PEAQ-based Psychoacoustic Model for Perceptual Audio Coder

Xiaopeng Hu, Guiming He, and Xiaoping Zhou
(School of Computer, Wuhan University, Wuhan, 430072, P.R.China)
E-mail: hxp_9703@yahoo.com.cn

Abstract — **For the purpose of improving the coding efficiency, this paper attempts to combine psychoacoustic model for perceptual evaluation of audio quality in BS.1387 with perceptual audio coder. The principle of this new psychoacoustic model is analyzed in theory, and corresponding improvements are proposed to make it be effectively applied to actual audio coder. Both this new model and MPEG psychoacoustic model 2 are implemented in the latest AVS reference coder of China, and comparison of output masking parameter and subjective hearing test between the two models are conducted. Experimental results show that the proposed psychoacoustic model is feasible.**

Keywords — **Psychoacoustic model, PEAQ, Masking threshold, Audio coding**

I. INTRODUCTION

PERCEPTUAL audio coding exploit the properties of human auditory perception to make coding loss perceptually indistinguishable as far as possible. Therefore psychoacoustic model is the key module determining audio quality and compression efficiency significantly. Particularly, since the design of psychoacoustic model is not involved in consistency of decoding format, this part is not defined in various audio coding standards.

Since middle of 20th century, human beings have realized the importance of human auditory perception for audio compression[1][2][3], and the theory foundation of psychoacoustic model has been developed gradually. MPEG-1 ISO/IEC11172-3[4] was released in 1992, in which two psychoacoustic models were present. Afterwards both ISO/IEC 13818-3 (MPEG-2 Audio part)[5] and 13818-7 AAC[6] adopted the basic framework of psychoacoustic model 2 firstly introduced as a whole by J.Johnston in [9]. In stead of simulating physiological structure of human ear, psychoacoustic model 2 makes the output results approaching to the examination data.

PEAQ (Perceptual Evaluation of Audio Quality) is the objective measurements Recommendation Standard of perceived audio quality established by ITU in 1998, which is also called BS.1387. It utilizes software to simulate perceptual properties of human ear, then integrates multi-indices to evaluate subjective quality of test audio. PEAQ has been considered as the most effective objective measurement schemes of perceptual audio quality. A large number of objective evaluating software such as EQUAL [7] conform to this standard. PEAQ has two options: a basic version and an advanced version. Psychoacoustic model is its core module and design emphasis. Compared to MPEG, the psychoacoustic model in PEAQ takes into account perceptual properties of human ear more comprehensively.

The goal of this paper is to combine objective measurement of audio subjective quality with audio coder, and design a more perfect psychoacoustic model in audio coder to improve coding efficiency and coding quality. This paper is organized as follows. In next section, the fundamental of psychoacoustic model adopted in the basic version is introduced. In section 3, algorithm implementation of proposed psychoacoustic model in audio coder is presented. In section 4, comparison experiments between proposed psychoacoustic model and MPEG psychoacoutic model 2 are conducted. Finally, the major works of this paper are summarized and the further research and direction of improvement is presented.

II. PRINCIPLE OF PSYCHOACOUSTIC MODEL

Although both are based on examination and observation, psychoacoustic models in PEAQ and in MPEG have been represented as different approach to psychoacoustic model design. The former simulates and approximates its auditory properties according to the physiological structure of human ear, while the latter emphasizes particularly on fit between model output and examination observation, which does not distinguish the relationship between this phenomena and physiological structure of human ear.

In MPEG psychoacoustic model 2, SMR (ratio between signal energy and masking threshold) is determined by experiential value of examination observation. This model can be easy to correct and adjust, moreover fit between model and examination result is satisfied. However, restricted in examination condition, so far most of examination results obtained were based on pure tone, only a few examination used dual tone[8]. In fact, a complex signal is combination of infinite sinusoids, which has much

more complicated masking phenomena than pure tone. This kind of model only fit the examination data to simple signals and is hard to be extended to complex signals in theory. Therefore the maximum masking threshold could not be obtained in this model unless the examination measure is improved.

Basic version of PEAQ adopted a design approach different from MPEG audio standard. It attempts to combine physiological structure of human ear with masking effect of simple signal represented from examination to find the inherent consequence, and then use mathematic model to emulate the structure of human ear. Fig. 1 illustrates the block diagram of this design[10], in which the function of outer/middle ear, inner ear, and audio perception related nerve cell and brain are emulated. This kind of psychoacoustic model could be extended conveniently to acoustic masking of complex signal.

