行政院國家科學委員會專題研究計畫　期中進度報告

HE-AAC 的編解碼研究(1/3)

計畫類別：個別型計畫
計畫編號：NSC94-2213-E-009-128-
執行期間：94 年 08 月 01 日至 95 年 07 月 31 日
執行單位：國立交通大學資訊工程學系(所)

計畫主持人：劉啟民

計畫參與人員：許瀚文，楊宗瀚

報告類型：精簡報告
報告附件：出席國際會議研究心得報告及發表論文
處理方式：本計畫可公開查詢

中　華　民　國 95 年 7 月 24 日

行政院國家科學委員會補助專題研究計畫成果報告

HE-AAC 的編解碼研究(1/3)

計畫類別：■個別型計畫　　□整合型計畫

計畫編號：94-2213-E-009-128-

執行期間：94 年 8 月 1 日至 95 年 7 月 31 日

計畫主持人：劉啟民 教授

共同主持人：

本成果報告包括以下應繳交之附件：

□赴國外出差或研習心得報告一份

□赴大陸地區出差或研習心得報告一份

□出席國際學術會議心得報告及發表之論文各一份

□國際合作研究計畫國外研究報告書一份

執行單位：交通大學 資訊工程系

中華民國 94 年 12 月 20 日

HE-AAC 的編解碼研究(1/3)

計畫編號：**94-2213-E-009-128**

執行期限：民國 94 年 8 月 1 日起至民國 95 年 7 月 31 日止

主持人：劉啓民　交通大學資訊工程系

主要參與研究生：許瀚文，楊宗瀚

上年度研究進度報告

由於所發展的技術已均有一展示系統，因此此計畫第一年將以架設軟體平台來整合出系統進行軟體重寫，以利未來的修改、發展、和技術轉移。另一方面將搜索專利已進行性能比較和專利圖建構並提出一基本方法供改進參考。本

第一年的計畫目標有是要

- 完成編解碼系統的軟體發展架構。
- 進行展示和測試系統設計。
- 完成理論研究。
- 對所發展技術進行專利圖研究。
- 完成編解碼系統的參考方法評估比較。

現第一年已完成第一年目標部份成果，並已發表於 AES Convention 118 和 119

- Chi-Min Liu, Li-Wei Chen, Han-Wen Hsu, Wen-Chieh Lee, “Bit Reservoir Design for HE-AAC,” Audio Engineering Society 118th Convention, Barcelona, Span, May 28-31, 2005
- Kan-Chun Lee, Chung-Han Yang, Han-Wen Hsu, Wen-Chieh Lee, Chi-Min Liu, Tzu-Wen Chang, “Design of Time-Frequency Stereo Parameter Sets for Parametric HE-AAC,” Audio Engineering Society 119th Convention, New York, Oct. 7-10, 2005

`並已申請發表以下論文到 AES Convention 120

- Efficient Time-Frequency Grid Decision in HE-AAC Encoder through Dynamic Programming
- Design for High Frequency Adjustment Module in MPEG-4 HE-AAC Encoder

將進行的第二年和第三年計畫如下

第二年計畫目標為根據第一年平台：

- 提出音訊高頻擴充壓縮系統─Spectral Band Replication(SBR)的壓縮方法。
- 提出 SBR 與 AAC 的系統整合方法。
- 提出多聲道音訊壓縮方法、低位元率高頻寬音訊壓縮方法。並提出專利圖的對應。

第三年計畫目標為根據計算複雜度、記憶體需求、和數質精準度等配合本實驗室的數百首音樂測試資料庫與其它著名軟體 Quick Time HE-AAC、Nero HE-AAC、Coding Technology HE-AAC 等作主觀和客觀品質分析。所考量的平台包含

- 擴展現 AAC 測試音樂資料庫 (http://psplab.csie.nctu.edu.tw/projects/index.pl/testbitstreams.html)來應對 HEAAC 。
- 自動測試和品質評估系統。
- 各壓縮方法的優缺點分析與品質改進方法設計。

以下附上以上四論文

Audio Engineering Society

Convention Paper

Presented at the 118th Convention
2005 May 28–31 Barcelona, Spain

Bit Reservoir Design for HE-AAC

Chi-Min Liu [1], Li-Wei Chen [1], Han-Wen Hsu[1], and Wen-Chieh Lee [1]

[1] PSPLab, Computer Science and Information Engineering, National Chiao-Tung Univiersity, Hsin-Chu, 33050, Taiwan
cmliu@csie.nctu.edu.tw

ABSTRACT

High Efficiency AAC (HE-AAC) has included the Spectral Band Replication (SBR) in combination with AAC to achieve high audio quality at bit rates lower than 96 kbits per second. SBR reconstructs high frequency signal through replicating the low frequency parts. The bits allocated to AAC encoder module and SBR module decides the quality and compression efficiency. In the past, we have designed the bit reservoir for AAC to reserve and predict the bits necessary for each time frame. The bit reservoir should be extended for HE-AAC especially for the SBR module. This paper considers the design of the bit reservoir for the HE-AAC. The efficiency of bit reservoir is verified through extensive objective tests.

Introduction

High Efficiency AAC (HE-AAC) is the extension of the conventional AAC codec by supporting the Spectral Band Replication (SBR) module [1]-[4]. The block diagram of the HE-AAC is illustrated in Figure 1. The audio signal is fed into the filterbank and split into high frequency signal $s_h(n)$ and low frequency signal $s_l(n)$ through a filterbank. The low frequency signal $s_l(n)$ is half the sampling rate of the original signal. The high frequency signal $s_h(n)$ is reconstructed through the band replication technique from the low frequency signal $s_l(n)$. The replication parameters are used to keep the reconstructed high frequency bands perceptually similar to the original high frequency bands. The bit reservoir finds the suitable bit distribution between the AAC encoder and SBR encoder according to the signal contents and the available bit budget.

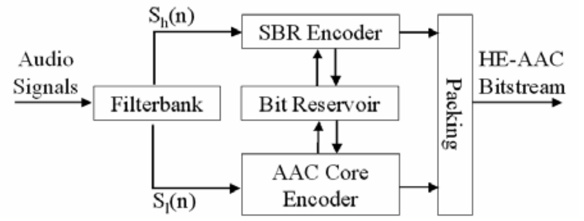

Figure 1 Block diagram of HE-AAC encoder

Bit reservoir has been adopted in MPEG-1 Layer 3 (MP3) [5] and MPEG-4 Advanced Audio Coding (AAC) [6] to control the bit variation among encoded frames. The mechanism provides the space to loan or deposit bits to control the audio quality under a bit rate constraint. The efficient bit reservoir design [8] for MP3 and AAC has been developed through the demand estimator and the budget regulator. This paper extends the design to the HE-AAC.

