Improved Psychoacoustic Noise Shaping
for Requantization of High-Resolution Digital Audio

CHRISTIAN R. HELMRICH [1], **MARTIN HOLTERS** [2], AND **UDO ZÖLZER** [3]

[1] *Technical University of Hamburg-Harburg, Germany*
christian.helmrich@tuhh.de
[2] *Helmut Schmidt University, Hamburg, Germany*
martin.holters@hsu-hh.de
[3] *Helmut Schmidt University, Hamburg, Germany*
udo.zoelzer@hsu-hh.de

ABSTRACT

Considering high-resolution, multi-channel audio, it is worth re-examining the concept of psychoacoustically optimized noise shaping to ensure that signal quality is preserved when word-lengths are reduced. In this paper, approaches in static (time-invariant) and signal-adaptive (time-variant) noise shaping are discussed. We identify problems occurring when equal-loudness level contours such as the inverse F-weighting curve are used as a model for noise shaping. As a remedy, we present two alternative time-invariant filter designs. Regarding adaptive requantization, we introduce a noise shaper based on work by Verhelst and De Koning with an improved design of the time-variant filter. The paper concludes with a comparative evaluation based on listening tests with different transducers.

0 INTRODUCTION

The popularity of high-resolution digital audio systems has renewed the interest in signal requantization issues. Today, many available recording and playback systems provide a level of performance that clearly exceeds the capabilities of the traditional Compact Disc Digital Audio format, which offers an amplitude resolution of 16 bits and a frequency range of 22 kHz for two independent channels. In particular, the entire signal chain, from recording over storage to reproduction, can now be realized using 24 bits per sample and channel. This allows for a dynamic range distinctly beyond that achievable in a 16-bit format.

However, there are numerous situations in which audio data of lower resolution is still preferred or required. In the consumer market, the 16-bit pulse code modulation (PCM) format remains the most widely used choice for audio storage and reproduction. Moreover, it should be noted that transform-based, compressed audio streams such as Ogg Vorbis or AC-3 (5.1-channel Dolby Digital on DVD or HDTV), are usually decoded using floating-point or high-precision integer arithmetic. This necessitates subsequent requantization to a fixed-point format of lower word-length (typically 16 bits) before sending the audio data to the digital-to-analog converter.

Consequently, high-resolution digital audio often needs to be converted into data of lower resolution. The most common example is the requantization of 24- or 32-bit recordings into "CD-quality" 16-bit PCM audio. Naturally, the dynamic range and transparency of the original signal should be conserved as much as possible even in lower resolutions. This is where some issues arise.

0.1 Problems in Audio Requantization

Requantization reduces the word-length, that is, the size (in bits) of the samples in a digital signal. This process inevitably causes an error $e(n)$ which, upon reproduction of the output signal, can become noticeable. In the case of audio signals, the following types of artifacts may be audible during playback:

- **Modulated distortion**. This linear or nonlinear, spectral distortion occurs when $e(n)$ is correlated with the original signal. As it appears modulated by the signal, it changes over time.

- **Broadband noise**. This effect occurs when $e(n)$ is mostly uncorrelated with the signal. The noise is permanently audible in the background and is similar to the noise in analog tape recordings.

At low signal levels, correlated quantization errors are considered much more serious than uncorrelated ones as they may cause pernicious harmonic distortion of the input signal [1], [2]. It is therefore generally agreed upon that a minimization of the correlation between the signal and $e(n)$ is desirable to "linearize" requantization effects and thus preserve signal fidelity [1]–[5], [14].

Decorrelation of $e(n)$ can be realized by adding a dither signal to the input signal before the requantization step. The type of dither most commonly used is random noise with a peak-to-peak amplitude of 2 least significant bits (LSB), that is, a range from –1 to +1 LSB [1], and a triangular probability distribution function (TPDF) [1]–[5]. Dither signals with such properties can be implemented in software by adding to each audio channel the output of two independent pseudo-random number generators, each with rectangular (uniform) probability distribution function (RPDF) and 1 LSB amplitude range [6, p. 38].

To reduce the word-length of an audio signal, numerous methods of varied computational complexity exist. As already mentioned, the error caused by requantization is inevitable, so each method generates some kind of error signal. The nature of these error signals, however, varies between the different processes. Two well-known, computationally simple word-length reduction methods are illustrated in **Figure 1**: requantization by pure rounding and requantization with TPDF dither added prior to the rounding process. The input is a 1-kHz sinusoidal waveform with a peak-to-peak amplitude of 2 LSB [1]. As discussed above, the dither signal also ranges over 2 LSB. The following phenomena can be observed:

- **Requantization without dither**. Simple truncation of the samples in the 1-kHz signal results in harmonic components at integer multiples of the base frequency. The noise level is still very low.

- **Requantization with TPDF dither**. With TPDF dither added, the harmonic distortion disappears at the expense of a noticeably higher noise floor. The noise is white (flat) and constant in level.

0.2 Requantization with Noise Shaping

The sensitivity of human hearing varies with frequency. This psychoacoustic principle can be exploited to minimize the audibility of $e(n)$. Regardless of whether or not dither is used during requantization, minimal audibility can be accomplished by noise shaping, a process which changes the shape of the error spectrum. The use of a properly chosen noise shaping function reduces the perceptual loudness of $e(n)$. Accordingly, it should be noted that inappropriate noise shaping functions will increase the loudness of $e(n)$.

Noise shaping can be applied in two fundamental ways:

- **Time-invariant noise shaping** tries to minimize the absolute audibility of the requantization error, that is, the loudness of $e(n)$ in the absence of an input signal. The significance of this approach is supported by the observation that errors are likely to be most audible and disturbing at low signal levels. Hence, a static noise shaping filter can be applied to shape the frequency spectrum of $e(n)$. Such filters have been proposed in [4], [5], [7].

