Pruebas subjetivas de audio-codecs durante el periodo 2010-2012 en la Universidad de San Buenaventura, Bogotá

Subjective audio-coding tests during the 2010-2012 period at the Universidad de San Buenaventura, Bogotá

Marcelo Herrera Martínez*, Darío Alfonso Páez Soto** y Miguel Ricardo Pérez***

Resumen

El presente artículo describe una serie de procedimientos y resultados que fueron obtenidos durante los últimos dos años en la evaluación subjetiva de audio-codecs. El método de testeo usado fue el «Método de doble estímulo con referencia escondida». Los resultados han sido validados estadísticamente con ANOVA de dos factores (Análisis de varianzas»). Se presenta la descripción estadística y los resultados, el valor medio de los codecs y los intervalos de confidencia.

Palabras clave

Compresión de audio, enmascaramiento, psicoacústica, codificadores perceptuales, evaluación de señales de audio, testeo

Abstract

The present paper deals with a set of test procedures and results which were obtained during the last two years on subjective coding evaluation. The testing method for all of

* Ingeniero Electrónico. Magíster en Radioelectrónica y Doctor en Acústica, Universidad Técnica de Praga. Profesor Titular de la Facultad de Ingeniería, Universidad de San Buenaventura, Bogotá. Líder del Semillero de Investigación de –Sistemas de Compresión Perceptual de Audio–. E-mail: mherrera@usbbog.edu.co.

** Ingeniero de Sonido, Universidad de San Buenaventura, Bogotá. Especialista en Gerencia de Proyectos de Ingeniería, Universidad EAN, Docente investigador del programa de Ingeniería de Sonido, Universidad de San Buenaventura, Bogotá. E-mail: dpaez@usbbog.edu.co.

*** Ingeniero en Control e Instrumentación Electrónica, Universidad Distrital Francisco José de Caldas, Especialista en Pedagogía y Docencia Universitaria de la Universidad San Buenaventura. Profesor adscrito programa de Tecnología Electrónica de la Facultad Tecnológica de la Universidad Distrital Francisco José de Caldas, Profesor de la Facultad de Ingeniería, Universidad de San Buenaventura. Investigador del Grupo de investigación Acústica Aplicada. Investigador del grupo de investigación en Robótica Móvil Autónoma (ROMA). Bogotá, Colombia. E-mail: mperez@udistrital.edu.co

Pruebas subjetivas de audio-codecs durante el periodo 2010-2012 en la Universidad de San Buenaventura, Bogotá • pp. 66-73

Investigación

them was the so-called «Double-Blind Stimulus with Hidden Reference». The results are statistically validated with two factor ANOVA (Analysis of Variances). For the statistic description and results the mean power of the codecs as well as confidence intervals are presented.

Keywords

Audio compression, Masking, Psychoacoustics, Perceptual coders, Audio signals, evaluation, test.

I.   Introduction

For broadcasting and storage purposes there is the need to reduce the size of audio files in order to save bandwidth and capacity, respectively. Therefore efficient algorithms are implemented, in order to meet the task, eliminating from the signal irrelevancies and redundancies. Coding technologies use, among other techniques, facts from signal processing, hearing physiology and perception. The masking phenomenon that takes place inside the inner ear, in the basilar membrane, is the key to new designs of audio coding devices, specifically their psychoacoustical models[1]. Nevertheless, there is a huge number of codecs available on the market and on the internet, and the need for their careful evaluation has become a necessity. Therefore, a definite number of audiophile groups have gotten involved in this evaluation process by both, subjective and objective approaches. The present work focuses on searching the correlation between audio subjective quality, codec-bit rate combination, CD quality and statistical results obtained with ANOVA methods.

The double-blind triple-stimulus with hidden reference method is used, in which listeners are asked to listen to three tracks of the same audio sequence, one of which was the reference one, labelled „R«, and the other two „A« and „B«. One of them (A or B) is the reference, and the other the compressed signal. Listeners are asked to recognize which is the compressed one and to judge its basic audio quality, on the scale from one to five, as is shown in the Tab. 2.

The listener could switch freely between the presentations 'Reference', 'A' and 'B', where 'A' and 'B' were the processed version and the hidden reference, randomly allocated for that trial. Each excerpt could be repeated as often as required. The listener was asked to judge the 'Basic Audio Quality' of the 'A' and 'B' versions in each trial and any difference from the reference as to be considered as an impairment.

