Filter Bank Analysis

VLADISLAV SKORPIL, ABDULHAKIM ABUZAHU
Department of Telekommunications,
Brno University of Technology,
Purkynova 118, 612 00 Brno,
CZECH REPUBLIC,
http://www.vutbr.cz/

Abstract: - In the MPEG audio coding, the input signal is divided into 32 subbands of equal bandwidth. In each subband, the signal is scaled and quantized in order to keep the quantizing noise below the masking curve. The result of the encoding is a series of scale factors, quantizer information and coded samples for each subband. According to analysis of subband filter flowchart the filter does a time to frequency mapping.

Key-Words: - MPEG, Filtr bank, subband

1 Introduction

Utilization of psychoacoustics knowledge has fundamental meaning for the lossy coding. Psychoacoustics is the scientific branch, which deals with research of human hearing. It studies principle of hearing, time frequency characteristics of ear, energetic viewpoints (threshold in quite and threshold of pain) and final interpretation depend on used measuring practising perception tests whose procedure and final interpretation depends on used measuring method. Measuring methods differ in psychoacoustics problem types.

Often used methods are Setting method – tested person s.ets e.g. level of pure tone as this tone is audible - Cueing method – tested person determines whether tone level increases or decreases - Yes – No method – tested person decides whether relevant signal is audible or not.)

An audio signal is usually in the analogy form (signal from microphone or another type of transducer) and has to be transferred into the digital form. Whole process begins with sampling must be fulfiled. It says that every time behaviour whose spectrum has maximum frequency fm can be expressed by its samples taken from the continuous behaviour in moments with period

$$T_s=1/fs=1/2fm \qquad (1)$$

Where fm is maximum frequancy and fs is the sampling frequency and T_s is time interval sampling. Sampling frequencies usually used in practice starts from 8 kHz for speech signal and ends with 96 kHz some time even more, 192 kHz for high-quality audio signals. Sampling frequency of 44.1 kHz is the most frequently used frequency for quality audio signals (sampling frequency used for recording of well-known CD audio disc).

The next operation attached to conversion from analogy to digital form is so-called quantization that transfers signal with continuous amplitude to the signal with final number of discrete values [2]. Whole range of input signals is divided into final number of quantization steps with different size. If it express the discrete quantized values in binary code, it discuss so-called PCM- pulse code modulation. Length of the assigned coding group depends on the number of quantizer discrete levels and it is in the range from n=8 bits (256 levels) to n=24 bits (16777216 levels). If size of the quantization steps is constant, we discuss so-called linear quantizing.

Quantization process causes irrecoverable signal distortion that is caused by so-called quantization noise. For the maximum signal-to-noise ratio SNR_{max} (we think the maximum amplitude of input signal given by quantizer parameters) with the linear quantizing and the PCM modulation following equation holds approximately [3]

$SNR_{\max}$ =6*n±C (dB) (2)

It must reserve space to avoid the over-excitation and the SNR value mustn't be lower than certain limit. If it think one channel with fs = 44.1 kHz and n =16 bits, the equation holds for resulting bit rate:

vp = fs *n = 44100 * 16 = 705600 bit/s (3)

The value is double for the stereophonic transmission. That is why various coding methods different from the PCM are used, which are designed for decreasing bit rate with unchanged quality.

Signal compression is a process that decrease amount of data in order in to it's storage or transmission. The compression efficiency is given by so-called compression ratio P_k :

$$P_k = \frac{K_0}{K_k} \qquad (4)$$

where K_0 is the original data amount and K_k is the data amount after compression. "Data amount" term relates to " bit rate" term if compressed signal is about to be transmited. High compression ratio causes low bit rate needed for transmission of the same information in the same time-in other words, high P_k is associated with strong reduction of the bit (transfer) rate.

