Perceptually motivated Sub-band Decomposition for FDLP Audio Coding*

Petr Motlicek[12], Sriram Ganapathy[13], Hynek Hermansky[123], Harinath Garudadri[4], and Marios Athineos[5]

[1] IDIAP Research Institute, Martigny, Switzerland
{motlicek,hynek,ganapathy}@idiap.ch
[2] Faculty of Information Technology, Brno University of Technology, Czech Republic
[3] École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
[4] Qualcomm Inc., San Diego, California, USA
hgarudad@qualcomm.com
[5] International Computer Science Institute, Berkeley, California, USA
msa24@columbia.edu

Abstract. This paper describes employment of non-uniform QMF decomposition to increase the efficiency of a generic wide-band audio coding system based on Frequency Domain Linear Prediction (FDLP). The base line FDLP codec, operating at high bit-rates (∼136 kbps), exploits a uniform QMF decomposition into 64 sub-bands followed by sub-band processing based on FDLP. Here, we propose a non-uniform QMF decomposition into 32 frequency sub-bands obtained by merging 64 uniform QMF bands. The merging operation is performed in such a way that bandwidths of the resulting critically sampled sub-bands emulate the characteristics of the critical band filters in the human auditory system. Such frequency decomposition, when employed in the FDLP audio codec, results in a bit-rate reduction of 40% over the base line. We also describe the complete audio codec, which provides high-fidelity audio compression at ∼66 kbps. In subjective listening tests, the FDLP codec outperforms MPEG-1 Layer 3 (MP3) and achieves similar qualities as MPEG-4 HE-AAC codec.

1 Introduction

A novel speech coding system, proposed recently [1], exploits the predictability of the temporal evolution of spectral envelopes of speech signal using Frequency Domain Linear Prediction (FDLP) [2, 3]. Unlike [2], this technique applies FDLP to approximate relatively long segments of the Hilbert envelopes in individual frequency sub-bands. This speech compression technique was later extended to

* This work was partially supported by grants from ICSI Berkeley, USA; the Swiss National Center of Competence in Research (NCCR) on "Inter active Multi-modal Information Management (IM)2"; managed by the IDIAP Research Institute on behalf of the Swiss Federal Authorities, and by the European Commission 6th Framework DIRAC Integrated Project.

high quality audio coding [4], where an input audio signal is decomposed into N frequency sub-bands. Temporal envelopes of these sub-bands are then approximated using FDLP applied over relatively long time segments (e.g. 1000 ms).

Since the FDLP model does not represent the sub-band signal perfectly, the remaining residual signal (carrier) is further processed and its frequency representatives are selectively quantized and transmitted. Efficient encoding of the sub-band residuals plays an important role in the performance of the FDLP codec and this is largely dependent on frequency decomposition employed.

Recently, an uniform 64 band Quadrature Mirror Filter (QMF) decomposition (analogous to MPEG-1 architecture [5]) was employed in the FDLP codec. This version of the codec achieves good quality of reconstructed signal compared to the older version using Gaussian band decomposition [4]. The performance advantage was mainly due to the increased frequency resolution in lower sub-bands and critical sub-sampling resulting in the minimal number of FDLP residual parameters to be transmitted. However, a higher number of sub-bands resulted in higher number of AR model parameters to be transmitted. Hence, the final bit-rates for this version of the codec were significantly higher ($\sim$136 kbps) and therefore, the FDLP codec was not competitive with the state-of-the-art audio compression systems.

In this paper, we propose a non-uniform QMF decomposition to be exploited in the FDLP codec. The idea of non-uniform QMF decomposition has been known for nearly two decades (e.g. [6, 7]). In [8], Masking Pattern Adapted Subband Coding (MASCAM) system was proposed exploiting a tree-structured non-uniform QMF bank simulating the critical frequency decomposition of the auditory filter bank. This coder achieves high quality reconstructions for signals up to 15 kHz frequency at bit-rates between $80 - 100$ kbps per channel. Similar to [8], we propose sub-band decomposition which mimic the human auditory critical band filters. However, proposed QMF bank differs in many aspects, such as in the prototype filter, bandwidths, length of processed sequences, etc. The main contrast between the proposed technique and conventional filter bank occurs in the fact that the proposed QMF bank can be designed for smaller transition widths. As the QMF operates on long segments of the input signal, the additional delay arising due to sharper frequency bands can be accommodated. Such a flexibility is usually not present in codecs operating on shorter signal segments.

