

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 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 the goal is transparency and sound quality when we are relating to a personalized preference. This implies that for audio quality we need a predefined ideal and we can force all subjects towards a unified opinion while for sound quality 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 sound quality judgments. The most banal characterization used in psycho acoustics for the characterization of steady state sounds is:

- Loudness
- Pitch
- Spaciousness
- Timbre (sound color) defined as everything that is not loudness, pitch or spaciousness

Now lets use these characterizations to find out What Do We Really Want!

The first characterization brings trouble, why?, because people love loudness! 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. Loudness is thus no problem in HiFi, it is a behavioral problem.

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 shocked by the 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. So our second conclusion is that we are happy with our perfect digital flutterless reproduction.

Timbre, the fourth 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 by all 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” and “illusion there and then”. In modern recordings however we see more and more artificially created spaciousness and we can do whatever we like. Some people like to add artificially generated spaciousness in their home HiFi system on top of the spaciousness as found in the recording and so our fourth conclusion is that the 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 de-localization requires a special recording / play back approach that forces one towards the use of a foley artist.

These psycho acoustic characterizations provide a good bases for high quality sound reproduction, but we still lack the characterization of unwanted degradations that can be introduced in a audio (re)production chain like noise, clicks, coding distortion, etc. We

therefore add two more characterizations, “background noise” and “nonlinear distortions”.

Background noise can be divided in two classes, noise like and pulse like. Natural additive noise is less disturbing than modulated noise as found in digital systems. Dithering prevents noise modulation and should always be used in quantization of the audio signal. Nobody likes noise modulation and fortunately in modern 16 or more bit audio systems noise modulations are inaudible [1]. Pulse like degradations are also seldom liked, some vinyl freaks who grew up with scratched discs may like them but in general they fall in the class of unwanted. In modern digital systems pulse like degradations are rare and very much unwanted. Our fifth conclusion is thus that we require a sound that is free of unwanted disturbing background noises.

The final extension, number six, the nonlinear distortion, is the audio garbage can, any nonlinear distortion introduced by the HiFi recording/reproduction chain is unwanted, including the ones introduced by low bit rate audio coding. Although musicians may use a non-linear distortion in their amplifier, no home system has a knob labelled “increase distortion”.

To summarize:

1. We like loudness but we get more than is good for us.
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 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.
4. We want sound with a natural timbre without disturbing resonances, the reproduction room dominates this.
5. We want sound that is free of unwanted disturbing background noises.
6. We want sound that is free of unwanted disturbing nonlinear distortions.

So we have reached our destination in the HiFi story ? No, one final point has to be discussed, telephony. Although a classical telephony 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 fact that when use 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. Lets discuss it in the final paper on [Telephony](#).

[1] E.B. Meyer and D. R. Moran, “Audibility of a CD-Standard A/D/A Loop Inserted into High-Resolution Audio Playback,” J. Audio Eng. Soc., vol. 55, pp. 775-779, (2007 Sep.).

John G. Beerends

Published in Hifi Video Test 6/2008 (in Dutch), translated and updated over the period 2012-2020.

[Part 1: Transparency and Perceptual Measurement Techniques](#)

[Part 2: Reproduction Philosophy “Here and Now” versus “There and Then”](#)

[Part 3: The Ideal Loudspeaker, Diffuse Field Equalization](#)

[Part 4: The Ideal Loudspeaker, Reflections and Resonances](#)

[Part 5: Audio Compression](#)

[Part 6: Subjective Testing](#)

[Part 7: What Do We Really Want](#)

[Part 8: Telephony](#)