Impairment description	Grade	Level
Imperceptible	5.0	
Perceptible but not annoying	4.0	
Slightly annoying	3.0	
Annoying	2.0	
Very annoying	1.0	

Table. 1 ITU-R five grade impairment scale.

II.   Statistical analysis

The first step in the analysis of the reliability and the validation of the results is to select the codec-bit rate combinations that have a very poor quality and which results could affect negatively the output of the results. In order to achieve this goal we will look for the codec-bit rate combination with a performance lower than –2 diffgrade according to ITU-R BS. 1116-1[2].

The second step is to evaluate the listeners performance. For that purpose we will set the hypothesis, H_0, affirming that a subject will be more reliable while his/her mean score will be different from zero and this to negative values (which will mean that the subject identifies correctly the differences between the compressed and original versions). This will be achieved applying a one-tailed t-test (Student test), searching for the significance value $p < 0.1$.

The hypothesis formulation can be expressed as, $H_0 : \mu \neq 0$, and the alternative hypothesis,

$$H_1 = 0.$$

A two factor ANOVA allows us to examine the interaction of the two factors, the codec-bit rate combination and the excerpt factor. The way we do this is to separate the between conditions sums of squares into the components due to the main effects of each factor and the interaction. We investigate a significant interaction further by looking at the simple main effects of one factor at the various conditions of the factor, taken one at a time. In this way we can discover the source of the interaction [3].

III.   Results

Results for the subjective testing of AAC, WMA and OGG formats

AAC (Audio Advanced Coding Technologies), which is equivalent with the MPEG-4 (Moving Picture Experts Group), was tested at 96 and 128 kbit/s as well as OGG and WMA .

The critical audio material chosen stresses different aspects of the codecs resolution. Mainly there were selected excerpts which stresses transient phenomena, and high frequency components. In the next table some characteristics of used excerpts are given.

Item	Audio excerpt	Description
1	Eric Clapton – Don't know which way to go	Charles drum evaluation
2	English female speaker, from mpeg web	Broadcast english female
3	French male speaker, critical mpeg material	Broadcast french male
4	Castanets, critical mpeg web material	Highly transient signal
5	Erotica, Grieg	Piano classical sequence
6	Alleluja, Adash	Female choral
7	Eric Clapton – Preludin fugue	Drum changes, Transient signal

Table. 2. Description of the used excerpts

Pruebas subjetivas de audio-codecs durante el periodo 2010-2012 en la Universidad de San Buenaventura, Bogotá • pp. 66-73

Investigación

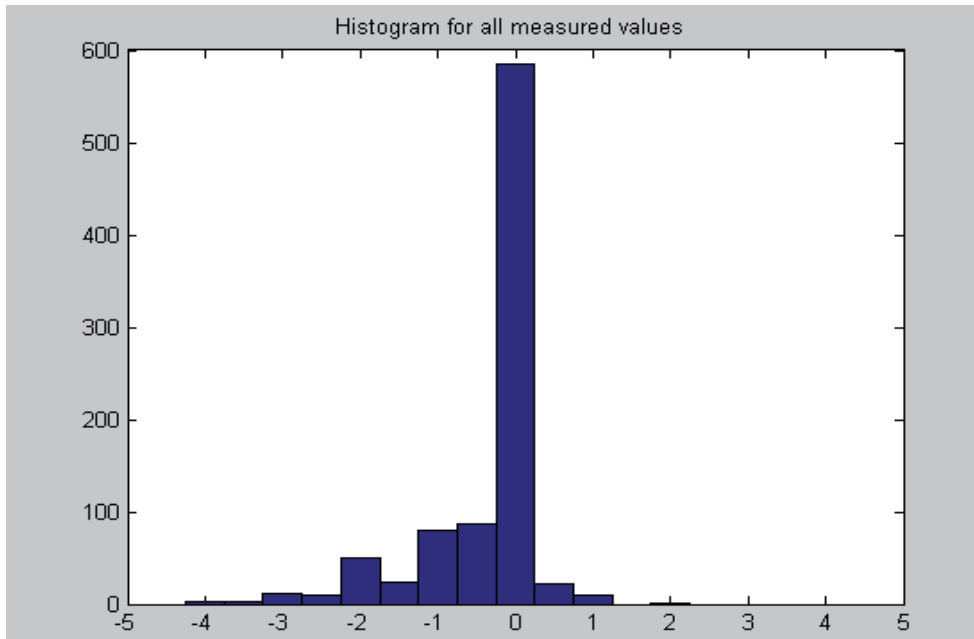The histogram for all measured values, computed in MATLAB, is presented.