Proposed version of non-uniform QMF replaces original uniform decomposition in FDLP audio codec. Overall, it provides a good compromise between fine spectral resolution for low frequency sub-bands and lesser number of FDLP parameters to be encoded. Other benefits of employing non-uniform sub-band decomposition in the FDLP codec are:

- As psychoacoustic models operate in non-uniform (critical) sub-bands, they can be advantageously used to reduce the final bit-rates.
- In general, audio signals have lower energy in higher sub-bands (above 12 kHz). Therefore, the temporal evolution of the spectral envelopes in higher sub-bands require only small order AR model (employed in FDLP). A non-

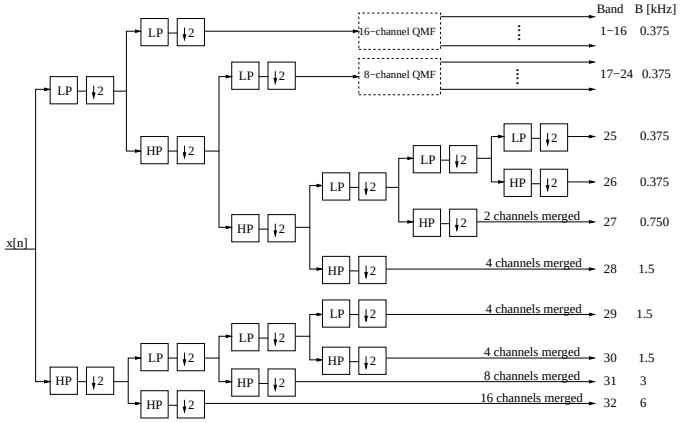


Fig. 1. *The 32 channel non-uniform QMF derived using 6-stage network. Input signal $x[n]$ is sampled at 48 kHz. LP and HP denote Low-Pass and High-Pass band, respectively. $\downarrow 2$ denotes down-sampling by 2. B denotes frequency bandwidth.*

uniform decomposition provides one solution to have same order AR model for all sub-bands and yet, reduces the AR model parameters to be transmitted.

– The FDLP residual energies are more uniform across the sub-bands and hence, similar post-processing techniques can be applied in all sub-bands.

The proposed technique becomes the key part of the high-fidelity audio compression system based on FDLP for medium bit-rates. Objective quality tests highlight the importance of the proposed technique, compared to the version of the codec exploiting uniform sub-band decomposition. Finally, subjective evaluations of the complete codec at ∼66 kbps show its relative performance compared to the state-of-the-art MPEG audio compression systems (MP3, MPEG4 HE-AAC) at similar bit-rates.

This paper is organized as follows. Section 2 describes proposed non-uniform frequency decomposition. Section 3 discusses the general structure of the codec and mentions achieved performances using objective evaluations. Section 4 describes subjective listening tests performed with the complete version of the codec operating at $\sim$ 66 kbps is given. Finally, Section 5 concludes the paper.

2 Non-uniform frequency decomposition

In the proposed sub-band decomposition, the 64 uniform QMF bands are merged to obtain 32 non-uniform bands. Since the QMF decomposition in the base line system is implemented in a tree-like structure (6-stage binary tree [4]), the merging is equivalent to tying some branches at any particular stage to form a non-uniform band. This tying operation tries to follow critical band decomposition

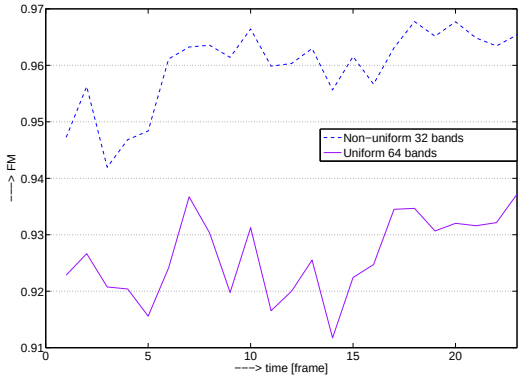


Fig. 2. *Comparison of the flatness measure of the prediction error* $\mathbf{E}$ *for 64 band uniform QMF and 32 band non-uniform QMF for an audio recording.*

in the human auditory system. This means that more bands at higher frequencies are merged together while maintaining perfect reconstruction. The graphical scheme of the non-uniform QMF analysis bank, resulting from merging 64 bands into 32 bands, is shown in Figure 1.

