INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11
MPEG2010/M17914
July 2010, Geneva, Switzerland

Source **Telcordia Poland Sp. z o. o.**
Poznan University of Technology, **Chair of Multimedia Telecommunications and Microelectronics**
Status **Contribution**
Title **Improved coding of tonal components in audio techniques utilizing the SBR tool**
Author Tomasz Żernicki **<tzernick@telcordia.com>** [1]
Maciej Bartkowiak **<mbartkow@multimedia.edu.pl>**
Marek Domański **<domanski@et.put.poznan.pl>**

1 Introduction

This document presents a new tool of efficient regeneration of high-frequency tonal components in an augmented MPEG-4 AAC HE [1] decoder. The basic idea is to synthesize tonal components using the technique called synthetic sinusoidal coding. This technique is already adopted in the MPEG-4 codecs in another context. Here, the idea is to mix this technique with standard Spectral Band Replication (SBR) [3,4], i.e. to add some control information to a standard MPEG-4 AAC HE bit-stream that is used to synthesize the high-frequency tonal components in a decoder. In that way, provided is proper synthesis of rapidly changing sinusoids as well as proper harmonic structure in the high-frequency band. The experiments show that the tool significantly improves the compression performance when added to an MPEG-4 AAC HE codec. This improvement has been confirmed by listening tests.

2 Main idea

The technique consists in augmenting the SBR tool by another tool based on sinusoidal modeling that is used to synthesize high-frequency harmonic components in a decoder. The additional tool is used to process signal components above the f_{SBR} – the SBR cut-off frequency. The signal components below f_{SBR} are encoded using core encoder (as described in MPEG-4 AAC) while the technique of SBR augmented by parametric coding of sinusoids is used for the frequencies above f_{SBR}.
The main task of the proposed tool is to eliminate the coding distortions caused by the SBR tool when fast frequency changes occur at high frequencies. Additionally, the proposed technique allows to keep the harmonic structure of tonal components, i.e. to preserve that frequencies of harmonics are integer multiples of the fundamental frequency from the low-frequency band.

Proposed tool uses the process of selection and separation of HF tonal components from the original signal x. As a result, we obtain parametric representation of HF tonal components which

[1] This research has been done during author work at the Poznań University of Technology.

are encoded using sinusoidal codec. Additionally, signal *y* with removed HF tonal components is encoded using the standard MPEG-4 AAC HE encoder (including SBR tool). However, signal *y* can still contain HF tonal components.

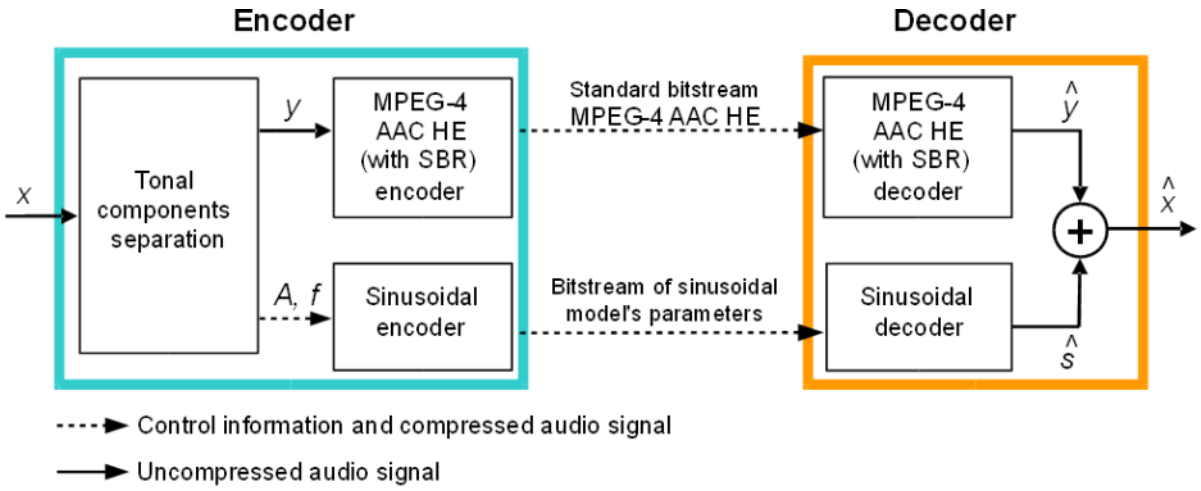


Figure 1. Idea of the proposed approach.