- **Time-variant noise shaping** considers the relative audibility of $e(n)$, that is, the error loudness in the presence of an audio signal. The approach here is to minimize audibility of $e(n)$ by applying an adaptive noise shaping filter. As the psychoacoustic properties of the signal (in this case, the spectral masking characteristics) vary with time, the filter coefficients need to be updated on-line and at frequent intervals. Adaptive noise shaping filters have been proposed in [7], [8].

Noise-shaping feedback around a requantizer in word-length reduction applications has been covered in detail in the scientific literature of the last two decades (one of the first publications on the subject is the paper by Gerzon and Craven [9] from 1989). Even outside the scientific press, noteworthy proposals and evaluations about psychoacoustically optimized requantization have been presented [10], [11]. The properties and advantages of noise shaping, especially in the time-invariant form, are therefore well understood. Nevertheless, we believe that in the face of high-resolution, multi-channel audio applications, there is still room for further optimization.

In this paper, we examine recent approaches in static and adaptive psychoacoustic noise shaping. Concerning minimally audible designs, we identify problems occurring when equal-loudness contours such as the threshold of hearing defined in ISO standard 226 [12] are used as a model for noise shaping. As a remedy, we propose two alternative filter designs. One is a modified approxima-

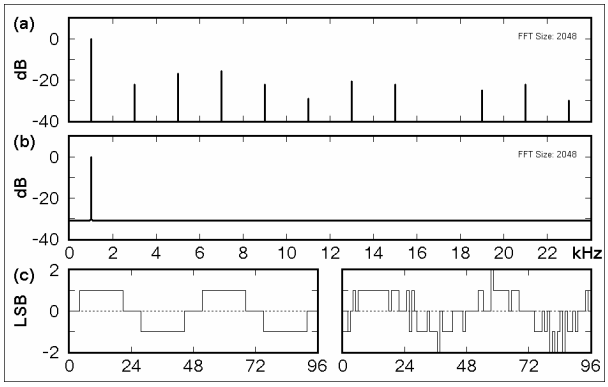

Figure 1. Frequency and amplitude effects caused by word-length reduction of a 1-kHz sine wave sampled at 48 kHz. (a) without dither, (b) with TPDF dither, (c) resulting waveforms without (left) and with dither (right).

[1] measured in least significant bits of the requantized output

tion of the noise weighting curve specified in ITU-R recommendation BS.468-4 [13], the other a diffused-field corrected variant of the uniformly-exciting noise (UEN) curve introduced in [14]. With regard to signal-adaptive noise shaping, we introduce a system based on work by Verhelst and De Koning [8] with improvements in the design of the time-variant noise shaping filter.

The remainder of this paper is organized as follows. In **section 1**, we review the fundamentals of noise shaping and relevant psychoacoustic principles, present our proposed time-invariant noise shaping filters and clarify the motivation behind their development. In **section 2**, we explain how signal-adaptive noise shaping can be achieved and describe our proposed modification of the noise shaping system presented in [8]. **Section 3** provides the essential details of the two listening experiments carried out to evaluate the performance of our filter designs in comparison to established solutions. **Section 4** follows with a discussion of the results gathered from the listening tests. Lastly, in **section 5**, we conclude the paper and present our own opinion about the use of adaptive noise shaping techniques.

1 TIME-INVARIANT NOISE SHAPING

As stated earlier, the performance of a (re)quantizer can be improved by adding for example a dither signal with triangular probability distribution (2-RPDF) prior to the actual quantization process. This method improves the quality especially of low-level signals as it renders the first two moments of $e(n)$ independent of the input signal. As a result, the power spectrum of $e(n)$ is rendered equal to the power spectrum of the dither signal plus a white "quantization noise" component [3].

Even though requantization using TPDF dither succeeds in making $e(n)$ virtually constant with respect to the system input, noise shaping can additionally be applied to minimize the total audibility of $e(n)$. The technique of noise shaping utilizes error feedback to spectrally shape $e(n)$, including the white noise component arising from the dither signal [3]. The general scheme for signal requantization with dither and noise shaping is illustrated in **Figure 2**. In this scheme, D denotes the dither signal generator, Q represents the (re)quantizer, and $H(z)$ is the error feedback filter. Due to the quantization error $e(n)$, the output (the requantized signal) $y(n)$ differs from the input (the initial signal) $x(n)$ and from $x(n) + e'(n)$. The parameters of $H(z)$ can be specified such that the difference between $y(n)$ and $x(n)$ becomes minimally audible.

1.1 Design of Optimal Noise Shaping Filters

As shown in [8], the requantization error $e'(n)$ has power spectrum

$$P_{E'}(e^{j\omega}) = \| 1 - H(e^{j\omega}) \|^2 \, P_E(e^{j\omega}) \,, \qquad (1)$$

so the filter shaping the spectrum of $e'(n)$ depends on the error feedback filter $H(z)$. If we assume a certain desired shape $P_{des}(e^{j\omega})$ of the requantization error spectrum, the coefficients of $H(z)$ have to be determined such that

$$\| 1 - H(e^{j\omega}) \|^2 \, P_E(e^{j\omega}) = \alpha P_{des}(e^{j\omega}) \,, \qquad (2)$$

where α is to be minimized. It was proven by Gerzon and Craven in [9] that the noise shaping filter $1 - H(z)$ that satisfies equation (2) with the lowest possible error power is the one that leaves the information capacity of the channel at its maximum, that is, unaffected. Furthermore, it was shown that this property is achieved if (and only if) $1 - H(z)$ is minimum-phase.

Verhelst and De Koning [8] noted that in order to avoid delayless loops in Figure 2, it is required that $1 - H(z)$, when realized by a finite impulse response (FIR) filter, can be written as

$$1 - H(z) = \sum_{n=0}^{M} a(n) z^{-n}, \qquad \text{where } a(0) = 1 \,. \qquad (3)$$

As demonstrated in [7], a minimum-phase FIR filter that satisfies (3) can be determined by approximating the inverse of the desired noise shaping spectrum with an M^{th}-order LPC synthesis filter and inverting the result. The interested reader is referred to [7] for details.

