Precise Psychoacoustic Correction Method Based on Calculation of JND Level

Z. Piotrowski*

Faculty of Electronics, Military University of Technology, gen. S. Kaliskiego 2, 00-908 Warsaw, Poland

This paper constitutes a proposal of a method of spectral shaping of the watermark signal without adjusting the interception point based directly on psychoacoustics of the human auditory system and the proposed calculation formula. The proposed calculation formula enables correction of watermark signal both as regards its spectrum shape and adjustment of watermark signal to the host's JND.

PACS numbers: 43.66.−x, 43.66.+y, 43.60.−c, 43.60.+d

1. Introduction

Information hiding is one of the techniques used in digital signal processing. Digital watermarking [1] is the information hiding technique used to hide additional digital content within original content (digital music, image, movie), for example to ensure identifiability of an acoustic signal among many other identical signals. In such an event the watermark signal (WS) inserted into the original host signal (HS) must be not only be resistant to intentional and unintentional degradation factors, but also perceptually transparent. In case of acoustic signals transparency of WS is ensured by means of the so--called psychoacoustic models, which are based on the human auditory system (HAS) and used in information hiding technique to compute the just noticeable difference (JND) level. In psychoacoustics the JND level is defined as the level of distortion that can be perceived in 50% of experimental trials. Of course, as calculation of the JND level is based on average and subjective perception of listeners with normal hearing and no loss of hearing is indicated by the audiogram in case of listeners with normal hearing, listeners with an exceptionally sensitive auditory system who are often referred to as "golden ears" are in most cases still able to perceive a difference between the original and the signal degraded to JND level. In the subjective fidelity audio test conducted in accordance with BS-1116 recommendation the listener is asked to assess which of the two presented similar audio stimuli has been degraded acoustically. Therefore the purpose of psychoacoustic models is to provide the basis for mathematical computation of a response to the acoustic stimulus perceived by the listener. The watermarked (WM) signal equals HS with embedded WS corrected to the JND level.

_____

* e-mail: Zbigniew.Piotrowski@wel.wat.edu.pl

Presented below is the mathematical formula for calculation of the spectral correction coefficient, which constitutes the basis of correction of WS to JND level. WS correction to JND level is calculated through incorporation of the human auditory system into the Mpeg Layer 1 lossy compression algorithm, which is commonly referred to as the mp3 format. The proposed spectrum correction coefficient differs from the standard method for calculation of the insertion point, by which the WS spectrum is corrected [2–4].

2. Human auditory system – psychoacoustic model

The objective of the acoustic perception model is to estimate the minimum masking threshold for each critical sub-band (n) within Bark scale, defined in [5–7] as:

$$LT_{\min}(n) = \min[LT_{\mathrm{g}}(i)] \text{ [dB]},$$

where $LT_{\mathrm{g}}(i)$ — global masking threshold computed for each spectral line (i), n — critical sub-band (Bark) specified as in [8, 9]:

$$z(f) = 13 \arctan(0.00076 f)$$

$$+ 3.5 \arctan\left[\left(\frac{f}{7600}\right)^2\right] \text{ [Bark]}. \tag{1}$$

Bark critical bands can be presented as a function of a Bark scale number and the corresponding linear frequency [Hz], as in Fig. 1.

The minimum masking threshold $(LT_{\min})$ constitutes the basis for calculation of the signal-to-mask ratio (SMR), as in Fig. 2.

SMR is the difference between the masker and the minimum value of the masking threshold within band measured in the Bark scale and SMR is expressed with the following formula:

$$\mathrm{SMR}_{\mathrm{sb}}(n) = L_{\mathrm{sb}}(n) - LT_{\min}(n) \text{ [dB]}, \tag{2}$$

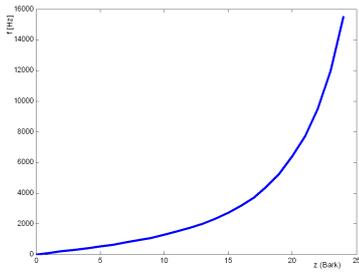where $L_{\mathrm{sb}}(n)$ — sound pressure level in [dB] for each crit-

