



Audio Engineering Society Convention Paper 6383

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

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Accurate Spectral Replacement

Aníbal J. S. Ferreira^{1,2}, and Deepen Sinha²

¹University of Porto, Portugal

²ATC Labs, USA

Correspondence should be addressed to A. Ferreira (ajf@atc-labs.com, a.j.ferreira@ieee.org)

ABSTRACT

Recent advances in perceptual audio coding are strongly based on the concept of bandwidth extension. Most techniques implementing bandwidth extension require an analysis/synthesis filter bank in addition to that used by the associated perceptual audio coder, which increases the overall system complexity and coding delay, and makes difficult the correct alignment between the operation of the audio coder and the operation of the bandwidth extension technique. We present a new Accurate Spectral Replacement (ASR) technique that is based on a suitable decomposition of the MDCT filter bank, and that implements synthesis of sinusoidal components with an accuracy much higher than the natural frequency resolution of the filter bank. The ASR technique is described, its performance is assessed with both synthetic and natural audio signals, and its main areas of application are addressed. Audio demos are available at <http://www.atc-labs.com/asr/>.

1. INTRODUCTION

Perceptual Audio Coding is a very successful approach in the compression of high quality audio signals. Several proprietary audio coding schemes such as ATRAC (Sony) [1] and PAC (Bell Labs, Lucent) [2], as well as standardized audio coding schemes such as the Dolby AC-3 [3], the MPEG-1 Layer 3 (MP3) [4] or the MPEG-2 AAC [5], are based on the perceptual audio coding paradigm [6] and are successfully used in many application areas.

Recent technological advances in perceptual audio

coding focus on approaches well known in speech coding, and specifically on the concept of bandwidth extension of speech or audio signals. Bandwidth extension consists in the reconstruction of a spectral region of the speech or audio signal and that is missing, using such techniques as frequency shifting (transposition), non-linear filtering, or band replication, and in a way that the bandwidth extended signal sounds as close as possible to the original full band signal, or at least sounds more pleasant than the bandwidth reduced signal.

Most frequently, bandwidth extension techniques are based on a controlled transposition of preserved spectral portions of the audio signal, so as to fill the missing spectral regions in a perceptually suitable way. Typically, bandwidth extension techniques imply the use of filter banks in addition to the natural filter bank of the perceptual audio coder the bandwidth extension technique operates with. This implies severe penalties in terms of system complexity, coding delay, and matching of the processing carried out between the perceptual audio coder and the bandwidth extension technique. One example of such mismatch is the fact that, most often, harmonicity is not achieved for the reconstructed spectrum when the original signal contains several sinusoids harmonically related, yet several studies indicate that the majority of all speech and audio material contains clear harmonic structures of sinusoids [7].

In this paper we present Accurate Spectral Replacement (ASR) as a new technique addressing the problem of accurate bandwidth extension and that is characterized by the following main properties:

- does not imply other filter bank than the MDCT, it just takes advantage of a suitable decomposition of the MDCT filter bank, with clear advantages in terms of system complexity and coding delay,
- it implements direct synthesis of sinusoids in the frequency domain in an extremely accurate way that could not be reached using frequency transposition,
- it paves the way to the independent control of the spectral tilt associated with the harmonic content of the signal (*i.e.*, the coherent part), and the spectral tilt associated with the stationary incoherent part (*i.e.*, the noise) of the audio signal.

This paper is organized as follows: in section 2 we present the main concept underlying the ASR technique. In section 3 we describe the actual implementation of the ASR technique that has been used to obtain all results presented in this paper. In section 4 we illustrate the performance of the ASR technique using both synthetic and natural audio signals that

are available on a Web page that complements this paper. Section 5 concludes this paper with a perspective on possible application areas.

2. THE ASR CONCEPT

The ASR concept is illustrated in Figure 1. It consists in a bandwidth extension technique that 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. In fact, the spectral representation of sinusoidal components exhibits a predictable and well defined magnitude and phase behavior and even a small perturbation in these characteristics is easily detected by the human auditory system. Furthermore, due to the sampled nature of a discrete time-frequency Fourier analysis such as the DFT or the Odd-DFT (ODFT) [7], a simple frequency displacement (or replication) of frequency bins (*i.e.*, samples in the discrete frequency domain) pertaining to a sinusoidal component, does not allow complete freedom and effective control over the frequency (and phase) of the displaced sinusoidal component. For this reason, when the audio signal exhibits a well defined harmonic structure of sinusoids, as it happens frequently in speech and audio signals, a simple displacement or replication of frequency bins breaks the harmonic organization of the sinusoidal components or partials, which is likely to be noticeable by the human auditory system in the form of a pitch shift, or the appearance of several pitches instead of a single one [8, 9, 10]. An appropriate procedure of frequency displacement of sinusoidal components on a discrete frequency scale *must provide sub-bin accuracy*.

