euronoise

Wavelet filter bank based wide-band audio coder

Jan Nováček

Czech Technical University, Faculty of Electrical Engineering, Technicka 2, 16627 Prague, Czech Republic
novacj1@fel.cvut.cz

New system for wide-band audio signals compression is presented. The system is based on a two stage wavelet filter bank principle. Special psychoacoustic model was designed for the system. Data compression is nowadays very topical issue. Due to the development of information technologies more sophisticated compression algorithms may be designed. Most of the present compression systems, e.g. MPEG, ATRAC or Vorbis, use a simple filter bank or DCT transformation. Both methods suffer from the origin of undesirable artefacts which are disturbing for the sound perception. Wavelet filter bank is used to suppress these artefacts and to increase data compression. Frequency dependant windowing function is used for improve of frequency resolution at low frequencies along with faster response at high frequencies. Wavelet transformation based coding algorithm can improve temporal masking effects over conventional perceptual coding algorithms for data compression.

1 Introduction

Coding of audio signals is a current topic in acoustics. Thanks to development and increase of computational power of processors more sophisticated coding algorithms may be used even in real-time digital signals processing. Most of the present wide-band audio coders use simple Pseudo-Quadrature Mirror Filter (PQMF) filter banks or Modified Discrete Cosine Transform (MDCT). Artefacts connected with both medhods impact disturbing.

Presented wide-band audio coder uses wavelet filter bank which should suppress these artefacts. Based on Generic audio coder [1], ISO/IEC MPEG-1 audio standard [2], Wavelet filter bank based psychoacoustic model [3] and Wavelet filter bank based sound signal analysis [4] was designed novel wavelet filter bank based wide-band audio coder. Block diagram of the designed coder is shown in Fig. 1.

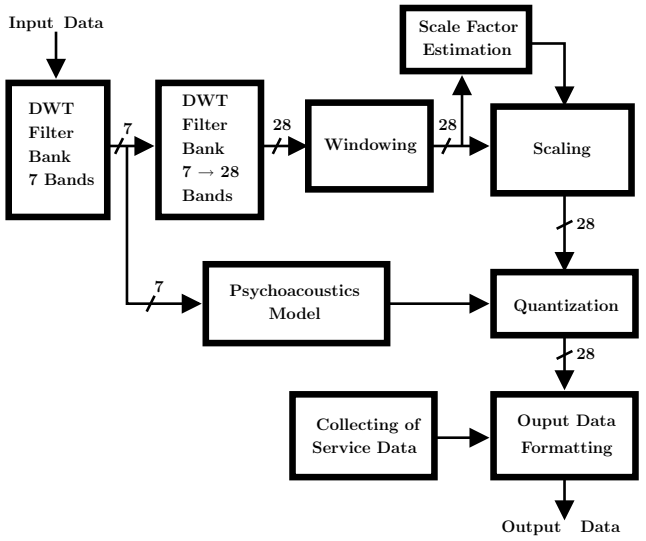

Figure 1: Block diagram of designed coder.

All the system blocks are described in the following chapters.

2 Wavelet Filter Bank

Designed wide-band audio coder is based on two-stages wavelet filter bank (WFB) [5]. In comparison with PQMF filter bank [6] used in ISO/IEC MPEG-1 audio standard [2], which has equal division of sub-bands, wavelet filter bank used in the designed system has non-equal sub-band division. Complete decomposition tree can be seen in Fig. 2. By the solid line is shown the first stage of the bank, the second stage is shown by the dashed line. This separation of the bank stages is because the output of the first stage wavelet filter bank is connected to the psychoacoustic model (see Fig. 1). In Tab. 1 can be seen description of 28 output frequency sub-bands of the wavelet filter bank calculated for 44.1 kHz sampling frequency. Note that WFB outputs are critically sampled because of the decimation therefore sampling frequencies varies along sub-bands. The theory of wavelet filter banks and wavelet decomposition can be found in [5].

Band No. [-]	Sampling frequency [Hz]	Window length [ms]
1 .. 4	5512	5.8
5 .. 8	2756	11.6
9 .. 12	1378	23.2
13 .. 16	689	46.4
17 .. 20	344	92.9
21 .. 28	172	185.8

Table 1: Wavelet filter bank sub-bands description.

