ON PSYCHOACOUSTIC NOISE SHAPING FOR AUDIO REQUANTIZATION

Dreten De Koning and Werner Verhelst

Vrije Universiteit Brussel, dept. ETRO-DSSP
Pleinlaan 2, B-1050 Brussels, Belgium
wverhels@etro.vub.be

ABSTRACT

Signal requantization to reduce the word-length of an audio stream introduces distortions. Noise shaping can be applied in combination with a psychoacoustic model in order to make requantization distortions minimally audible. The psychoacoustically optimal noise shaping curve depends on the time-varying characteristics of the input signal. Therefore, the noise shaping filter coefficients are to be computed and updated on a regular basis. In this paper, we present a least squares theory for optimal noise shaping of audio signals. It provides shorter and more straightforward proof of known properties, and in contrast with the standard theory, it does show how noise shaping filters that attain the theoretical optimum can be designed in practice.

1. INTRODUCTION

Signal requantization is applied in digital audio systems whenever the word-length of audio samples needs to be reduced. This is the case for instance when an audio signal has to be stored on a CD and was originally produced at the output of a digital audio system that operates with more than 16 bit precision. In some applications, like multimedia, gaming, or mobile communication devices, requantization to 8 bit or 12 bit could be an economically interesting alternative to other forms of data compression because requantized data can be send directly to the D/A converter, while encoded data requires a decoder.

Signal requantization inevitably introduces an error, which can cause two types of audible problems. The first is a background noise that may be audible by itself. It can usually occur when (part) of the error signal is uncorrelated with the original audio. When the error is correlated with the signal, linear or nonlinear distortions may cause alterations in the perceived quality of the signal itself. At low signal levels, this second problem is usually much more serious [1]. Dither noise can be used to remove the correlation between the error and the signal at the expense of increased noise energy. The standard choice for the dither signal is a random noise source with a triangular distribution between -1 LSB and +1 LSB [2]-[4].

With or without dither, the requantization error can be made minimally audible by proper noise shaping. It suffices to change the shape of the error spectrum such that it becomes minimally audible in the presence of the audio signal. The specifications for the optimal noise shaping filter can be determined from the input signal using a psychoacoustic model. However, these design specifications are time-varying since they depend on the global masking properties of the audio input. Thus, because the noise shaping filter coefficients need to be updated on-line, it is crucial that an efficient filter design technique be used.

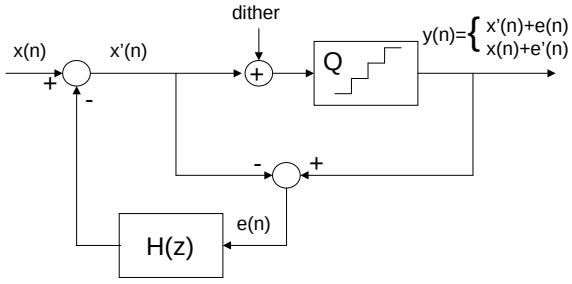

Fig. 1. Dithered requantization with error feedback filter $H(z)$ and requantization error $e'(n)$

In this paper, we present a Least Squares (LS) theory for optimal noise shaping of audio signals that provides shorter and more straightforward proof of known properties of dithered and non-dithered noise shaping. In contrast with the standard theory, this approach shows how noise shaping filters that attain the theoretical optimum can be designed in practice. We also present results from an experimental noise shaping system for minimally audible signal requantization that is based on our filter design method and a simple masking model. In listening experiments, this system was unanimously preferred over the alternatives which included straightforward requantization, dithered requantization and optimized fixed noise shaping [2], [3].

The rest of this paper is organized as follows. The standard theory of noise shaping [5] and minimally audible dither signals [2]-[4] is reviewed in section 2. The LS theory and the practical design method that it entails are described in section 3. The experiments described in section 4 show unanimous preference amongst listeners for our experimental noise shaping system. Finally, section 5 concludes the paper.

2. REQUANTIZATION AND NOISE SHAPING

2.1. General concept

While a white noise dither signal can already improve the quality of low level requantized signals, noise shaping can additionally be applied in order to make the requantization error minimally audible [2]. Fig. 1 illustrates the general scheme for signal requantization with noise shaping. In this scheme, Q represents the quantizer and $H(z)$ is the error feedback filter. Due to the requantization error $e(n)$, the output $y(n)$ differs from $x'(n)$ and from $x(n)$. The

error feedback filter has to be controlled such that the difference between $y(n)$ and $x(n)$ becomes minimally audible.

