The Basics of High Fidelity
Part 1: Transparency and Perceptual Measurement Techniques

With the introduction of audio compression (MP3, AAC, WMA, Ogg-Vorbis, …) on one side, and high resolution audio (Super Audio CD, DVD-Audio, Blue-ray Audio, …) on the other side of the HiFi spectrum, one can ask oneself what does "HiFi" stand for these days? And why do we need more than two audio channels if we only have two ears? The extension from 1 to 2, mono to stereo, was a clear improvement. But the extension from 2 to 4, from stereo to quadrophony, was a failure and the further extensions towards more than a dozen of channels (5.1, …..22.2), as found in many surround and object based audio systems (e.g. Dolby Atmos, Sony RA360, Auro3D, DTS-X, …), looks like some kind of overkill, especially in the reproduction of music. While for movies sound localization is important, the reproduction of music is seldom improved by adding more reproduction channels with the goal of an improved localization. In music reproduction feeling immersed by a natural sounding diffuse field is far more important than localization. In fact localization can lead to uncontrolled degradations that can be characterized as "hearing things jumping around". As a result of these complicated new audio developments some people have even decided to go back to good old "stereo on vinyl". Yes, in the modern audio world anything is possible so the question arises "what do we want". Are we seeking for "naturalness"?, and if so can we measure to what extent a reproduction sounds natural? In a series of eight papers I will answer these more or less philosophical questions.

In this first paper I will start with an important concept: "transparency". From a technical perspective a lot of HiFi equipment should behave transparently, i.e. nothing should be added to the sound and nothing should be left out. Many HiFi components have an electrical input and output in which case the quality is easily defined, if I subtract the output, after amplitude scaling, from the input and the result is zero the device under test is perfect. If the difference is not zero it can still be inaudible and thus the device under test can still be perfectly transparent from a perceptual point of view. For nearly all modern HiFi components with electric in/out the quality is perfect. Storage, transport and amplification need not to be an issue in HiFi since the introduction of the compact disc (CD) and its associated 44.1 kHz/16 bit digital audio format. Subjective tests have shown that higher sampling rates and/or more bits per sample do not lead to any perceptual relevant improvement [1]. And contrary to what some vinyl "voodoo" audio freaks may tell us the CD is a much better, more transparent, storage medium than the vinyl LP. For the "voodoo" audio freaks that still listen to LP's I will provide two arguments that prove the superiority of the CD over the LP.

The first argument is a mathematical argument that is related to the 44.1 kHz sampling of the time and the 16 bit quantization of the amplitude that is used with the CD. What kind of distortions does this introduce?, and how are they related to distortions as found in analog equipment? Well, as to the sampling of the time axis the Nyquist-Shannon theorem tells us that with a sampling rate that is higher than double the highest frequency of the signal that is being sampled the original signal can be exactly reconstructed. Using the upper limit of audibility of 20 kHz we can conclude that sampling above 44 kHz is high enough for perceptual transparency. Regarding the quantization of the amplitude the situation is slightly more complicated because of the statistical identity between the analog world of perception and the digital amplitude representation. Using a mathematical trick known as dithering we can construct a digital representation of which the error signal has the same mathematical properties as noise found in analog equipment (I would call it Roberts-Lipschitz dithering [2], [3]). If your sampling fulfills the Nyquist-Shannon criterion and if you use the correct Roberts-Lipschitz dithering, the digital signal has exactly the same properties as a band limited analog signal which contains some noise. This analysis shows a direct equivalent between the digital domain and the analog domain. If we make an analog representation with the same quality we need an audio bandwidth of about 20 kHz and a signal to noise ratio of about 95 dB. A beautiful old fashioned vinyl LP may be able to reproduce 20 kHz but a signal to noise ratio of 95 dB can never be achieved, listen to the noise floor, the rumble and scratches of your LP and decide if this medium fulfills the perceptual transparency requirement. If you are still not convinced then play a piano record and listen to the wow and flutter of your LP and then to the perfect CD reproduction with no audible wow or flutter.

A second argument to show superior transparency behavior of one storage format over another, without having an ideal storage medium, is to carry out the symmetric copying experiment. You copy the content of one format onto the other and vice versa and listen to the differences. So copy your LP to a CD and listen to the difference, its hardly detectable. Now copy your CD to an LP and you will clearly hear a difference showing that a CD is the better storage format. So you can make a CD to sound like an LP but the reverse is not true. As already stated the CD format is nearly perfect and subjective tests have shown that higher sampling rates and/or more quantization bits do not lead to any perceptual relevant improvement [1].