A similar approach, based on a least-squares interpretation of the above problem, is presented in [8]. There, the proposed theory for the design of optimal noise shaping filters is centered around the autocorrelation formulation of the LPC analysis of $v(n) = \mathcal{F}^{-1}[\, (W(\omega) P_E(e^{j\omega}))^{0.5} \,]$ and is given as the linear matrix equation

$$\mathbf{Ra} = -\mathbf{r} \qquad (4)$$

with

$$\mathbf{R} = \begin{pmatrix} r(0) & r(1) & \cdots & r(M-1) \\ r(1) & r(0) & \cdots & r(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(M-1) & r(M-2) & \cdots & r(0) \end{pmatrix}$$

$$\mathbf{a} = \begin{pmatrix} a(1) \\ \vdots \\ a(M) \end{pmatrix} \qquad \mathbf{r} = \begin{pmatrix} r(1) \\ \vdots \\ r(M) \end{pmatrix} \qquad \text{and}$$

$$r(i) = \sum_{n=-\infty}^{+\infty} v(n) v(n-i) \,.$$

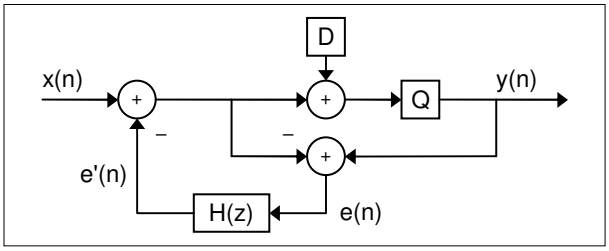

Figure 2. Requantization with dither signal generator D, quantizer Q, and time-invariant feedback filter $H(z)$.

$W(\omega)$ is also called the perceptual weighting function. It is stated that, in the absence of computational errors, the solution to (4) is guaranteed to be a minimum-phase filter. Furthermore, (4) can be solved easily and efficiently since **R** is a symmetric Toeplitz matrix. Lastly, when the requantization error spectrum is known to be white, as is the case in a properly dithered system, $r(i)$, $i = 0...M$ can be approximated by sampling the inverse $W(\omega)$ of the desired error spectrum and calculating the inverse FFT:

$$r(i) = \frac{1}{N}\sum_{k=0}^{N-1} W\left(\frac{2\pi}{N}k\right) e^{j\frac{2\pi}{N}ki}, \qquad i = 0 \ldots M, \qquad (5)$$

where N represents the number of frequency samples of $W(\omega)$ and M denotes the desired order of the FIR noise shaping filter.

Given M, N, and $W(\omega)$ such that $N >> M$, the approximation error can indeed be rendered sufficiently small. We therefore agree with the authors of [8] that solving equations (5) and (4) represents a computationally efficient method for designing optimal noise shaping filters. Subsequently, we shall apply this *Least Squares* method to compute time-invariant as well as time-variant noise shaping filters.

1.2 Psychoacoustic Considerations

As mentioned previously, the solution to (5) and (4) produces a noise shaping filter that is numerically optimal (minimum-phase). To also optimize the filter in terms of its psychoacoustic properties, $W(\omega)$ needs to be chosen such that the filtered spectrum is minimally audible. For now, we shall consider dithered requantization and time-invariant noise shaping, so $W(\omega)$ should represent the inverse of a noise spectrum which is minimally audible in silence. The critical psychoacoustic aspect is the ear's sensitivity to low-level broadband noise as a function of frequency, averaged over a large number of subjects.

Various perceptual weighting functions have been adopted to produce minimally audible requantization noise. Two prominent choices are examined below.

- The **F-weighting** curve represents a diffuse-field corrected variant of the 15-phon equal-loudness contour specified in ISO standard 226 [12]. This weighting curve for usage in the design of noise shaping filters was first introduced in [4], where it was referred to as "improved E-weighting". F-weighting was explicitly defined by Wannamaker in [5] and has been employed in the experimental requantizers in [8] and other implementations.

- The **inverted hearing threshold** represents the inverse of the 4-phon threshold-of-hearing curve (also known as minimum audible field, or MAF) defined in [12]. This weighting function has been applied in the Super Bit Mapping systems [7].

In addition, we recommend the review in [10] to readers interested in the noise shaping spectra of some commercial implementations.

Both weighting functions introduced above represent a so-called equal-loudness level contour. Adopting such a contour as $W(\omega)$ in a noise shaping system supposedly renders the requantization error perceptually uniform at loudness levels of 15 and 4 phon, respectively. It should be noted, however, that this does **not** guarantee minimal audibility for the case at hand. In fact, an investigation of the equal-loudness contours in [12] reveals two major issues which apparently have rarely been considered in connection with audio requantization in the past.

The first problem arising from the utilization of perceptual weighting functions based on equal-loudness curves from [12] can be identified by examining the scope of the standard. According to the abstract, standard 226, revision 2003 "*specifies combinations of sound pressure levels and frequencies of pure continuous tones which are perceived as equally loud by human listeners*" [12]. The noise spectrum produced by a dithered requantizer is dense and normally stretches over at least the audible range, hence we believe that it should not be regarded as a combination of pure tones. This view is supported by [14]–[16], where it is in fact suggested that the cochlea (inner ear) acts like a bank of narrow-band filters. The bandwidth of each filter (or "hair cell", to be precise) is proportional to its center frequency, so as the frequency increases, a wider portion (in absolute terms) of spectral energy is "collected" from a noise source. Accordingly, sensitivity to noise differs markedly from sensitivity to pure tones, especially at high frequencies.

