Audio Engineering Society

Convention Paper

Presented at the 120th Convention
2006 May 20–23 Paris, France

A Novel Integrated Audio Bandwidth Extension Toolkit (ABET)

Deepen Sinha[1], Anibal Ferreira[1, 2], and Harinarayanan E. V.[1]

[1] ATC Labs, New Jersey, USA

[2] University of Porto, Portugal

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

ABSTRACT

Bandwidth Extension has emerged as an important tool for the satisfactory performance of low bit rate audio and speech codecs. In this paper we describe the components of a novel *integrated audio bandwidth extension toolkit (ABET)*. The ABET toolkit is a combination of two bandwidth extension tools: (i) The Fractal Self-Similarity Model (*FSSM*) for signal spectrum; and, (ii) Accurate Spectral Replacement (*ASR*). Combination of these two tools, which are applied directly to high frequency resolution representation of the signal such as the Modified Cosine Transform (MDCT), has several benefits for increased accuracy and coding efficiency of the high frequency signal components. At the same time the combination of the two tools entails a number of important algorithmic and perceptual considerations. In this paper we describe the components of the *ABET* bandwidth extension toolkit in detail. Algorithmic details, audio demonstrations, and, *ABET* configuration details are presented. Additional information and audio samples are available at http://www.atc-labs.com/abet/.

1. INTRODUCTION

Audio coding at low bit rates has many established and emerging applications. These include Satellite and Terrestrial Digital Audio Broadcasting, audio delivery over the mobile network, high quality audio communication over the IP and mobile network, Internet music download and streaming, solid-state audio playback devices, etc. In many of these applications the demand for higher compression efficiency continues to grow. In fact there appears to be a proliferation of applications demanding CD like quality stereo at bit rates of 32-48 kbps and high quality FM grade mono audio at bit rates of 16-24 kbps. These in turn continue to spur the demand for newer algorithms for audio bit rate reduction. Audio Bandwidth Extension has emerged as a key technique for achieving higher compression efficiency at hence high subjective quality at low bit rates.

The rapid growth in the field of Perceptual Audio Coding has yielded a number of audio coding technologies based on the principle of Adaptive Transform Coding [1]. These include proprietary schemes such as PAC (Bell Labs, Lucent) [2] and

ATRAC (Sony) [3] as well as standard based codecs such as MPEG-1 Layer 3 (popularly known as MP3) [4], MPEG-2 AAC [5], Dolby AC-3 [6]. At best these conventional audio coding techniques are capable of producing full fidelity CD quality audio 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. For the viability of these coding schemes for new and emerging applications it is desirable to reduce the bit rate further without sacrificing the audio bandwidth.

A second class of coding schemes is geared primarily towards the coding of voice signal for two way communications [14][7]. At the lowest bit rates these typically employ a variation of the Code Excited Linear Prediction (CELP) technique. These coding schemes typically code a small audio bandwidth (< 4 kHz). For these existing [7] and emerging [14] low bit rate coding schemes it is attractive to improve the audio bandwidth significantly with as little bit overhead as possible.

In order to reduce the bit rate requirement of adaptive transform coding schemes further, or to provide increased audio bandwidth with very low bit rate CELP based codecs, it becomes necessary to 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 or another suitable 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 [9] and [10] we introduced two novel bandwidth extension techniques which are applied directly to the high resolution frequency representation of the signal. The first described in [9] is based on a Fractal Self Similarity Model (*FSSM*) for the MDCT representation of audio signal. It was shown that the *FSSM* model works across a wide class of natural audio and is capable of providing detailed and natural sounding audio reconstruction. The second scheme, *Accurate Spectral Replacement* (*ASR*), was introduced [10]. *ASR* is capable of an extremely accurate reconstruction of the tonal components and harmonic structures in the synthesized high frequency spectrum of the signal.

Further work with the ASR and FSSM bandwidth extension tools has led to the understanding that the two techniques have several complementary aspects. The two techniques have therefore been combined into an integrated bandwidth extension platform called the Audio Bandwidth Extension Toolkit (ABET). Some of the highlights of ABET are as follows:

- ABET works with virtually any baseband coding scheme. It is particularly suitable for use in conjunction with coding schemes that employ a high resolution filterbank for coding. However, ABET has also been used with considerable success in combination with other coding schemes such as low bit rate speech codecs.

- ABET makes use of both the *FSSM* and *ASR* algorithms in an adaptive framework and also allows for the combination of aspects of *FSSM* and *ASR* synthesis. In other words through ABET, ASR and FSSM may either be used independently or in combination to exploit their complementary nature.

- ABET bandwidth extension models (ASR and FSSM) are applied in the domain of a high frequency resolution filterbank such as the Odd Discrete Frequency Transform (ODFT) or the Modified Discrete Cosine Transform (MDCT) [8].

- ABET also incorporates a third essential tool "Multi Band Temporal Amplitude Coding" (MBTAC) (also described in [9]). MBTAC may (optionally) be employed when the time resolution of the primary MDCT/ODFT filterbank is too low to allow for suitable temporal shaping of the reconstructed high frequency components. For the computation and application of MBTAC signal is analyzed using a secondary Utility Filter Bank (UFB) that has a significantly better time resolution.

- The presence of efficient and high quality coding tools for the stereo envelope allows ABET to function as the main building block of a parametric audio coding scheme offering accurate reproduction of stereo envelopes.

