Principles in Surround Recordings with Height

v2.01, December 8, 2011, Günther Theile[1] and Helmut Wittek[2]

ABSTRACT

New multichannel sound formats extending conventional formats like 5.1 with height channels are adding the third dimension to recordings. They provide a much wider range of spatial sound effects and allow more realism of spatial reproduction in terms of direct sound, early and late reflections, reverberation and ambient sound. Using the example of two upper layer front and two upper layer surround complementary loudspeakers (5.1+4, also known as “Auro-3D 9.1”) the psychoacoustic principles in the perception of elevated phantom sound sources, spatial depth, spatial impression, envelopment, ambient atmosphere, as well as directional stability within the sweet area are discussed. Concrete proposals for microphone configurations are given evolving from these considerations.

1. SURROUND SOUND IN TRANSITION

After the international ITU-R BS.775-1 standard had been released in 1992, it took key-media vendors some time to implement the necessary techniques and to gain sufficient expertise in using them. In recording, switching from 2.0 to 5.1 was the first considerable step away from “pure” stereophony with two loudspeakers placed in front of the listener towards the realistic reproduction of an acoustic environment.

5.1, however, was just a compromise. It was necessary due to restrictions such as compatibility with 2.0 stereo and the fact that at that time cinema formats supported a maximum of six channels. Therefore, 5.1 essentially brought along not more but two improvements [1]:

- It increased the listening area and improved the stability and quality of stereo sound by subdividing the L/R basis, which is 60° in width, into two stereo sub-ranges with 30° each (L/C and C/R).
- Within certain limits, it allowed for creating a realistic acoustic environment by placing to additional surround speakers behind and on the sides of the listener.

A few years ago, we found that virtually the entire industry was ready for using 5.1 in production, distribution, and on end-user equipment. In addition, consumers today typically accept the presence of a larger number of speakers – at least when used as components of a home-cinema setup. On the other hand, we discovered that only a limited number of listeners are able to achieve the sound quality that can actually be realized using a surround system – or the quality they had hoped for. There are several reasons for this:

- The listening environment is unfavorable in terms of room geometry or acoustics, the arrangement of the speakers is not standards-compliant, or device settings are inappropriate.
- The recording quality is bad. This results either from economic constraints in production or from inappropriately chosen miking and mixing techniques.
- The 5.1 listening zone is too narrow. There are recordings that require a perfect listener placement, assuming that only the "sweet spot" matters.
- Limitations of the 5.1 format including improper 3-D imaging, proper speaker positioning in height and in relation to the listener's head, and imperfect distance imaging.

The above list is not necessarily in order of importance; however, it illustrates that problems arise mostly when it comes to practical application. This is equally true for the producer and the listener. Eliminating the issues just by increasing the number of channels and speakers is not possible; in fact, recently introduced enhancements and innovative systems ranging from various 7.1 formats to Higher-Order Ambisonics (HOA) and Wave-Field Synthesis (WFS) require new paradigms, new hardware, and special attention from recording engineers. Plus the listener still needs to accept a living room in home-cinema style. In this context, the current variety of formats and the lack of standards present an additional obstacle. The current DCI specification (or SMPTE 428M, respectively) specifies channel mapping and purposely allows for any use of 16 channels.

The ITU-R BS.775-1 standard already specified optional LL and RR speakers located between the front and surround speakers. This improves the stereo quality of side imaging, enlarges the listening zone, and fills the gap between frontal and side imaging. Altogether, this leads to more flexibility for reproducing stationary audio events at the side or the critical lateral reflections. In conjunction with new developments in film sound, companies such as DTS and Dolby follow this principle and promote various 7.1 formats. These use a similar array where four surround speakers are spread laterally and behind the listening zone while utilizing the same front-speakers arrangement (L/C/R). Today, several hundred Blu-ray discs offering 7.1 audio are available for home-cinema use. Those media excel with clear sound definition and stable directional imaging at the sides and behind the listener; however, there are hardly any music recordings [2].

All those surround formats are essentially based on stereophony, i.e. they use phantom sources between two adjacent speakers for source imaging. In surround, the direction of the phantom source greatly depends on the listening position and is highly unstable; therefore, directional imaging virtually relies on the physical speaker positions. The volume balances are position-dependent as well. This is particularly true for the relation between front and surround sources. Therefore, adding more channels on the horizontal plane aims at enlarging the listening zone and providing a more homogenous and more stable directional resolution.

There are alternative ways of using additional channels, leaving the horizontal plane. Arranging speakers above the listener's head complements the spatial area, allowing for creating a 3-D sound within certain limits. In 2001, *Werner Dabringhaus* published the first music recordings produced using his 2+2+2-recording technique. This approach is based on 5.1 but does without center and subwoofer speakers; instead, it uses two height speakers positioned on top of L and R [3]. This concept was designed with the DVD-Audio in mind. The objective was to reproduce the sound from the concert hall as realistically as possible, so it used speakers allowing for imaging height information rather than center and subwoofer speakers. Similarly, *Tom Holman* integrated the third dimension using two tilted height speakers placed in front of the listener; however, his 10.2 Channel Surround Sound setup requires eight channels on the horizontal plane and was originally created for cinema and home-cinema applications [4].

In 2006, *Wilfried Van Baelen* introduced the Auro-3D format that specifies four extra channels for height information. With the Auro-3D 9.1 basic version, the height speakers complement the 5.1 format – they

are positioned above the L, R, RH, and LH speakers (figure 1, [5]). Of course, similar formats such as 7.1 Surround can be complemented using four height speakers, for example, in a “quadraphonic” array.

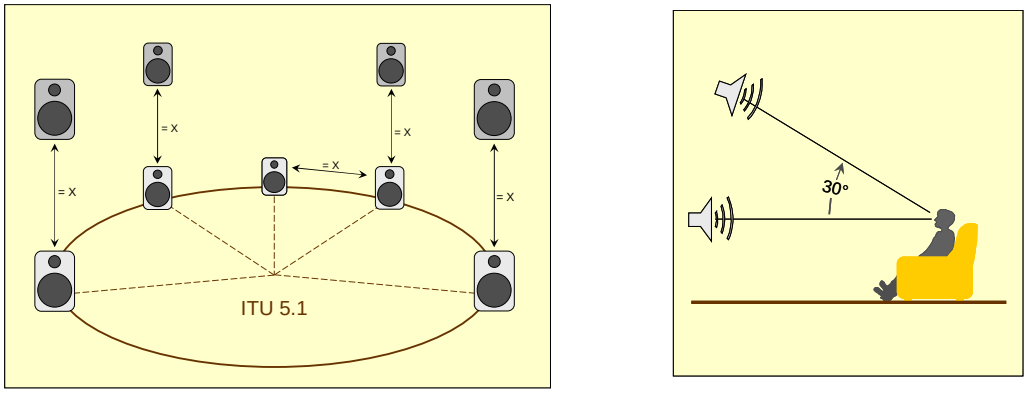

Figure 1: Auro-3D 9.1 basic setup (according to [5]), backward compatible with ITU-R BS.775-1

