The Basics of High Fidelity
Part 2: Reproduction Philosophy "Here and Now" versus "There and Then"

Part 1 dealt with the problem of transparency, our goal was to store and transport an audio signal without introducing any audible degradation. In part 2 we will deal with the problem of converting the audio signal into sound.

We have already been able to deal with a number of audio myths using the basic idea of "transparency", but we have made little progress in answering the question how the sound that finally reaches our ears should be constructed. We may be able to design audio equipment on the basis of perceptual transparency but the problem of transduction towards sound is difficult to solve on the basis of the idea of transparency. What is a transparent loudspeaker? We don't know, a loudspeaker may have a clear input terminal where we hook up our perfect transparent amplifier but we can define infinite many outputs, the loudspeaker creates a sound field that is different in each measurement location and that is also dependent on the reproduction room. This makes it nearly impossible to characterize a loudspeaker in a technical, perceptual relevant, manner. Of course, one can measure the on axis response in a anechoic room, the power response in a standardized living room, the off axis responses, etc., but how should we interpret all those characterizations?

And how about a headphone, it couples the sound directly into our ears and by using recordings made inside our own ears, we may be able to bring the problem back to the basics of transparency. In principle we can make a recording of the sound pressure at the left and right eardrum and reproduce it exactly using a correctly equalized headphone, creating a transparent link. Unfortunately this is not as simple as it sounds, we all have different Head Related Transfer Functions (HRTF's) that change when we move our head. We can try to use individualized HRTF's in combination with a head tracker, but what will we do with all the wonderful music that has been recorded directly in the electric domain? Will we just play it "dead", with an artificial spatial position, doesn't sound right to me. And even if we play acoustically recorded music with a head tracked headphone with optimized equalization we will miss part of the low frequency sensation which is for a large part produced by bone conduction.

So let's dive deeper into the underlying philosophical problem, what sound field are we trying to produce or reproduce? There are in fact two different approaches towards the reproduction of sound. In the headphone example we are trying to create the illusion "there and then" (virtual reality). This represents the situation that when you close your eyes you will experience the illusion that you are present at the place where the recording was made, e.g. a concert hall, i.e. we are striving towards a virtual reality illusion. However there is another ideal that we can pursue, the "here and now" (augmented reality), where we have the illusion that the sound source is present in the room where we are sitting. Unfortunately when you try to realize the two different idealizations "here and now" and "there and then" you will find that they lead to incompatible requirements on the recording and reproduction technology.

If you are pursuing the “here and now” ideal you have to make a recording in a room without any reflections, an anechoic room. And when you reproduce the signal you have to use a loudspeaker set up that has the same radiation pattern(s) as the original source(s). Let’s go for a simple example, you want to perfectly reproduce a single voice. You make a recording in an anechoic room and reproduce it over a single loudspeaker that has the same radiation pattern as the voice (see Figure 1).