- These techniques offer the promise of a more accurate reconstruction of the synthesized high frequency spectrum in comparison to previously reported approaches such as the Spectral Band Replication approach [9].

In this paper we describe the components of ABET and discuss its application to actual audio coding schemes. We have utilized (parts of) ABET in the building of three audio coding products. These include the TeslaPro codec [12], the Audio Communication Codec [13][14], and a new very low bit rate coding techniques for mixed contents [15].

The organization of the rest of the paper is as follows. In Section 2 we take a closer look at the ABET encoder followed by the ABET decoder in Section 3. The primary coding tools in ABET – the *FSSM*, the *ASR*, and the UFB/MBTAC are further described in Section 4. Sections 5 and 6 Audio present the functional description of the ABET Encoder and Decoder processing blocks respectively. Coding results and the codecs utilizing the ABET scheme are discussed in section 7.

2. THE ABET ENCODER

The ABET Encoder is shown in Figure 1. ABET works in conjunction with a baseband coding scheme which is expected to encode the low pass-filtered signal information. ABET encoder is configurable using a set of options. These options are used to invoke one or more of the ABET components, control the relative precedent of the ABET components, and control the level of detail in a particular component. In a complete audio coding scheme, selection of the configuration parameter is typically a function of the bit rates and may need to be carefully tuned in conjunction with other codec parameters. As noted above, the bandwidth extension tools inherent in ABET, i.e. ASR and FSSM, operate on high resolution frequency representation of the signal. The ABET encoder therefore incorporates an integrated MDCT/ ODFT computation module. If the baseband coding scheme also operates in the MDCT/ODFT domain the transform information may be shared between the ABET encoder and the encoder of the baseband coding scheme; ABET encoder therefore makes the low-pass-filtered MDCT/ODFT coefficients available as part of its output. ABET supports window switching; in other words it is possible to use a conventional window switching algorithm of the type *Long*→*Start*→*Short*→*Stop*→*Long* (e.g., as in [2][4]) and the ABET parameters are suitably adopted to the time varying filterbank resolution. However, as noted above ABET incorporates additional tools for temporal shaping, reducing (and/or in certain cases eliminating) the need for window switching in a conventional MDCT/ODFT based baseband coding scheme.

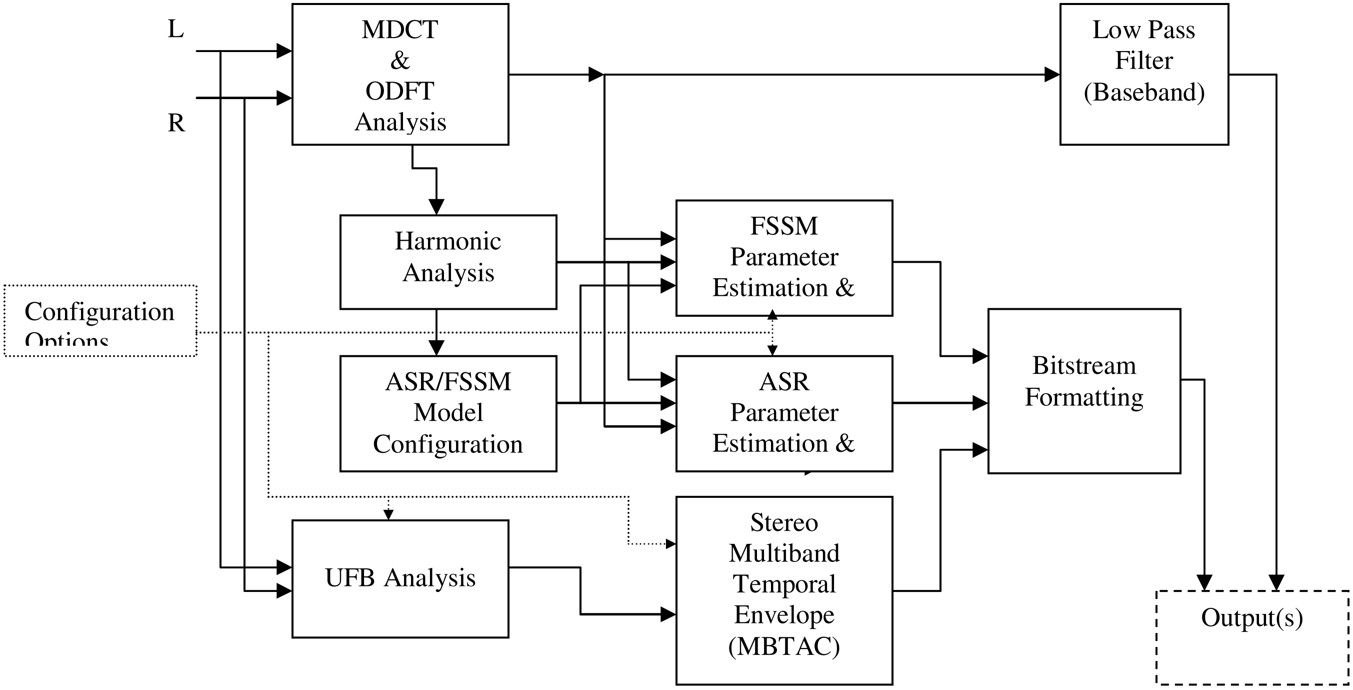

L
R
MDCT & ODFT Analysis
Low Pass Filter (Baseband)
Harmonic Analysis
FSSM Parameter Estimation &
Configuration Options
ASR/FSSM Model Configuration
ASR Parameter Estimation &
Bitstream Formatting
UFB Analysis
Stereo Multiband Temporal Envelope (MBTAC)
Output(s)