With signals defined as shown in Fig. 1, and using z-transforms, we have

$$X'(z) = X(z) - H(z)E(z) \quad (1)$$
$$X'(z) = X(z) + E'(z) - E(z) \quad (2)$$

where $E(z)$ represents quantizer Q's error signal and $E'(z)$ is the additive quantization distortion at the output of the noise shaping requantizer. Subtracting equation (2) from (1), one finds that

$$E'(z) = (1 - H(z))\, E(z) \quad (3)$$

The requantization error $e'(n)$ therefore has power spectrum

$$P_{E'}(e^{j\omega}) = \|1 - H(e^{j\omega})\|^2 P_E(e^{j\omega}) \quad (4)$$

Note that, if a properly dithered quantizer Q is used, the power spectrum of the quantization error $e(n)$ is constant.

In order to achieve minimal audibility of the requantization error, $H(e^{j\omega})$ can be designed to minimize the total amount of perceptually weighted noise power N_w:

$$N_w = \int_{-\pi}^{+\pi} P_{E'}(e^{j\omega}) W(\omega) \mathrm{d}\omega \quad (5)$$

$W(\omega)$ is a perceptual weighting function that approximates the relative audibility of noise power at the different frequencies.

2.2. The Gerzon-Craven theory [5]

Let us assume that the desired shape $P_d(e^{j\omega})$ of the error spectrum is given. Thus, from eq. (4), the noise shaping filter $H(z)$ has to be determined such that

$$\|1 - H(e^{j\omega})\|^2 P_E(e^{j\omega}) = \alpha P_d(e^{j\omega}) \quad (6)$$

with minimal α. In principle, there are several filters $H(z)$ that satisfy eq. (6); the different solutions correspond to noise shaping filters $1 - H(z)$ with a same power spectral shape and different phase characteristics. Based on information theoretic considerations, it was proven by Gerzon and Craven that the noise shaping filter $1 - H(z)$ that satisfies (6) with the smallest possible output error power is the filter that leaves the information capacity of the channel unaltered (maximum), and that this is attained when the filter is minimum phase.

Therefore, it was suggested that a practical method to design optimal noise shaping filters would be to use any filter design program to approximate the desired spectral shape and mirroring any zeroes that lie outside the unit circle (for reasons of stability and causality, all poles would already have to be inside the unit circle).

2.3. Some noise shaping implementations

Super Bit Mapping (SBM) [6] follows this suggested strategy and introduces a clever trick to design a minimum phase FIR noise shaping filter with a given power spectral shape.

Note that in order to avoid delayless loops in Fig. 1, it is required that the FIR noise shaping filter can be written as

$$1 - H(z) = \sum_{n=0}^{M} a(n) z^{-n}, \quad \text{where } a(0) = 1. \quad (7)$$

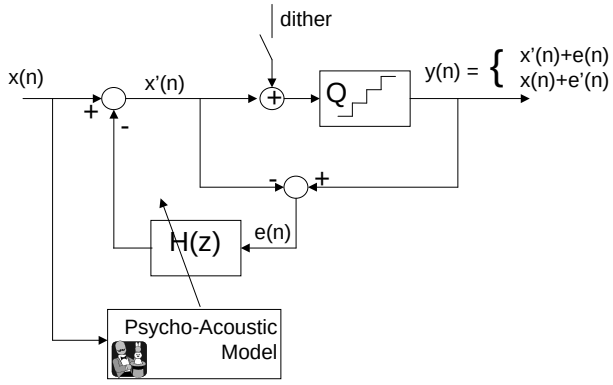

Fig. 2. Psychoacoustic requantization with time-varying noise shaping filter $H(z)$ derived from a perceptual masking model

It was observed in [6] that an M^{th}order inverse LPC filter is minimum phase (if obtained from the autocorrelation formulation [7]) and satisfies (7). Thus, the required minimum phase FIR noise shaping filter can be obtained by approximating the inverse of the desired noise shaping spectrum with an LPC synthesis filter and inverting the result.