Compression techniques can be separated into two groups according to principle of their operation- lossless compression and lossy compression. Backup of the computer programs and data is the best-know utilization of the lossless compression algorithms. In this case every loss of information causes the crash. These algorithms remove certain redundancy algorithms from the compressed data and they generally have low compression ratio P_k [4]. These algorithms are used rarely for the audio compression.

So-called lossy algorithms have generally much higher ratio P_k but at the expense of loss of the information portion. Only these algorithms are efficient for the effective audio compression. Loss of data portion is not

Bit number most often used today is n =16 (even 24), However the dynamic range is markedly smaller (e.g. 50 harmful because the audio signal has redundancy and irrelevancy whose omission dose not changes the whole subjective perception .

Great number of the compression techniques based on pieces of knowledge about perception has been designed which reach high P_k ratio with selectable subjective degradation of the signal quality. They are suitable for the digital television and sound broadcasting (e.g. all-European TV system DVB- digital Video Broadcasting), ISDN videoconferencing, Internet and data storage on media of DVD type (Digital Versatile Disc) or MD type (Minidisk)

In 1948 Shannon introduced the source theory where accuracy of signal source representation is the main criteria [16]. Bit rate – information amount necessary for representation of the signal characteristics – depends on the distortion that it allow for their transmission, i.e. the accuracy that it want to use for the signal source specification. Well-known equation holds for information capacity of the ideal frequency limited channel :

C = W$\log_2$ (1 + P/N)[bit/s] (5)

2 Hearing Area

Utilization of psychoacoustics knowledge has fundamental meaning for the lossy coding. Psychoacoustics is the scientific branch, which deals with research of human hearing. It studies principle of hearing, time frequency characteristics of ear, energetic viewpoints (threshold in quite and threshold of pain) and final interpretation depend on used measuring practising perception tests whose procedure and final interpretation depends on used measuring method. Measuring methods differ in psychoacoustics problem types[10].

Often used methods are:

Setting method – tested person sets e.g. level of pure tone as this tone is audible

Cueing method – tested person determines whether tone level increases or decreases

Yes – No method – tested person decides whether relevant signal is audible or not

The hearing area is a plane in which audible sound can be displayed. The normal form of the hearing area is usually plotted in as frequency logarithmic scale on the X-axis and sound pressure level in dB linear scale on Y-axis. In the usual display of human hearing area is the sound intensity measured in ($W/m^{2)}$, while the sound pressure is measured in (Pa). The sound level is given for free-field condition relative to $2X10^{-5}$ Pa. While the sound pressure level is plotted relative to 10^{-12} W/m^2 .

The actual hearing area limits, that, hold for pure tones in steady state condition, lies between the thresholds in quit and the threshold of pain . The components of music encompass a larger distribution in the hearing area are shown in figure by hatching. The high level border which known as the limit of damage risk that's very important in everyday life: reached at quite high sound pressure at very low frequencies is also indicated in the figure by the thin dotted line. area between threshold in quiet and threshold of pain. Also indicated are the areas encompassed by music and speech, and the limit of damage risk..

The threshold in quiet indicates as a function of frequency the sound pressure level of a pure tone that is just audible. The threshold in quiet can be measured for both experienced and inexperienced subjects precisely and quickly by Bèkèsy-tracking method, where the decrease or increase of the sound pressure level is recorded as a function of time, and as the frequency also is slowly changing it could be recorded as a function of frequency too.

At low frequencies, threshold in quiet requires relatively high sound pressure level reaching about 40 dB at 50 Hz. The thresholds in quiet for frequencies between 0.5 and 2 kHz remains almost independent on the frequency are indicated in the Fig. 3.2. In many cases threshold in quiet shows some excursions or small humps.

3 Masking

Masking plays a very important role in real life: for a conversation on the pavements of a quiet street, for example, little speech power is necessary for the speakers to understand each other. However, if a loud truck passes by, our conversation is severely disturbed: if we keep our speech power constant, our partner can no longer hear us, or cannot hear us as well.

