THE DESIGN OF A HYBRID FILTER BANK FOR THE PSYCHOACOUSTIC MODEL IN ISO/MPEG PHASES 1, 2 AUDIO ENCODER+

Chi-Min Liu, Member, IEEE, and Wen-Chieh Lee
Department and Institute of Computer Science and Information Engineering
National Chiao Tung University, Hsinchu, 30050, Taiwan
E-Mail: cmliu@csie.nctu.edu.tw

Abstract— **The ISO/MPEG phases 1 and 2 audio compression are receiving a wide range of applications. In the encoding process of MPEG, the psychoacoustic model exploits audio irrelevancy which is the key role to achieve high compression ratio without losing audio quality. However, the Fourier transform (FT) which has been used by the two psychoacoustic models suggested in standard draft requires high computational complexity, and hence leads to high hardware and software cost for real-time applications. This paper presents a new design named the hybrid filter bank to replace the FT. The hybrid filter bank can be integrated with the psychoacoustic models and provides a much lower complexity than the FT. Also, this paper shows that the hybrid filter is more suitable for the stereo coding and hence can provide a better quality for the intensity stereo coding, which is the key technology for the MPEG 1 to achieve near transparent quality lower than 96x2 kbits for two stereo channels.**

I. Introduction

LIKE most perceptual audio coders [1]-[3], MPEG audio encoder can be considered from four parts: the time-frequency mapper, the psychoacoustic model, quantization and frame packing as shown in Fig. 1. The psychoacoustic model exploits audio irrelevancy which is usually defined in frequency domain. The time-frequency mapper maps the time-domain signals into a frequency representation to reduce the data redundancy and provides the ease with the integration with the psychoacoustic model. The quantization quantizes the audio signals from time-frequency mapper based on the information from the psychoacoustic model. The frame packing packs the quantized signals with some synchronous information like sampling frequency for identified by

+This work was supported in part by Acer Laboratories Inc. under contract C85098.

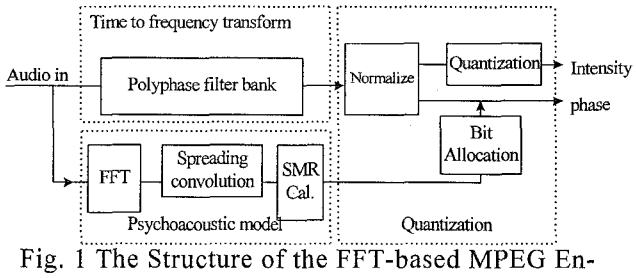


Fig. 1 The Structure of the FFT-based MPEG Encoder

MPEG decoders.

In the encoding process of MPEG, the 1024-point Fourier transform (FT) has been used by psychoacoustic models to analyze the frequency components in the 1152 samples of one frame. If the conventional real-data fast FT (FFT) [4] has been adopted for implementing the FT, the complexity has an order of (4*256*log(512)). Such a complexity leads to high implementation cost for real-time applications.

This paper presents a new design named the hybrid filter bank to replace the FT. The hybrid filter bank can be integrated with the psychoacoustic models and provides a much lower complexity than the FT. Also, this paper shows that the hybrid filter is more suitable for the stereo coding and hence can provide a better quality for the intensity stereo coding, which is the key technology for the MPEG 1 to achieve near transparent quality lower than 96x2 kbits for two stereo channels.

This rest of this paper is organized as follows: Section II illustrates the design of hybrid filter banks. The hybrid filter bank has problems in the phase shift and the aliasing components arising from the decimation in the 1st level filter bank. Section III provides the method to solve the two problems. Section IV considers complexity and the integration of the hybrid filter banks with the psychoacoustic models in MPEG. Section V evaluates the design through spectrum analysis, subjective measure, and objective

Manuscript received June 13, 1997

measure to show the feasibility of the hybrid filter bank. Section VI gives a brief conclusion.

II. Filter Response in Hybrid Filter Banks

The motivation of the hybrid filter banks can be considered from the two frequency analyzers in the time-frequency mapper and the psychoacoustic model. The MPEG has adopted a 32-band polyphase filter bank which can provide a frequency resolution $\pi/32$ with sidelobe attenuation 96 dB while the FT with Hann window a resolution $\pi/512$ with attenuation 32 dB. The approach of the hybrid filter bank is to cascade another filter bank, named the second (2^{nd}) level filter bank, to the output of the original polyphase filter bank, named the first (1^{st}) level filter bank, to achieve a high frequency resolution. The block diagram of the hybrid filter bank is shown in Fig. 2.