Figure 1: ABET Encoder

The ABET Encoder encompasses the following functional areas:

1. Frequency Analysis: MDCT/ODFT analysis as described above.
2. High Resolution Spectrum and Harmonic Analysis: detection of tones and harmonic features in a signal segment
3. *ASR/FSSM* Model Configuration: Selection of *ASR* and *FSSM* coding tools matched to the coding specific signal features. This is driven by the output of the spectrum/harmonic analysis block and the configuration parameters
4. *ASR* Parameter Extraction and coding
5. *FSSM* parameters extraction and coding
6. UFB analysis: a second time-frequency analysis of the signal with a higher time resolution than the primary MDCT/ODFT analysis.
7. Stereo MBTAC Coding: Joint encoding of stereo time-frequency envelope of the signal.
8. Huffman coding: noiseless coding of MBTAC, *ASR*, and, *FSSM* parameters.
9. Bitstream packing of all the encoded parameters.

3. ABET DECODER STRUCTURE

The ABET Decoder is shown in Figure 2. The primary job of the ABET decoder is to perform signal synthesis using the *FSSM* and *ASR* model in the ODFT domain. If the baseband coder utilizes MDCT representation, then a MDCT to ODFT mapping is utilized. In the cases where both *ASR* and *FSSM* models are simultaneously active additional processing is necessary to ensure that any harmonic pattern synthesized by *ASR* is not duplicated by the *FSSM* model. To ensure this partial *ASR* synthesis and subtraction in baseband is performed prior to the application of *FSSM* model synthesis. In the cases where the time resolution of the MDCT/ODFT filterbank is too high to allow for adequate temporal shaping, the MBTAC information is applied in the UFB domain.

The ABET decoder incorporates the following functional areas

1. Huffman decoding and de-quantization of *ASR*, *FSSM*, and, MBTAC information.
2. MDCT to ODFT transformation
3. *FSSM* synthesis module
4. *ASR* synthesis module, including the baseband

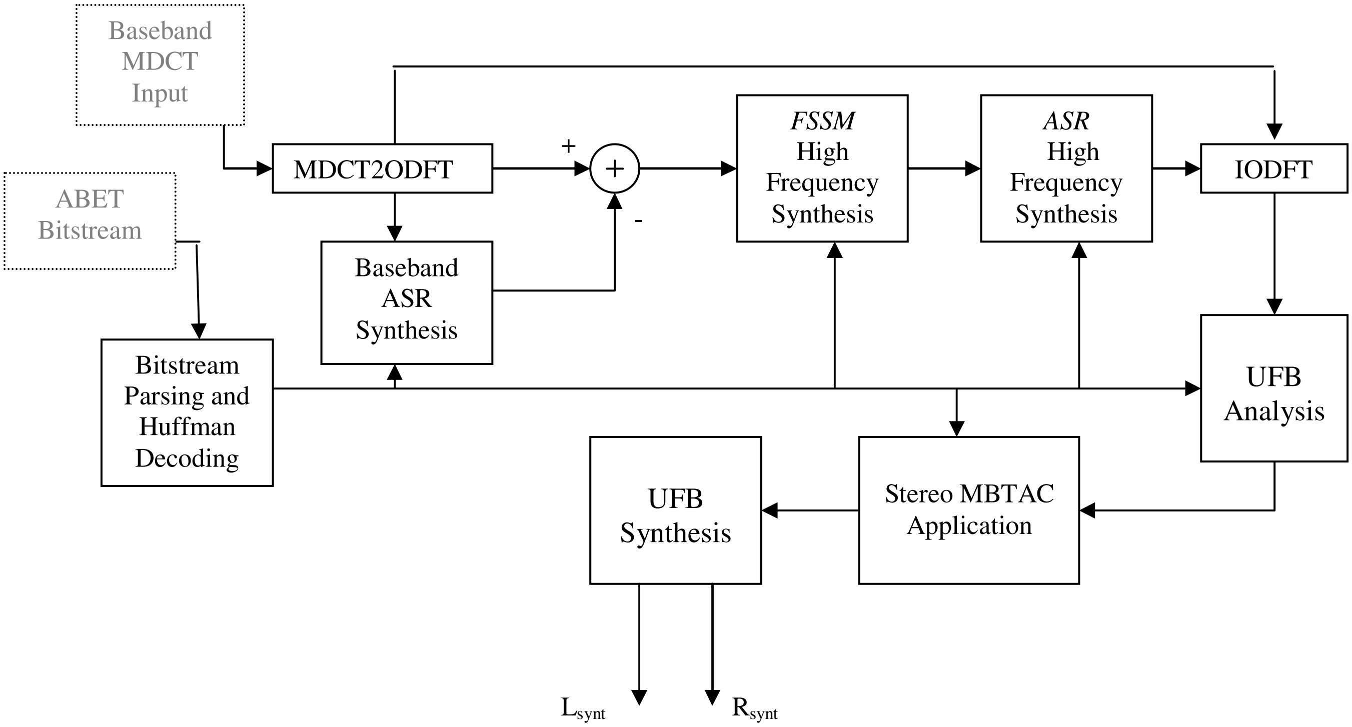


Figure 2: ABET Decoder

ASR synthesis and removal to ensure harmonious combinations of ASR and FSSM synthesized components.

