The Science Behind HiFi Audio

JOHN G. BEERENDS

TNO, The Hague, The Netherlands

RICHARD VAN EVERDINGEN

Delta Sigma Consultancy, Nunspeet, The Netherlands

ABSTRACT

According to Wikipedia, 'Hi-Fi' [1] is high-quality reproduction of sound without audible noise and distortion, based on a flat (neutral, uncolored) frequency response within the human hearing range. Technically, these requirements can be easily met today, even with moderately priced consumer equipment. Note that this definition does not mention the number of channels required in the recording and playback system. Because we only have 2 ears one would expect that 2 channels are enough but why then was quadrophony introduced, why did it fail, and why are more and more audio channels introduced in HiFi recording and play back? Is there a scientific, perception based, approach that allows us to assess the quality of HiFi audio systems, including transduction devices like microphones and loudspeakers? The short answer is yes but only up to a point that personal preferences dominate perception.

1 INTRODUCTION

The first author of this paper once attended a HiFi show where a small company demonstrated a very expensive tube amplifier. During that demo a soft hum was audible in the silent intervals of the music, but the showrunner could not perceive the hum, not even when he moved towards the best listening position. However, when he was asked to listen to the loudspeakers at closer distance, he could also perceive the soft audible hum. Nobody appreciates a humming amplifier and this experiment points towards a fundamental problem in the development of a scientific

approach towards HiFi audio, we all have different ears, and different brains, making it difficult to separate opinions from facts which often leads to fake news with fake quality claims.

This anecdote also points towards a good starting point for a scientific HiFi audio approach, the concept of transparency. For many audio devices, like amplifiers, audio compression schemes, storage and transport techniques, there should be no audible difference between the input of the device and the output. If we take the tube amplifier as an example, we could make a recording of the input and compare it with a recording of the output. If we then subtract the output from the input, after alignment of the amplitude and compensating for a possible delay, we should get an overall zero output. If the difference signal is not zero there still could be no audible difference between the in- and output, making the device under test transparent from a perceptual point of view. And if the device under test is not perceptually transparent, we should have some kind of interpretation algorithm that allows to quantify the impact of the audible degradation. This transparency approach can be used to construct a perceptual measurement system that allows to check for audible degradations in the storage, transport and amplification of audio signals using the basic set up of Figure 1.

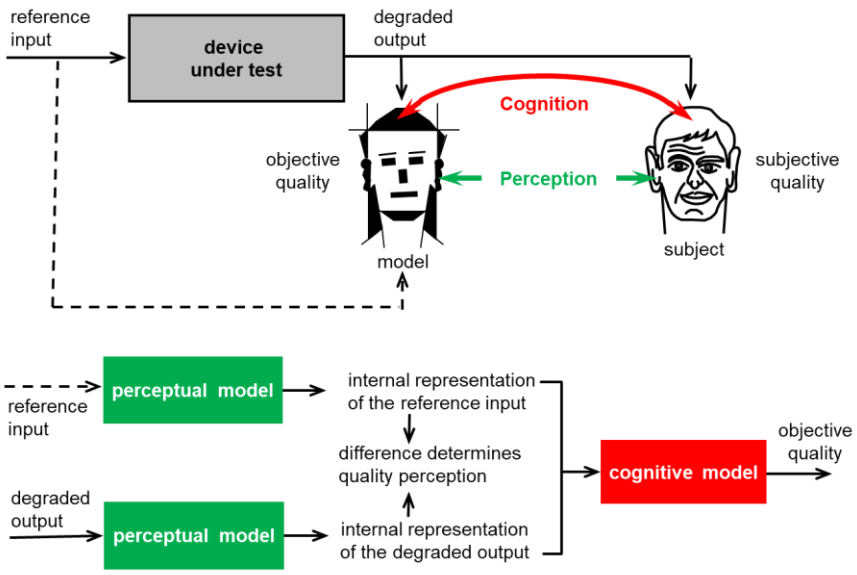


Figure 1. Basic approach in assessing the quality of HiFi audio devices that are designed for perceptual transparency. It uses a combination of a perceptual and cognitive model to predict the impact of audible differences introduced by the audio device.