On the contrary, the perceptual impact of incoherent components of an audio signal is essentially determined by their power distribution in frequency. As a consequence, a spectral displacement or replication of frequency bins pertaining to incoherent audio components affects essentially the perceived coloration of the noise. Therefore, in this case, sub-bin accuracy is not required in the displacement or replication process.

The ASR technique takes into consideration these aspects and requirements by segmenting the components of an audio signal into sinusoids and noise, and

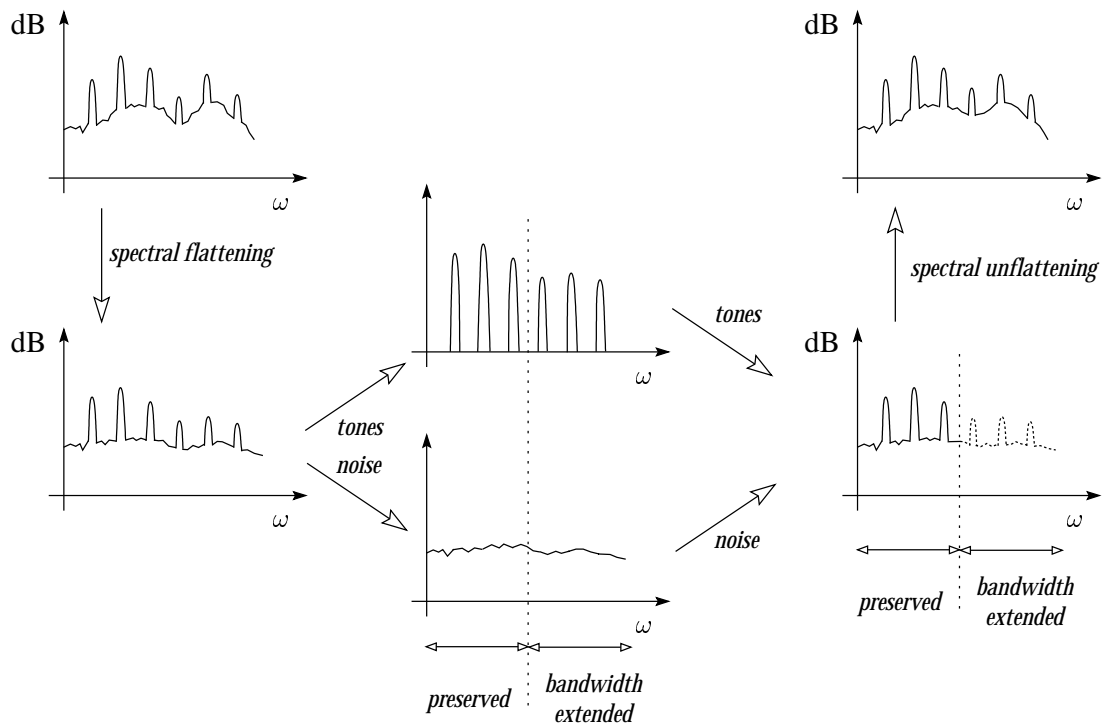


Fig. 1: Illustration of the Accurate Spectral Replacement concept: bandwidth extension is achieved by means of independent and accurate processing of both coherent and incoherent components of the audio signal.

processing them separately. This approach is therefore more accurate than other bandwidth extension techniques that perform spectral replication on both components simultaneously.

The major processing steps of the ASR technique are as follows:

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, magnitudes or phases) describing the original audio sinusoidal components,

4. synthesis and bandwidth extension of noise with bin accuracy and using a reduced (low-band) spectral portion of the original audio residual,
5. sum of both bandwidth extended components and inverse normalization in order to recover the spectral envelope model of the original spectrum.

This approach is highly flexible in that different criteria can be applied to the bandwidth extension of sinusoidal components and to the bandwidth extension of noise components. For example, the spectral tilt affecting the incoherent components and the spectral tilt affecting the sinusoidal components can be shaped and controlled in an independent way, which is not possible or at least is not trivially achieved using most known bandwidth exten-

sion techniques. On the other hand, the bandwidth extended sinusoidal components may not fall in the same frequency region of the bandwidth extended noise part of the reconstructed audio signal.

As another example, the reduction of sinusoidal parameters is itself highly flexible since this reduction may differently affect the frequencies, the magnitudes or the phases. In the case of harmonic sinusoids, only the fundamental frequency is strictly needed and in other cases phase information may be completely discarded.

All this added flexibility can be used to improve the tradeoff regarding subjective quality of the reconstructed audio signal and the bit rate of its compressed representation.

