Audio Engineering Society

Convention Paper

Presented at the 118th Convention
2005 May 28–31 Barcelona, Spain

A Fractal Self-Similarity Model for the Spectral Representation of Audio Signals

Deepen Sinha[1], Anibal J. S. Ferreira[1,2], and Deep Sen[1,3]

[1] *ATC Labs, New Jersey, USA*

[2] *University of Porto, Portugal*

[3] *University of New South Wales, Sydney, Australia*

Correspondence should be addressed to D. Sinha (sinha@atc-labs.com)

ABSTRACT

In the application of conventional audio compression algorithms to low bit rate audio coding one is faced with the unsatisfactory tradeoff between coarser quantization and audio bandwidth reduction. Frequency Extension has therefore emerged as an important tool for the satisfactory performance of low bit rate audio codecs. In this paper we describe one of a newer class of Frequency Extension techniques which are applied directly to the high frequency resolution representation of the signal (e.g., MDCT). This particular technique is based on a Fractal Self-Similarity Model (*FSSM*) for the short-term frequency representation of the signal. The *FSSM* model, which may include multiple dilation and translation terms, has been found to be effective for a wide variety of speech and music signals and provides a compact description for long term correlation that may exist in frequency domain. The high frequency resolution of MDCT aids in accurate parameter estimation for the model, which in turn has shown promise as a Frequency Extension tool that offers a detailed and natural sounding quality at low bit rates. Structure of the *FSSM* model, issues related to parameter estimation, and its application to audio coding for bit rates of 8-48 kbps is discussed. Audio demos are available at http://www.atc-labs.com/fssm.

1. INTRODUCTION

Audio coding at low bit rates has many established and emerging applications. These include Satellite and Terrestrial Digital Audio Broadcasting (DAB)[1], Internet music download and streaming, solid-state audio playback devices, high fidelity audio communication on

[1] Support for this work was provided by Sirius Satellite Radio (see Acknowledgements for details).

the cellular telephone network, etc. In many of these applications the demand for higher compression efficiency continues to grow. This is the case despite of availability of cheaper and higher capacity storage devices, and the availability of more efficient modulation schemes for transmission. In fact there appears to be a proliferation of applications demanding CD quality stereo at bit rates of 48 kbps and lower and high quality FM grade mono audio at bit rates of 20-24 kbps. These in turn continue to spur the demand for newer tools for audio bit rate reduction.

The field of Perceptual Audio Coding has matured over last several years and a number of audio coding technologies exist. These include proprietary schemes such as PAC (Bell Labs, Lucent) [1] and ATRAC (Sony) [2] as well as standard based codecs such as MPEG-1 Layer 3 (popularly known as MP3) [3], MPEG-2 AAC [4], Dolby AC-3 [5]. In general the established audio coding schemes fit into the framework of adaptive sub-band coding whereby the output of a filterbank is quantized using quantizers driven by a perceptual model. Although a few different options for the filterbank have been employed, the Modified Discrete Transformation (MDCT) [6] is especially popular. In particular the use of a high resolution MDCT (i.e., with 1024 frequency subbands) has led to higher coding efficiency in algorithms like PAC and AAC. The best of these techniques are capable of producing full fidelity CD quality audio (20 kHz audio bandwidth high stereo separation) in the range of 96-128kbps. Furthermore, near-CD quality audio with somewhat lower audio bandwidth (~ 15 kHz) and limited stereo is achievable in the range of 48-64 kbps.

In order to reduce the bit rate requirement further (to enable newer applications as noted above), several parametric approaches have been proposed. These rely on a compact parametric description of all or a portion of the audio signal. One such approach that has proven to be particularly effective is the so called “Bandwidth Extension” approach. In Bandwidth Extension only a low pass filtered version of the signal is directly coded using the conventional perceptual coding paradigm. The high frequency portion of the signal spectrum is recreated at the decoder by a mapping generated from the low frequency spectrum of the signal. Typically an attempt is made to match the reconstructed high frequency spectrum to the original high frequency spectrum as closely as possible. In practice significant mismatch may remain between the two. However, the philosophy is that increased naturalness of the higher audio bandwidth signal compensates for any other perceived distortion in the (reconstructed) higher frequencies.

The Spectral Band Replication (SBR) technique [7] is perhaps the best known of the bandwidth extension tools used in audio coding. Other attempts at bandwidth extension, some of which came out of the speech coding world include, e.g., [8]. In the SBR approach, the bandwidth extension for an audio signal is achieved by controlled replication of the base band signal towards higher frequencies. This replication is performed using a complex QMF filterbank that is different from the primary coding filterbank (e.g., MDCT).

In this paper we introduce a Fractal Self-Similarity Model (*FSSM*) for the representation and reconstruction of the higher frequencies. As will be seen below, this model serves as an effective low bit rate tool for bandwidth extension. The novel aspects of the proposed model are as follows

- The self-similarity model provides a broad description of the intra-spectral correlation structures that may exist in the signal. The model, which includes replication as a special case, therefore allows for a better approximation to the original spectrum for a wider class of signals.

