# On the audibility of "high resolution" digital audio formats and how to test this

**Dr. Hans R.E. van Maanen**

*Temporal Coherence*

Date of issue: 18 February 2010

---

**Abstract**

The introduction of the "high resolution" digital formats for the recording of sound has initiated a lot of discussion on the audibility of the improvements, compared to the CD-format, which has been around for almost 30 years now. In this paper, some of the initial setbacks of the digital recording are memorized as well as the paradoxes it entailed. Some results from the analog era are used to illustrate that the CD-format has audible deficiencies and the paradoxes are explained by phenomena, which can be attributed to the anti-aliasing filtering and the interaction of the temporal and amplitude quantisations. It is shown that the "normal" definitions of Signal-to-Noise Ratio (SNR) cannot be applied in the same way for digital systems as for analog systems.

Listening experiences in "Temporal Coherence" with "high-end" audio equipment reveal the improvements of the "high resolution" formats over the CD-format which are especially clear in detail, a more realistic, crisp, sparkling, reproduction of percussion with more brilliance and the "ease" of the sound. However, because of the limitations of many sound reproduction systems, it is not always easy to prove the quality of "high resolution" formats unambiguously in a scientific way: often the improvements are masked by the limitations of the audio equipment, even the best ones, which are still inferior to the very discriminating human hearing. The problem is further enhanced by the large number of variables, like the choice of D/A conversion and type of reconstruction filtering, which presumably all have an influence on the perceived sound and thus easily lead to either an unpractical large number of experiments or an endless discussion whether the differences are to be attributed to the CD-format or its implementation or both. The use of "circumstantial" evidence should therefore be considered where the large volume can compensate for the less rigid test conditions.

Some ideas for carrying out comparisons using listening tests are presented which will help to resolve the discussion on the audibility of the "high resolution" formats. The availability of "high resolution" formats to the consumer is essential for future developments of audio systems: only if there is a clear gap between the quality levels of the source and of the reproduction system, further improvement of the latter is worthwhile. A simple, yet convincing, way to demonstrate the improvements needs to be found.

As the major limitation of the current audio equipment lies in the tweeter responses, both in frequency and time domains, manufacturers of loudspeaker units are encouraged to improve both the frequency response of tweeters to 40 kHz or higher and to shorten the impulse response of the units, so as to reveal the improvements of the high-resolution formats.

---

## 1. Introduction

Until roughly 1980, the large majority of the audio equipment was analog based. In the early ’70-ies of the previous century, the development of digital systems, initially for recording and storing purposes, were developed. The main formats of those days were the 48 kHz, 16 bit PCM format for professional recording and the 44.1 kHz, 16 bit PCM, better known as the Compact Disk or CD, for the consumer. Although the "digital revolution" was advertised as a major improvement in sound quality over the analog era, there were some problems. Not all consumers were convinced and up to this day, many prefer the sound from an analog (vinyl) record over the CD-format. The sound from the digital systems is qualified by such people as harsh, unpleasant, less "musical", compared to the sound from analog sources. Are all these people just mad addicts who inhaled too much marihuana in the flower power age, are these people gifted with more subtle hearing than most of us or is there more to it and do they really have a point? The given that vinyl is popular especially in the "high-end" audiophile range (and rising!) is a good reason not to dismiss this as just the opinion of a number of weirdos who got stuck in the previous century, so let us have a better look first before we make a judgement. In order to do this, we will first look back at some paradoxical results from the early days of digital sound recording and some experiences from the analog era. To understand these paradoxes, we need a bit of background on the usual imposed limits of audio systems, which we will subsequently use to confront the digital formats.

## 2. Initial problems with the digital formats

When the development of the digital systems started, the engineers were optimistic. From some simple calculations, it showed that it would be a piece of cake to outperform the analog counterparts. Wow and flutter would be something from the past, a flat frequency response would come naturally, copying could be done as often as required without any losses and the miserable 65 or so dB dynamic range that the professional analog recorders made available (which even required Dolby-A systems to get that far!) would be no competition for the digital systems: a 12 bit resolution is already equivalent to 72 dB. But to their surprise, a 12 bit system sounded so bad that it was no pleasure to listen to, diplomatically put. In one of the reports of those days the developer exclaimed that he was surprised that analog systems sounded so good, taking their specifications into account. Actually, he put it the wrong way around: he should have been surprised that the 12 bit digital system sounded so miserable and he should have found out why. Later on, I will show underpinning of these problems and their causes. In practice, it showed that at least 16 bits amplitude quantisation were required to bring the quality to an acceptable level, but one can still question whether it subjectively rivals the best analog systems. It also brings up the question whether the quality of the CD-format is so high that it generates no audible deficiencies. As mentioned above, there are plenty of people who think it is far from perfect. But how come?

## 3. Experimental results from analog audio systems

In the ’70-ies of the previous century, the best source, available to the consumer, was the long-play (vinyl) record. The electrical signal was generated by pick-up cartridges and the high-end market was divided in "moving magnet" (MM) and "moving coil" (MC) types. The basic distinction between the two is that with MM’s the magnet is moved by the mechanical undulations in the groove, the pick-up coil being static, whereas with MC’s the magnet is static and the coil moves. As even accelerated motions are relative (at least since 1915), one would not expect a systematic difference between the two types, but that is not the case. When it came to measuring, the MM’s were usually superior but when it came to listening, exactly the opposite was the case. How come? This remained a mystery to me for a long time until I found out that the upper frequency limit of MM’s is around 20 kHz and of MC’s around 80 kHz. The reason is that in MM’s the magnet, mounted on the stylus, resonates around 18 - 19 kHz and this increases the response in the range of roughly 10 - 19 kHz. By creating an LRC low-pass filter using the pick-up coil, the cable capacity and the Ohmic input impedance of the pre-amplifier, the overall response is more or less ironed out and the whole system acts a bit similar as a fourth order Butterworth filter, as shown in fig. 1. Because the moving coil in MC cartridges is very small, its

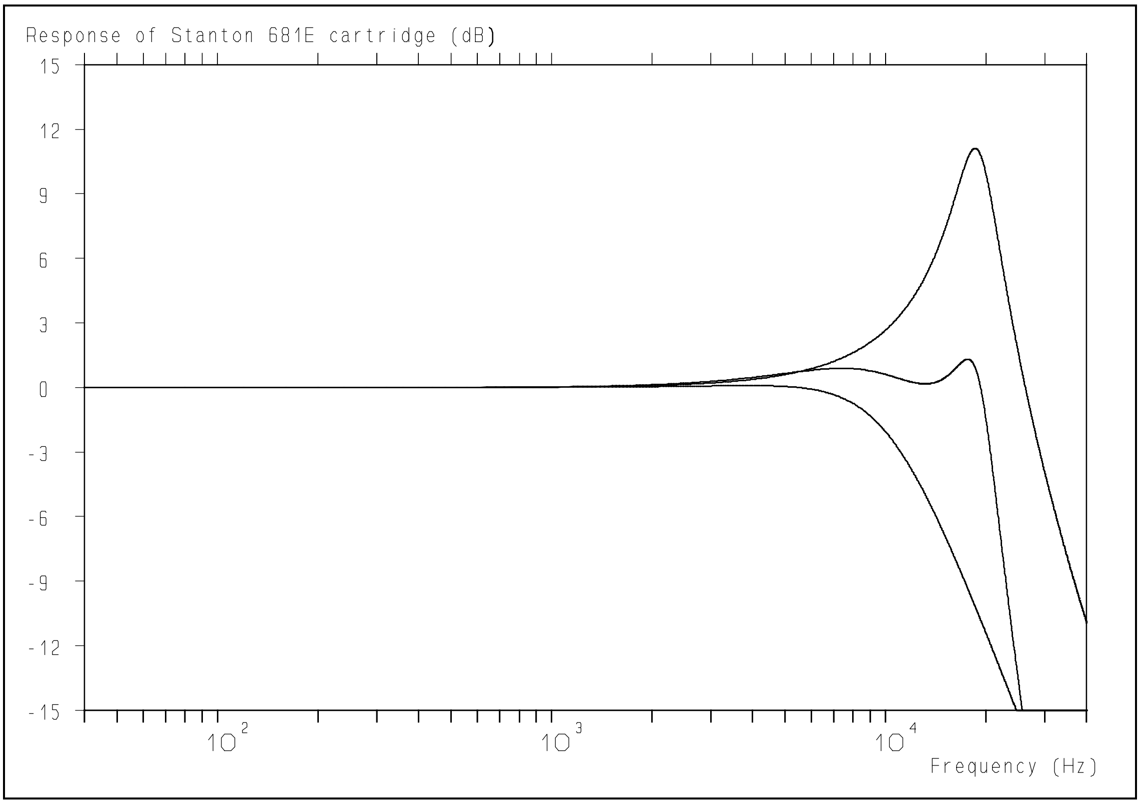

**Figure 1**: Response of Stanton 681 EEE cartridge. Lower curve = electrical response, upper curve = mechanical response, middle curve = overall response.

mechanical resonance lies at roughly 80 kHz and therefore no low-pass filtering by the electrical components in the audio band is applied. The "Butterworth" filtering as occurs with MM cartridges,however, has a dramatic effect on the temporal response and resolution of the reproduced sound. This is illustrated in fig. 2.