The bit reservoir deciding the dynamic of the available bits among frames is the quality buffer avoiding the severe quality degradation from critical frames. The previous work [8] considered the design basically through two modules: the demand estimator and the budget regulator. This paper extends the two modules to the HE-AAC.

On the bit allocation for the AAC encoder and SBR encoder, the problem leads to a closely dependent issue. The SBR which reconstructs the high frequency signal from the low frequency signal encoded needs to check the AAC encoding results to predict the bit required. However, the AAC also need to know the bit consumed by SBR to efficiently encode the signals based on the available bits. Furthermore, the bit reservoir should control the quality among frames with regulation on the bit variation. Although the kind of deadlock or interdependent issue can be approached through an iterative manner, the complexity would increases tremendously due to the inherent complexity in AAC encoder and SBR encoder. This paper proposes a single iteration approach on the bit reservoir. Based on the demand estimator and budget regulator, we design a SBR bit estimator through a recurrent mechanism. Also, on the budget regulator, we modify the budget regulator that was used in AAC to be the one for both AAC and SBR. The new reservoir was conducted through an immense objective and subjective tests to show the efficiency.

Bit Reservoir Control for AAC

The basic concept of bit reservoir in AAC is to deposit bits from easy frames and loan bits for difficult frames. In our previous work in AAC, an efficient bit reservoir design with demand estimator and budget regulator is proposed [8]. The demand estimator predicts the required bits according to the audio contents while the budget regulator controls the budget according the accumulated bits. The balance of the two modules maintains the average bit rate among frames and improves perceptual quality. This mechanism can automatically adapt with the various bit rates, the various encoders like MP3 and AAC, and the preferred bit rate scenario like the constant bit rate (CBR), the variable bit rate (VBR), and the average bit rate (ABR).

Demand Estimator

Johnston [9][10] used the perceptual entropy (*PE*) to reflect the bits required for transparent quality. But the PE does not reflect the bits required for the cases where the transparent quality is not achievable under limited bit rates. Therefore, another perceptual criterion, allocation entropy (AE) [11], is proposed. The AE could well reflect the bits required to have the graceful degradation and have put into consider the bandwidth proportional noise-shaping criterion. The bits required for the frame can be obtained from

$$AE(f) = \sum_q AE_q = \sum_q W_q * \log_{10}(SMR'_q + 1), \qquad (1)$$

where q is the index of quantization band, f is the frame index, W_q is the number of spectrum lines in quantization band q, $SMRq$ is the signal-to-noise ratio in quantization band q. In order to control the average quality, there is an average demand aligned to the average bit rates. The average demand $AE_{average}$ can be estimated through the average over the past N frames

$$AE_{average} = \frac{\sum_{f=1}^{N} AE(f)}{N}, \qquad (2)$$

Through the average AE, we could evaluate the demand ratio, *D*(*f*).

$$D(f) = \frac{AE(f) - AE_{average}}{AE_{average}}, \qquad (3)$$

D(*f*) represents the current demand over the previous *N* coding unit. The demand ratio should be transformed into R_{demand} *(f)* by a transform function to shape the curve and clip the upper/lower bounds:

$$R_{demand}(f) = \eta(D(f)) \qquad (4)$$

Figure 2 illustrates the example $\eta(\cdot)$ used in AAC encoder.

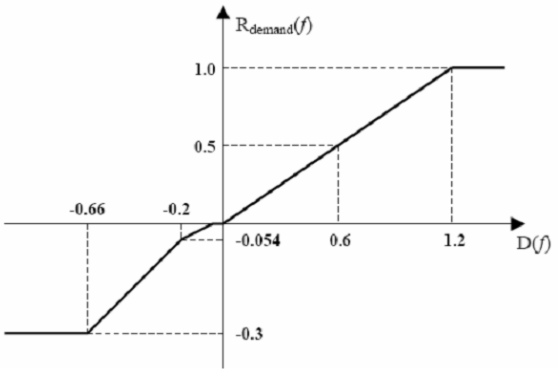

Figure 2 Demand Curve for AAC

Budget Regulator

The budget regulator decides the available bits according to the preferred scenario. We define a budget ratio and adjust the budget ratio with the fullness (denoted as *F*) of the bit reservoir as depicted in Figure 4. The fullness F is evaluated through

$$F = \frac{S}{S_{MAX}} \qquad (5)$$

where *S* is the current accumulated budget size, S_{MAX} is the maximum allowable reservoir size.

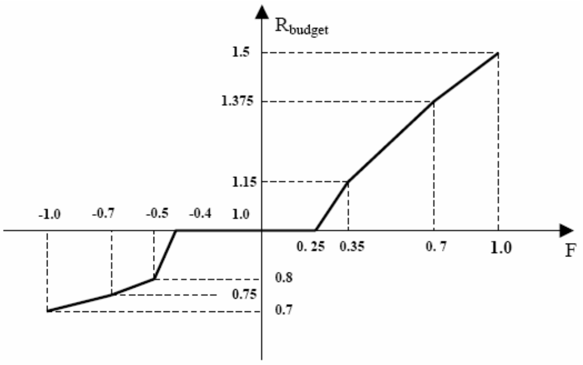

Figure 3 Budget curve for AAC

Hence the allocated bits for encoding unit of AAC encoder is derived from

$$\begin{aligned}&Allocated\_bits\_for\_AAC\\&= mean\_bits + R_{demand} * mean\_bits * R_{budget}\end{aligned} \tag{6}$$

where *mean_bits* is derived from the desired bit rate for AAC encoder, R_{budget} comes from the budget curve.