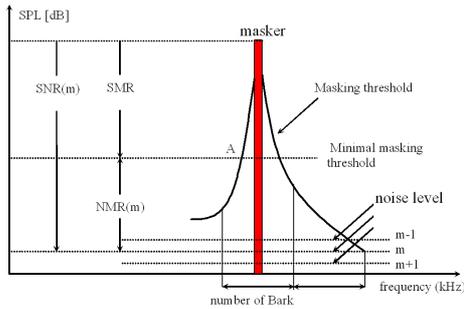Fig. 1. Critical bands in the Bark scale (x-axis [Barks], y-label frequency [Hz]).

Fig. 2. SMR ratio and minimum masking threshold in the audio perceptual model, y-axis sound pressure level (SPL), x-axis Bark scale.

ical band n (Bark), as against reference level $pp = 20\ \mu\text{Pa}$ and expressed with the following formula:

$$\text{SPL}(n) = 20\log 10(p/pp). \tag{3}$$

The minimum masking threshold is computed on the basis of the frequency masking effect. This means that, based on the so-called spreading function, it is possible to estimate both the frequency areas and the corresponding signal levels below which the signal will be inaudible, as the given frequency range is dominated by another signal that is stronger than the remaining signals, as in Fig. 3.
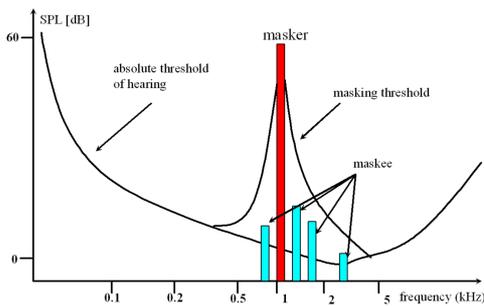

Fig. 3. Frequency masking phenomena — the masker and masking threshold computed based on spreading function.

The shape of the spreading functions depends on the Bark critical band (n) and it can be expressed with the following formula:

$$\text{SF}(n) = 15.81 + 7.5(n + 0.474)$$

$$-17.5\sqrt{1 + (n + 0.474)^2}. \tag{4}$$

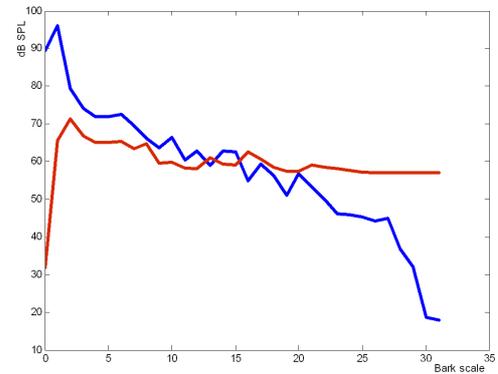The minimum masking threshold for the signal under consideration is presented in Fig. 4.

Fig. 4. Minimum masking threshold level (horizontal line) vs. host signal SPL curve (decreasing).

3. Watermark psychoacoustic correction procedures

Determination of the minimum masking threshold is useful in determination of the spectral shape and efficiency of the watermark. When WS is corrected just below the minimum masking threshold, the JND level is reached, which constitutes a guarantee that the corrected watermark embedded into HS will be inaudible for most listeners with normal hearing. This is possible due to the fact that any energy just below $LT_{\min}$ is inaudible to HAS.

Another well-known watermark correction method [2, 7, 9] is based on determination of the minimum masking threshold for HS. The watermark signal WS is generated in accordance with a well-known pseudo-random noise sequence and represented by white noise. The first stage of WS formation consists in spectral shaping of white noise to $LT_{\min}$ shape by frequency filtering. White noise undergoes fast Fourier transformation (FF) and then is combined with $LT_{\min}$ values for each subband.

The second stage consists in correction of WS against the $LT_{\min}$ curve interception point. This stage is completed by computation of the minimum difference (of the correction coefficient) between $LT_{\min}$ curve and shaped noise (coloured noise) in the logarithmic scale of the signal's amplitude. Upon application of inversion fast Fourier transformation (IFFT) coloured noise is corrected by means of the correction coefficient. The resulting noise is inaudible in the presence of the original signal. The stages of watermark correction are presented in Fig. 5.