If we have a perfect storage medium available we can carry out the ideal transparency test. Play a set of audio files over the system under test and compare the ideal input with the degraded output, if you cannot find any input for which you can hear the difference, the system under test is transparent. Such a test may seem simple but is unfortunately rather difficult to carry out, e.g. the loudness has to be equalized within 0.2 dB in order to prevent identification on the basis of loudness. The test should be carried out in a double blind fashion, i.e. both the subject who is judging the audio quality and the one who is running the test are not allowed to know whether the input or the output is played. Also the subject should have the possibility to listen to the input signal and compare it with the unknown stimulus in a time synchronous play back loop.

A way out of expensive and difficult to carry out subjective listening tests is to use a perceptual measurement approach. In this approach the input and output signals of a device under test are mapped onto an internal representation and subtracted. If the difference in internal representations is zero the device under test is perceptually transparent. If the difference is larger than zero we have to interpret this difference using a cognitive model (see Figure 1) and train it to provide quality scores as obtained in a subjective listening test. This may seem difficult but a simple implementation of this idea which I developed in the early 90's [4] at KPN Research, using speech coding degradations, was benchmarked by the International Telecommunication Union (ITU) and showed correlations above 0.97 in the comparison of subjective and objective quality measurement results. This method, PSQM [5], was accepted as ITU Recommendation P.861 in 1996, the first worldwide standard for measuring narrow band speech quality. After a number of updates and extensions in 2001 and 2011, it is now standardized as P.862 PESQ [6] [7] [8] (narrowband speech) and P.863 POLQA [9] [10] [11] (wide-, super wide- and full band speech). Also a modified version was used in the development of a measurement system for assessing the quality of music codecs such as MP3, AAC, WMA, OGG, … (ITU Recommendation BS.1387, PEAQ, 1998 [12] [13] [14]). Unfortunately the commercial interest in measuring HiFi audio quality is marginal compared to the commercial interest in measuring speech quality where the third generation of the measurement approach was standardized in 2011 (P.863 POLQA) while for HiFi music (BS.1387 PEAQ) no updates were made since 2001, despite its severe shortcomings.