A simple implementation of this perceptual measurement approach was developed in the early 90's at PTT Research NL [2] for assessing the audio quality of speech codecs. This method, called PSQM, was benchmarked by the International Telecommunication Union (ITU) in 1995 and showed correlations above 0.97 in the comparison of subjective and objective speech quality measurements and was accepted by the ITU as Recommendation P.861 in 1996 [3], the first worldwide standard for measuring narrow band speech quality using perceptual modelling. Later improvements in the modelling allowed to asses wider bandwidths and the current standard, P.863 POLQA [4] [5] [6] can deal with full audio band width speech [7].

The same perceptual measurement approach was used in the development of a system for assessing the quality of music codecs [8] such as MP3, AAC, WMA, OGG, (ITU Recommendation BS.1387, PEAQ, 1998 [9] [10]). However assessing the quality of audio codecs with music is far more difficult than with speech, especially when the codecs behave non transparently. Subjects tend to have more widely different opinions on the impact of degradations with music than with speech. Furthermore, individual differences in the properties of the perceptual and cognitive processing are far more important when listening to music then when listening to speech. Even

simple differences in the absolute perceptual threshold can lead to large differences in quality assessments of music codecs, especially with degradations in the high frequency region (above roughly 8 kHz) which are less important in speech coding. This is reflected in the correlation results obtained with perception-based measurement models which are significantly lower in experiments that use music in comparison to when speech is used.

A fundamental problem in the development of an objective perceptual quality assessment method is that in many use cases subjects do not have access to a predefined ideal reference signal, and only use the output of the device under test in the quality assessment, using an unknown internal ideal in their personalized opinion on the audio quality. If we have access to the individual ideal representation that a subject uses in his quality judgement, we could design a processing method that delivers a personalized perfect audio quality. This idealization idea turned out to be very important in the assessment of speech codecs where many signal processing techniques are used to enhance the perceived speech quality, even when high quality voice recordings are used. For music recordings, sometimes post processing techniques, like artificial reverberation, are used to enhance high quality audio recordings. For speech quality assessment the idealization approach is implemented in the latest ITU standard, POLQA P.863, which uses a cognitive model that allows to assess speech enhancement techniques. In POLQA an average ideal speech representation is constructed that is derived from the reference input signal on the basis of the analysis of a large set of speech quality assessments (see Figure 2).

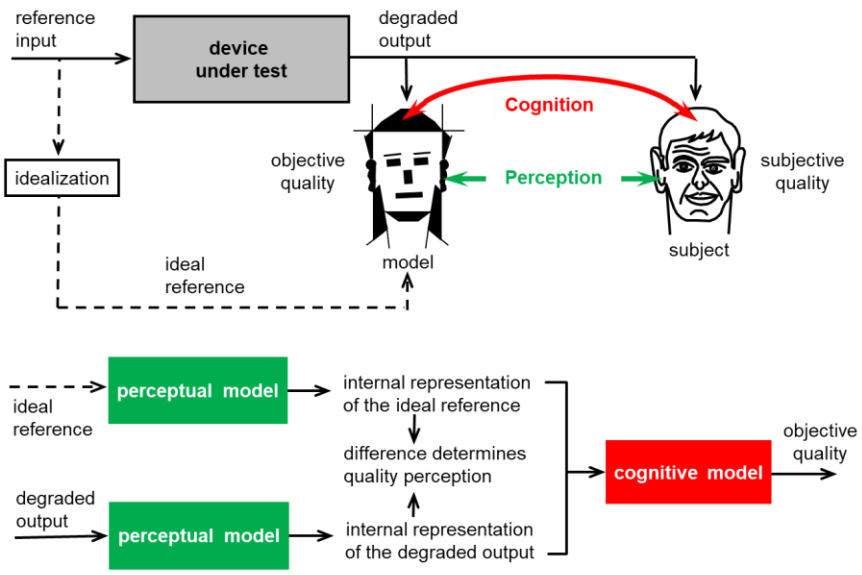


Figure 2. Basic approach in assessing the quality of speech codecs that are not necessary designed for perceptual transparency. It use a combination of a perceptual and cognitive model to predict the impact of audible differences based on an ideal reference that is derived from the reference input signal.

Because our ears operate in the acoustic domain subjective tests will always need a transduction device, a headphone or a loudspeaker, or even maybe a large set of loudspeakers, in the assessment of an audio device. The transduction device that we use has to be of superior quality in order to allow to hear out small degradations in the degraded output, especially if we want to assess high quality devices that are designed for perceptual transparency. In testing high quality audio devices for transparency subjects are allowed to directly compare the reference input with the degraded output, allowing them to detect very small degradations in the output signal [11].

