MEMORY REDUCTION TECHNIQUE OF SPREADING FUNCTION IN MPEG AAC ENCODER

Shih-Way Huang, Liang-Gee Chen

DSP/IC Design Lab, Grad. Inst. of Elect. Eng.,
Dept. of Electrical Engineering
National Taiwan University, Taipei, Taiwan, ROC
shihway@video.ee.ntu.edu.tw

Tsung-Han Tsai

Dept. of Electrical Engineering
National Central University,
Chung-Li, Taiwan, ROC
han@ee.ncu.edu.tw

ABSTRACT

In order to reduce the computational complexity of an MPEG AAC encoder, calculation of spreading function, which is the critical algorithm in the psychoacoustic model, can be replaced with look-up tables. However, the memory size of the table is considerably large in DSP or hardware implementation for portable devices. The paper deals with the methods to reduce the memory size. We modified a method originally used in an MP3 software encoder and adopted it in an MPEG AAC encoder. The result shows that only about one-third of the original size is required in programmable processor implementation such as DSP. Moreover, for hardware implementation, we analyzed the wordlength and reduced the size to only 3.6% compared to that for a 16-bit DSP while maintaining audio quality.

1. INTRODUCTION

Portable electronic devices such as smart mobile phones, digital cameras, PDAs, USB drives, and digital audio devices with audio playback and audio recording have been attractive because of both the prevalence of MP3 [1] audio files and the development of flash memory cards. The MPEG AAC [2][3] encoder, so-called MP3's successor, is to be embedded into those portable devices because of its higher coding efficiency. Therefore, low power and low cost (area) are important considerations in DSP and hardware implementations. In the previous work, Tsai et al. [4] proposed that the calculation of spreading function can be replaced with a look-up table to reduce the computational complexity dramatically. However, the size of the table is considerably large in the implementation. For example, Gayer et al. [5] showed that about 10000 words of data RAM and 7000 words of data ROM are required for a typical AAC LC encoder on DSP. On the other hand, the memory size of spreading function requires 6614 words (at sampling rate 44100 Hz) without optimization. Therefore, it is important to reduce the memory size of spreading function.

The paper proposes methods to reduce the memory size of spreading function. According to the characteristics of the table, we utilized two methods to reduce the size of the table. Referring to an MP3 software encoder, we employed two arrays to store the values. Only about one-third of the original size is required. Besides, for dedicated hardware implementation, where area can be further reduced with word length analyses of the table values, the memory size was reduced to only 3.6% compared to that for a 16-bit DSP.

This paper is organized as follows. In Section 2, the background of spreading function in psychoacoustic model is introduced and the problem of the table's large size is emphasized. In Section 3, the analyses and characteristics of the table of the spreading function are reported. In Section 4, the methods to reduce the size of the table are described. In Section 5, simulation and results are demonstrated and tabled. Finally, major contributions are summarized.

2. SPREADING FUNCTION IN PSYCHO-ACOUSTIC MODEL

In psychoacoustics, masking is an effect that one sound is made inaudible because of the presence of another sound. Since the simultaneous masking effects are not band-limited within the boundaries of a single band, inter-band masking also occurs. A masker centered within one critical band has effects across the other bands. The effect, predictable on detection thresholds in other critical bands, is known as the spread of masking and often modelled in coding applications by an approximately triangular spreading function [6]. A spreading function is adopted in ISO/IEC MPEG Psychoacoustic Model 2, which is employed in MP3 and AAC.

Figure 1 illustrates the spreading function of psychoacoustic model (PAM) in an MPEG AAC encoder. The above of the figure is a block diagram of MPEG AAC encoder, and the below are the 13 steps of PAM according to MPEG AAC [3]. In step 5, both partitioned energy and unpredictability are convolved with the spreading function in order to estimate the effects across the partitioned bands.

Spreading function, shown in Figure 2, is composed of a series of functions, including some complex operations such as square roots, divisions, exponential operations, etc, which increase the complexity. Tsai et al. [4] proposed that the repeated calculation of spreading function is dominant in PAM, and it can be replaced with a look-up-table memory because it is only dependent on the sampling rates and the block type used. Table I shows the reduction rate of computational complexity when the calculation of spreading function is replaced with a look-up table. The reduction rate is up to 36% for an MPEG AAC encoder. However, a problem is left–the size of the look-up table is considerably large. For example, if the memory for sampling rate 44100 Hz (CD-quality) is required, a two-dimension table due to convolution in step 5 is calculated as follows:

