SCALABLE PARAMETRIC AUDIO CODER USING SPARSE APPROXIMATION WITH FRAME-TO-FRAME PERCEPTUALLY OPTIMIZED WAVELET PACKETE BASED DICTIONARY

Convention paper 9264

Al. Petrovsky, V. Herasimovich, A. Petrovsky

Department of Computer Engineering,
Belarusian State University of Informatics and Radioelectronics,
Minsk, Belarus

1. Introduction

Presented work describes a new algorithm of parametric audio coding based on sparse approximation that used matching pursuit (MP) algorithm with optimized wavelet packet (WP) dictionary.

Main features are:

- High quality of reconstructed audio signal;
- Low speed rate of audio data transmission;
- Algorithm scalability to audio data coding;
- Universality for different nature of audio signals.

Main ideas of research are:

- Using MP algorithm as a core of encoder;
- Psychoacoustic optimized of time-frequency functions dictionary;
- WP – single transform domain.

2. MP Using WP Dictionary

Common MP procedure[1]

signal approximation ⟵ $$x(t) = \sum_{n=0}^{\infty} a_n g_{\gamma_n}(t)$$

window function ⟵ $$g_\gamma(t) = \frac{1}{\sqrt{s}} g\left(\frac{t-u}{s}\right) e^{i\xi t}$$

where s – scale, ξ – frequency modulation, u – translation.

WP based dictionary of time-frequency functions

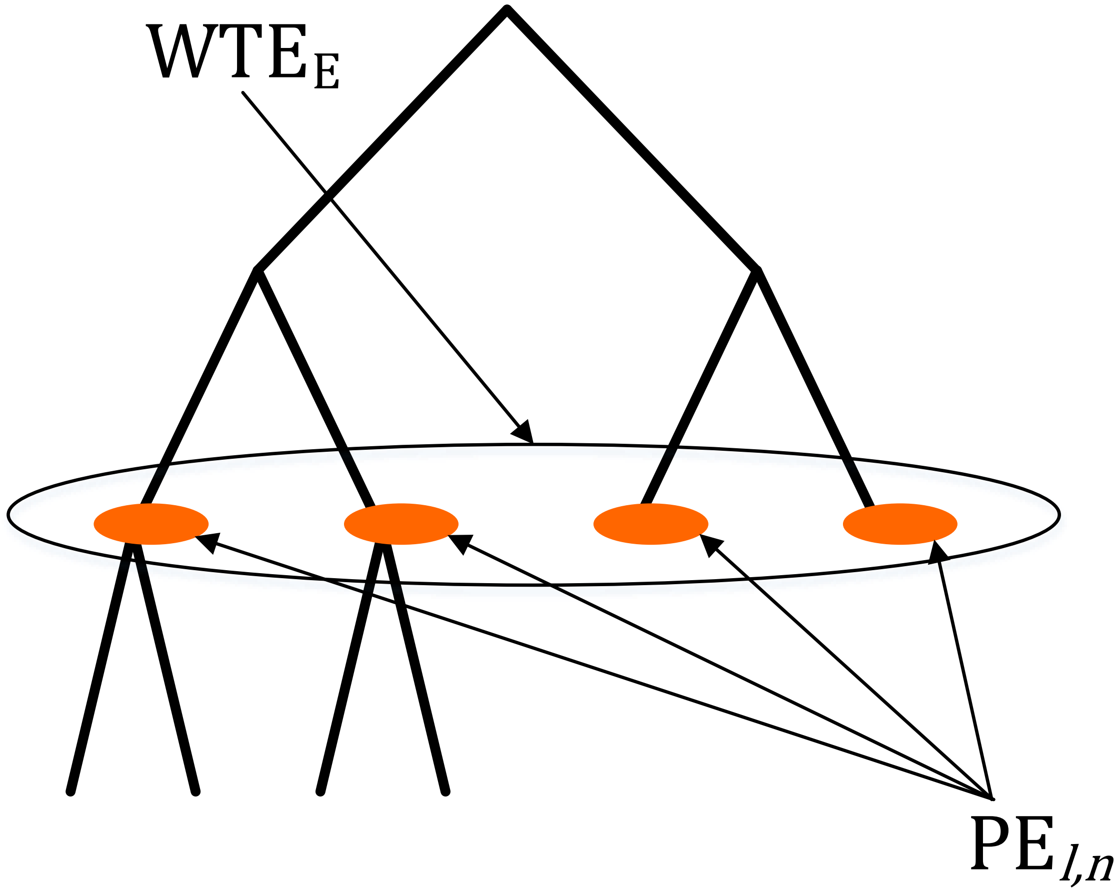
WP-based dictionary ⟵ $$g_\gamma \in D, \gamma = (l, n, k)$$