The main feature of this format is the cube-like arrangement of eight speakers. It allows for including the entire upper half space for the reproduction of (early) reflections and for appropriately reproducing the subjective spatial diffusion of the reverb part. The format has provided an excellent starting position for imaging parameters such as envelopment, spatial impression, and depth. In addition, the height speakers obviously offer the same possibilities for stereo imaging as the ITU setup without the center speaker. On the other hand, creating phantom sources between the lower and upper speakers (i.e. stable directions for stationary audio events with an elevation of 0 to 30°) as well as immediately above the listener’s head is practically not possible. We will discuss this shortly.
Some limitations of the 5.1 format can be eliminated or alleviated using Auro-3D 9.1; others cannot. Table 1 lists a number of attributes of reproduced sound. The first four parameters affect the direct portion (and are normally modified using panning); the other four attributes refer to the effects of indirect sound (designed using miking techniques and processing). These attributes allow for categorizing and comparing the profiles of the various techniques in a reasonably adequate fashion, provided that the reproduction recommendations have been implemented properly and appropriate miking and mixing techniques have been used on the recording side.

As the table shows, Auro-3D 9.1 offers some specific benefits compared to other speaker-arrangement techniques. This also applies to other formats complementing 2-D surround systems with quadraphonic speaker arrays on the plane above the listener. In section 4, we will describe arrangement options and limitations in detail with a focus on relevant miking techniques.

3-D SOUND FOR 3-D VIDEO

New developments in dummy-head recording (e.g. [6], [7]) marked the start of serious and partly successful efforts to establish 3-D in broadcast and on recording media. The original method of 3-D audio is a binaural reproduction of ear signals. Ideally, the reproduced dummy-head signals are identical with the ear signals the listener would perceive at the dummy-head position inside the recording room. In this case, the virtual listening experience would match the real sound inside the recording room. Unfortunately, binaural techniques are limited to special applications due to various practical reasons [8]. They are not compatible with speaker reproduction, that is, their 2-channel signals cannot be converted to multichannel speaker signals producing the same effect. On the other hand, the quality of 3-D imaging that can be achieved using binaural techniques may be used as a reference: The imaging zone includes the entire upper half space, and audio events of any elevation and distance can be represented.

In the interest of completeness, we also want to look at the intra-active perspective. This is a feature of natural auditory scene analysis. The way directions are perceived changes depending on the distance to the sources: when nearby sources move, they travel “further” than remote sources. WFS systems allow

for reproducing this behavior [9] – a fact of that may be interesting, for example, for future developments in gaming. However, we will not go into the details in this paper

.

ATTRIBUTES OF SOUND REPRODUCTION	2.0 STEREO	5.1 SURROUND	AURO-3D 9.1	WFS*	BINAURAL TECHNIQUES
Front direction	●	●●	●●	●●	●
Surround direction		●	●	●●	●●
Elevation			(●)***		●●
Height			●		●●
Distance/depth	(●)**	●	●●	●●	●●
Proximity to the head				●	●●
Intra-active perspective / [1]				●●	
Spatial impression	(●)**	●	●●	●	●●
Envelopment		●	●●	●	●●
Timbre	●●	●●	●●	●	●●

Table 1: Comparison of stereo/surround-format profiles
(requires appropriate recording and reproduction techniques)
*horizontal arrays; **emulated depth/spatial impression; ***unstable; in the “sweet spot” only

1.1. Speaker Reproduction

Which features of 3-D audio reproduction are appropriate for 3-D video? The first thing we noticed is that the initial situation is different compared to audio. The flat two-dimensional image is converted to 3-D video by creating a sense of depth using the means of stereoscopy within the limits of video reproduction[1]. Unlike, with audio, the third dimension is height (the other two being direction and distance). Regardless of the extent to which the possibilities of imaging are limited, 2.0 stereo, 5.1 surround, and conventional WFS are definitely 2-D techniques. This is particularly obvious with 2.0 stereo, which emulates distance and depth and limits the listening zone to a 60° angle; with 5.1 surro und and WFS, the limitations are not so clear [1], [9] (see table 1).

In principle, WFS and HOA allow for adding height channels in order to enable true 3-D audio reproduction. As the spatial resolution can be lower for sources out of the horizontal plane, usually single channels are added and driven with stereophonic signals, just as with Auro-3D.

Auro-3D (9.1 and above) can be considered the most efficient format for 3-D audio reproduction. It meets many of today’s requirements to a universal and compatible future-oriented standard for digital cinema, games, broadcast, and the music industry [5]. As we will describe in detail, engineers recording for an Auro-3D speaker array need to pay special attention to the phenomena of psychoacoustics in order to achieve good results when implementing specific creative ideas. After the introduction of the 5.1 surround channels, the inclusion of height has been the second step towards enhancing freedom in speaker stereophony. One of the most sophisticated tasks is recording music “realistically”. It requires the use of a special miking technique to control the four main attributes of 3-D recording at the same time – source direction and width, depth of the scene, spatial impression, and envelopment. Based on that recording situation, we will explore the new creative possibilities in the following sections.

/ [1] More precisely, a distinction is made between 2½-D reproduction (where the viewer moves to perceive depth) and 3-D reproduction (depth is intuitively perceived due to stereoscopy).

1.2. Headphone Reproduction

Current convolution methods allow for realistically imaging a virtual Auro-3D studio using headphones. Commercially available Binaural Room Synthesis (BRS) systems ensure virtual 5.0 speaker reproduction in professional quality. In addition, they can easily be modified to support additional height channels. A BRS system convolves surround signals with the sampled binaural room impulse responses (BRIRs) of a high-quality studio. Data suitable for convolution are selected using head tracking. This method takes the current head orientation into account, so the listener locates the virtual speakers regardless of the head posture (i.e. in relation to space) [10]. In 2007, the IRT released a BRS plug-in for VST-compliant host applications [11]. In the meantime, a cost-effective BRS standalone device capable of perfectly emulating the studio environment using individual equalization is available [12].

This technology allows for autonomously producing Auro-3D recordings on the OB truck and in any other scenario with unfavorable monitoring conditions. Engineers can take their familiar monitoring environments wherever they go. Several monitoring scenarios are available at the press of a key, allowing, for example, for checking the sound beyond the “sweet spot” or comparing various speakers or listening rooms. Using BRS, consumers can achieve significantly better reproduction quality with Auro-3D signals than living-room speakers would allow at all. In addition, BRS makes the listener completely independent from the selected speaker array: If fed with suitable material, a BRS processor can essentially emulate virtually any multichannel speaker setup. This eliminates all the practical problems that come up when placing speakers at home properly.