- The model is applied directly into the MDCT domain. The higher frequency resolution of MDCT helps in two ways: (i) it allows for a better reproduction of the harmonic structure in the signal (i.e., the harmonic frequencies in the reconstructed signal match the frequencies of original signal closely. This is an important consideration [10]); (ii) it allows for a better estimation of the model parameters.

The organization of the rest of the paper is as follows. In section 2 we introduce the *FSSM* model and the underlying Bandwidth Extension mechanism. In section 3 we look into the significance of various parameters in the model as well as techniques for estimating those parameters. In section 4 we look into issues related to the temporal envelope of the re-created high frequency components. In section 5 we summarize the full audio codec structure utilizing the proposed model. The performance of the model in audio coding is described in section 6, followed by conclusions in section 7, and acknowledgements in section 8.

2. FRACTAL SELF SIMILARITY MODEL (*FSSM*) FOR HIGH FREQUENCY REPRESENTATION

In this section we begin with the formulation of the Bandwidth Extension problem and then introduce the *FSSM* model, Bandwidth Extension procedure, and, codec considerations.

2.1. Reconstruction in the MDCT Domain

The bandwidth extension problem being addressed here assumes an element of *Backward Compatibility* with the structure of conventional MDCT codecs. In other words:

- It is assumed that the MDCT representation up to certain frequency f_c, denoted as $X_{LP}(f)$, is coded directly using standard quantization and coding techniques.

- The MDCT spectrum for frequencies $f > f_c$ is to be reconstructed using a mapping BE such that

  $$\overline{X}_{HP}(f) = BE(\overline{X}_{LP}(f)) \qquad (1\ )$$

  Where, $\overline{X}_{LP}$ is the quantized baseband and $\overline{X}_{HP}$ is the reconstructed higher frequencies in MDCT domain.

2.2. *FSSM* Model

We model the high frequency components of the signal as being reconstructed using an iterative sequence of *Expansion Operators* (EO). In other words,

$$\overline{X}_{HP}(f) = \cdots EO_i \circ (\cdots (EO_1 \circ (EO_0 \circ \overline{X}_{LP}(f))\cdots) \qquad (2\ )$$

Where each expansion operator EO_i is assumed to have the form

$$EO_i \circ \overline{X}_{LP}(f) = H_i \bullet X_{LP}(\alpha_i f - f_i) \qquad (3\ )$$

where α_i is a dilation parameter ($\alpha_i \le 1$) and f_i is a frequency translational parameter. H_i is a high pass (brick-wall) filter with a cutoff frequency $f_c{}^i = \alpha_i * f_c{}^{(i-1)} + f_i$, with $f_c{}^0 = f_c$. This sequence of nested expansion operators resulting in bandwidth expansion is graphically illustrated in Figure 1. We will see below that the *FSSM* model is able to reconstruct the high frequency spectral details with a high level of accuracy across a wide range of different audio signals.

2.3. Codec Considerations

The application of *FSSM* to coding of audio signals entails the following additional considerations:

- Estimation and application of the dilation and translation parameter. This is discussed in more detail in Section 3.

- Determining the fit of the model reconstructed signal to the original signal. For this frequency spectrum may be split into multiple slices and for each slice a determination can made to either apply the model or replace it by an independent signal such as synthetic noise. The *FSSM* model therefore, in general, is a *FSSM+Noise* model.

- The shape of the temporal and frequency envelope of the signal is an important consideration. The *FSSM* model, as proposed above, does not accurately reconstruct the coarse frequency envelope which needs to be coded separately. The temporal shape of the *FSSM* generated component has interesting characteristics as seen in Section 4 and may need further shaping under many situations.

2.4. Fractal Nature of the *FSSM* Model

It may be noted that the set of operators $\{EO_i\}$ denotes a set of "contractive mappings" [11] on a two-dimensional metric space $\{f, X(f)\}$. In the Bandwidth extension application, the set defined mapping cover only a sub-space of the corresponding *Hausdorff Space* $\widehat{\mathrm{H}}(\{f, X(f)\})$[11]. This is because the mappings have been defined only for frequency regions $f \ge f_c$. Extension and definition of similar mappings to baseband leads to a complete *Iterated Function System* (*IFS*) for which the original audio spectrum may be