It should be noted here that in analog audio equipment such as magnetic tape recorders, noise shaping has been performed since at least the early 1960s. In response to calls for a noise weighting curve providing satisfactory agreement with subjective evaluations, data from experiments performed by the BBC were incorporated into

CCIR recommendation 468-4, which is now maintained by the International Telecommunication Union, Radiocommunication Assembly (ITU-R) [13]. ITU-R 468-4 defines a weighting network whose magnitude response differs significantly from the F-weighting curve and the inverse threshold of hearing. Most notably, its spectrum, which is shown inverted in **Figure 3** (thin solid curve), peaks at 6.3 kHz, whereas the F-weighting and inverted MAF contours exhibit maximum gain around 3.5 kHz.

The second issue that should be considered when selecting a perceptual weighting function for noise shaping is the potential discrepancy between listening conditions. The equal-loudness contours specified in ISO standard 226 were measured in (and hence apply to) a free field [12], [17], that is, a non-reflective environment in which the sound source is directly in front of the listener. Free-field conditions can be attained either in open air or in anechoic chambers. Arguably, the most common transducer setups for "real-world" critical listening are

- diffuse-field equalized headphones and
- two or more loudspeakers distributed in a more or less reflective room.

In acoustic terms, both differ considerably from a free-field environment for which the contours in [12] apply.

In addition, we would like to emphasize that for multi-channel audio, the requantization error is typically not the same for all channels. Assuming non-identical input signals on each channel, the error signals introduced by word-length reduction will also be different. This is the case for both undithered and dithered requantization [1]. Hence, the requantization noise in multi-channel (particularly surround-sound) recordings exhibits, upon playback, a spatial character which resembles a diffuse field.

Indeed, free-to-diffuse-field correction has been applied to the 15-phon equal-loudness contour during the design of the F-weighting curve [5]. The effects of this equalization, however, are only moderate. The distinctive drop in the frequency region around 8 kHz remains. This pit is mirrored as a peak in the noise shaping spectrum and is believed to cause audible problems. In [10], perceptible noise at 8 kHz is reported for POW-r3, a commercial noise shaping algorithm developed by the POW-r (Psychoacoustically Optimized Word-length reduction) Consortium whose filter spectrum resembles the inverse of the F-weighting curve. In numerous preliminary experiments with noise shapers implementing the 9-coefficient FIR filters proposed in [4] and [5], we also independently observed distinct "hiss" in the same band.

Concluding this subsection, we would like to affirm our supposition that regarding noise-shaping requantization in multi-channel audio applications, neither the perception of noise versus pure tones, nor the characteristics of "real-world" diffuse-field listening conditions have been considered sufficiently in the literature. Hence, we shall proceed with the introduction of two alternatives to the F-weighting and inverse MAF curves which, to the best of our knowledge, have not been utilized in the design of noise-shaping requantizers to date.

1.3 Two Alternative Noise Shaper Designs

As stated previously, the response of the weighting network specified in [13] differs from the F-weighting and inverted MAF curves, especially in the frequency region around 8 kHz. There, the ITU-R network does not show the distinct drop in level which is prominent in the other two curves. We therefore devised a perceptual weighting function based on the ITU-R 468-4 curve, intended for use in noise-shaping requantization (and possibly other areas of application).

This curve, which we shall subsequently call "*HF (high-frequency) modified ITU 468 weighting*", is identical to the weighting curve in [13] for frequencies from 1 to 12 kHz. The response above 12 kHz, however, was customized and does not coincide with the ITU-R data. The motive behind this modification is the observation that the ear's sensitivity to high frequencies rapidly declines above approximately 12 kHz [4], [6, p. 46], [18, p. 18]. As this feature is not fully reflected in the original ITU-R curve, we decided to model a steeper progression for high frequencies by applying cubic spline extrapolation.

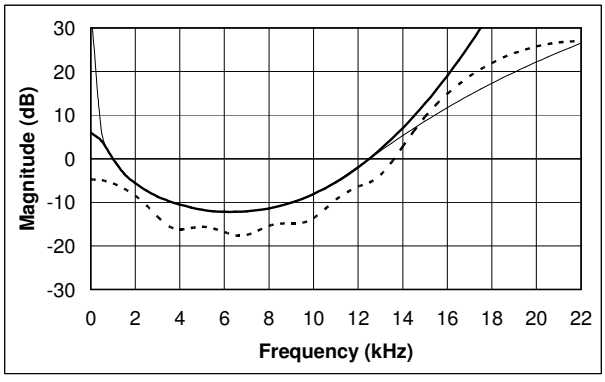

Figure 3. Derivation of *HF modified ITU-R 468* noise shaping filter. (—) inverse of original ITU-R curve [13], (**—**) proposed modification, (**- - -**) approximation by 8th-order FIR filter (see Table 1).

[1] As this might not seem intuitive to some readers, it should be added that dithered word-length reduction is commonly performed on a per-channel basis, that is, with independent dither signals for each channel in a multi-channel signal. In fact, this procedure is applied in all commercial implementations we know of and is recommended because it ensures that stereophonic separation will not be compromised [2].

The sound pressure levels from [13] for the defined frequencies from 31.5 to 12500 Hz (inclusively) were used to spline-interpolate equidistant values for multiples of 22.05/22 kHz. The resulting levels for 13.03 kHz up to the chosen Nyquist frequency of $f_S/2$ = 22.05 kHz thus represent a spline-extrapolation of the lower-frequency data which, when plotted, falls off more quickly than the actual response defined for this frequency range in [13]. In addition, we somewhat arbitrarily specified an SPL of –6 dB at 0 Hz. This correction intends to compensate for the dramatic roll-off at very low frequencies, a property which is unrealizable in noise shaping filters [4], [5].

Figure 3 shows the inverse of the resultant *HF modified ITU 468 weighting* curve (thick solid line) along with its approximation by an 8th-order FIR noise shaping filter (dotted line). Note that the high-frequency response of this filter was delimited to a maximum level of 27 dB in order to produce an unweighted noise power identical to that of the 9-coefficient F-weighting FIR filter proposed in [5]. The 8 coefficients of our filter were obtained via the *Least Squares* approach [8] discussed in section 1.1 and are listed in **Table 1** for convenience.

