Audioqualität: transparent channel

- analog (LP) besser als digital (CD,DAT) ?!
- Übertragungsqualität analysieren

=> Anforderungen an einen "transparenten Kanal" ?

J. R. Stuart: "Coding high quality audio"
www.meridian.co.uk/ara/index.html

Thesen:

- CD und DAT knapp unzureichend
- vor allem höhere Abtastrate nötig
- DVD-Audio / SACD ist Verschwendung

bits	f / KHZ
16	44.1 / 48
16	66
24	96

- bei Musikproduktion auf Quantisierungsfehler vermeiden
- Dithering/Noiseshaping unbedingt notwendig

"coding high quality digital audio"

ACKNOWLEDGEMENTS

This article is based on papers the author has presented to the Audio Engineering Society. It inevitably, draws on the work of others. Many of its insights have resulted from discussions between members of the Technical Subcommittee of the Acoustic Renaissance for Audio (Tony Griffiths, Professor Malcolm Hawksford, David Meares and Bob Stuart) and their advisors (Peter Craven, Michael Gerzon, Hiro Negishi, Francis Rumsey and Chris Travis). The assistance of Peter Craven, Takeo Yamomoto, Bike Suzuki, Malcolm Law and Adrian Farmer has also been particularly valuable.

CONTACT ADDRESSES

Acoustic Renaissance for Audio, c/o Meridian Audio, Stonehill, Stukeley Meadows, Huntingdon PE18 6ED, England. **Phone**: +44 1480 52144 **Fax**: +44 1480 451587

Email: ara@meridian-audio.com **Web:** http://www.meridian-audio.com/ara

Motivation: "CD-Qualität"

As digital audio has progressed, we have also evolved the capability to record and play back with resolution that exceeds that of Red Book CD[1] and current studio practise recognises this Red Book channel as a 'bottleneck'. High-quality recordings are routinely made and edited using equipment whose performance potential is considerably higher than CD. Figure 1 illustrates this conceptually, while figure 2 illustrates how resolution (in this case indicated by word-size) typically varies through a quality audio chain.

Along the way, some interesting ideas have been proposed to try to maximise the human-auditory potential of CD. One idea is noise-shaping. Noise-shaping was first proposed by Michael Gerzon and Peter Craven in 1989 [11] and successfully embodied in Meridian's 618, 518 [12, 22] and also in Sony's Super Bit Mapping [3]. This technique has been used on maybe a few thousand titles – but these include some of the very finest sounding CDs available today. Other proposals were interesting, but didn't get off the ground – like subtractive dither and schemes to add bandwidth or channels to CDs [4, 14, 19,].

The author has felt strongly for some time, that we are on the threshold of the most fantastic opportunity in audio. It comes from two directions. First, psychoacoustic theory and audio engineering may have progressed to the point where we know how to define a recording system that can be truly transparent as far as the human listener is concerned. Second, we will soon see the evolution of a high-density audio format, related to DVD, that has, if it is used wisely, the data capacity to achieve this goal.

Hörbereich: höchste Frequenzen?

IN-BAND NOISE SHAPING AND PRE-EMPHASIS

It is possible to exploit the frequency-dependent human hearing threshold by shaping the quantisation and dither so that the resulting noise floor is less audible.

Figure 19 shows how the Meridian 518 (an in-band shaper) can allow a 16-bit transmission channel to have a *subjective* noise floor more equivalent to a 20-bit 'simple' channel. If such a channel is to be useful, the resolution of the links in the chain before and after the noise-shaped channel must be adequate. In simple terms, this means mastering and playing back using well-designed converters offering at least 20-bit resolution.

It was the view of the ARA committee that noise shaping can be a linear process, and that it deserves serious consideration when distribution channels are to be matched to data-rate limitations.

FREQUENCY RANGE

The graphs to date have used the standard hearing threshold described in [20]. However, individuals can exhibit somewhat different thresholds [21 and 8]. The minimum audible field has a standard deviation of approximately 10dB.

Individuals are to be found whose thresholds are as low as –20dBspl at 4kHz. Similarly, although the high-frequency response cut-off rate is always rapid, certain people can detect 24kHz.

Abtastrate?

DO WE NEED MORE THAN 44.1KHZ?

