The Basics of High Fidelity
Part 7: What Do We Really Want?

In the previous six parts we have deepened our understanding of HiFi and come to the conclusion that HiFi stands for "naturalness" in the sense that we are striving for a certain amount of transparency between recording and reproduction. We have also seen that there are two different kinds of transparency, "here and now" (augmented reality) versus "there and then" (virtual reality), which require non compatible optimizations on the recording/reproduction chain. This leads us towards the big question What Do We Really Want?

I have carried out many subjective tests, especially regarding the assessment of speech quality, and have seen strange personal preferences regarding what we like. We also know that people tend to adjust to a certain sound, if you listen a long time to your own loudspeakers you may develop a personalized preference bias. Musicians also suffer from this and it is not a good strategy to ask a musician about his favorite loudspeaker, in general musicians listen to their own instrument at close distance which automatically results in a distorted view on the reality of listening to music. And to be clear, we are allowed to develop our own preference bias, but we should realize that this leads to a shift from science to art.

At this point it is wise to define more precise the terms audio and sound quality. We will use the term audio quality whenever it is related to the transparency goal and sound quality whenever it is related to a personalized preference. This implies that for audio quality, in either the electric or acoustic domain, we need a predefined ideal allowing us to force subjects towards a unified opinion. For sound quality, which is only defined in the acoustic domain, we have to deal with personal preferences which are sometimes difficult to average over large sets of subjects.

So let's go back to science and try to develop characterizations which can be used in both the audio and sound quality domain. In psycho acoustics the four most fundamental characteristics of a steady state sound are:

- Loudness (amplitude / volume control)
- Pitch (time / play back speed control)
- Timbre (frequency / bass-treble tone color control)
- Spaciousness (localization and reverberation / immersion control)

Now let's use these characterizations to find out What Do We Really Want!

The first characterization brings trouble, why?, because some people love loudness too much! Almost all modern HiFi equipment is capable of producing more than natural high loudness and if we leave it up to the user they will crank up the volume to a level where the loudness will damage their ears. And if you compare two HiFi loudspeakers the more loud one will almost always be preferred. In the world of telephony many loudness experiments have been carried out and they show that the preferred loudness of a telephone is about 20 dB higher (i.e. 100 time more powerful) than a natural voice at 1 meter distance. In fact this high loudness level is standardized by the ITU (International Telecommunication Union) as the optimal loudness level. Loudness is like a drug, you adapt and need more and more to be satisfied, in the end always leading to a damaging high loudness level. From a technical perspective loudness is thus no problem and the

volume knob on your HiFi system is by far the most important knob which you should manipulate with care.

Pitch, the second characterization, used to be a huge problem in the analog world. Play a single piano note and reproduce it over a classical HiFi system using a vinyl recording, or even worse over a cassette recording, you will be able to clearly perceive audible wow and flutter. In the old analogue world you needed an expensive studio quality tape recorder to hear no wow and flutter. But fortunately if you run the same experiment with a CD the wow and flutter will be inaudible. And although a tremolo represents a wanted flutter there is seldom a post processing in our HiFi system that adds flutter. The only control that is useful is play back speed control, which in its most advanced form is used to play back speech faster than live while maintaining the correct pitch using advanced PSOLA (Pitch Synchronous Over Lap Add) algorithms. Note that pitch is related to frequency content, but is dominated by repetition time, e.g. an harmonic sound with frequencies of 600, 800, 1000 and 1200 Hz will have a dominant virtual pitch of 200 Hz (inverse repetition time), while the four spectral pitches are less dominant. Our second conclusion is that we are happy with our perfect digital flutterless reproduction while play back speed control with pitch preservation opens up a new world of post processing possibilities.

Timbre, the third characterization, is predominantly determined by the frequency response of the complete recording/reproduction chain. And from this chain the responses of the recording and reproduction room dominate perception. While the response of the recording room can and is manipulated to provide high quality recordings, the response of the reproduction room is in most cases poor due to (room) resonances, leading to the conclusion that our room is the dominating factor in the final timbre. Only with headphone reproduction one can bypass the reproduction room degradation. However as explained in the previous papers headphone reproduction suffers from other, even more disturbing, degradations. We have also seen in the previous papers that resonances are the most disturbing timbre degradations and a first goal of HiFi reproduction is to suppress all unwanted resonances in the reproduction room as well as in the loudspeaker enclosure. When we are able to suppress all unwanted resonances we can easily optimize the timbre of our reproduction by balancing the low (20-200 Hz), mid (500-2000 Hz) and high frequencies (4000-20000 Hz) as was discovered by Baxandall who designed a timbre control still used in HiFi amplifiers. In practice we see that most modern recordings are already timbre optimized and need no, or little personal adaptation. So our third conclusion is that the optimal timbre hardly needs any post processing when using modern, well balanced, recordings.