As mentioned above the wavelet filter bank has two outputs. The first output of the wavelet filter bank is a decomposed signal for the psychoacoustic model. This output has seven sub-bands. The second output is a decomposed signal for the next coder blocks. The second output has 28 sub-bands.

3 Psychoacoustic Model

Based on ISO/IEC MPEG-1 Psychoacoustic Analysis Model 1 of MPEG 1 audio standard [2] described in [6] and [7] and Wavelet filter bank based sound signal analysis [4] was derived new psychoacoustic model for this wavelet filter bank based wide-band audio coder. Basic ideas of the psychoacoustic model are described in [3]. Block diagram of the designed psychoacoustic model is shown in Fig. 3.

As can be seen in Fig. 1 an input of the psychoacoustic model is the output of the first-stage wavelet filter bank of the audio coder. In Fig. 3 is this wavelet filter bank shown by the dashed box. An input signal of the psychoacoustic model consists of seven sub-band critically sampled signals, each may have different sampling frequency. These seven signals are analysed by the psychoacoustic model. Output of the model is a frequency dependant global masking threshold or signal to

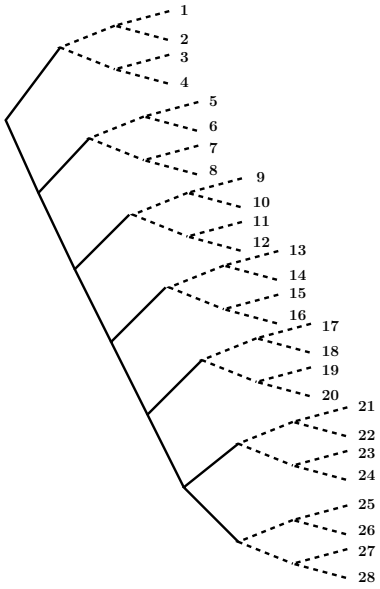

Figure 2: Decomposition tree of the used wavelet filter bank.

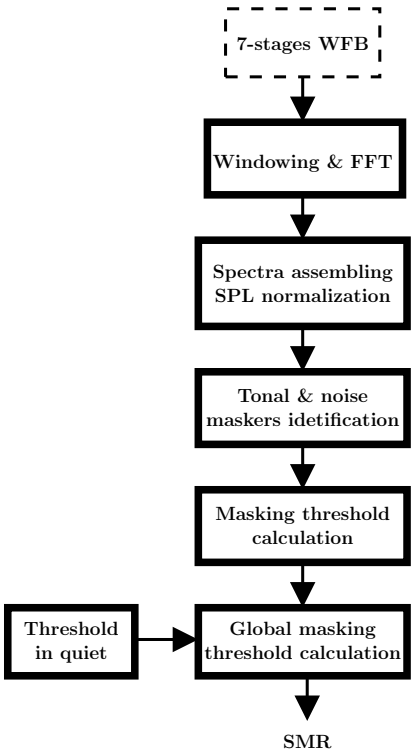

Figure 3: Block diagram of designed coder.

mask ratio (SMR) respectively.

Designed psychoacoustic model was implemented in the Matlab environment. Typical graphical output of this psychoacoustic model implementation can be seen in Fig. 4.

3.1 Windowing and FFT

Signal in each frequency band is analysed separately. Band dependant windowing function is used to best-fit required characteristics. Number of samples of window function is fixed along the bands in the present version

Band No.	Frequency [Hz] Low - High	Equivalent samples [-]	Window length [ms]
1	11 k - 22 k	128	2.7 ms
2	5.5 k - 11 k	256	5.3 ms
3	2750 - 5500	512	10.7 ms
4	1375 - 2750	1024	21.3 ms
5	687.5 - 1375	2048	42.7 ms
6	343.75 - 687.5	4096	85.3 ms
7	$<$ 343.75	4096	85.3 ms3

Table 2: Time lengths of the windows along the frequency bands.

of the model, which means that only one window vector is used in the model. Hamming window used in the current version of the model may be 32 or 64 samples long. For example 64 samples window in the first band (11 - 22 kHz) is 2.6 ms long (for sampling frequency 48 kHz), in the second band (5.5 - 11 kHz) is twice longer (5.3 ms) and in the last 7^{th} band ($<$343.75 Hz) is 85.3 ms long.

