

Requirements for loudspeakers and headphones in the "high resolution audio" era

Dr. Hans R.E. van Maanen
Temporal Coherence
www.temporalcoherence.nl
e-mail: info@temporalcoherence.nl

Abstract

The discussion about whether there is a perceived quality improvement with the "high resolution formats" for digital audio has been going on for a large number of years without having reached a final conclusion. Results of listening tests have resulted in conflicting conclusions, certainly for those tests which have mainly focused on the audibility of sonic contributions with frequencies above 20 kHz. Yet, there is a lot of circumstantial evidence¹ that the high resolution formats produce higher quality sound reproduction. In this paper we will bring forward a different explanation than the presence of ultrasonic components only which, however, will have important consequences for the requirements of loudspeakers and headphones (and microphones!) to bring the advantages of high resolution formats to full fruition.

0. Introduction

Ever since the coming of the CD as the digital audio format for the consumer, there has been criticism on its sonic quality. From numerous tests, evaluations and developments, it became clear that the digital format is not as straightforward as it seemed at first and that D/A linearity, absence of jitter and reconstruction filtering are crucial in the determination of the sonic quality of the CD (player). Especially the effect of the reconstruction filtering is -at first sight- surprising as the filtering occurs above 20 kHz, which is commonly accepted as the upper limit of hearing in humans (ref. 1).

A serious problem is that the underlying assumptions in the concept of the CD format are flawed, you might even say incorrect. It is true that humans are basically unable to hear continuous sine wave tones above 20 kHz. However, the conclusion that tones above this limit do not contribute to the sonic experience is "jumping to conclusions". Such a conclusion would only be correct for linear systems, but as we all know, human hearing is non-linear in virtually every aspect (refs. 2, 3 and 4). Therefore, a further and deeper analysis is necessary to determine the requirements for loudspeakers and headphones in the "high resolution" era.

In this paper, in section 1 we will review the current situation on the different digital formats and some of the different tests on the contributions of sounds with frequencies above 20 kHz and the high resolution formats, combined with our own experiences with these formats. In section 2 we will put forward a different explanation for the perceived differences and see whether this is in line with the available information in section 3. After discussion in section 4, we will derive the consequences for loudspeakers, headphones and microphones in section 5. We will come to our final conclusions and recommendations in section 6.

1. The deficiencies of the different digital formats

When we want to strive for the reproduction of sound in general and music in particular as close to reality as possible, we have to compare the reproduced sound with "live" sound. In order to avoid a

¹ In some publications, "circumstantial evidence" is called "anecdotal evidence". However, in Dutch, the native tongue of the author, "anecdotal" has the meaning of a funny story, so he prefers "circumstantial evidence" when the information cannot be called "scientific", but still has to be taken serious.

discussion about "live", it is defined in this paper as the sound as produced by instruments, voices and possibly other devices without electronic amplification or generation. So we will exclude all sound reinforcement systems, electric guitars, synthesizers etc. When comparing reproduced sound from a CD with "live" sound, several things can be noted:

- The level of detail in the reproduced sound is less than of "live" sound.
- Choirs and orchestras sound "fuzzy" (ref. 5).
- Metallic sounds from percussion like a triangle sound muffled.
- Reconstruction filters have been shown to have a major impact on the reproduced sound.

Detail in reproduced sounds is always a problematic aspect. In many reports in e.g. audio magazines, it is often mentioned how well (or not) a specific piece of equipment behaves in this respect. But using the CD format it is nowhere near the detail of a "live" concert. One visit to a good concert hall will convince you if you doubt this statement.

When listening to a CD, it often sounds quite pleasant as long as the music is rather "simple", e.g. just one voice and a few instruments. But as soon as it becomes a bit more complex (sometimes even within a single track!) the quality deteriorates. This reduction is very evident with the reproduction of the symphony orchestra and/or choirs. It is clearly less (but not absent!) with high resolution² recordings. But it is questionable whether this quality reduction can (solely) be attributed to intermodulation distortion in loudspeakers as this can also be heard with high-quality headphones. A typical example which points in another direction was the demonstration by Manger at an audio fair in The Netherlands: they played the same track in the CD format and at 96 kHz / 24 bit and the difference was obvious to everybody in the audience. Of course, you can have lengthy debates about the conversion from the high resolution format to the CD format, yet it is not easy to explain the difference between the two signals just using the "common wisdom" about human hearing.