5. Inverse ODFT/MDCT transformation
6. UFB analysis
7. MBTAC application in the UFB domain

4. PRIMARY CODING TOOLS IN ABET

As noted above, the proposed coding scheme utilizes two bandwidth extension tools. Here we provide a high level description of the two tools *FSSM* and *ASR*. For a detailed description of *FSSM* the reader is referred to [9], similarly, a detailed description of *ASR* may be found in [10]. In this section we describe the essential elements of both these models and also another important aspect of ABET, i.e. UFB filterbank and MBTAC.

The bandwidth extension paradigm may be formalized as below.

- It is assumed that in each audio frame, the spectral representation of the signal (such as the MDCT representation) up to certain frequency f_c, denoted as $X_{LP}(f)$, is coded directly using efficient quantization and coding techniques. It may be noted that it is not required that the baseband codec be a MDCT/ODFT domain coding scheme. What is required by ABET is that after decoding the signal is transformed into MDCT/ODFT domain.
- The MDCT/ODFT spectrum for frequencies $f > f_c$ is to be reconstructed using a mapping BE such that

$$\overline{X}_{HP}(f) = BE(\overline{X}_{LP}(f)) \qquad (3)$$

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

4.1. *The FSSM* Bandwidth Extension Model

In the *FSSM* technique high frequency components of the signal are reconstructed using an iterative sequence of *Expansion Operators* (EO) as below,

$$\overline{X}_{HP}(f) = \cdots EO_i \circ (\cdots (EO_1 \circ (EO_0 \circ \overline{X}_{LP}(f)) \cdots) \qquad (4)$$

Where each expansion operator EO_i is assumed to have the form

$$EO_i \circ \overline{X}_{LP}(f) = H_i \bullet X_{LP}(\alpha_i f - f_i) \qquad (5)$$

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 described further in [9]. The dilation/translation equations suggest a Fractal like Model for *FSSM* which is able to reconstruct the high frequency spectral details with a high level of accuracy across a wide range of different audio signals.

The significance of the dilation and translation terms in *FSSM* is illustrated with the help of coding examples in Figures 3 (a), (b), (c). For example, the translation term improves the accuracy of reconstruction for musical instruments with a pitch structure and also for voiced speech and vocal signals. For these classes of signals the lack of dilation terms results in a discontinuity in the pitch structure. This is illustrated in Figure 3 (a) and (b). Figure 3(a) shows the reconstructed spectrum superimposed over the original spectrum using a different bandwidth extension scheme (such as the spectrum replication approach of [11]).This is compared against the reconstruction using the *FSSM* model as shown in Figure 3(b).

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 3(c).

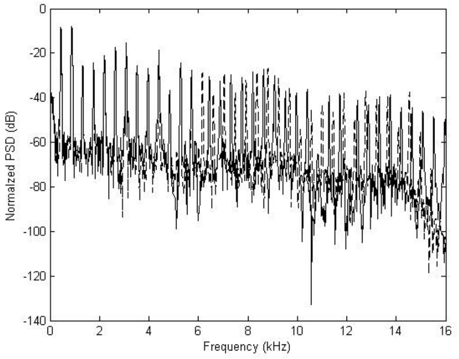


Figure 3(a): Reconstructed signal spectrum (solid line) and original spectrum (dashed line) using a spectrum replication approach.

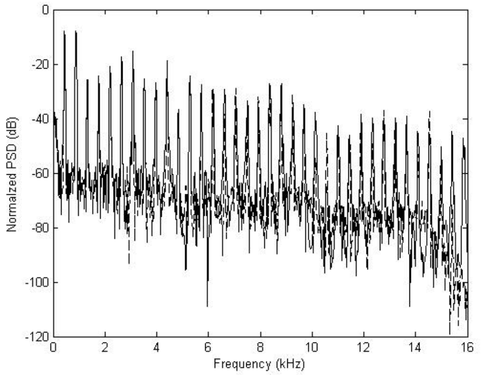


Figure 3(b): Reconstructed signal spectrum (solid line) and original spectrum (dashed line) with the *FSSM* model.

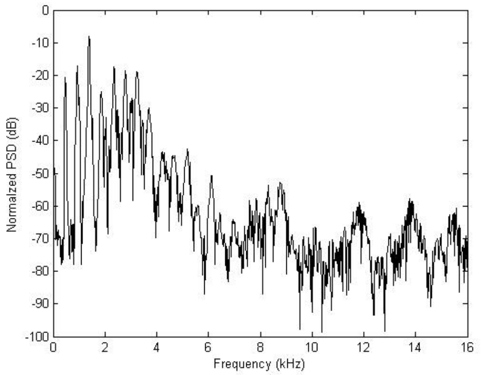


Figure 3(c): Example of a signal (short-term PSD) that benefits from the inclusion of the dilation term in the *FSSM* model.

