FREQUENCY WARPING IN LOW DELAY AUDIO CODING

Stefan Wabnik, Gerald Schuller, Ulrich Krämer and Jens Hirschfeld

Fraunhofer Institute for Digital Media Technology, Ilmenau, Germany

ABSTRACT

The goal of the schemes we present in this paper is to obtain an ultra low delay audio coder with a good performance even at low bit rates (around 64 kb/s). The problem to be solved is to gain sufficient frequency resolution at low frequencies for precise low frequency psycho-acoustics and quantization noise shaping, because the ear has a higher frequency resolution at lower frequencies. Our approach is to use a warped linear noise shaping pre- and post-filter, and a short DFT for the psycho-acoustic model (length 256), but with frequency warping. We compare four different psycho-acoustic versions: DFT with no warping, DFT without warping using warped pre- and post-filters, warping with the so-called NDFT (WDFT), and a DFT with an all-pass delay chain pre-processing. Listening tests show that the best performance is obtained using the WDFT.

1. INTRODUCTION

Our goal is an ultra low delay audio coding scheme (ULD) at low bit rates (around 64 kb/s) with good audio quality. It has 8 to 5.3 ms algorithmic delay at sampling rates of 32 to 48 kHz. To keep the delay low, there is no bit rate buffer, only a bit rate control loop.

The problem to solve here is to obtain a suitable frequency resolution for the psycho-acoustical model. In our approach a size 256 windowed and overlapping DFT is used to keep the delay short, but as a consequence the lower bands are spread too wide (about 1 Bark in the lower bands).

In standard audio coders, a longer DFT is used as input for the psycho-acoustic model. Typically they have 1024 sub-bands (as in MPEG-AAC [1]). The sub-bands are then grouped into 1/3 to 1/4 Bark bands.

To obtain a better frequency resolution also with a shorter length DFT, frequency warping can be used. Two different approaches are known in literature: the warped DFT using a line of all-pass filters as input [2],[3], and a DFT with non-uniformly spaced center frequencies, the WDFT [4],[5].

To determine if warping works and to find out which version of warping works best in a low delay audio coding environment, we conducted listening tests. We compared three different psycho-acoustic warping versions with each other and with the non-warped case: the DFT with no warping but warped noise shaping pre-filter, warping with all-pass filters, and warping with the so-called WDFT as a special case of the NDFT.

2. CODER DESCRIPTION

The ULD Coder separates the two aims of irrelevance and redundancy reduction by assigning them to different functional units [6]. Fig. 1 shows the main functional blocks of the ULD Coder.

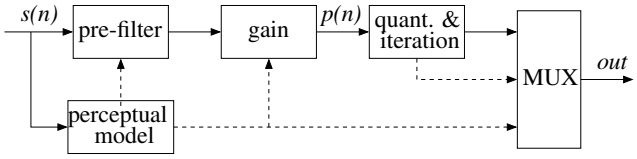

Fig. 1. Basic structure of our ULD Encoder

In the ULD encoder, a psycho-acoustically controlled adaptive linear filter is applied to the input audio signal $s(n)$ for the irrelevance reduction. The psycho-acoustic model incorporates a DFT filter bank with 256 bands and 50% overlap causing a delay of 128 samples. To synchronize the pre-filtering step with the psycho-acoustic model, a corresponding delay is introduced in the signal path. The output of the psycho-acoustic model is an estimation of the masking threshold. This estimation is then transformed into filter coefficients and a gain factor via the Levinson-Durban algorithm. Together, the filter operation and the following filter gain (see Fig. 1) can be interpreted as a normalization of the input signal $s(n)$ to the masking threshold. The decoder contains a post-filter, which is the inverse of the pre-filter, and hence has a frequency response like the masking threshold. Thus, the quantization noise introduced in the iteration and quantization module remains at or below the masking threshold, if $p(n)$ is not scaled by a factor smaller than 1.0 within the iteration module.

The quantization and iteration block employs quantization and redundancy reduction as well as a method to obtain a constant bit rate of the coded signal [7]. After scaling and quantization of the pre-filtered signal $p(n)$, a predictive lossless coding scheme consisting of a backwards adaptive predictor and an entropy coder is used to remove redundancy from the signal [8].

As a final step, both side information and entropy coded audio data are packed into a bit stream with the help of a bit multiplexer.