Tab. 2 summarises time-lengths of the windows in the sub-bands. Column "Equivalent samples" compares resolution obtained by the designed model with a resolution obtained by the "Equivalent samples" long Short Time Fourier Transform (STFT), length of the 7^{th} band may be twice extended to 8192 samples. Windowing of the band signals hints Eq. 1.

$$s_{wn}(i,j) = s_n(i \cdot \frac{N}{2} + j) \cdot W(j), \qquad j = 0 \ldots N, \quad (1)$$

where s_n represents signal in band n, W represents used windowing function (currently Hamming window), i is a possition in time and N is the length of the window.

Windowed signal is then transformed by Fourier transform as hints Eq. 2.

$$S_n(i) = FFT(s_{wn}(i)) \quad (2)$$

Output of this block is a frequency representation of the windowed seven-band signal under test.

3.2 Spectra assembling and SPL normalization

Global spectral representation of the signal under test is assembled from partial spectras calculated in each frequency band. Because of the decimation of the signal all the details obtained from wavelet filter bank are mirrored from the second Nyquist zone, spectral representations of the first six bands have to be inverted while the seventh not (it is an approximation not a detail). Only half of the spectra is taking into account.

This part of the model produces vector of the spectral components $S(i)$ and vector with the corresponding frequency positions. Vector of the spectral components S is then transformed into sound pressure levels by Eq. 3 (derived from [2]).

$$S_{SPL}(i) = 90.302 + 20 \cdot log_{10}(|S(i)|) \quad [dB] \quad (3)$$

Least significant bit (LSB) of the input signal corresponds with the threshold in the quiet at the most sensitive frequency (4 kHz) [8], therefore fulll-scale sinusoid signal has sound pressure level of 90 dB (for 16-bit representation). In Eq. 3 is this phenomenon expressed by the constant 90.302. Note that input signal has to be normalised. This normalisation is not part of the designed model and is expected to be performed externaly.

In graphical example in Fig. 4 can be seen output of this system block plotted by a dotted blue line.

3.3 Tonal and noise maskers identification

Masking curves of tonal and noise maskers have different shapes [8] therefore it is necessary to separate them. According to [7] spectral component which exceeds its neighbourhood in a given bark distance of 7 dB in minimum is tonal. To find tonal components it is necessary to find local maximas first and then compare them with their neighbourhood components. This action hints Eq. 4.

$$S_{SPL}(i) - S_{SPL}(i \pm \Delta_i) \geq 7, \tag{4}$$

where Δ_i represents examined neighbourhood. All spectral components that satisfy Eq. 4 are tonal. Δ_i is a parameter of the model and usually $\Delta_i = \{1, 2, 3\}$.

According to ISO/IEC MPEG-1 Psychoacoustic Analysis Model 1 of MPEG 1 audio standard [2] sound pressure level of the tonal masker is computed by Eq. 5 as a summation of the spectral density of the masker and its neighbours.

$$X_{TM}(i) = 10 \cdot \log_{10} \sum_{j=-1}^{1} 10^{\frac{S_{SPL}(i+j)}{10}} \quad [dB] \tag{5}$$

Our example in Fig. 4 shows identified tonal maskers in the right SPL possition marked by the green stars.

Noise maskers are according to [2] computed for each critical band from the spectra exclusive of tonal maskers and their neighbourhood components determined by Δ_i. Sound pressure level of the noise maskers is computed according to Eq. 6 as a summation of the sound pressure level of all spectral components in corresponding critical band.

$$X_{NM}(b) = 10 \cdot \log_{10} \sum_{i} 10^{\frac{S_{SPL}(i)}{10}} \quad , z(i) \in b \quad [dB], \tag{6}$$

where b represents critical band, i index spectral components that lies in the corresponding critical band. Noise maskers are placed in the middle of the corresponding critical band. Example in Fig. 4 shows identified noise maskers in their right SPL possition marked by the cyan circles.