The reproduction of percussion is a tough job. Actually, what I usually hear from a CD is a rather weak infusion of the actual "live" sound. The sound is muffled, the attack is notably deteriorated and there is insufficient detail to recreate the typical metallic sound of percussion. The best reproduction I have ever heard was from an analog recording using ionophones, the second best a ribbon tweeter³. Ordinary electrodynamic tweeters usually reproduce only a far cry from the "live" version. But with good tweeters (we will discuss later what "good" is in this respect) or a good headphone, the difference between a CD and a high resolution recording can be striking, although there are clear differences between recordings because of the experience of the Tonmeister.

Through the years, numerous concepts for the reconstruction filters of CD players have been tried. Although all try to avoid any degradation in the audio frequency band (20 Hz - 20 kHz), they all sound different. A nice example is the "Perfect Wave" D/A converter for the HRx format (a high resolution format running at 176 kHz sampling frequency). It lets the user choose between five different types of reconstruction filter and although these are only active above 20 kHz, all five sound different even to more senior listeners. Because the D/A conversion is the same, the audible differences can only be attributed to the reconstruction filtering.

In general, it can be stated that there is still a clear difference between reproduced sound and "live" sound. There must, of course, be an explanation for this phenomenon and this explanation will be essential to determine the requirements for sound reproduction systems in general and more specific for loudspeakers and headphones.

When we take a "high level" look at sound reproduction systems and the human hearing which receives and processes the sounds, we can notice several things:

² High resolution" will be defined as digital recordings with a sampling frequency of at least 96 kHz and 20 bit amplitude resolution. The comparisons, mentioned in this paper will, however, refer mostly to 192 kHz / 24 bit and SACD / DSD recordings.

³ I would like to mention that unfortunately I never had the opportunity to listen sufficiently long to the Manger loudspeakers to rank these for percussion reproduction. They might be comparable to the ribbons.

1. Any sound reproduction system is band-limited.
2. Any sound reproduction system introduces non-linear distortion.
3. Any sound reproduction system introduces linear distortion.
4. Human hearing is non-linear in virtually every aspect.

Ad 1.

It is, of course, banging on an open door that any sound reproduction system is band-limited. It starts with the microphone which picks up the sound and the microphone preamplifier which both have a limitation in the high frequency range. The carrier to the consumer is also a limiting factor: the CD breaks down at 20 kHz, vinyl comes at best to 50 kHz. Consumer amplifiers rarely have a cut-off frequency above 100 kHz and last-but-not-least the loudspeakers in general act as low-pass filters at or slightly above 20 kHz, although gradually tweeters and headphones with a wider bandwidth become available. But still, there is an upper limit. This has, however, two major consequences:

- Components of the original sound with frequencies outside the transmission band are suppressed. Do these components contribute to the sonic impression? The results so far are inconsistent (refs. 6 - 10). However, some critical remarks about this kind of tests can be made (ref. 11).
- The bandwidth limitation introduces a broadening of the impulse response. Is this broadening reducing the level of detail and/or increasing the perceived "fuzziness" of the reproduced sound (ref. 5)?

Unfortunately, there is a lot of misunderstanding on the relation between bandwidth limitations and impulse response widening. We will come back to this later.

Ad 2.

As we all know, no reproduction system is perfect, all introduce non-linear distortion. As a consequence, signal components are generated by harmonic and intermodulation distortions which are not present in the original signal. It is likely that these signal components will reduce the level of detail by masking subtle components of the original sound, but it is unclear to which extent. Distortion figures are very hard to interpret as the distortion of e.g. semiconductor amplifiers is often orders of magnitude smaller than those of loudspeaker units, yet can clearly be heard. On the opposite side, distortion figures of valve amplifiers are usually above those of loudspeaker units, yet these are often preferred over semiconductor amplifiers. The distortion figures of headphones are often small because of the small amounts of power, required to drive these, still the loss of detail occurs.

Ad 3.

Especially loudspeakers introduce linear distortion. As a consequence, the ratio between the harmonics of instruments is changed, which can, in some cases, give rise to a change in the tonal character of an instrument. But it is not likely to reduce the level of detail of the reproduction unless the higher frequencies are notably suppressed.

Ad 4.

Fourier theory is a well-founded mathematical framework for the analysis of signals in both time and frequency domains (e.g. refs. 12, 13 and 14). This enables the study of responses of systems to complex signals. However, such an analysis is only valid for linear systems and human hearing is non-linear in virtually all its properties (refs. 2, 3 and 4). Paradoxically, a non-linear system is able to extract more and probably better information from a signal than a linear system does, so one has to be very careful with the interpretation of signals in both time and frequency domains: the influence of the non-linearity has to be taken into account at every step in the analysis. As our ears process the incoming signals in time domain (ref. 15), Fourier analysis of the incoming signals without applying the non-linear response of human ears will lead to the wrong conclusions. An example is shown in figs. 2 and 3 in which the amplitude spectra of the signal of fig. 1 are shown without and with a non-linear responding system (not necessarily our ears).