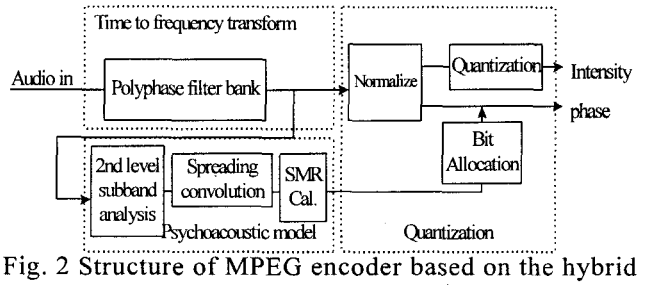


Fig. 2 Structure of MPEG encoder based on the hybrid filter banks

Fig. 3 shows the detailed structure of the hybrid filter bank. The structure adopts a 16-band filter bank based on the time domain aliasing cancellation (TDAC) filter bank [6] for each band of the 1^{st} level filer bank to achieve a frequency resolution as high as the FT. The input-output relation of the TDAC filter bank is

$$X_i(k)=\sum_{n=0}^{N-1}h(n)x_i(n)\cos\left[\frac{\pi}{2N}(2n+1+\frac{N}{2})(2k+1)\right] \quad \text{for } 0\le k\le\frac{N}{2}-1, \quad (1)$$

where $x_i(n)$ is the *nth* output of the band *i* from the 1^{st} level polyphase filter bank, $X_i(k)$ is the corresponding output of the 2^{nd} level filter bank and $h(n)$ is the window function deciding the band selectivity in the 2^{nd} level filter bank. To achieve a frequency resolution $\pi/512$ the same as the FT, the value of N is set to 32. Also, to have a frequency selectivity the same as the FT, we select the window function

$$h(n)=\sin(\frac{\pi}{N}(n+\frac{1}{2})) \quad \text{for n = 0, ..., N-1} \quad (2)$$

which has a sidelobe attenuation 24 dB as shown in Fig. 4. The function has the property

$$h(n)^2+h(n+\frac{N}{2})^2=1 \quad \text{for } 0\le n\le\frac{N}{2}-1 \quad (3)$$

which is a necessary condition leading to the perfect reconstruction filter banks [5]. Substituting (2) into (1) yields

$$X_i(k)=\sum_{n=0}^{N-1}\sin(\frac{\pi}{N}(n+\frac{1}{2}))x_i(n)\cos(\frac{\pi}{2N}(2n+1+\frac{N}{2})(2k+1)) \quad (4)$$

$$\text{for k = 0 to } \frac{N}{2}-1$$

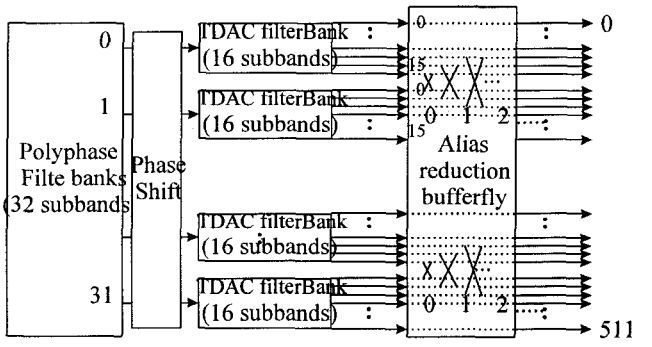


Fig. 3 Detailed structure of the hybrid filter bank

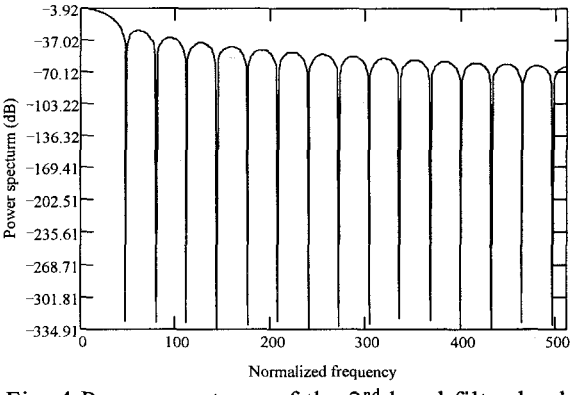


Fig. 4 Power spectrum of the 2^{nd} level filter bank

III. Phase Shifter & Alias Reduction

As mentioned in [7], [8], the hybrid filter bank has problems in the phase shift and the aliasing components arising from the 1^{st} level filter bank. We follow the similar concept in [7], [8] to design a phase shifter and an alias reduction butterfly to solve these two problems.