3.4 Masking thresholds calculation

When tonal and noise maskers are identified masking threshold for each masker is determined. As defined in ISO/IEC MPEG-1 Psychoacoustic Analysis Model 1 of MPEG 1 audio standard [2] tonal masker masking curve can be calculated by expression 7.

$$M_{TM}(i,j) = X_{TM}(i) + MF(i,j) - 0.275z(j) - 6.025 \quad [dB], \tag{7}$$

where X_{TM} is a SPL of the masker tone, $z(j)$ is the masking curve position on the bark axis, $MF(i,j)$ is a masking function defined by Eq. 8 (from [2]) and constant 6.025 represents SPL distance of the masker and the top of the masking curve.

$$MF(i,j) = \begin{cases} 17\Delta_z - 0.4X_{TM}(i) + 11 & \Delta_z \in < -3, -1) \\ (0.4X_{TM}(i) + 6)\Delta_z & \Delta_z \in < -1, 0) \\ -17\Delta_z & \Delta_z \in < 0, 1) \\ -(\Delta_z - 1) \cdot (17 - 0.15X_{TM}(i)) - 17 & \\ & \Delta_z \in < 1, 3 > \end{cases} \tag{8}$$

where $\Delta_z = z(i) - z(j)$ represents bark distance from the masker in barks. Note that outside the interval $< -3, 3 >$ is $MF = -\infty$. Example in Fig. 4 shows masking curves of tonal maskers plotted by a green line.

Masking curves of the noise maskers is by ISO/IEC MPEG-1 Psychoacoustic Analysis Model 1 of MPEG 1 audio standard [2] defined similarly to the tone one by the Eq. 9.

$$M_{NM}(i,j) = X_{NM} + MF(i,j) - 0.175z(j) - 2.025 \quad [dB], \tag{9}$$

where X_{NM} is SPL of the noise masker, $z(j)$ is the masking curve position on the bark axis, $MF(i,j)$ is a masking function defined by Eq. 8 with X_{TM} changed to X_{NM} and constant 2.025 represents SPL distance of the masker and the top of the masking curve. Outside the interval $< -3, 3 >$ is $MF = -\infty$ again. Example in Fig. 4 shows masking curves of noise maskers plotted by a cyan line.

3.5 Global masking threshold calculation

When individual masking curves of both tonal and noise maskers are determined, global masking threshold may be calculated. According to ISO/IEC MPEG-1 Psychoacoustic Analysis Model 1 of MPEG 1 audio standard [2] the global masking threshold can be calculated by Eq. 10. This equation shows addition of masking and taking into account threshold in the quiet.

$$\begin{aligned} G_{th}(i) \quad = \quad & 10 \cdot \log_{10} \sum_{j} (10^{0.1 \cdot M_{tm}(i,j)} + 10^{0.1 \cdot M_{nm}(i,j)} + \\ & + 10^{0.1 \cdot T(j)}) \qquad [dB], \end{aligned} \tag{10}$$

where $T(j)$ represents threshold in quiet at a given frequency.

Sample graphical output of the designed psychoacoustic model can be seen in Fig. 4. Calculated global masking threshold is in Fig. 4 shown by a solid purple line.