The corrected watermark signal is then embedded into HS. As a result the output shaped noise (WS) after the two-stage correction procedure is inaudible at HS presence. To sum up, the method proposed in [2] requires the following:
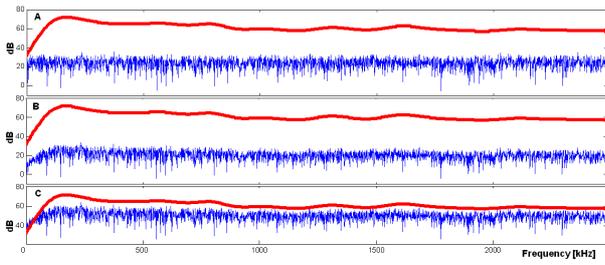
Fig. 5. Watermark signal two-stage correction procedure: (A) white noise (WM) and $LT_{\min}$ (dotted line), (B) stage 1: noise filtering (colored noise), (C) stage 2: noise scaling using interception point, just below $LT_{\min}$ curve.

In stage 1: one FFT operation and one multiplication of the mask with a complex signal (FFT for WS, multiplication of $LT_{\min}$ mask vector with WS spectrum).

In stage 2: computation of the minimum difference between corrected WS (coloured noise) and LT_{g} global masking threshold ($LT_{\min}$ set for all Bark subbands), computation of IFFT for corrected WS and multiplication of corrected WS by the correction coefficient.

The method proposed in this paper consists of one stage and is based on the following formula for computation of only one correction coefficient on the basis of the psychoacoustic model:

$$\mathrm{coef}_{\mathrm{WM}} = \mathrm{SPL}_{\mathrm{WM}} + (\mathrm{Delta}_{\mathrm{H}} - \mathrm{Delta}_{\mathrm{WM}})$$

$$- \mathrm{LT}_{\min} + C, \qquad (5)$$

where $\mathrm{coef}_{\mathrm{WM}}$ — correction coefficient values [dB] for WS, $\mathrm{SPL}_{\mathrm{WM}}$ — sound pressure level [dB] for WS, $\mathrm{Delta}_{\mathrm{H}}$ — normalization coefficient for HS, $\mathrm{Delta}_{\mathrm{WM}}$ — normalization coefficient for WS in reference to 96 dB SPL (for 16 bit A/C converter), C — (optionally) constant value added in order to determine SMR [dB].

Before WS is thoroughly processed in accordance with MPEG 1 Layer 1 psychoacoustic model [5] HS must be divided into frames and transformed into frequency domain by means of FFT. Then the power spectrum density of WS is computed on the basis of the following formula:

$$X = \max\left(20\log 10\left(\mathrm{abs}\left(fft(xw)\right)/512\right) - 200\right), \quad (6)$$

X — power spectrum density, x — input signal vector, w — windowing function, 512 — symmetric FFT points [-], 200 — minimum signal power [dB].

The windowing function, defined as normalization to 96 dB SPL reference level, is carried out in standard [5] in accordance with the following formula:

$$w = \sqrt{\frac{8}{3}} \cdot \mathrm{hanning}(512). \qquad (7)$$

Normalization to the reference level 96 dB SPL is carried out in standard [5] using the following formula:

$$\mathrm{Delta} = 96 - \max(X), \qquad (8)$$

$$X = X + \mathrm{Delta}. \qquad (9)$$

The result of normalization for power density function for one HS frame is presented in Fig. 6. The obtained value $\mathrm{Delta}_{\mathrm{H}}$ equals in this example 126.58 dB.
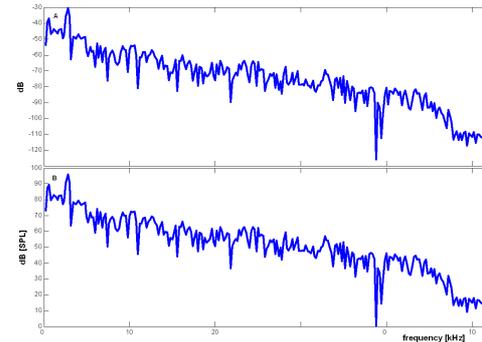

Fig. 6. Normalization of the power spectrum density to the reference sound pressure level of 96 dB. Signal spectrum (A) before normalization (maximum value: –30.58 dB); (B) after normalization (maximum value: 96 dB).