In SBM-1, the desired noise shaping spectrum is taken to be the hearing threshold in quiet. Although SBM-1 can be successfully applied to make the quantization error minimally audible in quiet, it could be preferable to use a spectral shaping that minimizes the audibility of the requantization error in the presence of the actual audio. As illustrated in Fig. 2, in SBM-2 the noise shaping filter is a time-varying filter that is designed to approximate the instantaneous masking threshold of the signal. In order to avoid artifacts from abruptly changing error feedback filters, smoothing of the time-varying filter characteristics is applied in the autocorrelation domain.

The design of $H(e^{j\omega})$ by optimization of (5) with numeric optimization techniques has also been considered [2],[3]. As this is a rather difficult problem, only fixed noise shaping filters have been designed in this way. Because the noise is obviously most audible in quiet fragments, the perceptual weighting function $W(\omega)$ was approximated by the inverse of an equi-loudness curve. The so-called E-weighting and F-weighting models for the 15-phon audibility function have been used as the perceptual weighting function in [2] and [3], respectively. As in SBM-1, this makes the noise shaping filter independent of the input signal, thus avoiding the need for on-line optimization of (5). Interestingly, [2] and [3] use a dithered quantizer, while [6] uses non-dithered quantization.

3. THE LEAST SQUARES THEORY

From equations (5), (4), and (7) the optimal noise shaping filter can be found by minimizing

$$E_{NS} = \int_{-\pi}^{+\pi} \| \sum_{n=0}^{M} a(n) e^{-j\omega n} \|^2 P_E(e^{j\omega}) W(\omega) \mathrm{d}\omega \quad (8)$$

Observe that $W(\omega)$ is a perceptual weighting function. Therefore, $W(\omega)$ is real and positive, such that we can define $V(\omega) = \sqrt{W(\omega) P_E(e^{j\omega})}$, and denote its inverse Fourier transform by $v(n)$.

Parseval's theorem can now be applied to transform the problem to time domain: E_{NS} equals the energy contained in the sequence that would be obtained by filtering $v(n)$ with the noise shaping filter, i.e.,

$$E_{NS} = \sum_{n=-\infty}^{+\infty} (\sum_{k=0}^{M} a(k)v(n-k))^2 \quad (9)$$

This quantity can be straightforwardly minimized by requiring that

$$\frac{\partial E_{NS}}{\partial a(k)} = 0, \quad k = 1 \dots M, \quad (10)$$

which leads to the so-called normal equations:

$$\mathbf{Ra} = -\mathbf{r} \quad (11)$$

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

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

$$r(i) = \sum_{n=-\infty}^{+\infty} v(n)v(n-i) \quad (12)$$

Equations (11) and (12) are the so-called autocorrelation formulation of the LPC analysis of signal $v(n)$. The properties of these equations have been studied for several digital signal processing applications in the past, e.g., for speech processing in [7]. For example, it is well known that the solution is indeed a minimum phase filter [8]. Thus, our theory provides an alternative and straightforward way to prove Gerzon and Craven's proposition. As another example, it is also known that the weighted and shaped error spectrum $\|1 - H(e^{j\omega})\|^2 P_E(e^{j\omega})W(\omega)$ will be maximally flat, such that the noise shaping filter spectrum will approximate the inverse of the weighted error spectrum $W(\omega)P_E(e^{j\omega})$.

Further, this least squares theory also shows how truly optimal noise shaping filters can be designed. Indeed, linear matrix equation (11) can be easily solved with any of a number of well-known methods and produces the required optimal noise shaping filter. Furthermore, $\mathbf{R}$ being a symmetric Toeplitz matrix, there is a choice of efficient and robust solution algorithms for solving (11) (such as [9], for example).

Traditionally, the quantizer has been assumed to produce a white error signal $e(n)$. In that case, the autocorrelation function of $v(n)$ is by definition equal to the inverse Fourier transform of the perceptual weighting function $W(\omega)$. In practice, $r(i), i = 0 \dots M$ can therefore be approximated by sampling $W(\omega)$ and computing the inverse FFT:

$$r(i) = \frac{1}{N} \sum_{k=0}^{N-1} W(\frac{2\pi}{N}k) e^{j\frac{2\pi}{N}ki}, \quad i = 0 \dots M \quad (13)$$

By using enough frequency samples $N >> M$, the total approximation error can be made arbitrarily small.

