IDIAP RESEARCH REPORT

Spectral Noise Shaping: Improvements in Speech/Audio Codec Based on Linear Prediction in Spectral Domain

Sriram Ganapathy [a b]   Petr Motlicek [a]
Hynek Hermansky [a b]   Harinath Garudadri [c]

IDIAP–RR 08-16

June 2008

published in
Interspeech 2008

[a] IDIAP Research Institute, Martigny, Switzerland
[b] Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[c] Qualcomm Inc., San Diego, CA, USA

IDIAP Research Institute **www.idiap.ch**
Av. des Prés–Beudin 20 P.O. Box 592 1920 Martigny – Switzerland
Tel: +41 27 721 77 11 Fax: +41 27 721 77 12 Email: info@idiap.ch

IDIAP Research Report 08-16

Spectral Noise Shaping: Improvements in Speech/Audio Codec Based on Linear Prediction in Spectral Domain

Sriram Ganapathy Petr Motlicek Hynek Hermansky
Harinath Garudadri

June 2008

published in
Interspeech 2008

Abstract. Audio coding based on Frequency Domain Linear Prediction (FDLP) uses auto-regressive models to approximate Hilbert envelopes in frequency sub-bands. Although the basic technique achieves good coding efficiency, there is a need to improve the reconstructed signal quality for tonal signals with impulsive spectral content. For such signals, the quantization noise in the FDLP codec appears as frequency components not present in the input signal. In this paper, we propose a technique of Spectral Noise Shaping (SNS) for improving the quality of tonal signals by applying a Time Domain Linear Prediction (TDLP) filter prior to the FDLP processing. The inverse TDLP filter at the decoder shapes the quantization noise to reduce the artifacts. Application of the SNS technique to the FDLP codec improves the quality of the tonal signals without affecting the bit-rate. Performance evaluation is done with Perceptual Evaluation of Audio Quality (PEAQ) scores and with subjective listening tests.

1 Introduction

A new speech/audio coding technique based on modeling the temporal evolution of the spectral dynamics was proposed in [1, 2]. The approach is based on representing Amplitude Modulating (AM) signal using Hilbert envelope estimate and Frequency Modulating (FM) signal using Hilbert carrier. Speech/audio signals are analyzed in time using a non-uniform Quadrature Mirror Filter (QMF) bank to decompose the signal into frequency sub-bands. For each sub-band signal, Hilbert envelopes are estimated using Frequency Domain Linear Prediction (FDLP), which is an efficient technique for auto-regressive (AR) modelling of the temporal envelopes of a signal [3]. The parameters of the AR model are transmitted along with a few spectral components of the residual. At the decoder, these steps are inverted to reconstruct the signal back.

The FDLP codec achieves good compression efficiency for speech/audio signals. However, there is need to improve quality of the reconstructed signal for inputs with tonal components. The technique of FDLP fails to model these signals because of the impulsive spectral content. Hence, most of the important signal information is present in the FDLP residual. For such signals, the quantization error in the FDLP codec spreads across all the frequencies around the tone. This results in significant degradation in the reconstructed signal quality.

In conventional codecs such as [4], the dual problem arises in encoding transients in the time domain. This is efficiently solved by Temporal Noise Shaping (TNS) [5]. Specifically, coding artifacts arise mainly in handling transient signals (like the castanets) and pitched signals. Using spectral signal decomposition for quantization and encoding implies that a quantization error introduced in this domain will spread out in time after reconstruction by the synthesis filter bank. TNS represents one solution to overcome this problem by shaping the quantization noise in the time domain according to the input transient. This technique is widely used in modern audio codecs such as [6].

In this paper, we propose a technique of Spectral Noise Shaping (SNS) to overcome the problem of encoding tonal signals in FDLP based speech/audio codec. The technique is motivated by the fact that tonal signals are highly predictable in the time domain. If a sub-band signal is found to be tonal, it is analyzed using TDLP [7] and the residual of this operation is processed with the FDLP codec. At the decoder, the output of the FDLP codec is filtered by the inverse TDLP filter. Since the inverse TDLP filter follows the spectral impulses for tonal signals, it shapes the quantization noise according to the input signal. Application of the SNS technique to the FDLP codec improves the quality of the reconstruction for these signals without affecting the bit-rate.