The additional tool of sinusoidal modeling is implemented by two additional blocks, one in an encoder and the other in a decoder. These new blocks are linked by an additional bitstream of parameters of sinusoids (Fig. 1). Application of this tool does not affect the standard MPEG-4 AAC HE bitstream that is augmented only by the above mentioned stream of parameters.

In the encoder (Fig. 2), the additional blocks identifies harmonic components in the spectrum above f_{SBR} and tracks them. Therefore, the parameters of sinusoids (A – amplitude, ω – radial frequency) are estimated in short windows in order to track their fast changes.

The parameters of all detected sinusoids (for the k-th sinusoid: A_k, ω_k) are compressed for consecutive sample blocks. The compressed parameters are transmitted to the decoder where they are used to synthesize the sinusoids (Fig. 2).

The sinusoidal synthesis is also performed in the encoder – the synthesized signal s has to be subtracted from the input to the AAC encoder with SBR (Fig. 2). In fact, the synthesized sinusoids s are used as an attenuation signal that attenuates the HF tonal components present in the input signal x (Fig. 1).

Proposed tool was tested with MPEG-4 AAC HE but can also be used with other bandwidth extension techniques including enhanced Spectral Band Replication (eSBR) tool, which is part of Unified Speech and Audio Coding (USAC) codec [2].

3 Coding of tonal components

Based on the idea presented on Figure 1 a new technique of high frequency bandwidth extension has been created (Figure 2). One of the main features of the proposed technique is application sinusoidal model [5,6] that represents a deterministic part of the signal as a sum of K quasi-sinusoidal components with time-variant parameters.

$$s_{\mathrm{det}}(t) = \sum_{k=1}^{K} A_k(t) \cos\left(\varphi_k + 2\pi \int_0^t f_k(\tau)\, d\tau \right), \qquad (1)$$

where $A_k(t)$, $\omega_k(t)$, $\varphi_k(t)$, k = 1, … , K denote time-variant amplitude, radial frequency and phase, respectively.

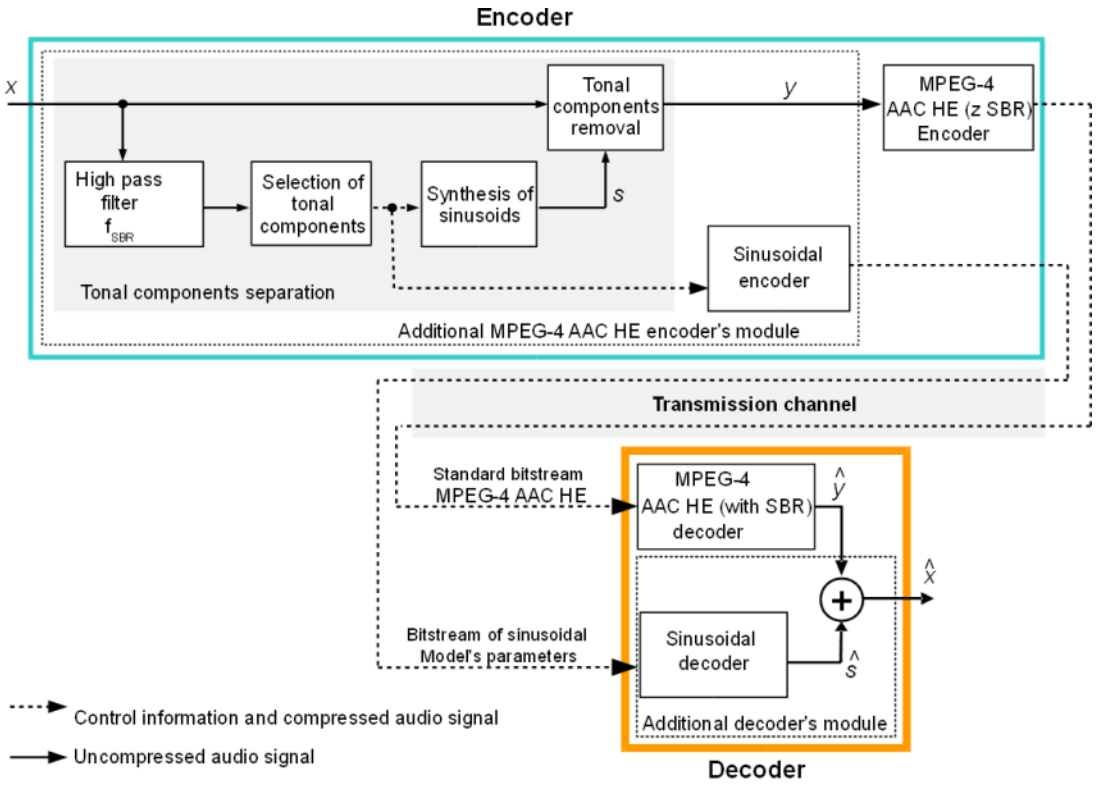


Figure 2. The modified MPEG-4 AAC HE codec with additional blocks related to sinusoidal modeling.