The high-frequency region of figure 21 is shown in detail in figure 22. It can be seen that an average listener will find little to criticise in the in-band amplitude response of the DAC. To acute listeners, a 44.1kHz sample rate (even with the extremely narrow transition band shown) means a potential loss of extreme HF (between 20kHz and 22kHz). Raising the sampling rate to 48kHz does a lot to remedy this.

However, the significance of this has to be questioned. Although there is an area of intersection between the channel frequency responses and the hearing thresholds, this region is all above 100dBspl. The author knows of no program material that has any significant content above 20kHz *and* 100dBspl!

Numerous anecdotes suggest that a wider-frequency response 'sounds better'. It has often been suggested that a lower cut-off rate would give a more appropriate phase response, and that the in-band response ripple produced by the kind of linear-phase high-cut-off-rate filter illustrated in figure 22 (DAC) and figure 23 (ADC) can prove unexpectedly easy to detect. It is also frequently asserted that the slower rate of fall-off in HF response found in an analogue tape recorder accounts for a preferred sound quality.

It has also been suggested that the pre-ringing produced by the very steep linear-phase filters used so far for digital audio, can smear arrival-time detection and impact stereo imaging. This pre-ringing shows up in nearly all reviews of CD players. It can be significantly reduced by making the filter less steep (which we could do by raising the sample rate) or by not using a linear-phase characteristic.

The literature can contribute very little to this discussion. One well-performed set of experiments by Ohashi has, however, strongly indicated that certain program material may benefit from a system frequency response extending beyond 50kHz [17, 18].

Quantisierung?

24-BIT PCM CHANNELS

There is no convincing argument for using 24-bit data in a distribution format. Figure 7 clearly implies that the noise floor and resolution limit of a 24-bit channel will be 24dB greater than is necessary.

Why do it, then? One reason would be in order to convey more data for the subsequent DSP processes to work with. This reasoning is superficially correct. However, the author believes it to be unlikely that A/D converters that deliver 133dB analogue SNR will ever be made, and therefore a 24-bit channel would be kept busy conveying its own input noise! Furthermore, the majority of DSP systems and interfaces use a 24-bit word size. It is very, very difficult at present to guarantee transparency when performing non-trivial DSP operations on 24-bit data in a 24-bit processing environment. Obviously we could develop DSP processors capable of handling larger words, but why should we? Not only is the combination of well-handled, carefully delivered 20-bit data and a 24-bit processing environment good enough, but to deliver anything *more* is virtually to *guarantee* a higher risk of inadvertent truncation in the average replay chain.

A more pragmatic reason not to distribute 24-bit data is that it is virtually certain that the overwhelming majority of DVD players will not pass 24-bit data correctly. Even if they were to use 24-bit conversion, truncation is virtually guaranteed, whereas 20-bit data in the same pathway will pass virtually unscathed.

Abtastrate?

PSYCHOACOUSTIC DATA ON HIGH-FREQUENCY HEARING

There is very little hard evidence to suggest that it is important to reproduce sounds above 25kHz. Instead there tends to be a general impression that a wider bandwidth can give rise to fewer in-band problems. However, there are a few points to raise before dismissing audible content above 20kHz as unimportant.

The frequency response of the outer and middle ear has a fast cut-off rate due to combined roll-off in the acoustics of the meata and in mechanical transmission. There also appears to be an auditory filter cut-off in the cochlea itself.

The cochlea operates 'top-down', so the first auditory filter is the highest in frequency. This filter centres on approximately 15kHz, and extrapolation from known data suggests that it should have a noise bandwidth of approximately 3kHz. Middle-ear transmission loss seems to prevent the cochlea from being excited efficiently above 20kHz.

Bone-conduction tests using ultrasonics have shown that supersonic excitation ends up in this first 'bin'. Any supersonic information arriving at above 15kHz therefore ends up here, and its energy will accumulate towards detection. It is possible that in some ears a stimulus of moderate intensity but of wide bandwidth may modify perception or detection in this band, so that the effective noise bandwidth could be wider than 3kHz.

The late Michael Gerzon surmised that any in-air content above 20–25kHz derived its significance from non-linearity in the hearing transmission, and that combinations of otherwise inaudible components could be detected through any resulting in-band intermodulation products.

There is a powerful caution against this. As far as the author knows, music spectra that have measured content above 20kHz always exhibit that content at such a low spl that it is unlikely that the (presumed) lower spl difference distortion products would be detectable and not masked by the main content.