In section 1.2, we noted that the 4-phon equal-loudness contour defined in [12] standardizes the human hearing threshold (MAF) exclusively for pure continuous tones as a function of tone frequency. An equivalent curve for noise (instead of tones) was derived by Stuart [14]. This curve, which is referred to as "*uniformly-exciting noise (UEN) at threshold*", defines the minimum audible SPL of critical-band-wide noise as a function of the noise's center frequency. Its spectrum is illustrated in **Figure 4** (thin solid line). We believe that *UEN at threshold* poses an excellent reference for noise-shaping requantization because at levels just below threshold, it supposedly represents the most intense in-band noise signal which we cannot hear [14]. Hence, we developed a corresponding noise shaping curve based on the information in [15].

The sound pressure levels defined for 4.2 phon (the minimum audible field) in [12] for all 29 frequencies up to 12.5 kHz were taken as the basis function. Correction to noise spectral density (NSD) was then accomplished by subtracting from the MAF curve the values defined for 0 dB SPL in [15, Table 1]. This "tone-to-noise" correction yields an equal-loudness curve which represents the power spectrum of *UEN at threshold* in a free field. To account for the more typical diffuse-field listening conditions, as discussed in the former subsection, additional free-to-diffuse-field equalization [12] was applied. As a final step, the curve was shifted upwards by 19.3 dB to achieve unity gain at 1 kHz and ease comparison to the MAF curve and our *HF modified ITU 468 weighting*.

From the resulting data, equidistant values were approximated analogously to the previous procedure by cubic splines. In particular, the response above 12.03 kHz was determined by direct extrapolation of the low-frequency data. As the *UEN at threshold* curve also exhibits an unfavorably steep run below 1 kHz, we defined a value of 6 dB SPL at direct current to facilitate filter design.

The resulting curve, which we shall refer to as "*diffuse-field (DF) corrected UEN at threshold*", is shown as the thick solid line in Figure 4. Note that to use this curve as a weighting function, it needs to be inverted. Following up on the foregoing derivation, we approximated the *DF corrected UEN at threshold* with an 8-coefficient filter (dotted line in Fig. 4), which is also included in Table 1. The high-frequency response of this filter was also limited to 27 dB to yield the same unweighted noise power as the F-weighting and *HF modified ITU 468* filters.

As a final note for this section, we must emphasize that both filters specified in Table 1 assume a sampling rate of f_S = 44.1 kHz. For higher values, the proposed weighting curves need to be extended and resampled and the filters recalculated, presumably with a higher order (for example, to f_S = 96 kHz with number of weighting curve samples N = 96 and filter order M = 18).

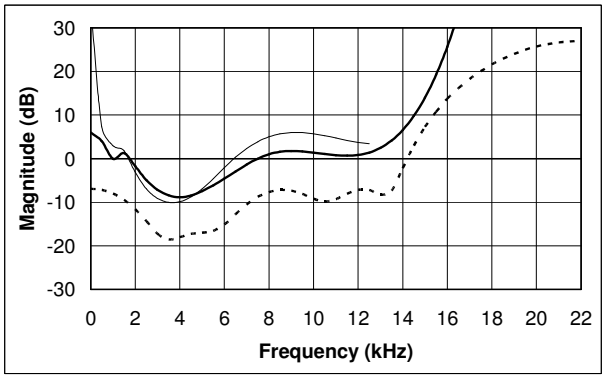

Figure 4. Derivation of *DF corrected UEN at threshold* noise shaping filter. (—) original free-field threshold for UEN [14], shifted for display, (**—**) proposed modification, (**- - -**) approximation by 8th-order FIR filter.

Coefficient	F-Weighting FIR [5]	HF Modified ITU-R 468	DF Corrected UEN at Thr.
a_1	2.412	2.312	2.259
a_2	–3.370	–3.839	–3.514
a_3	3.937	4.456	4.222
a_4	–4.174	–4.317	–4.308
a_5	3.353	3.242	3.391
a_6	–2.205	–2.040	–2.239
a_7	1.281	0.8933	1.095
a_8	–0.569	–0.2863	–0.3580
a_9	0.0847		

Table 1. Coefficients for discussed filter designs $H(z)$.

2 TIME-VARIANT NOISE SHAPING

Besides the frequency-dependent loudness of noise (see section 1.2), it can prove advantageous to consider two other psychoacoustic effects in the design of noise-shaping requantizers: simultaneous and temporal masking. The former occurs when a signal becomes inaudible in the presence of a masker, that is, another signal which is higher in level than the first. Temporal masking of a signal, on the other hand, takes place before (pre-masking) and after (post-masking) the presentation of the masker and is only of relatively short duration [18, p. 72 f.].

2.1 More Psychoacoustic Considerations

If the input to a noise-shaping requantizer is seen as the masker signal, simultaneous masking can intuitively be exploited as a means to minimize the relative audibility of the requantization noise. The effect of simultaneous masking (and to some extent, of temporal masking) depends greatly on the spectral content of both masker and masked signal [18, p. 56 ff.]. This implies that the noise shaping spectrum should resemble the input spectrum in order to achieve minimal audibility. The frequency content of typical signals such as music or speech, however, varies over time, so the noise shaping spectrum must be updated on a regular basis.

Figure 5 illustrates the general scheme for time-variant psychoacoustically optimized word-length reduction [8]. Contrary to time-invariant noise shaping (Figure 2), the error feedback filter $H(z)$ here is adaptive and designed to approximate the instantaneous masking threshold of the signal $x(n)$. For this purpose, (re)computation of the coefficients of $H(z)$ is controlled by a frequently updated psychoacoustic model.