3. IMPLEMENTATION OF THE ASR TECHNIQUE

The implementation of the ASR technique is illustrated in Fig. 2 for the encoder part, and in Fig. 3 for the decoder part. For illustration purposes, as will be detailed in section 4, this implementation denotes a specific example of reduction of sinusoidal parameters: all phase information is discarded at the encoder.

This implementation presumes that:

- an analysis overlap-add building block (not shown in Fig. 2) exists at the input using the sine window that is commonly used in audio coders [7],
- a synthesis overlap-add building block (not shown in Fig. 3) exists at the output using the sine window,
- an audio coding environment exists where the MDCT is the filter bank used for signal analysis, synthesis and shaping of the coding noise.

As Figures 2 and 3 address only the implementation of the ASR concept, the important building blocks of an audio coder and corresponding to the psychoacoustic model, the quantizer, the rate loop and entropy coding, are not included. These will be included in the future in an integrated ASR-aware audio coder.

3.1. The ASR Encoder

The ASR encoder takes advantage of the decomposition of the MDCT analysis filter bank as a complex ODFT filter bank and a simple (complex-to-real) ODFT2MDCT frequency mapping. The ODFT provides magnitude and phase information. A smooth spectral envelope model is derived by short-pass filtering the real cepstrum, similarly to the technique presented in [11]. This model is used to flatten the ODFT spectrum and is also transmitted to the ASR decoder. The frequency, magnitude and phase of sinusoidal components are estimated from the ODFT spectrum using techniques similar to those presented in [12]. In particular, the magnitude of sinusoidal components is estimated in the flattened ODFT spectral domain in order to anticipate and facilitate its differential coding and entropy coding.

Sinusoidal components are removed (*i.e.*, subtracted) from the flattened ODFT spectrum by direct synthesis of ODFT spectral bins using a model of the frequency response of the sine window [12] and the estimated frequency, magnitude and phase parameters. It has been concluded that only a small number of spectral bins per sinusoid are needed in order to generate a good quality sinusoid and to effectively remove it from the ODFT spectrum.

A complex residual is obtained after sinusoidal subtraction in the flattened ODFT spectrum. This (complex) ODFT residual is transformed to a (real) MDCT residual by means of a ODFT2MDCT mapping. The MDCT residual is essentially white noise. As a consequence, depending on the desired quality of the reconstructed audio signal after decoding, the MDCT residual can be transmitted to the decoder in several possible ways:

- as a full-band signal,
- as a bandwidth reduced signal,
- as a simple parametric description, for example, in the form of an average power spectral density.

In this paper, the results shown in section 4 presume that audio signals are sampled at 44100 samples/second and that the MDCT residual is reduced to the frequency range [0, 6kHz], *i.e.*, to about 1/3 of the original bandwidth. As suggested in Fig. 2, the