Quantisierungfehler vs. Dithering

Single undithered truncations at the 16-bit level are regrettably all too common in practice. Not only do inadvertent truncations arise in the hardware filters of very many converters, but the editing and mastering processes often include level shifts, mixing events or DC filtering processes that have not been dithered correctly. There have therefore been reasonable grounds to criticise the sound of some digital recordings – even though (as laboured earlier) this particular defect can be avoided by combining good engineering with good practice.

Figure 14 represents the audible significance of a channel in which a correctly dithered quantisation (perhaps in a word-length reduction from 20 to 16 bits) is followed by a minor undithered process, in this case a 0.5dB attenuation. This figure shows how just one undithered process can degrade a correctly converted signal. Once again it is predicted that detection of a raised and granular noise floor is highly probable.

Figure 15 shows how this effect could operate in practice. The upper curve represents the audible significance of the same –90dBFS tone with all the errors introduced by an original 'correct' 16-bit quantisation followed by four undithered signal-processing operations. Four operations may seem like a lot, but this figure actually illustrates a common case in which everyday analogue-to-digital and digital-to-analogue converters are used. (As has already been mentioned, the decimation/oversampling filters in hardware converters are rarely dithered).

This curve may be taken as a baseline of current bad practice in CD recording/replay. It is put in historical context in figure 16, which includes the audible significance of the playback noise in a silent LP groove.

Audioqualität: Fazit

CONCLUSIONS

This article has reviewed the issues surrounding the transmission of high-resolution digital audio. It is suggested that a channel that attains audible transparency will be equivalent to a PCM channel that uses:

- 58kHz sampling rate, *and*
- 14-bit representation with appropriate noise shaping, *or*
- 20-bit representation in a flat noise floor, i.e. a 'rectangular' channel

This conclusion has the following obvious implications:

- The CD channel with 44.1kHz 16-bit coding (even with noise shaping to extend the resolution) is inadequate
- Even 48kHz sampling is not quite high enough
- Sampling at 88.2kHz or 96kHz is too high, and therefore wasteful of data
- The use of sampling rates above 96kHz to convey a wider audio bandwidth cannot currently be justified

On the assumption that the industry will chose sample-rates based on 44.1kHz or 48kHz (i.e. 88.2kHz and 96kHz), we have looked at options for improving coding efficiency at these rates.

Noise shaping combined with a new pre-/de-emphasis characteristic for 96kHz (88.2kHz) applications can result in an effective addition of between 2 and 7 bits to the channel. In other words, at these sampling rates a 16-bit channel should be sufficient[10].

Audioqualität: CD

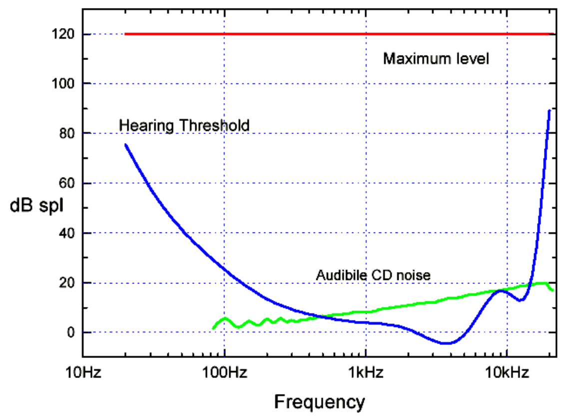

Figure 10. Dynamic range of CD.

Audioqualität: FM (UKW-Radio)

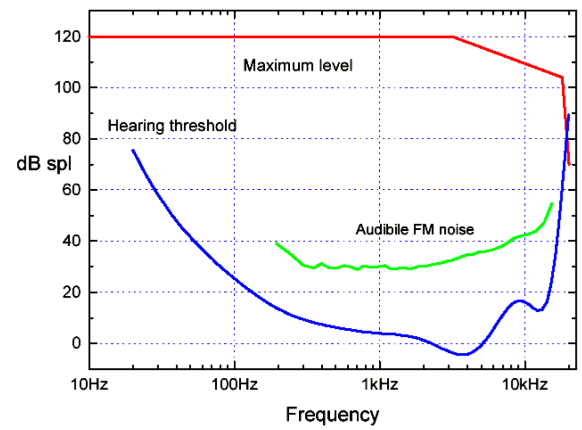

Figure 11. Dynamic range of FM.