The parameters $A_k(t)$, $\omega_k(t)$, $\varphi_k(t)$ are estimated in the process of sinusoidal analysis of the input signal x. Such an analysis has been already described in many papers [5-8]. The major steps of the analysis are shown in Fig. 3. The analysis starts with Short-Time Fourier Transformation (STFT). Then the tonal components must be identified (spectral peak detection and classification) and measured in order to estimate the parameters. In fact, the time-variant parameters are estimated in consecutive overlapping windows that contain N samples. In our implementation, there is $N = 2048$, and the windows overlap by 512 samples.

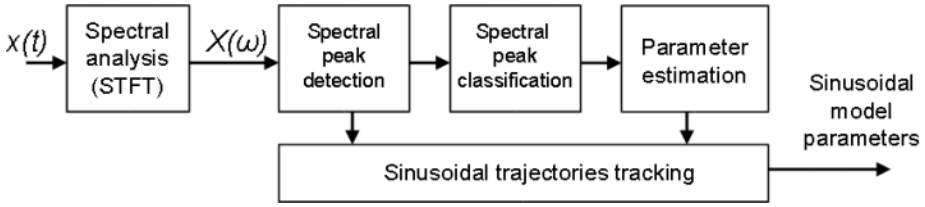


Figure 3. Sinusoidal analysis scheme.

In the next step, the sequences of parameters A_k, ω_k, φ_k are linked into sinusoidal trajectories that describe evolution of individual tones over time (Fig. 4). Tracking of sinusoidal trajectories has been already described in many papers, e.g. [5-8]. Here, in the implementation, a technique based on maximum likelihood is used.

Both in the encoder and in the decoder, the quasisinusoidal signals are synthesized from individual sinusoidal trajectories. For a given trajectory, values of frequency and amplitude are interpolated over time. In our implementation, the amplitude is linearly interpolated in the log (dB) domain, while the frequency is interpolated using cubic splines of degree 3.

In the encoder the signal *s* is synthesized. Such signal represents high-frequency harmonic components of the input audio signal *x*. The number of K depends on the accuracy of sinusoidal trajectory estimation. The encoder aims to find as much as possible sinusoidal components.

Next, the high-frequency tone components should be removed from the input to a classic MPEG-4 AAC HE encoder, which is done in tonal component removal block (Fig. 2). In our codec, such removal is modeled by attenuation of the high-frequency tonal components in the input signal *x*, but also other techniques can be used in that context [14]. This attenuation is implemented as Short Time Spectral Attenuation (STSA) [15] by synthetic signal *s*.

$$Y_m(k) = \frac{X_m(k)}{\min\left(\frac{|S_m(k)|}{\varepsilon}, 1\right) * h(k)}, \qquad h(k) = \begin{cases} 0, & |k| > R \\ 1, & |k| \le R \end{cases} \qquad (2)$$

where:

$Y_m(k)$	– spectrum of the signal *y* with attenuated HF tonal components,
$S_m(k)$	– spectrum of attenuation (synthetic) signal *s*,
M	– index of analyzed frame,
ε	– attenuation coefficient,
$R = 4$	– coefficient obtained through experimental tests.

The obtained signal *y* is encoded using the standard MPEG-4 AAC HE encoder, i.e. an encoder with SBR tool. In high frequencies, the signal *y* contains mainly noise components, which are less important, from the perceptual point of view, than tonal components. The noise-like components are very effectively encoded using the spectral band replication (SBR) technique. It needs to be outlined that SBR encoder does not transmit any information related to tonal components, e.g. scaling factors and parameters used for describing synthetic sinusoids. In this way the bitrate of SBR data stream is reduced.

The parameters of the sinusoidal model have to be transmitted to the decoder, therefore they also need to be compressed. In our implementation, uniform quantization, linear predictive coding (LPC) and Huffman coding has been used for that purpose. For the parameters of the sinusoidal model, Burg method of linear prediction is used [10]. The order of the predictor is 6 and there are 20 previous samples used to estimate the predictor coefficients. It should be noted that encoder does not transmit additional information about phase. Phase spectrum of high frequency components is less important, from the perceptual point of view and can be omitted. In this way the bitrate of additional sinusoidal model parameters data stream is reduced.