WP tree structure ⟵ $$E \in \{(l, n): 0 \le l \le L, 0 \le n \le 2^l\}$$

where l – WP tree level number, n – tree node number.

[1] S. Mallat, Z. Zang, "Matching Pursuits with Time-Frequency Dictionaries", IEEE Transactions on signal processing, vol. 41, pp. 3397-3415 (1993 December).

3. Adaptive WP Decomposition



Cost functions

Adaptation cost functions:

Wavelet time entropy estimation

$$WTE_{E_i} = -\sum_{\forall(l,n)\in E_i} \sum_{k} \frac{|X_{l,n,k}|}{\sum_{\forall(l,n)\in E_i}|X_{l,n,k}|} ln\left(\frac{|X_{l,n,k}|}{\sum_{\forall(l,n)\in E_i}|X_{l,n,k}|}\right)$$

Perceptual entropy estimation

$$PE_{l,n} = \sum_{k=1}^{K_{l,n}-1} \log_2(2[nint(SMR_{l,n,k})] + 1)$$

where $SMR_{l,n,k} = |X_{l,n,k}| \big/ \sqrt{12 \cdot T_{l,n}/K_{l,n}}$

Tree optimization procedure

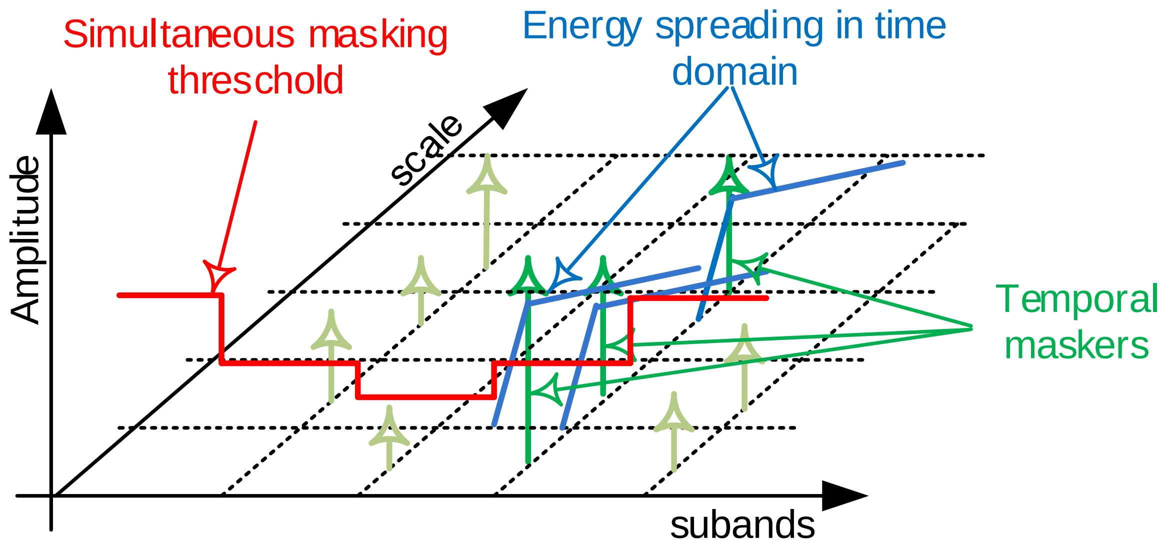
IF $WTE_{E_i} \leq WTE_{E_{i-1}}$ and $PE_{l,n} \geq PE_{l+1,2n} + PE_{l+1,2n+1}$,