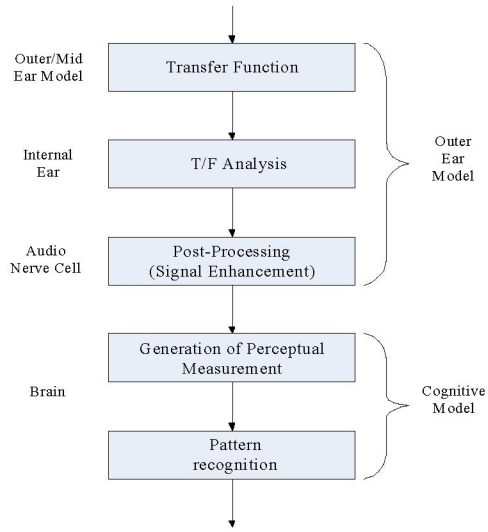

Fig. 1. Perceptual framework of psychoacoustic model based on PEAQ basic version

PEAQ was initially designed to measure audio quality, without special consideration for requirements of audio coder, such as window switching, unification of critical band scale, estimation of masking properties on transient signal, etc. Thus, we must take into account these requirements and correct the psychoacoustic model of PEAQ basic version so as that it could be used in audio coder.

III. ALGORITHMS IMPLEMENTATION OF MODEL

This section will discuss in detail the implementation of the psychoacoustic model of PEAQ basic version in audio coder. Parameter adjustments in implementation are all relative to the latest China AVS (Audio Video Coding Standard). Since both AVS audio standard and MPEG audio standard have the same partition of critical band, these parameters are also applicable to MPEG audio standard.

Fig. 2 illustrates detailed flow chart of implementation. Masking parameters are calculated frame by frame. Different from the original psychoacoustic model of PEAQ, this improved model used in audio coder takes into account the cases of both long window and short window.

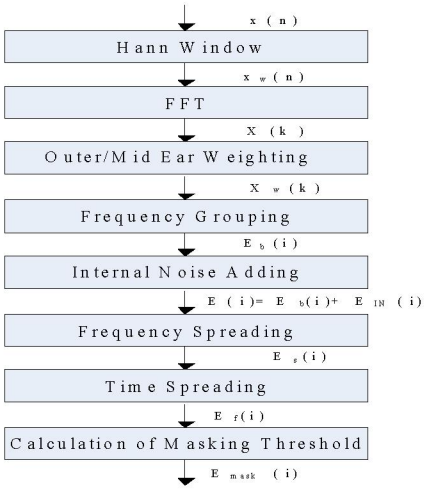

Fig. 2. Flow chart of implementation of psychoacoustic model of PEAQ basic version in audio coder

A. Windowing and Time-Frequency Transform

Firstly, input frame is windowed and then transformed into frequency domain via FFT. $x_w(n)$ is obtained by adding normalized Hann window to $x(n)$, function of Hann window is given by

$$h(n,N_F)=\begin{cases}\frac{1}{2}\left[1-\cos\left(\frac{2\pi n}{N_F-1}\right)\right] & 0\le n\le N_F-1,\\ 0 & otherwise\end{cases} \quad (1)$$

Where N_F is frame length of input signal, the value of N_F is assumed to be 2048 for long window, while N_F is 256 for short window.

Because this window function changes the original signal energy, it must be normalized [12]. The actual normalized window function is $h_w(n,N_F)=\sqrt{\frac{8}{3}}h(n,N_F)$, here we ignore the normalizing factor $\sqrt{\frac{N_F-1}{N_F}}$. Compared to the window function used in MPEG-2, here energy change along with windowing is calculated more precisely.

Then $x_w(n)$ is converted into frequency domain signal $X(k)$ using a Discrete Fourier Transform, as shown in (2).

$$X(k)=\frac{1}{N_F}\sum_{n=0}^{N_F-1}x_w(n)e^{-j2\pi nk/N_F} \quad (2)$$

B. Outer/Middle Ear Transfer Model

Human ear's sensitivity varies along with frequency. Signal is weighted by Emulating outer/middle ear transfer model, so as to be consistent with human ear sensitivity curve. There is no such consideration in MPEG standards.

Since auditory perception of human ear is pertinent to loudness, loudness correction for $X(k)$ must be done before weighting. Correction factor G_L is defined as [12] :

$$G_L = \frac{10^{L_p/20}}{\gamma(f_c)\sqrt{\frac{8}{3}\frac{A_{\max}}{4}\frac{N_F-1}{N_F}}} \quad (3)$$

Where L_P is the sound pressure Level (SPL) corresponding to sine wave with maximum amplitude. A_{max} is the max amplitude. $\gamma(f_c)$ is center frequency related factor altered from 0.84 to 1.