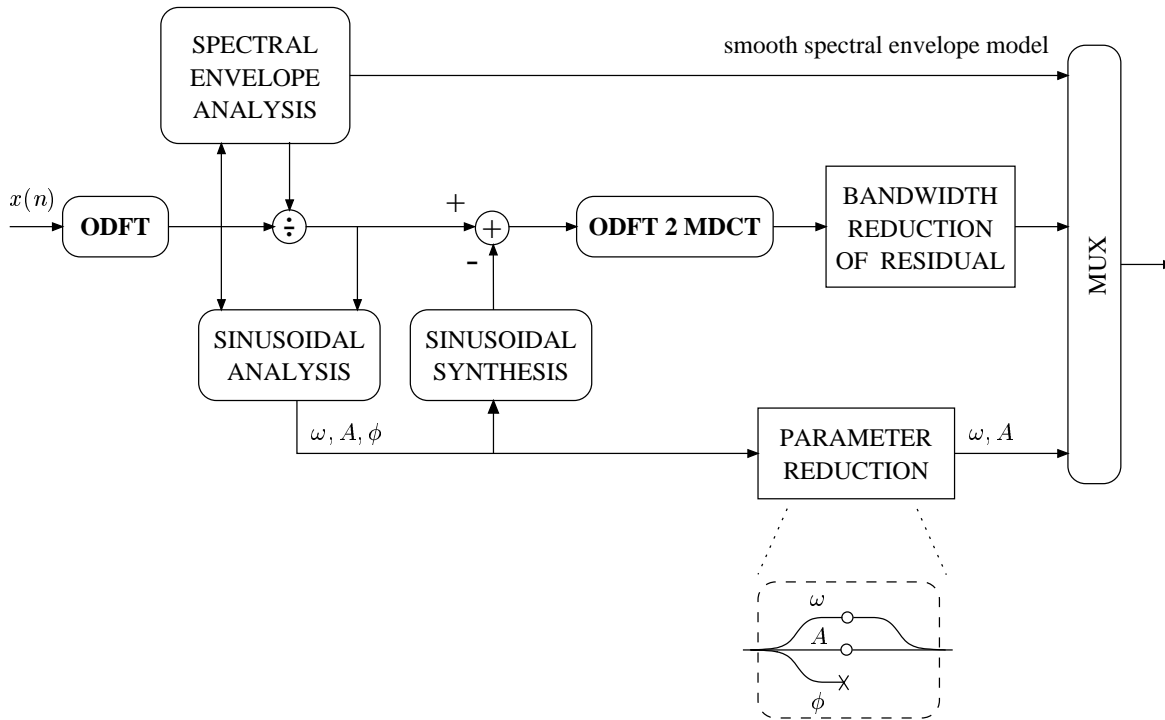


Fig. 2: A particular implementation of the ASR technique (encoder).

results presented in this paper presume that the reduction of sinusoidal parameters consists simply in discarding the phase of all sinusoidal components. This option has been taken so as to allow the assessment of the importance of phase in audio and so as to test the performance of the artificial phase generation algorithm in the decoder, as will be described next.

3.2. The ASR Decoder

The structure of the ASR decoder allows full flexibility regarding the independent control of the four major steps in the reconstruction of the audio signal:

1. bandwidth extension of the residual,
2. synthesis of sinusoidal components lying within the bandwidth of the transmitted MDCT residual,
3. synthesis of sinusoidal components lying outside bandwidth of the transmitted MDCT residual,

4. spectral denormalization by the transmitted smooth spectral envelope model.

Each one of these aspects is detailed in the following.

bandwidth extension of residual: the MDCT residual is bandwidth extended by means of a controlled transposition of the transmitted residual and its combination with synthetic noise,

in-band synthesis of sinusoids: the sinusoidal components that have been subtracted from the original (flattened) ODFT spectrum and that lie within the bandwidth of the transmitted MDCT residual, must be reconstructed in the MDCT domain and added back to the transmitted MDCT residual so as to respect the Time Domain Aliasing Cancellation property (TDAC) of the MDCT filter bank [13]. Strictly, this is only required if all parameters (phase, magnitude and frequency) describing a sinusoid

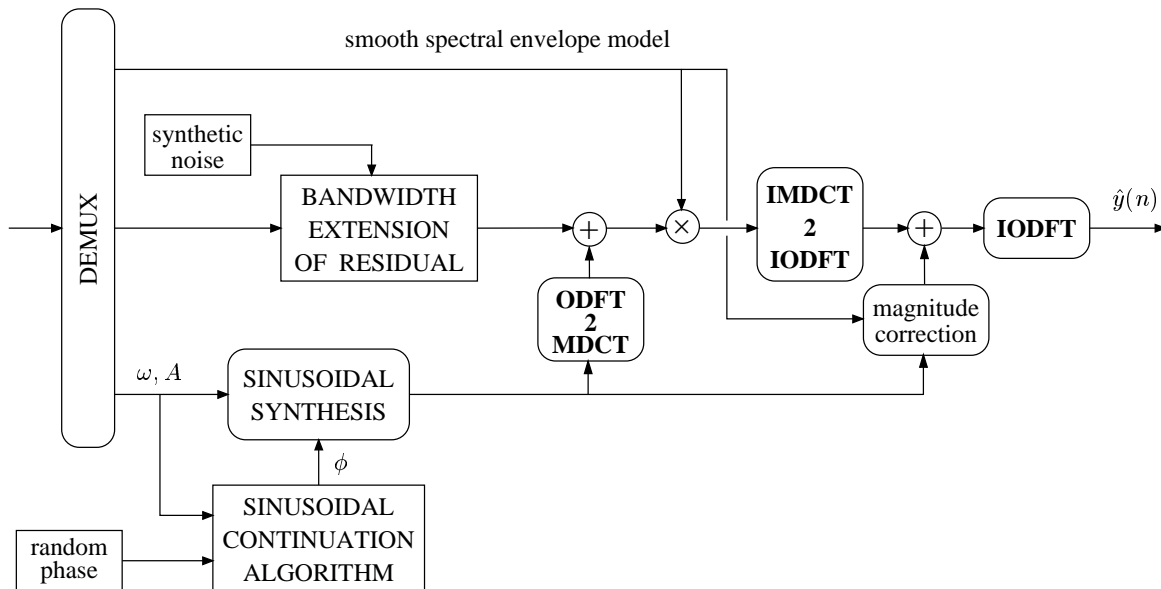


Fig. 3: A particular implementation of the ASR technique (decoder).