BRS will considerably speed up the acceptance of production quality, multichannel audio, and, in particular, Auro-3D in the market.

2. PSYCHOACOUSTIC REQUIREMENTS TO MULTICHANNEL AUDIO

The human ear evaluates various properties of the sound field and uses them for spatial hearing. Table 2 roughly outlines the meanings of direct sound, early reflections, reverb, and listener envelopment for each of the above sound attributes and the timbre. Enveloping sound includes both diffuse-field sound (background noise, “atmo”=ambience) and audibly decaying reverb.

The ear is typically capable of intuitively (or spontaneously) distinguishing between these three portions in natural sound; however, the more localization and timing are deteriorated due to inappropriate reproduction, the more difficult it is to achieve this intuitive distinction. A good example is a mono recording where direct sound, early reflections, and reverb sum up to a heavily colorized sound mush. In this case, spatial perception is exclusively based on conscious recognition. For example, a long reverb implies a large room, low-level direct sound means “far distance”, etc.

SOUND ATTRIBUTES IN THE HALL	DIRECT SOUND	EARLY REFLECTIONS	REVERB	BACKGROUND NOISE
Direction/elevation	●●	●		
Distance/depth		●●		
Spatial impression		●●	●	
Envelopment			●	●●
Timbre	●●	●	●●	

Table 2: Interrelation between sound attributes and sound-field types

2.1. Reflections and Reverb

Indirect sound portions allow for reproducing the recording space. The relation between direct and indirect sound determines the spatial attributes of a sound event. Figure 2 shows this interrelation. The natural pattern of early reflections occurring at a delay of 15 to 50 milliseconds plays a key role in spatial hearing. When it comes to recording, this portion of reflected sound deserves special attention as it critically affects attributes such as distance, depth, and spatial impression. The hearing takes spatial information from early reflections and converts it to a spatial event. With natural sound, the human ear performs this conversion spontaneously and with amazing robustness because that type of sound contains all properties of a reflection pattern in their original form. Key parameters include

- the timing structure in relation to direct sound
- levels and spectrums
- horizontal and vertical incidence directions

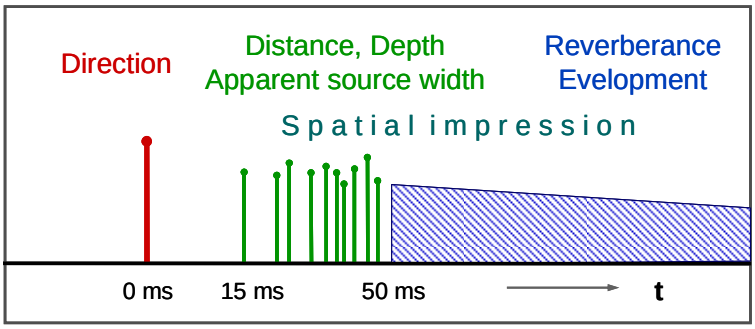

Figure 2: Influence of attributes on sound impression over time

Imaging a spatial environment is realistic when the ear is able to recognize and interpret the features of the reflected sound – that is, when it "understands" the reflection pattern. Therefore, the reproduction must be absolutely consistent with a real spatial environment. The same applies to the spatial distribution of the early reflections' incidence directions. Meeting this requirement using room microphones is hardly possible (see section 4) because one needs to keep acoustic crosstalk on the ambient-microphone channels as low as possible (approx. 10 dB at the most). A single reflection coming in from a specific direction – say, the top-right corner of the rear part of the hall – should be reproduced as such; it must not be picked up by the "wrong" mikes.

The perception of distance and depth mostly depends on early reflections. This can be proven by simply adding pure early reflections (without the reverb) derived from a real room to a source that has been dryly recorded. The source is perceived as distant, which is in correspondence with the reflection pattern. Perception is particularly stable when the reflections come in from the original directions of the upper half space. Reproducing depth requires careful handling of early reflections.

Adding appropriate reverb at a suitable level creates a natural sense of depth and realistic spatial impression /[2]. Even with short reverb times, the virtual reproduction of these two attributes creates a realistic spatial impression. Increasing the reverberation time, for example by using concert-hall or church reverbs, adds another attribute of spatial hearing: the envelopment.

2.2. Diffuse Sound

Ambient sound (or noise) consists of a large number of spatially distributed individual acoustic sources that cannot be separately localized. Rustling leaves in a wood, audience noise and response, and

/ [2] The term "spatial impression" refers to the effect of early reflections and early reverb on localization. Due to reverberation inside the room, the apparent source width (ASW) seems greater, and the source event appears to be "fuzzy" in time.

applause at a performance are typical examples. Unlike indirect sound, this portion of the surrounding sound cannot be created using effect units, so appropriate miking is essential.

When recording indoors, using room/ambience microphones for recording ambient sound as well is obvious. With some trying and testing, an experienced engineer can create a realistic balance (for example, upper/lower space) between reverb and background sound (applause, audience noise) by carefully selecting capsule and polar patterns and sensibly placing the microphones; however, there are situations where this cannot be achieved, and it should be avoided while actually recording. The use of an 8-channel reverb unit provides more flexibility: It allows for routing the background sound to the lower speakers while reverb is fed to all eight channels.

3. 3-D AUDIO WITH AURO-3D

The speakers on the upper plane obviously have the same imaging capabilities as those on the horizontal plane (except for the Center speaker). The stereo image in the L/C/R range is complemented by 2-channel stereo sound on the upper L_h/R_h base. Similarly, the additional height speakers can be used in the same way as those on the horizontal plane. This arrangement already enhances flexibility considerably. The possibilities resulting from the interaction of the two planes are interesting. In the following sections, we will describe source imaging using the five speakers in front of the listener and the reproduction of reflections and diffuse sound field in the 3-D surround array.

3.1. The Upper and Lower Representation Areas

Elevating Sources

Unfortunately, the familiar stereo imaging of localizable sources can be achieved only at the upper and lower edges of the area in front of the listener (i.e. between L-R and L_h-R_h). A localization of phantom sources between the upper and lower speakers is highly unstable due to propagation-delay differences and also depends on the spectrum. Elevation cannot be achieved just by using panning functions – this would affect sound and spatial perception in a way that cannot be controlled. Figure 3 shows a practical analysis of stereo-level relations between speakers arranged one above the other (0° and 45°) in front of the listener [16]. It is obvious that reliable localization cannot be achieved even from the „sweet spot“ and with correct delay relationships; this is similar to lateral phantom sources. Thus, stationary-source elevation cannot practically be accomplished.