A. Design of the phase shifter

Due to the decimation operation implied in the 1^{st} level filter bank, the 1^{st} filter bank has a phase shift

π in the odd-indexed subbands. The phase shift causes a reversed spectrum for the subband. If further spectral analysis is needed to achieve higher frequency resolution, this shift should be corrected. This phase shift can be corrected by multiplying $(-1)^n$ to the subband signal in the odd-indexed subbands; that is

$$X_i(k) = \begin{cases} \sum_{n=0}^{N-1}(-1)^n \sin(\frac{\pi}{N}(n+\frac{1}{2}))x_i(n)\cos(\frac{\pi}{2N}(2n+1+\frac{N}{2})(2k+1)) \text{ for even i} \\ \sum_{n=0}^{N-1} \sin(\frac{\pi}{N}(n+\frac{1}{2}))x_i(n)\cos(\frac{\pi}{2N}(2n+1+\frac{N}{2})(2k+1)) \text{ for odd i} \end{cases} \quad \text{for k = 0 to } \frac{N}{2}-1 \tag{5}$$

where odd/even stands for odd/even indexed subband of 1st level filter bank. The phase shifter can be combined into window function to avoid computation burden.

B. Design of the aliasing reduction butterfly

It has been well known that the decimation operation leads to aliasing and there are decimation in the hybrid filter banks. The aliasing effects indicate a many-to-one merging between the input frequency and output frequency of filter banks, and hence lead to the difficulty distinguishing the "many" frequency components from the "one" frequency component. The merged frequencies and the corresponding merging weights are decided by the filter bandwidth and the magnitude response of the filter in filter banks. For the filter bank designed in last section, since that the sidelobe attenuation is around 24 dB, the aliasing

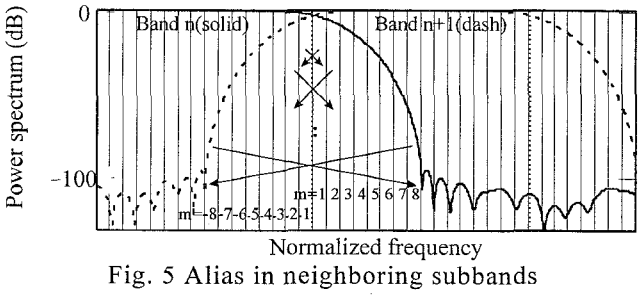


Fig. 5 Alias in neighboring subbands

term of the frequency in a filter band can be reasonably approximated by the frequency components from the nearest neighboring band. For the hybrid filter bank design in Fig. 3, aliasing arises from both the 1st filter banks and the 2nd filter banks. The aliasing terms in the 1st level filter bank lead to the merging of frequencies with distance as far as $\pi/32$ while that in the 2nd level filter bank $\pi/512$. Since that the psychoacoustic models in MPEG needs a frequency resolution $\pi/512$, the aliasing terms from the 1st level filter bank should be suitably corrected to increase the frequency resolution.

Fig. 5 shows the frequency responses for the two neighboring filters in the 1st level filter bank before decimation. The lattice lines in Fig. 5 show the resolution boundary for the 2nd level filter bands. The cross lines in Fig. 5 shows the merged bands from the decimation in the 1st level filter bank.

Edler [7] has designed the butterfly structure in Fig. 6 to ease the aliasing errors in hybrid filter banks. The hybrid structure in Fig. 3 has included the butterfly structure to compensate the aliasing terms. The butterfly operation is

$$\begin{aligned} u_i &= d_m(r_i - c_m r_j) \quad \text{with i=16·k-1-m} \\ u_j &= d_m(r_j + c_m r_i) \quad \text{with j=16·k+m} \\ &\text{with } d_m = 1/\sqrt{1+c_m^2},\ -N/2 \le m \le -1 \end{aligned} \tag{6}$$

where r_i and r_j are the band signals from the 2nd level filter bank indicated by the cross lines in Fig. 5. The u_i and u_j are resulted signals after the correction. The c_m and d_m are the two weighting factors designed to compensate the aliasing errors in each band. The values of these two factors vary with bands labeled as m indicated in Fig. 5. In the following we show the method to obtain the values for these weighting factors.