THEN perform decomposition of the current level $l = l + 1$ and corresponding nodes $(l, n) \in E_j$ and transfer to new tree structure $E_j = E_{j+1}$,

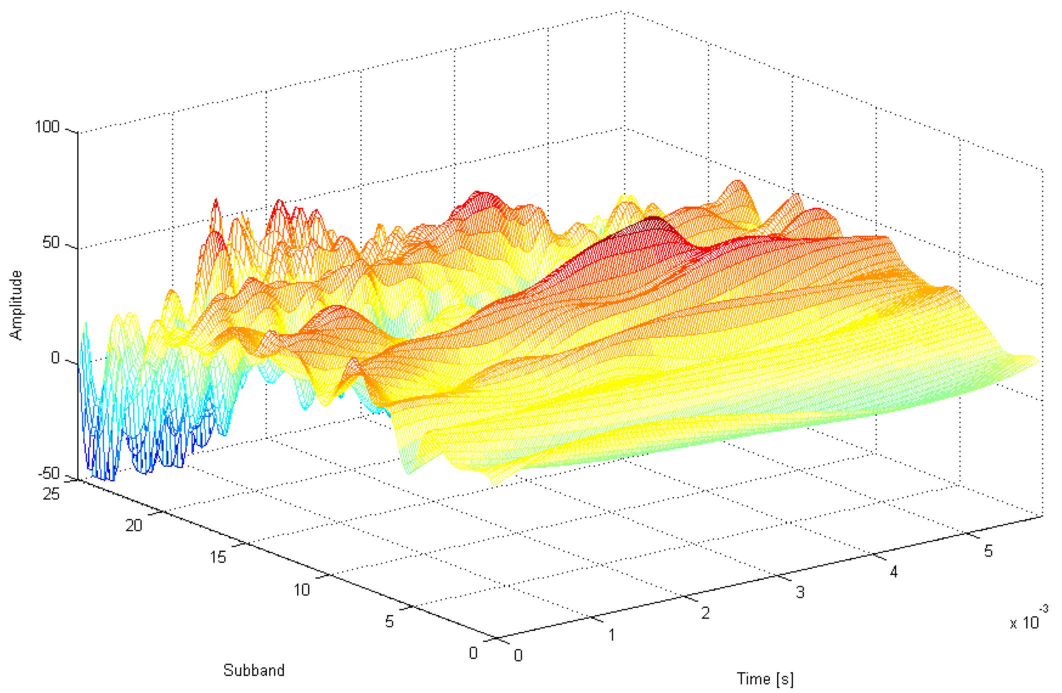
and repeat optimization procedure for next new tree structure E_j.