If we want to assess the quality of headphones and loudspeakers using perceptual modelling, we run into the problem that the output is an acoustic wave that we need to transform back to the electric domain in order to be able to feed it to a perceptual measurement model. This recording of the output of a loudspeaker or a headphone is difficult and can be carried out in many ways that all lead to different assessments of the device under test. This is the core problem in the science of HiFi audio quality assessment, subjective tests of microphones, headphones and loudspeakers are

all based on judgements that use an unknown internal ideal and developing an objective perceptual measurement approach is very difficult.

2 TRANSLATION OF THE TRANSPARENCY GOAL TO THE ACOUSTIC DOMAIN

Up to now we have used the objective perceptual measurement approach for devices with electric input and output. Can we extend this approach towards the assessment of a complete recording/play back audio chain including the transduction, carried out with microphones on the recording side and headphones/loudspeakers on the play back side? A first step is the construction of an experiment that uses our own ears in a “live versus recorded” transparency test in the acoustic domain. If you set up such an experiment you will notice that when you listen to a recorded sound, the room where the recording is made has a significant impact on the audio quality. Listen to a voice recorded in a bathroom and you will hear that the room dominates the audio quality. Also the way we reproduce the recording has a significant impact on the audio quality. If we listen over headphones the sound is unnatural and we need personalized HRTF (Head-Related Transfer Function) corrections, including compensations for head movements.

If we listen over loudspeakers the reproduction room has a significant impact on the perceived audio quality. The advantage of the loudspeaker approach is that the reproduction room has the same degradation on the audio quality as would have been perceived as we would listen to the live source in that room. This allows us to make a recording that will lead to a transparent recording/play back of a single acoustic source. We just make mono recording of the acoustic source in an anechoic room and play it back through a single loudspeaker with the same directional properties as the acoustic source. If we then listen to that recording in a room, we have a full Six-Degrees-of-Freedom (6DoF) [12] reproduction that is transparent in comparison to the live source. The whole chain of capture and reproduction is transparent.

Unfortunately this approach requires that the directional properties of the loudspeaker match the directional properties of the acoustic source. For a single voice this can be achieved, but If we try to apply this to a musical instrument we run into trouble (see Figure 3).