Likewise, in lossy compression algorithms, masking of signals plays a large role. We can most easily demonstrate the masking process when two pure tones are being listened to. In practice, however, this occurs but rarely. Therefore, we mainly need to describe the masking phenomenon when complex sounds (music, conversations) are being listened to.

The aim of all tests is to set the masking threshold (curve), which is a value for sound pressure of the masked sound at which it is audible over the masking sound.

When viewing masking as a whole, one must take into account the temporal viewpoint, which comes into play for sounds with a strong temporal dependence, such as e.g. music and speech. The phenomena in question are pre- stimulus masking (also known as pre- masking and backward masking) and post-stimulus masking (also known as post- masking and forward masking). Backward (pre) masking occurs prior to masker onset and lasts only a few ms; forward (post) masking may persist for more than 100 ms after masker removal. So far it has considered the simultaneous action of the masking and masked signal, with each having a length exceeding 200 ms. A masking effect also occurs, however, in cases where a masked shorter- lasting signal comes before a masking, longer-lasting signal. Masking likewise occurs in cases where a masked signal appears after a longer, masking signal

4 Critical Band

Fletcher presented the concept of critical bands. Masking is dependent on the frequency and level of the acoustic pressure of the masking tone and masked tone. This means that masking curves have differing shapes for different starting parameters. Therefore, a search began for a solution where they would have the same shape (i.e. the same width) throughout the entire audible band if the same masking level were used throughout.

Let us imagine pure tone masked by white noise. This tone will be masked by the noise, but only a certain portion of the noise will participate in the masking. The breadth of this effective frequency band of noise is called the critical width of the noise band for the given tone . The frequency axis is divided into 24 critical bands with a size of 1 Bark. (The name of the Bark unit is derived from the name of the German scientist Barkhausen.) The result is a linear frequency scale divided into Bark units that have differing sizes in Hz at different places throughout the frequency axis.

In lossy compression, one needs to determine how much the post-compression signal can differ from the original, so that this difference can be guaranteed as inaudible. As the distortion allowed is increased, the achievable compression ratio also rises. The task is a matter of finding an optimal compression level that captures the properties of human hearing. The so-called psycho acoustic model serves this end. This model determines, on the basis of signal properties, how large an error signal (generally this means quantization noise) will be masked and therefore inaudible for listeners. The quality of a psycho acoustic model can be evaluated by how well the prediction for various input signals approaches reality. it can, for example, expect that a model that factors is not in only frequency, but also time during masking will be higher in quality.

To demonstrate the use of a masking effect, it will use the following example. Three harmonic signals with frequencies of 0. 25, 1, and 4 kHz with a level of L = 60 dB make up an input, and thus masking, signal.. We must note that this is masking in the signals' frequency spectrum, as that is the only place where individual simultaneously playing sounds can be "separated" from each other. The result of compression also depends on **the** way the transfer from the time domain to the frequency domain is performed. In our case, a division into 32 equally broad bands is used. The masking curves thus unambiguously prescribe the maximum level that the masked

Band No.	Centre Freq. (Hz)	Bandwidth (Hz)
1	50	-100
2	150	100-200
3	250	200-300
4	350	300-400
5	450	400-510
6	570	510-630
7	700	630-770
8	840	770-920
9	1000	920-1080
10	1175	1080-1270
11	1370	1270-1480
12	1600	1480-1720
13	1850	1720-2000
14	2150	2000-2320
15	2500	2320-2700
16	2900	2700-3150
17	3400	3150-3700
18	4000	3700-4400
19	4800	4400-5300
20	5800	5300-6400
21	7000	6400-7700
22	8500	7700-9500
23	10500	9500-12000
24	13500	12000-15500
25	19500	15500-

Tab.1 The division of the frequency axis into critical/band for a collection of 25 critical bandwidth

signal for the given frequency can have without being audible. This is then used during audio compression, where the "noise spectrum" of the quantization noise is shaped according to the strongest spectral elements of the signal to be compressed. The maximum quantization noise in a given frequency interval yields a minimum number of bits needed for quantization of samples of the signal being transferred to the frequency spectrum. The possibility to reduce the number of bits is judged separately in each band. The narrower the sub bands are in frequency, the higher the achievable reduction masking effects are applied better.