The *FSSM* model in general is a *FSSM+Isolated Tones+Noise* model. In a subsequent sub-section we discuss that in the cases where FSSM is used as the primary bandwidth extension model, it is advantageous to encode secondary tonal components (e.g., a secondary harmonic sequence and isolated tones) using the *ASR* model. In general it may still be necessary to add synthetic noise for part or the entire short term spectrum that does not fit the *FSSM* (or ASR tonal model). In practice, however, if the dilation parameter in the FSSM model is suitably estimated, the occurrence of such cases is rather infrequent.

An interesting observation related to the *FSSM* model is that the temporal envelope of the reconstructed high frequency components using the *FSSM* model shows a high level of coherence with the temporal envelope of the base band components. This observation is illustrated with the help of a synthetic narrowband noise signal in Figure 4. The figure shows the base band 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).

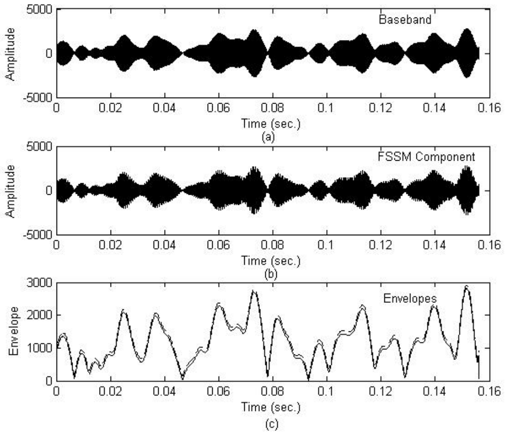


Figure 4: (a)Base band noise signal, (b)*FSSM* constructed high frequencies, (c) Envelopes of (a) & (b)

4.2. The *ASR* Bandwidth Extension Model

The *ASR* Model for bandwidth extension is described in detail in [10]. It takes into account the specificity of the coherent (i.e., sinusoidal) components of an audio signal, as well as the specificity of the incoherent (i.e., noise) components of an audio signal, namely with respect to their different perceptual impact and their different spectral nature and fine spectral structure. At the heart of *ASR* is a sinusoidal analysis and synthesis algorithm *with sub-bin accuracy*. The *ASR* model is particularly effective when the audio signal exhibits a well defined harmonic structure of sinusoids. In this case a bandwidth extension technique based on the replication of base band components may not provide satisfactory reconstruction of higher order partials. A replication model in this case, as noted above, has a significant deficiency in the sense that it may either break the organization of the harmonics in frequency which is likely to be noticeable to the human auditory system in the form of a pitch shift or the appearance of several pitches instead of a single one. *ASR* also allows sufficient and flexible control over the phase of the synthesized higher order partials which may not be possible in techniques utilizing mapping based on the lower frequencies (base band). The most general form of *ASR* processing consists of the following steps.

1. Normalization of the audio spectrum by a model of the smooth spectral envelope, the noise part of the resulting flattened spectrum is very approximately white.

2. Segmentation of the flattened spectrum into sinusoids and a residual (or noise), this residual results by removing (i.e., by subtracting) sinusoids directly from a complex discrete frequency representation of the audio signal, presuming that this representation is able to resolve all existing sinusoidal components.

3. Synthesis and bandwidth extension of sinusoids with sub-bin accuracy and using a reduced set of parameters (frequencies, magnitude, or phases) describing the original audio sinusoidal components.

4. Synthesis and bandwidth extension of noise with bin accuracy (in the next sub-section we discuss how it may be advantageous to extend the noise component using the *FSSM* model).

5. Sum of both bandwidth extended components and inverse normalization in order to recover the spectral envelope model of the original spectrum.

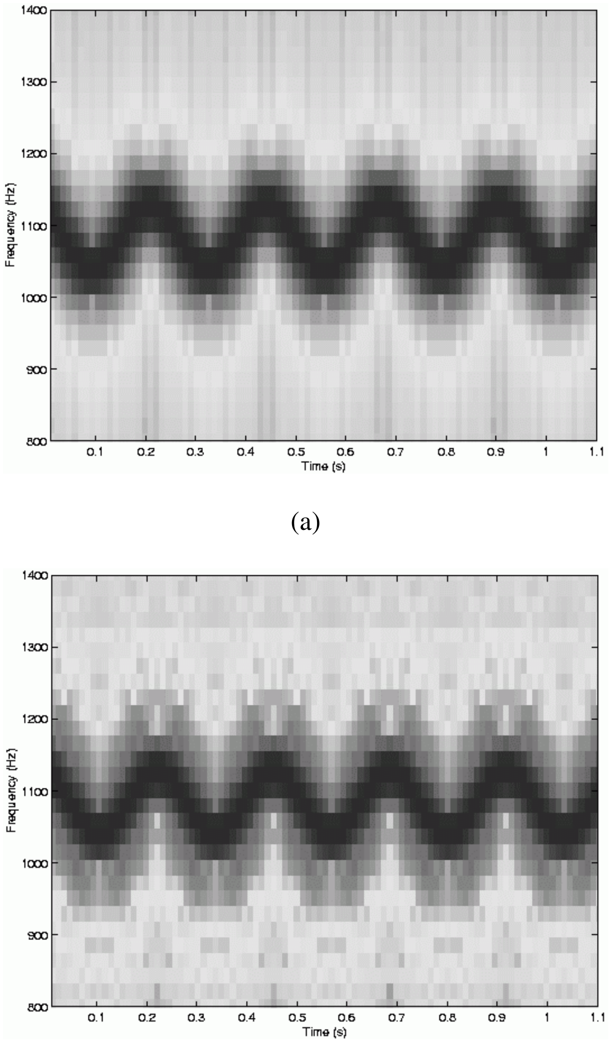
The *ASR* model is highly flexible in terms controlling the spectral balance of the reconstructed high frequency components. For example, the spectral tilt affecting the incoherent components, and the spectral tilt controlling the sinusoidal components can be shaped and controlled in an independent way.