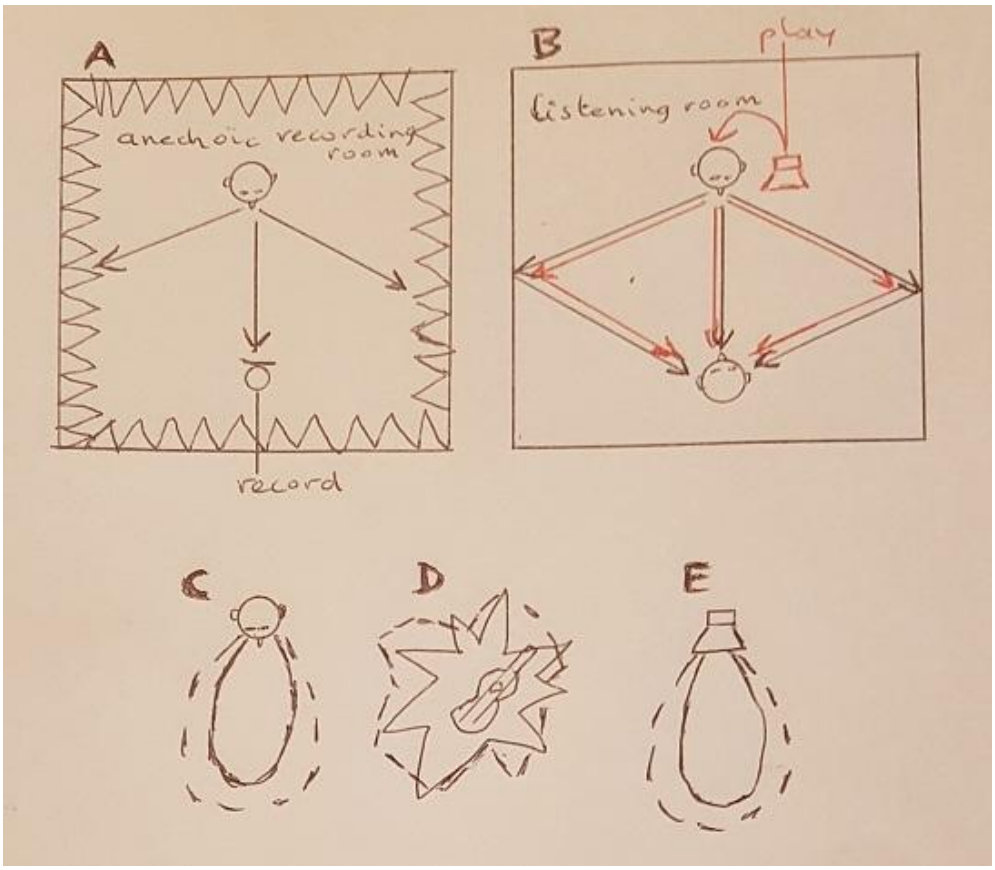


Figure 3. Transparent recording/play back of a single acoustic source. A mono recording of a voice in an anechoic room (A) can be played back transparently in any room over a loudspeaker that has the same directional properties as the voice (B). While a loudspeaker and a voice can have similar directional 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).

Musical instruments can have wild directivity patterns, 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.

For the recording of music events, one should try to capture a similar immersive feeling as one would have experienced at the live event, with the acoustics of the recording room providing proper acoustic integration across all instruments, including their directional patterns. In playback, we should then try to reproduce the sound field as would have been experienced in the live event, especially taking into account the feeling of immersion. This shows that there are two approaches in the Hi-Fi recording/playback chain: One focused on the transparency ‘Here and Now’, the augmented reality approach, and the other on the transparency ‘There and Then’, the virtual reality approach. Both require completely different, incompatible, recording/playback techniques.

If we aim for the illusion of ‘There and Then’, what is the minimum number of audio channels required for Hi-Fi quality loudspeaker reproduction? The introduction of stereo significantly improved the perceived loudspeaker reproduction quality of music events compared to mono. For headphone reproduction we do not need to extend the number of channels, although a high quality, transparent recording-reproduction chain requires personalized HRTF corrections, including compensations for head movements. In loudspeaker reproduction, personalized HRTF corrections are unnecessary, while the quality is determined by other characteristics of which the acoustics of the listening room is a dominating factor. For a high-quality experience, room modes should be damped and we should aim for a low reverberation time, preferably less than 0.5 seconds, allowing to hear the longer echoes of recording environment. If the reverberation time of the listening room is extremely low, such as in an anechoic room, we will experience a lack of feeling immersed by the sound and to compensate for that effect, an extremely large number of audio channels would be required.

The number of audio channels needed for Hi-Fi quality loudspeaker reproduction of music events in a living room environment is not clear. While the expansion of the number of recording/playback channels from 1 to 2 – mono to stereo – was a great improvement, the extension from 2 to 4 – stereo to quadrophony – was a failure. Further expansion to more than a dozen channels (5.1 up to 22.2 or even higher) as in many surround systems, seems inconsistent with the characteristics that improve music reproduction. Advanced systems like Dolby Atmos, Sony RA360, Auro3D, DTS-X, high order Ambisonics, Wave Field Synthesis or Object-Based Audio Coding ([13], [14], [15], [16], [17]), are mainly useful in films and games, where sound effects require a more exact localization. To ensure correct binaural and monaural decolorization in these environments, specific recording techniques are required that often forces one to use unconventional recording techniques, e.g. as used by foley artists.

While for movies and games sound localization is important, especially in the case of moving objects, the reproduction of music is seldom improved by adding more reproduction channels. The feeling of being immersed by a natural sounding diffuse field is much more important than an improved localization. In fact, adding more reproduction channels can lead to uncontrolled degradations that can be characterized as “hearing things jumping around”.

While the number of audio channels on consumer equipment has been growing, recording of music has remained mainly in stereo. Although many upscaling algorithms have been developed, it turns out to be difficult to create a high-quality, immersive diffuse field. Complex algorithms can be formulated (see e.g. [18], [19]), possibly extended with a center channel, but in general, two-to-five up mixing algorithms provide a poorer front image quality and only a marginal improvement in immersion. In most cases, the original stereo reproduction is preferred [20] although the feeling of immersion is less than in a live music event.