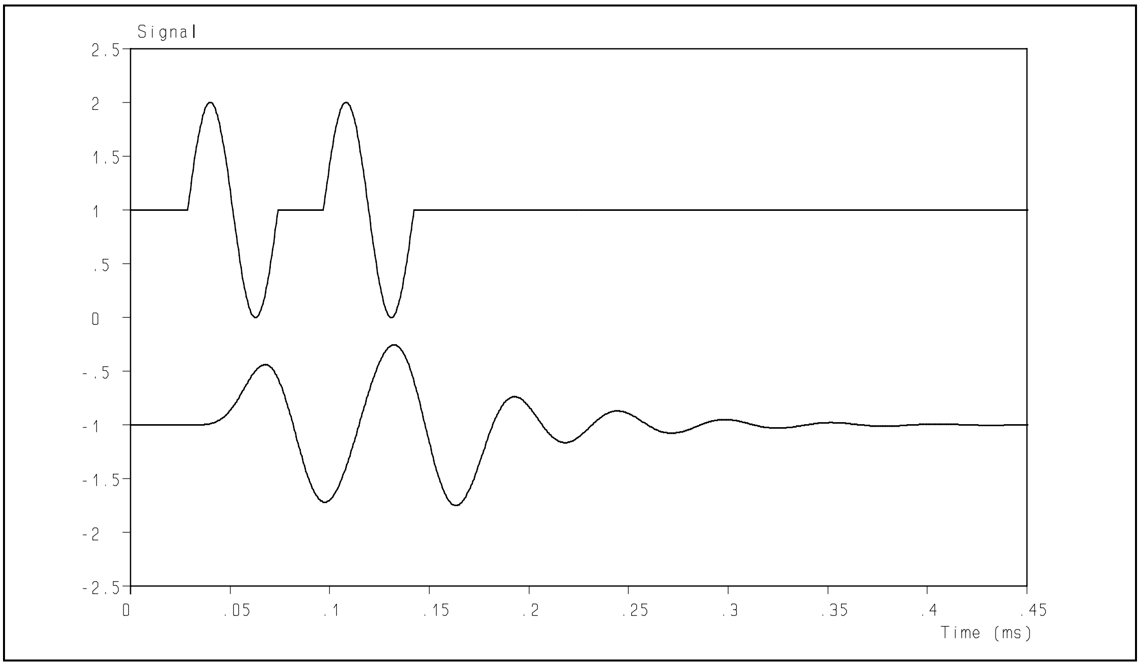

**Figure 2**: Response of a moving magnet cartridge to a complex signal.

But regarding the short interval in between the two individual sine waves (corresponding to 36 kHz), one can question whether this is audible or not. It is generally agreed that humans cannot hear continuous sine waves above 20 kHz, so one would expect these changes to the signal to be inaudible, but the generally agreed subjective difference between MM's and MC's points in the opposite direction. Because of this discrepancy, I was curious and I developed a correction amplifier to eliminate the mechanical resonance electronically as much as possible. Without going into details, I was able to improve the response to the level as is shown in fig. 3. This is not perfect, but it comes a lot closer to the desired

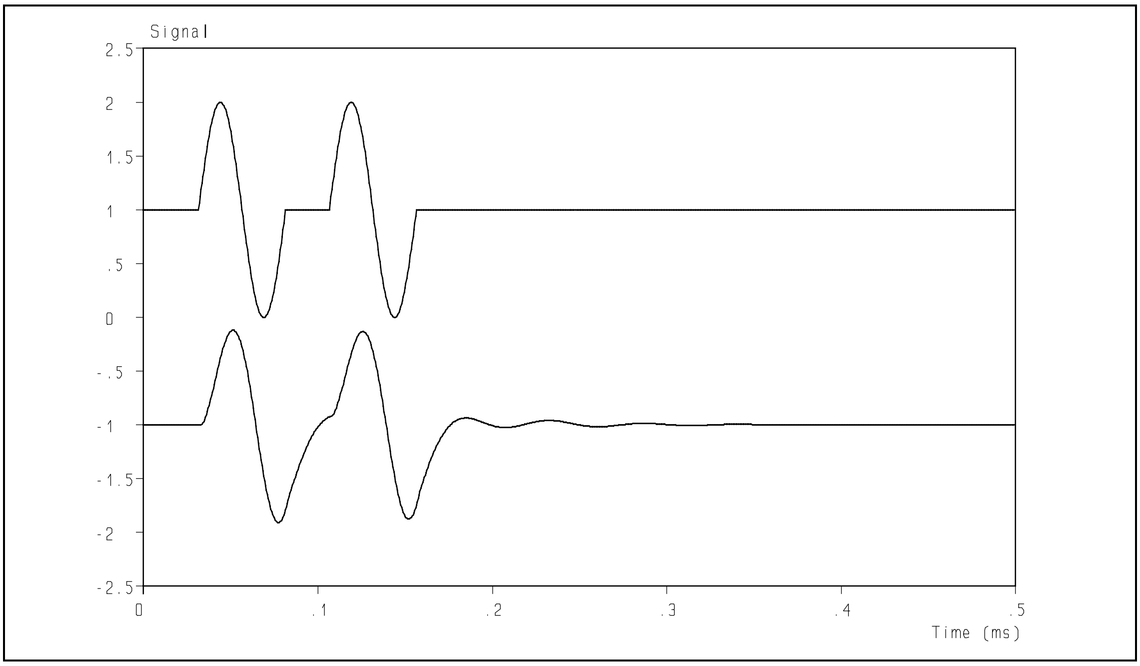


**Figure 3**: Response of a moving magnet cartridge to a complex signal after compensation.

response and the response of an MC cartridge. When I first heard the difference (**N.B.** there were **no** mechanical changes to the cartridge, this was solely a matter of analog signal processing!) I was stunned. The difference is not subtle, it is clearly audible and it is subjectively qualified as a significant improvement. These findings were confirmed by an A/B blind listening test in which all participants agreed within 30 seconds(!) on the major sonic improvement of the correction. I published the correction amplifier in a Dutch magazine (it can be found on our website www.temporalcoherence.nl for those who can understand Dutch) and the above mentioned findings have been repeated by many others[1], including an -in those days well known- reviewer of audio equipment. So we can conclude from this result that it is very likely that frequency components *above* the upper continuous sine wave hearing limit do contribute to the sonic impression, contrary to what many think. This, actually, has already been known for ages: in his famous book "Loudspeakers", G.A. Briggs (ref. 1) describes an experiment in which two elderly persons were tested: one was stone deaf above 10 kHz, the other above 11 kHz. But both could unambiguously tell when a 12.5 kHz low pass filter was inserted or not. So what is wrong with the reasoning of the upper hearing limit?

[1] In this respect it is interesting to note that around that time an MM cartridge was introduced which had a mechanical resonance of 24 kHz. It was by reviewers qualified as better than the "normal" MM cartridges, but not as good as MC cartridges. Application of the correction amplifier did still give an audible improvement, but -not surprisingly- not as spectacular as with "normal" MM cartridges. This further strengthens the evidence that signals above the continuous sine wave hearing limit contribute to the sonic experience.

## 4. The historical background of the upper sine wave limit

In the early years of the 19th century, a mathematical theory was developed by Jean Baptiste Fourier, in which he developed a transformation of a function of "*x*" into a function of "1/*x*" (refs. 2 - 4). The theory tells us that this transformation is a "one-to-one" projection. This means that any function of "*x*" has *only one* corresponding function of "1/*x*" and that no other function of "*x*" has the same corresponding function of "1/*x*", which means that also the reverse is true. In mathematical terms, the function of "*x*" and its transform (the function of "1/*x*") are therefore identical, albeit that they look completely different to ordinary human beings. The most common application of this theory (but certainly not the only one) is the transformation from "*t*" (time) to "1/*t*" (frequency). This transformation creates a lot of opportunities for the analysis and development of audio systems, but, unfortunately because of its complexity, also a lot of confusion. The application of the theory requires that a number of conditions are fulfilled and by now, too many people have forgotten when it can be applied correctly and when not. Let us see how some errors have crept in through the years.

One of the major consequences of the Fourier theory is that any[2] signal in time can be written as an infinite series of cosine and sine waves with frequencies from zero to infinity. It can also be expressed as an infinite series of cosine waves including a phase term. The advantage of this way to specify the Fourier series is that the signal in frequency domain can be split into the *amplitude* and the *phase* (ref. 2 - 4, see also the Appendix at the end of this paper). Numerous studies have shown that humans are not able to hear continuous (co)sine waves above 20 kHz, the upper hearing limit (which even decreases with age). The assumption that the contributions of the cosine waves above the upper hearing limit can be ignored or neglected seems obvious, but is this correct? Other tests have tried to determine the sensitivity of human hearing to phase by generating two cosines with different frequencies. Changing the phase relation between the two did not give any audible change in the sonic experience, so it was concluded that the ear is "phase deaf". Obvious at first sight, but correct?