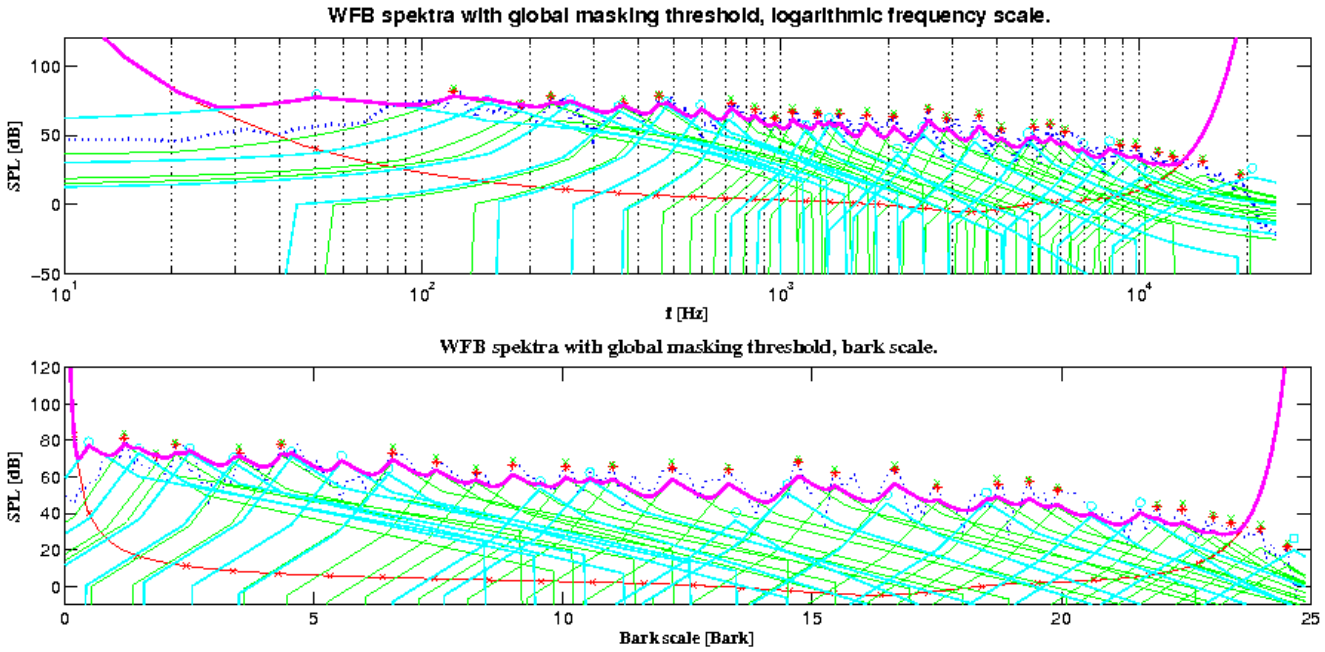

Figure 4: Example of the psychoacoustic model graphical output.

3.6 SMR calculation

Signal to mask ration (SMR) is calculated by the Eq. 11 as a subtraction of a sound pressure level and global masking threshold of the given spectral component.

$$SMR(i) = S_{SPL}(i) - G_{th}(i) \qquad [dB] \qquad (11)$$

Positive SMR indicates that the signal is above the masking threshold while negative SMR indicates that the signal is below the masking threshold and can be excluded from the transmission.

4 Windowing

Similarly to the psychoacoustic model coded signal has to be windowed before following data reduction is performed. In comparison with the psychoacoustic model Hamming window cannot be used because the signal has to be perfectly reconstructed on the decoder side. Using of the simplest rectangular window will cause rising of discontinuities at the border of the windows in a reconstructed signal. Using of overlap windows with special shape is needed to avoid rising of the artefacts connected with this discontinuities. Overlap regions of the windows has to satisfy conditions described in [6].

32 samples long sine window with fifty percent overlap is currently used in the coder. Sine window is defined by Eq. 12 [6]. Use of another window function or different length of the window is possible.

$$w_{sin}(n) = \sin(\pi \frac{n + 0.5}{N}) \qquad (12)$$

Windowing is then performed similarly to Eq. 1. Output of this system block is a band dependant windowed signal $s_w(n, i)$.

5 Scale factor estimation and scaling

Windowed signal is scaled to optimize quantization. In comparison with ISO/IEC MPEG-1 audio standard [2] no table of scale factors is used. Scale factor estimation is performed as a finding of the closest higher power of two of the maximal absolute value in the window. Coding of such scale factor in service information is very efficient as four bits only are needed (for 16 bit input signals) per one window.

$$S_f(n, i) = min(2^k \geq max|s_w(n, i)|), \qquad (13)$$

where k is an integer number, $n = 1 \ldots 28$ is indexing sub-bands, i is indexing windowed signals.

Scaling of the windowed signal is then easily performed as a division of the windowed signal with the scale factor as described in Eq. 14.

$$s_{ws}(n, i) = \frac{s_w(n, i)}{S_f(n, i)}, \qquad (14)$$

As scale factor is a k^{th} power of two an optimal implementation of the scaling is a simple bit rotation of k bits. For the bit allocation will be useful to convert scale factor into sound pressure level. This conversion is made by Eq. 15.

$$\begin{aligned} S_{fSPL}(n, i) &= 90.302 + 20 \log_{10} \left(\frac{2^k}{2^{16}} \right) \\ &= 6.02 \cdot k - 6.03 \qquad [dB] \end{aligned} \qquad (15)$$

6 Bit allocation and quantization

Task of the bit allocation system block is to find least number of bits that are necessary for coding of each windowed signal. Inputs of this block are scaling factors and global masking threshold. For each sub-band is found minimum of the masking threshold which describes maximal level of non-perceptible sound (Eq. 16).

$$G_{min}(n,i) = min(G_{th}(i,j)), \quad j \in band_n, \qquad (16)$$

where i indicate position in time and j is indexing spectral components of the masking threshold in a given sub-band.