4. Excitation Scalogram Creation

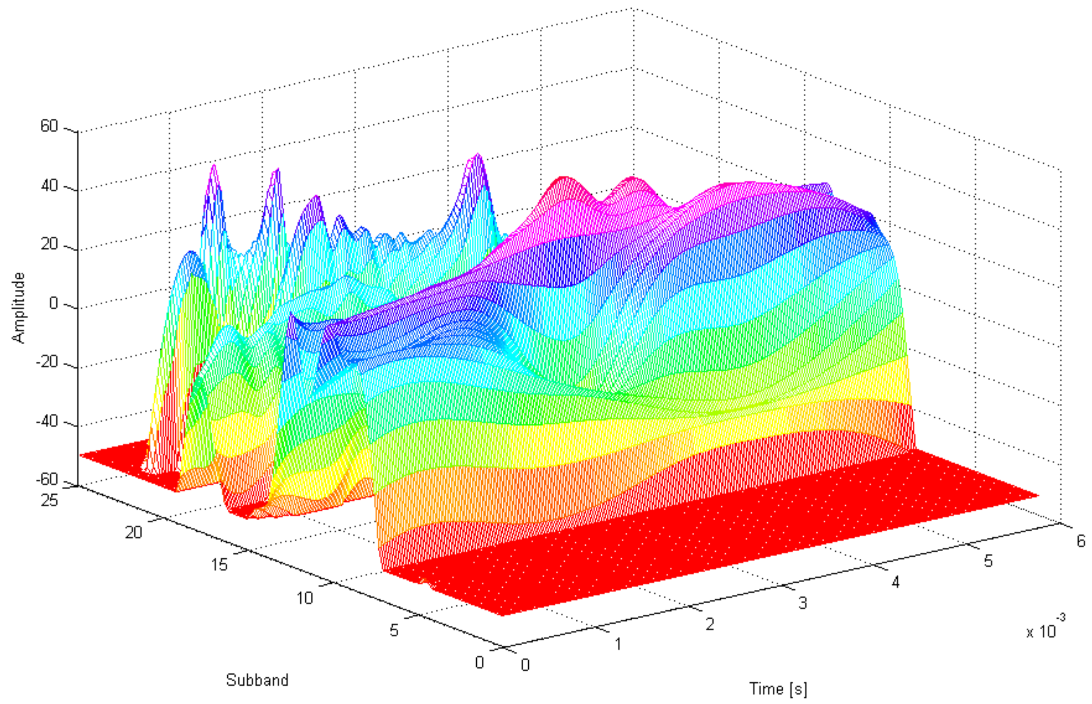
Masking thresholds[2] $T_{l,n}$ and temporal maskers[3] $F_{l,n}$ are used for excitation scalogram estimation;



Applying spreading energy functions in two domains



Excitation scalogram associated with original signal



Excitation scalogram associated with modeled signal with 5 atoms

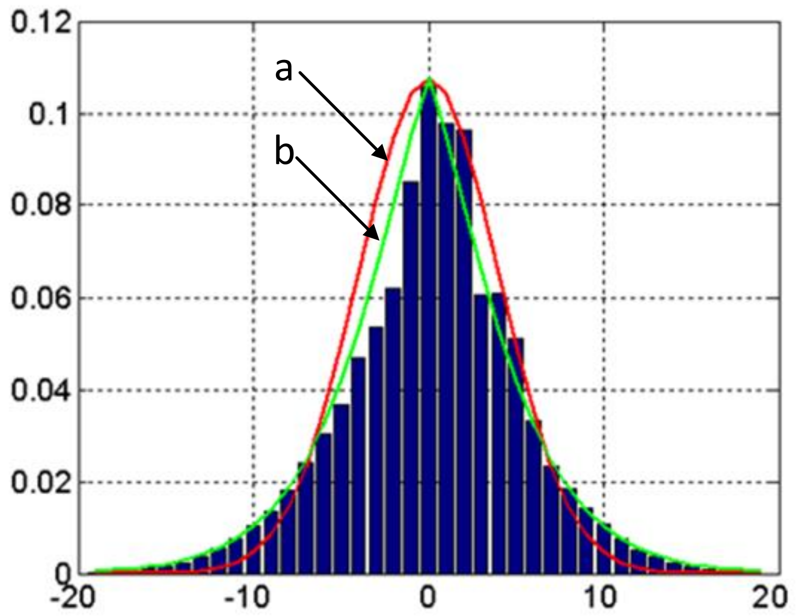
---

[2] A. Petrovsky, D. Krahe, A.A. Petrovsky, “Real-Time Wavelet Packet-based Low Bit Rate Audio Coding on a Dynamic Reconfigurable System”, presented at the AES 114th Convention, Amsterdam, The Netherlands, 2003 March 22-25.

