Fidelity Estimation of Watermarked Audio Signals According to the ITU-R BS.1116-1 Standard

Z. Piotrowski and P. Gajewski

Telecommunications Institute, Military University of Technology

gen. S. Kaliskiego 2, 00-908 Warsaw, Poland

The paper describes the method that enables to measure acoustic fidelity of audio signals degraded with digital watermarks. Fidelity measurements of degraded acoustic signals is suitable for such applications as distinguishing whether the acoustic record is signed with a digital watermark (watermarked) or not as well as evaluation of the degradation degree affecting the host signal. The test of subjective assessment according to the recommendation ITU BS.1116-1 is the standard that served as the basis for implementation of a dedicated application within the Matlab environment and for development of the concept how to make this application available as an IP network service to enable assessment of degraded signal by a wider auditorium.

PACS: 43.75.Zz, 43.72.Uv, 43.66.−x, 43.66.+y

1. Introduction

Wide range of methods that are used to measure the degradation degree of acoustic signals [1–9] includes also a commonly recognized standard, which is the subjective test for fidelity assessment BS 1116-1 [10]. It is the test that is used to fidelity estimations of acoustic signals by comparison to its standard pattern — the host signal — for such applications as improvements of the MPEG-4 AAC lossy compression [11, 12], assessment of the signal subjected to the phase randomization procedure during the perceptual coding of audio signals [13], evaluation how the number of lost data packages affects the perceived fidelity of audio signals [14] or effect of the oversampled discrete wavelet transform (ODTW) transform onto fidelity of signals subjected to compression with comparison to standard signal processing based on the discrete wavelet transform [15]. In this study it is proposed to use the ITU-R BS.1116-1 test for the subjective assessment how much a watermark affects the degradation degree of the host signal and for determination of the just noticeable difference (JND) level by means of human perception. Correction of the watermark signal to the level of the minimal masking threshold LT_{min} guarantees that the JND level is achieved when auditors are incapable to distinguish the signal with the watermark from the one without the watermark during 50% of audition tests. Therefore JND is the threshold for potential inaudibility of the watermark signal at presence of the host signal.

2. The problem of a hidden transmission detection

Modern systems of hidden transmission must be resistant to the stegoanalysis and to perceptual detection.

The system resistance to stegoanalysis means that confidential information hidden beneath the original acoustic layer is invulnerable to detection attempts. There exist various types of stegoanalytic attacks. One of such attacks consists in distinguishing of a data stream generated by a steganographic transmitter from the stream that has the perfectly random nature. When the vector code of noise is not perfect, the collision in the steganographic sequence (in the stego-code) occurs with a defined probability. The next type of the attack enforces initiation of clear linear relationships between internal states of the algorithm for the code generation.

The robustness of the system against already mentioned perceptual detection is also a very important pre-condition that must be met by correctly designed system. It is surprising but there are a very low, nearly negligible proportion of people who are capable to distinguish very insignificant acoustic distinctions between the host and modified signals, so called "golden ears". It is even more significant that the Human Auditory System still remains the unattainable pattern and model of an acoustic receiver, very difficult for computer modeling, even in times when the theory dedicated to operation of the auditory apparatus is being developed at exponential rates. Therefore the fact of hidden transmission can be detected with pretty high probability by such a user of the system who is gifted with such hearing. Such a probability is even higher than the chance that the stego-code is detected and decoded upon a computer-controlled analytical attempt. Frequently, it is enough to know that a hidden transmission occurs before further advanced methods for analytic detection and recognition of the hidden signal are undertaken.

3. Assessment of the perceptual transparency for transmitted signals

Correctly adjusted transmitter of the watermark (WM) signal supplies the signal at the JND level, which guarantees that the watermark is inaudible at presence of the host signal (HS). Simultaneously, the maximum power of the WM signal is achieved in order to enable error-free decoding of the WM signal signature. Watermark signal is embedded into the original host one according to the method described in [16].