3. WARPING

Signal processing techniques using warping are based on the application of a first-order (bilinear) frequency mapping on the z-transform of the signal.

$$z \mapsto A(z) = \frac{z - \lambda}{1 - \lambda z} \tag{1}$$

The parameter λ can be chosen such that the resulting mapping strongly resembles the Bark scale for given sampling frequency f_s (see [9])

$$\lambda(f_s) = 1.0674 \left[\frac{2}{\pi} \arctan(0.6583 f_s) \right]^{1/2} - 0.1916 \; . \tag{2}$$

3.1. Warped Filter Structures

Warping FIR and IIR filter structures is done by substituting all delay elements z^{-1} with $A(z)^{-1}$ and always produces IIR filter structures, as $A(z)$ is a first-order all-pass. For IIR filter structures, warping leads to delay-free loops, but in [10] methods to solve this problem were described. Given a spectrum on a warped frequency scale, the procedure to obtain filter coefficients for warped filters is identical to the non-warped case.

3.2. NDFT Filter Bank

The nonuniform DFT (NDFT) proposed in [4] can be used to construct a warped DFT as presented in [5],[11]. With $z_k, 0 \leq k \leq N-1$ denoting N distinct points in the z-plane, the $N-$point NDFT of the length-N sequence $x[n]$ is given by

$$\begin{aligned} X_{\mathrm{NDFT}}(k) &= X(z_k) \\ &= \sum_{n=0}^{N-1} x[n] z_k^{-n}, \qquad 0 \leq k \leq N-1 \,. \end{aligned} \tag{3}$$

With $\mathbf{X}_{\mathrm{NDFT}} = [X_{\mathrm{NDFT}}(0), ..., X_{\mathrm{NDFT}}[N-1]]^T$ and $\mathbf{x} = [x(0), ..., x(N-1)]^T$, the NDFT can be written in matrix form

$$\mathbf{X}_{\mathrm{NDFT}} = \mathbf{D}_{NDFT} \cdot \mathbf{x} \tag{4}$$

$$\text{with} \qquad \mathbf{D}_{NDFT} = \begin{bmatrix} 1 & z_0^{-1} & \dots & z_0^{-N+1} \\ 1 & z_1^{-1} & \dots & z_1^{-N+1} \\ \vdots & \vdots & & \vdots \\ 1 & z_{N-1}^{-1} & \dots & z_{N-1}^{-N+1} \end{bmatrix} .$$

The WDFT is obtained from (3) by setting $z_k = A(e^{j2\pi k/N}) = A(W_N^k)$, which maps N equally spaced points of the unit circle onto N nonequally spaced points.

$$\mathbf{X}_{\mathrm{WDFT}} = \mathbf{D}_{WDFT} \cdot \mathbf{x}, \tag{5}$$

$$\mathbf{D}_{WDFT} = \begin{bmatrix} 1 & A(W_N^0)^{-1} & \dots & A(W_N^0)^{-N+1} \\ 1 & A(W_N^1)^{-1} & \dots & A(W_N^1)^{-N+1} \\ \vdots & \vdots & & \vdots \\ 1 & A(W_N^{N-1})^{-1} & \dots & A(W_N^{N-1})^{-N+1} \end{bmatrix} .$$

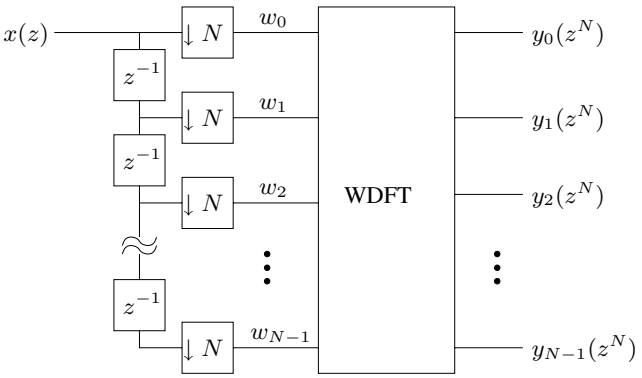

Fig. 2. WDFT Filterbank

To preserve signal power in corresponding parts of the unit circle, a correction factor c_k has to be applied to each WDFT frequency coefficient [11]

$$c_k = \frac{\sqrt{1-\lambda^2}}{1-\lambda W_N^k} \,. \tag{6}$$

The WDFT, together with a blocking structure and a window function, can be used to build a warped filter bank, as shown in Fig.2. Notice that the a blocking structure is not warped.