Figure 1. The Histogram for all measured values

The results for the ANOVA table are presented:

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Square	Variance Ratio (F)	Probability
Factor A	5	109,8867	21,9773	59,0312	p<0.05
Factor B	6	12,0919	2,0153	5,4131	
Factor A X B	30	47,5625	1,5854	4,2584	
Error	840	312,7176	0,3723		
Total	881	482,2587			

Table 3. ANOVA Results Table for the first test

We see that the value of the F-statistic is very high for the codec-bit rate factor, with a significance p less than 0.01. This means that there is a high correlation between format used and the listeners marks. Some formats have better performance than others, the same is true about the excerpts. In some excerpts the masking difference between the original and compressed version can be higher than in others. For the interaction of the two factors the same reasoning is valid. Some excerpts are critical for a determined format, but for others not.

Performance of the individual codecs on the chosen excerpts are graphically described, plotting the mean of the scores, and the confidence interval for each measurement.
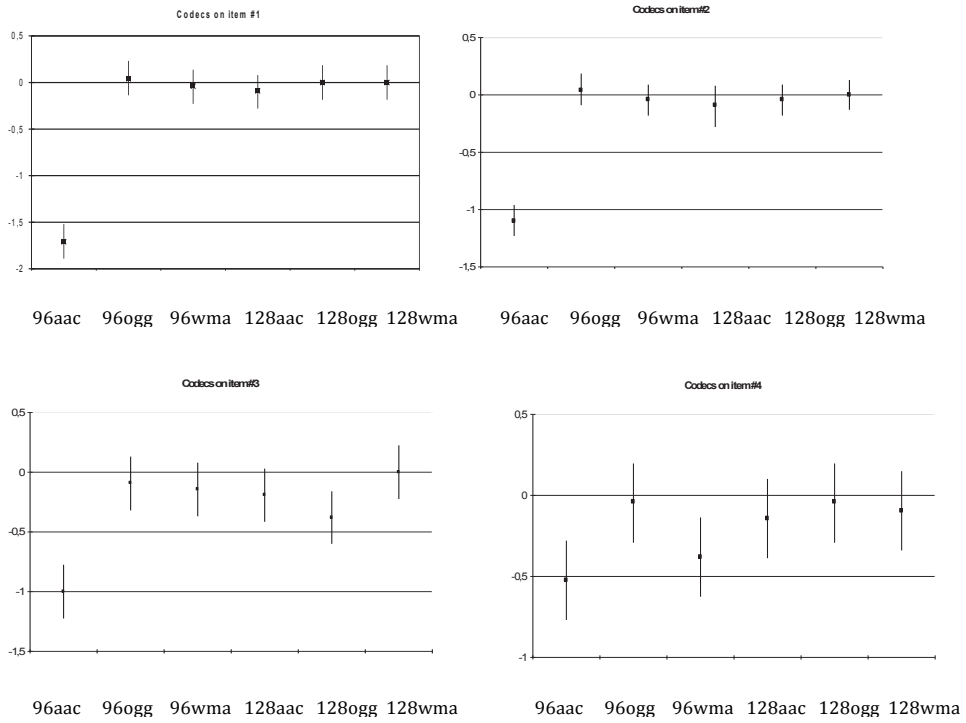
Figure 2. Average subjective results for codec comparison. a) Codecs on item No. 1. b) Codecs on item No. 2. c) Codecs on item No. 3. d) Codecs on item No. 4.
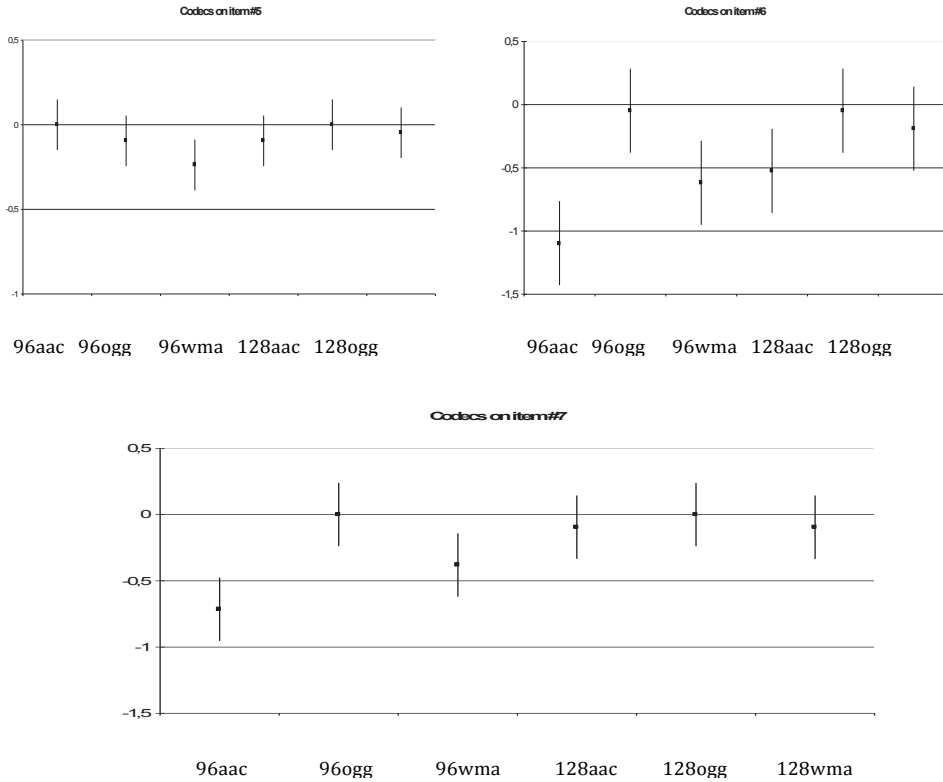