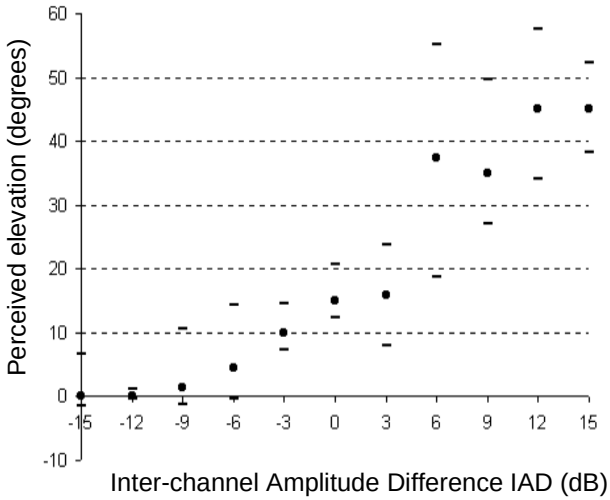

Figure 3: Stereo imaging on the median plane affected by differences in level (speaker angles: 0° and 45°), taken from [16].

In addition, very small differences in propagation delay result in the phantom source migrating upwards or downwards. A delay of just 0.5 milliseconds is sufficient to move the audio event to one or the other side. Coloration will occur as well. Thus, the listening zone is greatly limited regarding depth and height. Figure 4 shows the delay conditions in an Auro-3D home-cinema speaker array.

Elevating or upward-expanding a stationary source using the upper speakers is practically not achievable in a stable way. This is particularly true where a large listening zone is required. Trying to solve this issue using panning functions would not be successful and could also result in coloration (which would, however, be masked almost completely by the diffuse-field portion). This scenario is similar to using the L/LS and R/RS side-speaker pairs – the speakers are the only stable source positions. Moving sources can, however, be represented with panning within certain limits.

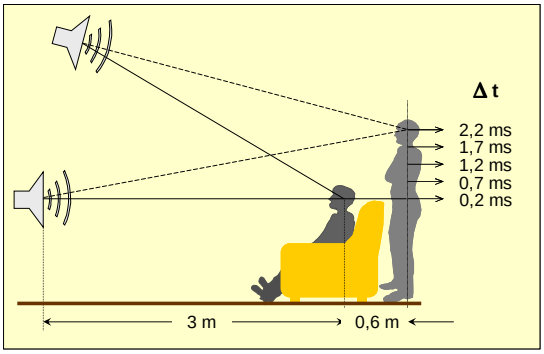

Figure 4: Delay differences occurring in listening positions beyond the sweet spot

Filling Up the Areas

Much better conditions exist for reproducing a large number of spatially distributed individual acoustic sources that cannot be separately localized. They have properties similar to those of a largely spaced A/B setup or a Decca tree: While directional imaging is not practicable due to the mapping curves being much too steep [8], reproducing a well-balanced imaging, for example, of a large orchestra and the reflections produced by it is possible. The risk of creating a “hole” in the center is controllable in many recording scenarios, in particular where the diffuse-field portion dominates the sound. Therefore, filling the areas in height is actually possible and an important creative element.

3.2. Reflections and Diffuse Sound

The approach allows for distributing, in particular, the early reflections in the upper plane. This is due to the delay differences of individual reflections on the capsules. Reflections come in naturally from upper directions, too.

The preferable distribution of the reflections reduces their spatial density, allowing the ear to better distinguish spatial information. Figure 5 shows the effect for the transition from 2.0 to 5.1 to Auro-3D. Another critical factor in this context is a positive effect on the timbre, which results in improved perception of reflections.

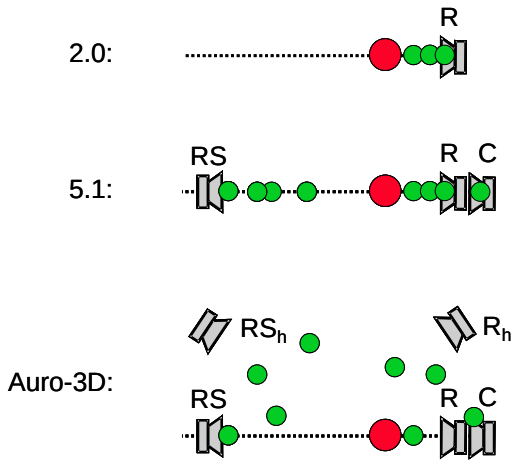

Figure 5: Spatial distribution of reflection patterns in 2.0, 5.1, and Auro-3D

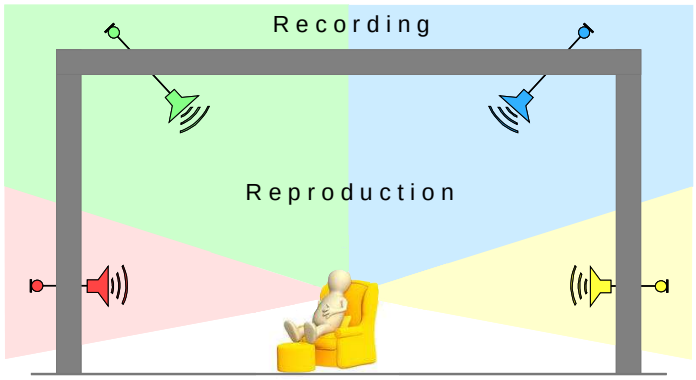

Figure 6: Reproducing the original incident directions requires strict channel separation during the recording process

4. RECORDING FOR AURO-3D

4.1. The Importance of Psychoacoustics for a Suitable Recording Technique

When looking for a suitable recording technique for Auro-3D, knowing the principles of natural hearing is quite helpful. Considering the complexity of the subject, one may decide that trying and testing would be appropriate – after all, it rarely sounds worse when feeding any portion to the height speakers. However, it soon becomes clear that this is not what we want. We know from our experience in searching the proper recording technique for 5.1 that a discrete recording needs to be considerably better than an automatic upmix (which Auro-3D already supports!).

However, practical investigation is needed to be able to verify, refine, and implement the conclusions that are based on the mentioned psychoacoustic expertise. Therefore, concrete methods and guidelines will evolve in the following years. The scientific approach will take the opposite direction as well: General guidelines may be specified on the basis of great recordings, and knowing why a recording sounds good

and why a guideline exists cannot be disadvantageous. After all, all guidelines are to be proven in practice. If one forgets why there is a guideline, it soon will become an antiquated custom.

The purpose plays a key role in determining the suitable recording technique. With 5.1, too, there are techniques that are more suitable for delivering convincing spatial imaging, and others that are better for use with spot microphones.

4.2. Channel Separation

To create the spatial resolution of direct, background, diffuse, and/or early-reflection portions as described above, microphone placement needs to ensure a sufficient level of acoustic channel separation (see figure 6); otherwise, spatial arrangement of multiple speakers as specified by Auro-3D would be hardly useful.