[3] Al. Petrovsky, E. Azarov, A., Petrovsky, “Hybrid signal decomposition based on instantaneous harmonic parameters and perceptually motivated wavelet packets for scalable audio coding”, Elsiver, Signal Processing, Special Issue “Fourier Related Transforms for Non-Stationary Signals”, vol. 91, pp. 1489-1504 (2011, June).

5. Parameters Quantization & Coding

Example of wavelet coefficients histogram and Gauss (a) and Laplace (b) probability distribution functions.



Laplace function parameters

level, l	α	β
1	0.00001	0.12
2	0.0008	0.2
3	0.15	0.36
4	0.13	0.7
5	0.27	1.25
6	0.26	1.3
7	0.35	1.8
8	0.6	1.8

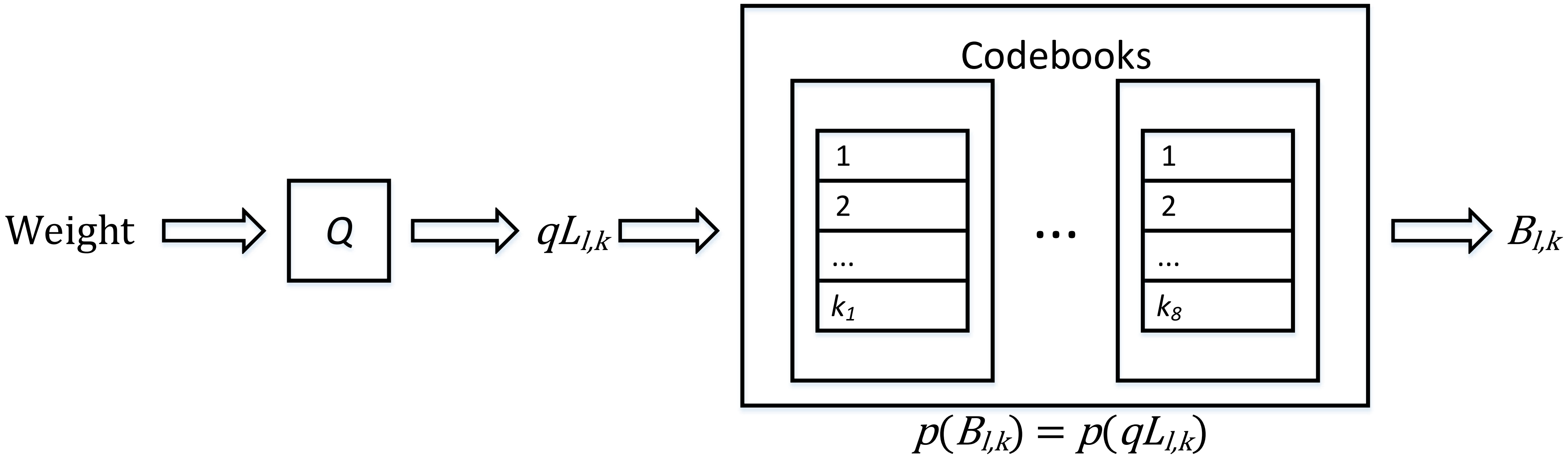
Weight quantization:

$$qL_{l,n,k} = 2\left| nint\left(\frac{|X_{l,n,k}|}{\Delta_{l,n}}\right)\right| + 1$$

$\Delta_{l,n} = \sqrt{12T_{l,n}/K_{l,n}}$ – quantization step

Quantized parameters coding:

$qL_{l,n,k}$ encoded using Huffman algorithm.



$$B_{l,k} = \left(b_{k,1}, b_{k,2}, b_{k,3}, \dots, b_{k,w_k}\right), b_{k,j} \in \{0,1\}, j = \overline{1, w_k}$$