The experiment consisted in the subjective quality assessment of the signals with intruded watermarks by listeners, where audibility tests were carried out according to the standard ITU-R BS.1116-1 [10]. For that purpose, the dedicated application "watermark quality estimator" (WQE) software was developed together with the relevant procedure that served as the guideline to carry out tests according to the standard BS.1116-1 with participation of 22 listeners demonstrating normal hearing capabilities. There were used 25 tracks in experiment based on the ITU-T repository of the speech samples [17]. The watermarking algorithm operated in frequency domain [16] to embed additional data payload in each audio track. The *.wav format were used with sampling frequency 44.1 kHz and resolution 16 bits per sample.

3.1. The ITU-R BS.1116-1 Standard

Description of the standard [10] comprises the detailed description how to measure quality of degraded signals and which acoustic conditions must be met when the audition tests are in progress. The test described in the standard BS.1116-1 is a typical procedure for subjective fidelity examination of signals subjected to the digital watermarking procedure. The presented method assumes that participants are asked to listen to three acoustic signals that correspond to the same fragment of a musical piece. The first signal is free of degradation with any factor and represents the pure host signal, two other signals are randomly degraded (either the second or the third one) but no information which acoustic signal is degraded is made available to the listener. The term of degraded signal stands here for the signal with the watermark included. The task of the listener is to compare the pure host signal against two other samples and to find out, which one is degraded and to which degree. The assessment of the degradation degree for the presented signals is done in the subjective way with the use of the five-point scale of the signal degradation: 5 — inaudible, 4 — audible but not irritating, 3 — slightly irritating, 2 — irritating, 1 — very irritating.

As a result of assessments performed by a specific participant the scale of diff grades (SDG) is developed. The scale is defined as the difference between the two values: the test signal and the reference one, where the possible values of results range from −4 to 4 and positive SDG values correspond to faulty assessments of listeners. The BS 1116-1 test is also used to determine the forced choice (FC) factor that brings information about the percentage of hits (correct answers), i.e. the signal assessed by the survey participant as a degraded one was actually degraded. The FC factor is calculated according to the following formula (1):

$$FC = \frac{A}{N} \cdot 100\%,\tag{1}$$

where A — number of correct answers, N — total number of answers.

For FC = 50% the assumption can be made that degradation of the signal with the watermark is inaudible and reached the level zero JND. Therefore, the BS 1116-1 can be considered as the test for the declared level of the JND transparency (the watermark inaudibility threshold) for the watermark embedder whether the declaration really matches the observable level. The audition material was chiefly composed of soundtracks made available by the European Broadcasting Union (EBU) organization. The measurements were carried out in a room with acoustic insulation, with use of a PC with the "WQE" application software and the closed headphones of the HiFi class type: RP-HT600 Technics, sound card type: Sound Blaster X-Fi Surround 5.1 (24 bits). Each track duration length was 10 s. Number of participants invited to the test was 22. The participants had normal hearing in a range of standard audiometric frequencies. The BS.1116-1 standard also describes how to carry out the statistic analysis of the acquired results.

3.2. Estimator for quality of watermarked signals — watermark quality estimator

The developed software named WQE (Fig. 1) is dedicated for comprehensive quality measurements of degraded signals according to the standard ITU-R BS.116-1 with further statistical analysis of measurement results. The software is capable to perform the following functions: logging the users in and out with recording of the user-specific data within the measuring system, visualisation of the process when user selects the degradation degree for acoustic signals, synchronisation of the measurement process with playing acoustic signals from soundtracks, calculation of measurement results in the form of a matrix of SDG values as well as the coefficients of forced choice FC [%]m and statistical analysis.

3.3. Statistical calculations for SDG values

The completed experiment with use of WQE application software made it possible to obtain the following statistical values for the SDG parameter (Fig. 2 and Fig. 3).