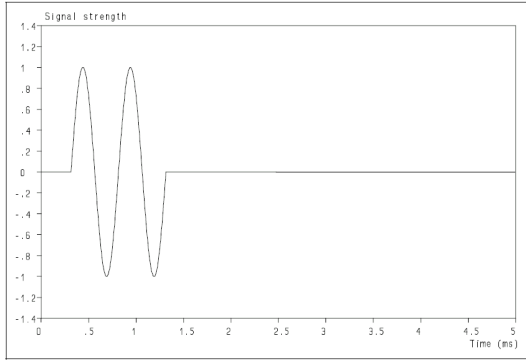


Figure 1: Test signal for demonstration of changes due to non-linearity (see fig. 2 and 3).

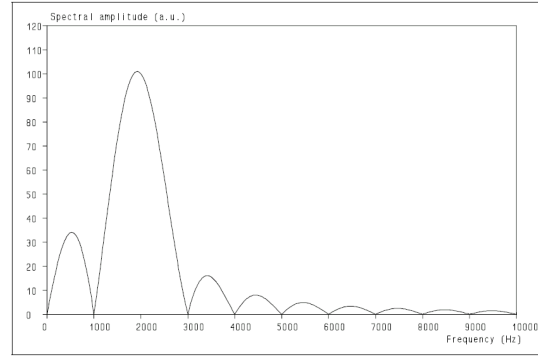


Figure 2: Amplitude spectrum of the signal of fig. 1 when no non-linearity is introduced.

As digital systems are also highly non-linear (quantisations both in time and amplitude!), one is in for a number of unpleasant surprises (ref. 1). If this is ignored, incorrect conclusions are easily drawn. An example is the sensitivity of the human ear for phase: when two *continuous* sine waves are added with a variable phase difference, shifting the phase does not lead to an audible difference. As in general phase has its influence on the *envelope* of a *non-steady* signal, the envelope of such a signal is modified when phase errors occur and this can be detected by a non-linear sensor like our ears, which has been confirmed by an experiment by Menno van der Veen (ref. 16). This may seem surprising at first sight, but is a logical consequence of the non-linearity of human hearing. But as there are many non-linearities involved, it is hard to predict the response of our ears to complex signals. However, it is safe to conclude that the measurement of a simple response to continuous sine waves is insufficient to come to correct conclusions.

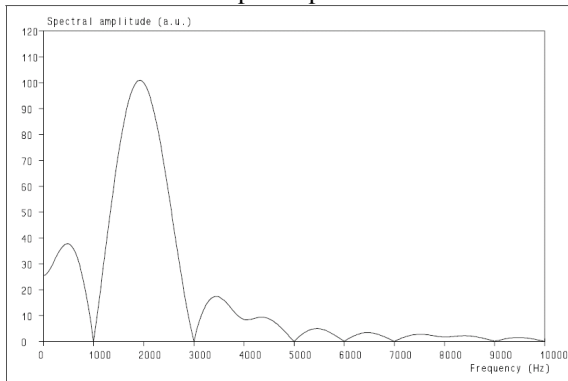


Figure 3: Amplitude spectrum of the signal of fig. 1 when non-linearity is introduced. Compare with fig. 2.

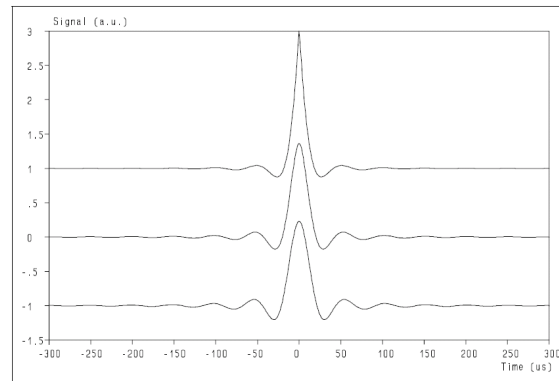


Figure 4: Impulse response of three different low-pass filters at 20 kHz but with different slopes. Top: lowest decay rate, bottom highest decay rate. For details: see text.