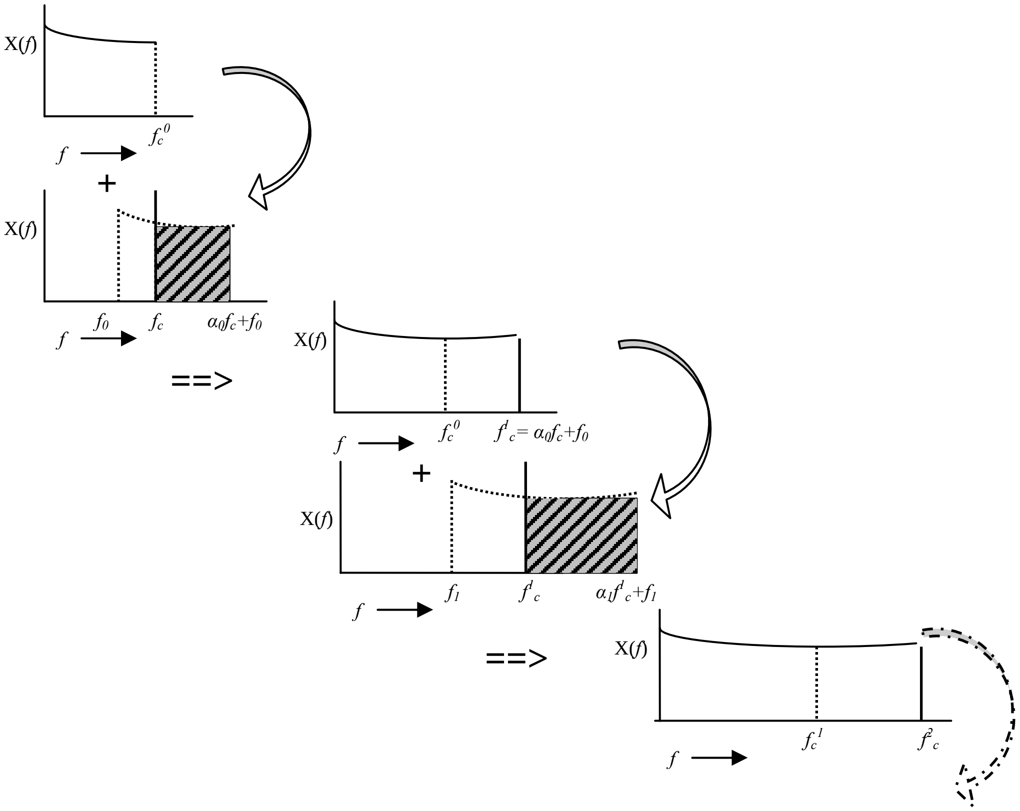

Figure 1: Bandwidth expansion using *FSSM*. Iterative application of the *Expansion Operators* (dilations and translations)

viewed as the corresponding "fixed point". The model therefore has application to the coding of the full signal spectrum (including the baseband). Such extension is beyond the scope of this presentation. In the context of bandwidth extension, the modeling and decoding problem is reduced to the case where subset of the "fixed point" (i.e., the baseband of the signal) is available during the decoding stage, resulting in quick convergence of the iterative decoding. However, the fractal nature has implications for the estimation of model parameters suggesting that the parameters $\{\alpha_i, f_i\}$ should be estimated in a way such that the nested operators in equation (2), when applied to the original spectrum will leave the spectrum (or corresponding frequency slice) approximately invariant.

3. SIGNIFICANCE AND ESTIMATION OF MODEL PARAMETERS

In this section we discuss the importance of various parameters in the operators $\{EO_i\}$ and the means for estimating these parameters. It may be noted that in the absence of the dilation parameter α_i and translation parameter f_i the model is equivalent to the spectrum replication model as in [7], albeit with the difference that this operation is being performed in the MDCT domain. Inclusion of these parameters in the model offers significant advantages as will be seen shortly.

3.1. Significance of the Translation and the Dilation Parameter

As alluded to earlier, the translation parameter helps in aligning pitch structure of the reconstructed signal with the one in the original signal (if the signal indeed has a well defined pitch structure). This has implication for two classes of audio signals. The first of this class consists of musical instruments with a pitch structure and the second class includes voiced speech and vocal signals. For these classes of signals the lack of dilation terms results in a discontinuity in the pitch structure at each boundary point $\left\{f_c^i\right\}$. This is illustrated in Figure 2a where the spectrum of a reconstructed pitchpipe signal is shown superimposed on the original spectrum.

The absence of translation term can lead to a certain audible ringing in the reconstructed signal. This cause of this ringing is understood easily by considering the original spectrum as a superposition of a low frequency and a high frequency slice and modeling the pitch pattern distortion as exhibited in Figure 2a as a frequency translation of the high frequency slice which is equivalent to multiplication by a sinusoid in time domain due to the Fourier duality [17].

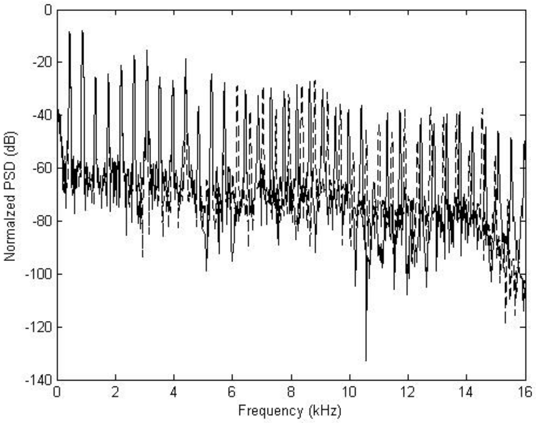

Figure 2a: Reconstructed signal spectrum (solid line) and original spectrum (dashed line) when no translation term is included in the *FSSM* model.

