Spatial perception and information of recordings

Holoplot
Author: Helmut Oellers , Text: Sebastian Böldt *

Ringbahnstraße 12, 12099 Berlin
info@holoplot.com

January 11, 2018

Abstract

This white paper investigates the spatial perception and information of recordings. Without an established status quo on existing channel bounded speaker systems and their spatial representation, it is challenging to reach a uniformed conclusion on a suitable playback system. The goal is to summarize the existing perspectives and methods, examine the most important acoustical principles, and to find an approach on new technologies based on wave field synthesis.

I Introduction

There are strong opinions about solutions for the perfect playback chain in high-end audio setups and large amounts of money are being spent to ensure the system delivers a faultless reproduction of the recording. In these cases, a linear frequency response is being reached as well as tonal accuracy, yet the important spatial information and an ability to precisely locate the source in the sound is lost. In order to understand this in more details, an analysis of at how sounds are being recorded and how humans perceive spatial information is required. If sound is produced in a recording room, one would hear the direct sound and additionally the reflections coming from all directions of the room. Due to a combination of those reflections and direct sound waves, interaural time differences (ITD) occur between the ears which are used in order to localize a sound source. Furthermore, the reflections and shadowing at the listener's head as well as resonances inside their ears, ear canal, the shape of nasal and oral cavities lead to level differences in both of the eardrums. These effects create an individual frequency perception which characterizes how one receives sounds from a point in space – also known as head-related transfer function (HRTF). In combination with the ITD’s, the HRTF provides all required information for allocating sound sources in the

*Thanks to Helmut Oellers for providing a lot of profound research knowledge

space - in horizontal and vertical angulation. Thereby waves between a frequency of 100Hz to 3,6kHz are just being analyzed. Waves below that range differ too little in phase and making it nearly impossible to evaluate. Testing alternative theories, it could be suggested that a setup with two microphones creates a recording that mirrors the exact hearing impression. But since multiple microphone characteristics always differ in their directionality, this assumption does not hold true. Microphones with spherical characteristics are nearly independent from where the sound is coming from. Other microphones with more directional characteristics totally differ from the complex, human perception and filtering of the head-torso-ear system, hence are not really able to receive sounds as the eardrums do. Another issue is, that during the recording the spatial information of the studio room are already reduced to the horizontal axis. Furthermore, recordings of the first reflections coming from the side, above or even from the back do not really match with the one that we perceive. Thus, they do not represent the ITD’s which would have been created as if the sound is being heard “live”. However, since these first reflections have a strong impact on our subjective listening perception, it is important to reproduce them correctly. They improve the perceived loudness, the speech intelligibility and more important they carry information about the size of the room and distance to the sound source. The most common and conventional way to achieve this is by using reverbs, where the recorded sound is fist positioned via a panpot (on the mixing desk) in the panorama, and spatial information of an artificial room are added afterwards. With this setup, the direct wave and first reflections are coming from the same direction (the loud speaker itself) and in order to keep the spatiality, the reverb is applied. By creating an artificial reverberation, the spatiality of the room is added and „inserted back” into the actual recording. Although this solution is considered "good enough", in reality a reverb provides only the information about the room size, but not about where the source is coming from - this information is lost.