Figure 3. Average subjective results for codec comparison. a) Codecs on item No. 5. b) Codecs on item No. 6. c) Codecs on item No. 3. d) Codecs on item No. 7.

Pruebas subjetivas de audio-codecs durante el periodo 2010-2012 en la Universidad de San Buenaventura, Bogotá • pp. 66-73

Investigación

Results for the subjective testing of MP3-FhG, MP3-Lame, AAC, OGG and WMA

These formats were evaluated at 128, 96, and 64 bit rates. The histogram of the diffgrades is presented in the following figure:
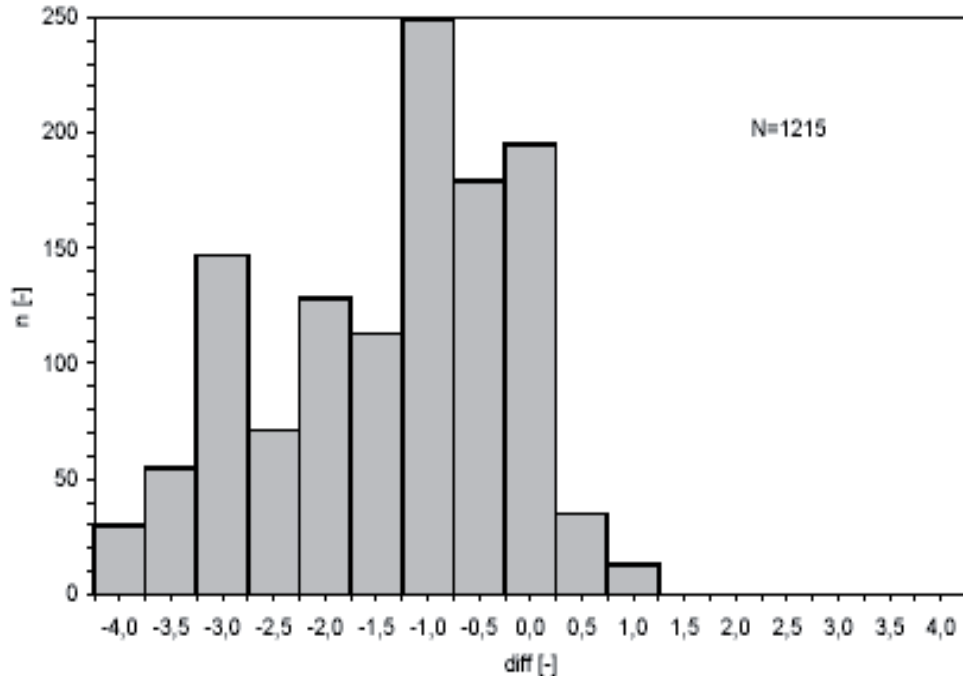

Figure 4. The Histogram for all measured values

Č.	Autor	Skladba
1	Alice in Chains	No Excuses
2	Maceo Parker	We're On The Move
3	-123 min.	Runnin' On

Table 4. Critical material for the second test

The first excerpt of the audio material chosen stresses the codec influence on drums, cymbals and the acoustical guitar. For the second one, the same affirmation is true. For the last excerpt the critical element is the vocal track.