We will now address the widening of the impulse response due to the bandwidth limitation. There are many misunderstandings about this phenomenon. First of all, it is a direct consequence of the Fourier theory: it is impossible to maintain the impulse response in a band-limited system: a Dirac impulse requires an infinite bandwidth. As the output signal of a band limited system is the convolution of the input signal with the impulse response, the output signal will always be longer than the input signal. But it is not only dependent on the bandwidth itself, *it is also dependent on the way it is limited*. A steep roll-off above the cut-off frequency will lead to a wider impulse response as is illustrated in fig. 4. Here we compare three filters which have 0 (zero) attenuation between 0 Hz and 20 kHz and above 20 kHz roll off with 20, 40 and 60 dB/dec (2nd, 4th and 6th order low-pass filters). The phase shift is 0 (zero) for all frequencies. Although the bandwidths and phase characteristics of the filters are identical, their impulse responses are strikingly different. This is further illustrated in fig. 5, where the impulse responses of three Butterworth filters of orders 2, 4 and 6 are shown which are also different, but are all "causal". Again, the bandwidths are identical, but the widths of the impulse responses are not. Comparison of figs. 4 and 5 shows that the width of the impulse response of the phase linear filters is wider than of these causal filters. Also, the decay of these causal filters is after the impulse (where with "live" instruments also the decay of the sound occurs), whereas with non-causal filters "pre-ringing" takes place. We can compare these filters by using the "Temporal Decay" (ref. 17) of these filters, which results in the following table:

Filter order \ type	LP	BTW
2	0.36 dB/ μ s	0.72 dB/ μ s
4	0.25 dB/ μ s	0.48 dB/ μ s
6	0.20 dB/ μ s	0.35 dB/ μ s

Table 1: Temporal decays of the filters as shown in figs. 4 and 5. LP = Linear phase⁴, BTW = Butterworth. Table 1 clearly shows the above mentioned filter properties. Although there is no consensus on which type of reconstruction filter "sounds" the best, some reviewers have criticized the linear phase filters because of the pre-ringing properties which were blamed for an unnatural reproduction.

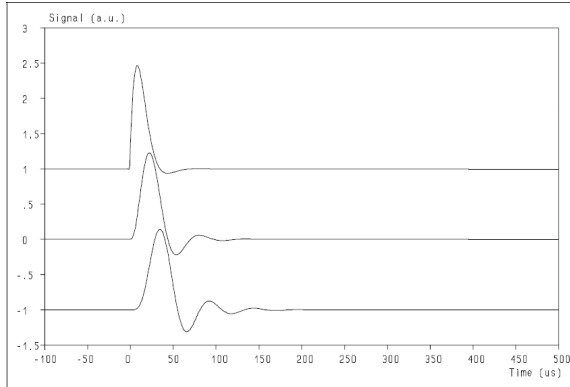


Figure 5: Impulse response of three different causal low-pass filters at 20 kHz but with different slopes. Top: lowest decay rate, bottom highest decay rate. For details: see text.

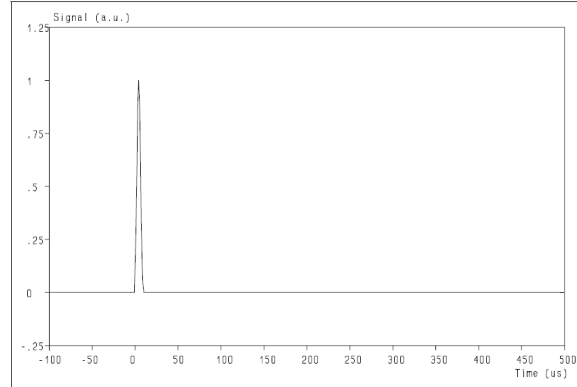


Figure 6: Impulse response of the 70 kHz low-pass filter without phase shift.

But this is not the whole story. Maintaining the bandwidth itself is insufficient to avoid widening of the impulse response. In fig. 6, we see the impulse response of a low-pass filter of 70 kHz which has a linear phase shift. In fig. 7, however, we see the impulse response of this filter, but it also introduces a phase shift identical to that of a 6th order Butterworth low-pass filter at 20 kHz. This filter creates a widening of the impulse response and thus time smear⁵, in this case with a temporal decay of 0.2 dB/ μ s, about as bad as an ordinary CD reconstruction filter (ref. 17). So in conclusion: a narrow impulse response requires

- a wide frequency response
- a moderate roll-off above the transmission band
- a phase response as close to linear as possible, but maintaining a causal response.

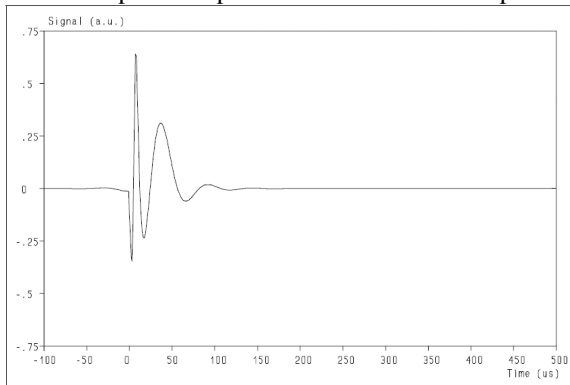


Figure 7: Impulse response of the 70 kHz low-pass filter with a phase shift of a 6th order Butterworth low-pass filter of 20 kHz introduced.

