Proceedings of the

International Conference on Spatial Audio

ICSA 2014

in

Erlangen, Germany

Conference Chairs:

Emanuël Habets
Jürgen Herre
Malte Kob
Bernhard Grill
Andreas Silzle
Günther Theile

ISBN 978-3-98 12830-4-4

Room signals – properties and influence on Auro 3D recordings

L. Nipkow

Silent Work LLC, Switzerland, Email: lasse.nipkow@silentwork.com

Abstract

The relevance of room signals in a recording, especially a classical recording, for 3D reproducing systems like Auro 3D is essential: The listener shall have a spatial impression of being in the recording room, and generally, to perceive a natural sounding environment. For music, usually, direct sound comes from the front loudspeakers only and room sound from everywhere. Therefore, a listener hears many more details of the room sound compared to stereo. Relevant parameters for controlling the spatial impression of the recording are:

- Energy spreading of the room signal between the involved loudspeakers.
- Proportion of direct sound in the room microphone signals.
- Sound Content Correlation – common features of uncorrelated signals.

The sense of depth of room sound reproduced by Auro 3D is perceived in a different way than room sound reproduced by a stereo setup. When the proportion of direct sound is decreased in a stereo recording, sound sources are reproduced with more depth. In an Auro 3D recording, the same change will lead to a more diffuse impression of the sound sources.

1. Introduction

Classical music performed in concert halls or churches can be, acoustically speaking, divided into two components: direct sound from the instruments and room sound arising from reflections in the concert hall. From the view of good listening seats, the instruments of the orchestra are horizontally in front of the listeners, while room sound arrives from all directions. A natural sounding recording should therefore be mixed according to these insights.

Direct sound will be reproduced by Auro 3D from the front loudspeakers L/C/R and all speakers are used for the reproduction of room sound. In this case, a listener has a spatial impression of being in the recording room (see **Figure 1**) [1].

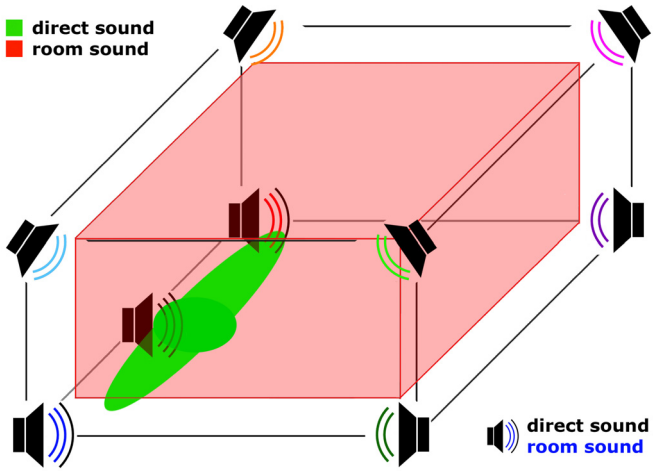

Figure 1: If direct sound is reproduced from the front loudspeakers L/C/R and all loudspeakers reproduce uncorrelated room sound, a spatial impression of being in the recording room results.

2. Directional perception of room sound

2.1. Differences between room sound reproduction of 2 channel stereo and Auro 3D

A traditional stereo system is consisting of only two loudspeakers, which means the entire information of a concert, i.e. imaging of the instruments and the envelopment of the recording room, is contained in two channels only. The consequence is that direct and room sound are reproduced horizontally frontally only within a ±30° sector.

In Auro 3D, room sound can be reproduced from different directions than direct sound: from the height and surround loudspeaker positions. That means the room sound without direct sound is not only reproduced in the horizontal frontal plane, but also from the rear (like 5.1) and additionally front above. The question is, however, how do we perceive the direction of reproduced room sound? Is it in a way comparable to phantom sources or different?

2.2. Directional perception of room sound

The room sound component of the loudspeaker signals, and therefore the room microphone signals, have to be uncorrelated between each other in order to get a natural character of the room sound [2]. Only in this case do no phantom sources between loudspeakers occur.