Correction of WS in accordance with the proposed procedure requires performance of two FFT operations and one multiplication of the complex vector by correction coefficient $\mathrm{coef}_{\mathrm{WM}}$ (FFT is required for determination of normalized power spectrum density function, multiplication of WS spectrum vector by correction coefficient $\mathrm{coef}_{\mathrm{WM}}$ and IFFT in order to obtain the corrected WS vector in time). Figure 7 presents the results of one-stage correction.
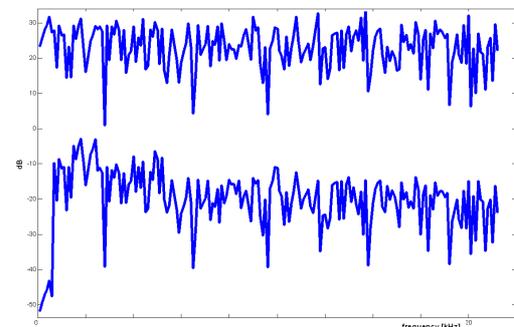

Fig. 7. Watermark signal one-stage correction procedure using proposed formula. Upper curve — signal spectrum before correction, bottom one — after correction procedure.

4. Comparison of computational effectiveness of the two proposed watermark signal correction procedures

The computational requirements for the method proposed in [2], according to the two-stages signal processing, are as follows:

Stage 1 is dedicated to frequency domain filtering and requires performance of the following procedures: one FFT of the white noise, one multiplication of the mask vector with the noise spectrum. Successful completion of

stage 1 results in the so-called coloured noise with spectral shape that resembles the curve of minimum masking threshold (JND level). Stage 2 consists in normalization of the signature to signal level and requires calculation of the minimum difference between WS (coloured noise) and the calculated LT_{min} level, as well as performance of one IFFT and one multiplication of WS in time domain with the correction coefficient.

The procedure proposed in this paper requires only one stage of processing of watermark signal consisting in one FFT, one multiplication of complex watermark signal by the correction coefficient $coef_{WM}$, and one IFFT to get filtered WS in time domain.

5. Experimental results

The basic metrics of the watermarked signal, using proposed WS correction method based on (5) formula, were calculated. The computed metrics are described by the following formulae:

$$\text{SNR} = \sum_n A_n^2 \Big/ \sum_n (A_n - A_n')^2$$

signal-to-noise ratio, (10)

$$\text{MSE} = \frac{1}{N} \sum_n |A_n - A_n'|$$

mean square error, (11)

$$\text{NMSE} = \sum_n (A_n - A_n')^2 \Big/ \sum_n A_n^2$$

normalized mean square error, (12)

$$\text{PSNR} = N \max_n A_n^2 \Big/ \sum_n (A_n - A_n')^2$$

peak signal to noise ratio, (13)

$$\text{AF} = 1 - \sum_n (A_n - A_n')^2 \Big/ \sum_n A_n^2$$

audio fidelity. (14)

In Table the metrics computed for corrected WS embedded into HS are presented.

TABLE

Basic metrics calculation for HS, WS after the correction procedure using formula (5), WM for audio *.wav format tracks.