One of the underlying requirements for the application of the Fourier theory is that it is applied to *linear* systems. For non-linear systems, first the non-linearity needs to be accounted for before the Fourier decomposition may be applied. As the human eardrum is highly non-linear (it distorts approximately 25 %), application of the Fourier decomposition before the introduction of the non-linear response of the human ear leads to erroneous results and thus to incorrect conclusions. An important (but not the only) consequence of the non-linearity of human hearing is that we are not only able to hear the actual sound, but also its *envelope*. It also means that we can hear intermodulation products -even if the two intermodulating frequencies are above the continuous sine wave hearing limit- when the difference frequency ends up in the audio band. Both phenomena are happily put to use by the ultrasonic disco sound system in which each listener can pick her or his own tune, which is radiated towards the person, modulated on an ultrasonic carrier wave (ref. 5). The ear demodulates the signal and the receiver hears the selected tune. So this proves beyond any doubt that sound with frequencies above the upper hearing limit are detected and processed by the human ear. I see no reason why this would only happen with carrier waves and not with sound waves above the upper hearing limit when these are part of a "normal" instrument (e.g. a triangle). There is also proof that other mechanisms in the human ear create a response to ultrasonic sounds and in some well done experiments, it was shown that the relation between time resolution and frequency response, which is valid for linear systems is invalid for human hearing: the temporal resolution of our ears is significantly better than the continuous sine wave limit suggests (refs. 6 and 7). This explains how sounds with frequencies above the upper hearing limit[3] contribute to the sonic experience. But what are the consequences for the phase? It is easy to create signals, which have the same spectral composition, yet have a different

[2] Of course, mathematicians have invented functions which cannot be transformed, but *all* sound signals, periodic and non-periodic, can be transformed.

[3] Some people claim that ordinary musical instruments do not or hardly produce any frequencies above 20 kHz. This is incorrect, especially percussion instruments like a triangle, cymbals, etc. produce significant amounts of energy above 20 kHz (ref. 8).

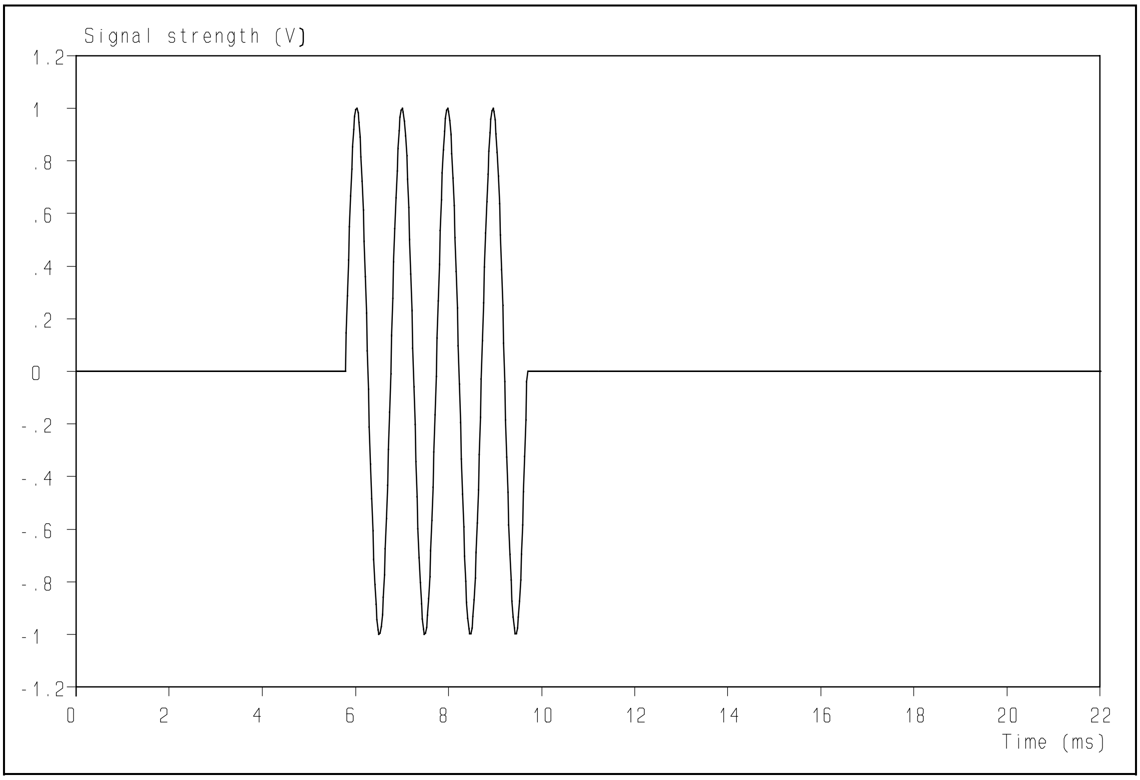


**Figure 4**: Signal #1.

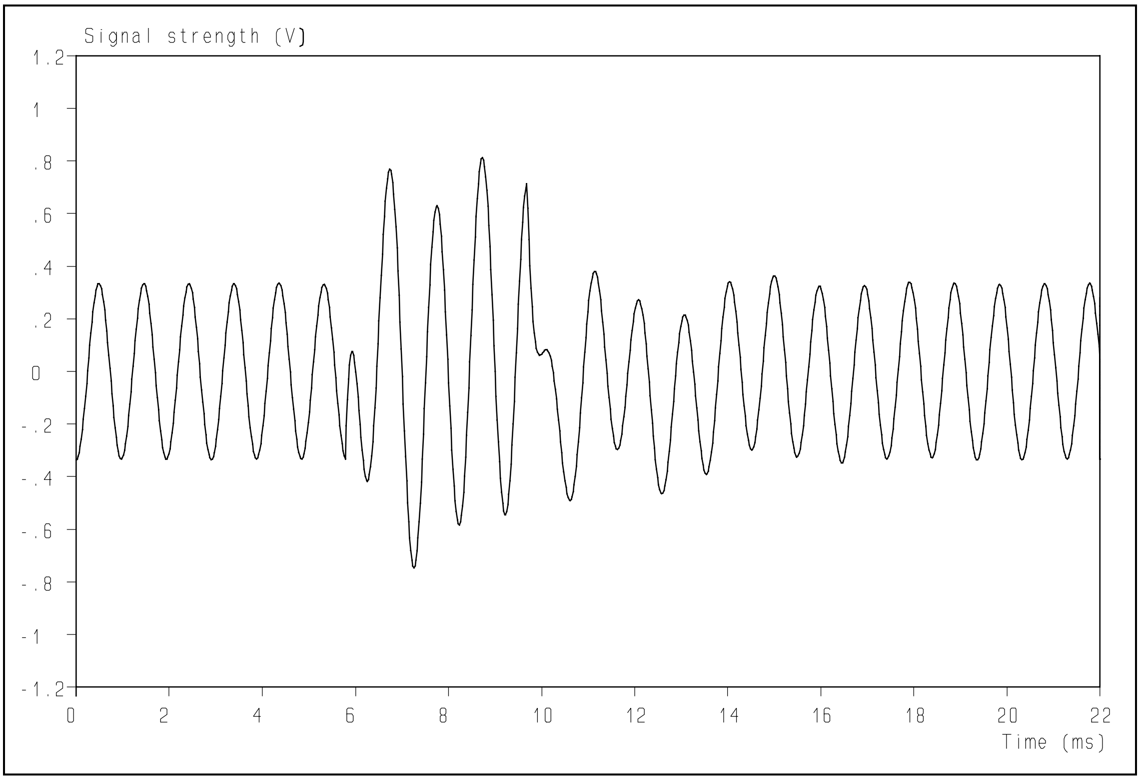


**Figure 5**: Signal #2, which has exactly the same spectral composition as signal #1, shown in fig. 4.

envelope, like the signals, shown in figs. 4 and 5 (for details: see Appendix). When the ear is completely "phase deaf", both signals should sound completely identical to us. Ask any designer of synthesizers and (s)he will tell you they won't. And that is because we also "hear" the envelope. In this, albeit indirect, way, the ear is sensitive to phase.

So there is a lot of evidence which shows us that sound waves above the upper hearing limit do contribute to the sonic experience, I think it even proves that the frequency response of high quality audio systems should be extended to at least 50 kHz, but preferably higher and with a moderate roll-off as the decay rate above the upper limit is of prime importance for the resulting "time smear" of the system (ref. 9).