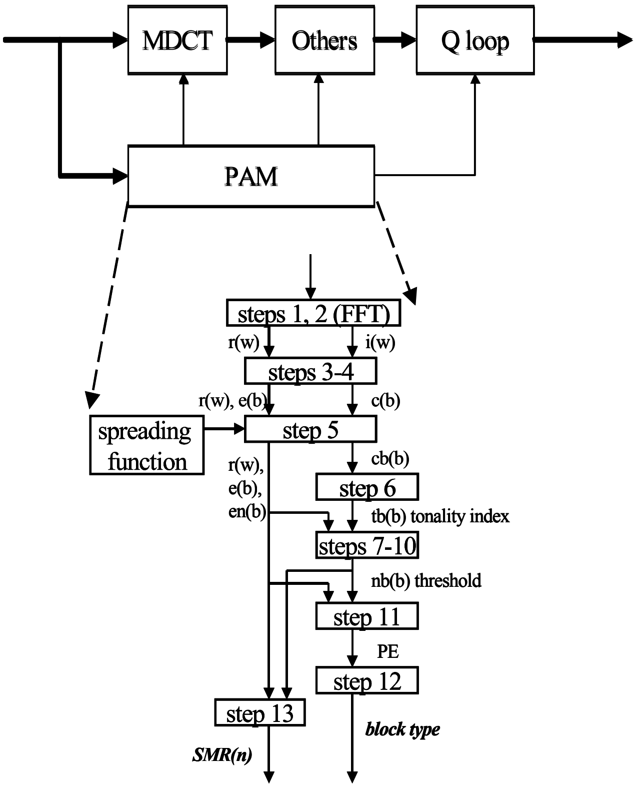

Figure 1: *The spreading function of psychoacoustic model in an AAC encoder.*

```
Spreading function (bark value i, bark value j)
{
if (j>=i)   tmpx = 3.0*(j-i)
      else   tmpx = 1.5*(j-i)
tmpz = 8 * minimum(((tmpx-0.5)^2-2(tempx-0.5)),0)
tmpy = 15.811389 + 7.5(tmpx+0.474) – 17.5(1.0+(tmpx+0.474)^2)^0.5
if(tmpy < -100)   then return 0
      else   return 10^((tmpz+tmpy)/10)
}
```

Figure 2: *The calculation of spreading function.*

LONG block type: 70 (partitioned bands) x 70 = 4900
SHORT block type: 42 (partitioned bands) x 42 = 1764
Total table size = 6664 words

For the implementation on a programmable processor such as DSP, the table is usually stored in ROM or RAM because it is repeatedly accessed. If the table is stored in ROM, it should be pre-calculated for all possible configuration parameters, that is, all the supported sampling rates besides 44100 Hz. If the table is stored in RAM, which is a rather expensive type of memory, it should be calculated initially and stored with other dynamic data. Gayer et al. [5] showed that about 10000 words of data RAM and 7000 words of data ROM are required for a typical AAC LC encoder, but they did not specify the implementation details of memory usage. Thus, the table (6664 words) is large wherever it is stored in RAM or ROM without optimization. Moreover, for the implementation on a dedicated hardware, area can be further reduced with word length analyses of the table. Therefore, it is important to reduce the table size and its according access bandwidth in an efficient way.

Table I: *The reduction rate by replacing the calculation of spreading function with a look-up table.*

	Original (MOPS)	By look-up table (MOPS)	Reduction Rate
PAM step 5	40.3	7.5	81.4%
PAM total steps	52.1	19.3	63.0%
AAC whole system	90.8	57.9	36.2%

Table II: *The number of zero and non-zero values in the table of spreading function (sampling rate 44100 Hz).*

	LONG	SHORT	Total	%
Original	4900	1764	6664	100.0%
Zeroes	3623	1198	4821	72.3%
Non-zeroes	1277	566	1843	27.7%

Table III: *The number of values for storage (sampling rate 44100 Hz).*

Reduction method	Array of value	Array of index	Total	%
Original	6664	0	6664	100.0%
Storage in two arrays	1843	224	2067	31.0%

3. ANALYSES OF SPREADING FUNCTION TABLE

In order to minimize the size of the two-dimension table, we pre-calculated all the values (outputs) of the spreading function and then analyzed them. We found that the table was a sparse table with the following characteristics. Figure 3 illustrates a table at sampling rates 44100 Hz and LONG block type. The size is 70 x 70 as the square number of partitioned bands. The table is characterized as follows:

- Most of the values are exactly zero. This is because the function (Figure 2) returns zero when the condition tmpy < -100 is true, which is often satisfied.
- The distribution of the zero values is regular in both the upper triangular parts and lower ones.
- The number of the non-zero values is very small (in the shape of necktie in Figure 3). Table II shows that it occupies only 27.7% of the table.
- The non-zero values are positive and smaller or equal to one. Most of them, however, are very small up to approximately zero.