The rest of the paper is organized as follows. Sec. 2 describes the FDLP technique for AR modelling of Hilbert Envelopes. The basic structure of the FDLP codec is described in Sec. 3. Sec. 4 explains the technique of SNS in detail. The objective and subjective evaluations are reported in Sec. 5.

2 Frequency Domain Linear Prediction

Typically, auto-regressive (AR) models have been used in speech applications for representing the envelope of the power spectrum of the signal by performing the operation of TDLP [7]. This paper utilizes AR models for obtaining smoothed, minimum phase, parametric models of temporal rather than spectral envelopes. The duality between the time and frequency domains means that AR modeling can be applied equally well to discrete spectral representations of the signal instead of time-domain signal samples. The block schematic showing the steps involved in deriving the AR model of Hilbert

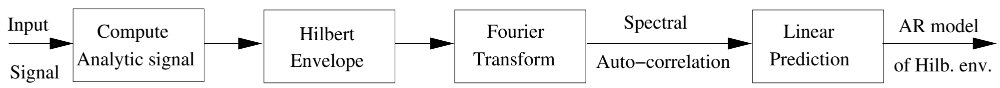

Figure 1: *Steps involved in FDLP technique for AR modelling of Hilbert Envelopes.*

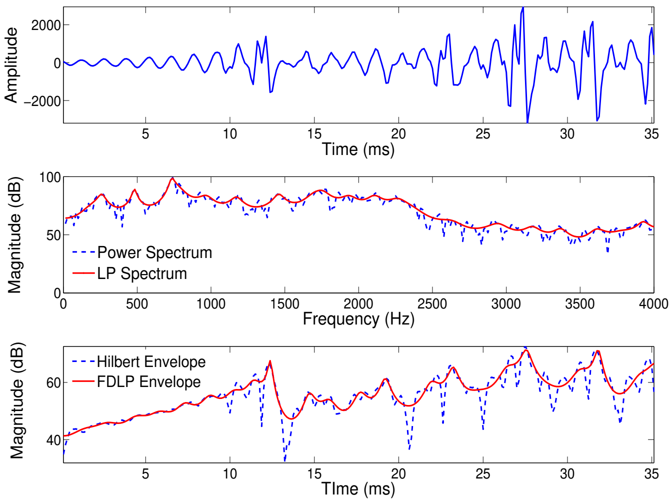

Figure 2: *Linear Prediction in time and frequency domains for a portion of speech signal: (a) input signal, (b) Power Spectrum and LP spectrum and (c) Hilbert Envelope and FDLP envelope.*

envelope is shown in Fig. 1. The first step is to compute the analytic signal for the input signal. For a discrete time signal, the analytic signal can be obtained using the Fourier Transform [8]. Hilbert envelope (squared magnitude of the analytic signal) and spectral auto-correlation function form Fourier transform pairs [5]. This relation is used to derive the auto-correlation of the spectral components of a signal which are then used for deriving the FDLP models (in manner similar to the computation of the TDLP models from temporal autocorrelations [7]).

For the FDLP technique, the squared magnitude response of the all-pole filter approximates the Hilbert envelope of the signal. This is in exact duality to the approximation of the power spectrum of the signal by the TDLP, as shown in Fig. 2.

3 Speech/Audio codec based on FDLP

Long temporal segments (1000 ms) of the input speech/audio signals are decomposed into 32 non-uniform QMF sub-bands which approximate the critical band decomposition in auditory system. 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 residual signals are processed in spectral domain with the magnitude spectral parameters quantized using VQ. Phase spectral components are scalar quantized (SQ) as they were found to have a uniform distribution. Graphical scheme of the FDLP encoder is given in Fig. 3.

In the decoder, shown in Fig. 4, quantized spectral components of the FDLP residual signals are reconstructed and transformed back to the time-domain using inverse Discrete Fourier Transform (DFT). The reconstructed FDLP envelopes (from LSF parameters) are used to modulate the corresponding sub-band residual signals. Finally, sub-band synthesis is applied to reconstruct the full-band signal.

4 Improvements in FDLP codec

For improving the reconstruction quality of tonal signals, we include the tonality detector and the SNS module to the FDLP codec.