The major problem with immersion is the fact that it is a highly cognitive term that was only recently introduced in the world of sound reproduction over loudspeakers. The feeling of being Immersed is related to the perceived sound quality and therefore difficult to define and measure [21]. In general, quality has two different dimensions: function and beauty. Quality optimization usually starts with the former. An excellent car, for example, should never fail in its function of

transportation; it must fulfill it with high reliability. Once that is achieved, the focus shifts towards the beauty dimension. But because that is in the eye of the beholder, it is difficult to quantize and optimize.

In sound quality research, the focus has thus been more on the functional quality aspects like localization and to a lesser extent on beauty aspects like immersion. The first studies related to immersion were carried out in the context of speech perception, where being immersed in a (cocktail) party leads to functional difficulties in the understanding of a single voice [22], [23]. This aspect of immersion is thus related to the functional localization aspect, where one tries to optimize speech intelligibility. The same could be said about the increased number of audio channels in the aforementioned advanced audio systems that focus on localization accuracy, without taking into account the beauty aspect.

The beauty aspect of immersion has only been studied more recently. Eaton and Lee [24] asked experts to rate ten aspects of sound quality in relation to immersion. Horizontal envelopment was found to be more important than vertical. What remains unclear, however, is to what extent subjects prefer to be immersed by a sound. Take the example we gave in the introduction: the reproduction of an anechoic recorded single voice, reproduced by a single loudspeaker, leading to full Six-Degrees-of-Freedom (6DoF) [12] reproduction. If we play that recording over a standard stereo setup, or over four loudspeakers using quadrophony (the big mono approach), the feeling of immersion improves with the number of loudspeakers used, but not the perceived sound quality. An increase in immersion is expected to be observed if a natural surrounding diffuse field is created. Four direct-radiating loudspeakers are, due to their properties, not very suitable for this purpose.

Many designers recognize the importance of wide spread directivity of loudspeakers for high-quality music reproduction. In general, the sense of immersion is less with classic direct radiating speakers than with omnidirectional designs. To improve the feeling of immersion, designers have used additional drivers that do not radiate directly towards the listener. The best-known is probably Amar Bose who designed an enclosure that has additional drivers in the back panel to produce reflections against the walls, thereby improving the balance between the direct and diffuse fields [25]. Kantor and De Koster from Acoustic Research extended this idea and developed an enclosure

that use extra backwards radiating drivers to equalize the diffuse field room response independently from the direct field [26]. The first author of this paper demonstrated the quality improvement of this approach in 1988 for manufacturer BNS, at the Dutch Firato exhibition [27]. An extra set of back radiating loudspeakers that can be added to any regular stereo setup was used to equalize the diffuse field response. However, the weakness of all these setups is that they primarily create a frontally localized diffuse field.

3 OPTIMIZATION OF HiFi LOUDSPEAKER REPRODUCTION

While new standards focus on an ever increasing number of reproduction channels [28], allowing improved localization, excellent quality of immersion can be achieved using ordinary stereo recordings that are reproduced by a regular stereo loudspeaker setup, complemented by only two additional surround loudspeakers of omnidirectional design, that project most of their sound energy towards the walls. These additional speakers should be designed to ideally contribute only to the diffuse field, so that the degree of immersion can be easily controlled without introducing localization errors. This also reduces undesired comb filtering effects that arise from the interference between front and rear output signals. A simple, but effective way to create a diffuse radiating surround speaker is to use a cone-shaped diffuser that creates, for a substantial part, a 360 degrees horizontal radiation pattern. In the optimal construction, the speaker is designed to minimize its contribution to the direct field, for example by limiting the radiation to about 300 degrees (see Figure 4).