C. Design of the weighting factors in the butterfly

For the bands other than those labeled as m=-1 and 1, the weighting factors are calculated using the ratio between the filter response energy in the signal band and that of the aliasing band:

$$c_m = \sqrt{\frac{\text{Energy of alias band}}{\text{Energy of signal band m}}} = \sqrt{\frac{\int_{aliasing\ band} |H(\omega)|^2\, d\omega}{\int_{signal\ band} |H(\omega)|^2\, \omega}} \tag{7}$$

where $H(\omega)$ is the frequency response of one filter in the 1st filter bank.

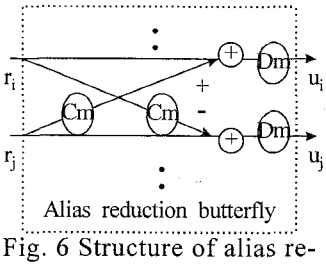


Fig. 6 Structure of alias reduction butterfly

m	c_m	d_m
-1	-0.56859	0.86930
-2	-0.49539	0.89607
-3	-0.28182	0.96251
-4	-0.14189	0.99008
-5	-0.05942	0.99824
-6	-0.01952	0.99981
-7	-0.00429	0.99824
-8	-0.00049	1.00000

Table 1 Eight weighting factors of alias reduction butterfly

However, the compensation should be modified for the bands labeled as m=-1 and 1. As described above, there are aliasing from the 2nd level filter bank. For example, the band labeled as m=2 have aliasing terms from the band labeled as m=1 and m=3. However, the aliasing terms for m=-1 and m=1 are only from the band m=-2 and m=2, respectively. To take the special effect into the butterfly, the weighting factors for m=-1, 1 are calculated as

$$c_1 \text{ or } c_{-1} = \sqrt{\frac{\text{Energy of alias band (1-r)}}{\text{Energy of signal band m}}} \qquad (8)$$

where γ is the ratio between the filter response energy of the signal and the aliasing terms in the 2nd level filter bank. Table 1 summarizes the values of the weighting factors.

IV. Complexity Analysis and the Integration with the MPEG

This section analyzes the complexity of the hybrid structure and shows various aspects and advantages of replacing the FT by the hybrid filter banks in MPEG audio coding.

A. Complexity analysis

The substitution of the hybrid structure for the FT in the psychoacoustic models of MPEG provides two advantages in complexity. First, since that the two frequency analyzers in Fig. 1 can be merged into the hybrid structure in Fig. 2, the complexity can be reduced. The second advantage in complexity is from the flexible tuning of frequency resolution in hybrid structure for the different perceptual resolution. If the perceptual resolution (which is the bandwidth of the critical band) is considered in Fig. 7, only 12 TDAC filter banks with alias reduction butterfly structures are required for low frequency range.