Audioqualität: LP

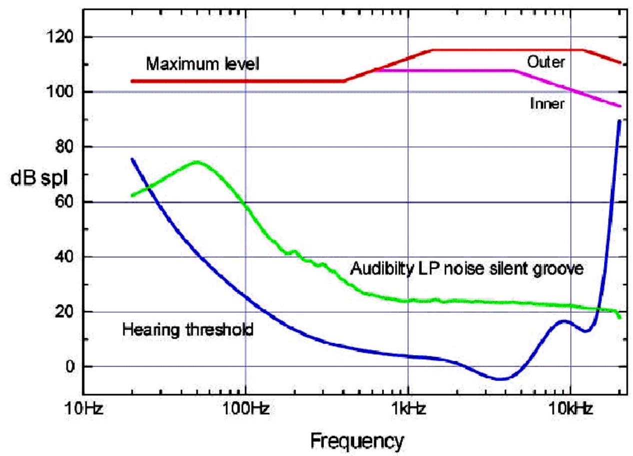

Figure 12. Dynamic range of LP.

Audioqualität: 16/18/20 bit

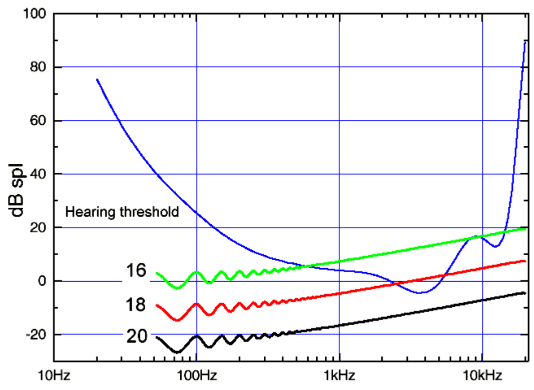

Figure 7. Audible significance of the noise created by a single white-spectrum TPDF-dithered quantisation in channels using 16, 18 and 20 bits. Audibility has been plotted against the average human hearing threshold assuming that a full-scale signal can attain 120dB spl at the listening position.

Audioqualität: 16/18/20 bit

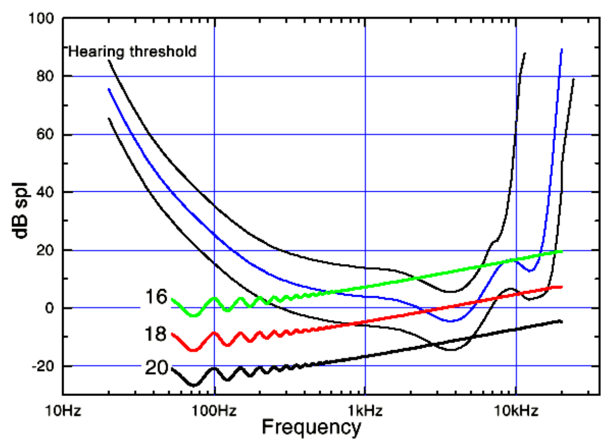

Figure 20. As figure 7, but showing how individual hearing thresholds can vary.

Audioqualität: Verzerrungen

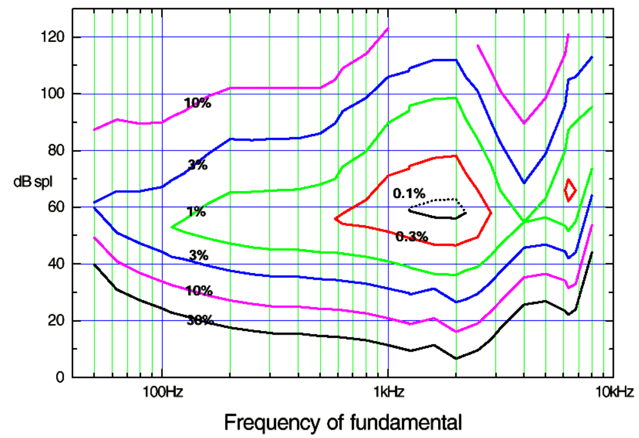

Figure 3. A contour map showing the existence regions for detecting the presence of an added second-harmonic tone. The spl is of the fundamental frequency. Inside a contour, 2nd-harmonic distortion of the marked percentage should be audible.