II Channel-bound speaker systems and phantom sources

Every speaker system which is based on a specific number of speakers (e.g. stereo or surround) creates a phantom source, that is perceived to be located somewhere in between the speakers. Yet this impression is a psychoacoustical effect - it is not real. One cannot get closer to the "source", as this source is ever-moving upon the listeners changing their position. It should be mentioned that one phantom source is created by two real sources (real sources on left/right, creating one phantom source in the center). In order to locate a sound from one direction, there must be a significant level difference between the different sources. This ratio is described by the Interaural Cross Correlation Coefficient (IACC). Giving an example, if two signals are identical and perceived in the middle from the front/back, the coefficient is +1, and lowering the values if the perceived source is off to one side. In case the signals are not equal at all, the value is 0. In free field conditions, the IACC reaches its maximum at around 0,3 where the sound waves impinge in an angle of 55 degrees (left or right seen from the median axis). Rooms with good acoustical properties are characterized by first reflections coming from that angle. If the actual (visually located) source is on the other side then the perceived reflection, a so called „acoustical attraction“ is created. In a usual setup, the speaker is positioned on a listening angle of 30 degrees, already brining the first limitations. The more the source is located in the middle, the less „attractive“ the hearing perception is. Thus, speaker setups usually never reach an IACC value below 0.6. This is also the reason why tools were introduced in order to spread the stereo image and create a „wider“ hearing impression, a downside of this being a manipulation of phase and creation of unwanted reflections in the playback room. Besides a less impelling hearing impression, there is also the problem of estimating the distance to the source. Since every phantom source is located on one axis between the speakers there is no real depth positioning. For example - if two musicians are playing in front of the listener, they are perceived as coming from the same direction and position. When moving to one side, it’s possible to hear that one musician is standing behind the other (one is perceived more on the left, one more on the right, when looking from the side). Reproducing this scenario with phantom sources coming from a speaker is difficult or even impossible – the perception of the acoustical position of the musicians will remain the same, as will the starting point and the depth. This scenario also proves that phantom sources are always located between speakers, and never in front or behind. Furthermore, they cannot create a realistic hearing impression reproducing the original distance of the sound source since the source never moves and the volume of the sound is used in order to estimate distances. Another important property is the ratio of direct waves and diffuse field that describes the amount of direct wave propagation and diffuse field that is perceived by the listener. But this ratio cannot be higher as produced by the speaker itself in the playback room - where the diffuse field amount is quite high and just a few direct waves reaching the listener. Meanwhile, in real free field conditions, a close sound source nearly transmits direct waves only (the diffuse field amount is not really perceivable). Due to the fact that most of the time the distance of the listener position to the loudspeaker (e.g. in a living room) is bigger than the reverberation radius, the diffuse field amount dominates – and just a few direct waves are reaching the listener. The reflections coming from the playback room, and characterizing the diffuse field, are going over and covering the direct waves. Thus, a sound source which is aimed to appear closer will be always located further away than it is - "behind the speaker". A lot of times this effect is being underestimated but it is worth paying attention to, since proximity is an important part of an emotional hearing.

III Interferences with the Playback Room

When examining the interferences of the sound waves with the playback room, "notch filter effects" occur - cutting or boosting certain frequencies up to 20dB, due to overlapping of reflections and direct waves, as mentioned in the chapter before. Naturally, these notch filter effects also occur in the recording room, where overlapping or cutting of frequencies creates a specific sound which characterizes the room itself, in a helpful mode. The human ear is trained to recognize patterns and compare with memories, making the estimation of the recording room size possible. Nevertheless, the playback room sticks a wrong stamp on actual acoustics of the recording, as the human hearing perception is used to imperfect acoustics of the living room and these additional effects are being depressed. A conventional way to avoid these effects, is achieved by acoustical improvements such as acoustic panels, base traps and other installations, in order to minimize possible reflections and resonances, as well as by improving the directivity of the speaker. Having said that, it is important not to create a totally "dry" hearing impression since some reflections can also be helpful – especially when the room size of the recording room and playback room do not differ that much. Here a certain interaction and interference of the playback room and the recording can be of help. A good example was shown by Acoustic Research at Carnegie Hall in London end of 1980 where loud speakers where put on a stage and dry recordings of singers where played back. The experiment showed that it was very hard for a listener to distinguish between the reproduction or an actual singer being on stage. Apparently, the distribution of reflections and interferences with the room are more important than an accurate frequency response of the playback chain, measured up to the last two decibels. Summing up these findings, it becomes clear that in order to achieve an authentic reproduction of the signal, acoustical improvements of the playback room help to avoid unwanted reflections and higher the ratio between direct waves and diffuse field. The important part is, however, that first reflections have to be created from the correct directions in order to re-create the actual setup in the recording room. As we have seen beforehand, this cannot be solved with the conventional model of phantom sources. A new approach is needed in order to create such virtual sources and to re-create a physical and sonic sound field.