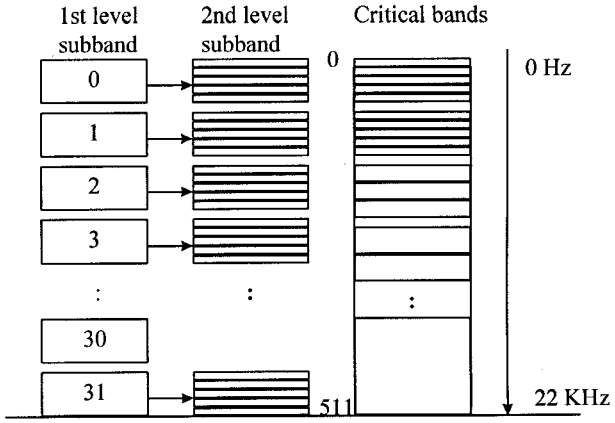


Fig. 7 Hybrid filter bank resolution vs. critical band

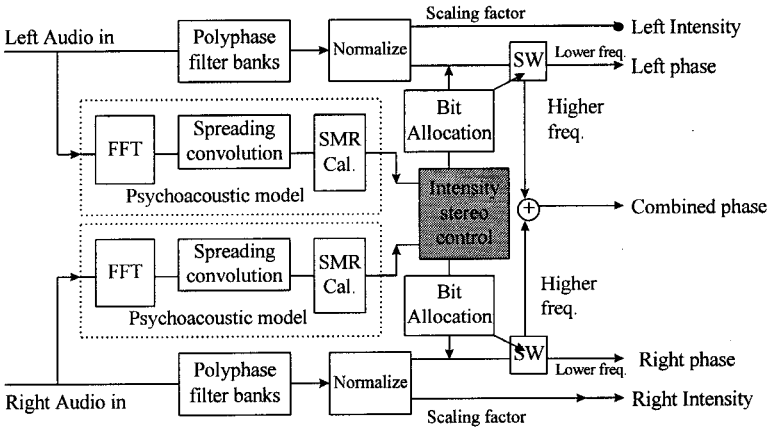


Fig. 8 Conventional intensity stereo coding scheme

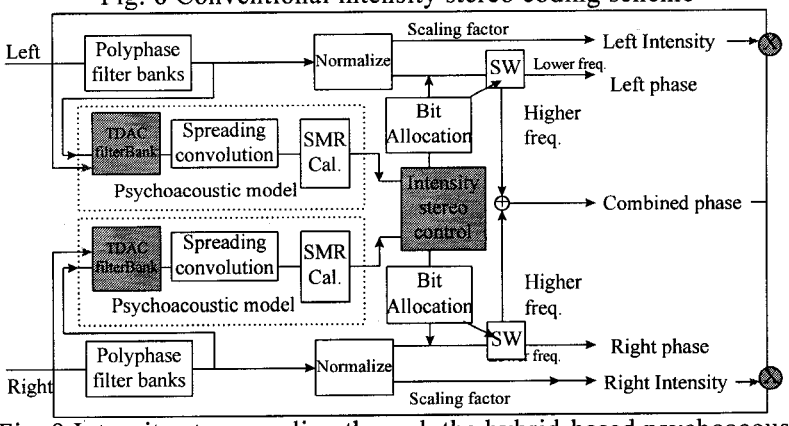


Fig. 9 Intensity stereo coding through the hybrid-based psychoacoustic model

Algorithms of frequency mapping in psychoacoustic model	# of multiplications per 1152 samples	# of additions per 1152 samples
1024 pt. FFT (real FFT) + Hann window	4*256*log(512)+512=9728	2*256*log(512)+2*512*log(512)=9216
32 (32 pt. TDAC filter bank + window)	32*16*log(32)+32*32=3584	32*32*log(32)=5120
32 (32 pt. TDAC filter bank + window + Alias cancellation)	3584+32*6*2*2=4352	5120+32*6*2=5504
12 (TDAC + window + Alias cancellation) + critical bands	12/32*(4352)=1632	12/32*(5504)=2064

Table 2 Complexity comparison between FFT and hybrid filter bank.

Table 2 shows the complexity of the hybrid structure compared with the FFT. The 1024-point real-data FFT requires 256*log(512) complex multiplications and 512*log(512) complex additions with Hann window of 512 multiplications, while 32 2nd level TDAC filter banks with the 6 aliasing cancellation butterfly structures require only an order of 32(16*log32+32 +6*2*2) when the fast algorithm of the TDAC filter bank [10] is applied. Further reduction from the perceptual resolution can reduce the complexity as indicated in row 4 of Table 2.

B. Cooperating with the intensity mode

The other advantage of the substitution of the hybrid structure for the FT in the psychoacoustic models of MPEG is on the stereo encoding. As mentioned in our previous paper [9], the intensity stereo coding is the key technology for layer 2 in MPEG 1 to achieve a near transparent quality at a bit rate as low as 96x2 kbits for the two stereo channels. However, the original FT analysis has problems in maintaining a consistent frequency analysis with the stereo signals. When the high frequency parts of the two stereo channels are combined into one channel in intensity stereo coding or the scheme mentioned in [9] as shown in Fig. 8, original FT analysis result is not representative for the frequency analysis of the combined channels.

One way to overcome this inconsistent problem is to recalculate the FT analysis and the psychoacoustic model for the two channels somehow based on the combined channels. This recalculation leads to heavy computing load. On the other hand, when these stereo coding schemes are applied, the hybrid structure can be easily tuned to a consistent analysis. Modification of the frequency analysis and the corresponding psychoacoustic model can be performed only on part of the frequency range for the combined channels through the hybrid structure. The hybrid filter bank cooperating with the intensity stereo coding scheme is shown in Fig. 9.