### 5. Consequences for digital systems based on the CD-format

Can the above mentioned results be used to explain the subjective quality assessment of the CD-format? Partly, yes. Let us, as a first step, apply the same complex signal, used to reveal the responses of pick-up cartridges, to the anti-aliasing filtering of a CD-like digital format[4]. The result is presented in fig. 6. Ignoring the amplitude quantisation for the moment, a similar result is obtained for

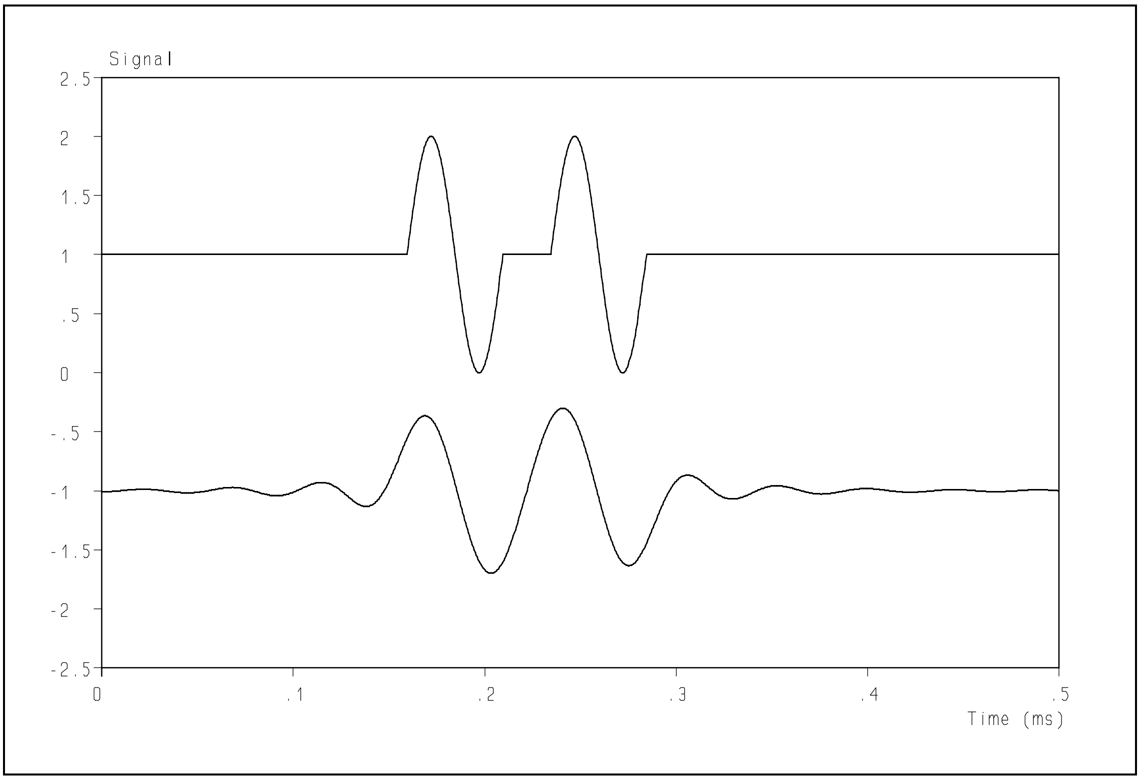


**Figure 6**: Response of a CD reconstruction filter to a complex signal.

the reconstruction filter. I hope you can understand that I would be very surprised if such effects on the signal would be very clearly audible in analog systems and completely inaudible in digital systems. My suspicions were confirmed when it turned out that the reconstruction filtering in CD players plays a major role in the subjective quality assessment. This has been the Achilles heel of CD players through the years. Note that at this stage we have not yet included the amplitude quantisation and the temporal quantisation (sampling in time domain). Let us look at the amplitude quantisation first.

---

[4] The anti-aliasing filter used has a transfer function which is 1 (one) from zero to 20 kHz, rolls-off with 360 dB/octave ($60^{th}$ order) above 20 kHz and has zero phase shift for all frequencies. Such a filter suppresses frequencies above 22.05 kHz sufficiently to reduce aliasing to low levels: suppression at 22.05 kHz = -51 dB, at 24.1 kHz = -97 dB.

The definition of the Signal-to-Noise Ratio (SNR) of a digital system is not really straightforward. One could argue that the amplitude quantisation introduces a sort of noise. We will define "noise" here as the difference between the actual signal value (in Volts) and its quantised value. In this way, the SNR can be determined. As with the PCM technique each bit increases the resolution by a factor of 2 (which equals 6 dB), the SNR of a digital system is roughly *N*, the number of bits of the amplitude quantisation, multiplied by 6 dB. So a 12 bit quantisation would be equal to 72 dB, 16 bit to 96 dB. For practical reasons, which I will not describe here, the commonly cited value is ($N$ - 1)·6 dB, which explains the 90 dB commonly listed for the CD-format. The 66 dB, remaining for the 12 bit version would still outshine the top of the consumer equipment around 1980, so why did it sound so miserable? The root cause of this paradox is the *interaction* between the steep anti-aliasing filtering on the one hand and the reconstruction filtering and the amplitude quantisation on the other. Or, if you prefer, an interaction between the quantisations in time and amplitude.

This can best be illustrated by an example using a simulation. I ran it in 1984 and published the results in a Dutch magazine, but it received very little attention (it can be found on our website www.temporalcoherence.nl for those who can understand Dutch). In the following discussion we will define "noise" as the difference between the signal *after* the anti-aliasing filter (regarding this as the input to the digitising) and the signal which is obtained after amplitude quantisation & reconstruction filtering[5] (effected by the digitising). It shows that the noise is significantly larger when low sampling frequencies are applied than with high sampling frequencies, *even though the same filtering is applied and all sampling frequencies fulfil the Nyquist / Shannon criteria*, as is illustrated in figs. 7 and 8.

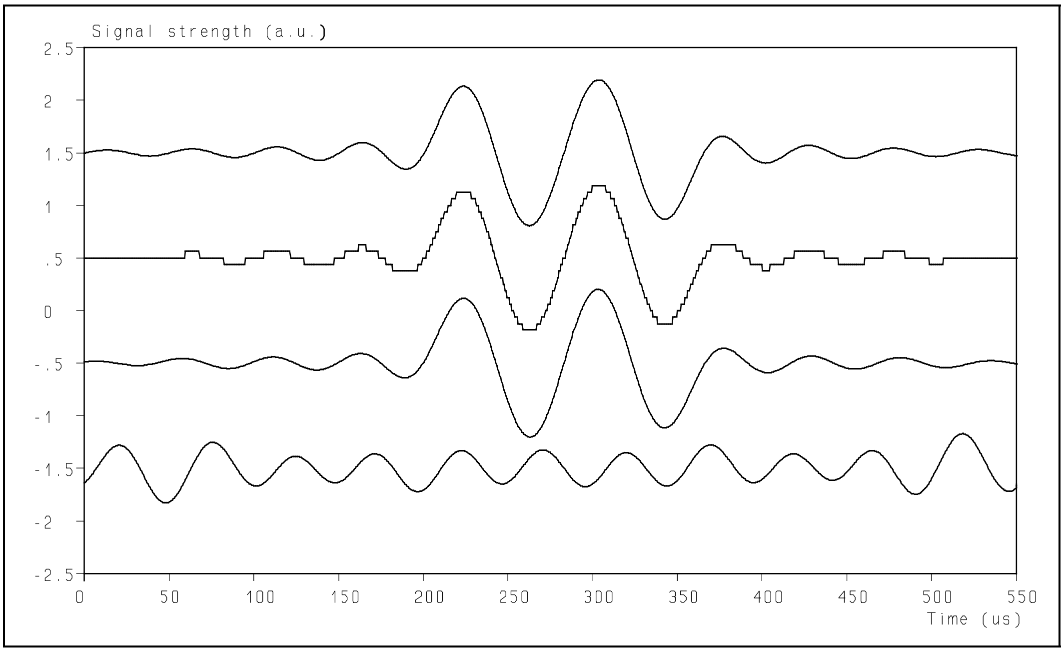


**Figure 7**: Signals in the digital signal processing. Sampling frequency 1800 kHz.

*From top to bottom:*
*Output of anti-aliasing filter (= input to digitizing system)*
*Output of amplitude quantisation (= input to reconstruction filter)*
*Output of reconstruction filter*
*Difference between input to digitizing system and output of reconstruction filter (x 10)*

The noises also have quite a different structure in spectral terms: the presence of noise in the 10 - 20 kHz range with low sampling frequency compared to the virtual absence of it with the high sampling

[5] To avoid problems, the reconstruction filter has a phase shift of zero degrees over the entire frequency range used.

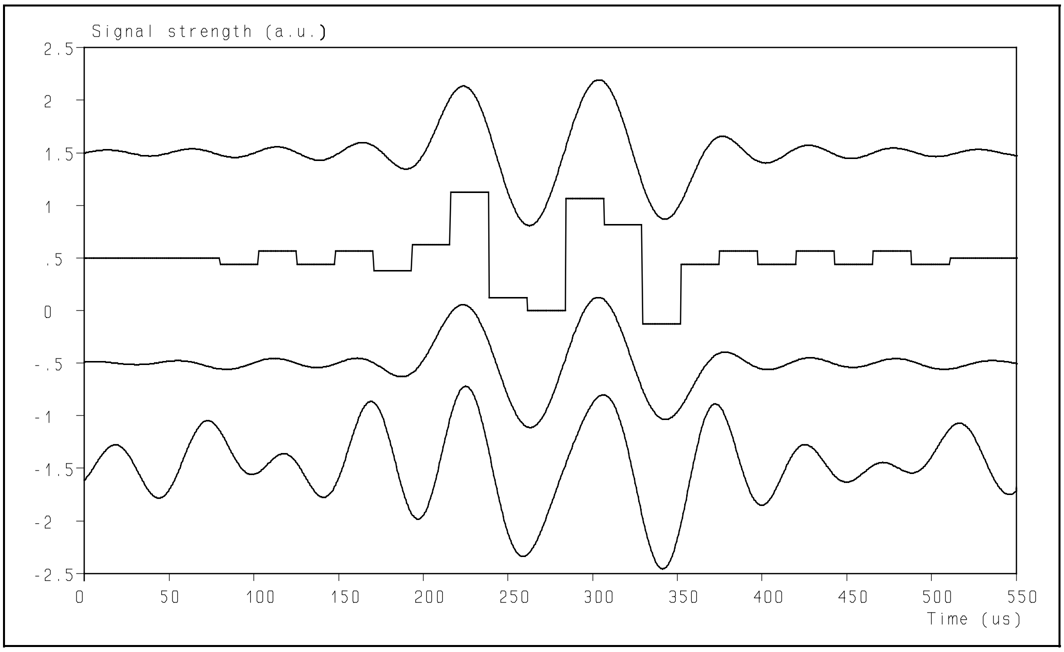


**Figure 8**: Signals in the digital signal processing. Sampling frequency 44.1 kHz.

*From top to bottom:*
*Output of anti-aliasing filter (= input to digitizing system)*
*Output of amplitude quantisation (= input to reconstruction filter)*
*Output of reconstruction filter*
*Difference between input to digitizing system and output of reconstruction filter (x 10)*

frequency is striking, as can be seen from figs. 9 and 10. Note that these phenomena show up with