Weighting transfer function is defined as

$$A_{dB}(f_{Hz}) = -2.184(f/1000)^{-0.8} + 6.5e^{-0.6(f/1000-3.3)^2} - 0.001(f/1000)^{3.6} \quad (4)$$

$$W(k) = W(kF_s/N_F) = 10^{A_{dB}(kF_s/N_F)/20} \qquad 0 \le k \le N_F/2$$

Thus spectrum energy after weighting is $|X_w(k)|^2 = G_L^2 W^2(k)|X(k)|^2 \quad (0 \le k \le N_F/2)$.

C. Frequency Grouping

In order to be consistent with the critical band partition of MPEG audio coder, frequency domain must be regrouped according to new critical band. Then Energy E_b is calculated for each band. Since critical band partition is defined by DFT bins, and energy of each DFT bin represents continuous frequency energy distributed over bin width. Thus specific processing for maximum and minimum frequency is necessary. The distribution width of energy for bin k in band i is given by (5), and the energy of the band i is given by (6), where F_s is sampling rate, $f_l(i), f_u(i), k_l(i), k_u(i)$ are upper limit, lower limit and bin in band i respectively, and $U_l(i), U_h(i)$ are the energy distribution corresponding to upper limit and lower limit respectively.

$$U(i,k) = \frac{\max\left[0, \min(f_u(i), \frac{2k+1}{2}\frac{F_s}{N_F}) - \max(f_l(i), \frac{2k-1}{2}\frac{F_s}{N_F})\right]}{\frac{F_s}{N_F}} \quad (5)$$

$$E_b(i) = U_l(i)|X_w(k_l(i))|^2 + \sum_{k=k_l(i)+1}^{k_u(i)-1}|X_w(k)|^2 + U_h(i)|X_w(k_u(i))|^2 \quad (6)$$

D. Adding of Internal Noise

Internal noise caused by blood flow within human ear, is added to the input signal. This noise is simulated on the base of physiological structure of human ear, whereas it is represented as absolute hearing threshold in MPEG psychoacoustic model 2. Internal noise function is given by

$$E_{IN}(i) = 10^{0.1456(f/1000)^{-0.8}} \quad (7)$$

E. Energy Spreading of Frequency Domain

Energy spreading of frequency domain is the important phenomenon of auditory properties. In MPEG psychoacoustic model 2, energy spreading curve is added linearly with independent of loudness. But it was discovered that energy spreading curve at a frequency point varied according to its loudness [13], and nonlinear adding was much coincidental to auditory properties of human ear [14]. Thereby, the loudness is taken effect into energy spreading curve, as shown in (8).

$$S_{dB}(i,l,E) = \begin{cases} 27(i-l) & i \le l \\ [-24 - \frac{230}{f_c(l)} + 2\log_{10}(E)](i-l) & i \ge l \end{cases} \quad (8)$$

Where i, l are the center frequencies respectively of the l_{th} excitation band and i_{th} spread band in unit of bark. $f_c(l)$ is center frequency of the l_{th} excitation band in unit of Hz. $log_{10}(E)$ relates loudness and spreading function. In addition, (8) is required to be normalized, as shown in (9).

$$S(i,l,E) = \begin{cases} \frac{1}{A(l,E)} 10^{2.7(i-l)} & i \le l \\ \frac{1}{A(l,E)} (10^{(-2.4-23/f_c(l))E^{0.2}})^{i-l} & i \ge l \end{cases} \quad (9)$$

Where $A(l,E)$ is normalizing factor, $A(l,E) = 10^{0.1\sum_i S_{dB}(i,l,E)}$.

Then we make use of nonlinear adding to calculate distribution curve spread energy as follows.

$$E_s(i) = \frac{1}{B_s(i)}\left(\sum_{l=0}^{N_c-1}(E(l)S(i,l,E))^{0.4}\right)^{2.5} \quad (10)$$

$$B_s(i) = \left(\sum_{l=0}^{N_c-1}(S(i,l,E))^{0.4}\right)^{2.5} \quad (11)$$

Where N_c is the number of band groups. $B_s(i)$ is normalizing factor, which is the nonlinear sum of spreading functions of the i_{th} band.

F. Time Domain Spreading

Time domain spreading concerns temporal masking effects. As for auditory properties, there are pre- and post-masking effects in time domain besides frequency domain masking [8], while such temporal masking effects are not involved in psychoacoustic model of MPEG. In this model, temporal masking effect is considered by means of first-order smooth filtering.