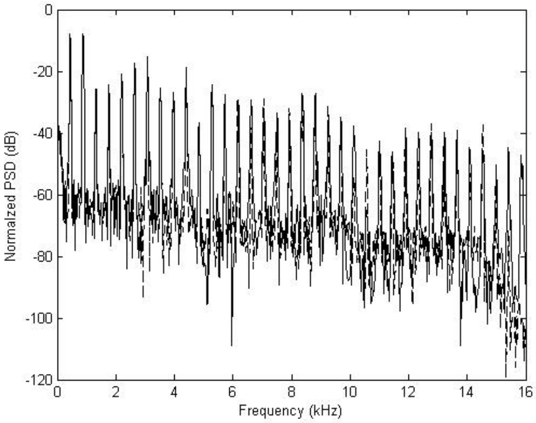

Figure 2b: Reconstructed signal spectrum (solid line) and original spectrum (dashed line) when a translation term is included in the *FSSM* model.

$$X_{HF}\left(f-f_d\right) \leftrightarrow \cos\left(2\pi f_d t\right) \cdot x_{HF}\left(t\right) \qquad (4)$$

where f_d represent the amount of pitch discontinuity. The time domain modulation by the cosine is often audible as ringing or beating of high frequency reconstruction depending upon the value of the distortion frequency and the nature of $x_{HF}\left(t\right)$. The presence of the translation term in conjunction with the high frequency resolution of MDCT allows for an accurate pitch alignment.

The inclusion of dilation parameter on the other hand leads to accurate signal spectrum reconstruction for a different class of audio signals, in particular for cases when the pitch structure is either not present in (part of) high frequencies or is more diffuse towards the higher frequencies. Example of a signal ("Aria") that benefits from the inclusion of the dilation terms in *FSSM* is shown in Figure 3a. The corresponding reconstruction is shown in Figure 3b, where the reconstruction with dilation (dashed line) has more of the features in high frequency signal spectrum.

.

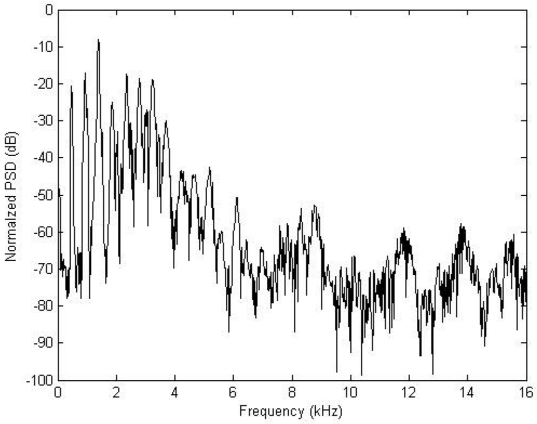

Figure 3a: Example of a signal (short-term PSD) that benefits from the inclusion of the dilation term in the *FSSM* model.

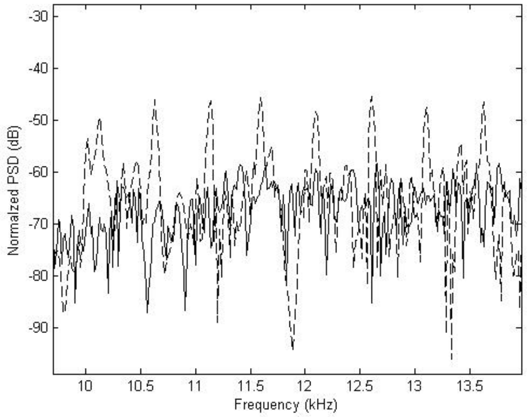

Figure 3b: Signal reconstruction with (dashed line) and without (solid line) the dilatation term for the example shown in Figure 3a.

3.2. Estimation of the Dilation and Translation Parameters

Good estimate of the dilation and translation terms may be found by searching for the maxima of the self-similarity coherence (SSC) function for the MDCT spectrum as defined below:

$$\Phi(\alpha_i, f_i) = \left\langle X\ (f) \cdot X(\alpha_i f - f_i) \right\rangle \qquad (5\ )$$

The optimal values for α_i and f_i denoted as $\bar{\alpha}$ and $\bar{f}$ are then found by maximizing the above function; i.e.,

$$\Phi(\bar{\alpha}_i, \bar{f}_i) = \max \Phi(\alpha_i, f_i), \ \ \forall \alpha_i \in A, f_i \in F \qquad (6)$$

Where A is a set of possible values for α_i and F is the set of possible values for the translation frequency. For the model to be meaningful for bandwidth extension the range of A and F should be restricted such that $\alpha_i f_c + f_i > f_c + C\ , \quad \forall \alpha_i \in A \ \& \ f_i \in F$ for some suitably chosen minimum extension band C.