6. WP Tree Structure Coding

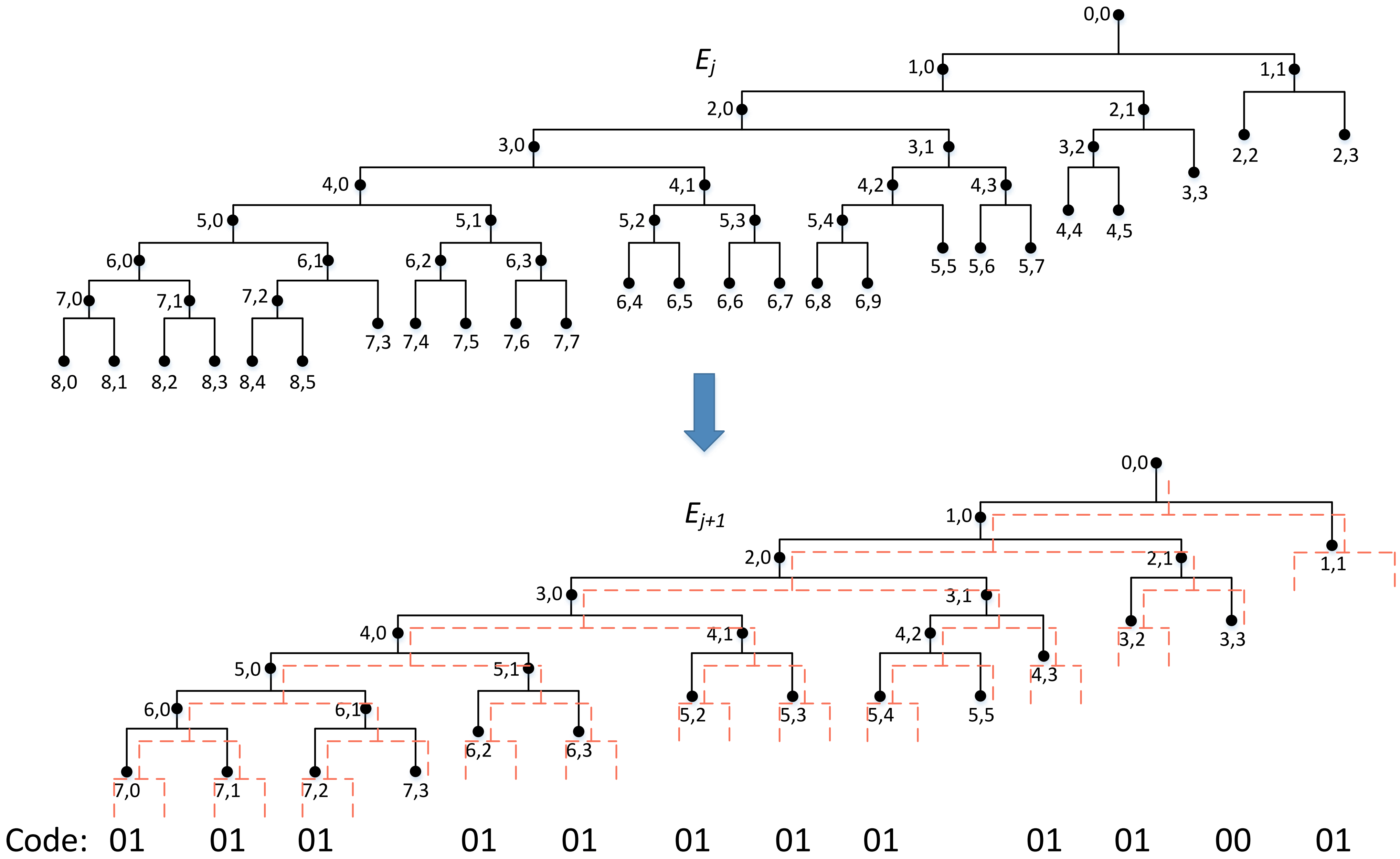
- 121 bits for straight *CB-WPD* tree coding – too many;
- Frame dependent WP tree structure coding;
- UP-DOWN one level WP tree structure grows.