Time domain spreading is calculated in unit of band partitioned above as shown below:

$$E_f(i,n) = \max(\alpha(i)E_f(i,n-1) + (1-\alpha(i)E_s(i,n), E_s(i,n)) \quad (12)$$

Where $E_f(i,n)$ is energy of time domain spreading of band n at frame i. $\alpha(i)$ is the decaying coefficient of band i controlled by time constant, as shown in (13):

$$\alpha(i) = \exp(-\frac{1}{F_{ss}\tau(i)})$$
$$\tau(i) = \tau_{\min} + \frac{100}{f_c(i)}(\tau_{100} - \tau_{\min}) \quad (13)$$

Where $F_{ss} = F_s/(N_F/2)$ is the frame rate, F_s is the sampling rate. τ_{100} and $\tau_{\min}$ are time decaying constants.

After tuning for AVS audio coder, $\tau_{100} = 0.03$ and $\tau_{\min} = 0.008$ for long window, while $\tau_{100} = 0.003$ and $\tau_{\min} = 0.001$ for short window.

G. Calculation of Masking Parameters

After processing described above, signal energy curve concerned both temporal masking and simultaneous masking is obtained. The energy distribution of masking threshold can be calculated utilizing masking curve. Masking property and masking energy of each critical band are given by (14).

$$E_{mask(dB)}(i) = E_{f(dB)}(i) - m_{dB}(i)$$
$$m_{dB}(z) = \begin{cases} 3 & z \le z_L + 12 \\ 0.25(z - z_L) & z > z_L + 12 \end{cases} \quad (14)$$

Where z is center frequency of the i_{th} band in unit of bark , and $z_L = 0.8594$.

IV. Experimental Results

This new psychoacoustic model has been implemented in reference coder of AVS audio standard[11]. Fig. 3 illustrates the distribution curve of masking threshold for tone at 1 kHz calculated by utilizing the two different psychoacoustic models respectively. Fig. 4 shows an examination result of masking properties for signal at 1 kHz given by Zwicker in 1976 [15], in which the lowest curve is absolute hearing threshold. From Fig. 3, we can observe that distribution of masking curve in new psychoacoustic model matches better with that in Fig. 4. Whereas masking curve in MPEG psychoacoustic model 2 is more sensitive to distribution of spectrum energy, and the slope of pre-energy spreading at peak energy is nearly equal to that of post- energy spreading. This is different from observed acoustic phenomena. In contrast, the new psychoacoustic model represents preferably the hearing property that pre-energy spreading effect is greater than post- energy spreading effect.

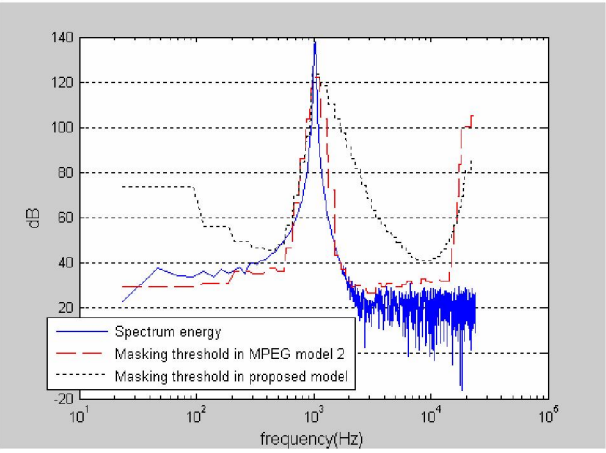

Fig. 3. distribution curve of masking threshold for tone at 1 kHz

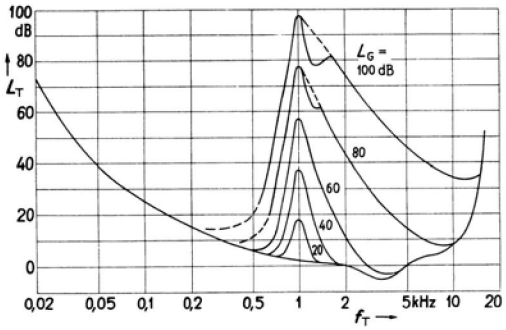

Fig. 4. Masking curve for masker at 1 kHz [15]

Fig. 5 shows distribution of masking threshold for a real complex signal calculated by utilizing the two different psychoacoustic models respectively. Similarly, comparing with MPEG psychoacoustic model 2, wave crest of spectral energy in our new psychoacoustic model has more intensive masking effect against energy trough in high frequency band. As for masking effect against energy trough in low frequency band, both models are similar. The psychoacoustic model designed in this paper is more practical and more coincident with hearing observation and examination phenomena than MPEG psychoacoustic model 2.