ANOVA resulting table is presented next:

	SS	v	MS	F	p
Signal	5,4280	2	2,7140	8,1309	0,000
Compression	111,8308	9	12,4256	37,2259	0,000
Signal xCompression	21,5291	18	1,1961	3,5833	0,000
Error	200,2743	600	0,3338		
Total	339,0622	629			

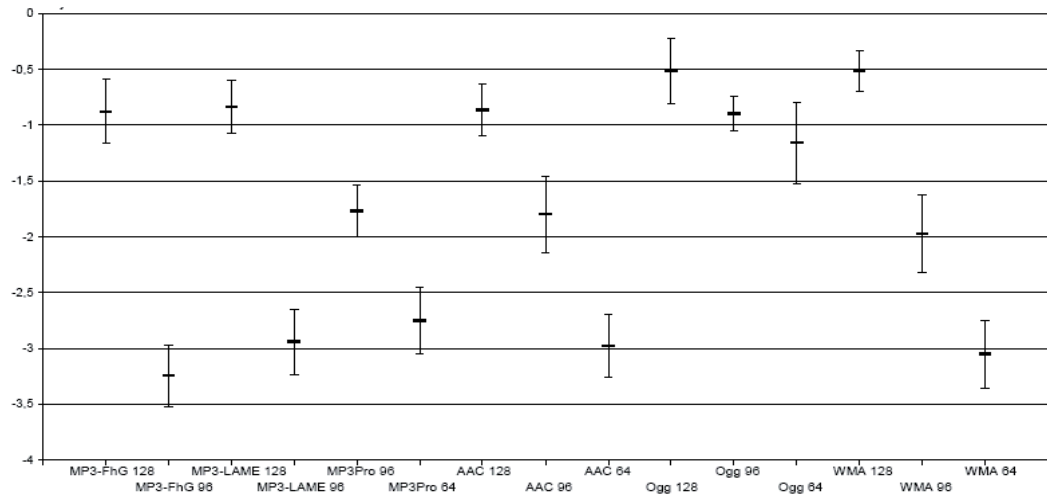Table 5. ANOVA Results Table for the second test
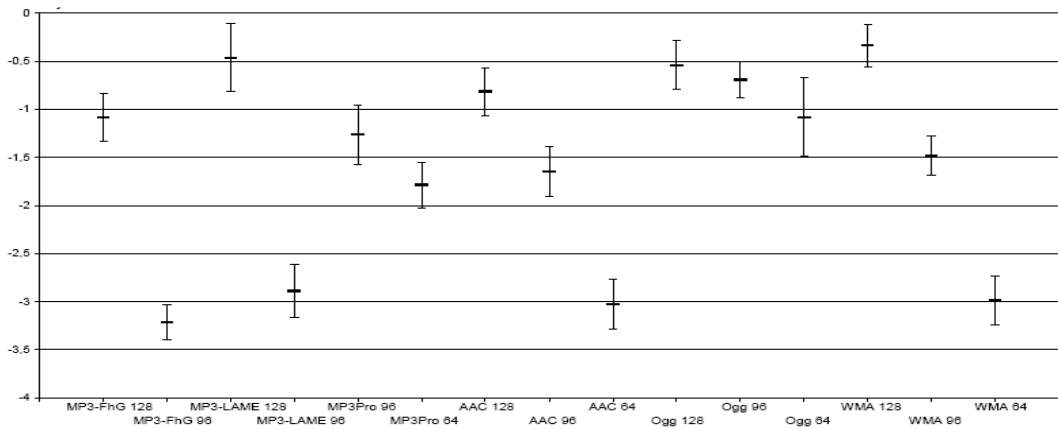
Figure 5. Results for the Excerpt #1