If uncorrelated room sound of two or more loudspeakers is reproduced at about the same level by each loudspeaker, the positions of each speaker can no longer be perceived. This is similar to the effect of a phantom sound source between two loudspeakers, which means that the positions of the used loudspeakers are no longer perceived. But since the room sound is uncorrelated between the loudspeakers, it appears as a blurred band of sound between the loudspeakers. If all loudspeakers of an Auro 3D setup reproduce uncorrelated room sound at about the same level, a listener has a spatial impression of being in the recording room [2].

2.3. Selective listening of direct sound and room sound

Room sound often is masked strongly by direct sound in 2 channel stereo recordings. Even if a comparable level of the room sound and direct sound is given, the character of the room sound is mostly only perceived at when direct sound level abruptly decreases, e.g. at the end of a movement. Due to the so-called binaural masking level difference, room sound is masked significantly less when reproduced from different directions than direct sound [3]. In this case, a listener will perceive the character of the room sound much better.

The consequence of this is, that the characteristics and quality of the room sound, detached from the direct sound, can be easily judge by a listener. From an aesthetic perspective, a first class concert hall could be reproduced in a much more impressive way. On the other hand, the weakness of a lower quality recording room is easier to hear. The same is true for the use of digital reverberators: Artificial reverberation can be identified much easier in an Auro 3D recording than in a 2 channel stereo recording.

3. Analysis of room signals and their perception

In spatial hearing, there are three components according to specific temporal patterns: direct sound, early reflections, and reverberation (late reflections). These three components also occur in recordings and must follow the same rules as in a room, to ensure a high quality:

- Direct sound from a sound source determines what instrument a listener hears and where it is localized. In stereo recordings direct sound appears either as one or more phantom sources or as a real sound source from just one speaker. Direct sound is primarily correlated between loudspeakers.
- Early reflections can be thought of as individual reflections from walls, floor, and ceiling. Especially from impulsive sources, like percussion or acoustic guitar, they can be identified in wide rooms as separate components of the acoustic signal. They give the listener a sense of room size and involvement of the sound source in the room (sense of depth). The lateral reflections consist of uncorrelated proportions while reflections from the ceiling and floor consist of correlated proportions [4].
- Reverberation in a concert hall is the result of a high number of reflections of direct sound. Reverberation sounds like a sound carpet that seems to come simultaneously from everywhere. The reverberation level depends on the characteristics of the recording room. At the critical distance, the early reflections and reverberation level is the same as the direct sound. In concert halls this distance is only a few meters away from the sound source [4]. Therefore, reverberation represents for listening of classical music a dominant element and leads to envelopment and gives the listeners a spatial impression of being in the recording room. Ideally, room sound is as far as possible uncorrelated between loudspeakers [5].

Some other signals have similar characteristics as early reflections and reverberation. The following chapter discusses such signals.

3.1. Sound Content Correlation – common features of uncorrelated signals

Not in any case uncorrelated signals reproduced by multiple loudspeakers leads to a sense of a spatial impression. If several voices are used as speaker signals, each of them is perceived as a single, localized sound source from the respective loudspeaker. This means, therefore, uncorrelated signals, which do not have a similar sound (when listening to them separately) don't lead to a sense of spatial impression.

Room sound has its origin in sound sources that are located in the concert hall. The sound reproduced by the sources is distributed as regularly as a water-soluble dye that has been dropped in a jar of water. Experimental recordings of the author in first class concert halls have revealed that the room sound with a very low proportion of direct sound is perceived as nearly identically at different points, although the correlation between all those signals is about zero. Therefore, the routing of an Auro 3D playback system can be changed during playback of room sound without any significant noticable changes in the room impression. In other words: Different room microphone signals ideally sound the same, but have a correlation about zero. This property is denoted Sound Content Correlation in this paper.

A similar phenomenon arises if instruments like guitar, piano, or string instruments are recorded by a small AB spot microphone system at a very small distance to the resonating body. It seems as if the resonating body would sound as an own room. Especially, if recording double bass in this way the instrument is nearly not locatalizable in the stereo image, but seems to float indefinable between left and right (see **Figure 2**). The complex dispersion behaviour is responsible for this phenomenon. [6]. Pad sounds of synthesizers have a similar character.