Fig. 3 shows the logarithmic magnitude response of a WDFT-filterbank for $N = 32$, a sine window and a warping parameter $\lambda = -0.6865$, chosen to match the bark scale for a sampling frequency of $f_s = 32kHz$. It can be seen that for $\lambda = -0.6865$, the lower bands overlap while there are gaps between higher filter bands. For the perceptual model these gaps can be a potential problem, because signals in these gaps cannot contribute to the calculation of the masking threshold, which hence could be too conservative at those frequencies.

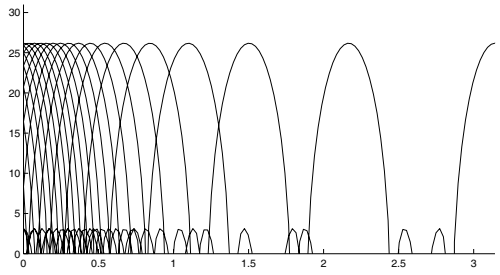

Fig. 3. WDFT-filterbank with $N = 32$, $\lambda = -0.6865$

3.3. Pre-processed DFT Filter bank

Oppenheim et. al [2] suggested to use a warped delay line as a blocking structure together with a DFT (hereafter named Warped Delay Line DFT, WDLDFT) to implement a filter bank with nonuniform frequency resolution (see Fig.4). Additionally, to preserve power or magnitude of the signal spectrum, an additional filter before the filter bank can be used [3],[12].

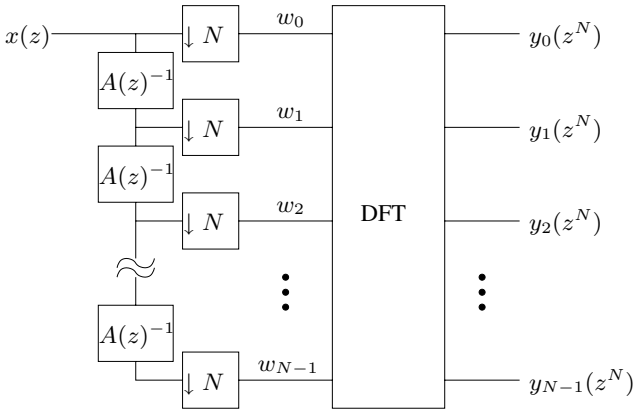

Fig. 4. WDLDFT Filterbank

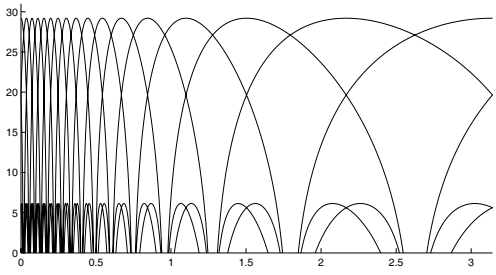

Fig. 5. WDLDFT-filter bank with $N = 32$, $\lambda = -0.6865$

Figure 5 shows the logarithmic magnitude response of a DFT-filterbank with warped delay elements $A(z)^{-1}$ for $N = 32$, a rectangular window and a warping parameter$\lambda = -0.6865$. In contrast to the WDFT-filter bank, the overlap ratios of all bands follow their relative bandwidths.

3.4. Temporal Properties

Figures 6 and 7 show the impulse responses of a WDLDFT-filter bank and a WDFT-filter bank with $N0256$ for a lower ($k = 10$) and an upper ($k = 100$) filter band. Whereas for the WDFT the length of all impulse responses are of length N, for the WDLDFT the length varies from about $0.25N$ for high bands to about $5N$ for low bands. The long impulse responses of the WDLDFT at low frequencies can be a problem in audio coding. They result in a low time resolution and hence temporal smearing of the quantization error. This can result in audible pre-echo artifacts.

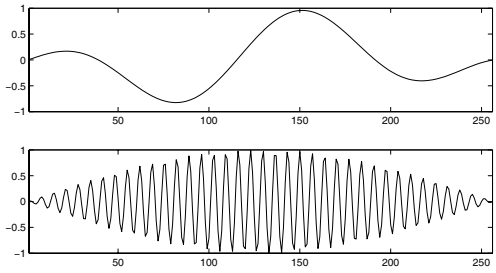

Fig. 6. WDFT-filter bank,$N = 256, \lambda = -0.6865$, impulse response of $10th$ band (upper) and $100th$ band(lower)

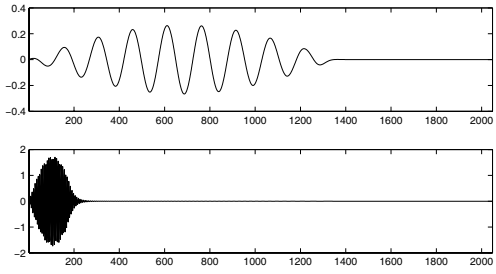

Fig. 7. WDLDFT-filter bank,$N = 256, \lambda = -0.6865$, impulse response of $10th$ band (upper) and $100th$ band(lower)

4. WARPING AND ULD

We can see that both the WDFT and WDLDFT versions have their advantages and shortcomings. The interesting question is: is any of them better in audio coding than the un-warped case, and if so, which one? To find out, we implemented them in the ULD coder and conducted a listening test.