There is no doubt that realizing acoustic channel separation for room miking becomes more difficult, the larger the number of playback channels is. There is an increasing risk of undesired crosstalk, i.e. correlated contents on three or more speakers. This again results in clear coloration (sometimes called “phasing”) that also depends on the listener’s position within the listening environment. Placing the microphones in a way that no unwanted crosstalk occurs is very difficult with nine channels! There are two solutions that also work with 5.1: either using optimized techniques such as OCT surround or increasing the distances between the microphones and thus the propagation delays in order to alleviate crosstalk.

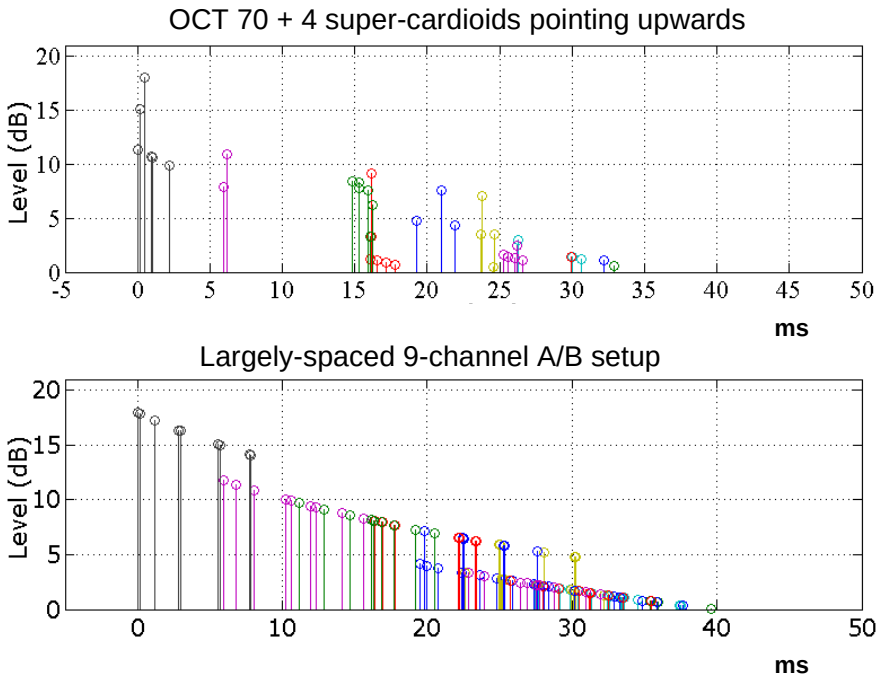

Figure 7: Reflection patterns in the “sweet spot” of an Auro-3D speaker array generated using 2 different microphone arrays. The microphone arrays record the same source. A shoebox-shaped recording room was produced for emulation purposes. The source produces a Dirac impulse. Each peak color corresponds to a (1st order) image source.

Figure 7 shows two sample arrangements and the highly different reflection patterns they produce. The example simulates a source (with first-order image sources) reproducing a Dirac impulse in a rectangular

parallelepiped room (a “shoebox”). The figure shows the first 50 ms of the resulting signal at the „sweet spot“ of an Auro-3D speaker arrangement.

The upper image contains the reflection patterns generated by a 9-channel arrangement similar to OCT (OCT70 plus four supercardioid microphones pointing upwards). Direct sound (black peaks) and the reflections produced in the recording room are reproduced with highest clarity and without any crosstalk from the direction that is consistent with the recording room. The second image shows a largely-spaced 9-channel A/B setup. The conditions are entirely different: Obviously, there are hardly any utilizable discrete reflections, and reverb builds up very quickly. Even the direct signal has a wide and reverberant character; however, this may actually be desired: Recording in long-reverb spaces where the diffuse-field (the envelopment) dominates the listening experience – for example, in a church – results in a great surround sound; presence and imaging stability can still be enhanced using spot microphones. Achieving a degree of imaging, depth, and distance perception corresponding to the recording room will definitely not be achieved.

4.3. Using Artificial or Convolution Reverb

Modern technologies would also allow for alternative approaches based on convolution. The necessary spatial information is gained either by sampling the physical recording room or existing rooms of high acoustic quality, or by using calculated models. Basically, the concept uses convolution algorithms for several locations in the area of the sources to be imaged (e.g. a stage). This allows for convolving signals from separate microphones or microphone groups with the room’s IRs from specific room directions. For Auro-3D 9.1, this requires eight convolutions per source signal (with the IRs from the eight corners of the room). Figure 8 shows the principle for a specific stage area (microphone group A).

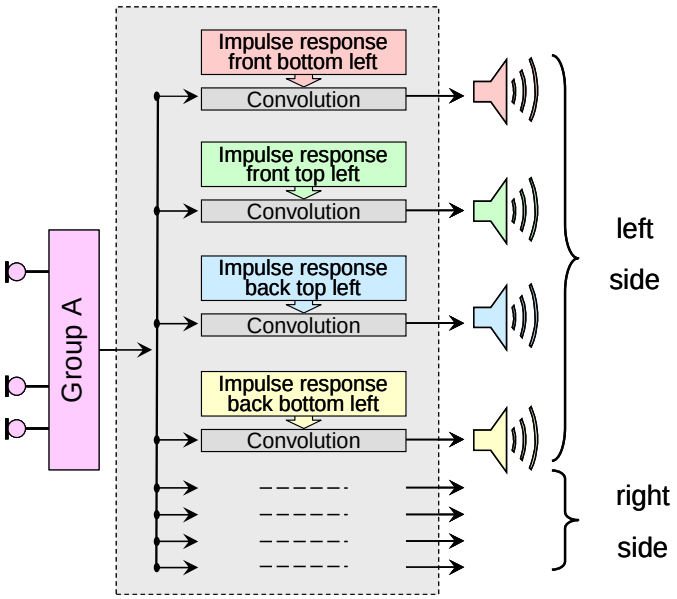

Figure 8: Concept of a convolution processor producing 8-channel early reflections

If one decides not to use model-based IRs in order to ensure realistic imaging, the IRs need to be sampled in advance using suitable directional microphones. In addition, if the microphone’s directivity is not adequate, unwanted sound-incidence directions likely to cause crosstalk can be shaded. (This might also include direct sound.) Afterwards, any future recordings made in that room can be convolved with the sampled impulse IRs. If desired, the engineer might do the mix without using convolution reverb and record the diffuse-field (including background noise) using room microphones. This allows for creating a

realistic balance between the reverb and applause / audience noise. The use of convolution, however, eliminates a number of practical recording problems and also provides more freedom of creativity.