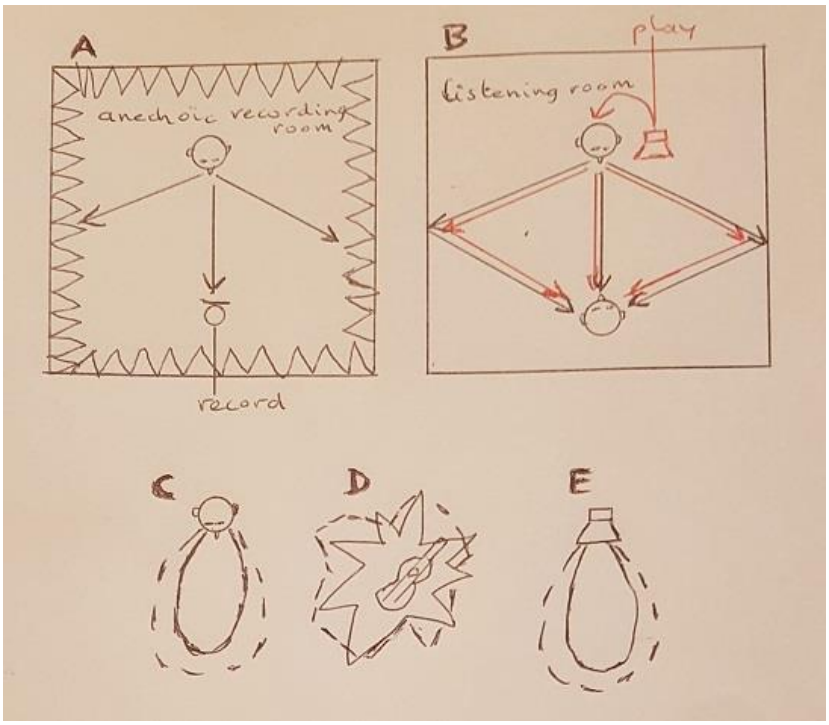

Figure 1. Transparent recording/play back of a single acoustic source. A mono recording of a voice in an anechoic room (A) can be played back transparently in any room over a loudspeaker that has the same directional properties as the voice (B). While a loudspeaker and a voice can have similar directional properties a musical instrument almost always will have wild varying directional properties. Figures C, D and E give the directional pattern of a mid (dashed line) and a high frequency (solid line).

For voices we have a simple radiation pattern with an ever more bundling characteristic with increasing frequency. A loudspeaker having about the same radiation pattern is easy to make and your anechoic voice recording will sound perfect. If we try to apply this to a musical instrument we run into trouble, musical instruments can have wild radiation patterns (see Figure 1), so recording in an anechoic room will result in a spectral imbalance. This inequity will also propagate in the artificial reverberation that is often added in recordings that have a “dry” overall sound quality. Also we run into further problems if we apply this approach to the recording of events with multiple acoustic sources such as an orchestra. We would need a large number of anechoic mono recordings that have to be played back over at least the same number of correctly placed loudspeakers. And also, we should ask the question whether this way of thinking is the correct Hi-Fi approach for large orchestras that do not fit in our living room. The illusion “here and now” is not something that you want when reproducing a symphony orchestra, you will miss the wonderful acoustics of the concert hall. These acoustics provide the optimal

integration of the irregular radiation patterns of musical instruments and the optimal integration over all instruments of the orchestra.

If you are pursuing the "there and then" ideal you can make a recording in a hall where the sound field is according to your wishes, e.g. in the Concertgebouw in Amsterdam. You make a sound recording at the ideal spot using your own ears in combination with in-ear microphones. You also make a recording of your head movements and when playing back the sound you use an individualized HRTF equalized headphone in combination with a head track system. This does not sound very practical, and as mentioned before you still miss something in the low frequency range. You can also try to recreate the sound field of the Concertgebouw in your own room with a set of loudspeakers, maybe using a special kind of surround recording technique. However creating the "there and then" illusion with loudspeakers is difficult due to two effects, room reflections and de-colorization. The reflections in your own room have to be eliminated, otherwise you are listening to the "unwanted" acoustics of your own room and not solely to the "wanted" acoustics of the room where the recording was made. De-colorization is dominated by binaural processing but also takes place in single ear listening (binaural and monaural de-colorization) and reduces the impact of room reflections. In both modes head movements are important and this de-colorization processing is always carried out when listening to an acoustic source in a room or a concert hall. This leads to the phenomenon that when we listen to a loudspeaker in a room we can de-colorize the sound on the listening spot to a large certain extent, but we cannot de-colorize the room where the recording was made due to imperfections in the reproduced sound field. So almost all sound recordings are made on spots where we would not prefer to sit when listening to a live performance, there where the sound is too "dry". This effect can be shown to the extreme by making a mono voice recording in a room and play it back over headphones. Due to the unnatural reproduction of the room reflections over the headphone the recording will sound "wet", a normal room will sound like a bathroom.

Both illusions, "here and now" and "there and then", are thus difficult to create and they are not compatible, a system that produces a perfect "here and now" illusion can never produce a high quality "there and then" illusion, and vice versa. So we can never speak about "the" sound quality of a loudspeaker or a headphone, we need to know the context. So what is the most realistic acoustic reproduction compromise? For a long time the idea was that a headphone with individualized HRTF reproduction and head tracking is the ultimate solution for the "there and then" illusion. However the individualization of the transfer and the head tracking are extremely difficult to implement, the role of binaural and monaural de-colorization is under estimated and the low frequency reproduction is not satisfactory, people want to "feel" the low frequencies. The final answer remains something with loudspeakers, but what is the correct approach, Surround? Wave Field Synthesis? I will give the answer in the upcoming parts, starting with Part 3: The Ideal Loudspeaker, Diffuse Field Equalization.

John G. Beerends
Published in Hifi Video Test 11/2007 (in Dutch), translated and updated over the period 2012-2023.