Bit Reservoir for HE-AAC

To extend the bit reservoir to HE-AAC, we need to have the estimator and the regulator for the SBR encoder. The estimator needs to predict the bits required for SBR part while regulator needs to leave a budget for the SBR encoder. Also, the estimator and regulator in SBR should be suitably combined with the estimator and the regulator in AAC to have global control. The block diagram of our new design is depicted in Figure 5. The SBR bit estimator predicts the bits required for the SBR part. Also, AAC bit estimator is the same as the estimator in the previous section. The budget regulator assigns bits to AAC encoder leaving some budget for SBR encoder with the following equation

$$\begin{aligned}&Allocated\_bits\_for\_AAC = mean\_bits +\\&R_{demand} * mean\_bits * R_{budget} - SBR\_bits,\end{aligned} \tag{7}$$

where *SBR_bits* is the number of bits from the SBR demand estimator. On the coding sequence, we proceed first with the AAC encoder based on the allocated bits and then the SBR encoder. The features of the algorithm can be considered from two aspects. We use one common budget regulator while two demand estimators for the AAC and SBR encoders. Second, we leave the budget for the SBR without directly regulating the encoded bits in SBR encoder. The section considers the design of the mechanism.

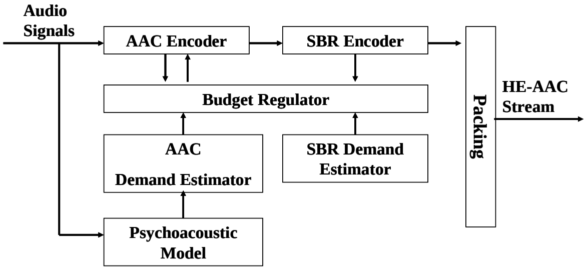

Figure 4 Block diagram of bit reservoir for HE-AAC.

Bit Estimator for SBR

80 kbps	Min		Max		Avg		Standard Deviation	
	Bits	Percentage	Bits	Percentage	Bits	Percentage	Bits	Percentage
es01	151	4.29%	423	12.03%	276.02	7.85%	47.69	1.36%
es02	151	4.29%	391	11.12%	256.91	7.30%	51.49	1.46%
es03	151	4.29%	455	12.94%	265.76	7.56%	47.53	1.35%
sc01	151	4.29%	279	7.93%	214.58	6.10%	29.82	0.85%
sc02	151	4.29%	327	9.30%	230.14	6.54%	31.15	0.89%
sc03	151	4.29%	383	10.89%	261.44	7.43%	41.31	1.17%
si01	151	4.29%	399	11.34%	244.71	6.96%	48.09	1.37%
si02	151	4.29%	439	12.48%	267.62	7.61%	57.53	1.64%
si03	151	4.29%	319	9.07%	220.96	6.28%	30.92	0.88%
sm01	151	4.29%	343	9.75%	249.37	7.09%	33.42	0.95%
sm02	151	4.29%	503	14.30%	241.33	6.86%	73.31	2.08%
sm03	151	4.29%	399	11.34%	248.02	7.05%	44.74	1.27%

Table 1 The minimum, maximum, average, and standard deviation of bits usage at 80 kbps. The “Min” and “Max”, “Avg”, and Standard Deviation columns denote respectively the minimum, the maximum bits, the average bits, and the standard deviation used in the SBR encoder among all the frames in the correspondent track. The percentage in each the above category column is the bit percentage for the budget in a frame at bit rate 80 kbps.

The bit rate of control parameters of SBR encoder varies with the SBR encoder modules inside, e.g. Grid and High Frequency Generation (HF Generation) [1]. The minimum, maximum, average, and standard deviation value of bits usage of each track at 80 kbps is shown in Table 1. We also list the percentage of these values in the mean bits at 80kbps. The standard deviation is derived by

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2} \quad \text{for } \forall x_i > 151 \qquad (8)$$

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \qquad \text{for } \forall x_i > 151 \qquad (9)$$

where x_i is the number of bits in each SBR frame, n is the number of frames exclusive the silence frames where a threshold 151 bits have been used as the selection criterion. The twelve test tracks recommended by MPEG are commented in Table 2. These tracks include the critical music balancing on the percussion, string, wind instruments, and human vocal. Figures 6-9 illustrates the consumed SBR bits with respect to audio frames for three test tracks. From the data, the bits used in SBR have small deviation. Also, the percentage of the SBR is small compared to AAC encoder. Due to the relatively stable on the bit variation, the proposed bit reservoir in Figure 4 has not regulated the bits used in a frame. But the bits consumed will be taken into account into the left budget. Although this kind of mechanism may lead to a situation that the bits used may be greater than the allowable budget originally used in the bit reservoir but the exceeding amount is small and can be consumed in the budget of the proceeding frame. So, from the viewpoint of two frames, the budget has been constrained to the allowable amount.

We have included three designs for the SBR demand estimators. The first method is to have a fix estimation by average. We can just provide the average SBR value for the SBR demand estimator. The second method is to predict the demand from the consumed SBR bits in the previous frame. The third one is to have the estimation from the average of a period of the previous frames.

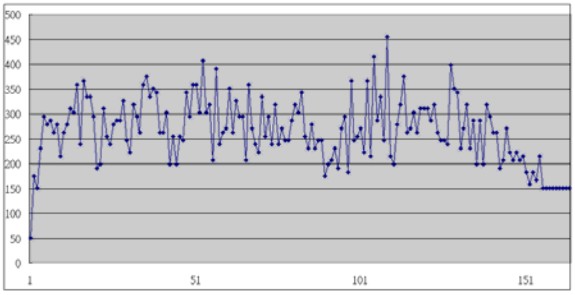

Figure 6 Natural vocal (es03).

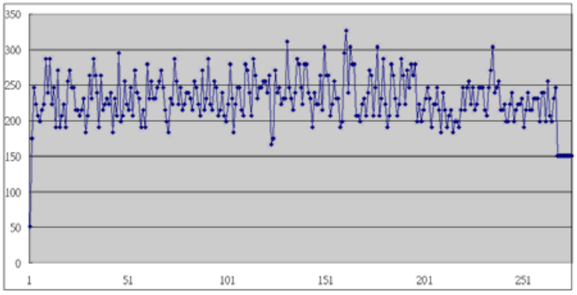

Figure 7 Complex sound (sc02).

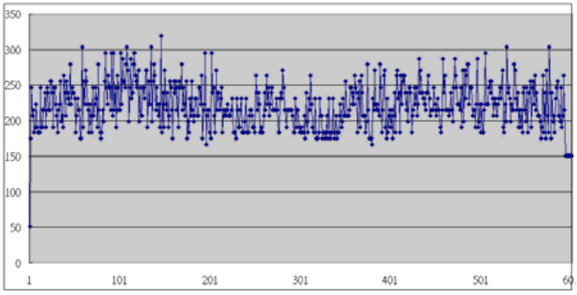

Figure 8 Harmonic (si03).