Dynamics necessary for signal coding in a given window is than calculated as a subtraction of S_{fSPL} and G_{min}.

$$D_{min}(n,i) = S_{fSPL}(n,i) - G_{min}(n,i) \qquad [dB] \qquad (17)$$

Number of bits that is necessary for coding of each windowed signal is than computed by Eq. 18.

$$B(n,i) = \left\lceil \frac{D_{min}(n,i)}{6.02} \right\rceil, \qquad (18)$$

where $\lceil\rceil$ means the nearest higher or equal integer number.

Value $B(n,i)$ is used by the quantization block which performs reduction of the data. Note that all other system blocks do not reduce data (except scaling) but prepare signal for effective quantization and analyse coded signal in this block. Thanks to this service quantization block has easy work. Data from each windowed and scaled signal are reorganized into stream of bits. Only $B(n,i)$ most significant bits (MSB) are taking into account and are serially arrange in sequence of output data.

7 Service data

Block named service data collects all the information about coding which is necessary for successful decoding. For each windowed signal is necessary to collect scaling factor and number of bits. Necessary is also information about audio source like sampling frequency, length, etc. All this information is collected by the service data block. Concrete format of the output is not yet designed as the model is in developing process.

8 Output data formatting

Output data has to be stored to keep coded audio data for future transmission or decoding. Output data formatting block performs this task. As there is not defined any output data format, this block just stores output data stream and service information on the hard drive separately. Concrete format for the data will be designed after the system optimization.

9 Conclusion

Based on ISO/IEC MPEG-1 audio standard [2], generic audio coder [1], wavelet filter bank based sound signals analysis [4] and wavelet filter bank based psychoacostic model [3] novel wavelet filter bank based wide-band audio coder was designed.

Wavelet filter bank usage allows frequency dependant resolution in psychoacoustic model and frequency dependant windowing for quantization. Masking effects may be better involved by such procedure and artefacts like pre-echo may be suppressed. Matlab implementation of the designed wavelet filter bank based wide-band audio coder is beeing developped. Coder outputs will be analysed by listening tests for future improvements of the designed system.

10 Acknowledgements

Research described in the paper was supported by the Czech Technical University in Prague under grant No. CTU 0807413.

References

[1] Rund, F., Nováček, J.: *Experimentální zvukový kodér*, Technical Computing Prague 2007 [CD-ROM], HUMUSOFT, Prague, 2007, ISBN 978-80-7080-658-6.

[2] ISO/IEC 11172-3: *Information Technology - Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s – Part 3: Audio*, 1993.

[3] Nováček, J.: *Wavelet Filter Bank Based Psychoacoustic Model*, In Poster 2008 [CD-ROM], CTU, Faculty of Electrical Engineering, Prague, 2008.

[4] Nováček, J.: *Matlab GUI for WFB spectral analysis*, Technical Computing Prague 2007 [CD-ROM], HUMUSOFT, Prague, 2007, ISBN 978-80-7080-658-6.

[5] Mallat, S.: *A Wavelet tour of signal processing*, Academic Press, San Diego, 1998, ISBN 0-12-466605-1.

[6] Bosi, M., Goldberg, R., E.: *Introduction to digital audio coding and standards*, Kluwer Academic Publishers, Boston, 2003, ISBN 1-4020-7357-7.

[7] Painter, T., Spanias, A.: *Perceptual Coding of Digital Audio*, Proceedings of the IEEE, Volume 88, Issue 4, Pages: 451 - 515, April 2000.

[8] Zwicker, H., Fastl, H.: *Psychoacoustics*, Springer-Verlag, Berlin, 1990, ISBN 0-387-52600-5.