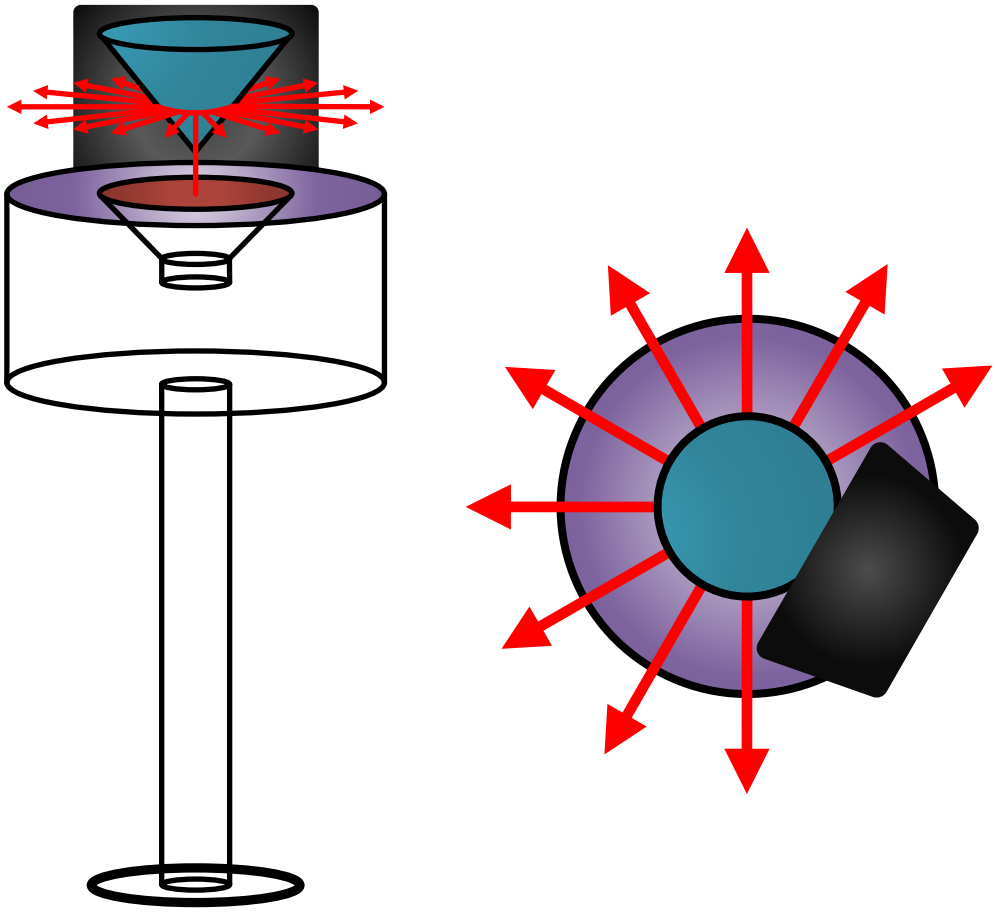


Figure 4. Side and top view of the surround loudspeaker used to create a diffuse field behind the listener. The grey block is attached to the mounting of the cone and is covered with absorbent material that provides shielding of the direct sound, so that dispersion is mainly limited to around 300 degrees omnidirectional.