A time-variant noise shaper based on the *Least Squares* approach is presented in [8]. There, the following simplified psychoacoustic model is applied to obtain $W(\omega)$:

$$W(\omega_k = \frac{2\pi k}{N}) = \frac{1}{\beta\, P_{XX}(\omega_k) + (1-\beta) P_{TQ}(\omega_k)}, \quad (6)$$

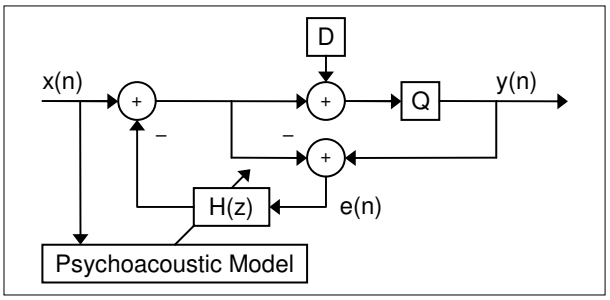

Figure 5. Time-variant requantization. Adaptive feedback filter $H(z)$ is controlled by a psychoacoustic model.

where $P_{XX}(\omega_k)$ represents the power spectrum of a 512-point Hanning-windowed segment of input signal $x(n)$. $P_{TQ}(\omega_k)$, the threshold in quiet, denotes the spectrum of the 9th-order filter approximating the inverse F-weighting curve (see Table 1). The scaling factor β depends on the window size (N = 512 is used) and the word-length (in bits) of input $x(n)$. $W(\omega_k)$ is updated every 256 input samples. Likewise, the noise shaper coefficients are recalculated every 256 samples by solving (4) with a 512-point inverse FFT of $W(\omega_k)$ as $r(i)$.

While this adaptive psychoacoustic approach was found to weaken requantization noise during signal portions of moderate and high level [8], we observed that a significant amount of noise shaping potential is left unused. In relatively loud segments, the unweighted requantization noise power tends to decline considerably — the louder and more spectrally complex a signal part is, the lower the requantization noise power tends to be for this part. The reason for this phenomenon is the strong adaptation of the noise shaping filter to high-level signals. In typical signals such as music or speech, most of the spectral energy is contained in the lower frequencies up to about 10 kHz [18, p. 15 f.]. High levels in this region therefore cause the requantization noise to rise above the absolute threshold in quiet, as approximated by $P_{TQ}(\omega_k)$. Consequently, the noise shaping spectrum produced by the filter determined with $W(\omega_k)$ from (6) shows a decrease in high-frequency gain relative to its logarithmic average [1]. In other words, the noise spectrum tends to "flatten".

In section 1.3, we pointed out the precipitous decline in hearing sensitivity at high frequencies. In this regard, it should be added that at frequencies above 18 kHz, only very few subjects detect any signals [18, p. 18]. Practical power-constrained noise shaping filters such as those proposed in this paper or in [4], [5] thus clearly remain below the hearing threshold at such high frequencies.

From this observation, it is safe to conclude that the requantization noise produced by an adaptive noise shaper can be kept at a relatively high level for the range from approximately 15 kHz up to $f_S/2$. Furthermore, it can be assumed that audible consequences will not occur at any reasonable playback level. In fact, an overall **reduction** in noise loudness can be expected as more noise energy is shifted to the less audible upper end of the spectrum.

2.2 An Improved Time-Variant Noise Shaper

To realize and verify the above concept, a modification of the time-variant noise shaper in [8] was undertaken.

[1] Remember that the filter is minimum-phase. The noise spectrum thus averages to a constant level which depends solely on the output word-length of the requantizer and the type of dither used.

A full description of our implementation would exceed the scope of this paper, hence only the essential aspects are presented. In principle, the achievement of constant requantization noise power by fixation of the high-frequency filter response represents a numerically complex problem. We therefore opted for an approximation of the above properties by "stretching" the amplitude spectrum of $P_{TQ}(\omega_k)$ based on the spectral content of $P_{XX}(\omega_k)$. For this purpose, the portion of the spectral power in $P_{XX}(\omega_k)$ lying above the threshold defined by $P_{TQ}(\omega_k)$ was determined by

$$S_k = \sum_{k=0}^{\frac{N}{2}-1} \left(\log_{10} d(k),\ d(k) > 1\right), \quad N = 512\,, \qquad (7)$$

where

$$d(k) = \frac{P_{XX}(\omega_k)}{P_{TQ}(\omega_k)}\,.$$

$P_{TQ}(\omega_k)$ was then expanded in magnitude range to yield the fundamental spectrum for the instantaneous masking threshold

$$P'_{TQ}(\omega_k) = P_{TQ}(\omega_k)^{1+\frac{S_k}{S_{max}}}, \quad S_{max} = 2.7\frac{N}{2}\,. \qquad (8)$$

Finally, $P_{TQ}(\omega_k)$ was replaced with $P'_{TQ}(\omega_k)$ in equation (6) to attain $W(\omega_k)$, that is, the inverse of the desired improved filter spectrum. A brief discussion of additional important changes to the system in [8] (and [7]) follows.

- The filter order used for $H(z)$ was increased from 9 to 12. It was observed that a 12th-order design yielded slightly better overall sonic performance. Raising the order even more did not result in noticeable improvements, though. In fact, orders in excess of roughly 20 degraded the sound quality because low-frequency noise became audible.

- To avoid artifacts caused by abrupt changes in the filter structure, interpolation between successive filter computations was performed. In contrast to the systems in [7], [8], however, the filter coefficients for each sample were interpolated directly. While this obviously does not guarantee a minimum-phase filter design for each output sample, (as the case when smoothing is applied in the autocorrelation domain) it yields a significant speed improvement due to reduced numerical complexity without causing any audible impairment.

- A different filter design was used to approximate the threshold in quiet. Two alternatives to the F-weighting FIR filter [5] are presented above. To identify the one producing the least audible noise spectrum, that is, the optimal choice for $P_{TQ}(\omega_k)$, we conducted a listening test, as described below.