5 Test signal

The outputs of the software calculation in MATLAB are displayed here in pictures. These outputs are the outcomes of the input-test signal. This signal was connected at the input in MATLAB programming. With these signals, a lot of main parts of compression principles are introduced. In addition, it will be introduced on real program in future, which is possible to run in real time, but now will be used step by step in developing software. The input chosen signal consists of audio signal which take it from CD. The name of audio signal (gedeon.wav) shown in Figure 1. For program, the same 16bit input signal was demonstrated in mono quality, 44.1 kHz sampling frequency.

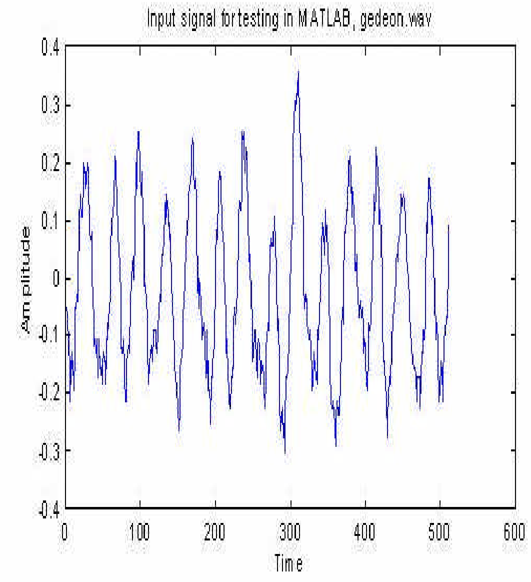

Fig.1 Input test signal for programming in MATLAB

6 Filter Bank Analysis

In the MPEG audio coding, the input signal is divided into 32 subbands of equal bandwidth. In each subband, the signal is scaled and quantized in order to keep the quantizing noise