Spaciousness, the fourth characterization is build up from the spaciousness of the recording room, the artificially added spaciousness and the spaciousness of the reproduction room. Regarding our transparency goal we have two different goals to strive for, “illusion here and now” (augmented reality) and “illusion there and then” (virtual reality). In modern recordings however we see more and more artificially created spaciousness and we can do whatever we like and there is no reality to strive for. Some people like to add artificially generated spaciousness in their home HiFi system on top of the spaciousness as found in the recording, so optimal spaciousness may require post processing and/or recording/play back techniques that require a multi-channel approach. We should be aware of the fact that multi-channel approaches tend to produce more problems than they solve due to the problem of localization stress. This is especially true

for full 6 degrees of freedom audio reproduction. For music reproduction the feeling of immersion, restricted to small head movements and focused on 3 degrees of freedom, is more important than the accurate localization of musical instruments. Advanced systems like Dolby ATMOS, high order Ambisonics or object based audio coding are only necessary in films where sound effects may require a more exact localization and where binaural and monaural decolorization requires a special recording / play back approach that forces one towards the use of a foley artist. In contrast, music reproduction requires a well-balanced diffuse field which by definition cannot be localized. So our fourth conclusion is that spaciousness is a problem and in years of trying to get control over this issue my conclusion is that the optimal additional spaciousness processing in home systems is strongly dependent on the recorded material and should be focused on allowing to control the diffuse field in such a manner that it can easily be adapted to the content that is played.

A trivial post processing strategy that improves the perceived envelopment of stereo recordings is to just add two extra (surround) loudspeakers that simply reproduce a slightly lower volume of the left and right loudspeaker [1]. They should preferably be positioned in such a way that they only contribute to the diffuse field, thus allowing for a simple control over the amount of immersion. Further improvement and control over the amount of immersion, requiring only marginal post processing, can be achieved by taking the difference signal of the Left and Right loudspeaker (L-R) and reproduce this signal over diffuse radiating Left and Right Surround speakers. A more complex algorithm that also provides a center channel was formulated by Irwan and Aarts [2]. In general two-to-five up mixing algorithms provide a poorer front image quality and sometimes a small improvement in immersion but in most cases the original stereo reproduction is preferred [3]. A simpler extension, that also improves low frequency performance, was formulated by my best friend, Ton Willekes; just send the low frequency part of the Left and Right loudspeaker towards the Left and Right Surround loudspeaker and send the high frequency part of the difference signal (L-R) to the Left and Right Surround loudspeaker. This approach prevents the audible degradation caused by the audibility of the antiphase between the Right and Left-Right signal and also reduces the impact of low frequency room modes. The Left and Right Surround signals can be derived directly from the Left and Right loudspeaker signals using an extra amplifier to allow volume balancing between the direct and diffuse field that follows the volume setting of the Left and Right loudspeaker (see Figure 1). Also the Left and Right Surround speaker are radiating towards the walls of the listening room instead of directly radiating towards the listener as used in standard surround set ups [3], [4]. In the ideal postprocessing this immersion control is combined with diffuse field processing of the Left and Right front loudspeakers to correct directivity timbre problems as explained in Part 3.

With this advanced diffuse field reproduction approach we can achieve levels of immersion that sound better than advanced multichannel recording/playback systems, or wave field synthesis approaches. Note that in contrast to standard surround setups, that are focused on localization, this setup does not use surround loudspeakers that radiate directly towards the listener, diminishing localization errors and optimizing the diffuseness of the surround sound field. For wave field synthesis the same applies, it is also focused on the optimization of localization and not on the optimization of the diffuse field [5]. Furthermore wave field synthesis may introduce spatial aliasing degradations and the current systems are not favorable in home music reproduction.

An extra strong point in the diffuse field approach as presented in this paper is that it allows to have control over the feeling of immersion by the diffuse field in such a manner that it can easily be adapted to the content that is played.