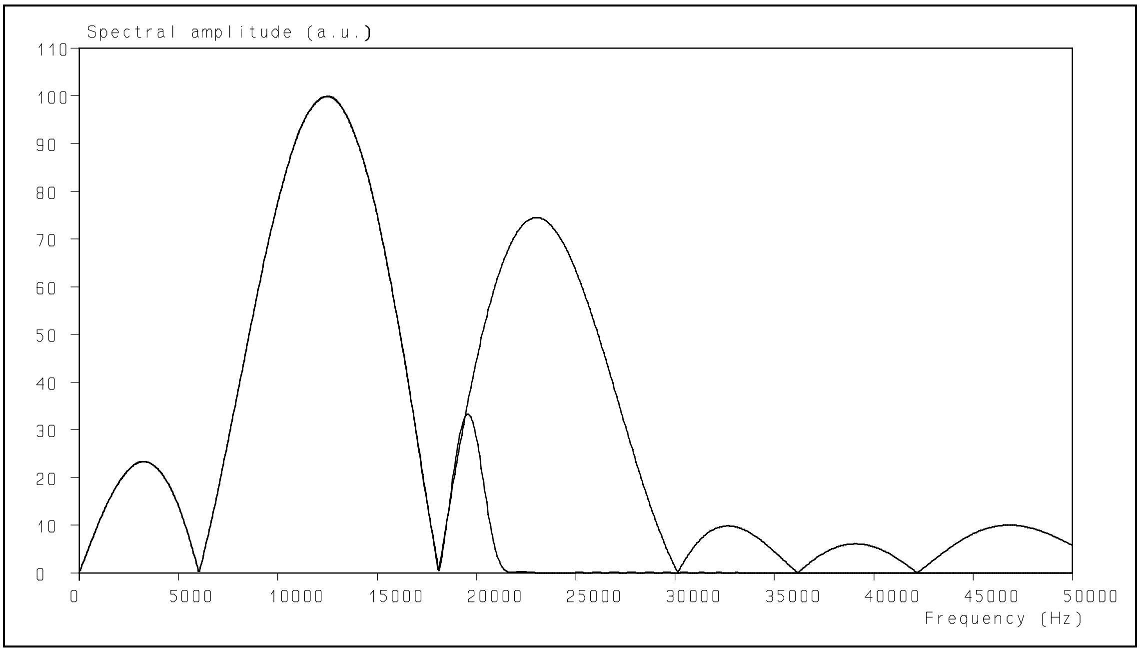


**Figure 9**: Spectra of the input signal (upper trace) and the output of the anti-aliasing filter (lower trace). Note that the frequency of the individual single sine waves (18 kHz) is absent in the spectra!

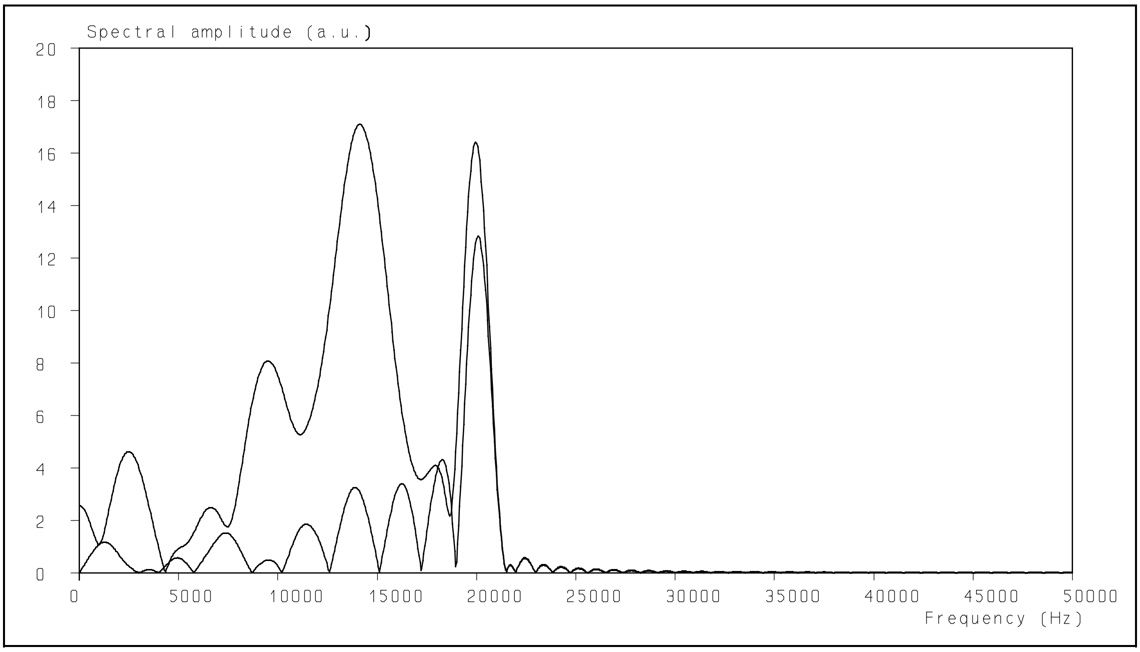


**Figure 10**: Spectra of the noise signal of the 44.1 kHz sampling (upper trace) and the 1800 kHz sampling, using the same anti-alias and reconstruction filtering. Note the vertical scale (compare with fig. 9, the units are the same!).

discontinuous signals and not with "infinite" sine waves. Fortunately, music is not monotonous, so the response to discontinuous signals is more interesting[6]. Under such "worst case" conditions, the increase in noise can amount 15 dB (and even more when "A" weighted)! So it is not really surprising that the 12 bit system sounded so bad: the noisy artifacts, created by the interaction of the amplitude quantisation & the reconstruction filtering are clearly audible and very disturbing. The introduction of another 4 bits will certainly reduce the problem, but claiming that these effects will then be inaudible is, diplomatically put, questionable. I am sure that such phenomena will have a negative effect on resolution and thus a negative impact on details in the reproduced sound.

The first time I heard a demonstration of the SACD was at a meeting of the Dutch section of the Audio Engineering Society (AES) in Eindhoven at the Philips premises. They played a track both from the CD and from the SACD on high quality equipment and the difference, including the above mentioned modulated noise, was clear to the audience. So the simulation is not just an academic result, it is there in real life (and explains why we need at least 16 bits). These results show that the application of linear theory to digital systems is incorrect.

This leaves us with the knowledge that the steep anti-aliasing filtering results in a time-smear of the input signal which is worse than with MM pick-up cartridges (which has been proven to be clearly audible) and that the interaction of amplitude and time quantisation leads to irregular modulated noise contributions, which are likely to give rise to loss of detail. To give a similar subjective quality as analog systems, digital systems require a wider theoretical dynamic range. It is, however, most questionable whether the currently used 16 bit / 44.1 kHz is sufficient to be inaudible using the top of the current reproduction systems. In my view (or should I say "in my hearing"?), it is not, as is indicated by the above mentioned examples and evidence.

---

[6] In several publications (e.g. ref. 10) is stated that the 16 bit resolution is enough to resolve sound at *any* level, provided correct dithering (the addition of noise with preferably a triangular amplitude distribution) is applied to the signal. Although this statement is theoretically correct, it has limited practical impact as this is only valid for infinitely long signals. The loss of detail in short impulses, as e.g. are common in percussion instruments, is *not* resolved by dithering.