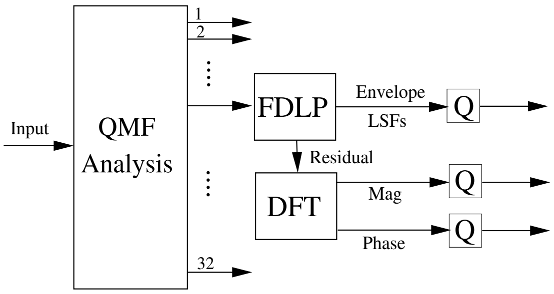

Figure 3: *Scheme of the FDLP encoder.*

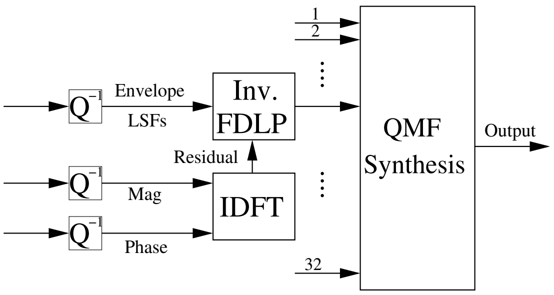

Figure 4: *Scheme of the FDLP decoder.*

4.1 Tonality Detector

The task of the tonality detector is to identify the QMF sub-band signals which have strong tonal components. Since the FDLP codec efficiently encodes non-tonal and partly tonal signals, highly tonal signals are alone processed using SNS. For this purpose, a global and a local tonality measure is computed and the tonality decision is taken based on both these measures. The global tonality measure is based on the Spectral Flatness Measure (SFM defined as the ratio of the geometric mean to the arithmetic mean of the spectral magnitudes) of the full-band signal. If the SFM is below a threshold, all the sub-bands for that input frame are checked for tonality locally. The local tonality measure is determined from the spectral auto-correlation of the sub-band signal (used for estimation of FDLP envelopes in Fig. 1). The ratio of the maximum spectral auto-correlation in higher lags (from lag 1 to the FDLP model order) to the zeroth lag of the spectral auto-correlation forms the local tonality measure. If the sub-band signal is highly tonal, its spectrum is impulsive and therefore, the spectral auto-correlation is impulsive as well. On the other hand, if the higher lags of spectral auto-correlation (within the FDLP model order) contain significant percentage of the energy (zeroth lag of spectral auto-correlation), the spectrum of the signal is predictable and the base-line FDLP codec (without the SNS) is able to model this signal structure.

4.2 Spectral Noise Shaping

As explained earlier, the tonal sub-band signals are applied to a TDLP filtering block. For the tonal signals, the TDLP and the FDLP model order are made equal to 20 as compared to a FDLP model order of 40 for the non-tonal signals. Hence, there is no increase in the bit-rate by the inclusion of the SNS. At the decoder, inverse TDLP filtering applied on the FDLP decoded signal gives the sub-band

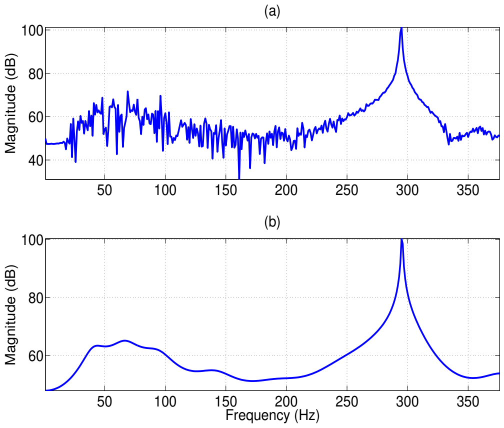

Figure 5: *Inverse TDLP filter used for spectral noise shaping: (a) Power Spectrum of tonal sub-band signal, and (b) Magnitude response of the inverse TDLP filter in SNS.*

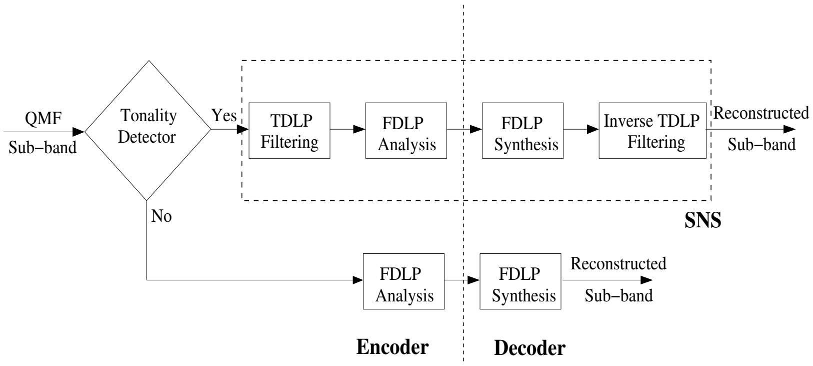

Figure 6: *FDLP codec with SNS.*

signal back.