The self similarity coherence maximization criterion as indicated above works well in many cases. However, in certain instances special considerations need to be taken into account as discussed below.

- In signals containing prominent harmonic structures the SSC maximization criterion is not the best suited from a perceptual point of view. For such signals the presence of a harmonic structure as well as the fundamental frequency of the dominant harmonic can be accurately estimated using the techniques developed in [12] [13]. In most such cases the translation parameter is best chosen as a value that ensures the continuity of the harmonic structure and the best value for the dilation parameter is often unity.

- Because of the nature of the MDCT filterbank fluctuation in translation parameter f_0 from one MDCT frame to next can cause aliasing distortion, an “unsteady” perception for the high frequency harmonics may result. This is particularly true for signals for which a strong and steady harmonic structure is present. Therefore, some smoothing or locking mechanism is necessary to avoid this problem.

- The quality of the estimates improves significantly if the MDCT spectrum is normalized by the coarse envelope prior to the estimation of these parameters.

4. TEMPORAL ENVELOPE SHAPING CONSIDERATIONS FOR THE *FSSM* COMPONENTS

As noted above the *FSSM* model is able to reproduce the detailed spectral fine structure of the high frequency components, and consequently is able to match faithfully the original spectrum. However the spectral envelope either is frequency or time is not produced by the *FSSM*. The spectral envelope in frequency (MDCT) domain may be coded and transmitted separately using conventional envelope coding techniques. The shape of the envelope in time is an important consideration and will be examined in this section.

4.1. Temporal Envelope of the *FSSM* components

Understanding the shape of the temporal envelope of the *FSSM* components is important in deciding the type of further processing that may be necessary to shape the envelope. Towards this end the following interesting observation may be made.

Observation[2]: The temporal envelope of an *FSSM* reconstructed signal using MDCT domain expansion operators exhibits a high correlation with the temporal envelope of baseband signal irrespective of the value of α_i 's and f_i 's.

For temporal envelope in this discussion the conventional definition in terms of the magnitude of the complex Hilbert Transform of signal is assumed [14]. The temporal envelope is examined at a time resolution that is several times higher than the time resolution of the MDCT. This is a somewhat surprising result given the coarseness of the MDCT time resolution. This observation is illustrated with the help of a synthetic narrowband noise signal in Figure 4. The figure shows the baseband signal (Figure 4a), the *FSSM* constructed high frequency signal (Figure 4b) and the Hilbert envelopes of the two signals superimposed on each other (Figure 4c).

The above behavior of *FSSM* constructed signal envelope is useful in most cases since high correlation between the envelopes of higher frequencies and baseband is generally desirable for a “quieter” perception. The same observation, however, cannot be made for the added “noise” components (i.e., when the model does not fit the actual spectral slice), which may have very little correlation with baseband envelope particularly for larger MDCT transform sizes. Further shaping of temporal envelope is therefore needed in two cases: (i) when noise is added, if the temporal envelope of the corresponding original signal spectral slice has a significantly different shape; and, (ii) in the case of *FSSM* reconstruction when the temporal envelope of *FSSM* slices is significantly different from the baseband envelope.

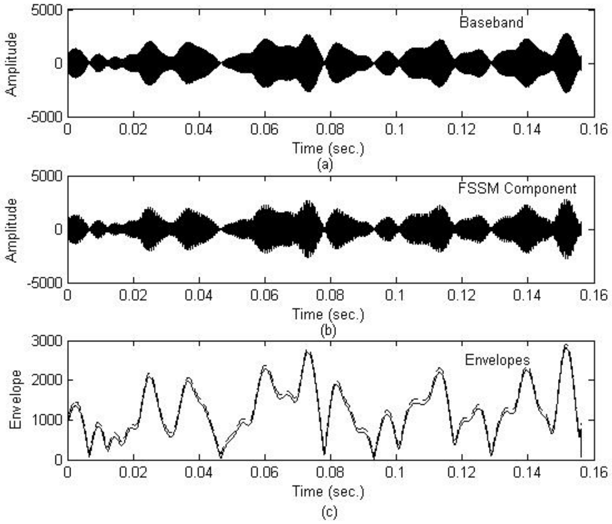

Figure 4: (a) Baseband noise signal, (b) *FSSM* constructed high frequency signal, (c) Envelopes of (a) and (b).

4.2. Generation of Temporal Envelopes: Filterbank Options

To perform the task of shaping the temporal envelope of the reconstructed higher frequency components (in those cases when it is needed) we need to examine time trajectories of the spectral energy in multiple frequency bands. Furthermore, these time trajectories need to be examined at a time resolution that is substantially higher than that afforded by the high frequency resolution MDCT filterbank. For accurate temporal shaping for voiced speech and dynamic musical instruments a time resolution of 4-5 msec (or lower) is desirable. The desired temporal shaping can be computed by utilizing a separate higher time resolution “Utility Filter Bank” (UFB). It is desirable for the UFB to be a complex, over-sampled modulated filterbank because of several desirable characteristics of such filterbanks such as very low aliasing distortion [16]. The magnitude of the complex output of the filterbank provides an estimate of the instantaneous spectral magnitude in the corresponding frequency band. Since UFB is not the primary coding filterbank its output may be suitably oversampled at the desired time resolution. Several options exist for the choice of the UFB. These include

[2] This observation can be converted into a theorem under some fairly general conditions. However, the mathematical formulation and the resulting proof is unnecessary for the sake of this presentation.