### 6. Experiences with the "high resolution" formats

One of the most striking improvements when playing the "high resolution" format recordings is the ease, the relaxed way the sound is reproduced. It is far less tiring than playing a "normal" CD. In my view, this is caused by the reduction of the modulated noise and the strongly reduced level of time-smear: another 8 bit extra resolution and anti-aliasing and reconstruction filtering shifted by a factor 4 virtually eliminate the above listed problems, as can be seen from figs. 11 - 13. Yet these results show

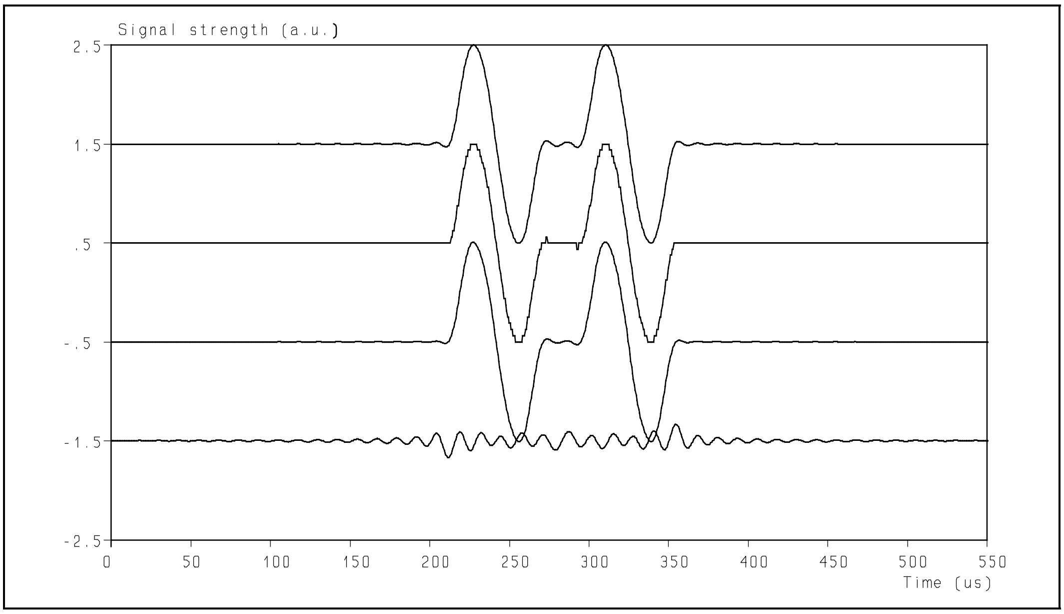


**Figure 11**: Signals in the digital signal processing. Sampling frequency 1800 kHz, anti-aliasing and reconstruction filters at 80 kHz.

*From top to bottom:*
*Output of anti-aliasing filter (= input to digitizing system)*
*Output of amplitude quantisation (= input to reconstruction filter)*
*Output of reconstruction filter*
*Difference between input to digitizing system and output of reconstruction filter (x 10)*

why SACD is superior to the 24 bit, 192 kHz PCM format, in agreement with the results of ref. 11, albeit that the 24 bit, 192 kHz PCM is far superior to the CD-format.

It is also clear that the non-continuous signals in the high frequency range are most vulnerable for these kinds of signal degradation (to put it diplomatically), like small bells, "silver threads", triangle, tambourine and the like as these also contain frequencies above 20 kHz (ref. 8). When played from an SACD, these sound brighter, more crisp, more detailed, sparkling and with more brilliance. So it should not come as a surprise that through the years, my ordinary analog recordings were rated superior to the CD version of the same recording (I am old enough to have both) even though I have used high quality CD players. This was even more clear with the half-speed master recordings[7] which were released in those days. To give an example: "Abraxas" from Carlos Santana was recorded in 1970 and the opening song is "Singing Winds, Crying Beasts". It is full of all kind of silverware, if you see what I mean, and I have four recordings of it: *(continued on page 13)*

[7] "Half speed master records" were cut using the original master tape at half the turntable speed. In this way, the high frequency response was improved by a factor of 2. In general terms, the amplitude response went to 50 kHz, the phase response was linear to 30 kHz, albeit that these specifications were slightly dependent on the manufacturer.

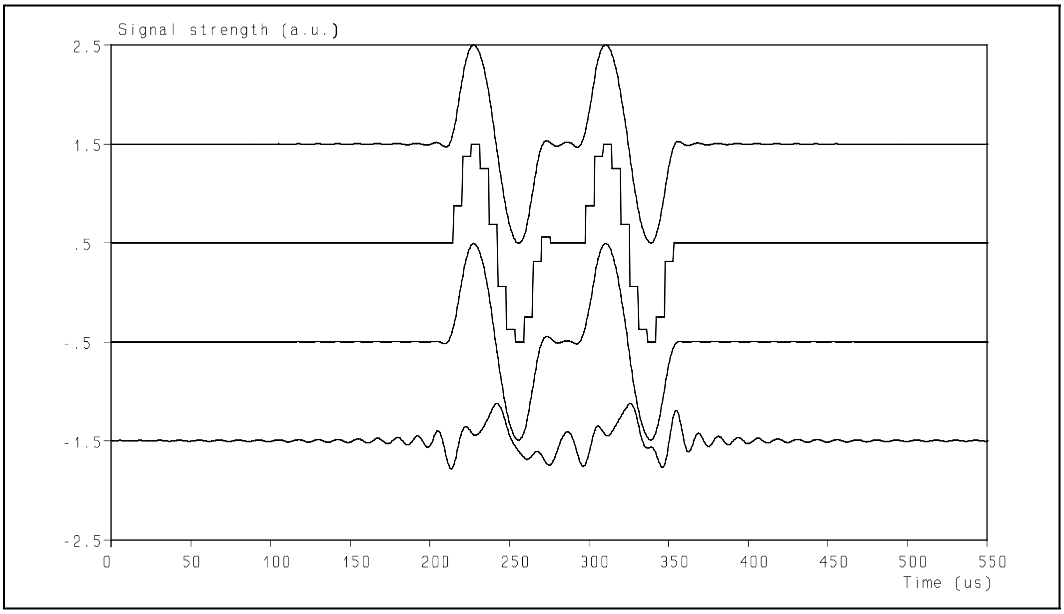


**Figure 12**: Signals in the digital signal processing. Sampling frequency 180 kHz, anti-aliasing and reconstruction filtering at 80 kHz.

*From top to bottom:*
*Output of anti-aliasing filter (= input to digitizing system)*
*Output of amplitude quantisation (= input to reconstruction filter)*
*Output of reconstruction filter*
*Difference between input to digitizing system and output of reconstruction filter (x 10)*

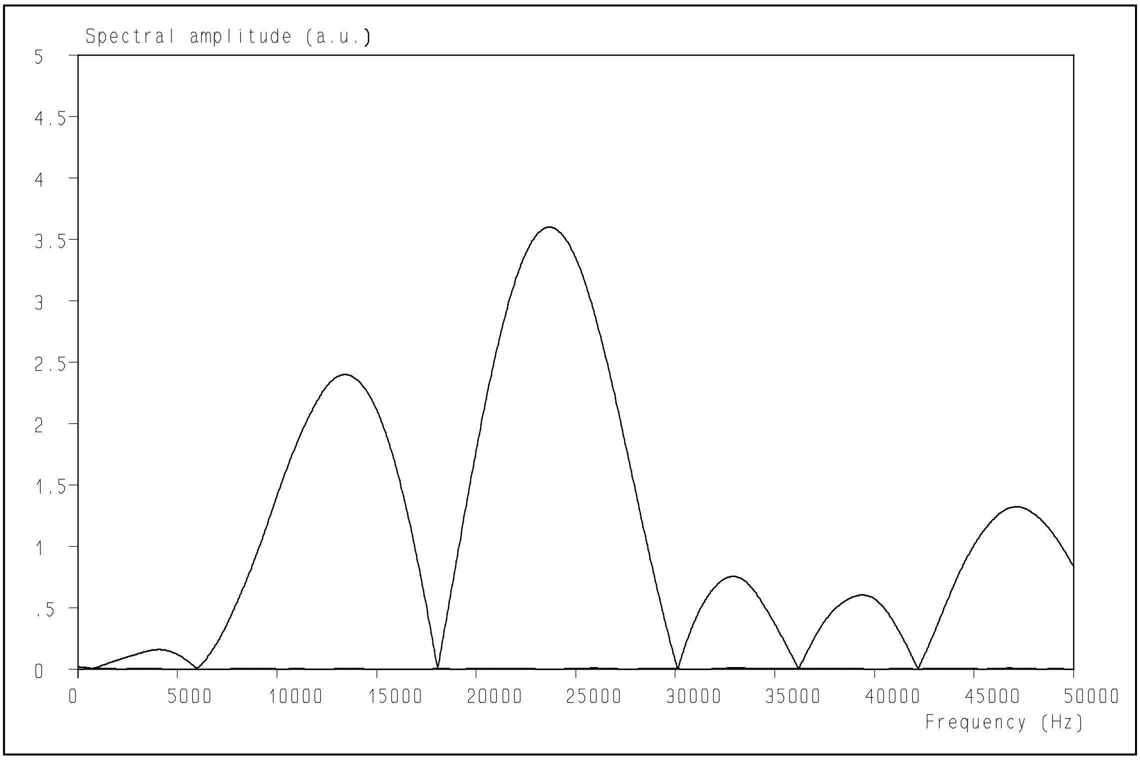


**Figure 13**: Spectra of the noise signal of the 180 kHz sampling, the noise of the 1800 kHz sampling is not visible. Note the vertical scale (compare with figs. 9 and 10).

- The original "normal" long-play record.
- The half-speed master long-play record.
- The CD version.
- The SACD version.