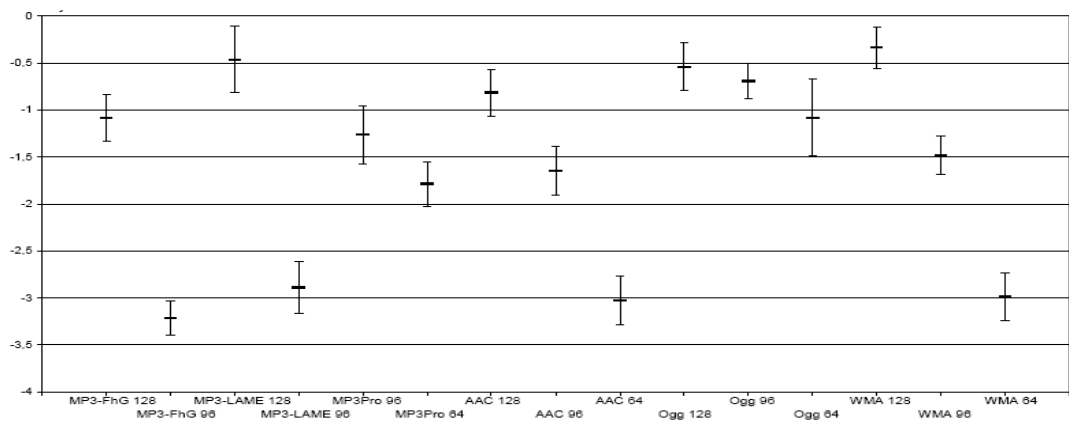Figure 6. Results for the Excerpt #2

Figure 7. Results for the Excerpt #3

Examining the results of the second experiment, we may affirm that no compressed version is transparent on the excerpts applied.

Pruebas subjetivas de audio-codecs durante el periodo 2010-2012 en la Universidad de San Buenaventura, Bogotá • pp. 66-73

Investigación

IV.   Conclusions

The evaluation of audio compression systems has been performed by the DBTS psychometric methodology recommended by the ITU.

The statistical analysis of the test was performer, showing a high degree of inter/intra reliability of the psychoacoustic experiment.

Ogg Vorbis is the only format that attains CD-quality already at 96 kbit/s specially on excerpts formed by percussion elements, and on choral excerpts. WMA Microsoft performed fine for speech radio broadcasting.

References

[1]   H. Zwicker, Fastl:«Psychoacoustics», *Springer Verlag* 1990.

[2]   P. Hinton, «Statistics Explained», *Routledge* (2013), 2nd. ed., 2013.

[3]   Recommendation ITU-R BS. 1116-1: *Methods for the Subjective Assessment of Small Impairments in Audio Systems including Multichannel Sound Systems*. (1994-1997).

[4]   M. G. Bosi, Goldberg: *Introduction to audio coding and standards,* Boston, 2003.

[5]   J.G. Proakis and D.G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, Prentice-Hall, NJ, Third Edition, 1996.

[6]   S.G. Mallat. *A Wavelet Tour of Signal Processing.* 2nd Edition. Academic Press, 1999. ISBN 0-12-466606-X

[7]   C. K. Pohlmann *Principles of Digital Audio*, Mc Graw-Hill, Fourth edition, 2000.

[8]   P. Srinivasan and L. H. Jamieson. «High Quality Audio Compression Using an Adaptive Wavelet Packet Decomposition and Psychoacoustic Modeling», *IEEE Transactions on Signal Processing*, Vol 46, No. 4, April 1998.

[9]   T. von Petersdorff, «How to use the Matlab wavelet toolbox» *(web page)*, Scientific Computing I (MAPL/CMSC 660), University of Maryland, College Park, MD, USA, 1999. Disponible: http://www.glue.umd.edu/~tvp/660/matlabwav.html.

[10]   R. A. Devore and B. J. Lucier, *Wavelets*, University of South Carolina, Columbia, SC and Purdue University, West Lafayette, IN, 50 pp., 1992.

[11]   SO/IEC 11172-3, «Information technology - Coding of moving pictures and associated audio for digital storage media at up to 1.5 Mbit/s» - (Part 3),1992.

[12]   P. Srinivasan, L. H. Jamieson, «High-quality compression using an adaptive wavelet packet decomposition and psychoacoustic modeling», *IEEE Trans. on Signal Processing*, Vol. 46, No. 4, pp. 1085-1093, April 1998.