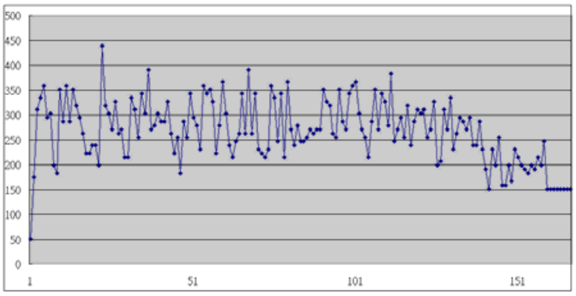

Figure 9 Transient (si02).

HE-AAC Bit Allocation

For HE-AAC bit allocation, the formula in (6) must be modified to consider both AAC encoder and SBR encoder. Hence the allocated bits for whole HE-AAC encoder is derived from

$$Allocated\_bits\_for\_HEAAC$$
$$= mean\_bits' + R'_{demand} * mean\_bits * R_{budget}$$

(10)

where *mean_bits'* is derived from the desired average

Track		Signal Description			
		Signal	Mode	Time(sec)	Remark
1	es01	Vocal(Suzan Vega)	Stereo	10	(c)
2	es02	German speech	Stereo	8	(c)
3	es03	English speech	Stereo	7	(c)
4	sc01	Trumpet solo and orchestra	Stereo	10	(d)
5	sc02	Orchestral piece	Stereo	12	(d)
6	sc03	Contemporary pop music	Stereo	11	(d)
7	si01	Harpsichord	Stereo	7	(b)
8	si02	Castanets	Stereo	7	(a)

9	si03	pitch pipe	Stereo	27	(b)
10	sm01	Bagpipes	Stereo	11	(b)
11	sm02	Glockenspiel	Stereo	10	(a) (b)
12	sm03	Plucked strings	Stereo	13	(a) (b)

Remark:

(a) Transients: pre-echo sensitive, smearing of noise in temporal domain.

(b) Tonal/Harmonic structure: noise sensitive, roughness.

(c) Natural vocal (critical combination of tonal parts and attacks): distortion sensitive, smearing of attacks.

(d) Complex sound: stresses the Device Under Test.

Table 2 The twelve test tracks recommended by MPEG.

bit rate for HE-AAC encoder, R_{budget} comes from the budget curve depicted in Figure 4, and R'_{demand} is derived from

$$R^{'}_{demand}(f) = \eta(D^{'}(f)) \qquad (11)$$

$$D^{'}(f) = \frac{AE(f) - AE_{average} + B}{AE_{average} + B}, \qquad (12)$$

where $\eta(\cdot)$ is the transform function comes from the demand curve as shown in Figure 3, *AE(f)* is derived from (1), $AE_{average}$ is derived from (2), and *B* is derived from

$$B = AE(f)\frac{SBR\_bits}{mean\_bits - SBR\_bits}, \qquad (13)$$

where *SBR_bits* is calculated by our SBR bit estimator. The flowchart of the bit reservoir design for HE-AAC is illustrated in Figure 10.

Results

For the objective quality evaluation, we mainly adopt the PEAQ system (perceptual evaluation of audio quality) which is the recommendation system by ITU-R Task Group 10/4. The system includes a subtle perceptual model to measure the difference between two tracks. The objective difference grade (ODG) is the

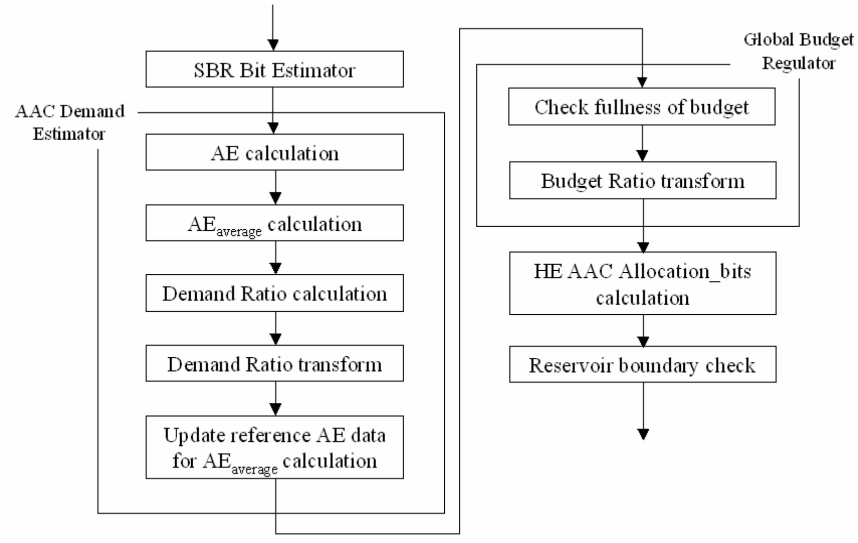

Figure 10 Flowchart of bit reservoir design for HE-AAC.

output variable from the objective measurement method. The ODG values should range from 0 to -4, where 0 corresponds to an imperceptible impairment and -4 to impairment judged as very annoying. The improvement up to 0.1 is usually perceptually audible. The PEAQ has been widely used to measure the compression technique due to the capability to detect perceptual difference sensible by human hearing systems. The following experiments are based on this PEAQ system [12].

We first illustrate three kinds of different bit estimator designs in our bit reservoir at bit rate 48kbps, 64kbps, 80kbps, and 96kbps. The results are shown in Figure 11-14.

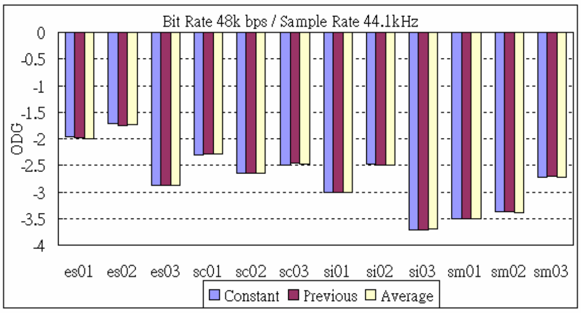

Figure 11 Objective measurements through the ODGs for three kinds of different SBR bit estimator designs. This test is under 48kbps and sample rate at 44.1kHZ. Constant: the method of using fixed value, the constant value is set as 225. Previous: the method of referring to value of previous one frame. Average: the method of referring to average value of previous frames, the reference length is set as 5.