Figure 2: Recording of a double bass by a small AB spot microphone system: The resonating body seems to sound as an own room.

The previous sections might motivate to set up the following rule: *if two signals have a similar sound and are uncorrelated, then they give an impression of a stereophonic spatiality and lead to envelopment.* However, experience shows that more attributes have to be considered. In some cases, this rule doesn't apply: If one person speaks two different texts and they will be reproduced at the same time by two speakers, two sources are perceived at the loudspeaker positions. Therefore, it is not sufficient enough for create a sense of spaciousness only by similarity of the sound. The presence of transients in the signals and the similarity of the temporal structure of the signals will also take influence.

3.1.1. Similar temporal structure

To prevent hearing two separate sound sources from left and right transient impulsive signals, such as language, in terms of their temporal structure have to be about the same. This is given when a speaker will be recorded speaking the same text twice with the same speed and the two sections will be synchronized and reproduced on a left and right loudspeaker discretely afterwards. By replaying the same text synchronized the waveforms and envelopes of left and right channel are not identical. Primarily time differences occur in the fine structure, which leads to uncorrelated signals [7] [8].

3.1.2. Transient-free room signals

Particularly direct sound of percussive instruments such as drums, guitar, or harpsichord contains fast transients. This corresponds to the attack of those instruments and is a very important factor of their sound. Transients help the listener to localize precisely [9]. In contrast, however, reverberation doesn't contain transients but sonically smoothed signals that make localization of sound sources more difficult or impossible. Reverberation and similarly shaped uncorrelated signals contain little phase relationships that aid localization of sources. The lack of a clearly defined temporal contour leads to a variance in the time axis between the individual room sound signals: Room signals of instruments with low-impulse behaviour such as slowly bowed strings or slowly played church organs could be varied in time about 0 to 250ms delay or more compared to the direct sound without getting disturbing echoes. When instruments play staccatissimo and fast transients occur, the range of delay until echoes are disturbing is reduced significantly. There are some possibilities to optimize room microphone setup to avoid fast transients from direct sound. One method is the use of a towel between the room microphone and the sound source (see **Figure 3**). Another method suppresses direct sound by the use of cardioid microphone aligned with its null angle to the sound source (see **Figure 3**).

Figure 3: Use of a towel between the room microphone and the sound source or a cardioid microphone aligned with its null angle towards the sound source avoids fast transients from direct sound.

There are further signals with similar properties as reverb, but they do not have any relation to a direct sound. Uncorrelated noise is an example of such signal: It is possible to delay one channel by any value without making a noticeable change. The same is true in certain conditions for ambient signals, i.e. soundscapes with continuously running sounds that contain neither transient nor striking events, such as an isolated horn of a car or bells of a church. Thus, by channel duplicating far away sounding city noise from a stereo recording and mutual time delay up to 1s an 8.0 Auro 3D atmosphere could be created (see **Figure 4**). The result is a spatially impressive sounding atmosphere.

Figure 4: An Auro 3D 8.0 atmosphere sound can be generated from mono or stereo far-away sounding city noise by channel duplicating and mutual time delays of up to 1s.

3.1.3. Short time fast transient direct sound

Another group of signals that leads to spatial impression or atmosphere consists of a variety of short time fast transient direct sounds that arrive from different directions to the listener, like individual early reflections. Due to the amount and the short duration of the individual sounds, no concrete direction of the direct sound can be perceived. One single impulse sound, such as a crackling or burst of an eggshell, could be used to generate an 8 channel atmosphere sound with a DAW. It must be noted that the individual sounds should only be played from time to time simultaneously through more than two loudspeakers. Otherwise, disturbing effects caused by phantom sound source formation could occur.