within the bandwidth of the transmitted MDCT residual, are also transmitted to the decoder. The important advantage in this case is that perfect reconstruction is achieved if the MDCT residual is not quantized, despite the fact that all side information (*i.e.*, the spectral envelope model, phases, magnitudes and frequencies of all sinusoids) is quantized. This advantage is important because it allows to shape the overall coding noise as desired by means of a suitable quantization of the MDCT residual. Although the ASR decoder structure, as illustrated in Fig. 3, permits the realization of this scenario, we have enforced in this paper a tougher operational mode by not transmitting any phase information. Phase is artificially generated by means of a sinusoidal continuation algorithm that picks a random phase at sinusoidal birth and insures phase coherence (in the frequency domain) as long as the sinusoidal trajectory remains alive.

out-of-band synthesis of sinusoids: the sinusoidal components that fall outside the spectral region of the transmitted MDCT residual, can be synthesized directly in the ODFT domain,

avoiding the ODFT2MDCT transformation as well as the associated TDAC mechanism, and avoiding the denormalization by the smooth spectral envelope model, which is replaced by a simple magnitude correction. Therefore, as illustrated in Fig. 3, it is sufficient to take advantage of the sinusoidal continuation algorithm in order to generate a correct sinusoidal trajectory using only the transmitted frequency and magnitude parameters. In the limit, if the magnitude, frequency and phase parameters are transmitted, the sinusoidal continuation algorithm is not needed. The opposite situation is also possible: neither the magnitude, frequency or phase parameters are transmitted and the bandwidth extension of sinusoidal components uses only the information of the transmitted low-frequency sinusoids and smooth spectral envelope model, in order to synthesize a correct sinusoidal trajectory. The significant advantage in this case (which makes our solution competitive) is that sub-bin frequency accuracy can be achieved which means that an existing harmonic structure of sinusoids will precisely replicate the harmonic organization avoiding perceptual distortions due to

harmonic misalignment which is common in other bandwidth extension techniques.

spectral unflattening: the spectral denormalization by the smooth spectral envelope model is highly flexible since it allows the independent control of the MDCT residual and of the sinusoidal components. This paves the way for example to the implementation of special effects in audio such as enhancing the timbre of a music signal or suppressing the singing voice from a song.

4. PERFORMANCE EVALUATION

The ASR bandwidth extension technique has been implemented according to the algorithms depicted in Figures 2 and 3. Audio signals are sampled at 44100 samples/second. The size of the ODFT/MDCT transform is 1024 samples, *i.e.*, the number of subbands of the analysis/synthesis filter bank is 512. The bandwidth of the preserved MDCT is only 6 kHz, the frequencies and magnitudes of sinusoids are transmitted to the decoder, but not the phases.

It should be noted that in an actual sinusoidal coder, typically phases are coded using 6 bits per sinusoid [14], which represents a significant fraction of all side information. Thus, by suppressing the transmission of phase in our experiments, we are evaluating to which extent important savings in side information can be made without incurring in a significant penalty on the subjective audio quality.

4.1. Results with a Synthetic Signal

We have created a synthetic signal resulting by frequency modulation (FM) of a single sinusoid. The corresponding Matlab code is displayed next.

```
samples = 512*100;
temp=5000*sin(2*pi*25.1*[0:samples-1]/1024 +
0.01*1024*sin(2*pi*[0:samples-1]/10/1024));
fidw = fopen('testfile.pcm','w');
fwrite(fidw, temp, 'short');
fclose(fidw);
```

The spectrogram of the FM signal is depicted in Fig 4. This figure reveals, as can be concluded from the previous Matlab code, that the center frequency of the FM signal is about 1081 Hz, and that the maximum frequency deviation is about 50 Hz. It should

be noted that the natural frequency resolution of the analysis/synthesis filter bank is $44100/1024 \approx 43$ Hz.

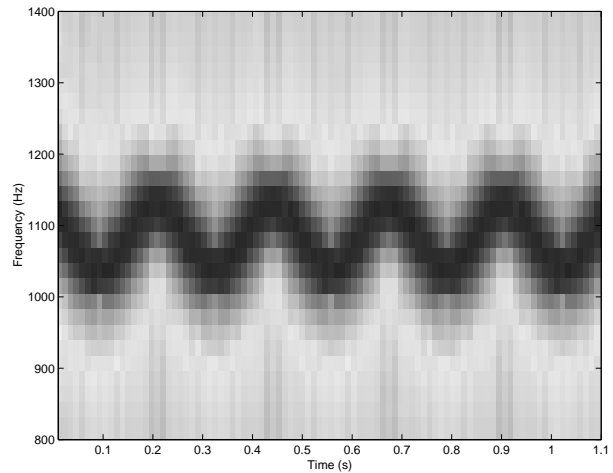


Fig. 4: Spectrogram of an FM modulated test sinusoid at the *input* of the ASR algorithm.

