.

EQUIPMENT

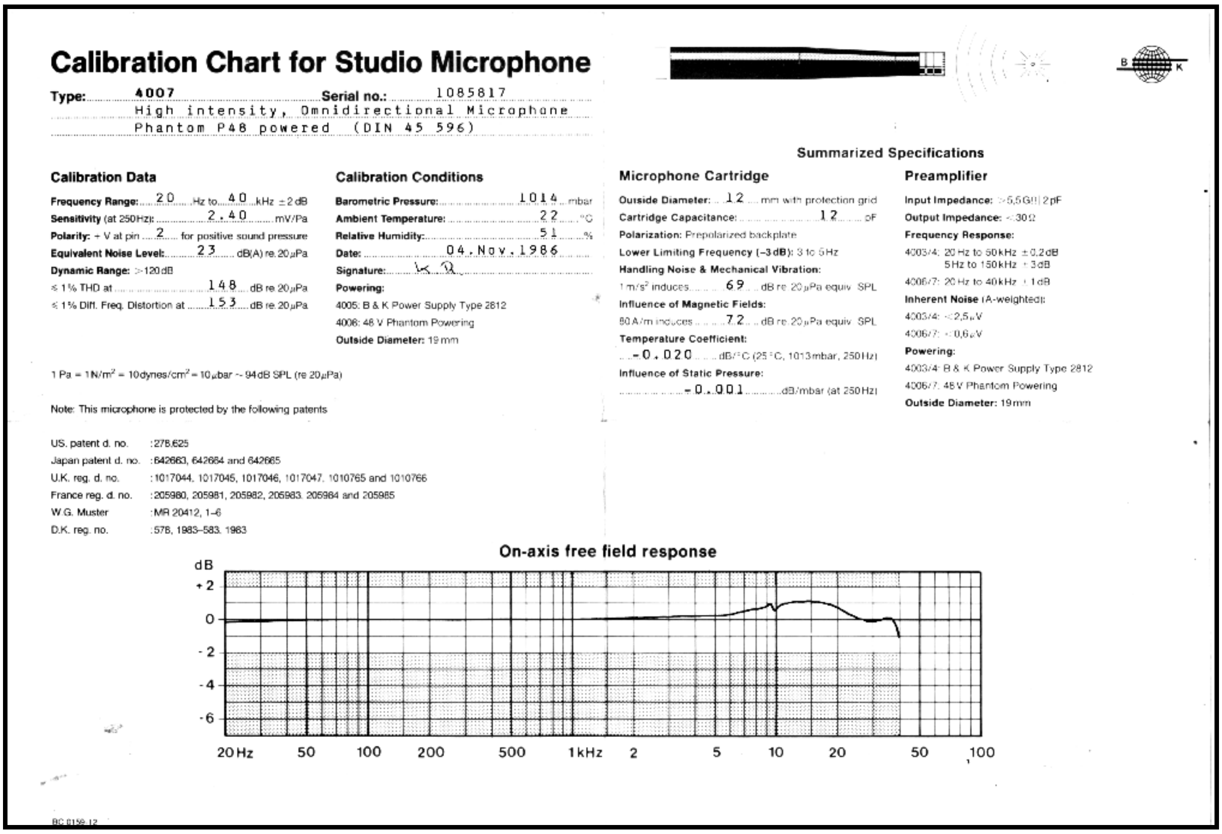
In none of the recordings has been made use of equalization, compression or whatever sound-altering processes one could think of. The set-up was as straightforward as possible, the signal flow as short as possible. The recordings supposedly are as pure and authentic as possible.

signal flow

B&K 4007 microphone
↓
12 meters of Tasker cable with Neutrik XLR connectors
↓
MAO MAMP1 pre-amplifier with MAO MPSU1 External power supply
↓
MusicNet AD24 analogue to digital (ADAT) converter
↓
StudI/O ADAT PCI computer interface
↓
Macintosh G4/ OS 9.1/ Logic Audio Platinum 4.5.1 (24 bits/48 kHz)

MICROPHONE

The microphone that was used, was a Brüel and Kjær type 4007. This is a (very) small-membrane microphone. It’s characteristic is omni-directional, it has a very steady frequency-response (20 Hz-40kHz ± 2 dB) and is capable of dealing with very loud signals. The trade-off for the latter two is a rather high noise-floor.