3 LISTENING EXPERIMENTS

To analyze and evaluate the subjective performance of the time-invariant and time-variant noise shaper designs presented in the foregoing, two listening tests on different types of transducers were performed. The tests were prepared and executed according to the MUSHRA (MUlti-Stimulus test with Hidden Reference and Anchor) methodology [19], [20]. It should be noted, however, that the following changes were applied to the recommended procedure in order to decrease the variance of the results:

- Instead of the 3.5-kHz low-pass filtered anchor stimulus, a dithered requantized signal with a flat error spectrum [1] was chosen. Details follow.

- The test subjects were asked to assign a grade of zero to the stimulus which they judged as having the lowest overall quality. The recommended instructions [19] do not require this explicitly.

Due to its high sonic quality and diversity, an 11-second excerpt from the violin solo in Rebecca Pidgeon's rendition of "Spanish Harlem" [21] was selected as material for the test. To reduce the duration of the tests, no other material was presented to the listeners.

For both listening tests, the absolute reproduction level used for stimulus presentation was defined as the sound pressure level at which the noise caused by dithered requantization to 12 bits [1] was barely audible in quiet to all three authors of this paper. Individual adjustment of this level by the test subjects was not allowed. The tests took place in a quiet room free from air conditioning noise or ground hum. All signals (reference, anchor, and systems under test) were stored as 16-bit, 44.1-kHz, stereo, PCM WAVE files and played from a Samsung YP-U2 Digital Audio Player. **Table 2** specifies the individual noise shapers used in the generation of the test signals.

3.1 Test 1: Time-Invariant Noise Shapers

In the first listening experiment, we compared the performance of the non-adaptive filter designs from section 1.3 with respect to the 9th-order F-weighting design [5] and to simple requantization without noise shaping [1]. 17 subjects, aged 20 to 49, participated in the test. For each person, the test signals were sorted in random order and presented over two Genelec S30D monitor loudspeakers directly connected to the YP-U2 player. To comply with the recommendation [22], the speakers were positioned to form an equilateral triangle with the subject. The subjects were allowed to move their head during playback. Playback was controlled by each subject via the YP-U2.

[1] requantization using TPDF dither without noise-shaping feedback

After the test, the judgments from two individuals were excluded from analysis due to the following reasons:

- One subject was unable to detect the hidden reference signal reliably.
- One subject did not give the lowest grade to the hidden anchor (no noise shaping). All of the other 16 subjects did.

3.2 Test 2: All Discussed Noise Shapers

The objective of the second listening experiment was to assess the subjective performance of the signal-adaptive noise shaper proposed in section 2.2 in comparison with time-invariant filter designs. For this purpose, all signals from the first test were included, with the addition of the time-variant requantized version (see Table 2 for details).

10 experienced listeners, aged 20 to 49, were chosen for this test. The average age was 27 years. The test signals were played back on Sennheiser HD 590 open-air headphones directly connected to the YP-U2. All results were consistent, thus no data had to be omitted from analysis.

The second test was conducted after the first was completed. This allowed us to use the insights gained in the first test to select the "best" threshold spectrum $P_{TQ}(\omega_k)$ for the psychoacoustic model in the time-variant system.

Signal	Bit Res.	Applied Noise Shaper Design
Reference	16-bit	none (original signal from CD)
Anchor	8-bit	none (no noise shaping applied)
F-Weighting	8-bit	9th-order inverse F-weighting [5]
HF ITU 468	8-bit	8th-order HF modified ITU-R 468
DF UEN Thr.	8-bit	8th-order DF corrected UEN at Thr.
Time-Variant	8-bit	12th-order signal-adaptive design

Table 2. Characteristics of the signals presented in the tests. All noise shapers are FIR filters. See also Table 1.

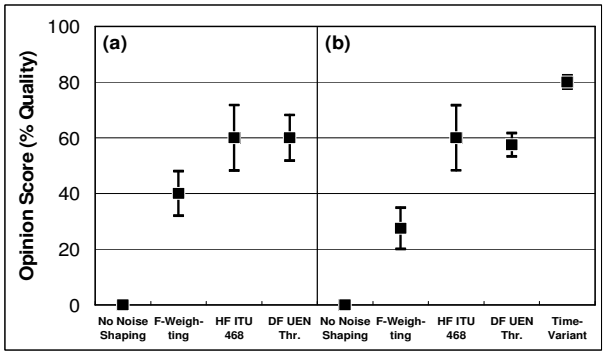

Figure 6. Results of the two listening experiments. (a) loudspeaker test, (b) headphone test, squares show mean scores, bars denote respective 95% confidence intervals.

4 EXPERIMENTAL RESULTS, ANALYSIS

Figure 6 illustrates the results of the two listening experiments. For each noise shaper under test, the arithmetic average of the individual assessments (the mean opinion score, MOS) is given. A significance level of 95% was used to determine the confidence interval for each MOS.

4.1 Results 1: Time-Invariant Noise Shapers

While the F-weighting noise shaper yields a significant perceptual improvement over plain requantization without noise shaping, it is clearly outperformed by the two static filter designs presented in this paper. In fact, only one individual graded the F-weighting filter higher than the proposed noise shapers. The subject reported disturbing hiss at very high frequencies as the reason for this anomalous judgment. This indicates that the subject has excellent upper-band hearing and implies that the high-frequency response of our filter designs might need to be slightly adjusted to satisfy a wider audience.

The time-invariant noise shapers proposed in this paper attained almost identical MOS figures. Their confidence intervals, however, differ considerably. Because the *DF corrected UEN at threshold* design produced the smaller interval of the two, we conclude that it represents a closer approximation of the average auditory threshold for noise when loudspeaker reproduction is chosen. Hence, we decided to apply the *DF corrected UEN at threshold* curve as the fundamental threshold spectrum $P_{TQ}(\omega_k)$ in the signal-adaptive requantizer, which was evaluated in the second listening test.

4.2 Results 2: All Discussed Noise Shapers

The headphone test continues the trend observed in the loudspeaker test. Requantization without noise shaping was scored lowest by all subjects. The F-weighting filter achieved satisfactory results, but its overall quality was rated lower than that of the three filter designs presented in this paper, most likely because of the aforementioned spectral peak around 8 kHz.