Audioqualität: Differenztöne

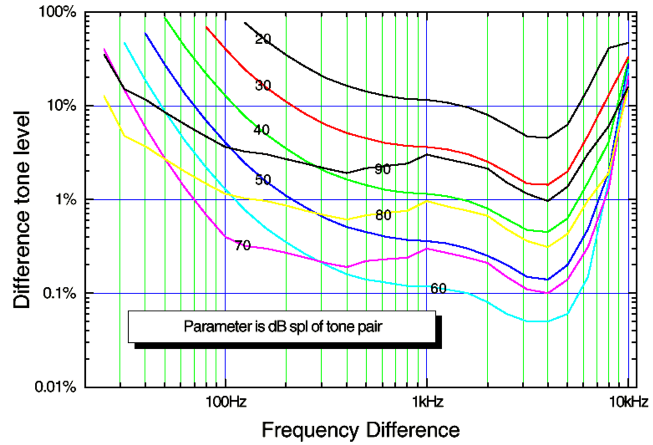

Figure 4. Illustrating predicted detectability of a difference-tone resulting from non-linear processing. See text. The parameter is spl of the combination.

Dither: Sinus ohne Dithering

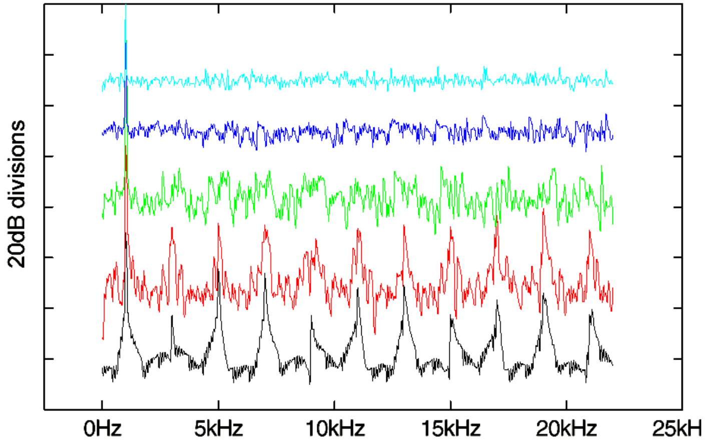

Figure 5. FFT analyses of an undithered 16-bit quantisation of a 1kHz tone at –20, –40, –60, –80 and at –90dBFS (top to bottom). Curves offset by 25dB.

Dither: Sinus mit Dithering

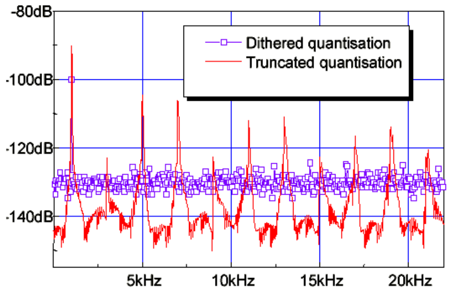

Figure 6. FFT measurements of the spectrum that results when a –90dBFS 1kHz tone is quantised to a 16-bit format, with and without correct (triangular probability distribution) dither.

Dither: ein Quantisierungsschritt

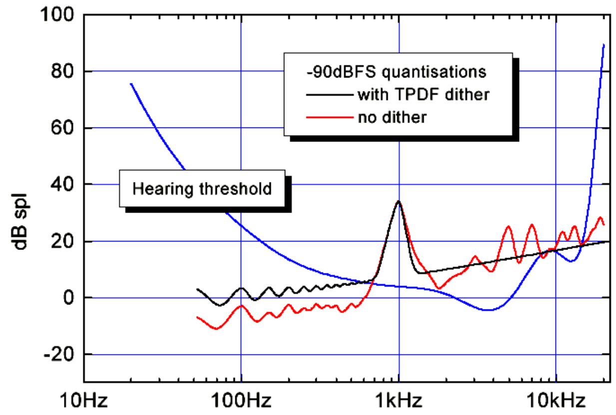

Figure 13. Audible significance of dithered and undithered 16-bit 44.1kHz sampling of a 1kHz –90dBFS (i.e. 30dBspl) tone. (0dBFS ≡ 120dBspl.)