Calibration Chart for Studio Microphone

Type: 4007 Serial no.: 1085817
High intensity, Omnidirectional Microphone
Phantom P48 powered (DIN 45 596)

Calibration Data

Frequency Range: 20 Hz to 40 kHz ±2 dB
Sensitivity (at 250Hz): 2.40 mV/Pa
Polarity: + V at pin 2 for positive sound pressure
Equivalent Noise Level: 23 dB(A) re. 20µPa
Dynamic Range: >120 dB
≤ 1% THD at 148 dB re 20µPa
≤ 1% Diff. Freq. Distortion at 153 dB re 20µPa

Calibration Conditions

Barometric Pressure: 1014 mbar
Ambient Temperature: 22 °C
Relative Humidity: 51 %
Date: 04.Nov.1986
Signature: K. R.
Powering:
4005: B & K Power Supply Type 2812
4006: 48 V Phantom Powering
Outside Diameter: 19 mm

1 Pa = 1N/m² = 10dynes/cm² = 10µbar ~ 94dB SPL (re 20µPa)

Note: This microphone is protected by the following patents

US. patent d. no. :278.625
Japan patent d. no. :642663, 642664 and 642665
U.K. reg. d. no. :1017044, 1017045, 1017046, 1017047, 1010765 and 1010766
France reg. d. no. :205980, 205981, 205982, 205983, 205984 and 205985
W.G. Muster :MR 20412, 1–6
D.K. reg. no. :578, 1983–583, 1983

Summarized Specifications

Microphone Cartridge

Outside Diameter: 12 mm with protection grid
Cartridge Capacitance: 12 pF
Polarization: Prepolarized backplate
Lower Limiting Frequency (–3 dB): 3 to 5 Hz
Handling Noise & Mechanical Vibration:
1 m/s² induces 69 dB re 20µPa equiv. SPL
Influence of Magnetic Fields:
80 A/m induces 72 dB re 20µPa equiv. SPL
Temperature Coefficient:
–0.020 dB/°C (25 °C, 1013mbar, 250 Hz)
Influence of Static Pressure:
–0.001 dB/mbar (at 250 Hz)

Preamplifier

Input Impedance: >5,5 GΩ || 2 pF
Output Impedance: <30 Ω
Frequency Response:
4003/4: 20 Hz to 50 kHz ±0,2 dB
5 Hz to 150 kHz ±3 dB
4006/7: 20 Hz to 40 kHz ±1 dB
Inherent Noise (A-weighted):
4003/4: <2,5 µV
4006/7: <0,6 µV
Powering:
4003/4: B & K Power Supply Type 2812
4006/7: 48 V Phantom Powering
Outside Diameter: 19 mm

Figure 19 - chart provided by manufacturerer

PRE AMPLIFIER

The pre-amp was a specially built MAO “MAMP1”. This pre-amplifier was ordered by the NOB, the Dutch Broadcasting Company and designed according to their specifications. Its characteristics are extremely “puristic”: very low noisefloor and flat frequency-response.

CONVERTER

The analogue to digital converter was a “MusicNet AD24”. It converted the signal into a 24 bit, 48 kHz ADAT signal.

COMPUTER (SOFTWARE)

The “StudI/O” dual-ADAT interface interfaced between the converter and Emagic Logic Audio Platinum, version 4.5.1.

ADDENDUM 6 - Sweet spot versus sweet area

Basically, there are two schools of thought in “audio world”. One group is after the most perfect audio recreation for as many people at the same time as possible, the other is after the most perfect audio recreation “no matter what the restrictions might be”. The latter typically results in a very small sweet spot, so only one listener can benefit from the perfectness of the system at a time.

Typically, the systems with a small sweet spot (in its ultimate version: headphones with head tracking, auralization, etc.) achieve by far the best results. Systems that aim for a larger sweet spot (a sweet area), typically have to compromise a lot, and still spend a lot of money, time and effort in the system, making it more impractical and less viable, the more perfect the system becomes.

In the anechoic room, several tests were conducted, mainly by mixing multitrack recordings and “playing around” with placement and reverb. In a critical environment as the anechoic room is -probably the most critical environment there can be-, it was found that creating a perfect sweet area in this place is virtually impossible.

Because of the lack of “obscuring” by reflections, most signals coming from the speakers maintain their “pinpiontedness”, there is hardly any “blurring” or blending. So in real life, it is probably easier to get a sweet area, but it’s probably safe to say that if a sweet area is achieved in an anechoic room, it is certainly possible to do it anywhere.

ADDENDUM 7 - Speaker setups

Although not in the scope of this research and thesis anymore, before making the decision to restrict to the 3/2 speaker setup, some experiments were conducted trying to find an optimal speaker setup. In the anechoic room 8 active speakers and a computer capable of outputting 8 channels where available. The setups that were tested are described and commented on below.

Cube

By far the worst setup was putting the 8 speakers in a cube-like setup, one speaker on every corner of the cube, pointing toward the center of the cube. The ribs of the cube were approximately 3 meters, giving it a volume of about 9 cubic meters.

It can be seen as an extended quadrophonic setup, now with some periphonic capabilities. The problems encountered with a quadrophonic setup are expanded in this setup, making it even worse.