	PHS	PWS	PWM	SMR	MSE	NMSE	PSNR	AF
Track 01	42.868	16.642	42.877	26.226	−38.253	−26.226	−17.682	0.9976
Track 02	40.218	15.474	40.233	24.744	−41.384	−24.744	−19.636	0.9966
Track 03	32.198	9.7676	32.221	22.43	−47.469	−22.43	−22.484	0.9943
Track 04	50.166	30.078	50.157	20.288	−26.567	−20.288	−29.878	0.9906
Track 05	48.246	29.485	48.278	18.831	−27.029	−18.831	−29.415	0.9869
Track 06	39.633	16.304	39.654	23.329	−42.928	−23.329	−17.473	0.9954
Track 07	37.774	12.968	37.788	24.807	−44.616	−24.807	−19.361	0.9967
Track 08	27.552	3.4705	27.569	24.081	−56.196	−24.081	−12.857	0.9961
Track 09	32.575	12.595	32.619	19.98	−43.849	−19.98	−18.972	0.99
Track 10	14.017	−8.0366	14.025	22.053	−64.895	−22.053	−7.8181	0.9938
Track 11	23.979	4.3197	24.023	19.659	−53.264	−19.659	−19.915	0.9892
Track 12	36.691	16.985	36.74	19.706	−40.252	−19.706	−21.25	0.9893
Track 13	40.218	15.474	40.233	24.744	−41.384	−24.744	−19.636	0.9966
Track 14	37.336	17.227	37.379	20.109	−39.631	−20.109	−23.375	0.9902
Track 15	46.338	21.718	46.351	24.656	−34.763	−24.656	−21.682	0.9966
Track 16	43.337	20.05	43.357	23.288	−36.395	−23.288	−20.049	0.9953
Track 17	37.842	17.785	37.885	20.057	−38.659	−20.057	−18.216	0.9901
Track 18	47.369	26.377	47.385	21.079	−30.155	−21.079	−26.29	0.9922
Track 19	43.937	22.218	43.971	21.733	−34.241	−21.733	−22.204	0.9933
Track 20	48.121	27.949	48.142	20.255	−28.578	−20.255	−27.866	0.9906
Track 21	47.205	29.383	47.256	17.861	−27.101	−17.861	−29.344	0.9836
Track 22	17.435	−4.5751	17.462	22.01	−55.169	−22.01	−12.526	0.9937
mean	37.957	16.075	37.982	21.905	−40.580	−21.905	−20.815	0.992
min	14.017	−8.036	14.025	17.861	−64.895	−26.226	−29.878	0.983
max	50.166	30.078	50.157	26.226	−26.567	−17.861	-7.818	0.997
std	9.919	10.369	9.918	2.307	10.249	2.307	5.602	0.0036

6. Conclusion

The proposed method of psychoacoustic correction of watermark signal proves to be efficient as against existing methods. In real-time processing mode the actual cost of computation constitutes the key element as regards implementation of the digital signal transformation algorithm. The proposed method of correction of the signal to JND level was implemented in real-time watermarking system in order to hide additional information in an acoustic signal. The method discussed above may be implemented in state-of-the-art applications used for watermarking acoustic content in order to fulfil the needs of, among others, Digital Rights Management Systems.

References

[1] Z. Piotrowski, I. Zagoździński, P. Gajewski, L. Nowosielski, in: *Proc. European DSP Education & Research Symp. EDERS 2008*, Texas Instruments, Tel Aviv 2008, p. 201.

[2] Y. Casuto, M. Lustig, S. Mizrachy, *Real-Time Digital Watermarking System for Audio Signals Using Perceptual Masking*, Signal and Image Processing Lab, Technion IIT, Haifa 2006.

[3] Z. Piotrowski, Polish Patent Office, P-367757, Wydawnictwo Urzędu Patentowego, Warsaw 2004.

[4] Z. Damijan, J. Wiciak, *Mol. Quant. Acoust.* **28**, 66 (2007).

[5] ISO CD11172-3, *Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s Part 3 Audio.*

[6] K. Pohlmann, *Principles of Digital Audio*, McGraw-Hill, New York 2000.

[7] J.G. Beerends, *Audion Quality Determination Based on Perceptual Measurement Techniques*, Kluwer Academic Press, Boston 1998, p. 1.

[8] T. Powalowski, Z. Trawinski, J. Wojcik, *Mol. Quant. Acoust.* **28**, 279 (2007).

[9] E. Zwicker, H. Fastl, *Psychoacoustics Facts and Models*, Springer-Verlag, Berlin Heidelberg 1990.