Application of non-uniform QMF decomposition is supported by the Flatness Measure (FM) of the prediction error power E_i (energy of the residual signal of AR model in each sub-band i) computed across N QMF sub-bands. FM is defined as:

$$FM = \frac{Gm}{Am}, \tag{1}$$

where Gm is the geometric mean and Am is the arithmetic mean:

$$Gm = \sqrt[\frac{1}{N}]{\prod_{i=1}^{N} E_i}, \quad Am = \frac{1}{N}\sum_{i=1}^{N} E_i. \tag{2}$$

If the input sequence is constant (contains uniform values), Gm and Am are equal, and $FM = 1$. In case of varying sequence, $Gm < Am$ and therefore, $FM < 1$.

We apply flatness measure to determine uniformity of the distributions of the prediction errors E_i across the QMF sub-bands. Particularly, for a given input frame, vector $\mathbf{E_u} = (E_1, E_2 \ldots E_N)$ obtained for uniform QMF decomposition ($N = 64$) is compared with the vector $\mathbf{E_n} = (E_1, E_2 \ldots E_N)$ containing the FDLP prediction errors for non-uniform QMF decomposition ($N = 32$). For each 1000 ms frame, FM is computed for the vectors $\mathbf{E_u}$ and $\mathbf{E_n}$. Figure 2 shows the flatness measure versus frame index for an audio sample. In case of uniform QMF decomposition, E_i is relatively high in the lower bands and low for higher bands. In case of non-uniform QMF analysis, E_i at higher bands is comparable to those in the lower frequency bands. Such FM curves can be seen for majority of audio samples.

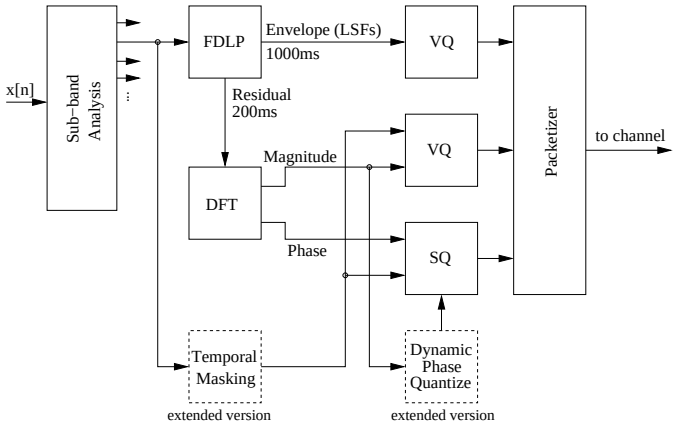


Fig. 3. *Scheme of the FDLP encoder.*

A higher flatness measure of prediction error means that the degree of approximation provided by FDLP envelope is similar in all sub-bands. Therefore, non-uniform QMF decomposition allows uniform post-processing of FDLP residuals.

3 Structure of the FDLP codec

FDLP codec is based on processing long (hundreds of ms) temporal segments. As described in [4], the full-band input signal is decomposed into frequency sub-bands. In each sub-band, FDLP is applied and Line Spectral Frequencies (LSFs) approximating the sub-band temporal envelopes are quantized using Vector Quantization (VQ). The residuals (sub-band carriers) are processed in Discrete Fourier Transform (DFT) domain. Its magnitude spectral parameters are quantized using VQ, as well. Phase spectral components of sub-band residuals are Scalar Quantized (SQ). Graphical scheme of the FDLP encoder is given in Figure 3.

In the decoder, quantized spectral components of the sub-band carriers are reconstructed and transformed into time-domain using inverse DFT. The reconstructed FDLP envelopes (from LSF parameters) are used to modulate the corresponding sub-band carriers. Finally, sub-band synthesis is applied to reconstruct the full-band signal.

3.1 Objective evaluation of the proposed algorithm

The qualitative performance of the proposed non-uniform frequency decomposition is evaluated using Perceptual Evaluation of Audio Quality (PEAQ) distortion measure [9]. In general, the perceptual degradation of the test signal with respect to the reference signal is measured, based on the ITU-R BS.1387 (PEAQ)

standard. The output combines a number of model output variables (MOV's) into a single measure, the Objective Difference Grade (ODG) score. ODG is an impairment scale which indicates the measured basic audio quality of the signal under test on a continuous scale from -4 (very annoying impairment) to 0 (imperceptible impairment). The test was performed on 18 challenging audio recordings sampled at 48 kHz. These audio samples form part of the MPEG framework for exploration of speech and audio coding [10]. They are comprised of speech, music and speech over music recordings. Furthermore, the framework contains various challenging audio recordings indicated by audio research community, such as tonal signals, glockenspiel, or pitch pipe.