Dither: 1+4 Quantisierungsschritte

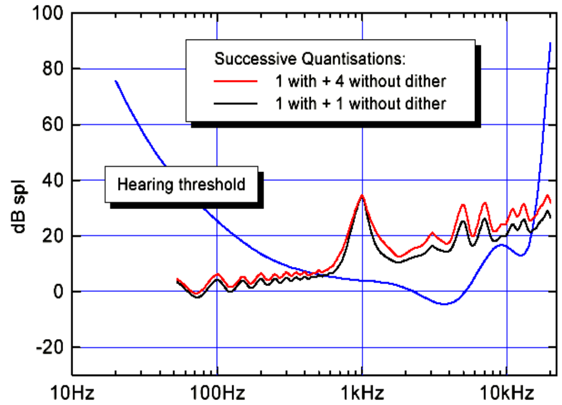

Figure 15. Audible significance of one (lower) and four (upper) successive undithered 16-bit 44.1kHz resamplings of a 1kHz –90dBFS (i.e. 30dBspl) tone on a signal already correctly quantised to 16 bits.

Dither: CD mit Quantisierung, vs. LP

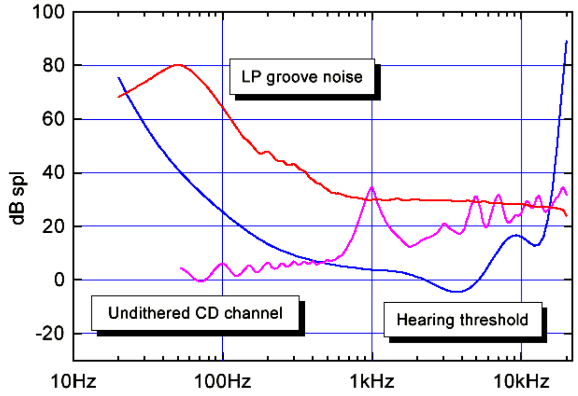

Figure 16. Audible significance of four (lower) successive undithered 16-bit 44.1kHz resamplings of a 1kHz –90dBFS (i.e. 30dBspl) tone on a signal already correctly quantised to 16 bits, contrasted with the audible significance of noise floor measured on a silent LP groove.

Dither: Quantisierung, 20 bit

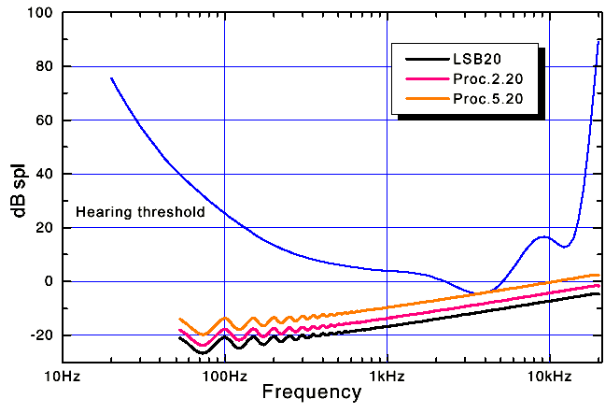

Figure 17. Audible significance of the noise created by 1, 2 and 5 successive TPDF-dithered quantisations in a 20-bit channel.

Noise-Shaping, 16 bit

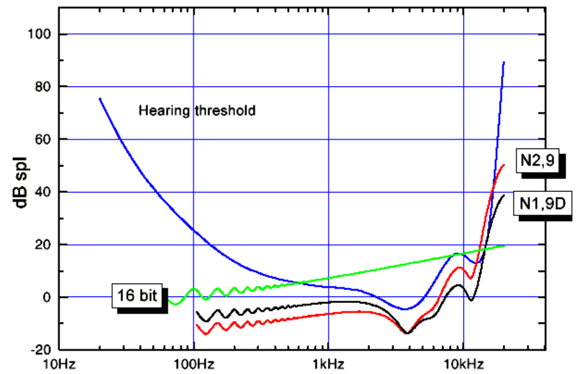

Figure 19. Audible significance of a simple 16-bit channel, with two examples from [24] of the audible significance of noise shaping in a 16-bit channel.

Audioqualität: CD/DAT Grenzfrequenzen

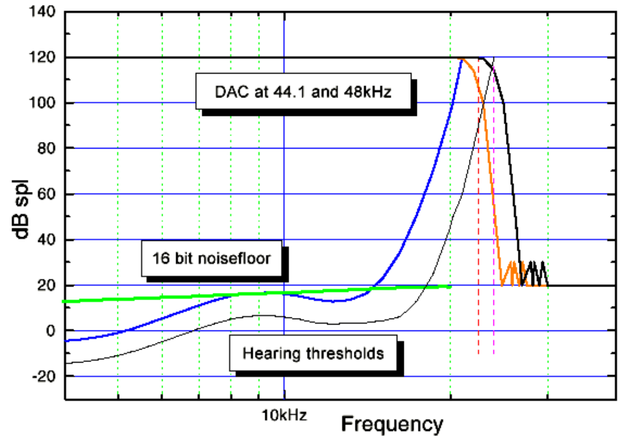

Figure 22. The high-frequency range of figure 21.