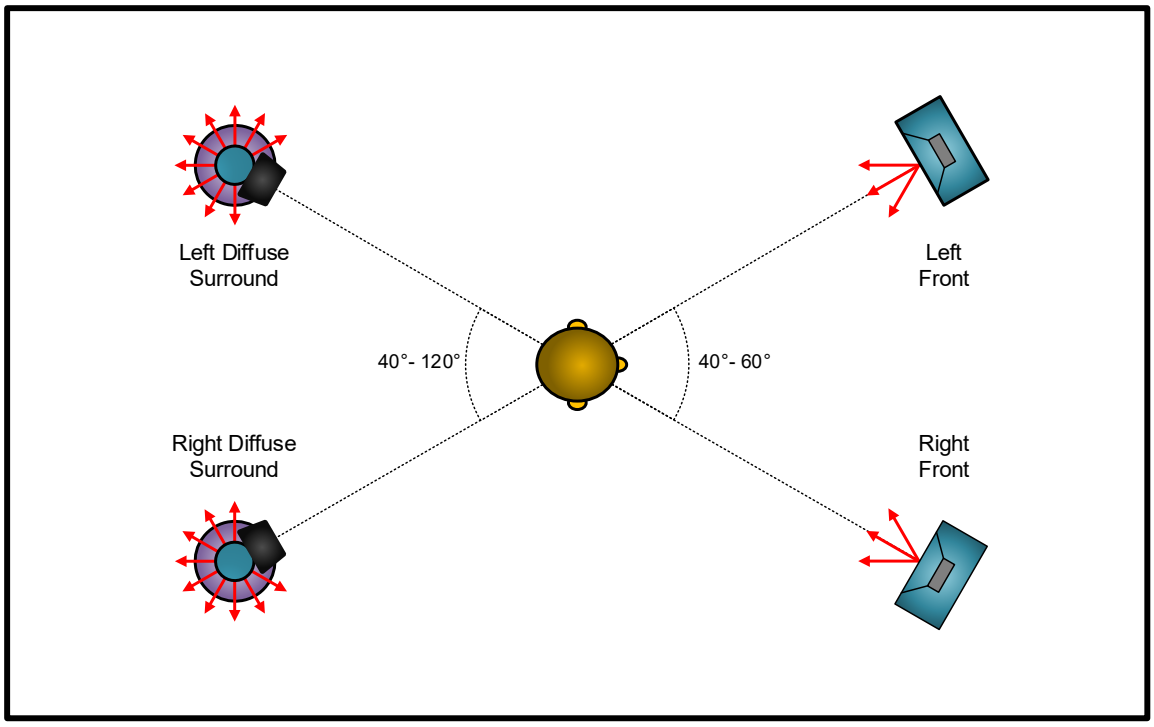


Figure 5. Loudspeaker setup that allows for Immersion Control by use of the cone-shaped diffuser of Figure 4, that mainly contributes to the surround diffuse field.

Figure 5 shows the layout of a complete speaker setup designed for an optimal sense of immersion that can be adapted to one's own personal preferences. Note that the Left and Right Diffuse Surround speaker mainly radiate towards the walls of the listening room, opposed to the standard surround setups where the surround speakers radiate towards the listener [11], [20]. This approach prevents localization degradations that can be characterized as "hearing things jumping around". The setup is limited in its capability to create a perfect diffuse field due to its focus on the horizontal plane, but in general the feeling of immersion is dominated by sound in that direction, compared to vertical [24]. To keep the front stereo image stable, and avoid localisation of the direct field in the rear loudspeakers, the audio signal for the surround speakers should be given an extra delay. In a co-operation with a number of small HiFi companies in The Netherlands a series of experiments were carried out where experts could adjust the delay and level of the diffuse surround speakers. For the delay the optimal value was between 10 and 20 ms depending on the acoustic properties of the room where the recording was made. Roughly speaking, more delay could be

allowed for recordings that are made in large concert halls than for dry pop recordings. The optimal level of the additional surround speakers was also dependent on the properties of the recording but also differed largely between experts. When the system was evaluated with naïve listeners the preferred level of the extra diffuse field speakers showed an even larger variation. Some subjects set the level close to the just noticeable difference, about 20 dB below the level of the direct field loudspeaker, while others choose to set it above the level of the direct field loudspeakers. From the 24 subjects that were used 23 preferred to switch on the extra diffuse field speakers for the majority of the used music fragments while 16 subjects always switched it on. One subject only switched it on in 43% of the fragments. The most interesting conclusion from the experiments was the significant increase in perceived overall sound quality when the diffuse surround speakers were switched on. Using a five point scale ranging from 1, a very small improvement to 5, a very big improvement, the experts judged the overall sound quality improvement around 3 while the naïve listeners judged the quality improvement even bigger with scores around 4.

The advantage of the diffuse field approach given in this paper is that only two small additional surround speakers are needed for a significant increase in overall perceived sound quality, without introducing degrading localization errors as found in many surround setups. The setup allows for a simple 'Immersion Control' that can be adapted to the characteristics of any standard stereo recording, and to one's own personal immersion preferences, through simple playback optimization.

The extreme dependence of the optimal perceived immersion on personal preferences makes it difficult to design an objective perceptual measurement system that allows to assess the overall sound quality of a system. For mono speech and music, and to some extent stereo music, perceptual models have been developed that show good correlation between objective measurement and subjectively perceived speech [5] [6] and music quality [9]. Models for spatial audio quality have been developed, but do not take into account the personalized immersion optimization as presented in this paper [29] [30] [31]. Thus there seems no possible way forward to allow for objective perceptual audio quality measurements without taking into account highly

personalized preferences that are also dependent on the room where the reproduction takes place [32].

4 REFERENCES

[1] High Fidelity, Wikipedia.

[2] 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).

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

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

[5] 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).

[6] 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).