Again, the *DF corrected UEN at threshold* filter showed a lower confidence interval than the *HF modified ITU-R 468* design. Further investigation revealed that the performance of the latter approach strongly depends on the playback SPL. This confirms our notion that the *DF corrected UEN at threshold* filter is the better choice.

The adaptive noise shaper design was judged superior to all other implementations by most subjects. Nonetheless, a clear advantage over the static filter designs was only observed at moderate and high signal levels. As reported by most subjects, low-level noise similar to that in other stimuli became audible during quiet signal passages.

5 CONCLUSION

This paper explored the principles of psychoacoustically optimized requantization with regard to high-resolution multi-channel digital audio. Some approaches have been presented, along with a discussion of the potential drawbacks of these designs. Through careful analysis of key psychoacoustic concepts, we were able to derive a set of improved noise-shaping requantizer designs. As alternatives to the F-weighting and similar noise shapers, two static filter structures were presented. The superior subjective performance of these designs was confirmed via two listening tests. In addition, we proposed an adaptive (time-variant) system which, during signal segments of moderate or high level, demonstrably reduces the apparent loudness of the requantization noise even further.

One aspect of time-variant requantization, however, must not be underestimated. The loudness of the noise created by a signal-adaptive noise shaper is, by design, modulated by the input signal. In surround-sound applications, the implication is that when spectral content differs distinctly between channels, the noise will tend to roam the stereophonic image. While partial unmasking will occur only at very low input powers (the noise cannot be fully masked by signals which are too quiet), signal fidelity may still be affected. We therefore believe that adaptive noise-shaping requantization should be used with care.

6 ACKNOWLEDGMENTS

The authors wish to thank the numerous participants of the listening experiments for their support and valuable insights into their perception of the sound excerpts they had to listen to so many times.

7 REFERENCES

[1] C. Roads, *The Computer Music Tutorial*, 6th printing, Cambridge, MA: MIT Press, 2002.

[2] B. Katz, "Dither," Digital Domain, FL, Apr. 2007. <http://www.digido.com/bob-katz/dither.html>

[3] R. A. Wannamaker, S. P. Lipshitz, J. Vanderkooy, and J. N. Wright, "A Theory of Nonsubtractive Dither," *IEEE Trans. Signal Processing*, vol. 48, no. 2, pp. 499–516, Feb. 2000.

[4] S. P. Lipshitz, J. Vanderkooy, and R. A. Wannamaker, "Minimally Audible Noise Shaping," presented at the 88th AES Convention, Montreux, Switzerland, Mar. 1990, *J. Audio Eng. Soc.*, vol. 39, no. 11, pp. 836–852, Nov. 1991.

[5] R. A. Wannamaker, "Psychoacoustically Optimal Noise Shaping," presented at the 89th AES Convention, Los Angeles, Sep. 1990, *J. Audio Eng. Soc.*, vol. 40, no. 7/8, pp. 611–620, July 1992.

[6] U. Zölzer, *Digital Audio Signal Processing*, 3rd printing, Chichester, England: Wiley, 1997.

[7] M. Akune, R. M. Heddle, and K. Akagiri, "Super Bit Mapping: Psychoacoustically Optimized Digital Recording," presented at the 93rd AES Convention, San Francisco, Oct. 1992.

[8] W. Verhelst and D. De Koning, "Least Squares Theory and Design of Optimal Noise Shaping Filters," presented at the 22nd AES International Conference, Espoo, Finland, June 2002.

[9] M. Gerzon and P. G. Craven, "Optimal Noise Shaping and Dither of Digital Signals," presented at the 87th AES Convention, NY, Oct. 1989.

[10] A. Lukin, "Comparison of Word Length Reduction Systems for Digital Audio," 2nd edition, 2002. <http://audio.rightmark.org/lukin/dither/dither.htm>

[11] 24-96 Mastering, "The Great Dither Shootout," May 2006. <http://www.24-96.net/dither/>

[12] International Org. for Standardization, "ISO 226: 2003: Acoustics – Normal Equal-Loudness-Level Contours," Geneva, Switzerland, Aug. 2003.

[13] International Telecomm. Union, Radiocommunication Assembly, "Recommendation ITU-R BS. 468-4: Measurement of Audio-Frequency Noise Voltage Level in Sound Broadcasting," 1986.

[14] J. R. Stuart, "Coding High Quality Digital Audio," Meridian Audio Ltd., UK, Dec. 1997. PDF, <www.meridian-audio.com/ara/coding2.pdf>

[15] J. R. Stuart, "Noise: Methods for Estimating Detectability and Threshold," presented at the 94th AES Convention, Berlin, Germany, Mar. 1993, *J. Audio Eng. Soc.*, vol. 42, no. 3, pp. 124–140, Mar. 1994.

[16] The Wikimedia Foundation Inc., "Equal-loudness contour," *Wikipedia, the free encyclopedia*, 2007. <http://en.wikipedia.org/wiki/Equal-loudness_contour>

[17] Y. Suzuki et al., "Precise and Full-range Determination of Two-dimensional Equal Loudness Contours," Tohoku University, Japan, 2003. <http://www.nedo.go.jp/itd/grant-e/report/00pdf/is-01e.pdf>

[18] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*, 2nd edition, Berlin, Germany: Springer-Verlag, 1990.

[19] International Telecomm. Union, Radiocomm. Assembly,"Recommendation ITU-R BS.1534-1: Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems," 2003.

[20] E. Vincent, "MUSHRAM: A MATLAB interface for MUSHRA listening tests," 2005. User guide, <http://www.elec.qmul.ac.uk/people/emmanuelv/mushram/>

[21] R. Pidgeon, "Spanish Harlem," from the album *Retrospective*, Chesky Records, 2003. CD track 1.

[22] International Telecomm. Union, Radiocommunication Assembly, "Recommendation ITU-R BS. 1116-1: Method for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems," 1997.