The subjective rating in quality is:

1. Half-speed master recording and SACD (maybe with a fully optimized analog playback system, the half-speed master might win)
2. The "normal" long-play record
3. The CD version

The differences are obvious to any listener with a reasonable listening experience and you may call this not scientifically proven, but this has been reported by so many listeners that I am not going to waste any time and money on it: the volume of evidence is so large that it is no use to deny it. And there is also significant scientific proof for the problems, limitations and audible deficiencies of the CD-format for digital systems (see above), so it should not come as a surprise that these are heard.

Before I continue, I want to put two statements on the table:

1. Currently, there is no sound reproduction system in the world which can reproduce what our hearing can detect. In other words: the performance of our hearing outshines the capabilities of *any* current sound reproduction system.

2. If there is not a significant gap between the quality of the sound sources, available to the consumer, and the reproduction equipment there is no real incentive to improve the quality of sound reproduction systems as better systems will only reveal the limitations and artifacts of the source more clearly.

<u>Ad 1.</u>

I am still amazed by the properties of our hearing and what it can reveal. I have been working on the improvement of audio systems for a long time and still further reduction of even small imperfections is audible, makes the sound more "natural", increases the level of detail which can be distinguished, etc. Yet, even the best systems which I have heard come nowhere near e.g. the live performance of a symphony orchestra. This is frustrating and I don’t expect that in my lifetime a system will be created which does come close. You should always keep this limitation in the back of your head.

<u>Ad 2.</u>

We, at Temporal Coherence, are designing, developing and building "high-end" audio systems for the consumer market and we ran into this problem: the limitations and artifacts of the CD become a nuisance when the quality of the reproduction system becomes high enough to reveal these to the listener. Customers are not really interested in systems which unveil the misery of the sound source so clearly. If the quality of the sound source becomes the limiting factor, it is basically the end of the development and we, the audio industry, would be shooting in our own foot. Therefore, we were very happy with the coming of the "high resolution" formats, albeit that to our experience the SACD was superior to the DVD-A format. But the subjective sonic improvement is large: the music is more "relaxed", more "musical", less stressed, metallic percussion instruments like the triangle, tambourine, etc. sound more natural, realistic, crisp, metallic if you want, and complex (multi-instrument) sounds like a symphony orchestra, reveal more detail. The high resolution formats create a significant widening of the quality gap between source and reproducer.

## 7. Listening tests to show the audibility of the "high resolution" formats

First of all, I want to challenge the "scientific" listening tests. With "scientific" I mean a test which is an attempt to unambiguously prove a certain phenomenon, contrary to what I call "circumstantial" evidence. With the latter I mean that there are clear indications that a certain phenomenon is real, but there is no "scientific" proof of it, yet the volume of evidence is so large that it cannot be denied. I will give an example later on. But "scientific" listening tests have a number of serious problems:

1. No sound system equals the quality of our hearing.
2. A listening test can only work on one single parameter
3. Usually, the number of independent parameters is large
4. Listening tests are time consuming and costly
5. There is (almost) always "that other explanation"

Ad 1.

In my view, the quality of the current sound reproduction systems is so low, compared to the abilities of our hearing, that statements that certain phenomena are "inaudible" are basically incorrect: the phenomena might not be resolved by the current audio systems and therefore go unnoticed. So the consequence of the limitations of the current systems is that no listening test can give a "negative" result: when a certain phenomenon is inaudible, it *cannot* be concluded that this is because of a limitation of our hearing, it might as well be a limitation of the sound system, used for the test. The author of refs. 6 and 7 ran into similar problems and he had to improve his reproduction equipment significantly to avoid this pitfall. Let me elucidate this a bit further. In 1993, the AES conference in Berlin was dedicated to "lossy" compression techniques. All of the presenters claimed that their compression technique was "inaudible". But one of the presenters came with a technique which threw away details which I knew, from my own experience, are clearly audible on our systems. The point was that it was "inaudible" on the audio system he used for the trials, but to generalise this statement is incorrect. By now, we all know that MP-3 introduces artifacts which are clearly audible on good audio systems, especially the lack of depth of the stereo image is striking. This is caused by the loss of detail in the reproduced sound. But at the introduction, it was claimed to be "inaudible". Maybe it was on the audio systems of those days (which I doubt), but not the current ones.

Ad 2.

In order to avoid confusion, during a listening test the effect of only one single parameter can be investigated. If not, differences in the experienced sound could be attributed to either parameter or to the combined effect of the separate changes of different parameters.

Ad 3.

This point can best be illustrated by the problem at hand. When we want to determine whether the CD format introduces audible deficiencies, we might want to send an (analog) signal through the "CD-chain" of signal processing. Which includes the anti-aliasing filter, the A/D conversion, the D/A conversion and reconstruction filtering. So we have already four independent parameters which can influence the result. But there is no "the" anti-aliasing filter: several different concepts with their specific pro's and con's are available (e.g. causal, non-causal, minimised time smear, maximum suppression of mirror frequencies, etc.). But for each concept, numerous realisations have been built. The same holds for the A/D converters, for the D/A converters and for the reconstruction filters. Now let us restrict ourselves to three concepts of each with each two different realisations. Which means that we have 6 different anti-aliasing filters, 6 different A/D converters, etc. To compare these all would require 6 x 6 x 6 x 6 = 1296 tests, which is completely unrealistic. But still, even if all these tests were done and the conclusion would be that the CD format does introduce degradation of the sound quality with all combinations, somebody could easily argue that her combination of CD processing, not included in the series of tests, is better and eliminates the degradation. In other words, because of the large number of independent parameters, it is impossible to unambiguously conclude that the CD format introduces an audible degradation of the sound quality.

Ad 4.

If listening tests could be set up and done in half an hour, some extensive test programs (see above) might be performed. In reality, listening tests require a lot of time to set up and to perform correctly and are therefore costly. So the volume of "scientific" listening tests is small and certainly does not cover the whole arena.

Ad 5.

The number of "scientific" listening tests which are not disputed is very small. There is (almost) always "the other explanation". To give an example: when it is determined that frequencies above 20 kHz do contribute to the sonic experience, the "other" explanation is that the non-linearities in the amplifier and / or the loudspeaker have generated intermodulation products below 20 kHz and thus explain the difference. Such arguments, correct or not, disable virtually the validity of (almost) all "scientific" listening tests as it is next to impossible to eliminate all imperfections in a test, with a few very well done tests as exceptions (refs. 6 and 7).

It is easy to see that because of the problems, listed above, "scientific" listening tests often encounter scepticism because of the ambiguity in the outcome or the interpretation of the outcome. So what else can we do? In my view, there is a lot of "circumstantial" evidence around which can be used for this purpose. Let me come back to the example of the MM and MC cartridges. To the best of my knowledge, a "scientific" test has never been done to distinguish between the perceived quality of MM and MC cartridges. Because of the huge number of cartridges this would have been completely impossible anyway (see "Ad 3" above). What happened was that the reviewers honestly reported what they heard and gradually the systematic separation between the two types surfaced. The technical explanation, as described in sec. 3, came afterwards. Until that moment, anybody could have argued that there was no "scientific" proof that MC's were superior to MM's and based on that "argument" discard the reports of the reviewers. Yet, in the end, it showed that the reviewers were right without a "scientific" listening test ever performed. So by analysing the "circumstantial" evidence, a lot can be learned. Therefore I respect the comments from everybody and I try to find the reason(s) of their criticism(s) and not to dismiss it because it is "not scientific". This has, through the years, provided a lot of information which we have used for further improvement of our systems. And my own "reference" is and remains the "Concertgebouw" in Amsterdam, even though it is absolutely useless for "scientific" listening tests. Yet I can guarantee you that no audio system comes near the live performance of a symphony orchestra!

The problem we are facing when trying to do a listening test in a "scientific" way is similar to the question how to find out how sharp our eye-vision is when you only have blurred pictures, covered with a haze, available for the test. The probability of misjudgment is pretty high. So in order to settle the question about the audibility of the improvements of the "high resolution" formats with such a test, we will have to make sure that we use the best available camera, lenses with the highest possible contrast, the CCD chip with the highest resolution and a professional photographer or we will just be fooling ourselves. If these requirements are not fulfilled, there is no use in doing such a test to begin with.

So how can such a test be done in a useful way, i.e. what are the requirements for the audio system which will be crucial in the comparison? In my view, these are:

- A linear frequency response up to at least 50 kHz with a moderate roll-off above the upper limit of the tweeter.
- A very short impulse response of the tweeter to minimise the time smear as much as possible. Ionophones are probably the best choice for this requirement, but these might fail with the following requirement.
- The change-over between different loudspeaker units must be realised in such a way that the time smear, introduced by this change-over is minimised. Time smear is introduced by cross-over filtering (refs. 12 and 13, see also fig. 14, taken from ref. 13) and transit time differences between the units and the listener. Because the latter is to a large extent dependent on the actual size of the loudspeaker unit and the cross-over frequency, this limits the choice of especially tweeters.
- The Sound Pressure Level (SPL) at the listening position should be 90 dB or higher without a high level of distortion, certainly no clipping (neither "hard" nor "soft") at any time during playback of the signals used for the test.