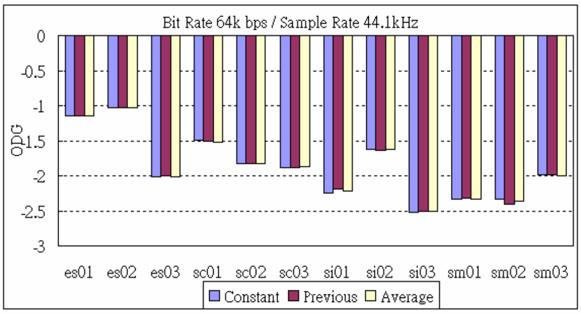

Figure 12 Objective measurements through the ODGs for three kinds of different SBR bit estimator designs. This test is under 64kbps and sample rate at 44.1 kHZ. Constant: the method of using fixed value, the constant value is set as 225. Previous: the method of referring to value of the previous frame. Average: the method of referring to average value of previous frames, the reference length is set as 5.

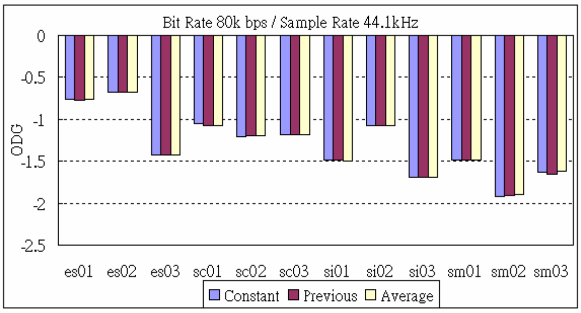

Figure 13 Objective measurements through the ODGs for three kinds of different SBR bit estimator designs. This test is under 80kbps and sample rate at 44.1 kHZ. Constant: the method of using fixed value, the constant value is set as 225. Previous: the method of referring to value of the previous frame. Average: the method of referring to average value of previous frames, the reference length is set as 5.

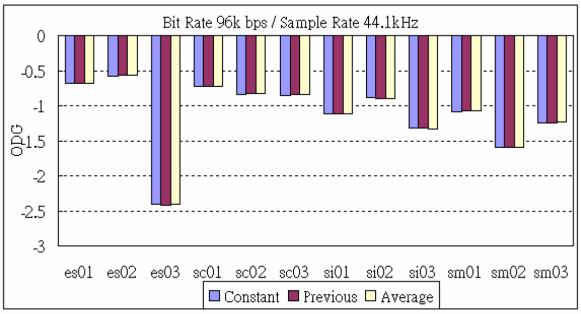

Figure 14 Objective measurements through the ODGs for three kinds of different SBR bit estimator designs. This test is under 96kbps and sample rate at 44.1 kHZ. Constant: the method of using fixed value, the constant value is set as 225. Previous: the method of referring to value of the previous frame. Average: the method of referring to average value of previous frames, the reference length is set as 5.

From the testing data shown above, we found that the track with worst ODG value in bit rate 48kbps is si03 and the ODG value is -3.71, -3.71, and –3.7 in Constant, Previous, and Average method. The track with worst ODG value in bit rate 64kbps is si03 and the ODG value is –2.52, -2.51, and –2.5 in Constant, Previous, and Average method. The track with worst ODG value in bit rate 80kbps is sm02 and the ODG value is –1.92, -1.91, and –1.9 in Constant, Previous, and Average method. The track with worst ODG value in bit rate 96kbps is es03 and the ODG value is –2.41, -2.42, and –2.41 in Constant, Previous, and Average method.

The average ODG value of the twelve test tracks in Constant, Previous, and Average method is –2.7325, -2.7308, and –2.7383 at 48kbps, –1.87, -1.8683, and –1.8692 at 64kbps, –1.3, -1.3033, and –1.3 at 80kbps, –1.1092, -1.1083, and –1.1075 at 96kbps. Although the average ODG values of different bit estimator methods are similar at different bit rates, we choose Average method as our default strategy in order to avoid the risk from some critical tracks.

In order to illustrate the improvement of the SBR encoder in our HE-AAC design, we first show the result of encoding by HE-AAC encoder and decoding without the SBR part of NCTU-HEAAC, Coding Technologies [13], and Nero 6.6.0.8 [14] at different bit rates in Figure 15. Also, we compare the objective quality of HEAAC encoder of NCTU-HEAAC, Coding Technologies, and Nero 6.6.0.8 at different bit rate in Figure 16. For each statistics line in Figure 15-16, the top arrow represents the maximum ODG value, the down cross represents the minimum ODG value, and the middle square represents average ODG value among the twelve test tracks.

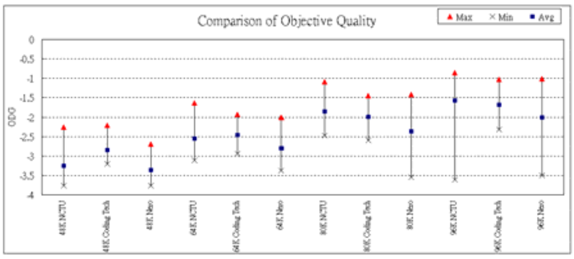

Figure 15 The ODG range comparison of NCTU-HEAAC, Coding Technologies, and Nero 6.6.0.8 with encoding by these HEAAC codecs but decoding without SBR part under bit rate 48kbps, 64kbps, 80kbps, and 96kbps.

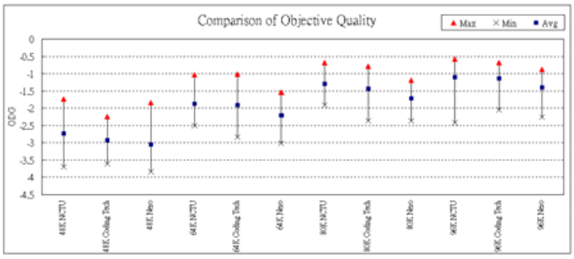

Figure 16 The ODG range comparison of NCTU-HEAAC, Coding Technologies, and Nero 6.6.0.8 under bit rate 48kbps, 64kbps, 80kbps, and 96kbps.