The objective quality performances are shown in Table 1, where we compare the base line FDLP codec exploiting 64 band uniform QMF decomposition (QMF_{64}) at 136 kbps with the FDLP codec exploiting the proposed 32 band non-uniform QMF decomposition (QMF_{32}) at 82 kbps. Although, the objective scores of QMF_{32} are degraded by 0.2 compared to QMF_{64}, the bit-rate reduces significantly by around 40%. QMF_{32} is further compared to QMF_{64} operating at reduced bit-rates 88 kbps (bits for the sub-band carriers are uniformly reduced). In this case, the objective quality is reduced significantly. The block of quantization, described in [4], was not modified during these experiments.

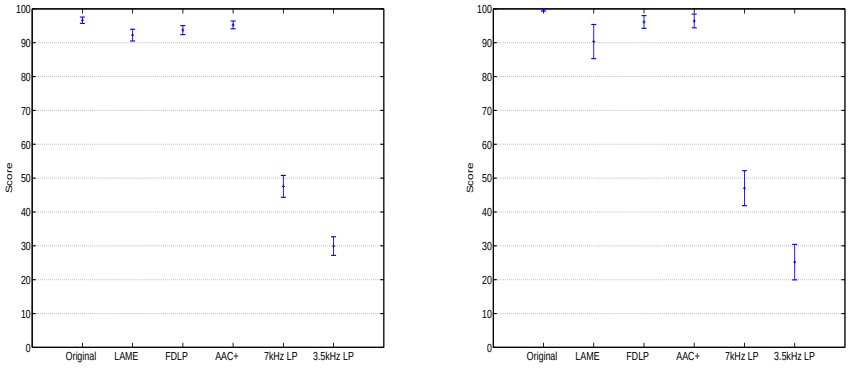
4 Subjective evaluations

For subjective listening tests, the FDLP codec (described in Section 3) exploiting non-uniform QMF decomposition (described in Section 2) is further extended with a perceptual model, a block performing dynamic phase quantization and a block of noise substitution.

The qualitative performance of the complete codec utilizing proposed non-uniform QMF decomposition is evaluated using MUSHRA (MUlti-Stimulus test with Hidden Reference and Anchor) listening tests [12] performed on 8 audio samples from MPEG audio exploration database [10]. The subjects' task is to evaluate the quality of the audio sample processed by each of the conditions involved in the test as well as the uncompressed condition and two or more degraded Anchor conditions (typically low-pass filtered at 3.5 kHz and 7 kHz).

bit-rate [kbps]	136	88	82
system	QMF_{64}	QMF_{64}	QMF_{32}
	Uniform	Uniform	Non-Uniform
ODG Scores	-1.04	-2.02	-1.23

Table 1. *Mean objective quality test results provided by PEAQ (ODG scores) over* 18 *audio recordings: the base line FDLP codec at* 136 *kbps, the base line codec operating at reduced bit-rates* 88 *kbps, and the codec exploiting proposed 32 band non-uniform QMF decomposition at* 82 *kbps.*



(a) 22 listeners. (b) 4 expert listeners.

Fig. 4. *MUSHRA results for 8 audio samples using three coded versions (FDLP, HE-AAC and LAME MP3), hidden reference (original) and two anchors (7 kHz and 3.5 kHz low-pass filtered).*

We compare the subjective quality of the following codecs:

- Complete version of the FDLP codec at ∼66 kbps.
- LAME - MP3 (MPEG 1, layer 3) at 64 kbps [13]. Lame codec based on MPEG-1 architecture is currently considered the best MP3 encoder at mid-high bit-rates and at variable bit-rates.
- MPEG-4 HE-AAC, v1 at ∼64 kbps [14]. The HE-AAC coder is the combination of Spectral Band Replication (SBR) [15] and Advanced Audio Coding (AAC) [16] and was standardized as High-Efficiency AAC (HE-AAC) in Extension 1 of MPEG-4 Audio [17].