C. Tonality measure

The determination of the tonality of a spectrum line or a band is important in the psychoacoustic model to calculate the sensitivity of the human on the lines or bands. The psychoacoustic model 2 indicated in MPEG draft consider the tonality through a simple prediction calculated in polar coordinates in the complex plane[2]. The tonality detection above is originally designed based on the complex numbers in the output of the Fourier transform. Since that the output of the hybrid filter bank presented in this paper is real data, the detection mechanism should be suitably modified. The predicted magnitude for a spectrum lines is denoted as $\tilde{r}(t,f)$, which is calculated from the two preceding magnitudes $r(t-1,f)$, $r(t-2,f)$:

$$\tilde{r}(t,f)=r(t-1,f)+(r(t-1,f)-r(t-2,f)) \quad (9)$$

where t and f represent the index of time and frequency, respectively. The tonality factor c(t ,f) used in psychoacoustic model 2 can now be obtained as

$$c(t,f)=\frac{\sqrt{r(t,f)^2-\tilde{r}(t,f)^2}}{r(t,f)+abs(\tilde{r}(t,f))} \quad (10)$$

For tone signals, the prediction turns out to be very good, and *c(t, f)* will have a value near zero. On the other hand, for very unpredictable signal such as noise signals, c(t, f) will have a value near 1.

V. Quality Measure

The effects of the hybrid filter bank and the corresponding modification can be illustrated by comparing the spectrum from the FT and that from the hybrid filterbank. The spectrum analysis for signals with five components at frequencies 400Hz, 800Hz, 1600Hz, 3200Hz and 6400Hz are shown in Fig. 10 through the FT (dotted line), the hybrid filter bank without alias reduction (dashed line with 100dB shifting up) and the hybrid filter bank with alias reduction (solid line with 200dB shifting up). The location of each frequency of the hybrid filter bank are almost the same as the one of FT and the alias component of the hybrid filter bank with alias reduction can effectively reduce the aliasing terms.

Several audio segments has been adopted to meas-

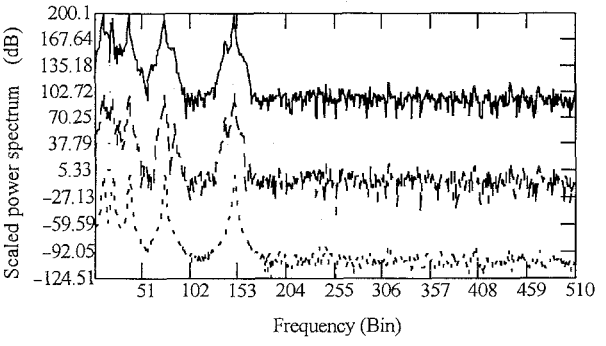


Fig. 10 Signal with frequency located at 400Hz, 800Hz, 1600Hz, 3200Hz and 6400Hz analyzed by 1024 pt. FT (dotted line), the hybrid filter bank (dashed line) and the hybrid filter bank with alias reduction butterfly (solid line)

ure the signal-to-masking ratio [9] from the FT and the various hybrid filter bank. Two of the results are shown in Fig. 11 and Fig. 12 where the FT is denoted by the solid line, the hybrid filter bank with alias reduction by dotted line, and the hybrid filter bank with only 12 bands in the 2nd level by dashed line. The results show that the hybrid filter bank with low

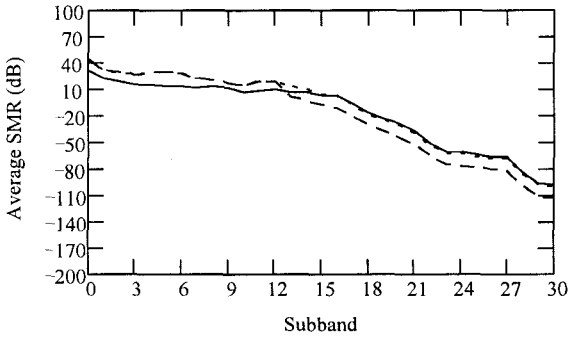


Fig. 11 Average signal-to-masking ratio of each subband for female vocal sound

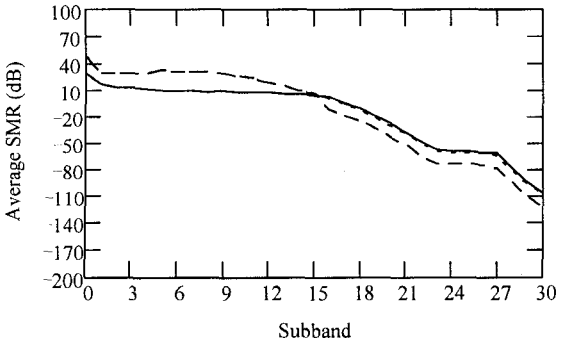


Fig. 12 Average signal-to-masking ratio of each subband for classical symphony orchestra

complexity can provide a result similar to the FT.