Synthetic test signals are very useful because they facilitate the design and performance optimization and verification of some of the main block modules of the ASR encoder and decoder, namely the sinusoidal analysis, the sinusoidal synthesis and the sinusoidal continuation algorithm.

One particularly interesting result in the case of the FM synthetic signal regards the output of the sinusoidal continuation algorithm, which is represented in Fig. 5. In this figure, a star represent the instantaneous frequency estimated by the sinusoidal analysis block at each frame, and the solid lines connect those stars for which the sinusoidal continuation algorithm has decided phase coherence should be enforced across adjacent frames.

The synthesis of the FM modulated signal at the ASR decoder has been performed in the ODFT domain (this has been achieved ignoring the transmission of the MDCT residual) and the corresponding spectrogram is depicted in Fig. 6. Both input and output audio files are available at <http://www.atc-labs.com/asr/>.

Comparing both input and output audio files, either objectively (and taking into consideration that the

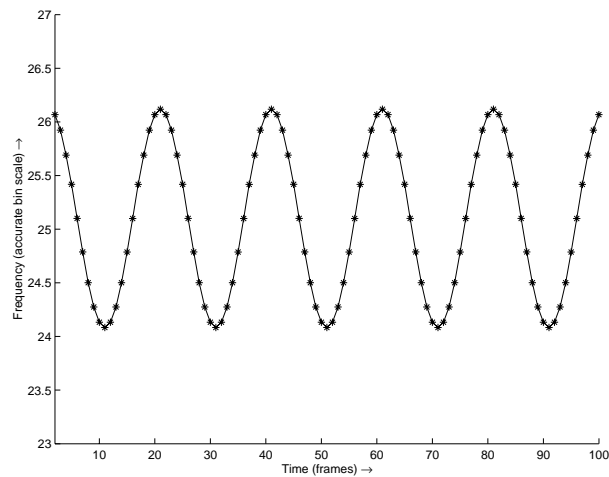


Fig. 5: Instantaneous frequency estimation regarding an FM modulated sinusoid.

phase of the output file is totally artificial), or subjectively, leads to the conclusion that the ASR analysis/synthesis process is not only correct from the point of view of signal integrity, but is also of very high quality since no noticeable artifacts are heard.

4.2. Results with a Natural Music Signal: *pitchpipe*

The complete ASR encoder and decoder has been tested using a short segment of a natural music signal: *pitchpipe solo*. It consists of a music signal with a well-defined harmonic structure. In order to have a reference of the operation and impact of a known bandwidth extension technique such as Spectral Band Replication (SBR) [15], the short *pitchpipe* music file has been coded with MP3Pro at 64 kbps, using the encoder/player available at <http://www.mp3prozone.com>. Fig. 7 represents the short-time Power Spectral Density (PSD) of the original signal and that of the MP3Pro encoded signal. It can be easily concluded that the SBR technique has implemented bandwidth extension for frequencies above 8 kHz. Ignoring the magnitude differences between some of the partials of the original audio file and those of the bandwidth extended signal, it can be seen that the SBR technique has introduced two sources of incorrectness:

1. there is an obvious frequency misalignment between the original partials and those replicated

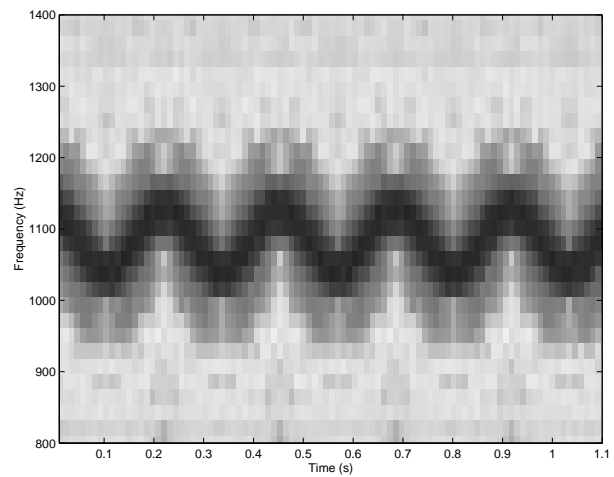


Fig. 6: Spectrogram of an FM modulated test sinusoid at the *output* of the ASR algorithm.

by the SBR technique,

2. the noise floor of the replicated signal is substantially different from that of the original signal.

These two reasons lead to the introduction of audible artifacts in the band replicated audio file.

The same original file has been processed using the ASR encoder and decoder. It should be noted that in this case:

- the bandwidth of the preserved MDCT residual is 6 kHz,
- *all* partials are synthesized at the decoder using *only* the frequency and magnitude parameters characterizing the original partials,
- phase information is discarded at the encoder.