The bit-rate of the sinusoidal model's parameters (B_{sin}) is about 2 kb/s. It means that the bit-rate of MPEG-4 AAC HE ($B_{AAC\text{-}HE}$) must be reduced to obtain the total bit-rate (B_T):

$$B_T = B_{AAC-HE} + B_{sin}. \qquad (3)$$

In our implementation we have used rate-distortion loop where the B_{sin} was controlled by removing sinusoidal trajectories of the lowest energies.

Proposed technique improves reconstruction of the high-frequency tonal components as compared to the standard MPEG-4 AAC HE coding with the SBR tool (see Figs. 4). For example, a violin produces high tones that are lost at an output of a standard MPEG-4 AAC HE decoder while being well reconstructed by the decoder proposed in this paper.

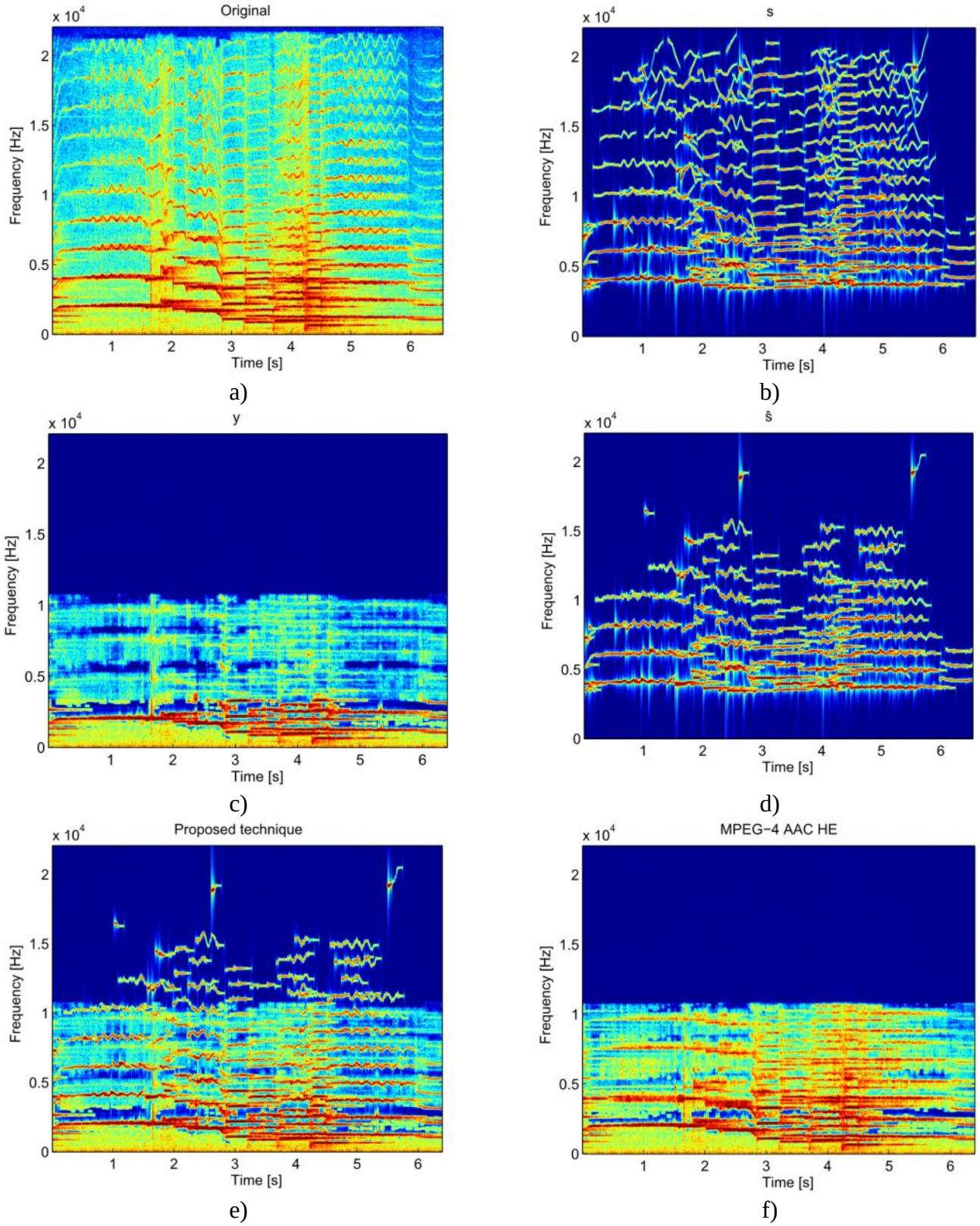


Figure 4. Exemplary spectrograms for signal *violin2* and target bit-rate of 16 kb/s, from the top: a) input signal, b) synthesized sinusoids from the encoder side – signal *s*, c) attenuated signal *y* obtained from the spectral attenuation block, d) synthesized sinusoids from the decoder side – signal $\hat{s}$, e) full reconstructed signal in the decoder – proposed technique, f) full reconstructed signal in the decoder – MPEG-4 AAC HE. Please note that the SBR codec operates up to only 12 kHz at this low bitrate.