Also, informal listening tests show that the audio segments coded by the psychoacoustic model of the FT and the hybrid filter bank are almost imperceptible.

VI. Concluding Remarks

This paper has presented a new design named hybrid filter banks to replace the FT adopted in the psychoacoustic model suggested in the draft on the MPEG phases I and II audio coding. This paper has given the means to solve the phase shift and aliasing problems in the hybrid structure. The hybrid filter bank can be well integrated with the psychoacoustic model and provide a much lower complexity than the FT. We have also shown that the hybrid filter bank can cooperate with intensity stereo coding scheme to obtain higher audio quality. Due to the flexibility of the hybrid filter bank, a consistent psychoacoustic model with the intensity stereo coding channel can be obtained with little computation increasing. The hybrid filter bank is tested through spectrum analysis, subjective measure, and objective measure to show the feasibility.

References

[1] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE Journal on Selected Area in Communications*, vol. 6, no. 2, pp. 314-323, Feb, 1988.

[2] K. Brandenburg, J. D. Johnston, "Second level perceptual audio coding: the hybrid coder," *The 88th Convention of AES*, March 13-16, 1990.

[3] R N. J. Veldhuis, "Bit rates in audio source coding," *IEEE Journal on Selected Areas in Communications*, vol. 10, no. 1, pp. 86-96, Jan, 1992.

[4] E. O. Brigham, "The fast Fourier transform and its application," *Prentice Hall Inc.*, 1988.

[5] P. P. Vaidyanthan, "Multirate digital filters," *Prentice Hall Inc.*, 1993.

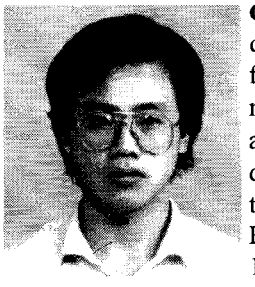
[6] J. Princen, A. Johnson, A. Bradley, "Subband/ transform coding using filter banks designs based on time domain aliasing cancellation," *Proc. of the ICASSP* 1987, pp. 2161-2164.

[7] B. Edler, "Aliasing reduction in sub-bands of cascaded filter banks with decimation," *Electronic Letters* vol. 28, no. 12, pp. 1104-1106, Jun., 1992.

[8] K. Brandenburg, E. Eberlein, J. Herre, B. Edler, "Comparison of filterbanks for high quality audio coding," *IEEE International Symposium on Circuit and Systems* vol. 3, pp. 1336-1339, 1992.

[9] C. M. Liu and J. C. Liu, "A new intensity stereo coding scheme for MPEG audio encoder- layer I and II," *IEEE Trans. on Consumer Electronics*, vol. 42, pp. 535-539, Aug., 1996.

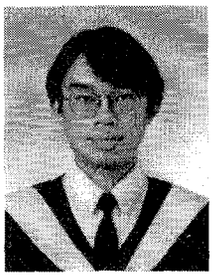
[10] T. Sporer, K. Brandenburg, B. Edler, "The use of multirate filter banks for coding of high quality digital audio," *The 6th European Signal Processing Conference*, vol. 1, pp. 211-214, Jun., 1992.



Chi-Min Liu received the B.S. degree in electrical engineering from Tatung Institute of Technology, Taiwan, R.O.C. in 1985, and the M.S. degree and Ph. D. degree in electronics from National Chiao Tung University, Hsinchu, Taiwan, in 1987 and 1991, respectively.

He is currently an Associate Professor of the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan. His research interests include video/audio

compression, speech recognition, radar processing, and application-specific VLSI architecture design.



Wen-Chieh Lee received the B.S. degree from the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan in 1995. He is currently a Ph.D. student of the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan. His research interest is in the area of audio compression