The testing results illustrate that the quality of NCTU-HEAAC without SBR part are worse than Coding Technologies without SBR part under low bit rate 48kbps and 64kbps. But with SBR part, the quality of NCTU-HEAAC gains improvement on average up to 0.515 at 48kbps and 0.6808 at 64kbps. Hence the quality of NCTU-HEAAC on average is 0.1925 better than Coding Technologies at 48kbps and 0.0458 better than Coding Technologies at 64kbps. Under other bit rates, the quality of HEAAC in NCTU-HEAAC is 0.1275 better than that in Coding Technologies at 80kbps and 0.0267 better than in Coding Technologies at 96kbps. With comparing to Nero 6.6.0.8, the quality of NCTU-HEAAC is 0.3059 better at 48kbps, 0.3341 better at 64kbps, 0.4142 better at 80kbps, and 0.2892 better at 96kbps.

conclusion

This paper has extended the bit reservoir design in AAC to HE-AAC and proposed the mechanisms of SBR bit estimator and global budget regulator. The experiments have shown that the bit reservoir is well fit the encoder for various bit rates and preferred scenario.

acknowledgements

This work was supported by National Science Council under NSC91-2622-E009-003 and InterVideo Digital Tech. under 792171.

REFERENCES

[1] ISO/IEC, “Text of ISO/IEC 14496-3:2001/FPDAM 1, Bandwidth extensions,” ISO/IEC JTC1/SC29/WG11/N5203, October 2002, Shanghai, China.

[2] M. Dietz, L. Liljeryd, K. Kjörling, O. Kunz, “Spectral Band Replication, a novel approach in audio coding,” at the 112th AES Convention, Munich, May 10–13, 2002.

[3] M. Wolters, K. Kjörling, D. Homm, H. Purnhagen, “Acloser look into MPEG-4 High Efficiency AAC,” at the 115th AES Convention, New York, USA, October 10–13, 2003.

[4] H.W. Hsu, C.M. Liu, and W.C. Lee, “Audio Patch Method in MPEG-4 HE-AAC Decoder,” at the 117th AES Convention, San Francisco, USA, October 28~31, 2004.

[5] ISO/IEC JTC1/SC2/WGII MPEG, International Standard ISO 11172-3 “Coding of moving pictures and associated audio for digital storage media at up to about 1.5Mbit/s”

[6] ISO/IEC 14496-3, “Information Technology–Coding of Audiovisual objects, Part3: Audio.”

[7] 3GPP TS 26.410 V6.0.0 (2004-09), http://www.3gpp.org

[8] C.M. Liu, L.W. Chen, M.T. Su, W.C. Lee, Y.H. Hsiao, Z.W. Li, and C.T. Chien, “Efficient Bit Reservoir Design for MP3 and AAC,” at the 117th

AES Convention, San Francisco, USA, October 28~31, 2004.

[9] J.D. Johnston, “Estimation of Perceptual Entropy Using Noise Masking Criteria,” *ICASSP*, 1988, pp.2524-2527.

[10] J.D. Johnston, “Transform coding of audio signals using perceptual noise criteria,” *IEEE J. Select. Areas Commun.*, vol. 6, Feb. 1988, pp.314-323.

[11] C.M. Liu, W.C. Lee, and Y.H. Hsiao, "M/S Coding Based on Allocation Entropy," *Digital Audio Effect*, London, UK, September 8-11, 2003.

[12] ITU Radiocommunication Study Group 6, “Draft Revision to Recommendation ITU-R BS.1387-Method for objective measurements of perceived audio quality”.

[13] Coding Technologies, aacPlusEval Evaluation Package, https://portal.codingtechnologies.de/eval/aacPlusEval/.

[14] Nero, http://www.nero.com

eering Society
Paper

t the 119th Convention
er 7–10 New York, New York USA

Efficient Design of Time-Frequency Stereo Parameter Sets for Parametric HE-AAC

Kan-Chun Lee, Chung-Han Yang, Han-Wen Hsu, Wen-Chieh Lee, Chi-Min Liu, and Tzu-Wen Chang

PSPLab, Computer Science and Information Engineering, National Chiao-Tung University, Hsin-Chu, 33050, Taiwan
cmliu@csie.nctu.edu.tw

ABSTRACT

Parametric Stereo Coding (PS) tool is used to reconstruct stereo signal from the monaural signal. The tool can be jointly used with the HE-AAC to have high compression ratio and is referred to as the parametric HE-AAC in this paper. The PS tool is able to capture the stereo image of the audio input signal into a limited number of parameters, requiring only a small overhead. In MPEG-4 HE-AAC, the PS tool segments a frame into several regions in time domain and into stereo bands in frequency domain to deliver stereo parameter sets. This paper considers the design of the stereo parameters. These methods are integrated in the NCTU-HE-AAC and the objective experiments are conducted to check the quality.

INtroduction

The Parametric Stereo Coding (PS) is a tool in the MPEG-4 audio parametric coding scheme for compressing high quality stereo audio at bit rates around 24 kbps. From the original stereo input signal and the monaural downmix of the stereo input signal generated by the parametric stereo coding tool, the PS module extracts the stereo parameter sets. The parametric stereo decoding can reconstruct the stereo signal by using these parameter sets and the monaural downmix signal. For the coding of the monaural downmix signal, it can operate in combination with any monaural coder such as SSC [15] or MPEG-4 HE-AAC [16][17][18][19]. This paper will focus on the PS tool with MPEG-4 HE-AAC. The block diagram of the PS encoder with HE-AAC is illustrated in Figure 1. Figure 2 shows the coding efficiency of AAC, HE-AAC, and Parametric HE-AAC. The parametric HE-AAC extends the high quality audio coding of AAC and HE-AAC to bit rates 24-48 kbps.