The cumulative MUSHRA scores (mean values with 95% confidence) are shown in Figures 4(a) and (b). MUSHRA tests were performed independently in two different labs (with the same setup). Figure 4(a) shows mean scores for the results from both labs (combined scores for 18 non-expert listeners and 4 expert listeners), while Figure 4(b) shows mean scores for 4 expert listeners in one lab. These figures show that the FDLP codec performs better than LAME-MP3 and closely achieves subjective results of MPEG-4 HE-AAC standard.

5 Conclusions and Discussions

A technique for employing non-uniform frequency decomposition in the FDLP wide-band audio codec is presented here. The resulting QMF sub-bands closely follow the human auditory critical band decomposition. According to objective quality results, the new technique provides bit-rate reduction of about 40% over the base line, which is mainly due to transmitting few spectral components from

the higher bands without affecting the quality significantly. Subjective evaluations, performed on the extended version of the FDLP codec, suggest that the complete FDLP codec operating at ∼66 kbps provides better audio quality than LAME - MP3 codec at 64 kbps and gives competitive results compared to MPEG-4 HE-AAC standard at ∼64 kbps. FDLP codec does not make use of compression efficiency provided by entropy coding and simultaneous masking. These issues are open for future work.

References

1. P. Motlicek, H. Hermansky, H. Garudadri, N. Srinivasamurthy, "Speech Coding Based on Spectral Dynamics", Proceedings of TSD 2006, LNCS/LNAI series, Springer-Verlag, Berlin, pp. 471-478, September 2006.
2. Herre J., Johnston J. H., "Enhancing the performance of perceptual audio coders by using temporal noise shaping (TNS)", *in 101st Conv. Aud. Eng. Soc.*, 1996.
3. M. Athineos, D. Ellis, "Frequency-domain linear prediction for temporal features", *Automatic Speech Recognition and Understanding Workshop IEEE ASRU*, pp. 261-266, December 2003.
4. P. Motlicek, S. Ganapathy, H. Hermansky, and Harinath Garudadri, "Scalable Wide-band Audio Codec based on Frequency Domain Linear Prediction", *Tech. Rep., IDIAP*, RR 07-16, version 2, September 2007.
5. D. Pan, "A tutorial on mpeg audio compression", IEEE Multimedia Journal, vol. 02, no. 2, pp. 60-74, Summer 1995.
6. A. Charbonnier, J-B. Rault, "Design of nearly perfect non-uniform QMF filter banks", *in Proc. of ICASSP*, New York, NY, USA, April 1988.
7. P. Vaidyanathan, "Multirate Systems And Filter Banks", Prentice Hall Signal Processing Series, Englewood Cliffs, New Jersey 07632, 1993.
8. G. Theile, G. Stoll, M. Link, "Low-bit rate coding of high quality audio signals", *in Proc. 82nd Conv. Aud. Eng. Soc.*, Mar. 1987, preprint 2432.
9. T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerends, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg, B. Feiten, "PEAQ - The ITU Standard for Objective Measurement of Perceived Audio Quality", J. Audio Eng. Soc., vol. 48, pp. 3-29, 2000.
10. ISO/IEC JTC1/SC29/WG11: "Framework for Exploration of Speech and Audio Coding", MPEG2007/N9254, Lausanne, Switzerland, July 2007.
11. S. Ganapathy, P. Motlicek, H. Hermansky, H. Garudadri, "Temporal Masking for Bit-rate Reduction in Audio Codec Based on Frequency Domain Linear Prediction", *Tech. Rep., IDIAP*, RR 07-48, October 2007.
12. ITU-R Recommendation BS.1534: "Method for the subjective assessment of intermediate audio quality", June 2001.
13. LAME MP3 codec: *<http://lame.sourceforge.net>*
14. 3GPP TS 26.401: "Enhanced aacPlus General Audio Codec", General Description.
15. M. Dietz, L. Liljeryd, K. Kjorling, O. Kunz, "Spectral Band Replication, a novel approach in audio coding", in AES 112th Convention, Munich, DE, May 2002, Preprint 5553.
16. M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, Y. Oikawa, "ISO/IEC MPEG-2 Advanced Audio Coding", J. Audio Eng. Soc., vol. 45, no. 10, pp. 789814, October 1997.
17. ISO/IEC, "Coding of audio-visual objects Part 3: Audio, AMENDMENT 1: Bandwidth Extension", ISO/IEC Int. Std. 14496-3:2001/Amd.1:2003, 2003.