4 Experiments

In order to assess the efficiency of the technique, full encoder and decoder have been implemented as software that processes audio excerpts off-line. The experimental software was built on top of the 3GPP codec that is compliant with the MPEG-4 Audio standard, i.e. AAC High Efficiency Profile (with SBR). In the experimental software, the proposed additional blocks have been implemented by routines written in Matlab.

As described before, the new technique switches on when the audio analysis procedures identify strong tones above the SBR cutoff frequency f_{SBR}. In the absence of such tonal components in audio signal, the proposed technique has no impact on the efficiency of an MPEG-4 AAC HE audio codec. Therefore, the detailed listening tests have been performed for records of instruments with strong tonal elements in higher frequencies, e.g. the excerpts of violin, accordion, and trumpet (Table 1) the EBU SQAM Compact Disk [11], the University of Iowa online database [12], MPEG [13] and other sources.

Table 1. Description of audio excerpts used during the experiments.

Signal name	Description	Duration
accordion	single instrument	9,3 s
altosax	single instrument	8,4 s
glissando	single instrument	4,6 s
violin2	single instrument	5,6 s
discofunk	music	7,2 s
philipspop	music	9,5 s
strings	music	10,4 s

In the experimental implementation, encoded parameters of high-frequency sinusoidal trajectories needed about 2 kbps. The average number of encoded sinusoidal trajectories per frame is about 20. This is a sufficient number of trajectories to get higher audio quality than from MPEG-4 AAC HE (with SBR) at the same total bit-rate.

The main objective of the listening tests is to compare the compression efficiency of the two audio codecs: the original MPEG-4 AAC HE (with SBR) and the same codec augmented with the tools proposed in this paper. The testing procedure was compliant with the ITU-R Recommendation BS.1534 [9] (MUSHRA – „Multi Stimulus test with Hidden Reference and Anchors"). This subjective quality methodology was chosen because it was developed for assessment of medium-quality audio sequences with significant distortions, e.g. resulting from compression. All subjective tests have been made for a group of 17 listeners who assessed 7 audio excerpts with sounds of musical instruments and music. Each excerpt was presented in 6 versions, i.e. decoded by the standard decoder and the proposed one, by 16, 20 and 24 kbps in both cases. The original excerpts as well as their lowpass-filtered versions have been also heard by the listeners. Two types of lowpass-filtered excerpts have been used: with bandwidth limited to 3.5 kHz and 7 kHz, respectively. The MUSHRA score is the statistical measure of subjective quality. During the tests we have used five grade scale where 5 means imperceptible difference with the original signal.

The results (Table 2 and Table 3) prove clearly that the proposed codec (HFR-SIN) outperforms significantly the standard MPEG-4 AAC HE codec for such musical sequences. Significant improvement was observed for all 7 excerpts and all bitrates (Table 2 and Figure 5a). Moreover, very good results are obtained for audio excerpts contains recording of single instrument which

has strong tonal components with varying frequencies in the high frequency part of the signal (Table 3 and Figure 5b). For that kind of signals subjective MUSHRA score is above 4.0 out of 5.0. For example for 16 kbps proposed technique obtained 4.23 of the MUSHRA score whereas MPEG-4 AAC HE got 2.79.