Both warped noise shaping pre- and post-filter structures and warped DFT for psycho-acoustics are implemented in the ultra low delay coding scheme (ULD). The delay-free loops in the warped IIR noise shaping post-filter are treated as described in [10]. For this paper, four different setups have been implemented. Setup *A* uses a windowed linear DFT filter bank for the psycho-acoustic model and linear noise shaping pre- and post-filters. Setup *B* uses the same filter bank as the first setup, but re-samples the calculated masking threshold to obtain a warped frequency resolution, and uses warped pre- and post-filters. Setup *C* uses a WDLDFT-filter bank and warped noise shaping filters, and setup *D* uses an WDFT-filter bank and warped noise shaping filters.

5. LISTENING TEST

This section presents the results of our subjective listening test conducted according to the MUSHRA standard [13]. The MUSHRA test was implemented on a Laptop computer with external DA-converter and STAX amplifier/headphones in a quite office environment. Our group of eight test listeners consisted of expert and non-expert listeners. Before the subjects started with the listening test, they had the possibility to listen to a test set.

The tests were conducted with 12 mono audio files of the MPEG test set: es01 (Suzanne Vega), es02 (male speech, german), es03 (female speech, english), sc01 (trumpet), sc02 (orchestra), sc03 (pop music), si01 (cembalo), si02 (castanets), si03 (pitch pipe), sm01 (bagpipe), sm02 (glockenspiel), sm03 (plucked strings). The audio files, with a sampling frequency of $32\,kHz$, were coded at a constant bit rate of $64\,kb/s$ with the four setups. In Fig. 8 the results of the MUSHRA listening test, including 95%-confidence intervals as bars, reference file and anchors, are presented.

The most surprising result of the listening test is that there is no sound item for which the frequency gaps of the WDFT lead to a reduced performance. On the other hand, it has the best performance for sound item si02 (castanets), which has the most pronounced attacks. The long low frequency impulse responses of the WDLDFT leads to significant pre-echos in this case. Overall, the WDFT is the clear winner, compared to the WDLDFT and both the setups of only noise shaping filter warping and no warping. These results coincided with additional PEAQ measurements [14].

6. CONCLUSIONS

Both WDFT and the delay-line warped DFT yield higher frequency resolution at lower frequencies, compared with the non-warped case. The WDFT has the disadvantage, that it can lead to gaps in the spectrum which are not covered by it. The delay-line warped DFT has the disadvantage of long impulse responses at low frequencies, and hence a low time resolution.

We found the gaps of the WDFT do not lead to a deterioration of the audio quality, but the reduced time-resolution of the delay-line warped DFT does lead to a reduced performance. In