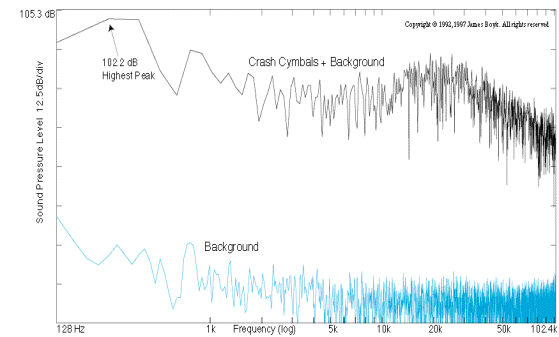


Figure 8(a) (Amplitude vs frequency) **Crash Cymbals**, 100.3 dB at 86k at B&K 4135 microphone with grid off, approximately one and a half feet away. Upper trace: Cymbals + background, connected to 100 kHz. Lower trace: Background alone.

Figure 8: Amplitude spectrum of cymbals (courtesy of J. Boyk).

⁴ "Linear phase" means that the phase shift is linearly proportional to the frequency. Zero phase shift filters are the most well-known, but a linear phase shift will only result in a time shift of the original signal without a change in shape.

⁵ "Time smear" is the phenomenon that due to the broadening of the impulse response each individual contribution to the signal is no longer attributed to a specific moment in time, but is spread out over a certain interval around this moment in time. So signal contributions which were originally separated are fused, thus degrading the quality of the original sound.

2. Possible explanation for the discrepancies

Overlooking the abundant amount of information, some remarks can be made:

- Tests on the contribution of signals above the hearing limit of 20 kHz to the sonic experience give conflicting results, although there are indications that these do contribute (ref. 18, who used senior listeners with a lower upper hearing limit).
- Many high resolution recordings have little content above 20 kHz (which is not really surprising as most studio microphones are band-limited at 20 kHz).
- Especially percussion instruments produce (lots of) signals with frequencies above 20 kHz. Cymbals actually produce most of their sonic energy above 20 kHz (see fig. 8, taken from ref. 19).
- The temporal resolution of human hearing far outshines its continuous sine wave response. Even more senior people are able to determine differences of as little as 6 μ s and this figure increases far less rapidly than the upper limit for continuous waves decreases with age (refs. 20 and 21).

In my view, all these findings indicate that the explanation of the observed phenomena rather needs to be found in the time domain and less in the frequency domain. Functions in time and frequency domains are not identical for non-linear systems, so there could be (and I think there are!) differences in the outcome for human hearing. Therefore it will be necessary to focus on aspects like the impulse response and the "time smear", introduced by the band-limitation of audio systems, recording and storage systems and the digital formats. Is my view in agreement with the experiences reported? This will be studied in the next section.

3. Experiences supporting the possible explanation

The earliest indication that the impulse response / time smear is important I got when I improved the response of moving magnet pick-up cartridges in the late '70's. I designed an analog processor to virtually eliminate the mechanical resonance of the magnet on the stylus (at around 19 kHz) which resulted in a major improvement in the perceived sound quality (ref. 1). As a bonus, the generally mentioned superiority of moving coil cartridges could easily be explained as these have a far better impulse response than the moving magnets, when operated as prescribed by the manufacturers.

The second clear message was the demonstration of the reproduction of percussion by an ionophone. So far, I have never heard a better reproduction and as the ionophone has the best impulse response because of its minimal moving mass, this is in full agreement with the hypothesis. Also, ribbon tweeters do a pretty good job; these also have a relatively small moving mass and a subsequently good impulse response.

The demonstration of Manger loudspeakers with the two versions of the same sound track also supports the hypothesis as Manger has paid a lot of attention to the impulse response of their mid/high range unit as is shown in fig. 9. This is in line with our own experience: lately we replaced the tweeters in our own systems with a newer type which has a better impulse response. Although we do not have a direct measurement of its impulse response, its characteristic can be approximated by a 4th order Butterworth low pass filter of 33.5 kHz as is illustrated in fig. 10. The resulting impulse response is shown in fig. 11 and it has a temporal decay of 0.80 dB/ μ s, far better than that of a CD reconstruction filter. This modification resulted in a clearly improved reproduction of e.g. percussion when using high resolution recordings.