4.4. Diffuse Sound

Diffuse sound (i.e. reverb or background noise) needs to be reproduced diffusely. This can be achieved using Auro-3D if appropriate signals are fed to the extra speakers. Diffuse signals must be sufficiently different on each speaker, that is, they need to be decorrelated over the entire frequency range. A sufficient degree of independence is necessary, in particular, in the low-frequency range as it is the basis of envelopment perception (for an example, see [14]). However, increasing the number of channels that need to be independent makes recording more complex. It is a tough job to generate decorrelated signals using first-order microphones – for example, a coincident array such as a “Double MS” array or a Soundfield microphone allows for generating a maximum of four channels providing a sufficient degree of independence [17]. Therefore, the microphone array needs to be enlarged to ensure decorrelation.

It is worth noting here that measuring diffuse-field correlation is not trivial. There are two reasons for this: First, measuring the correlation requires the diffuse sound level to be much higher than direct and reflection levels, so the distance from the source needs to be sufficiently long. Secondly, considering the degree of correlation is not sufficient; this does not account for the fact that low-frequency (de)correlation is particularly important (see [13]).

A study on the effects of diffuse-field correlation may be useful for determining the required minimum spacing and angles of microphone pairs. Coincident, equivalent, and delay-based techniques might be suitable for eliminating diffuse-field correlation (see [13]). Figure 9 shows the interrelation between the DFI predictor (a frequency-weighted degree of coherence) and the subjectively perceived stereo width. Mono portions in the diffuse-field often distract listeners due to its narrowness and the coloration they produce. Several coincident, equivalent, and delay arrays were simulated. We assume that only those arrays causing low diffuse-field correlation will be acceptable as they do not restrict the perception of spatial width (i.e. quantitation > 2). There are six arrays meeting this requirement: the Blumlein pair array (2 coincident figure-eight microphones, ±45°), two equivalent cardioid arrays, and three omnidirectional arrays where the microphones were spaced more than 35 cm.

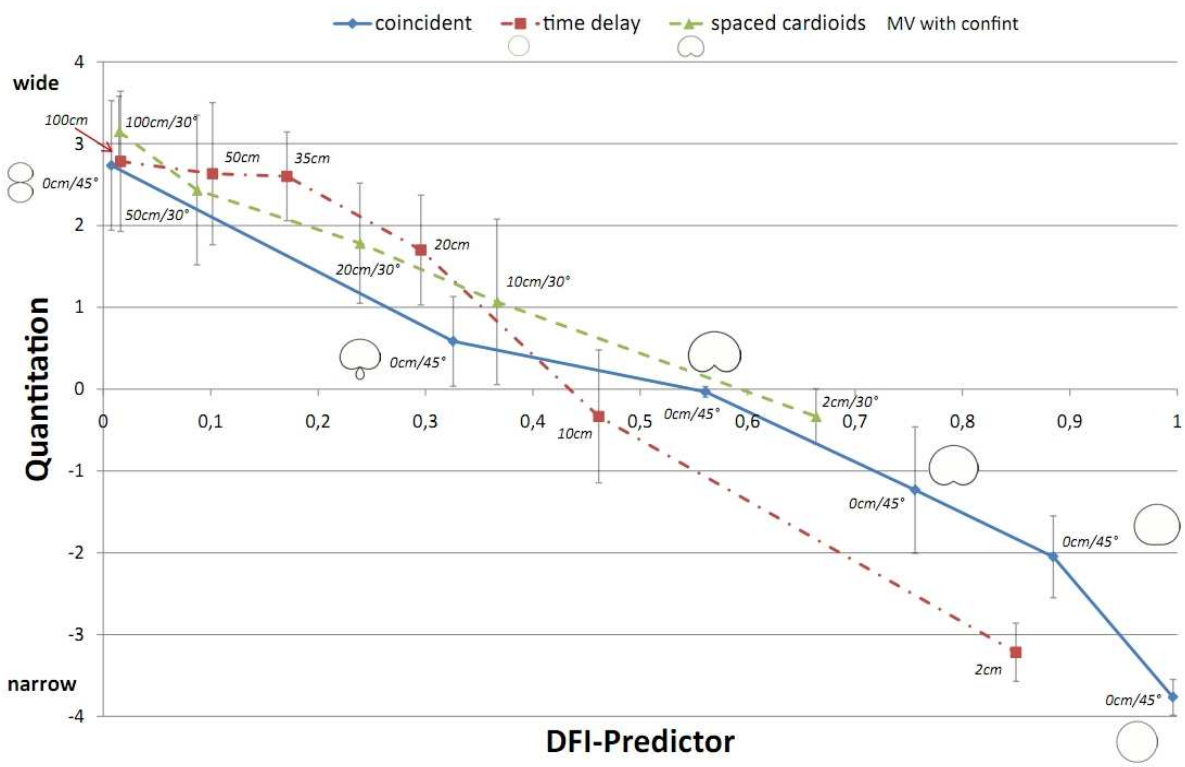

Figure 9:

Interrelation of the DFI predictor and the perception of spatial width (taken from [13]).
Arrays include (from left to right):

Coincidental (blue): r=0 cm, ±45°, omnidirectional portion: 0 (Blumlein), 0.4, 0.5 (cardioid, X/Y), 0.6, 0.7, 1 (mono)
Equivalent (green): Cardioid array, ±30°, spacing: 1m, 50cm, 20cm, 10cm, 2cm;
Delay (red): Omnidirectional array; spacing: 1m, 50cm, 35cm, 20cm, 10cm, 2cm

4.5. Design of an Auro-3D Main Microphone

The design of an Auro-3D main microphone needs to account for a number of aspects that result from the preceding considerations. A large number of basic requirements are already known – for example, the laws of directional imaging and of spatial perception. Because of the large number of speakers and the resulting interaction between them, it has become more difficult to find a suitable miking technique meeting all requirements.

On the other hand, following a trial-and error-approach is perfectly legitimate and will often lead to success. This is also because yet by acoustically exciting the upper half of the reproduction room a positive effect is generated.

Summarizing the preceding sections, these 4 main physical differences can be expected when adding height loudspeakers, see also figure 5:

1. More possible directions for discrete sources
2. More possible directions for reflections
3. Lower source/reflection density
4. Higher diffuseness of the diffuse portions, more evenly distributed diffuse field

Consequently, these 4 perceptual differences may be hypothesized after adding height loudspeakers:

1. Enhanced distribution of sound sources
2. More natural perception of distance/depth
3. Less coloration
4. More natural spatial impression; larger “diffuse field sweet spot”

From this, we conclude for the microphone recording technique:

Directional Imaging (direct sound, early reflections):
Stereophonic rules (ΔL, Δt) apply in general for all loudspeaker pairs. However, as the height loudspeakers are potentially displaced by more than 1-2 ms, this has to be taken into account in microphone placement. For example, it doesn’t make sense to use too closely spaced omnis between L and Lh. The imaging between upper and lower loudspeaker plane can be realized only roughly, yet it is important to “fill up the area”.