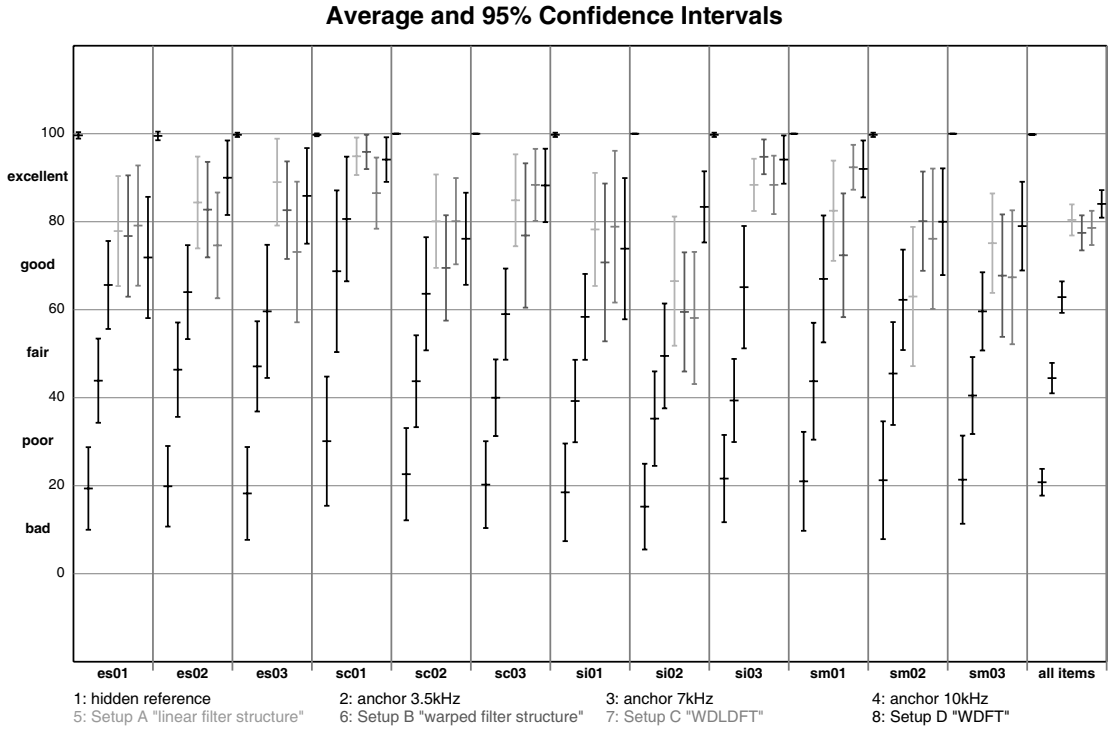

Fig. 8. Listening test results for the different setups in the listed order

our listening test, the use of the WDFT yields statistically significant better results for attack like signals, as castanets, because of the short impulse responses and good time resolution. For stationary signals like bagpipes, also the delay-line warped DFT is better than the un-warped case, because here the improved low frequency resolution is important. Warping of the noise-shaping pre- and post-filter only, without warping the psycho-acoustics DFT, does not give an advantage.

7. REFERENCES

[1] ISO/IEC JTC1/SC29/WG11 (MPEG), "Generic Coding of Moving Pictures and Associated Audio: Advanced Audio Coding," International Standard ISO/IEC IS 13818-7, 1997.

[2] A.Oppenheim, D. Johnson, and K. Steiglitz, "Computation of spectra with unequal resolution using the fast Fourier transform," *Proc. IEEE*, vol. 59, no. 6, pp. 299 – 301, 1971.

[3] C.Braccini and A. V. Oppenheim, "Unequal bandwidth spectral analysis unsing digital frequency warping," *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol. 22, pp. 236 – 244, 1974.

[4] S. Bagchi and S. K. Mitra, *The Nonuniform Discrete Fourier Transform and its Applications in Signal Processing*, Kluver Academic Publishers, Boston, Dordrecht, London, 1999.

[5] Anamitra Makur and Sanjit K.Mitra, "Warped Discret-Fourier Transform: Theory and Applications," *IEEE Transactions on Circuits and Systems*, vol. 48, no. 9, pp. 1086 – 1093, September 2001.

[6] B. Edler, C. Faller, and G. Schuller, "Perceptual Audio Coding Using a Time-Varying Linear Pre- and Post-Filter," *109th AES convention*, September 2000, Los Angeles, CA, USA.

[7] U. Krämer, G. Schuller, S. Wabnik, J. Klier, and J. Hirschfeld, "Ultra Low Delay audio coding with constant bit rate," *117th AES Convention*, October 2004.

[8] G. Schuller and A. Härmä, "Low Delay Audio Compression using Predictive Compression," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2002, Orlando, FL, USA.

[9] II Julius O. Smith and Jonathan S. Abel, "Bark and ERB Bilinear Transforms," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 6, pp. 697 – 708, November 1999.

[10] Aki Härmä, "Implementation of frequency-warped recursive filters," *Signal Processing*, vol. 80, no. 3, pp. 543 – 548, February 2000.

[11] Marek Parfieniuk and Alexander Petrovsky, "Warped DFT as the Basis for Psychoacoustic Model," *ICASSP*, vol. IV, pp. 185 – 188, 2004.

[12] T. von Schroeter, "Frequency Warping with Arbitrary All-pass Maps," *IEEE Signal Processing Letters*, vol. 6, pp. 116 – 118, May 1999.

[13] ITU-R BS.1534-1, "Method for the subjective assessment of intermediate quality levels of coding system," January 2003.

[14] ITU-R BS.1387-1, "Method for objective measurements of perceived audio quality," November 2001.