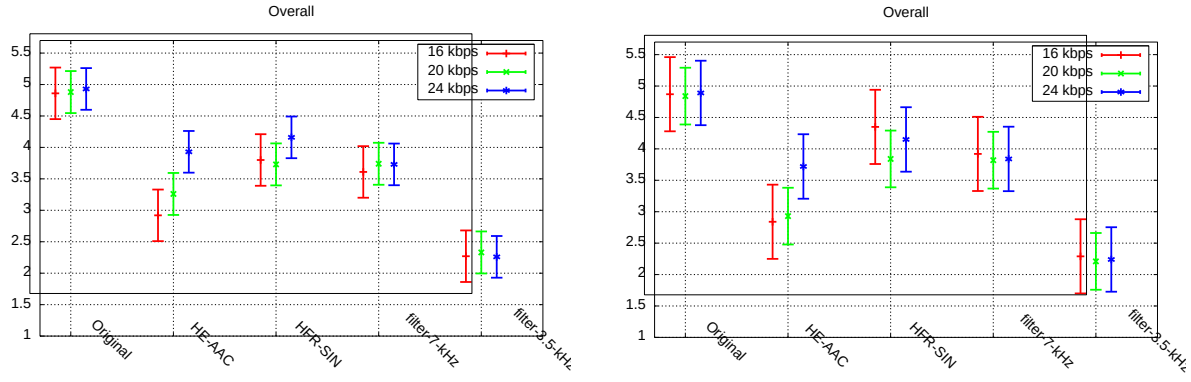
Figure 4 proves that the proposed technique is well suitable for signals with strong, stable and quickly varying sinusoidal trajectories where pure MPEG-4 AAC HE is unable to reconstruct properly those signal components.

Table 2. Results of subjective listening tests.

Signal	MUSHRA score [%]					
	16 kbps		20 kbps		24 kbps	
	MPEG-4 HE-AAC	Proposed technique HFR-SIN	MPEG-4 HE-AAC	Proposed technique HFR-SIN	MPEG-4 HE-AAC	Proposed technique HFR-SIN
accordion-ebu	2.68	3.83	3.18	3.60	3.68	4.03
altosax	3.38	3.93	2.78	3.64	4.11	4.02
discofunk	3.42	3.68	3.53	3.67	4.17	4.44
glissando1	2.78	4.57	3.33	3.79	3.69	3.97
philipspop	3.03	2.73	3.78	4.00	4.23	4.01
strings	2.81	3.27	3.54	3.33	4.29	4.21
violin2	2.35	4.57	2.66	4.08	3.35	4.46
AVERAGE	2.92	3.80	3.26	3.73	3.93	4.16

Table 3. Results of subjective listening tests for signals contains recording of single instrument.

Signal	MUSHRA score [%]					
	16 kbps		20 kbps		24 kbps	
	MPEG-4 HE-AAC	Proposed technique HFR-SIN	MPEG-4 HE-AAC	Proposed technique HFR-SIN	MPEG-4 HE-AAC	Proposed technique HFR-SIN
accordion-ebu	2.68	3.83	3.18	3.60	3.68	4.03
altosax	3.38	3.93	2.78	3.64	4.11	4.02
glissando1	2.78	4.57	3.33	3.79	3.69	3.97
violin2	2.35	4.57	2.66	4.08	3.35	4.46
AVERAGE	2,79	4,23	2,99	3,78	3,70	4,12

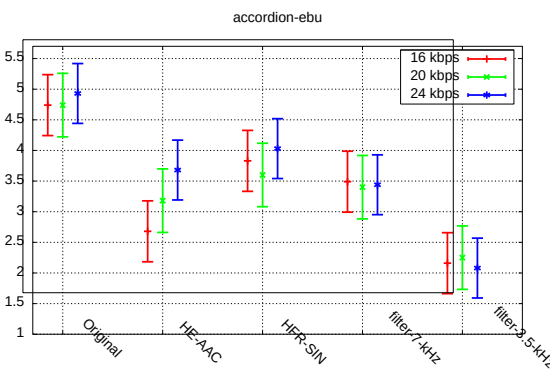


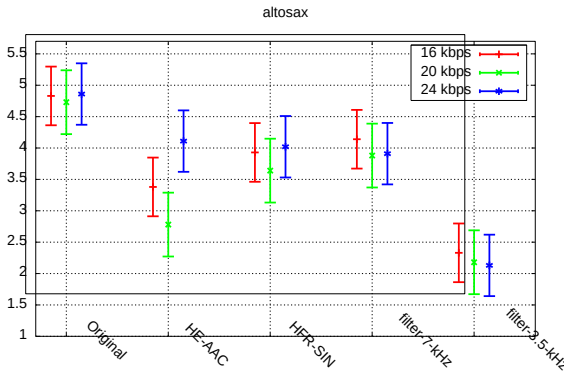
a) Comparison for all signals under test

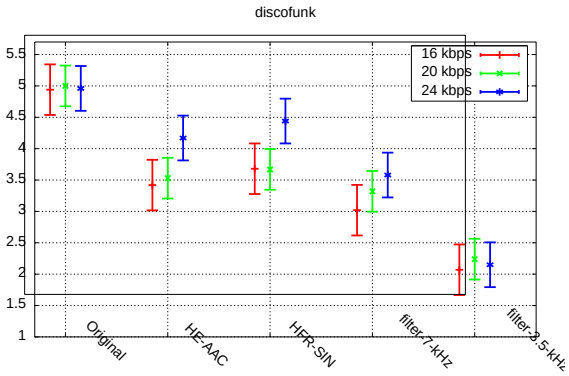
b) Recordings of single instruments.