#	Action	Code
1	no changes	00
2	delete node	01
3	node grows	10
4	double “no changes”	11

<u>Example:</u>

One level up of terminal nodes from E_j – *CB-WPD*, to E_{j+1}.

Up to 22 bits required.



7. An Objective Assessment[4] of the Audio Quality

	250 atoms	450 atoms	AAC
Bitrate, *kbps*	**45**	**80**	100
Compression rate	**15.6**	**8.8**	7.0

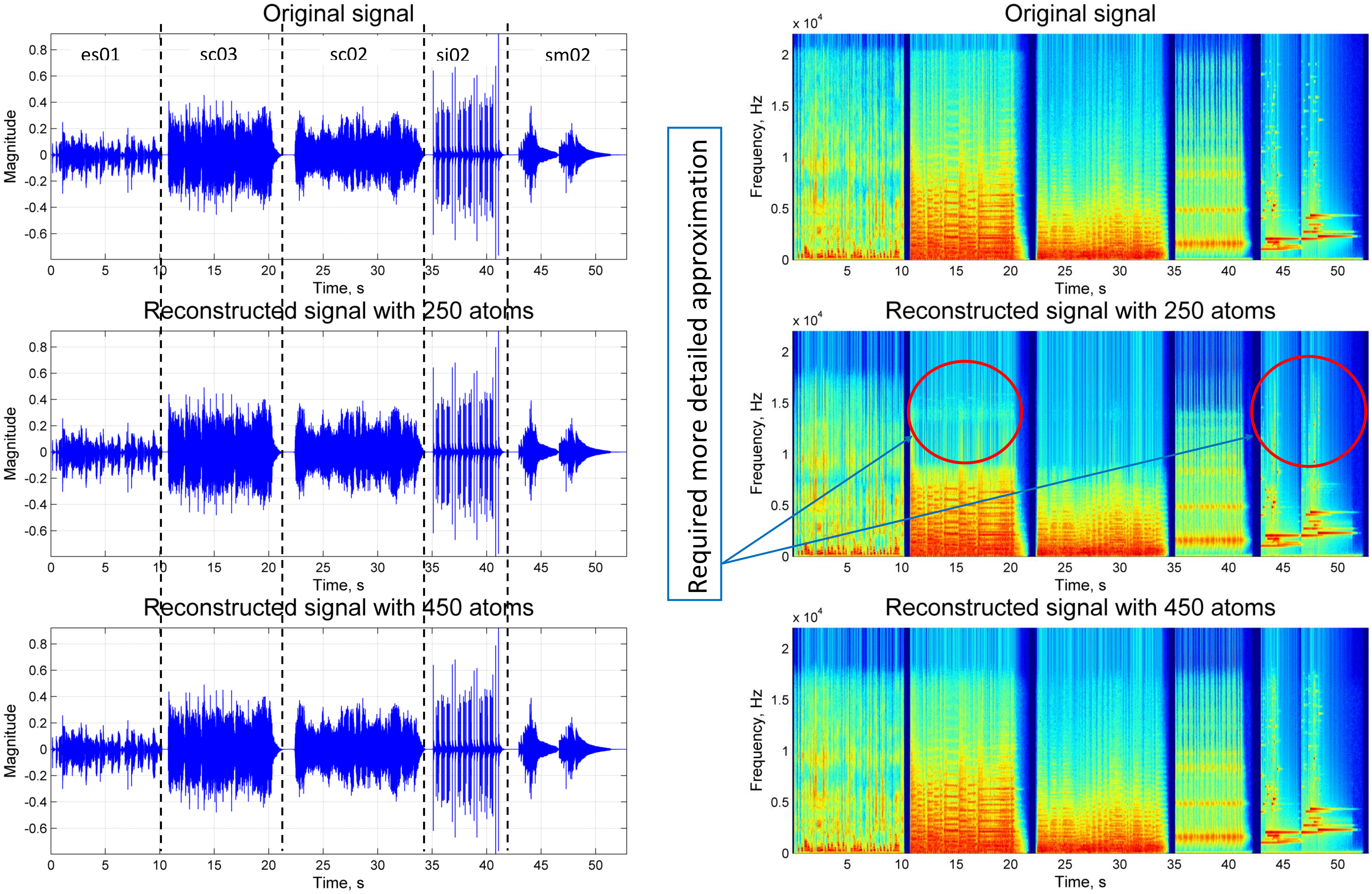
Impairment description	ODG
Imperceptible	0.0
Perceptible, but not annoying	-1.0
Slightly annoying	-2.0
Annoying	-3.0
Very annoying	-4.0