The technique of SNS is motivated by the fundamental property of the linear prediction: For AR signals, the inverse TDLP filter has magnitude response characteristics similar to the Power Spectral Density (PSD) of the input signal [7]. As an example, Fig. 5 shows the power spectrum of a tonal sub-band signal and the frequency response of the inverse TDLP filter for this sub-band signal. Since the quantization noise passes through the inverse TDLP filter, it gets shaped in the frequency domain according to PSD of the input signal and hence the name, spectral noise shaping. Fig. 6 shows the block schematic of the FDLP codec with SNS. The additional side information involved is only the signalling of the tonality decision to the decoder (32bps).

5 Results

The subjective and objective evaluations of the proposed audio codec are performed using challenging audio signals (sampled at 48 kHz) present in the framework for exploration of speech and audio coding [9]. It is comprised of speech, music and speech over music recordings. For purpose of detailed evaluation, more tonal signals from [10] are also used.

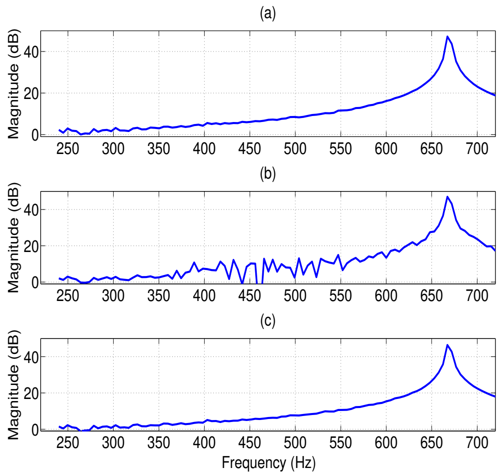

Figure 7: *Improvements in reconstruction signal quality with SNS: A portion of power spectrum of (a) a tonal input signal, (b) reconstructed signal using the base-line FDLP codec without SNS, and (c) reconstructed signal using the FDLP codec with SNS .*

Input	Base-line codec	With SNS
Flute 1	-0.49	-0.41
Flute 2	-1.86	-1.66
Violin	-0.43	-0.32
Organ	-2.82	-1.12
Alto-flute	-2.10	-2.04
Avg.	**-1.54**	**-1.10**

Table 1: *PEAQ scores for tonal files with and without SNS.*

bit-rate [kbps]	66	64	64
system	FDLP	LAME	AAC
Avg.	-1.11	-1.61	-0.77

Table 2: *Average objective quality test results provided by PEAQ for* 27 *files.*

5.1 Objective Evaluations

The objective measure employed is the Perceptual Evaluation of Audio Quality (PEAQ) distortion measure [11]. 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 - Objective Difference Grade (ODG) score. ODG is an impairment scale which indicates the measured audio quality of the signal under test on a continuous scale from -4 (very annoying impairment) to 0 (imperceptible impairment).

Table 1 shows the comparison of the PEAQ scores for some tonal files with and without SNS. The objective quality score (average PEAQ scores) is improved by the application of SNS (on the average by about 0.4), without affecting the bit-rate. The improvement provided by SNS for tonal signals is also illustrated in Fig. 7, where we show a portion of power spectrum of the (a) input signal, (b) the reconstructed signal using the base-line FDLP codec, and (c) reconstructed signal using the FDLP codec with SNS. This figure illustrates the ability of the proposed technique in reducing the artifacts present in tonal signals.

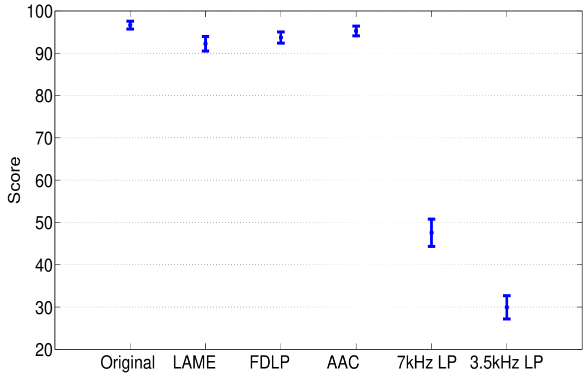

Figure 8: *MUSHRA results for 8 audio files with 22 listeners using three coded versions (FDLP, AAC and LAME), hidden reference (original) and two anchors (7 kHz low-pass filtered and 3.5 kHz low-pass filtered).*