Because the angle between two speakers is always 90 degrees, it is really hard to create a phantom image right in between two speakers, because the angle simply is too large. In the 1920's it was already found that a separation up to 30 degrees could still provide an 'acceptable" phantom image, anything beyond that gets too critical.

Also, there are no speakers situated "at ear level", which is the area in which humans are most sensitive and best at hearing. So all speakers are rather far outside the area in which they can do their work properly. And sounds supposed to come from "ear level", are always phantom-images, so inherently compromised. And since the ears of human beings typically are at the same height, this phantom-imaging-trick doesn't work as good as it should for correct imaging.

Four of the speakers are located almost on the floor, probably even in a corner, capable of providing for quite some trouble. Roommodi can be addressed rather easily, making almost sure annoyingly loud standing waves will occur. When there is more than one listener, or when objects have to be placed within the cube, it is very likely the image will get distorted, because of the blocking of some speakers.

Circle

When putting the 8 speakers in a symmetrical circle, a far better compromise (according to the author) is achieved. Although there are no periphonic possibilities anymore, the image created "at ear level" is much more stable, coherent and pleasing. The speakers are separated by 45 degrees, which could be regarded as "too much", but when using 8 speakers this is actually the only option.

Being completely symmetrical, a rather large sweet spot (sweet area) can be created, and the listener can be virtually looking in any direction without additional compromises. For Ambisonics this setup is recommended, and it could also be seen as a heavily stripped-down version of WFS.

2+2+2

As an alternative to the 5.1 system, the 2+2+2 setup was proposed by some audiophile engineers. The 5.1 implies the use of a, for musical purposes redundant, subwoofer, and the proponents for this system suggested this channel could be used for better purposes.

The stereo speakers remain in the same spot, as do the two rear speakers. The subwoofer is replaced by another “regular” speaker, similar to the ones already being used. It is, together with the usual center speaker, placed on top of the two stereo-speakers, some two meters higher. If possibl the could be placed pointing down somewhat, towards the center of the setup.

This adds some periphonic possibilities, but for the trade off of having a smaller sweet spot, because of the lack of a center speaker. It also can be argued that, in a typical concert-situation, no useful sound or reflections come from the direction of where the speakers are located. The author doesn’t find it a typically “bad” setup, but wouldn’t opt for the trade-off when loosing the center speaker.

3/2

As already described in the main work, the 3/2 setup is the one most likely to get widely used, and actually already is some kind of standard. Coming from the film-industry, where it is used as: “speech coming from the center, music from the stereos, special effects and ambience from the rear, and dinosaurs from the subwoofer”.

For musical purposes, and more so for “accurately recreating a soundfield”, this is a rather compromised setup, most importantly because of its non-symmetrical nature. Only when a source is located right in front of the listener, a rather “correct” soundfield can be created, but when coming from any other direction, things are rather heavily compromised.

7.1

As proposed by Lexicon, for their “Logic 7” surround system, the typical 3/2 setup is expanded by the use of two “side-speakers”. This on the contrary of the 7.1 system used by Sony’s SDDS, where two speakers are added in between the center and left and right speakers.

In Lexicon’s 7.1 setup, the rear speakers can be located somewhat more behind the listener, providing for a better envelopment. The two side speakers take care of a better localization, and make the system capable of reproducing rather important side reflections, which typically play an important role in a typical concert-situation.

The author finds this one of the best setups encountered so far, but because of its lack of “general support” by the audio community, it isn’t used for any further experimentation.

ADDENDUM 8 - Periphony (over-the-head-localization)

As already described in some other parts of this writing, humans capabilities of localizing sounds coming from “higher-than-earlevel” locations, is a whole study by itself. Because of the chosen restriction to only go for the 3/2 setup, this subject can be largely disregarded. But in order to come to this choice, some experiments were conducted, in order to get some clues about the importance of periphony.

It is widely known that humans are best at localizing sounds at ear level, right in front of them. At the rear things get a bit less accurate, but still are quite reliable. Up until recently, little was known about localization and how the human hearing system works outside of the horizontal plane.

It is most likely that the most important cues for localizing sound outside the horizontal plane is done by using the filtering by the pinna. This implies it is an asset which has to be trained, can be learned, and can be altered by training. Therefore it isn’t a very reliable cue.

In the anechoic room, a “pillar of speakers” was created, with speakers located from the floor up till about 3,5 meters. At about the same mutual distance, speakers where placed in the horizontal plane. Subjective listening tests where done with placing monaural sources in either the vertical or the horizontal plane. The assumption that localization in the horizontal plane is more adequate, more sensitive to changes and incorrectness than in the vertical plane, were affirmed.

Off course, when applied correctly, recreation of sound not only in the horizontal plane can greatly improve the overall performance of a sound system. But because of the many improvements that still have to be made only in the horizontal plane, it would be wise to focus on that first. Besides, it is not very likely that people will start using periphonic sound systems on a large scale in the near future.