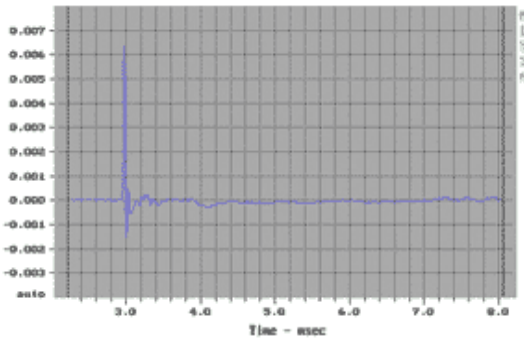


Figure 9: Impulse response of Manger unit (courtesy of Manger).

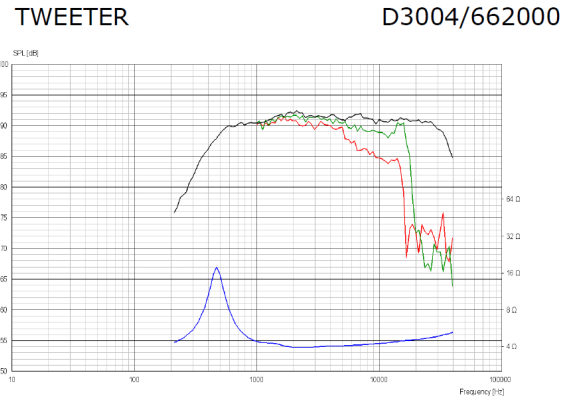


Figure 10: Response of the ScanSpeak D3004/662000 tweeter. 5dB/large division & 1 dB/small division (courtesy of ScanSpeak).

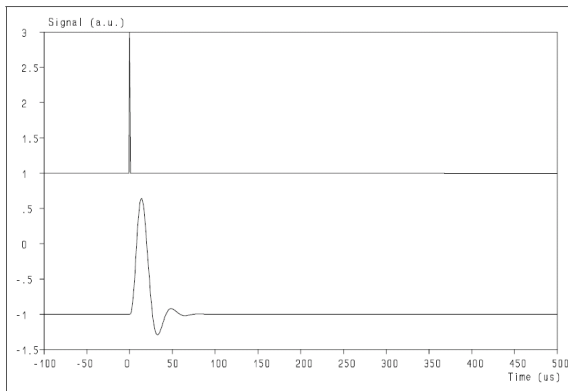


Figure 11: Impulse response (simulation by author) of the ScanSpeak D3004/662000 tweeter when approximated by a 4th order Butterworth low-pass filter of 33.5 kHz (as derived from the data of fig. 10).

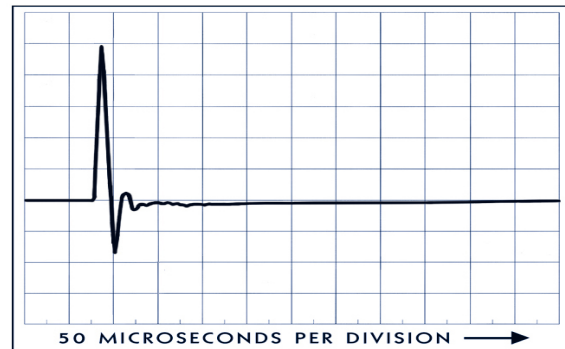


Figure 12: Typical impulse response of the earthworks M50 microphone (courtesy of Earthworks).

On the recording side of the playback chain, Earthworks, a microphone manufacturer, has independently arrived at the same conclusion and has focused on the impulse response of their devices as illustrated in fig. 12, which has resulted in a demonstrable improvement in the sound reproduction, especially clear (but not solely) with percussion.

The "fuzzy" sound which occurs with choirs and complex orchestra reproduction by the CD format can also be explained by the time smear, introduced by a broad impulse response. With the temporal resolution of the human hearing of 6 μ s or less, the decay of an ordinary CD reconstruction filter is in the order of 1 - 1.2 dB during this time lapse, so the different components are not separated, resolved, resulting in loss of detail. So it should not come as a surprise that the sound becomes "fuzzy", especially with complex music signals.

4. Discussion

The perceived differences between "live" and reproduced sound can be explained by the imperfect impulse response and the related "time smear" of reproduction systems. There are clear indications that the temporal response is more important than the frequency response. The -at first sight- surprisingly short time differences that the human ear can distinguish (refs. 20 and 21) can be used to estimate the requirements for sound reproduction systems. Taking the highest and lowest values for the temporal decay from table 1 and a time discrimination of 5 μ s, we find that the filters have decayed by 1 - 3.6 dB during this time. It should be obvious that certainly the decay of 1 dB is insufficient to resolve different sound contributions and is therefore likely to result in loss of detail and to smear out the "attack" of e.g. a triangle. Whether the 3.6 dB is sufficient remains to be seen (I severely doubt it), but the filter is certainly insufficient as reconstruction filter for a CD player. When looking at the problem from this angle, the earlier stated problems do not come as a surprise.