Audioqualität: 20 bit

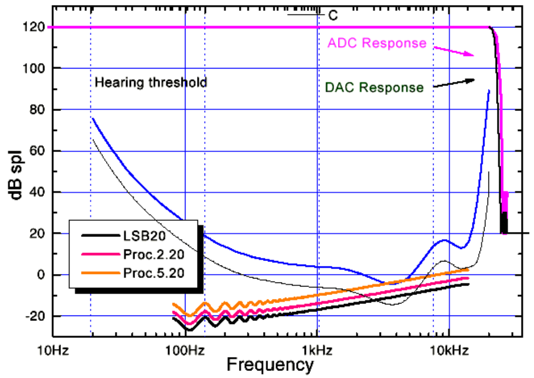

Figure 23. Useful operating region of a well-engineered 20-bit channel. The audible significance of noise created by 1, 2 and 5 successive quantisations is shown.

Audioqualität: "Shannon-Space"

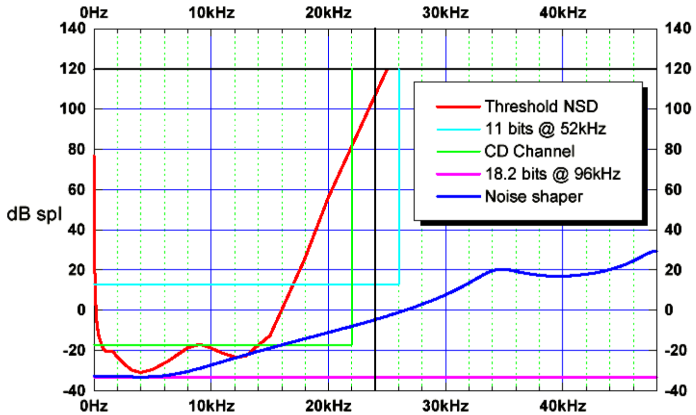

Figure 25. The 'Shannon space' for human hearing.

Audioqualität: 96 KHz, 66 KHz

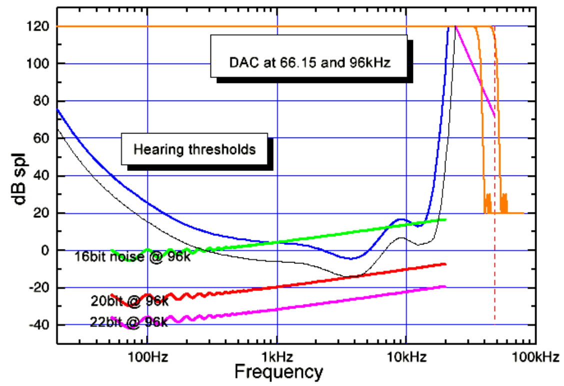

Figure 26. Useful operating regions of channels using 96kHz and 66.15kHz sampling. The figure shows that both rates allow for a near-audible HF region in which more gentle filtering could be used. The audible-significance channel-noise curves are given for 96kHz and for 16, 20 and 22-bit word lengths.

Audioqualität: 66 KHz, noise-shaping

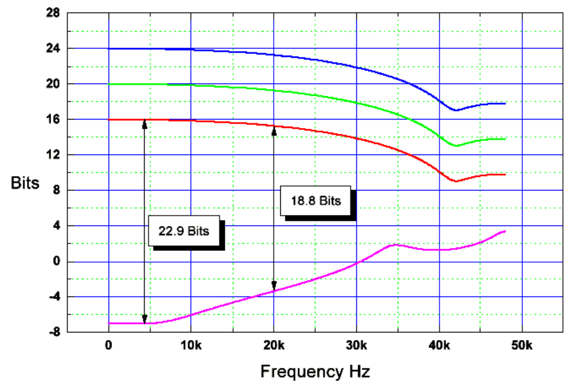

Figure 33. Output noise spectrum and headroom for a channel after the example 6th-order noise shaper has been combined with the proposed pre- and de-emphasis. The graph expresses the dynamic range in bits. The example illustrates a capacity of almost 23 bits at 4kHz for a 16-bit channel, i.e. a perceptual gain of 7 bits.