A useful tool for designing stereophonic microphone setup with regard to optimal imaging characteristics is the “!mage Assistant” [15].

Channel separation:

- for discrete signals (direct sound, early reflections):
  One signal on more than two loudspeakers leads to coloration and thus should be avoided.
- for diffuse signals:
  the more de-correlated the diffuse sound is reproduced between the channels, the more diffuse is the resulting sound field.

Channel separation can be successfully achieved by using directive microphones, which generate a level difference by aiming them in different directions. Furthermore, channel separation is also achieved by spacing the microphones apart. When the spacing is sufficiently large, the diffuse sound is decorrelated at enough low frequencies and even further level differences are produced by the inverse square law.

Derived from these postulations as well as derived from the existing practical experience, the following two proposals for a microphone setup for Auro-3D can be given:

A. ORTF-like recording techniques

These techniques consist of relatively closely spaced (<= 1m), directive microphones. The typical sonic properties of these techniques are a proportional and clear directional imaging and a natural spatial impression.

These recording techniques are applied when in the aesthetical approach an emphasis is put on a proportional directional imaging and a natural (=close-to-real) spatial impression. The typical applications include chamber music, drama, sports, ambience recording.

A specific proposal for an ORTF-like setup is the “OCT 9” technique, which is derived fro the well-known OCT Surround technique for 5.1 Surround [1]. Its advantages are the low inter-channel crosstalk, very smooth and balanced frontal localization and the absence of direct sound in the rear and height channels leading to a large listening area.

The diffuse sound correlation is optimally low.

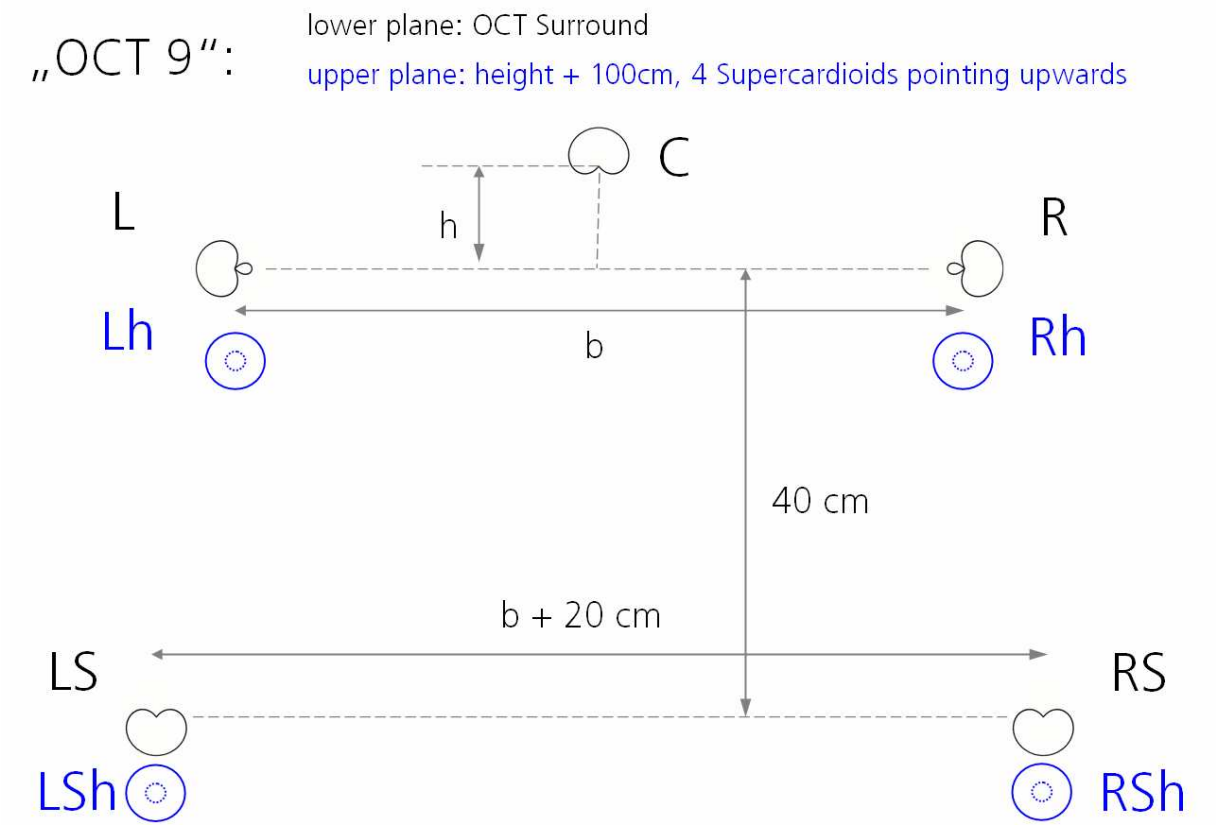

Figure 10: OCT 9 main microphone technique for Auro-3D 9.1
The technique is derived from the OCT Surround technique [1]. The spacing “b” determines the recording angle. A typical value for “b” is 70 cm, the Center displacement to the front “h” is 8 cm. The 4 height channels are fed with 4 supercardioids facing upwards. Their position is about 1m higher than the 4 microphones for L, R, Ls, Rs.

B. Wide a/b-like recording techniques

These techniques consist of widely spaced, omni-directional microphones. These techniques typically create a stable, but not a proportional directional imaging. Furthermore, the spatial impression usually is desirable and impressive. The spatial impression is different to that created by the ORTF-like techniques as it is rather impressive than natural, which nevertheless is often desired. Furthermore, these techniques often are chosen due to their good sound color properties as they are using (mainly) omnidirectional microphones. The typical application is classical music recording and film music.