Given a desired filter order M, the solution of matrix equation (11) produces the filter that does minimize (5). While numerical optimization of (5) is too slow for on-line applications and may fail to converge in practice, our method is computationally efficient and robust. Furthermore, this method can also be applied in case of non-white quantization error, such as in the non dithered case. Indeed, in that case it suffices to perform the same operations on $W(\omega)P_E(e^{j\omega})$ (instead of $W(\omega)$).

In SBM the inverse noise shaping filter is approximated by applying an LPC modellisation to the inverse of the desired noise shaping spectrum. If we let $W(\omega)$ be equal to the inverse of the desired noise shaping spectrum, then the above theory proves that the SBM filter is the optimum one. However, this is only valid under the assumption that $P_E(e^{j\omega})$ is constant. Because SBM uses a non-dithered quantizer, it is doubtful that this would be satisfied for the more critical low level signals. This suggests that SBM could be improved by applying the LPC modellisation to the inverse of the desired noise shaping spectrum *multiplied by the quantizer's error spectrum* $P_E(e^{j\omega})$ or an estimate thereof.

4. EXPERIMENTS

4.1. Experimental setup

A software version of the psychoacoustical noise shaping requantizer (Fig. 2) was implemented. 44.1 kHz sampling frequency and 9^{th} order FIR designs for $H(z)$ were used. The filter coefficients were obtained by solving (11) as described above. $r(i)$ was approximated by a 512 point inverse FFT of the sampled weighting function

$$W(\omega_k = \frac{2\pi k}{N}), k = 0 \dots N-1; N = 512$$

$W(\omega_k)$ was updated every 256 input samples and corresponded to the inverse masking curve of a simplified psychoacoustic model:

$$W(\omega_k) = \frac{1}{\beta P_{xx}(\omega_k) + (1-\beta)P_{TQ}(\omega_k)} \quad (14)$$

$$\beta = \frac{1}{1 + 10^{-6} N^2 2^{2B-2}} \quad (15)$$

Here $P_{xx}(\omega_k)$ represents the energy spectrum of a 512 point hanning windowed input segment. Its spectral resolution is comparable to the critical bandwidth of hearing at about $3kHz$. $P_{TQ}(\omega_k)$ is the hearing threshold in quiet, and B is the number of bits of the input signal representation. In our experiments $B = 16$ and $P_{TQ}(\omega_k)$ was approximated by the energy spectrum of the 9^{th} order F-weighting filter. As with other perception models, β depends on the expected sound pressure level of the input signal. The value in (15) compensates for the scale factor incurred with our choice of $P_{TQ}(\omega_k)$ and is appropriate when the loudest signal portions are played at 84 dB SPL.

Sixteen bit test data from good-quality CDs was used in the experiments. Six different fragments were used, containing different types of instruments, vocals, and musical styles. They had a total duration of 1 minute 33 seconds. The fragments were first requantized using straightforward requantization (i.e., rounding) to a precision where the requantization errors became clearly audible. The fragments were then requantized to the same precision (between 5 and 7 bits, depending on the fragment) with three additional methods, resulting in four different versions: (1) straightforward requantization; (2) requantization with standard dither; (3) non-dithered requantization with fixed noise shaping (F-weighting); (4) non-dithered requantization with adaptive psychoacoustic noise shaping.

4.2. Informal diagnostic evaluation

Two adults with normal hearing participated in this informal evaluation. They were allowed to listen to the different quantized and original sound fragments in any desired order and as often as they desired. The experiment was performed in a quiet office and the sound files were played from a multimedia PC (Compaq presario 4810) over headphones (AKG K-300). The experimenters discussed their findings with one another and eventually reported the following conclusions.

Version 2 (standard dither) was systematically judged to have lowest quality due to the very audible presence of the high level dither signal. In version 1, signal distortions and quantization noises were also clearly audible, especially in softer segments. Because the noise was less prominent than in version 2 and distortions occured only intermittently, version 1 was prefered over version 2 for all cases. Versions 3 and 4 were found to be of much higher quality than versions 1 and 2. No nonlinear distortions were perceived.[1] In version 3 a weak high-frequency noise was permanently audible. In version 4, a similar but weaker noise could be heard during the softer signal portions only. Version 4 was judged to be better in general than all other versions. Differences between version 4 on the one hand and versions 1 and 2 on the other hand were deemed so clear that they could be evaluated in a classroom experiment.