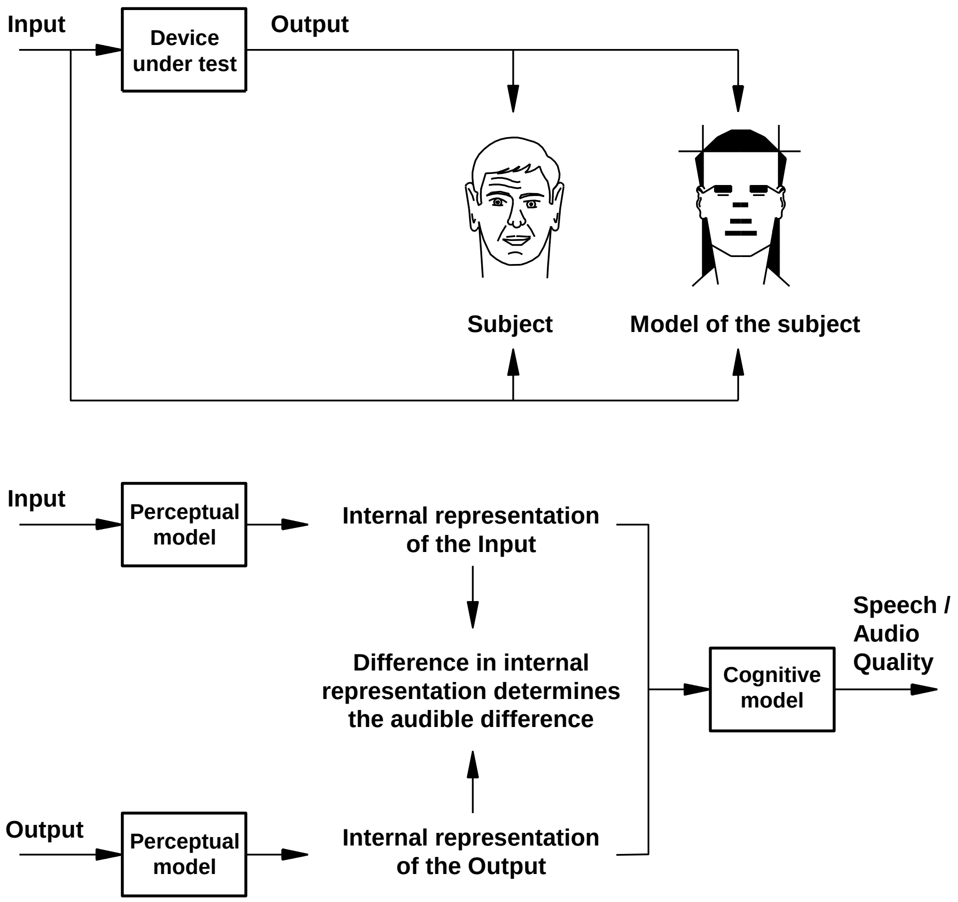

Figure 1. Basics of the perceptual measurement approach. Top, equivalence between a subjective test and an objective perceptual measurement using a model of the subject. Bottom, the internal representation of the input is compared with the internal representation of the output using a cognitive model. If the difference in internal representation is zero, the device under test is perceptually transparent (perfect reproduction).

Although we have acquired basic insights into the problem of transparency we still cannot answer the question whether or not something sounds good. There is a difference between sound quality and audio quality, the first one deals with a wide variety of problems where personal taste comes to play a role, the second one deals with the representation of sound for which we can strive towards transparency. So we end up with the ultimate HiFi question: "can we define a transparency goal for a loudspeaker or headphone reproduced audio signal"? The answer is given in Part 2: Reproduction Philosophy "Here and Now" versus "There and Then".

[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.).

[2] L. G. Roberts, "Picture Coding Using Pseudo-Random Noise," IEEE Transactions on Information Theory. 8 (2), pp. 145–154 (Feb. 1962).

[3] S. P. Lipshitz, R. A. Wannamaker, and J. Vanderkooy, "Quantization and Dither: A Theoretical Survey," J. Audio Eng. Soc., vol. 40, pp. 355–375 (1992 May).

[4] J. G. Beerends and J. A. Stemerdink, "A Perceptual Speech Quality Measure Based on a Psychoacoustic Sound Representation," J. Audio Eng. Soc., vol. 42, pp. 115-123, (1994 March).

[5] ITU-T Recommendation P.861, Objective Quality Measurement of Telephoneband (300-3400 Hz) Speech Codecs (1996 Aug.).

[6] A. W. Rix, M. P. Hollier, A. P. Hekstra and J. G. Beerends, "PESQ, the New ITU Standard for Objective Measurement of Perceived Speech Quality, Part I - Time Alignment," J. Audio Eng. Soc., vol. 50, pp. 755-764 (2002 Oct.).

[7] J. G. Beerends, A. P. Hekstra, A. W. Rix and M. P. Hollier, "PESQ, the New ITU Standard for Objective Measurement of Perceived Speech Quality, Part II - Perceptual Model," J. Audio Eng. Soc., vol. 50, pp. 765-778 (2002 Oct.).

[8] ITU-T Rec. P.862, "Perceptual Evaluation Of Speech Quality (PESQ): An Objective Method for End-to-end Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs," International Telecommunication Union, Geneva, Switzerland (2001 Feb.).

[9] J. G. Beerends, C. Schmidmer, J. Berger, M. Obermann, R. Ullman, J. Pomy and M. Keyhl, "Perceptual Objective Listening Quality Assessment (POLQA), The Third Generation ITU-T Standard for End-to-End Speech Quality Measurement Part I – Temporal Alignment," J. Audio Eng. Soc., vol. 61, pp. 366-384 (2013 June).

[10] J. G. Beerends, C. Schmidmer, J. Berger, M. Obermann, R. Ullman, J. Pomy and M. Keyhl, "Perceptual Objective Listening Quality Assessment (POLQA), The Third Generation ITU-T Standard for End-to-End Speech Quality Measurement Part II – Perceptual Model," J. Audio Eng. Soc., vol. 61, pp. 385-402 (2013 June).

[11] ITU-T Rec. P.863, “Perceptual Objective Listening Quality Assessment,” Geneva, Switzerland (2011 Jan.).

[12] J. G. Beerends and J. A. Stemerdink, “A Perceptual Audio Quality Measure Based on a psychoacoustic sound representation,” J. Audio Eng. Soc., vol. 40, pp. 963-978 (1992 Dec.)

[13] T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerends, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg, B. Feiten, “PEAQ - The ITU-Standard for Objective Measurement of Perceived Audio Quality,” J. Audio Eng. Soc., vol. 48, pp. 3-29 (2000 Jan./Feb.).

[14] ITU-R Rec. BS.1387, “Method for Objective Measurements of Perceived Audio Quality,” International Telecommunication Union, Geneva, Switzerland (2001 Nov.).

John G. Beerends
Published in Hifi Video Test 10/2007 (in Dutch), translated and updated over the period 2012-2022.

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?, Immersion!

Part 8: Telephony