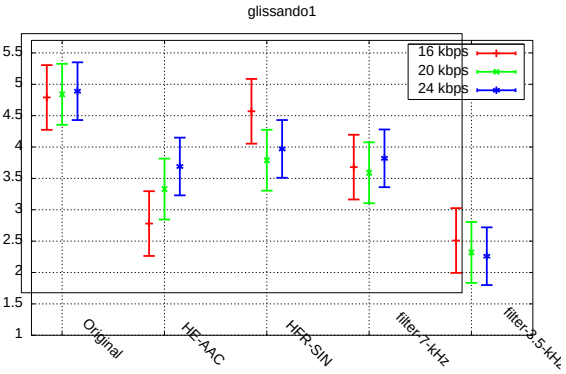
Figure 5. Average results of MUSHRA subjective listening tests for bitrates of 16, 20 and 24 kbps. The scores for the hidden reference (Original) and two anchors are given for comparison.

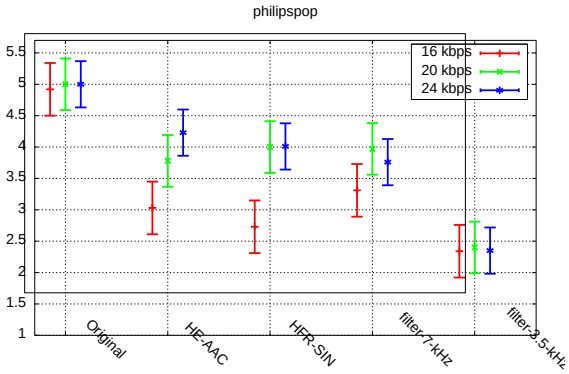
The 95 % confidence intervals (17 listeners) are plotted for all cases. HFR-SIN – proposed technique.

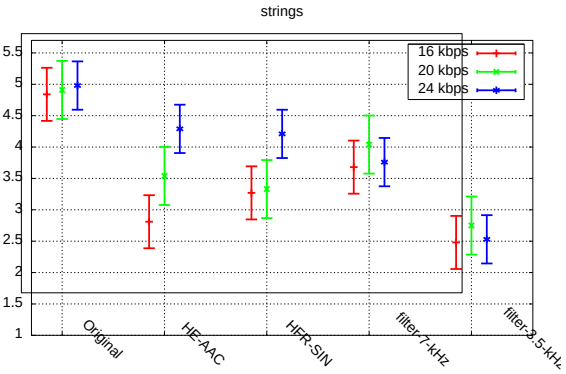












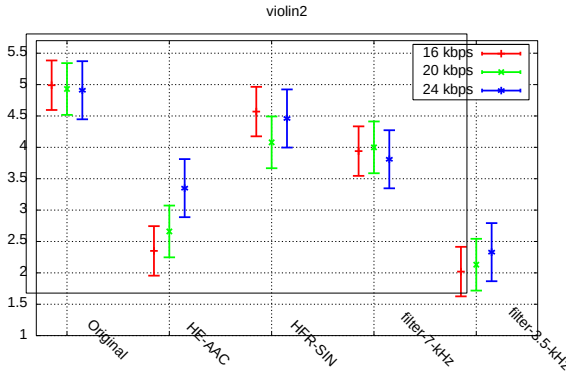
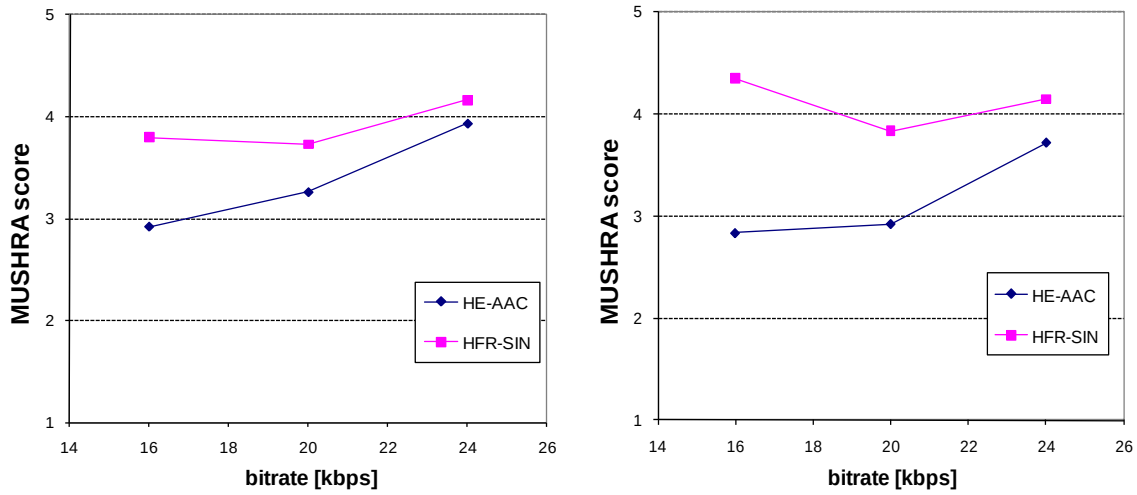