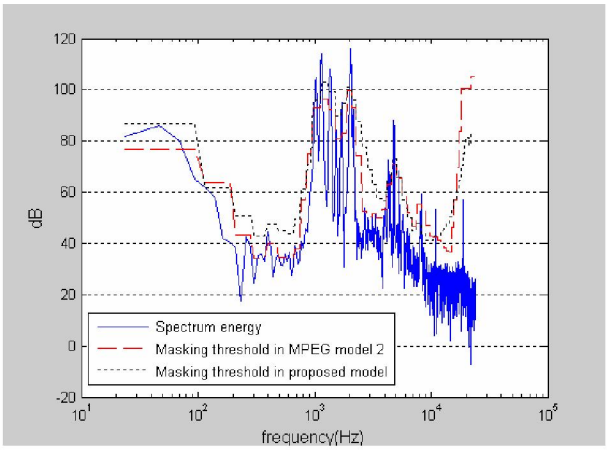

Fig. 5. distribution curve of masking threshold for complex signal

Of course, we also found during experiments that this new psychoacoustic model is not mature enough so far. Since this proposed model integrates directly the psychoacoustic model in PEAQ, it can achieve quite high score of objective evaluation on PEAQ based testing software, but this score is meaningless. So subjective listening test is necessary. We discovered that the subjective quality of new model was not better than that of MPEG psychoacoustic model 2. On the one hand this is due to that analysis of masking properties for short window and process during window transition need to be improved further; On the other hand, it is necessary to adjust original rate control and bit allocation strategy due to changes of psychoacoustic model. Subjective quality has been

remarkably improved after adjustment. For test sequences with few window switching, the new model can provide better quality than MPEG psychoacoustic model 2. All these examination results demonstrate the potential prospect of this new psychoacoustic model.

V. CONCLUSION

In this paper, we discussed application of a new psychoacoustic model based on PEAQ to perceptual audio coder. We first described in detail the principles of this new psychoacoustic model, then presented the implementation of algorithms. In order to meet requirements of audio coder, we proposed improving methods as well as corresponding parameters setting. Compared to hearing examination and observation, this new model is proved to be practical. Due to consideration of more acoustic properties, the proposed psychoacoustic model can characterize the auditory properties of human ear more precisely than MPEG psychoacoustic model 2. Thus the work in this paper provides new approach to audio coder's optimization. In addition, the results of subjective listening test show the proposed model is faced with some problems for short window and window transition, which is one of our further research area.

REFERENCES

[1] Fletcher, H., "Loudness, masking, and their relation to the hearing process and the problem of noise measurement", *J. Acoust. Soc. Am.,* 9: 275-293, 1938.

[2] Fletcher, H. and Munson, W.A., "Relation between loudness and masking", *J. Acoust. Soc. Am.,* 9: 1-10, 1937.

[3] Green, D.M. and Swets, J.A., "Signal detection theory and psychophysics", John Wiley & Sons, New York, 1966.

[4] ISO/IEC JTC1/SC29, "Information technology—Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s–IS 11172 (Part 3, Audio)", 1992.

[5] ISO/IEC JTC1/SC29, 'Information technology—Generic coding of moving pictures and associated audio information–IS 13818 (Part 3, Audio)", 1994.

[6] ITU-R Recommendation BS.1387, "Method for objective measurements of perceived audio quality", Dec. 1998.

[7] A. Lerch, EAQUAL – Evaluation of Audio Quality, Software repository: http://sourceforge.net/projects/eaqual, Jan. 2002.

[8] J. Hall, "Auditory psychophysics for coding applications," CRC/IEEE DSP Handbook, 1996.

[9] Johnston, J. D. "Estimation of perceptual entropy using noise masking criteria" ICASSP, A1.9, pp 2524-2527, 1998

[10] Dipl. Ing. Thilo Thiede, "Perceptual audio quality assessment using a non-linear filterbank", PH.D. thesis, 1999

[11] AVS N1201 "Information technology: Advanced Audio-Video Coding Part 3: Audio (FCD)", 2005

[12] P. Kabal, "An examination and interpretation of ITU-R BS.1387: perceptual evaluation of audio quality", TSP Lab Technical Report , 2002

[13] Terhardt, E., "Calculating virtual pitch". Hearing Research, Vol. 1, 1979, pp. 155-182.

[14] Humes,L. E., Jesteadt, W., "Model of the additivity of masking", Journal of the Acoustical Society of America, Vol. 85 (3), March 1989, pp. 1285-1294.

[15] Zwicker, E., Feldkeller, R., "Das Ohr als Nachrichtenempfänger". Stuttgart: Hirzel Verlag, 1967.