In next Section, we will discuss how to reduce the size of the table according to these characteristics.

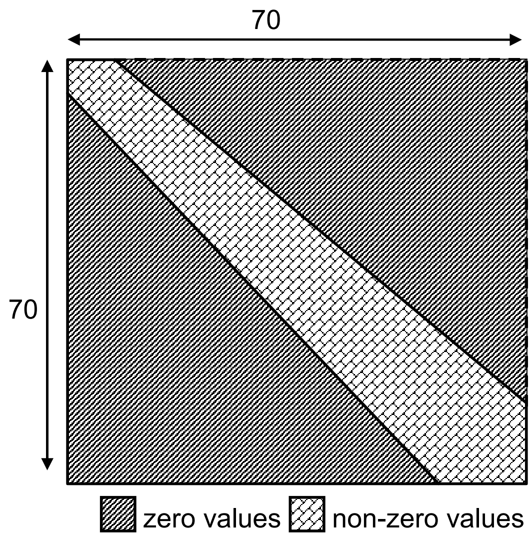

Figure 3: *The distribution of zero values and non-zero values.*

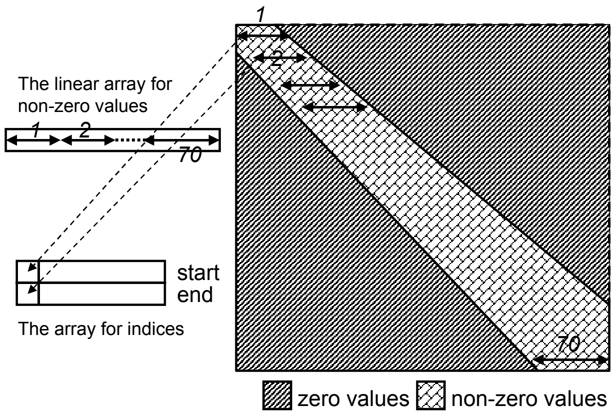

Figure 4: *The reduction method "Storage in two arrays."*

4. MEMORY REDUCTION METHODS

According to the characteristics in the last Section, we employed two methods to minimize the size of the table. First, we referred to an MP3 software encoder, which utilizes a linear array to store values and an array to store indices. Secondly, we analyzed the wordlength of the values and further reduced the size in hardware implementation.

4.1. Storage in two arrays

Since ISO/IEC MPEG Psychoacoustic Model 2 is employed in both MP3 and AAC, there is no essential difference between them in the utilization of spreading function. However, the psycho-acoustic parameters such as bark values and partitioned bands are different. Because the partitioned bands in AAC are larger than those in MP3 are, the size of the spreading function table in AAC is larger than that in MP3. Besides, in the MP3 reference software [7], the spreading function is by use of look-up tables, whereas in the AAC reference software [7], the spreading function is by use of repeated calculation. Therefore, referring to the MP3 software encoder, LAME [8], where a linear-based array is utilized to store the values of the spreading function and an array is used for indices, we modified the table of the spreading function in AAC. It is illustrated in Figure 4. The non-zero values are stored into the linear-based array segment-by-segment and row-by-row. The start indices and the end indices of each row are also stored in the array of indices, which is an overhead. Compared to the table, the indices array (overhead) is very small. The number of values for storage is shown in Table III, and only 31% of the original number is required. Table IV shows the reduction of tables at different sampling rates. Most of the table's size could be reduced to about one-third of the original.

4.2. Reduction in wordlength

We know that if the wordlength of the values of the table decreases, the area of the table would decrease while the sound quality would degrade. This is a trade-off. However, because of the last characteristic that most of the non-zero values are approximately zero, they have inferior impacts on the convolution in step 5 and thus they can be replaced with zeroes in smaller wordlength.

Table IV: *The reduction of table's size at different sampling rates.*

Sampling Rate (Hz)	Original	After Reduction (with overhead)	%
8000	4304	1732	40.2%
11025	5345	1963	36.7%
12000	5553	2002	36.1%
16000	5809	1993	34.3%
22050	6085	2007	33.0%
24000	6472	2071	32.0%
32000	6292	1980	31.5%
44100	6664	2067	31.0%
48000	6525	2053	31.5%
64000	6010	1958	32.6%
88200	6553	2125	32.4%
96000	6337	2082	32.9%

5. SIMULATION AND RESULTS

In order to verify our methods to reduce the size of the table, we simulated our design by an MPEG AAC encoder with MDCT-based psychoacoustic model [4], which was derived from the reference software. In addition to subjective listening, NMR (Noise-to-Mask-Ratio) and ODG (Objective Difference Grade) [9] were employed as objective audio quality measurement. EAQUAL [10] was used to calculate NMR and ODG. The sound quality would not be affected when only the first method “Storage in two arrays” was employed. Table V shows the wordlength reduction versus the degradation of sound quality. The wordlength of indices (overhead) is calculated as 7 bits, and the number of the values is calculated as

2 x (70(LONG)+42(SHORT)) = 224.