Interpretation of received results is the following. In Fig. 2 each track from 25 were estimated by 22 listeners, thus each numbered record represents averaged value. According to the SDG value, the positive values represent false-positive type error, mean that tacks were assessed as watermarked (degraded) signals rather than original one. Values lower than −1 SDG mean that watermark signal is not transparent for Human Auditory System at
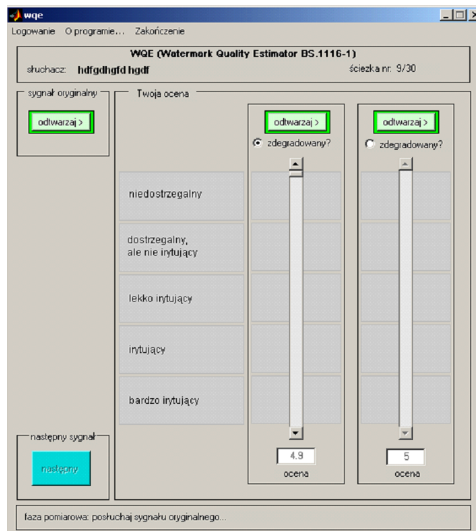
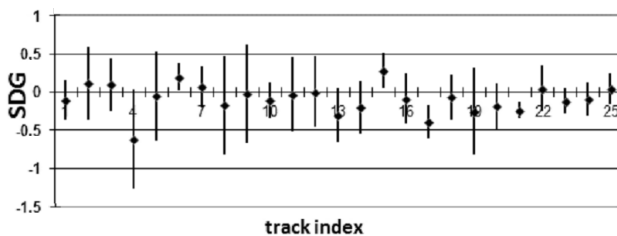Fig. 1. Graphical user interface (GUI) of the WQE software in Matlab environment.

Fig. 2. Statistical values for the SDG parameter (with the confidence interval of 95%).

presence of the original audio host signal (power of watermark signal was too high to be inaudible at presence of host audio signal). In Fig. 3 the correct answers in percentage value can be observed. Taking into account track no. 1, 67% correct answers were given (listeners distinguished with success if track is or not degraded by watermark signal).
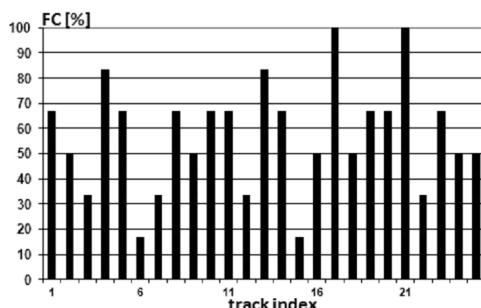

Fig. 3. Coefficients of forced choice (FC) [%].

4. The concept for measurements of the signal fidelity via the IP network

Transfer of signal fidelity measurements according to the test procedure ITU-R BS.1116-1 to a IP network of distributed participants offers a number of advantages: increased number of participant who can listen to test signals and therefore reduction of the measurement uncertainty, comfortable conditions of the audition test (individual participants perform the on-line test at home, e.g. with use of their home PCs). However, such tests performed via IP networks have also a number of drawbacks: it is impossible to be sure whether the participant has healthy Human Auditory System (without deficiencies) [18, 19], location of tests can be unsuitable for experiments (noise in the test room) [20]. It is assumed that implementation of the WQE application software as a web tool shall substantially facilitate the selection process of participants and cuts down duration of experiments. Prior to the actual test of the signal fidelity according to the standard ITU-R BS.1116-1 carried out via the IP network it is necessary to perform a remote (via the internet) examination of hearing by collection of an audiogram from each participant, for instance with use of the on-line System for Broad Examination of Hearing "Słyszę" ("I can hear") [21, 22]. This concept can be used mainly for screening purposes.

5. Summary

The completed experiments made it possible to find out that the watermark signal embedded into the host signal by a watermark embedder is transparent for human perception (inaudible). Participants of the survey assessed that the WM signal is unnoticeable for them (inaudible). Only for a single audition track the value of SDG < −1 and it also happened that the degraded (watermarked) and host signals were mutually mismatched, i.e. the SDG values are > 0.

Use of the statistical analytic tools diminishes dispersion of estimated values around the arithmetical mean. In case, if for majority of audition tracks the SDG values would be less than (−1), it would serve as the confirmation that the signal is audible but not irritating. Obviously, in such cases it is necessary to increase the signal-to-mask (SMR) ratio within the watermark embedder. It will lead to inaudibility of the WM signal but, unfortunately, it also entails drop of the probability that the watermark signature would be correctly detected.