The short-time PSD of both input and output audio files is represented in Fig. 8. It can be concluded from this figure that there is a perfect frequency alignment between the original partials and the partials that have been synthesized by the ASR technique, and that the replaced noise floor follows closely the original noise floor. The fact that there is also a perfect match between the magnitudes of the original partials and those of the replaced partials is

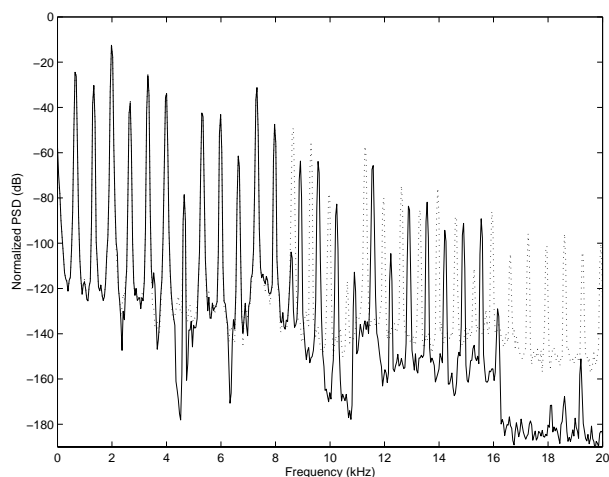


Fig. 7: Short-time PSD of the original *pitchpipe* audio excerpt (dotted line) and of the resulting MP3+SBR encoded audio at 64 kbit/s (solid line).

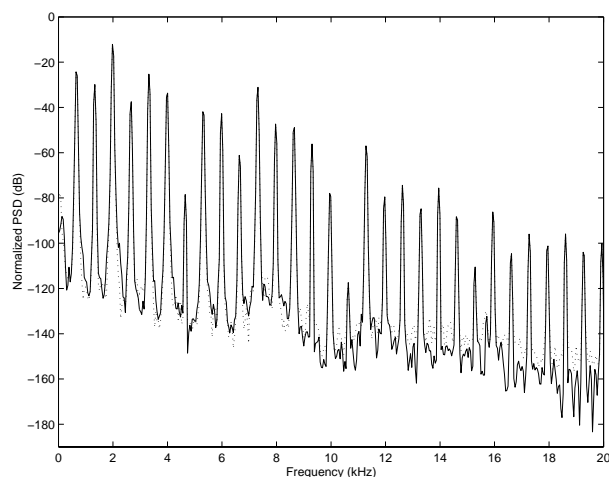


Fig. 8: Short-time PSD of the original *pitchpipe* audio excerpt (dotted line) and of the resulting ASR processed audio (solid line).

due to the fact that we are using the exact original magnitudes but we are not claiming in this regard any advantage of the ASR technique over the SBR technique.

Both SBR (MP3Pro) and ASR encoded audio files are available on-line at <http://www.atc-labs.com/asr/>. Contrarily to the SBR encoded audio file, the ASR processed audio file does not exhibit audible distortions due to frequency misalignment.

4.3. Results with a Natural Music Signal: *singing voice*

The natural audio signal considered previously was not especially difficult to the ASR technique because its pitch is fairly stable. More difficult signals are those for which there is a vibrato effect since this introduces a higher stress on the sinusoidal continuation algorithm. In order to illustrate the operation of the ASR technique under such difficult conditions, we have used as natural audio signal a short segment of a female singing voice whose spectrogram is represented in Fig. 9. This figure highlights a clear vibrato effect particularly evident in the time region between 0.4 sec. and 0.7 sec. The operation of the sinusoidal continuation algorithm is reflected on the results depicted in Fig. 10. As explained previously, stars represent instantaneous frequencies of sinusoidal components at each time frame, and con-

nected starts denote sinusoidal trajectories in which phase continuity and coherence has been enforced (in the discrete frequency domain) by the ASR decoder in the synthesis process. Since phases are not transmitted, *all* partials are synthesized using artificial phase. The noise floor is obtained as a combination of bandwidth extension from the 6 kHz bandwidth MDCT residual, and synthetic noise. The spectrogram of the output file is depicted in Fig. 11 and essentially reveals that the vibrato effect has been preserved in the output file and that sinusoids have been correctly replaced.

Both input and output files are available at <http://www.atc-labs.com/asr/>. It is quite a good surprise that the output file sounds with an excellent quality despite the fact that the original phases are discarded. This suggests that the ASR technique may also find application in speech modification or coding.