Figure 6. Results of MUSHRA subjective listening tests for all excerpts using during experiments. The 95 % confidence intervals (17 listeners) are plotted. HFR-SIN – proposed technique.



a) Comparison for all signals b) Single instrument

Figure 7. Rate distortion curves of tested codecs.

5 Conclusions

- We have proposed a new tool for efficient compression of tonal components for applications in low-bitrate coding of wideband audio.
- The improvement of the "rate-distortion" performance has been well proved using the standardized subjective listening tests.
- The proposed codec outperforms significantly the standard MPEG-4 AAC HE codec for bitrates below 24 kbps, especially for signals which has strong HF tonal components.

Recommendation: The technique is proposed as an additional tool for MPEG-4 AAC HE or USAC codecs in order to improve their performance in higher frequencies.

6 Acknowledgments

This work was supported by the research grant N N516 228 435 of the Polish Ministry of Science and Higher Education.

References

[1] ISO/IEC International Standard 14496-3: "Coding of Audio-Visual Objects – Part 3: Audio", 3rd Edition, 2005.

[2] ISO/IEC JTC1/SC29/WG11 MPEG/N11213, "Working draft 6 of Unified Speech and Audio Coding," Kyoto, Japan, January 2010.

[3] M. Dietz, L. Liljeryd, K. Kjörling, O. Kunz, "Spectral Band Replication, a novel approach in audio coding", *112th AES Convention*, Munich, May 2002.

[4] D. Homm, T. Ziegler, R. Weidner, R. Bohm, „Bandwidth extension of audio signals by spectral band replication", *Proc. 1st IEEE Benelux Workshop on MPCA*, Louvain 2002.

[5] R.J. McAulay, T.F. Quatieri, "Speech analysis/synthesis based on sinusoidal representation", *IEEE Trans. on ASSP*., vol 34, no. 4, 1986

[6] X. Serra, "Musical sound modeling with sinusoids plus noise", in C. Roads et al (eds) *Musical Signal Processing*, Sweets & Zeitlinger, 1997, pp. 91-122.

[7] M. Lagrange, S. Marchand, J. B. Rault, "Enhancing the Tracking of Partials for the Sinusoidal Modeling of Polyphonic Sounds", *IEEE Trans. Audio, Speech and Language Proc.*, Vol. 15, July, 2007.

[8] M. Bartkowiak, T. Żernicki, "Improved partial tracking technique for sinusoidal modeling of speech and audio", *Poznańskie Warsztaty Telekomunikacyjne - PWT'07, Poznań, Polska, 2007. [Online]. Available: http://www.multimedia.edu.pl/publications/*

[9] ITU-R, BS.1534, "Method for the subjective assessment of intermediate quality levels of coding systems," 2003.

[10] M. Lagrange, S. Marchand, M. Raspaud, J. B, Rault, "Enhanced partial tracking using linear prediction", *Digital Audio Effects (DAFX-03) Conference*, London, UK, 2003.

[11] European Broadcasting Union, "Sound Quality Assessment Material" compact disk, a part of an EBU publication Tech 3253.

[12] University of Iowa Electronic Music Studios, "Musical Instrument Samples Database," 1997. [Online].Available: http://theremin.music.uiowa.edu/

[13] ISO/IEC JTC1/SC29/WG11 MPEG w9099, "Final Spatial Audio Object Coding Evaluation Procedures and Criterion," Apr.2007.

[14] S.Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Transactionson Acoustics, Speech and Signal Processing, vol.27, no.2, pp. 113–120, 1979.

[15] O. Cappeand J. Laroche, "Evaluation of short-time spectral attenuation techniques for the restoration of musical recordings," IEEE Transactionson Speech and Audio Processing,vol. 3, no. 1, pp. 84–93, 1995.