- Discrete Fourier Transform (DFT) with a higher time resolution (compared to MDCT): A DFT with 64-256 size *power complementary* window may be used in a sequence of overlapping blocks (with a 50% overlap between 2 consecutive windows)

- A complex modulated filterbank with sub-band filters of the form

$$h_i = h_0 \cdot e^{j\frac{2\pi}{N}\cdot(i-1)\cdot n} \qquad (6\ )$$

  where h_0 is a suitably optimized prototype filter. The DFT is a sub-class of this type of filterbanks. The more general framework allows for selection of longer windows (compared to the down-sampling factor).

- A complex non-uniform filterbank; e.g., one with two or more uniform sections and transition filters to link the 2 adjacent uniform sections [16]. The frequency response of one such filterbank with 2 uniform sections is shown in Figure 5.

The exact choice of the UFB may be application dependent. The complex-modulated filterbanks with a higher over-sampling ratio offer superior performance when compared to the DFT but at a cost of higher computational complexity. The non-uniform filterbank with higher frequency resolution at lower frequencies is useful if envelope shaping at very low frequencies (1.2 kHz and lower) is desirable.

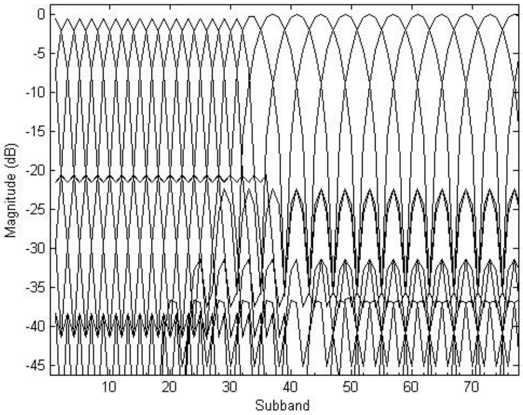

Figure 5: Frequency response of a non-uniform filterbank with 2 uniform sections

4.3. Coding of Temporal Shape Information

Multi-band temporal envelope information to perform the temporal shaping as discussed in Section 4.1 is computed by analyzing the output of the UFB and transmitting a suitable representation as side information. The overhead for this information can be reduced by utilizing the temporal shape that may already exist (as described in Section 4.2) and by grouping the information in adjacent time and frequency bands.

5. CODEC STRUCTURE UTILIZING THE *FSSM* MODEL

The structure of a complete codec utilizing the *FSSM + NOISE* model is illustrated in Figure 6a (Encoder) and Figure 6b (Decoder). At the encoder *FSSM* parameters are extracted from the MDCT representation of the signal. The input is also fed to the UFB in parallel and the output analyzed for the temporal envelope of the audio signal in multiple frequency bands. The output and decisions from the *FSSM* block are used in deciding which parts of the temporal envelope are to be transmitted to the decoder. The *FSSM* model parameter and additional temporal coding information is multiplexed and forms the side information for the band-width extension scheme. In parallel, the low pass filtered portion of MDCT is quantized and coded using conventional perceptual coding.

The total cost of transmitting the *FSSM* model and temporal envelope information depends on a number of factors such as, frequency limit from which bandwidth extension is applied as well as the desired accuracy of the temporal envelope information. Based on these design decisions the net overhead can range anywhere from 2-4 kbps per audio channel.

The decoder with *FSSM* bandwidth extension performs the conventional decoding functions such as Huffman decoding, inverse quantization, and, inverse MDCT. However, prior to the inverse MDCT, *FSSM* model is iteratively applied to the MDCT coefficient to extend the audio bandwidth to the specified limit. The extended bandwidth MDCT is inverse transformed and fed to the UFB. The output is UFB is further shaped in the time domain using the temporal shaping information available to the decoder. The final shaped output is inverse filtered through the UFB synthesis filters to generate the final audio output.