5. CONCLUSION

We have presented a new bandwidth extension technique that operates directly on the high resolution ODFT/MDCT frequency domain. The concept and operation of the technique has been described and its performance has been illustrated using both synthetic and natural audio signals. The main advantages of the technique are structural since no addi-

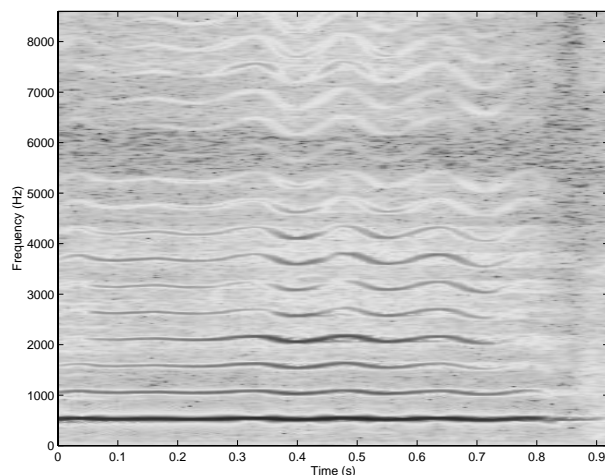


Fig. 9: Spectrogram of a short excerpt of a female singing audio signal.

tional filter banks are required besides the MDCT, and also operational since high-quality signal replacement is achieved for both coherent and incoherent components of the audio signal. The main application areas foreseen for ASR are low-delay, low-complexity and low bit rate high-quality audio communication¹, speech coding and transformation, and special effects in audio, including multichannel audio effects.

6. REFERENCES

- [1] Kyoya Tsutsui, Hiroshi Suzuki, Mito Sonohara Osamu Shimoyoshi, 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.
- [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] Mark Davis, "The AC-3 Multichannel Coder," *95th Convention of the Audio Engineering Society*, October 1993, Preprint n. 3774.

¹In addition to audio streaming and downloading, where time delay is not critical.

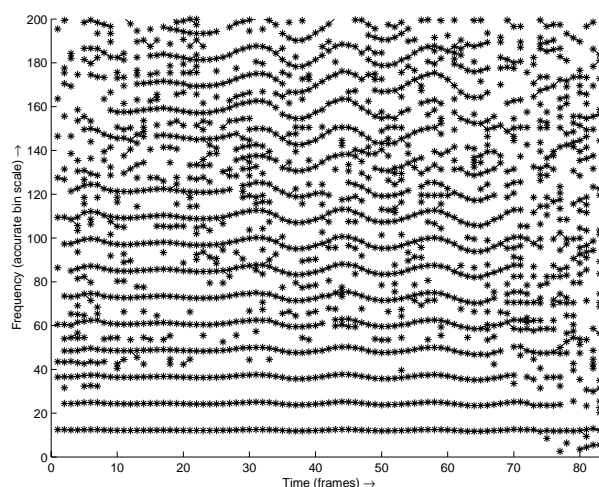


Fig. 10: Output of the ASR sinusoidal continuation algorithm: stars represent instantaneous frequencies and solid lines identify sinusoidal trajectories.

- [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] 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.
- [7] Aníbal 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.
- [8] William M. Hartmann, "Pitch, periodicity, and auditory organization," *Journal of the Acoustical Society of America*, vol. 100, no. 6, pp. 3491–3502, December 1996.
- [9] Juan G. Roederer, *The Physics and Psychophysics of Music-An Introduction*, Springer-Verlag, 1995.

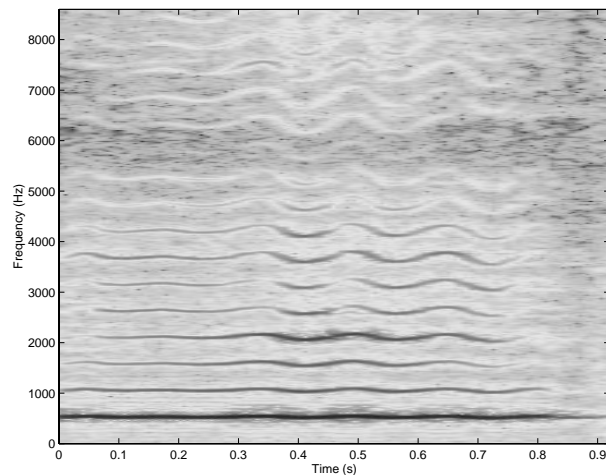


Fig. 11: Spectrogram of the ASR processed short excerpt of a female singing.

in audio coding,” *112th Convention of the Audio Engineering Society*, May 2002, Paper 5553.

- [10] Brian C. J. Moore, *An Introduction to the Psychology of Hearing*, Academic Press, 1989.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] Anibal J. S. Ferreira, “Perceptual Coding Using Sinusoidal Modeling in the MDCT Domain,” *112th Convention of the Audio Engineering Society*, May 2002, Paper 5569.
- [15] M Dietz, L. Liljeryd, K. Kjorling, and O. Kunz, “Spectral Band Replication, a novel approach