3.1.4. Conclusions

As discussed in the preceding sections, spatial impression and atmosphere can be created in various ways. The methods must respect the following rules: Signal between loudspeakers shall be uncorrelated but feature Sound Content Correlation. The spaciousness is perceived as more impressive the more dimensions are used during playback [1].

4. Sense of depth in stereo and Auro 3D

4.1. Sense of depth in recordings

The sense of depth in recordings primarily depends on the ratio of level between direct sound and room sound. If the level of room sound is low, sound sources seem to be near. If the level of room sound is high, sound sources seem to be farther away. The room microphone signals consist of a combination of early reflections and reverberation. The early reflections are responsible for distance perception [10]. The sense of depth is especially pronounced when the nearest sound source can be heard from a short distance. Otherwise, the differences of the distance are weak.

The positions of phantom sound sources in the stereo image are perceived directly – level differences can't be perceived as such. In contrast, sense of depth is determined on the basis of characteristics of the sound. In other words, it is possible up to a certain degree to hear and to describe the individual cues, which are responsible for the perception of distance. The perception of distance has various causes. The most obvious difference between close and distant sound sources is the specific sound of a distance of the sound source: Sound sources recorded with spot microphones contain direct sound with a high proportion of transients. These signals sound very clear, have a distinct localization and low sound coloration. If the microphones are very close, sound sources also have the tendency to use a lot of space in the stereo image. This is due in part to the effect of the instruments with resonant body described in the chapter Sound Content Correlation. This phenomenon decreases as the distance increases from the sound source. In other cases, high proportion of low frequencies compared to more distant sound sources leads to an increased spreading of close sounding phantom sources. Furthermore, the spectral content of the direct sound changes, partly due to the dissipation with increasing distance from the sound source [10]. From a geometrical point of view, it can be concluded: Instruments perceived extended such as string instruments and physical extended instruments such as piano and drums appear less extended as the distance increases from a main microphone.

4.1.1. Sense of depth in stereo

As said before, the sense of depth in stereo recordings is primarily dependent of the ratio of level between direct sound and room sound. If the influences of the reproduction room may be neglected, the sense of depth will be influenced by the specific used microphone setups or during the mixing process. In the case of a mixing process, the following thoughts could be taken: The mix consists of room sound and a spot microphone signal. If the direct sound is muted, the instrument is perceived as diffuse only, at a large distance. When direct sound is increasingly added to the mix, the instrument appears from its location in the stereo image. When the level of the direct sound is increasing, the perceived distance from the instrument is reduced. The reproduction room can't influence the ratio, because it is already adjusted electrically or by software in a DAW. Thus, the distance information of the recording is independent of the position of the listener in the reproduction room (see **Figure 5**). Therefore, the audience can stay closer or further away from the stereo speakers and hear the sound sources always at the same simulated distance.

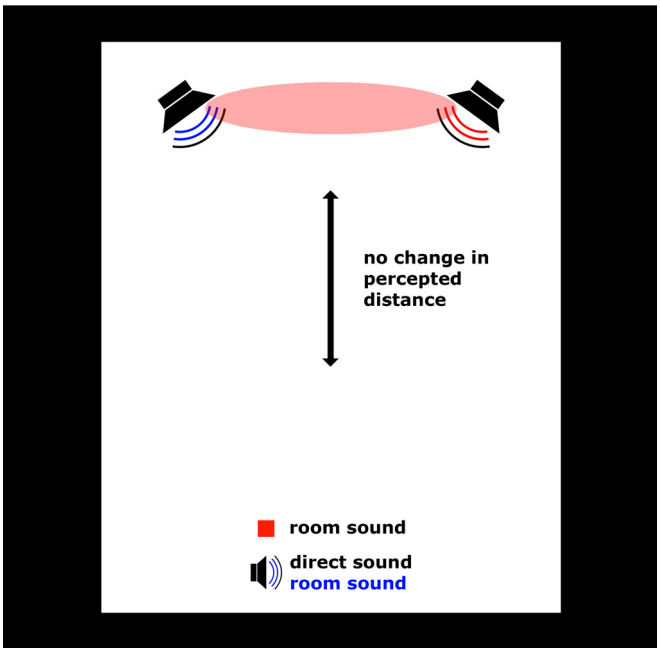

Figure 5: The distance information in stereo recordings is independent of the position of the listener in the reproduction room.