IV Spatial representation in channels

Since Alan Blumlein invented the concept of stereo recording, all the acoustical information regarding room size and characteristics are transmitted, as a result of signal differences between left and right channel. With the acceleration of modern technology, high-quality recordings and mastering technologies became more prevalent and the listener is getting accustomed to this subjective and appealing hearing impression – where every instrument in the mix is very clear and easy to identify. However, it has to be noted that this experience often differs to the sound of a live performance. As a conclusion, it can be said that the aim became less to preserve a certain recording environment, but rather to create an artificial, spatial hearing impression. As described in the previous section, this exact re-production remains one of the biggest challenges. Often, a sound of an instrument is represented as a mono sound, positioned in the panorama through panning (left/right) and reverb is added to the dry recorded signal in order to add spatial information of the desired room. As long as the aim of a production is to produce a well-sounding (stereo) panorama by creating phantom sources, this is a reasonable approach. However, if the aim is to (re)produce the clearance of a voice in a specific acoustical environment — the sum of a bee directly in front of listener's nose or an emotional effect of a sound field inside a well-known concert hall — this channel-bounded approach that creates phantom sources, is not sufficient anymore. Additionally, trying to compensate for the aforementioned issues with more microphones, does not solve the problem of being unable to capture the exact sonic information. The sonic sound field is not fully reproducible just by the signal differences between two points and in order to reach that, the original sonic structure of all wave fronts has to be recovered/ reproduced. This also includes a change of the listening position, where it alters the impression like in the recording room — requiring a new approach for the recording and playback.

V Mirror sound sources

The idea is to assume that the sound source itself (instrument or singer) doesn’t have an own sound field. Certainly, the radiate characteristics of each instrument differs, but the spatial information is restored in reflections of the source signal in the recording room. These reflections are created by mirror sound sources. Their positions are defined by the reflection surfaces and the position of the primary sound source – independently from the listening position. For example, for reflections coming from the wall, the apparent origin of these reflections seems as it would be outside of the recording room. Going from there, each first (early) reflection is the origin of a further, second (later) reflections. Each time, every reflection is just sending a part of the pure and original signal, retrieved by the primary source, according to the acoustical properties of the room. Furthermore, the signal sent from the primary source, is being damped by reflections (e.g. from a wall) and the pressure level decreases, with the addition of the high frequencies being filtered due to the air resistance. However, providing that all reflections of the recording room are described as well as the positioning and direction of the source, positioning and pressure levels of the mirror sound sources can be retrieved. With this method, the sonic sound field of specific room can be described and recovered in an anechoic environment — just a pure, mono recording of the signal(s) that is going to be played back is needed. Every mirror sound source is then reproduced by positioning a speaker at its origin. The play-back signal is specified by the primary source or the previous reflection (including filtering, damping, etc.). With a bigger distance, the number of speakers is increasing because they represent reflections of a higher order. This would imply an infinite number of speakers, but starting at one point the sound would become damped and filtered so much, that it would not be audible any more. Although these speakers can be neglected, it is nearly impossible to realize such a setup because the number of required speakers is still too high, still meaning just one position of a source in the recording room would be covered and reproduced. If the source was moving during the recording, this movement would not be captured – the whole setup would have to be changed with every action. This concept works only if the listener is the one moving, creating the same impression as if the source was moving in the recording room. The signal is getting louder and dryer when coming closer to the source. When moving further away, the ratio between reflections and direct wave changes immediately. This is why the listener can already estimate the distance of the primary source much better. The same Doppler effects also occurs in the recording room. This illustrates that from a theoretical perspective, it’s possible to create a virtual copy of the sound field. In order to realize such a reproduction, the next step is to create solutions on the technology side. Wave field synthesis is not only a complex approach, but also a possibility of creating a virtual sound field copy of a specific room.

In conclusion, this paper examined existing principles and acoustic conditions of conventional playback systems. The most important basics of physical circumstances and the human perception were pointed out and led to the overall conclusion that a correct reconstruction of the recording needs to include not only volumes, and size of the room. It is very important to also restore every single positioning of an instrument, first reflections and creating a feeling for proximity as it is perceived live. In order to achieve that, first approaches were suggested and the wave field technology was introduced as an encouraging method.