[7] J. G. Beerends, N. M. P. Neumann, E. L. van den Broek, A. Llagostera, J. Torres Menendez, C. Schmidmer, and J. Berger, "Subjective and Objective Assessment of Full Bandwidth Speech Quality," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 440-449 (2020 Jan.) .

[8] 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.)

[9] 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.).

[10] ITU-R Rec. BS.1387, “Method for Objective Measurements of Perceived Audio Quality,” International Telecommunication Union, Geneva, Switzerland (2023 May).

[11] ITU-R BS.1116, “Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems”, International Telecommunication Union, Geneva, Switzerland (2015 Feb.).

[12] Six-Degrees-of-Freedom, Wikipedia.

[13] M. A. Gerzon, “Periphony: With-Height Sound Reproduction”, J. Audio Eng. Soc., vol. 21, 2-10 (1973 Feb.).

[14] M. A. Gerzon, “Ambisonics in Multichannel Broadcasting and Video”, J. Audio Eng. Soc., vol. 33, 859-871 (1985 Nov.).

[15] Furness, Roger K. “Ambisonics - an overview,” Audio Engineering Society Conference: 8th International Conference: The Sound of Audio, pp. 181-190 (UK, May 1990).

[16] 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).

[17] S. Spors, H. Wierstorf, A. Raake, F. Melchior, M. Frank, & F. Zotter, “Spatial Sound With Loudspeakers and Its Perception: A Review of the Current State,” Proceedings of the IEEE, 101(9), pp. 1920-1938 (2013 Sep.).

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

[19] C. Faller, “Multiple-Loudspeaker Playback of Stereo Signals*," J. Audio Eng. Soc., vol. 54, pp. 1051-1064 (2006 Nov.).

[20] 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.).

[21] A. Raake and J. Blauert, "Comprehensive modeling of the formation process of sound-quality," *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, Klagenfurt am Wörthersee, Austria, 2013, pp. 76-81, doi: 10.1109/QoMEX.2013.6603214.

[22] E. C. Cherry: Some experiments on the recognition of speech, with one and with two ears. J. Acoust. Soc. Am. 25, pp 975-979 (1953).

[23] A. W. Bronkhorst, "The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions," Acta Acustica united with Acustica, Volume 86, Number 1, pp. 117-128 (2000).

[24] C. Eaton and H. Lee, "Quantifying Factors of Auditory Immersion in Virtual Reality", Audio Engineering Society International Conference on Immersive and Interactive Audio (UK, March 2019).

[25] A. G. Bose, "On The Design, Measurement, and Evaluation of Loudspeakers", presented at the 35th Convention of the Audio Engineering Society (USA, Oct. 1968).

[26] K. L. Kantor and A. P. Koster, "A Psychoacoustically Optimized Loudspeaker," J. Audio Eng. Soc., vol. 34, pp. 990-996, (1986 Dec.).

[27] J. G. Beerends, "BNS Energetic Diffuse Field System", presented at the FIRATO (NL, Aug. 1988).

[28] J. Herre, J. Hilpert, A. Kuntz and J. Plogsties, "MPEG-H 3D Audio—The New Standard for Coding of Immersive Spatial Audio," in *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 770-779, Aug. 2015, doi: 10.1109/JSTSP.2015.2411578.

[29] R. Conetta, T. Brookes, F. Rumsey, S. Zielinski, M. Dewhirst, P. Jackson, S. Bech, D. Meares, S. George, "Spatial Audio Quality Perception (Part 1): Impact of Commonly Encountered Processes", J. Audio Eng. Soc., vol. 62, pp. 831-846 (2014 Dec).

[30] R. Conetta, T. Brookes, F. Rumsey, S. Zielinski, M. Dewhirst, P. Jackson, S. Bech, D. Meares, S. George, "Spatial Audio Quality Perception (Part 2): A Linear Regression Model", J. Audio Eng. Soc., vol. 62, pp. 847-860 (2014 Dec).

[31] P. M. Delgado and J. Herre, "Objective Assessment of Spatial Audio Quality Using Directional Loudness Maps," *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, 2019, pp. 621-625, doi: 10.1109/ICASSP.2019.8683810.

[32] J. G. Beerends, K. van Nieuwenhuizen, and E. vd Broek, “Quantifying Sound Quality in Loudspeaker Reproduction”, J. Audio Eng. Soc., vol. 64, pp. 784-799 (2016 Oct.).