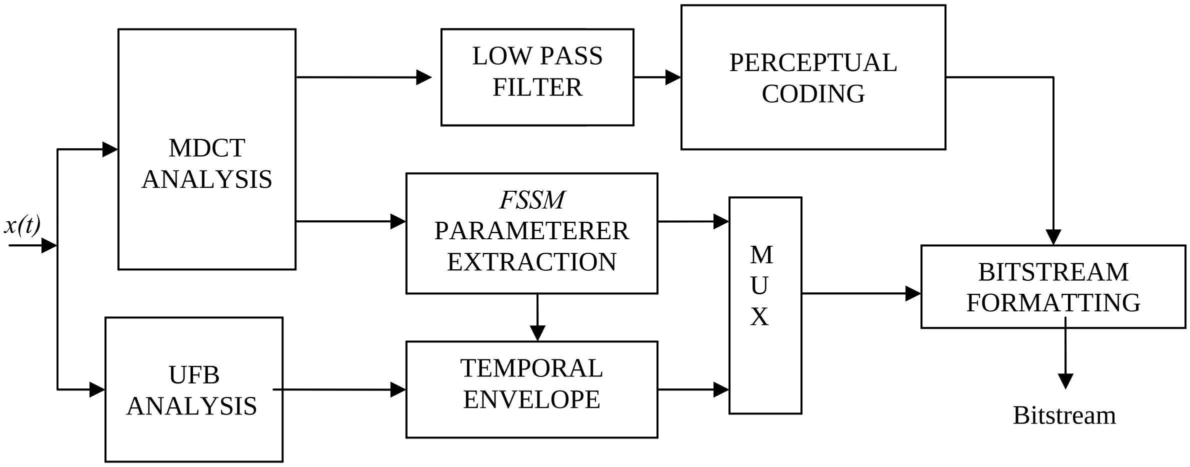

Figure 6a: Encoder Architecture with *FSSM* Bandwidth Extension

6. PERFORMANCE EVALUATION

A complete perceptual codec incorporating the *FSSM* bandwidth extension model was built and evaluated for audio quality at multiple bit rates. In the first part of this evaluation we attempt to subjectively quantify the audio quality gain obtained by utilizing the *FSSM* bandwidth extension. We include here two sets of subjective test results using the ITU MUSHRA subjective testing methodology [18]. The tests were conducted using 10 listeners and critical audio samples categorized in different genres. In the first test we compared the performance of a generic MDCT based “Perceptual Coder A” (PCA) at stereo bit rates. We compared PCA operating at 48 kbps and 64 kbps with PCA at 48 kbps but operating with the benefit of *FSSM* bandwidth extension. The results are summarized in Table 1. It is clear that there is a substantial improvement in audio quality due to the use of *FSSM* bandwidth extension. The quality in the codec with *FSSM* at 48kbps is often higher than the quality PCA (without *FSSM*) operating at 64kbps.

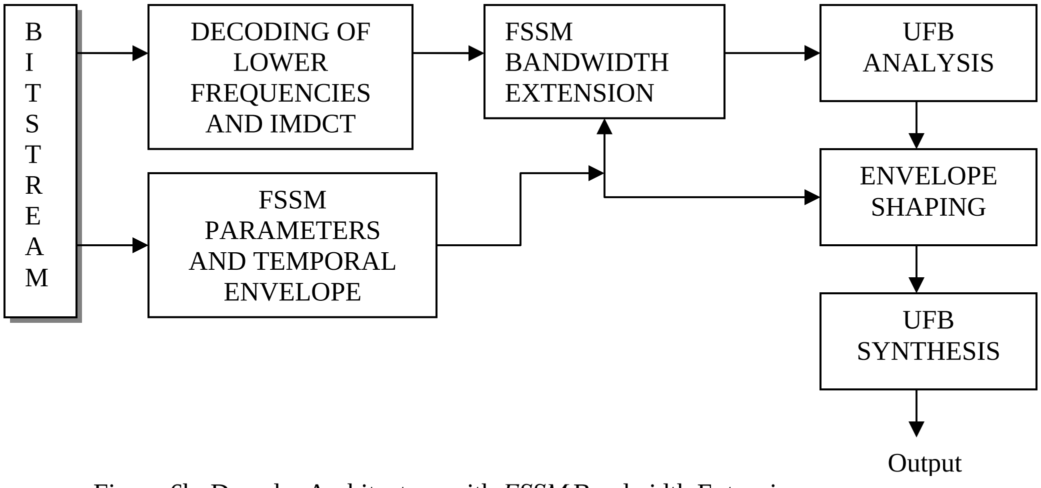

Figure 6b: Decoder Architecture with *FSSM* Bandwidth Extension

Table 1: MUSHRA SCORES (in %) FOR 3 STEREO CODECS

Music Genre	PCA with *FSSM* @ 48kbps	PCA @ 48kbps	PCA @ 64kbps
Oldies	69	51	73
Pop	75	52	68
Classical	71	58	69
Rock	83	65	77
Vocal	79	60	71

In the second test we compared the performance of PCA at 24 kbps with that of PCA at 24 kbps but with the benefit of *FSSM* bandwidth extension and also with PCA at 32 kbps. Once again there is substantial improvement due to the use of *FSSM* bandwidth extension.