4.1.2. Sense of depth in Auro 3D

During Auro 3D playback, the room sound of the recording is reproduced both from front and behind. As already described earlier, room microphone signals contain reverberation and early reflections. The level, spectrum, and timing of the individual reflections and recorded direct sound are dependent on both, the position in the recording room as well as the directional pattern and orientation of the room microphone relative to the body of sound. Direct sound picked up from room microphones is perceived in the mix, in the presence of spot and main microphone signals, as early reflection, because room microphones have a larger distance from the sound source as the other microphones and thus direct sound in the room signal arrives later in the mix [1]. Room signals from large distant microphones, for music played by instruments with fast transients such as percussion or choir, are often temporally aligned with the main microphone, to avoid disturbing echoes. This inevitably means that the direct sound and the subsequent early reflections in the room microphones occur almost simultaneously as the main microphone signal and are reproduced from each speaker in the Auro 3D setup.

4.2. Changing distance perception within the listening area

When the surround speakers reproduce room sound, the ratio of direct sound to room sound shifts depending on the position of the listener. Accordingly, when a listener is placed close to the front speakers the body of sound is perceived as more direct than at a position closer to the surround loudspeakers (see **Figure 6**). This corresponds on one hand mostly to natural hearing in concert halls. On the other hand, the sound engineer is requested to optimize the sound aesthetic with this new parameter of distance perception within the listening area.

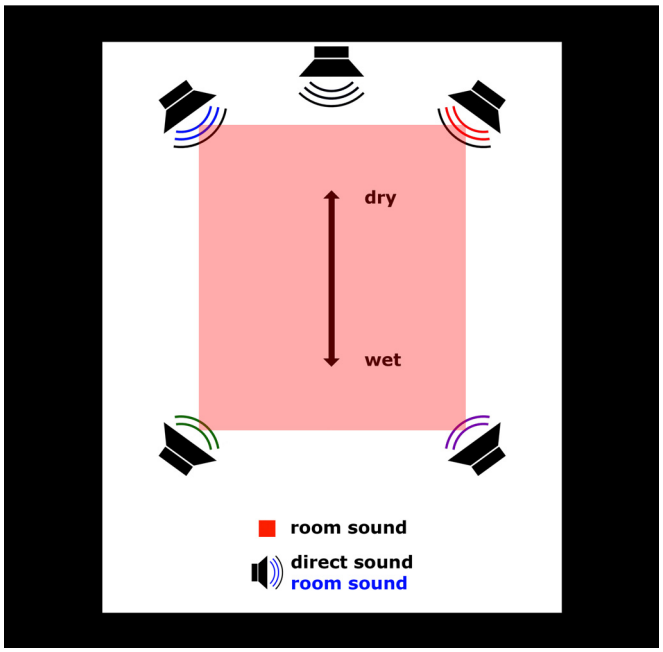

Figure 6: The distance information in Auro 3D recordings is dependent of the position of the listener in the reproduction room.

4.3. Distortion of the sense of depth

The sense of depth in Auro 3D is fundamentally different than in 2 channel stereo (see **Figure 7**). Similarly as above, the following thoughts could be taken: The mix consists again of room sound and direct sound from spot microphones. If the direct sound is muted, the listener is surrounded by room sound and is hearing the instrument diffusely from everywhere. When direct sound is increasingly added to the mix, the instrument appears from its intended location. When the level of the direct sound is increasing, the clearness of the instruments' sound is increasing, as well as loudness. The simulated distance perception is weak only, because the early reflections are reproduced from different directions. If the main microphone contains early reflections in a way that the sound of the body of sound is strongly colored, the additional reflections from the room lead in many cases to muddiness. If the room signals, reproduced from the surround speakers, contain direct sound components of one or more instruments as the front speakers, the affected instruments may even unintentionally move backwards, causing unnatural effects (see **Figure 8**).