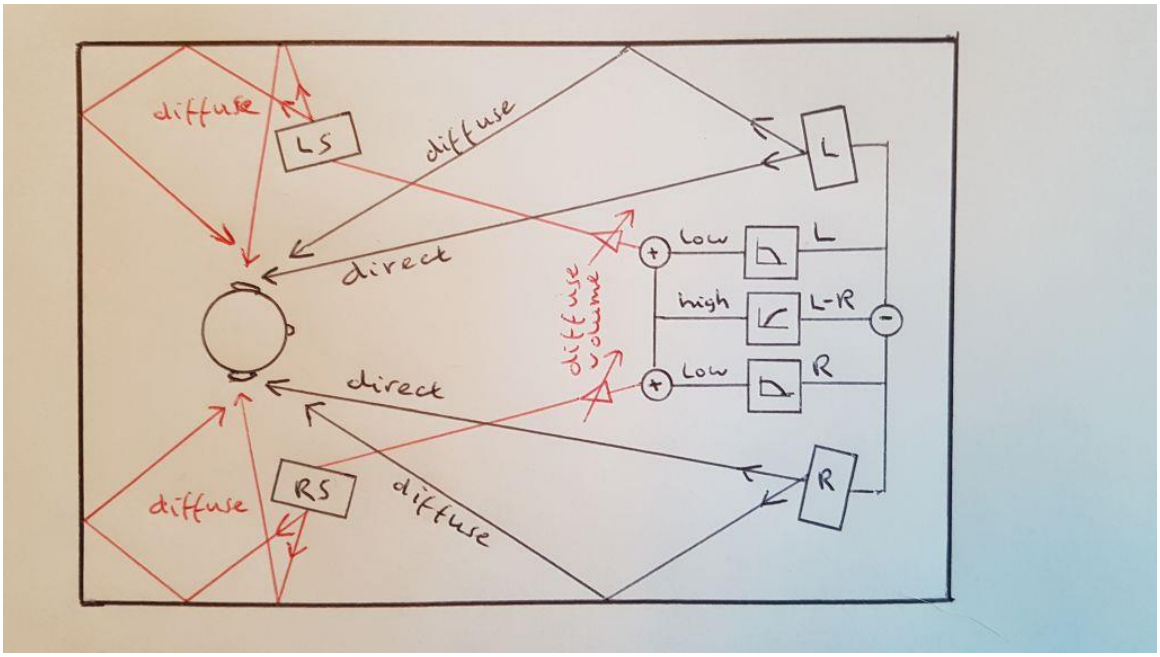

Figure 1. Loudspeaker setup that allows for control of the spaciousness by using the Left-Right difference signal in the high frequency band that is added to the Left and Right signals in the low frequency band. The combined signals are send to the Left and Right Surround speakers (LS, RS). Depending on the characteristics of the recording the diffuse volume control can be used to optimize the feeling of immersion.

Summarizing:

1. We like loudness, but we get more than is good for us, the volume control is the most important knob on our HiFi system.
2. We don't want unnatural pitch variations (wow and flutter in the old analogue world) and all modern HiFi systems can provide this.
3. We want sound with a natural timbre without disturbing resonances, the reproduction room dominates this. Global timbre optimization is possible with the Baxandall approach. The Baxandall knobs "bass" and "treble" are the second important knobs on our HiFi system.
4. We have to choose between the spaciousness of "here and now" (augmented reality), "there and then" (virtual reality) or "anything goes" (better than reality?) and have to adapt our recording/reproduction chain accordingly. The amount of spaciousness of stereo recordings can be optimized with a diffuse field volume control using the setup of Figure 1. This knob should replace all the difficult surround choices with their set up knobs that are currently available.
5. We could add a final point that we want the reproduced sound to be free of unwanted disturbing (background) noises, signal interruptions and nonlinear distortions. From a technical point of view this requirement is not difficult to fulfill.

So we have reached our destination in the HiFi story ? No, one final point has to be discussed, telephony. Although a classic telephone connection is considered to be the rock bottom in HiFi there are some interesting observations to be made regarding the conversational speech quality of a voice link. It starts with the observation that with a telephone we are not only dealing with listening but also with talking and interacting. When you talk you can hear your own voice and when you hear your own voice in the wrong manner the conversational quality is terrible even if the listening quality is perfect. This will be discussed in the final paper, part 8 on Telephony.

[1] M. Tohyama and A Suzuki, “Interaural cross-correlation coefficients in stereo-reproduced sound fields," J. Acoust. Soc. Am., vol. 85, pp. 780-786, (1989 Feb.).

[2] R. Irwan and R. M. Aarts, “Two-to-Five Channel Sound Processing," J. Audio Eng. Soc., vol. 50, pp. 914-926, (2002 Nov.).

[3] F. Rumsey, “Controlled Subjective Assessment of Two-to-Five Channel Surround Sound Processing Algorithms," J. Audio Eng. Soc., vol. 47, pp. 563-582, (1999 July/Aug.).

[4] ITU-R BS.1116, “Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems”, International Telecommunication Union, Geneva, Switzerland (1997).

[5] A. J. Berkhout, D. de Vries and P. Vogel, “Acoustic control by wave field synthesis,” J. Acoust. Soc. Am., vol. 93, pp. 2764-2778, (1993 May).

John G. Beerends
Published in Hifi Video Test 6/2008 (in Dutch), translated and updated over the period 2012-2022.