The concept to transfer measurements of the signal fidelity for the test ITU-R BS.1116-1 to a IP network potentially offers a number of advantages, where reduction of the measurement uncertainty owing to extension of the survey participants seems to be the most important one. The BS.1116-1 test is a good estimator for quality of watermarked signals. It is based on calculation of values for FC coefficients [%] and for the SDG parameter for each of the signals under estimation. Designers of digital watermarking systems based in the standard BS.1116-1 are

therefore capable to check whether the threshold of perceptual transparency for watermarks determined by the process of psychoacoustic correction is in line with the assumed value and, consequently, whether signals with watermarks are undistinguishable for the Human Auditory System from the host signal without the watermark.

References

[1] M.T. Ali, M. Saleem, in: *Proc. ICEE*, Ed. A.H.M. Zahirul Alam, International Islamic University Malaysia, Kuala Lumpur 2008, p. 1.

[2] M.T. Ali, M.S. Mian, in: *Proc. TENCON*, IEEE Region 10, Hong Kong 2006, p. 1.

[3] M.T. Ali, M.S. Mian, in: *Proc. WASPAA*, Ed. W. Kellermann, IEEE Computer Society, New York 2005, p. 223.

[4] T. Painter, A. Spanias, *Proc. IEEE* **88**, 451 (2000).

[5] Standard ITU-R BS.1116-1, *Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems*, Geneva, Switzerland 1997.

[6] J.E. Voldhaug, E. Hellerud, U.P. Svensson, in: *Proc. ICSPC*, Ed. M. al-Mualla, IEEE Region 8, Dubai 2007, p. 536.

[7] F. Sinaga, T.S. Gunawan, E. Ambikairajah, in: *Proc. ICSP*, Eds. K.H. Li, J. Ngamwiwit, IEEE, Bangkok 2005, p. 191.

[8] A. Czyżewski, H. Skarzyński, www.slysze.telezdrowie.pl .

[9] J.C. Hardin, C.D. Creusere, in: *Proc. ACSSC*, Ed. M.B. Matthews, IEEE Signal Processing Society, Pacific Grove 2008, p. 1245.

[10] S. Kandadai, J. Hardin, C.D. Creusere, in: *Proc. ICASSP*, Ed. A.H. Sayed, IEEE Signal Processing Society, Las Vegas 2008, p. 221.

[11] C.D. Creusere, K.D. Kallakuri, R. Vanam R., *IEEE Trans. Audio, Speech Language Proc.* **16**, 129 (2008).

[12] Standard ITU-R BS.1387-1, *Method for objective measurements of perceived audio quality*, Geneva, Switzerland 2000.

[13] A.Z. Jusoh, R. Togneri, B. Rohani, S. Nordholm, in: *Proc. APCC*, Ed. Y. Zhen, IEEE Signal Processing Society, Shanghai 2009, p. 694.

[14] E. Skrodzka, D. Hojan-Jezierska, *Acta Phys. Pol. A* **118**, 155 (2010).

[15] E. Skrodzka, B. Czechyra, F. Tomaszewski, *Acta Acustica united with Acustica Supplement 1*, **96**, 72 (2010).

[16] P. Malecki, J. Wiciak, *Acta Phys. Pol. A* **118**, 118 (2010).

[17] B.W. Ramsey, B.E. Mullins, R.W. Thomas, T.R. Andel, in: *Proc. IPCCC*, Eds. M. Cheng, Y. Zhang, IEEE Computer Society, Phoenix 2008, p. 469.

[18] Z. Piotrowski, *Proc. Reg. Conf. Military Commun. Inform. Systems* Eds. W. Burakowski, A. Wieczorek, Military Communication Institute, Zegrze (Poland) 2002, p. 147.

[19] C. Creusere, J. Hardin, *IEEE Trans. Audio, Speech Language Proc.* **19**, 711 (2010).

[20] A. Czyżewski, B. Kostek, J. Kotus, in: *Multimedia Services Applied to Noise and Hearing Monitoring and Measuring*, Ed. J. Kacprzyk, Springer, Berlin 2008, p. 275.

[21] http://www.itu.int/net/itu-t/sigdb/genaudio/AudioForm-g.aspx?val=1000050 .

[22] Z. Piotrowski, P. Gajewski, *Computational Methods and Experimental Measurements XIII*, WIT Press, Southampton 2007, p. 791.