The background of these requirements is obvious when you look at the properties of the "high resolution" formats: extended frequency response and strongly reduced time-smear and noise, compared to the CD-format (ref. 11). But this is still not be enough......

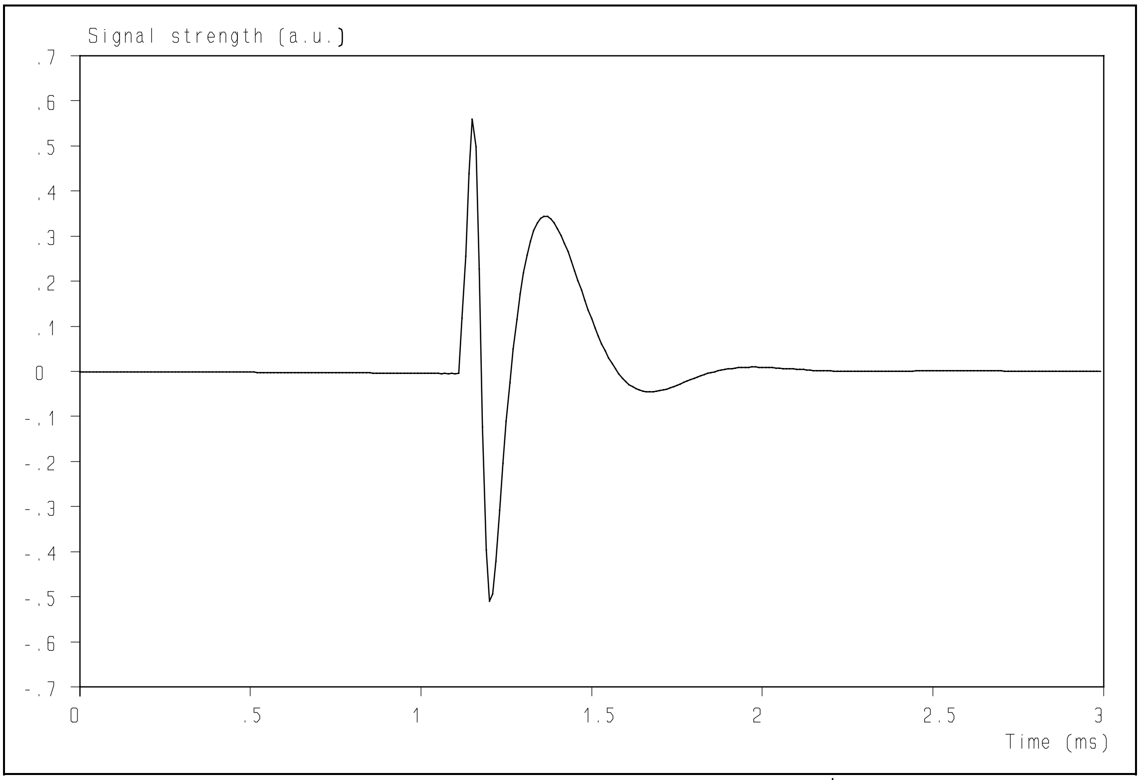

**Figure 14**: Dirac impulse response of idealised two-way system with 3rd order Butterworth filtering and the units operating in phase.

It also requires a listening team which is experienced and has a wide base of "live" reference[8]. A large variety of music selections needs to be compared and ranked for quality. As others have more experience in setting up listening tests in a "scientific" way, I will refrain from further suggestions.

### 8. Conclusions

There is quite a large volume of "circumstantial" evidence that ultrasonic frequencies contribute to the sonic experience in human hearing. This evidence is further supported by experiences from analog audio equipment. The contribution of ultrasonic sounds to the sonic experiences is not surprising when the non-linear properties and other complex mechanisms of the human ear are taken into account. But ingenious additional experiments have shown that the non-linearity of our hearing is insufficient to explain all its additional abilities. This will require more research, but for now, we can assume that we will be in for more surprises.

The interaction of amplitude and temporal quantisations lies at the root of unpleasant and masking "noise" contributions with the CD-format. This is confirmed by listening to CD’s and the major consequence is that CD-like formats require a significantly larger SNR than comparable analog systems. It also invalidates figures of SNR which are based on continuous sine waves as this kind of noise is strongly modulated by discontinuous signals.

Listening tests with "high-end" audio equipment have provided sufficient "circumstantial" evidence to show that the high resolution formats result in a significant improvement of the sound quality over the CD-format, completely in agreement with the above stated results and the outcome of some very well done "scientific" tests.

---

[8] With "live" sounds I mean pure mechanical instruments and voices, none with electronic amplification. Classical music is perfect, a modern pop concert forbidden.

Listening tests require the use of the best sound reproduction equipment, both in frequency and time domains. Especially the time-smear of most systems is an underestimated problem, which can easily mask the phenomena one wants to detect. However, very few "scientific" listening tests are undisputed because of a variety of reasons. Therefore, a simple, yet convincing, way to demonstrate the improvements need to be found.

The availability of sound sources to the consumer which are significantly better than the sound reproduction systems is a prerequisite for the improvement of consumer systems. But in order to improve on the sound reproduction systems for the consumer, we especially need better tweeters which have a wider range and can reproduce frequencies up to 40 - 50 kHz and have a very short impulse response. Manufacturers are encouraged to develop such units a.s.a.p.

As a final statement I would like to propose to ban all "lossy" compression systems as there is no need for it (anymore) and it brings up people with a flawed reference, which will hamper their judgement.

## References


1. G.A. Briggs, "Loudspeakers", Fifth edition, October 1958, Rank Wharfedale ltd., Idle, Bradford, Yorkshire, U.K.

2. A. Papoulis, "The Fourier Integral and its Applications", McGraw-Hill Book Company, New York (1962).

3. A. Papoulis, "Signal Analysis", McGraw-Hill Book Company, New York (1984).

4. V. Čižek, "Discrete Fourier Transforms and their Application", Adam Hilger Ltd., Bristol (1986).

5. www.ted.com/talks/lang/eng/wwody_norris_invents_amazing_things.html

6. Milind N. Kunchur, "Temporal resolution of hearing probed by bandwidth restriction", Acta Acustica, Vol. 94 (2008), pp. 594 - 603

7. Milind N. Kunchur, "Audibility of temporal smearing and time misalignment of acoustic signals", Technical Acoustics, Vol. 17 (2007)

8. 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.

9. 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)

10. J. Robert Stuart, "Coding for High Quality Digital Audio", www.meridian-audio.com/ara/

11. Derk Reefman and Erwin Jansen, "Signal processing for Direct Stream Digital", Internet publication, www.superaudiocd.philips.com/Assets/Downloadablefile/wp-2323.pdf (Dec. 2002).

12. Dr. Hans R.E. van Maanen, "The audibility of the temporal characteristics of audio systems", 14 May 2006, www.temporalcoherence.nl

13. Dr. Hans R.E. van Maanen, "Time smear at unexpected places in the audio chain and the relation to the audibility of high-resolution recording improvements", 18 January 2009, www.temporalcoherence.nl

## APPENDIX

Generation of signals with the same amplitude spectrum, but different in time domain

---

A signal can be decomposed in its (co)sine constituents using the Fourier theory (ref. 2 - 4). The basic equation (for the sake of simplicity ignoring scaling factors of the time axis) is:

$$f(t) = \sum_{1}^{\infty} a_n \cos(2\pi\ n\ t) + b_n \sin(2\pi\ n\ t) \qquad \textbf{[A.1]}$$

in which:

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos(nt)\ dt \qquad \textbf{[A.1}^{\textbf{a}}\textbf{]}$$

and

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin(nt)\ dt \qquad \textbf{[A.1}^{\textbf{b}}\textbf{]}$$

Eq. [A.1] can also be re-written as:

$$f(t) = \sum_{1}^{\infty} c_n \cos(2\pi\ n\ t + \phi_n) \qquad \textbf{[A.2]}$$

in which:

$$c_n = \sqrt{a_n^2 + b_n^2} \qquad \textbf{[A.2}^{\textbf{a}}\textbf{]}$$

and

$$\phi_n = \arctan\left(-\frac{b_n}{a_n}\right) \qquad \textbf{[A.2}^{\textbf{b}}\textbf{]}$$

In practice, this is the more common way to look at signals, $c_n$ is the amplitude spectrum of the signal and $\phi_n$ represents the phase. We will now use eq. [A.2] in the following procedure to obtain two signals which have the same amplitude spectrum, but are different in time domain:

- Select a signal (e.g. the signal from fig. 4)
- Determine its $a_n$ and $b_n$ coefficients
- Convert this to eq. [A.2]
- Add a phase term to $\phi_n$ (in this case, the phase of an 8th order Butterworth low-pass filter has been used.
- (Optional:) Select the frequency with the highest value of $c_n$ and add $\pi$ radians to the phase at that particular frequency.
- Synthesize the new signal using the modified set of $\phi_n$ and the same set of $c_n$, using eq. [A.2]. This results in the signal of fig. 5.

Because the values of $c_n$ have not been touched, the amplitude spectrum of the signal from fig. 5 is identical to that of fig. 4. Using this procedure an infinite number of signals with identical amplitude spectra, which differ in time domain, can be created.