4.3. Classroom experiment

The classroom experiment was set up as a blind pairwise comparison experiment with forced choice. Twenty-one students, aged 20-25, participated in the listening experiment. Eighteen reported normal hearing, one better than normal and two less than normal hearing. The same six audio fragments as in the informal diagnostic evaluation were used. Two fragments were used to compare version 4 with version 2 (one for each presentation order), and the remaining four fragments were used to compare version 4 with version 1 (two for each presentation order). The audio was played in a normal and quiet classroom from a portable PC (Toshiba Satellite Pro CDT-420) using small active loudspeakers (Philips SBC 8237). The original 16 bit version was always played first, as a reference, before the two test versions. The subjects indicated their preference for one of the test versions by crossing the corresponding column of a table on their response forms.

Twenty properly filled-out response forms were received. They showed that version 4 had been systematically preferred by all listeners in every pair that had been presented. Thus, in a total of 120 comparisons (six fragments, twenty subjects), the adaptive psychoacoustic requantization method was never defeated.

5. CONCLUSION

In this paper, we presented a short review of the state of affairs in noise shaping for audio requantization. We discussed the Gerzon-Craven theory and two classes of noise shaping that are based on it (fixed and adaptive psychoacoustic noise shaping).

We introduced a Least Squares theory for optimal noise shaping of audio signals, and we described an efficient method that allows for on-line design of optimal noise shaping filters and that also applies in case of a non-white quantizer error signal $e(n)$. It was noted that some of the existing noise shaping systems use a dithered quantizer and others use non-dithered quantizers. With our proposed theory, we are basically able to cope with both situations.

In our experiments signal-adaptive psychoacoustic noise shaping without dithering was clearly preferred over straightforward requantization, requantization with standard dithering and requantization with fixed noise shaping. Further work should be done to decide which perception models can best be applied, what type of quantizers should be used (e.g., dithered or non-dithered), what noise shaping filter orders to use, etc. It is our impression that different optimal conditions could apply for different applications (bit precision, signal type and bandwidth, ...). Above all, the dynamic behavior of adaptive noise shaping should be studied.

[1] This lead us to decide to use perceptual noise shaping without dither in the classroom experiment.

Acknowledgments

Support from the Flemish Community through grants from IWT and FWO is gratefully acknowledged. Many thanks to Robert Bristow-Johnson from Wave Mechanics for the stimulating e-mail correspondence that we had on the subject.

6. REFERENCES

[1] Roads, C., The Computer Music Tutorial, third printing, pp. 33-38, MIT Press, Cambridge, Massachussets, 1998.

[2] Lipshitz, S.P., Vanderkooy, J., and Wannamaker R.A.,"Minimally Audible Noise Shaping", J. Audio Eng. Soc. 39 (11): 836-852, 1991.

[3] Wannamaker, R.A., "Psychoacoustically Optimal Noise Shaping", J. Audio Eng. Soc. 40 (7/8): 611-620, 1992 - presented at the 89th AES Convention, Los Angeles, September 21-25, 1990.

[4] Wannamaker, R.A., Lipshitz, S.P., Vanderkooy, J., and Wright J.N., "A Theory of Nonsubtractive Dither", IEEE Trans. on Signal Processing 48 (2): 499-516, February 2000.

[5] Gerzon, M., and Craven, P.G., "Optimal Noise Shaping and Dither of Digital Signals", AES preprint 2822, presented at the 87th AES Convention, New York, October 18-21, 1989.

[6] Akune, M., Heddle, R.M., and Akagiri, K., "Super Bit Mapping: Psychoacoustically Optimized Digital Recording", AES preprint 3371, presented at the 93rd AES Convention, San Fransisco, October 1-4, 1992.

[7] Markel, J.D., and Gray, A.H.Jr., *Linear prediction of speech*, Springer Verlag, New York, 1976.

[8] Deller, J.R. Jr., Proakis, J.G., and Hansen, J.H.L., Discrete-Time Processing of Speech Signals, Chapter 5, Macmillan, New York, 1993.

[9] LeRoux, J., and Gueguen, C., "A fixed point computation of partial correlation coefficients", IEEE Transactions on Acoust., Speech and Signal Processing: 257-259, June 1977.

[10] Verhelst W., and De Koning D., "Noise Shaping Filter Design for Minimally Audible Signal Requantization", proceedings of WASPAA-01, 21-24 October 2001, New Paltz, New York.