For comparison with state-of-art codecs, the following 3 codecs are considered:

1. FDLP codec with SNS at ~ 66 kbps denoted as FDLP.

2. LAME MP3 (MPEG 1, layer 3) [4] at 64 kbps denoted as LAME.

3. High Efficiency Advanced Audio Coding (AAC+v1) with Spectral Band Replication (SBR) [6, 13] at ~ 64 kbps denoted as AAC.

For the 27 speech/audio files from [9], the results of objective quality evaluations are shown in Table 2.

5.2 Subjective Evaluations

MUSHRA (MUltiple Stimuli with Hidden Reference and Anchor) is a methodology for subjective evaluation of audio quality. It is defined by ITU-R recommendation BS.1534 [12]. We perform the MUSHRA tests on 8 audio samples from the database with 22 listeners. The results of the MUSHRA tests are shown in Figure 8. It is found that the proposed version of the codec, with SNS, is competitive with the state-of-art codecs at similar bit-rates.

6 Conclusions

We identify the problem of encoding tonal signals in codecs based on spectral dynamics in sub-bands. We propose the technique of spectral noise shaping to overcome this issue. The technique relies on the fact that tonal signals are temporally predictable and the residual of the prediction can be efficiently processed using the FDLP codec. Without increasing the bit-rate, the quantization noise at the receiver can be shaped in the frequency domain according to the spectral characteristics of the input signal. For some audio samples like Alto-flute and Trumpet, the current version of the SNS module does not significantly improve the performance as the inverse TDLP filter is unable to completely capture the signal dynamics. Further refinements of the SNS module for these signals, form part of the future work.

References

[1] P. Motlicek, H. Hermansky, S. Ganapathy and H. Garudadri, "Non-Uniform Speech/Audio Coding Exploiting Predictability of Temporal Evolution of Spectral Envelopes", *Proc. of TSD*, LNCS/LNAI series, Springer-Verlag, Berlin, pp. 350-357, September 2007.

[2] S. Ganapathy, P. Motlicek, H. Hermansky, and H. Garudadri, "Autoregressive Modelling of Hilbert Envelopes for Wide-band Audio Coding", *Audio Engineering Society*, 124th Convention, Amsterdam, Netherlands, May 2008.

[3] Marios Athineos and Dan Ellis, "Autoregressive Modeling of Temporal Envelopes", *IEEE Trans. on Signal Proc.*, Vol. 55, Issue 11, Nov. 2007 pp. 5237 - 5245.

[4] LAME MP3 codec: http://lame.sourceforge.net

[5] J. Herre and J.D Johnston, "Enhancing the Performance of Perceptual Audio Coders by using Temporal Noise Shaping (TNS)", *Audio Engineering Society*, 101st Convention, Los. Angeles, USA, November 1996.

[6] "3GPP TS 26.401: Enhanced aacPlus general audio codec; General Description", 2004.

[7] J. Makhoul, "Linear Prediction: A Tutorial Review", in *Proc. of the IEEE*, Vol 63(4), pp. 561-580, 1975.

[8] L.S. Marple, "Computing the Discrete-Time Analytic Signal via FFT", *IEEE Trans. on Acoustics, Speech and Signal Proc.*, Vol. 47, pp. 2600-2603, 1999.

[9] "ISO/IEC JTC1/SC29/WG11: Framework for Exploration of Speech and Audio Coding", *MPEG2007/N9254*, July 2007, Lausanne, Switzerland.

[10] "Musical Instrumental Samples", http://theremin.music.uiowa.edu/MIS.html.

[11] "ITU-R Recommendation BS.1387: Method for objective psychoacoustic model based on PEAQ to perceptual audio measurements of perceived audio quality", December 1998.

[12] "ITU-R Recommendation BS.1534: Method for the subjective assessment of intermediate audio quality", June 2001.

[13] Martin Dietz, Lars Liljeryd, Kristofer Kjorling and Oliver Kunz, "Spectral Band Replication, a novel approach in audio coding", *Audio Engineering Society*, 112th Convention, Munich, Germany, May 2002.