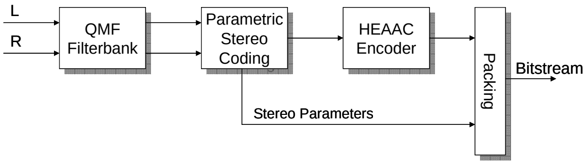

Figure 1 Block diagram of the PS encoder with HE-AAC

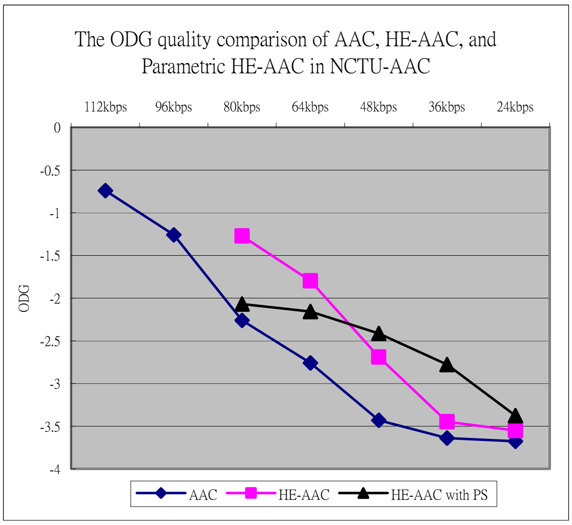

Figure 2 The ODG quality comparison of AAC, HE-AAC, and Parametric HE-AAC in NCTU-AAC

To regenerate the stereo signal, the PS decoder extracts one parameter set for each stereo band of each time region in a frame. Each parameter set contains four parameters: inter-channel intensity difference (IID), inter-channel coherence (ICC), inter-channel phase difference (IPD) and overall phase difference (OPD). In decoder, the mixing module transforms the parameter set into matrices to regenerate the stereo signal. The PS tool is designed at bit rates around 24kbit/s, the number of parameter sets must be controlled according to the signal contents and the available bits. The parameter sets associated with the bands and regions reflect the time-frequency resolution required for the signal contents. However, the number of sets also reflects the bits required. For the time resolution defined in PS tool, it allows a number of time regions up to four. For the frequency resolution, there are three configurations, i.e. 10, 20, and 34 stereo bands. In other words, the number of parameter sets can range from 0 to 136 in a frame. Apparently the decision of the number of parameter sets and the evaluation of the associated parameters will be the kernel deciding the coding efficiency and coding quality in the parametric stereo coding. This paper has considered the efficient algorithm for the design of the parameter sets which contain IID and ICC values. The block diagram of PS modules discuss in this paper is illustrated in Figure 3.

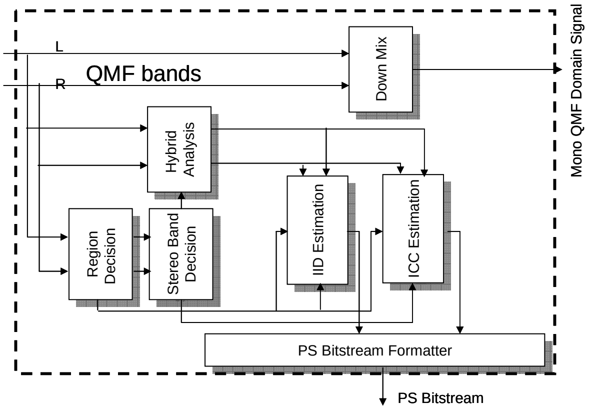

Figure 3 Block diagram of the PS modules

Time/frequency stereo parameter set

In the literature, there have been limited researches on the determination of the stereo parameter sets. The code used in 3GPP 0 uses basically a fixed parameter set, which always uses fixed stereo band resolution with two time regions, each region is half of frame, in each frame. We consider the design of parameter sets by the two modules in the PS encoder: "Region Decision (RD)" for time domain and "Stereo Band Decision (SBD)" for frequency domain. Region Decision is designed to find the time regions for band-limited signals to share the same stereo parameter. The other module is Stereo Band Decision which decides the frequency resolution sharing the same stereo parameters. The band decision module informs "Hybrid Analysis" module to split lower QMF subbands to achieve a higher frequency resolution decided. The bands after hybrid analysis are generally referred to as sub-subbands. The stereo bands sharing the same stereo parameters contain one or multiple of sub-subbands. In this paper, RD module uses QMF subband information as input, and SBD module uses QMF subband information and time region information as input.

Region Decision

First, we consider how the PS decoder reconstructs the stereo signal by using the monaural signal and the delivered parameter sets. In each stereo band of a time region, there is only one stereo parameter set and this parameter set is assigned to the last time slot in the region. The parameter sets of other time slots in the region are assigned by means of interpolation. Our algorithm is to find the time regions and its border positions by the information of QMF subbands, such that the interpolation error will be small enough. Because the QMF subband has higher frequency resolution than the stereo band, the time regions calculated in stereo band domain are also calculated in QMF subband domain Therefore, RD can be done in QMF subband domain.

Dynamic Programming Applied to Deciding Time Regions

Dynamic programming (DP) is a subset of the general theory concerned with discrete sequential decisions. The varied history and an extensive bibliography can be found in the article by Silverman and Morgan [21]. The prolific application of the DP to various fields has been credited to Professor Richard Bellman [22]. The basic principle of DP is to break an optimization problem down into stages of decisions that follow a criterion leading to a recurrence relations. For audio compression, we have already applied the DP algorithm to efficient design the MS coding, and Huffman code book search [23]. In this paper, we will consider the applying of the

DP algorithm to the region decision.

Applying DP to RD, we need to define the problem into a decision problem and break the problem into stages of the recursive decision of sub-problems. Let $e_{i,j}^{s_1,s_2,\dots,s_k}$ denote the reconstruction error of parameters in the range from time slot i to time slot j under the time region set with k inner borders, say $s_1, s_2, \dots, s_k$, and there is a border at time slot *i-1*. Also, let $E_{i,j}^{(k)}$ denote the minimum construction error of parameters in the range from time slot i to time slot j among all possible time region sets with k inner borders and there is a border at time slot *i-1*. That is,

$$E_{i,j}^{(k)} = \min\left\{e_{i,j}^{s_1,s_2,\dots,s_k} \mid i \le s_1 < s_2 < \dots < s_k < j\right\}, \forall k > 0$$

Furthermore, let $E_{i,j}^{(0)}$ for the case whose no inner border. The optimum sub-structure of $E_{i,j}^{(k)}$ can be explored as follow. Assume the optimum k borders is $s'_1, s'_2, \dots, s'_k$, we have

$$E_{i,j}^{(k)} = E_{i,s'_1}^{(0)} + e_{s'_1+1,j}^{s'_2,\dots,s'_k}$$

By the definition of $E_{s'_1,j}^{(k-1)}$, it gives

$$E_{i,j}^{(k)} = E_{i,s'_1}^{(0)} + e_{s'_1+1,j}^{s'_2,\dots,s'_k} \ge E_{i,s'_1}^{(0)} + E_{s'_1+1,j}^{(k-1)}$$

Since $E_{i,j}^{(k)}$ is the optimum solution, the equality must hold,

$$E_{i,j}^{(k)} = E_{i,s'_1}^{(0)} + E_{s'_1+1,j}^{(k-1)}$$

Hence, to inspect all the possible s'_1, $E_{i,j}^{(k)}$ is determined in (5) and Figure 4 illustrates this condition.