As can be seen from table 1, the phase linear filters have a temporal decay of roughly half of the Butterworth filters. And thus have an inferior impulse response, as can also be seen from figs. 4 and 5. So the inferior impulse response, in combination with the pre-ringing, could be responsible for the critical remarks about this type of reconstruction filters by the reviewers.

The temporal approach also provides an alternative explanation for the "fuzzy" sound of choirs and orchestras (ref. 5). Such complex sounds, the combination of numerous instruments and voices, require a very high temporal resolution to avoid that the inherent time smear of the system will "homogenize" the different sounds and the individual contributions can no longer be distinguished.

The reasoning can be reversed in order to get an indication of the required bandwidth of a sound reproduction system with inaudible deficiencies (ignoring non-linear and linear distortions for the moment). We found that a 33.5 kHz, 4th order Butterworth low-pass has a temporal decay of 0.8 dB/ μ s. Putting the temporal resolution of human hearing at 5 μ s and requiring a decay of 10 dB (an -at this moment- bit arbitrary value) during this time span, we need a temporal decay of 2 dB/ μ s, which is 2.5 times the value of the 33.5 kHz filter. So we would need an 85 kHz low-pass filter or better. Although this is not an accurate figure, the striving for an overall bandwidth of 100 kHz with a moderate roll-off rate above it is not unrealistic. For microphones and loudspeakers this has not been achieved yet, electronics are able to do this.

Any "theory" can be challenged and this is good practice in sciences. So, based on the above, predictions can be made. The most obvious is, of course, that an improvement of the impulse response (or, if you prefer, an increase in the temporal decay) should lead to an improvement in the perceived sound quality. Related to this is the prediction that less steep reconstruction filtering will lead to an improvement of the perceived sound quality. The Wadia converters use this principle already for tens of years and have been qualified as very good in literature. Finally, it can be predicted that choirs and orchestras should sound less "fuzzy" with systems with a better impulse response and these do sound better with high resolution recordings.

It is interesting to conclude that the extended frequency response is not required because our ears can detect waves of such frequencies, but *is required to resolve the reproduced sound sufficiently in time to satisfy the temporal resolution of our hearing*. This is a nice example where non-linear theory is required to explain certain phenomena.

5. Consequential requirements for loudspeakers / headphones

It is now no longer hard to formulate the requirements for loudspeakers, headphones and microphones in the "high resolution audio era": a lot more attention should be paid to the impulse response and consequentially the frequency range, the phase characteristic and the roll-off rate above the upper frequency limit. All three can be caught in the temporal decay figure and our crude estimate of 2 dB/ μ s should, for the time being, be the goal to strive for. This will be hard enough, but fortunately, we have seen progress in this performance and manufacturers can be stimulated to progress on this path. The availability of the components, described above, can hopefully pave the way.

An objection against the reproduction of ultra high frequencies is that these are very strongly directed towards the listener and thus further limiting the "sweet spot" where the sound is optimal (ref. 5). In my view, this is an indecent approach: if the strong beaming of certain sound contributions is a problem, the problem should be solved, not avoided by eliminating such contributions. At "Temporal Coherence" we have solved the problem by creating a 360° degree radiating loudspeaker as shown in fig. 13, which also has several other advantages. But also other clever designs could be applied, there is absolutely no reason to eliminate the reproduction of ultra high frequencies because of beaming. Yet, in general, the beaming is an aspect to be improved for loudspeakers.



Figure 13: Picture of the 360° radiating loudspeaker system of Temporal Coherence.

6. Conclusions and recommendations

The reproduction of signal components with frequencies above 20 kHz alone is insufficient to achieve a "natural" reproduction of music. When the "time smear" of the total system is too high, this contribution will go unnoticed and it might even be less realistic than a system which does not reproduce such frequencies, but has a higher temporal decay. It would be interesting to evaluate the Wadia converters in this respect.

The apparent differences between "live" and reproduced sound can be explained by looking at the impulse response of audio components. The characterization of the impulse response by the "temporal decay" is a (semi-)quantitative way to compare different impulse responses and helpful to steer developments in the right direction.

An estimation of the minimal required temporal decay, based on the available information, is 2 dB/ μ s and a related bandwidth of 100 kHz, but this needs (a lot more) research to establish these figures more accurately. For the time being it is a goal to strive for as we are currently far below these boundaries.