Further details on ASR may be found in [10] and at http://www.atc-labs.com/asr. In the *ASR* model the parameters necessary for the synthesis of harmonic partials are suitably reduced. For example in many cases the phase information may be completely discarded, or in other cases it is transmitted only at the time of harmonic birth and used in conjunction with a synthesis technique that insures phase continuity from frame to frame.

The primary filterbank domain for ASR processing is the Odd-DFT (ODFT). At the encoder sinusoidal components are estimated from the ODFT spectrum and removed by direct synthesis of ODFT spectral bins using a model of the frequency response of sine window [23, 24] and the estimated frequency, magnitude, and phase parameters. It has been concluded that only a small number of frequency bins per sinusoidal component are needed to generate a good quality sinusoid and to effectively remove it from the ODFT spectrum. The sinusoidal components are further analyzed to detect the presence of one or more harmonic patterns (including harmonics with missing fundamentals) as well as isolated (non-harmonic) sinusoidal components. Parameters necessary for the synthesis of high frequency sinusoidal components are then analyzed and suitably reduced (e.g., by discarding the phase components). The reduced parameters are forwarded to the decoder.

In the decoder the high frequency sinusoidal components can be synthesized directly in the ODFT domain, avoiding the TDAC mechanism associated with MDCT. A sinusoidal continuation algorithm is used to generate sinusoidal trajectory using only the transmitted frequency and magnitude parameters. In most cases phase information is only needed at the time of harmonic birth. Furthermore, in most cases a reduced level of magnitude information in the form of a smooth spectral envelope is needed for the sinusoidal continuation algorithm.

The accuracy of sinusoidal synthesis using the ASR model is depicted in Figure 5 using a synthetic FM modulated sinusoid . The synthesis accuracy for natural audio is highlighted in [10] with additional illustrations.



(a)

(b)

Figure 5: Spectrogram of FM Modulated Sinusoid; (a) Original (b) ASR Synthesized

4.3. Adaptive Combination of *FSSM* and *ASR* Models

As noted above, ABET allows for the flexible application of *FSSM* and *ASR* bandwidth extension models either independently or in combination with each other. Practical combinations of the two models include (but are not limited by) the scenarios described below. It may also be noted that ABET allows for the model configuration to change on a frame to frame basis.

1. In this case the ASR model is utilized for the encoding and synthesis of the dominant harmonic sequence in the signal and also isolated (inharmonic) tones. The secondary harmonic sequence (if one is present) and the non-tonal components are modeled by the FSSM algorithm.

2. In this case all the harmonically related tonal components and inharmonic signal components are coded by the FSSM model. The ASR model is then used to encode and synthesis isolated (inharmonic) tones. In this configuration the FSSM model estimation algorithm emphasizes accurate reconstruction of the dominant harmonic tone sequence in the signal.

3. In this case all the tonal components (up to two harmonic sequences and isolated tones) are coded using the ASR model. The non-tonal noise-floor is then modeled using the FSSM approach.

In audio frames where both *FSSM* model and *ASR* sinusoidal synthesis model is active, it is important to ensure that the harmonics synthesized by the *FSSM* and the *ASR* models do not interfere with each other. This may happen for example in a case when *FSSM* models the dominant harmonic in the signal and *ASR* is used to synthesize the secondary harmonic. Unless care is taken, the *FSSM* synthesis will also create high frequency partials corresponding to the secondary harmonics (albeit with inaccurate frequency location). These will then interfere with partials generated by the *ASR* model. To eliminate this problem the partials due to the secondary harmonic pattern are subtracted from the baseband before the application of the FSSM model (this process is illustrated in the ABET decoder block diagram, Figure 2).

4.4. Utility Filter Bank (UFB) and Multiband Temporal Amplitude Coding (MBTAC)

Since the frequency resolution of the primary coding and bandwidth extension filter bank is typically quite high ABET incorporates additional tools for the shaping of the temporal envelopes of the signal in multiple frequency bands which may be optionally invoked. This aspect is discussed in more detail below.

At the heart of the temporal shaping tools in ABET is the "Utility Filter Bank" (*UFB*).The *UFB* is a complex, over-sampled modulated filter bank [9]. An over sampling ratio between (and including) 2 and 8 is permitted by ABET. Depending upon configuration parameters (e.g., based on the complexity profile of the decoder and bit rate of operation) the *UFB* may take one of the following 2 forms.

- A complex modulated filter bank with an over-sampling ratio between 2 and 8 and sub-band filters of the form

$$h_i = h_0 \cdot e^{j\frac{2\pi}{N}\cdot i\cdot n} \qquad (7)$$

  where h_0 is an optimized prototype filter. N = 128 and N = 256 are allowed.