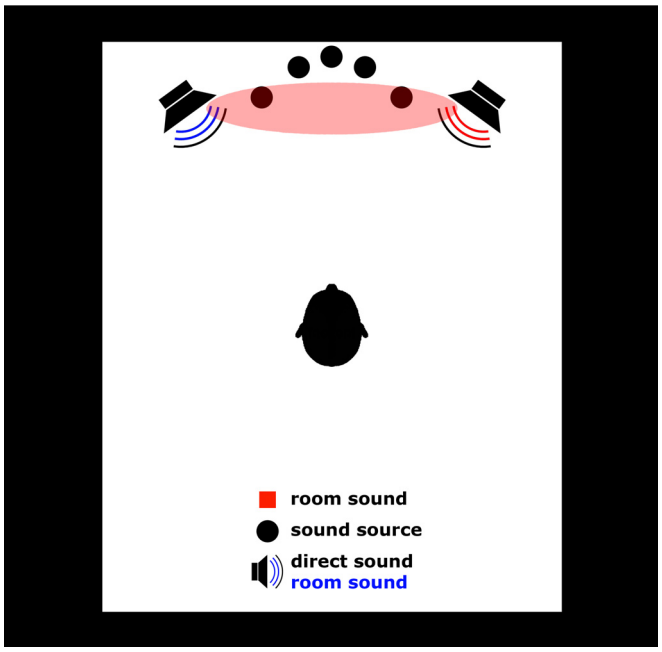

Figure 7: The sense of depth in stereo. The sound sources are perceived behind the line joining the two loudspeakers.

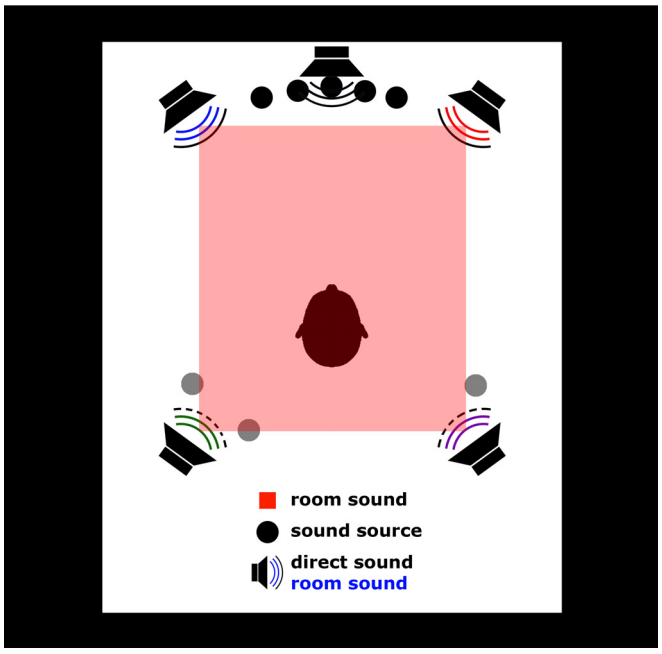

Figure 8: The simulated distance perception is weak only, because the early reflections are reproduced from different directions. Room signals with direct sound could lead to backwards perception.

4.4. Aesthetic corollary in the mix

Auro 3D recordings of classical music need more room sound in total than the corresponding stereo version, to exploit the advantage of nice room sound and envelopment rendering. The level of room sound signals has to be balanced between the used loudspeakers, except the center loudspeaker, to ensure perceived connection between front and rear loudspeakers [1]. Furthermore, those room microphone signals with the lowest proportion of direct sound should be taken for the upper surround speakers to avoid irritation of the listener by reason of hearing direct sound coming from that direction [1].