The beaming of loudspeakers at high(er) frequencies should be reduced, preferably eliminated, by clever designs or configurations to avoid small "sweet spots" for listeners.

The required bandwidth is not because human hearing is able to detect such high frequencies "as such", but to satisfy its temporal resolution, which degrades less with age than the upper limit of hearing.

"High resolution" formats should therefore be able to achieve at least a temporal decay of 2 dB/ μ s but it might be that this is easier achieved with a narrower pass-band width in combination with a moderate roll-off than the other way around. See e.g. the characteristics of the "minimized time smear filter" as described in ref. 17. This needs further analysis and development.

The development of loudspeakers, headphones and microphones which fulfill the above mentioned properties for temporal decay should be stimulated as much as possible in order to come to realistic reproduction of music.

References

1. Dr. Hans R.E. van Maanen, "On the audibility of 'high resolution' digital audio formats and how to test this", 18 February 2010, www.temporalcoherence.nl
2. V.M. Eguíluz, M. Ospeck, Y. Choe, A.J. Huspeth and M.O. Magnasco, "Essential Nonlinearities in Hearing", *Physical Review Letters* 84, 22 (May 2000)
3. Thomas Duke and Frank Jülicher, "Active Travelling Wave in the Cochlea", *Physical Review Letters* 90, 15 (April 2003)
4. W.S. Rhode, *Journal of the Acoustical Society of America* 49, 1218 (1971)
5. David Griesinger, "Perception of mid frequency and high frequency intermodulation distortion on loudspeakers, and its relation to high-definition audio", www.davidgriesinger.com/intermod.ppt
6. Kaoru Ashihara, Kenji Kurakata, Tazu Mizunami and Kazuma Matsushita, "Hearing threshold for pure tones above 20 kHz", *Acoust. Sci. & Tech.* 27, 1 (2006), pp. 12 - 19.
7. Ashihara Kaoru and Kiryu Shogo, "Detection threshold for tones above 22 kHz", AES Convention paper 5401, presented at the 110th Convention, Amsterdam (Netherlands) May 12 - 15, 2001.
8. Toshiyuki Nishiguchi, Kimio Hamasaki, Masakazu Iwaki and Akio Ando, "Perceptual Discrimination between Musical Sounds with and without Very High Frequency Components", AES Convention paper 5876, presented at the 115th Convention, New York, New York (U.S.A.) October 10 - 13 2003.
9. Kimio Hamasaki, Toshiyuki Nishiguchi, Kazuno Ono and Akio Ando, "Perceptual Discrimination of Very High Frequency Components in Musical Sound Recorded with a Newly Developed Wide Frequency Range Microphone", AES Convention paper 6298, presented at the 117th Convention, San Francisco, CA, U.S.A., October 28 - 31, 2004.
10. Wieslaw Woszczyk, Jan Engel, John Usher, Ronald Aarts and Derk Reefman, "Which of the two digital audio systems best matches the quality of the analog system?", AES 31st International Conference, London, UK, 2007 June 25 - 27.
11. Hans R.E. van Maanen, "The case for subjective listening tests", *Linear Audio*, Volume 4, September 2012, pp. 5 - 9.
12. A. Papoulis, "The Fourier Integral and its Applications", McGraw-Hill Book Company, New York (1962).
13. A. Papoulis, "Signal Analysis", McGraw-Hill Book Company, New York (1984).
14. V. Čížek, "Discrete Fourier Transforms and their Application", Adam Hilger Ltd., Bristol (1986).
15. Auditory demonstrations, IPO / NIU, Eindhoven Netherlands, Philips 1126-061, track 54, 55 and 56.
16. Menno van der Veen, private communication.
17. Hans van Maanen, "Temporal Decay": A useful Tool for the Characterisation of Resolution of Audio Systems?", paper no. 3480 (C1-8), presented at the 94th AES Convention, March 16 - March 19 1993, Berlin (Germany)
18. G.A. Briggs, "Loudspeakers", Fifth edition, October 1958, Rank Wharfedale Ltd., Idle, Bradford, Yorkshire, U.K.
19. J. Boyk, "There's life above 20 kilohertz! A survey of musical instrument spectra to 102.4 kHz", <http://www.cco.caltech.edu/~boyk/spectra/spectra.htm>.
20. Milind N. Kunchur, "Temporal resolution of hearing probed by bandwidth restriction", *Acta Acustica*, Vol. 94 (2008), pp. 594 - 603
21. Milind N. Kunchur, "Audibility of temporal smearing and time misalignment of acoustic signals", *Technical Acoustics*, Vol. 17 (2007)