- A complex non-uniform filter bank; e.g., one with two uniform sections and transition filters to link the 2 adjacent uniform sections as described in [9]. This filter bank is designed using the technique described in [27]. The sub-bands in the lower sections have ½ the bandwidth of the sub-bands at higher frequencies. The higher frequency resolution at lower frequencies is useful, for example, in parametric stereo coding.

MBTAC information to perform the temporal shaping 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 and by grouping the information in adjacent time and frequency bands. The highlights of the MBTAC algorithm are as follows.

- Supports non-uniform time-frequency tiling for the computation of signal envelope. The initial frequency resolution is configurable into bands which are either full, half, or, quarter critical band wide.
- Incorporates several tools for the efficient coding of envelope which look for typical and/or perceptually significant patterns in the time-frequency envelopes. These include techniques for noiseless coding and grouping based on perceptual criterion.
- Efficient techniques for the coding of stereo envelopes.

5. FUNCTIONAL DESCRIPTION OF ABET ENCODER PROCESSING BLOCKS

In this section we present additional details regarding several functional blocks of the ABET Encoder.

High Resolution Frequency Analysis (MDCT/ODFT) is the first block in the ABET encoder. It simultaneously computes the MDCT and ODFT representation (for two channels). The MDCT/ODFT analysis is computed for two frequency resolutions: (i) a Long window which is typically 2048 sample long (with 1024 sample overlap between two consecutive windows), (ii) a Short window which is typically 256 samples long (with 128 sample overlap between two consecutive windows). ABET includes its own window state detector. This information needs to be shared and synchronized with the baseband coding scheme in the case where the frequency analysis is common.

Accurate Harmonic Analysis is the next functional block in the encoder it involves the detection of all the tonal components in signal using the ODFT representation. The frequencies of the tonal components are accurately estimated using the algorithm described in [23]. The tonal components are further analyzed to determine if these fit into a harmonic structure (the possibility of missing harmonics up to a 7th order is allowed). The output of the accurate harmonic analysis block is the parameters corresponding to one or more detected harmonic patterns as well as the parameters of isolated (inharmonic) tonal components in the high frequency region.

ASR/FSSM Model Configuration: Based on the user selected parameter (e.g., *ASR/FSSM* model order, number of harmonic patterns to be coded etc.) and the output of accurate harmonic analysis, a decision is made regarding the frequency structures (harmonics and tones) which are to be coded by *FSSM* and *ASR* respectively.

Accurate Spectral Replacement (*ASR*) model parameter estimation is the next functional block at the encoder. For the harmonic patterns coded by the ASR model, the transmitted information consists of the fundamental frequency as well as the envelope of the high frequency partials computed using a suitable frequency band structure. This envelope is differentially coded and further compressed using noiseless coding. For stereo signal same harmonic pattern is present in both the channels, the parameters are jointly coded for higher efficiency. For isolated tones transmitted information consists of the frequency and magnitude of the tone.

Fractal self-similarity model (*FSSM*) follows the *ASR* functional block. The FSSM model parameters are estimated using a combination of 3 criteria: (1) Maximization of a Self-similarity coherence (*SSC*) function as defined below:

$$\Phi(\alpha_i, f_i) = \left\langle X\ (f) \cdot X(\alpha_i f - f_i) \right\rangle \qquad (6)$$

(2) A harmonic continuity criterion to ensure the accuracy of the dominant harmonic structure in the signal, (3) Consistency criterion over time (multiple audio frame) to ensure steady alias-free reconstruction of steady harmonics. Furthermore, the quality of the estimates improves significantly if the MDCT spectrum is normalized by the coarse envelope prior to the estimation of these parameters.

The Utility Filterbank (UFB) as described above is a complex modulated filterbank with several times over-sampling. It allows for a time resolution as high as 16/Fs (where Fs is the sampling frequency) and frequency resolution as high Fs/256. It also optionally supports a non-uniform time-frequency resolution.

Multi Band Temporal Amplitude Coding (MBTAC) involves efficient coding of two channel (stereo) time-frequency envelopes in multiple frequency bands. The resolution of MBTAC frequency bands is user selectable. The envelope information is grouped in time and frequency and jointly coded (across two channels) for coding efficiency. Various noiseless coding tools are used to reduce bit demand.

6. FUNCTIONAL DESCRIPTION OF ABET DECODER PROCESSING BLOCKS

The MDCT coefficients from the encoder are mapped to ODFT coefficients using a mapping described in [24].The low pass spectrum is analyzed for the presence lower order partials corresponding to the harmonic structure(s) which are designated to be encoded by the ASR model. The identified partials are synthesized and subtracted from the input low-pass spectrum to get a flattened spectrum.

FSSM reconstruction is applied on the flattened spectrum. On applying dilation and translation parameters with spectral norm values, the high frequency flattened spectrum is approximately reconstructed.

ASR at the decoder involves synthesizing the chosen harmonic structure and high frequency tones from the encoder information. The synthesized sinusoids are added to the *FSSM* full band spectrum to reconstruct the original spectrum.

MBTAC application in the UFB domain ensures that the temporal envelope approaching the original signal is maintained after the reconstruction from the bandwidth extension technique. MBTAC application involves suitable smoothing techniques.