No precise proposals exist for these techniques as the microphone placement is not found due to calculated localization curves. A possible variant looks like this:

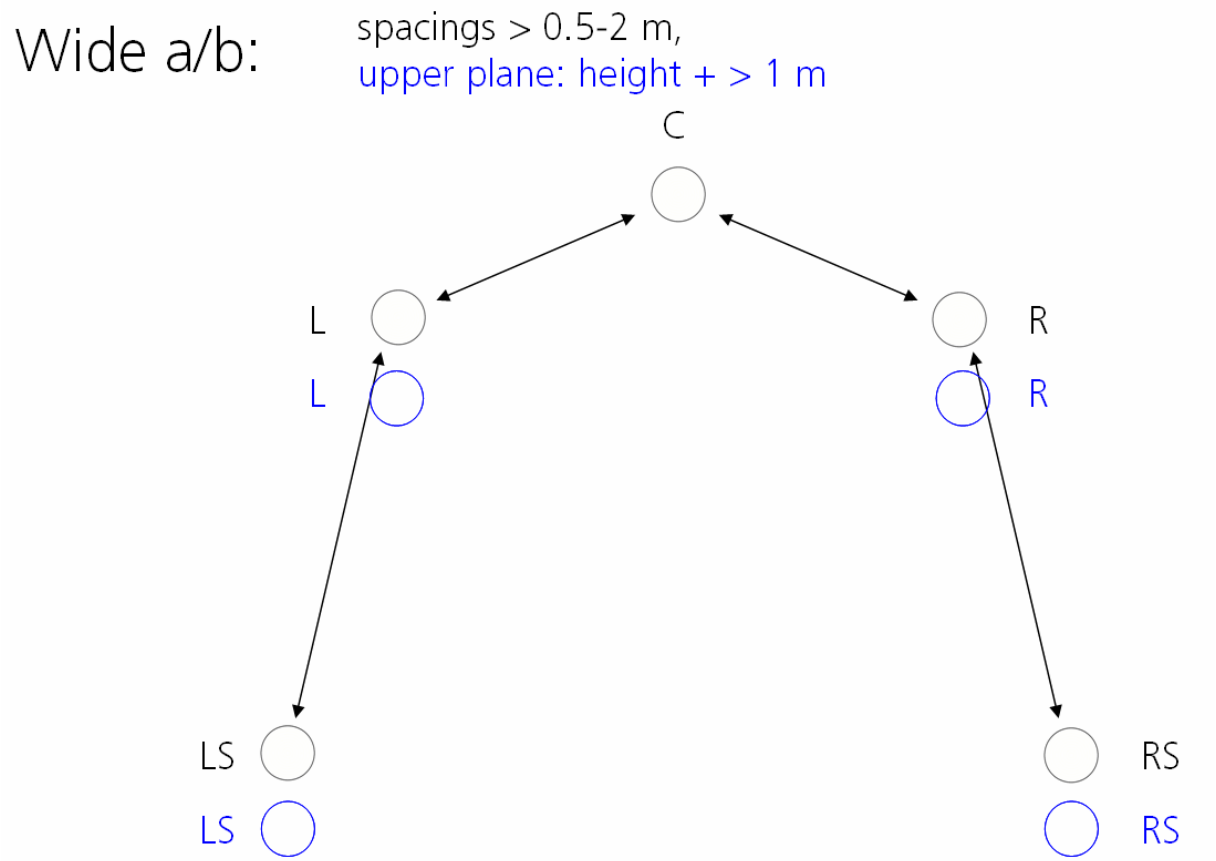

Figure 11: A possible variant of a Wide a/b-like recording technique
9 omnidirectional microphones are placed not dissimilar to the actual Auro-3D 9.1 loudspeaker setup.

Further experience is needed to refine these microphone placement proposals and to carve out the maximum effect of the height loudspeaker plane. One should, however, refrain from trying to compare 5.1 and Auro-3D by just switching the height speakers off and on. Such a comparison would be misleading due to the loudness difference occurring and spatial information missing. After all, the listener/consumer needs to be convinced of the true added value offered by the individual professional performance that goes along with this innovative reproduction technique. For that purpose, we need not only to improve spatial reproduction but also require new ideas for an aesthetic use of the height channels.

5. REFERENCES

[1] Theile, G.: “Natural 5.1 Music Recording Based on Psychoacoustic Principles”. Nordic Sound Symposium XX, BOLKESJØ, 2001.
www.hauptmikrofon.de/theile/Multich_Recording_30.Oct.2001_.PDF

[2] Wikipedia: “7.1 surround sound”. http://en.wikipedia.org/wiki/7.1_surround_sound

[3] Dabringhaus, W.: “2+2+2 Aufnahme-Verfahren”. www.mdg.de/frame2.htm

[4] Holman, T.: “10.2 channel surround sound”. http://en.wikipedia.org/wiki/10.2

[5] Van Baelen, W.: “Challenges for Spatial Audio Formats in the near Future”. In: “26. Tonmeistertagung 2010. Tagungsbericht” (ISBN 978-3-9812830-1-3), pp. 196-205

[6] Theile, G.: “Zur Kompatibilität von Kunstkopfsignalen mit intensitätsstereofonen Signalen bei Lautsprecherwiedergabe: Die Klangfarbe”. In “Rundfunktechn. Mitteilungen 4/1981”, pp. 146-154

[7] Steickart, H.: “Kopfbezogene Stereofonie – neuere Erfahrungen bei Produktion und Rezeption”. In: “15. Tonmeistertagung 1988. Tagungsbericht”, pp. 316-331

[8] Dickreiter, M., Dittel, V., Hoeg, W., and Wöhr, M.: “Handbuch der Tonstudiotechnik”, vol. 1, chapter 5. K. G. Saur Verlag Munich, 2008 (ISBN 978-3-598-11765-7)

[9] Wittek, H.: “Räumliche Wahrnehmung von virtuellen Quellen bei Wellenfeldsynthese”. In: “23. Tonmeistertagung 2004. Tagungsbericht”, pp. 268-297. http://hauptmikrofon.de/HW/Wittek_TMT2004_Paper_final.pdf

[10] Horbach, U., Pellegrini, R., Felderhoff, U., and Theile, G.: “Ein virtueller Surround Sound Abhörraum im Ü-Wagen”. In: “20. Tonmeistertagung 1988. Tagungsbericht”, pp. 238-245

[11] IRT: “Binaural Room Synthesis BRS”. www.irt.de/de/produkte/produktion/binaural-room-synthesis-brs.html

[12] Fey, F.: “Ohrenbetörend. Smyth Research SVS Realiser A-8”. In: “Studio Magazin 12/2009”, pp. 24-34

[13] Riekehof-Böhmer, H., Wittek, H., and Mores, R.: “Voraussage der wahrgenommenen räumlichen Breite einer beliebigen stereofonen Mikrofonanordnung”, In “26. Tonmeistertagung 2010”

[14] Griesinger, D.: “General overview of spatial impression, envelopment, localization, and externalization”. In “Proceedings of the 15th International AES Conference”, Copenhagen, 1998, pp. 136-149.

[15] Wittek, H.: “Image Assistant” JAVA applet available on www.hauptmikrofon.de (last visited 2/23/2011)

[16] Barbour, J.: “Elevation Perception: Phantom Images in the Vertical Hemi-sphere”. In: “Proceedings of the 24th AES Conference on Multichannel Audio”, The New Reality, June 2003

[17] Wittek, H., Haut, C., and Keinath, D.: “Doppel-MS – eine Surround-Aufnahmetechnik unter der Lupe”. In: “24. Tonmeistertagung 2006”, Leipzig