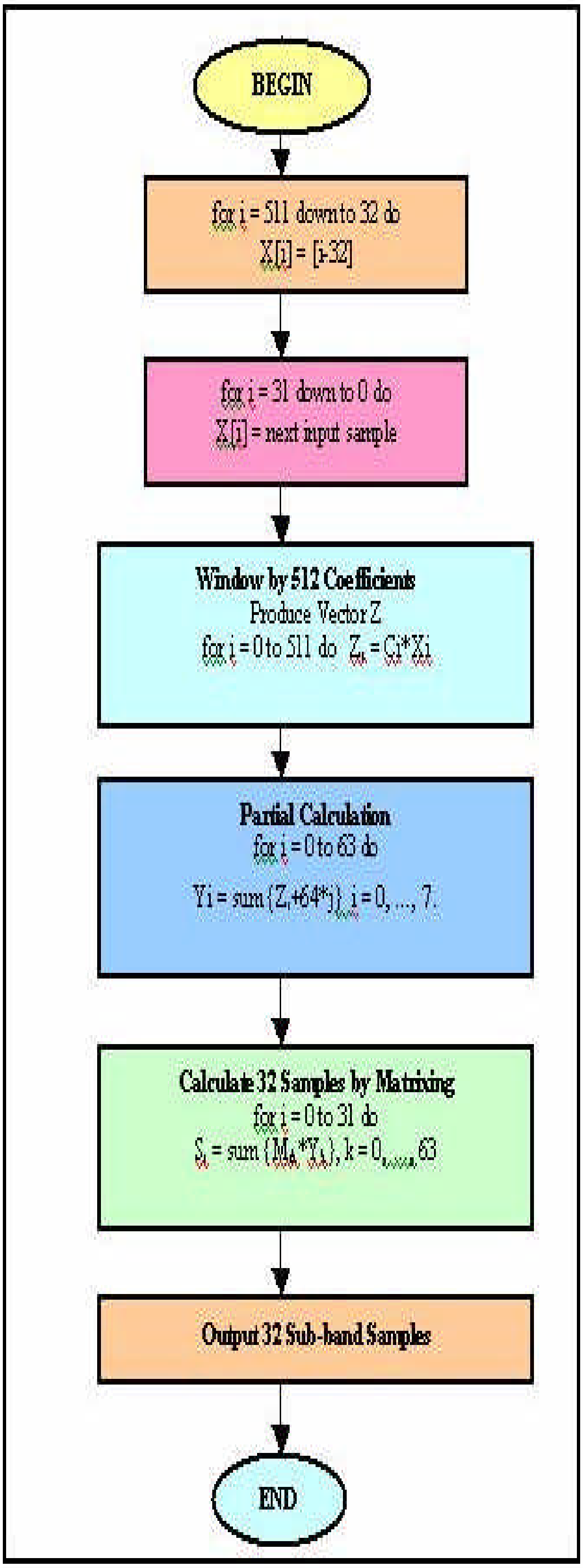

Fig.2 Analysis of subband filter flowchart

below the masking curve. The result of the encoding is a series of scale factors, quantizer information and coded samples for each subband. According to analysis of subband filter flowchart (Figure 2), the filter does a time to frequency mapping. The 32 subband polyphase filter presents optimized characteristics with respect to a performance complexity ratio. The filter, of order 511, with side lobe rejection better than 96 dB, is a compromise between two competing specifications of the filter response: the spectral resolution and the transient impulse response (TIR). The spectral resolution of the filter bank is important, because it corresponds to the critical bands found in the human ear.

7 Conclusion

The information yielded by the filter, when compared with empirical perceptual data, facilitates the reduction of the bit information by eliminating masked, unnoticed spectra, and reduces the number of bits allocated for spectra carrying low amounts of information. The time-frequency mapping of the filter allows a reasonable emulation of the critical bands of the ear, which correspond to a width of about 100 Hz in frequencies below 500 Hz, and width of about 20 % of the center frequency at higher frequencies.analysis-synthesis prototype implementation of the filter provides efficient performance.

References:

[1] ISO/IEC 13818-7 Information Technology – Generic coding of moving pictures and Associated Audio Information – part 7: Advanced Audio Coding (AAC), 1996, 131s.

[2] BOSI M. and RANDENBURG, K. ISO/IEC MPEG-2, Advanced Audio Coding. Journal of the Audio Engineering Society, Vol.45, 1997, No. 10,PP. 789-814

[3] SERANTES, C. PENA A., Prelcic N. A Fast Noise-scaling Algorithm for Uniform Quantization in Audio Coding schemes. IEEE Proceedings, 1997 , PP 339-342

[4].ABUZAHU,A. Digital Audio Compression. Ph.D. Thesis, BUT, Brno 2003

Acknowledgement:

This research was supported by the grants: No 102/03/0434 Limits for broad-band signal transmission on the twisted pairs and other system co-existence. The Grant Agency of the Czech Republic (GACR)

No 102/03/0260 Development of network communication application programming interface for new generation of mobile and wireless terminals. The Grant Agency of the Czech Republic (GACR)

No 102/03/0560 New methods for location and verification of compliance of quality of service in new generation networks. The Grant Agency of the Czech Republic (GACR)

No MS 1850022 Research of communication systems and technologies (Research design) Grant 2811 F1 Advanced Technology of Transport Networks in Education (grant of the Czech Ministry of Education, Youth and Sports)

Grant 3112 F1 Inovation of Education of Last Mile Data Transmission (grant of the Czech Ministry of Education, Youth and Sports)