Since the more positive of ODG and the more negative of NMR mean that the sound quality is better, we estimated the sound quality degradation as the difference between the tested and the original sound and normalized them into positive values, namely, larger positive degradation stands for worse sound quality. As has been noted, the reduction of wordlength of the values has little influence on sound quality. Even if the wordlength of the values is decreased to only 5 bits, the sound quality just degrades a little, which is almost not perceivable. When the wordlength is 5 bits, the area of table is only 3808 bits, which is only 3.6% of the reference, which is assumed stored in a RAM or ROM with 16 bit/word for a 16-bit DSP.

6. CONCLUSION

In order to reduce the computational complexity, calculation of spreading function in an AAC encoder is replaced with a look-up table, which is considerably large in the data RAM or ROM on a programmable processor such as DSP for portable devices. In the paper, we employed two methods to reduce the size of the table of the spreading function. Referring to the MP3 software encoder LAME [8], we employed a linear-based array to store the non-zero values and an array to store the indices, by which only one-third of the original size was required. This efficiently reduces the memory usage on the programmable processor. Moreover, for dedicated hardware implementation, we further reduced the memory size by word length analyses of the table values. The size (at sampling rate 44100 Hz) was reduced to only 3.6% compared to that for a 16-bit DSP while audio quality was still maintained.

7. REFERENCES

[1] MPEG. Coding of moving pictures and associated audio for digital storage media at up to 1.5 Mbit/s, part 3: Audio, International Standard IS 11172-3, ISO/IEC JTC1/SC29 WG11, 1992.

[2] MPEG. MPEG-2 Advanced Audio Coding, AAC, International Standard IS 13818-7, ISO/IEC JTC1/SC29 WG11, 1997.

[3] MPEG. Information technology – Coding of audio-visual objects – Part 3: Audio, International Standard IS 14496-3, ISO/IEC JTC1/SC29 WG11, 1999.

[4] Tsung-Han Tsai, Shih-Way Huang, Liang-Gee Chen, “Design of a low power psychoacoustic model co-processor for MPEG-2/4 AAC LC stereo encoder,” in *Proceedings of the 2003 IEEE International Symposium on Circuits and Systems*, vol. 2, pp. 552–555, May 25-28, 2003.

[5] Marc Gayer, Markus Lohwasser, Manfred Lutzky, “Implementing MPEG Advanced Audio Coding and Layer-3 encoders on 32-bit and 16-bit fixed-point processors,” presented at the *AES 115th Convention*, New York, Oct. 10-13, 2003.

[6] Ted Painter, Andreas Spanias, “Perceptual Coding of Digital Audio,” *Proceedings of the IEEE*, vol. 88, issues 4, pp. 451–515, pp. 451-515, Apr. 2000.

[7] `http://www.tnt.uni-hannover.de/project/mpeg/audio/ftp/`

[8] LAME v3.93.1, “LAME Ain’t an MP3 Encoder.” Available: `http://www.mp3dev.org`, Dec. 1, 2002.

[9] ITU-R Recommendation BS. 1387: “Method for objective measurements of perceived audio quality,” July 2001.

[10] `http://www.mp3-tech.org/programmer/sources/eaqual.tgz`

Table V: *The wordlength reduction vs. sound quality degradation (sampling rate 44100 Hz).*

Word length (bit)	LONG	SHORT	Array of value	Array of index	Area (bit)	Area (%)	Quality degrade (ODG)	Quality degrade (NMR)
16 (Ref.)	4900	1764	6664	0	106624	100.0%	none	none
13	625	295	920	224	13528	12.7%	0	0
11	576	255	831	224	10709	10.0%	0	0
9	494	232	726	224	8102	7.6%	0.01	0.01
7	422	195	617	224	5887	5.5%	-0.02	0.02
5	305	143	448	224	3808	3.6%	-0.01	0.01
3	199	96	295	224	2453	2.3%	-0.01	0.02
2	169	91	260	224	2088	2.0%	0.04	0.04