As described in the Section 3.1 the distance impression of the frontal sound sources varies depending on the distance the listener has to the front loudspeakers. In order to ensure a similar impression of distance as the corresponding stereo mix at the place of the sweet spot, as little as possible early reflections should be reproduced from the front speakers. In other words: If all other speakers are switched off, the mix has to sound more direct than in stereo. Therefore, it's important to use a microphone technique for the body of sound to avoid early reflections as much as possible. If spot microphones are used, they should be placed as close as possible to the individual instruments to keep the crosstalk with other instruments low – the crosstalk is perceived as early reflections and leads to an increased distance perception of the corresponding sound source. If a body of sound is recorded with a main microphone system, high directive microphones should be used, like good sounding hypercardioid microphones or high quality shotgun microphones with a sphere microphone as an acoustic separator [1] [2]. In return, however, the upper frontal loudspeakers should reproduce more room sound, to bring the level of room sound signals into a good balance between the used loudspeakers in front and rear. This could be done with a combination of high-frequency direct sound from cymbals or brilliant string sounds of an orchestra and additional early reflections from the body of sound. The result is a very natural sounding and amazing listening experience [1].

One challenge is the balance between the individual instruments of an ensemble. While direct sound is captured by spot microphones and can be selectively precisely inserted into a mix, the balancing of the individual instrument levels in the room sound has narrow limits: If an instrument is played at a louder level than others, it is louder in the recording room too, practically independent of the position of the room microphones. In that case, the instrument is not only louder, but also at the same time is perceived with larger distance, because the proportion of the room sound of that instrument is higher than of other instruments. This can't be corrected by spot microphones without affecting the balance of the instruments. Therefore, at this point, the recording director is requested to be attentive during the recording session and take corrective action if necessary.

4.5. Conclusions

Room sound gets a much higher importance in Auro 3D than in 2 channel stereo. On one hand, spatial sound experiences can be created in Auro 3D, which can not be realized in any way in 2 channel stereo. On the other hand, the demands on the sound engineers who make such recordings are multiplied. It requires not only knowledge in the field of psychoacoustics, but also a lot of recording experience to fully explore the potential offered by Auro 3D.

5. References

[1] Nipkow, L.: Eigenschaften von mikrofonierten Raumsignalen bei 3D Audio / Auro 3D. 27. Tonmeistertagung, Köln (2012), Proceedings (ISBN 978-3-9812830-3-7), 757-770

[2] Nipkow, L.: Angewandte Psychoakustik bei 3D Surround Sound Aufnahmen. 26. Tonmeistertagung, Leipzig (2010), Proceedings (ISBN 978-3-9812830-0-6), 786-795

[3] Blauert, J.: Räumliches Hören. S. Hirzel Verlag, Stuttgart, 1974, ISBN 3-7776-0250-7

[4] Meyer, J.: Akustik und musikalische Aufführungspraxis. Verlag Erwin Bochinsky (1999), ISBN 3-923639-01-5

[5] Griesinger, D.: The Theory and Practice of Perceptual Modeling - How to use Electronic Reverberation to Add Depth and Envelopment Without Reducing Clarity. 21. Tonmeistertagung, Hannover 2000, Tagungsbericht Verlag K.G. Saur, München (2001), S.766–795

[6] Cremer, L.: Physik der Geige: Der Streichvorgang, der Instrumentenkörper, der abgestrahlte Schall. S. Hirzel Verlag Stuttgart (1981), ISBN 3-7776-0372-4

[7] Helmholtz, H. v.: Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik. Esther von Krosigk (Hrsg.), Edition Classic VDM Verlag Dr. Müller, (2007), ISBN 978-3-8364-0606-2

[8] Terhardt, E.: Akustische Kommunikation, 1998: Grundlagen mit Hörbeispielen, Springer, ISBN 978-3-642-58733-7

[9] Friesecke, A.: Die Audio-Enzyklopädie: Ein Nachschlagewerk für Tontechniker. K. G. Saur Verlag (2007), ISBN 978-3598117749

[10] Potratz, U.: Untersuchung der Gestaltungsmöglichkeiten früher Reflexionen mit Hilfe eines raumakustischen Modells. Diplomarbeit, Erich-Thienhaus-Institut der Hochschule für Musik Detmold, Studiengang Musikübertragung, Erstgutachter: Prof. Thomas Görne (2005).