Description (44,1 *kHz*, 16 *bit*, mono)	**Proposed coder**					AAC
	250 atoms	300 atoms	350 atoms	400 atoms	450 atoms	
es01 – Vocal (Suzan Vega)	-2.1254	-1.8778	-1.4908	-1.1021	-0.8544	-0.218
es02 – German speech	-0.8315	-0.6183	-0.4641	-0.4494	-0.3854	-0.100
es03 – English speech	-1.5974	-1.4694	-1.3776	-0.8741	-0.5970	-0.132
sc01 – Trumpet solo and orchestra	-0.2116	-0.1738	-0.1706	-0.1656	-0.1624	-0.085
sc02 – Orchestra piece	-0.8659	-0.7642	-0.3608	-0.2509	-0.2165	-0.154
sc03 – Contemporary pop music	-2.5894	-1.6039	-0.8315	-0.3985	-0.3001	-0.236
si01 – Harpsichord	-2.4671	-1.8598	-1.0398	-1.0578	-0.9873	-0.483
si02 – Castanets	-3.0789	-2.7877	-1.7663	-1.4170	-1.0545	-0.918
si03 – Pitch pipe	-1.0545	-0.9036	-0.7675	-0.6380	-0.6380	-0.542
sm01 – Bagpipes	-3.3451	-2.6495	-1.1496	-0.7118	-0.6265	-0.485
sm02 – Glockenspiel	-3.2723	-2.7661	-2.3840	-2.1697	-2.1484	-0.269
sm03 – Plucked strings	-1.4563	-0.8216	-0.4543	-0.2952	-0.2099	-0.151

Additional 50 atoms add 8.6 *kbps* for the bitrate.

[4] R. Huber, B. Kollmeier, "PEMO-Q – A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception", IEEE Transactions on audio, speech, and language processing, vol. 14, pp. 1902-1911 (2006 November).

8. Encoded Audio Samples Example Results



9. Conclusions & Future Research

Conclusion:

- Perceptually optimized WP dictionary approach allows to adapt psychoacoustically WP tree structure to each signal frame and provides WP dictionary with less number of functions;
- The nonlinear nature of the algorithm leads to compact signal representation;
- Proposed scalable parametric audio encoder using sparse approximation as a core provides:
  - more than twice increase compression ratio (for some sequences 250 atoms variant provides comparable results with AAC).
  - decreasing bitrate depending on signal type.

    (For example, sc01 - 250 atoms, sc02 - 300 atoms (~53 kbps, CR – ~13), si01 - 350 atoms (~62 kbps, CR – ~11) and so on.

Future Research:

- Further parameters selection optimization and quantization algorithm improvement to increase quality of reconstructed audio signal;
- Hardware implementation of scalable parametric audio coder using sparse approximation with frame-to frame perceptually optimized wavelet packet based dictionary as a field programmable system-on-chip (FPSoC).