Table 2: MUSHARA Scores (in %) for 3 Mono Codecs

Music Genre	PCA with *FSSM* @ 24kbps	PCA @ 24kbps	PCA @ 32kbps
News-Female	65	28	52
News-Male	60	34	60
Voiceover	63	26	55

The performance of the proposed bandwidth extension schemes in comparison to other bandwidth extension schemes is in progress. For evaluation purposes several audio clips coded at 48 kbps with a perceptual coder utilizing *FSSM* are available at

http://www.atc-labs.com/fssm

For reference and benchmarking purpose files encoded with MP3Pro at 64 kbps using the encoder/player available at http://www.mp3prozone.com are also included at the same location. These clearly illustrate the promise of the proposed scheme.

7. CONCLUSIONS

The proposed *FSSM* model and bandwidth extension methodology offers a promising new approach to the problem of reconstructing higher frequencies from a low pass filtered baseband. By working in the MDCT domain the technique is able to achieve a higher accuracy (compared to the original) bandwidth extension, resulting in a more natural and higher fidelity audio reproduction. Extension of the technique as a coding tool for the low frequency components as well is a possibility and will be investigated in future.

8. ACKNOWLEDGEMENTS

The authors will like to gratefully acknowledge the role of Sirius Satellite Radio in making this work a possibility. Several members of the Sirius codec team made contribution towards the validation, testing, and, refinement of the ideas presented in this paper. In particular, the contributions of Paula Muller, Lin Lin, Bhadresha Dedhia, Alex Cabanilla, and Jim Tracey are acknowledged. The support and feedback from Mr. Mark Kalman, and Mr. Ray Lowe is also gratefully acknowledged.

9. REFERENCES

[1] J. D. Johnston, D. Sinha, S. Dorward, and S. R Quackenbush, "AT&T Perceptual Audio Coding (PAC)," *in AES Collected Papers on Digital Audio Bit-Rate Reduction*, N. Gilchrist and C. Grewin, Eds. 1996, pp. 73-82.

[2] Kyoya Tsutui, Hiroshi Suzuki, Mito Sonohara Osamu Shimyoshi, Kenzo Akagiri, and Robert M.Heddle, "ATRAC: Adaptive Transform Acoustic Coding for MiniDisc," *93rd Convention of the Audio Engineering Society*, October 1992, Preprint n. 3456.

[3] K. Bradenburg, G. Stoll, et al. "The ISO- MPEG-Audio Codec: A Generic-Standard for Coding of High Quality Digital Audio," in *92nd AES Convention*, 1992, Preprint no. 3336.

[4] Marina Bosi et al., "ISO/IEC MPEG-2 Advanced Audio Coding," *101st Convention of the Audio Engineering Society*, November 1996, Preprint n. 4382.

[5] Mark Davis, “The AC-3 Multichannel Coder,” *95th Convention of the Audio Engineering Society*, October 1993, Preprint n. 3774.

[6] J. P. Princen, A. W. Johnson, and A. B. Bradley, “Subband/Transform Coding Using Filter Bank Designs Based on Time Domain Alias Cancellation," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1987, pp. 2161-2164.

[7] M Dietz, L. Liljeryd, K. Kjorling, and O. Kunz, “Spectral Band Replication, a novel approach in audio coding,” *112th Convention of the Audio Engineering Society*, May 2002, Paper 5553.

[8] P. Jax and P. Vary, “On Artificial Bandwidth Extension of Telephone Speech,” *Signal Processing*, Vol. 83, pp 1707-1719, August 2003.

[9] Nikil Jayant, James Johnston, and Robert Safranek, “Signal Compression Based on Models of Human Perception,” *Proceedings of the IEEE*, vol. 81, no. 10, pp. 1385-1422, October 1993.

[10] Anibal J. S. Ferreira, *Spectral Coding and Post-Processing of High Quality Audio*, Ph.D. thesis, Faculdade de Engenharia da Universidade do Porto-Portugal, 1998, http://telecom.inescn.pt/doc/phd_en.html.

[11] Michael Barnsley, “Fractals Everywhere”, *Academic Press, Inc*., ISBN 0-12-079062-9, 1988.

[12] Anibal J. S. Ferreira, “Combined Spectral Envelope Normalization and Subtraction of Sinusoidal Components in the ODFT and MDCT Frequency Domains,” in 2001 *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, October 21-24 2001, pp. 51-54.

[13] Anibal J. S. Ferreira, “Accurate Estimation in the ODFT Domain of the Frequency, Phase and Magnitude of Stationary Sinusoids," in 2001 *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, October 21-24 2001, pp. 47-50.

[14] John G. Proakis, *Digital Communications*, 4th ed, McGraw-Hill, 2001.

[15] Anibal J. S. Ferreira, “Perceptual Coding Using Sinusoidal Modeling in the MDCT Domain," *112th Convention of the Audio Engineering Society*, May 2002, Paper 5569.

[16] Z. Cvetkovic and J. D. Johnston, “Nonuniform Oversampled Filter Banks for Audio Signal Processing,” *IEEE Transactions on Speech and Audio Processing*, Vol. 11, No. 5, September 2003.

[17] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*, Prentice-Hall, 1975.

[18] ITU-R Recommendation BS.1534, “Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems,” June 2001.