7. CODING RESULTS

The ABET toolkit (or its subset) has been employed in three audio codecs developed by ATC Labs. In first of these products *TeslaPro [12]*, which is geared towards broadcast applications, ABET is employed in its full strength with adaptive combination of the *FSSM*, *ASR*, and, MBTAC tools. Coding results at multiple bit rates (between 20 – 48 kbps) using TeslaPro are available at http://www.atc-labs.com/teslapro. In this codec ABET is used to encode up to 75% of the audio bandwidth.

In a second audio codec geared towards two-way audio communication, the ASR and FSSM models are used for bandwidth extension. The shorter block length of this codec, called the Audio Communication Codec (ACC) [13][14] obviates the need for additional temporal envelope shaping, hence UFB/MBTAC is not employed. Coding results using ACC are available at http://www.atc-labs.com/acc.

In a third audio coding product geared towards very low bit rate coding of voice and mixed content, ABET is employed for bit rates as low as 4-6 kbps. Coding results using this recently introduced codec [15] may be found at http://www.atc-labs.com/lbrcodec.

The bit overhead due to ABET is a function of the model configuration parameters and the fraction of bandwidth coded ABET. The table below summarizes the overhead for a few preferred configurations. It typically ranged between 2-3 kbps/channel.

% BW Coded by ABET	ASR/FSSM Config	MBTAC Config	Overhead per channel
50	1st har-FSSM 2nd har & iso tones - ASR	Very Detailed Envelope	3.1 kbps
50	1st har-FSSM 2nd har & iso tones - ASR	Moderately Detailed Envelope	2.5 kbps
50	1st ha-FSSM iso tones - ASR	Moderately Detailed Envelope	2.1 kbps
75	1st har-FSSM iso tones - ASR	Moderately Detailed Envelope	2.6 kbps
50	1st har & iso tones - ASR 2nd har & noise floor - FSSM	Moderately Detailed Envelope	3.5 kbps
50	1st har & iso tones - ASR 2nd har & noise floor - FSSM	No Envelope	2.5 kbps

8. CONCLUSIONS

We described a novel audio bandwidth extension toolkit with application to low bit rate audio and speech coding. The proposed toolkit, called ABET, allows for flexible combination of the *FSSM* and *ASR* bandwidth extension models. It incorporates additional tools for accurate shaping of the time-frequency envelope of the signal. Coding results indicate that ABET allows for a very high quality reconstruction of the high frequency signal components that is significantly more accurate than other similar techniques.

9. REFERENCES

[1] A. Gersho and R. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Press, 1992.

[2] 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.

[3] 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.

[4] 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.

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

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

[7] ITU-T Recommendation G.729,- "Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP)", March 1996

[8] 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.

[9] Deepen Sinha, Anibal Ferreira, and, Deep Sen “A Fractal Self-Similarity Model for the Spectral Representation of Audio Signals,” *118th Convention of the Audio Engineering Society*, May 2005, Paper 6467.

[10] Anibal J. S. Ferreira and Deepen Sinha, "Accurate Spectral Replacement," *118th Convention of the Audio Engineering Society*, May 2005, Paper 6383.

[11] 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.

[12] Deepen Sinha and Anibal Ferreira "A New Broadcast Quality Low Bit Rate Audio Coding Scheme Utilizing Novel Bandwidth Extension Tools," *119th Convention of the Audio Engineering Society*, October 2005. Paper 6588.

[13] Anibal J. S. Ferreira and Deepen Sinha, "A New Low-Delay Codec for Two-way High-Quality Audio-Communication," *119th Convention of the Audio Engineering Society*, October 2005, Paper 6572.

[14] Anibal J. S. Ferreira and Deepen Sinha, "Audio Communication Coder," *in the preprints of 120th Convention of the Audio Engineering Society*, May 2006.

[15] Raghuram A., Anibal J. S. Ferreira, and Deepen Sinha, "A New Low Bit Rate Speech Coding Scheme for Mixed Content," *in the preprints of 120th Convention of the Audio Engineering Society*, May 2006.

[16] Joseph L. Hall, *"Auditory Psychophysics for Coding Applications,"* Section IX, Chapter 39, The Digital Signal Processing Handbook, CRC Press, Editors: Vijay K. Madisetti and Douglas B. Williams, 1998.

[17] B.C.J. Moore, *An Introduction to the Psychology of Hearing, 5th Ed.*, Academic Press, San Diego (2003).

[18] Eberhard Zwicker, and Hugo Fastl, *Psychoacoustics: Facts and Models*, Springer Series in Information Sciences (Paperback), Second updated edition.

[19] 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.

[20] D. Sinha, *Low bit rate transparent audio compression using adapted wavelets.* Ph.D. thesis, University of Minnesota, 1993.

[21] Hall JW, Grose JH, Mendoza L (1995) Across-channel processes in masking. In: Hearing (Moore BCJ, ed), pp 243–266. San Diego:Academic.

[22] Jesko L. Verhey, Torsten Dau, and Birger Kollmeier "Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation filter bank model" Journal of the Acoustical Society of America, 106(5), p. 2733-2745.

[23] Anibal J. S. Ferreira and Deepen Sinha, "Accurate and Robust Frequency Estimation in ODFT Domain," *in the proceedings of the 2005 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, October 16-19, 2005.

[24] 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.

[25] 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.

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

[27] 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.

[28] 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.

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