$$E_{i,j}^{(k)} = \min_{t \in \{i,i+1,\dots,j\}} \left\{E_{i,t}^{(0)} + E_{t+1,j}^{(k-1)}\right\}$$

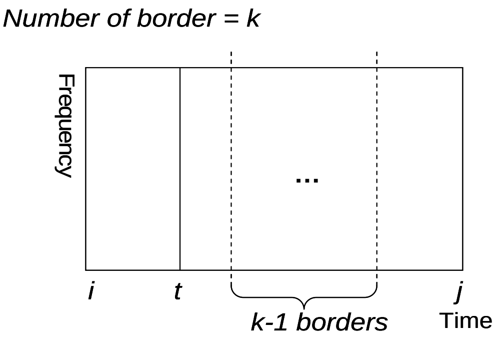

Figure 4 dynamic programming in Region Decision

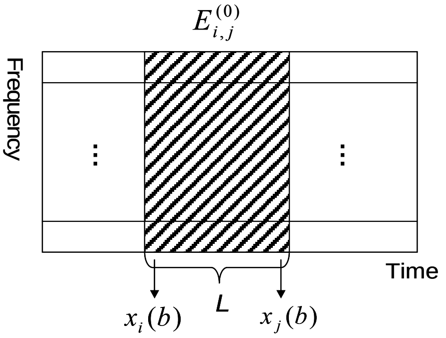

Figure 5 $E_{i,j}^{(0)}$, the time region from time slot i to time slot j

Each $E_{i,j}^{(k)}$ can be recursively constructed for all k>0. Therefore, we need to calculate $E_{i,j}^{(0)}$ for all i and j>i at initialization of dynamic programming. Figure 5 illustrates $E_{i,j}^{(0)}$, where b is the QMF subbands index, x is the variable which can be IID or ICC value, and L is the length of $E_{i,j}^{(0)}$.

To simulate what the PS decoder does, we define the slope as

$$\Delta_{i,j}(b) = \frac{x_j(b) - x_{i-1}(b)}{L}$$

By $\Delta_{i,j}(b)$, we can calculate the reconstruction error.

$$E_{i,j}^{(0)} = \sum_{b}\sum_{l=0}^{L-2}\left|x_{i+l}(b) - x_{i-1}(b) - \Delta_{i,j}(b)\times(l+1)\right|$$

Because the reconstruction error is combined with IID and ICC reconstruction errors, each IID and ICC must be normalized before to calculate the reconstruction error of each time region. The normalized IID and ICC values, $\tilde{iid}_{t,b}$ and $\tilde{icc}_{t,b}$, of time domain index *t* and QMF subband index *b* can be obtained by the following equations:

$$\tilde{iid}_{t,b} = \frac{iid_{t,b} - mean(iid_{t,b})}{\sigma(iid_{t,b})}$$

and $$\tilde{icc}_{t,b} = \frac{icc_{t,b} - mean(icc_{t,b})}{\sigma(icc_{t,b})},$$

where *σ(iid$_{t,b}$)* and *σ(icc$_{t,b}$)* are the standard derivation of $iid_{t,b}$ and $icc_{t,b}$. The dynamic programming algorithm calculates the minimum errors from the case without inner border, up to the case with four inner borders in the frame.

Instead of finding the regions with minimum reconstruction error, we use a threshold to detect the regions whose region number among all the regions with reconstruction error less than the threshold is minimum. This is because in a frame, the number of regions increases, the total reconstruction error decreases, but the number of bits used in a frame increases. Under the requirement of quality and the limited available bits, the threshold here provides this tradeoff condition.

Stereo Band Decision

Since only one stereo parameter set can be sent for a stereo band in a region, the QMF subbands in the same stereo band is supposed to have similar characteristics. In other words, these QMF subbands should have similar stereo parameter sets.

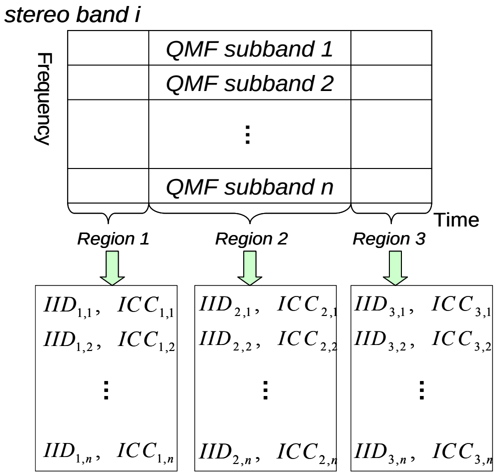

Figure 6 Stereo Band Decision in stereo band *i*

Figure 6 illustrates a stereo band *i*, which contains QMF subbands 1, 2, …, *n*. There are three time regions in stereo band *i*. For each time region and QMF subband, we calculate a parameter set, including IID and ICC, *IID$_{T,b}$* and *ICC$_{T,b}$*, for region *T* and QMF subband *b*. For each time region, the IIDs and ICCs must be normalized. The normalized *IID$_{T,b}$* and *ICC$_{T,b}$*, $\tilde{IID}_{T,b}$ and $\tilde{ICC}_{T,b}$, are obtained through

$$\tilde{IID}_{T,b} = \frac{IID_{T,b} - mean(IID_{T,b})}{\sigma(IID_{T,b})}$$

and $$\tilde{ICC}_{T,b} = \frac{ICC_{T,b} - mean(ICC_{T,b})}{\sigma(ICC_{T,b})}.$$

Using these normalized IIDs (ICCs), we first calculate the variance of IIDs (ICCs) in each stereo band containing more than one QMF subband. If the variance of IIDs (ICCs) in this stereo band is small enough, this QMF subbands in this stereo band have the similar IIDs (ICCs) and can be combined. Since the IIDs and ICCs are both normalized, we can sum up all variances under three kinds of stereo band resolutions: 10, 20, or 34 stereo bands. Finally, the stereo band resolution with the smallest summation value is the output of this module.

Results

For objective quality evaluation, we mainly adopt the PEAQ system (perceptual evaluation of audio quality) which is the recommendation system by ITU-R Task Group 10/4. The system includes a subtle perceptual model to measure the difference between two tracks. The objective difference grade (ODG) is the output variable from the objective measurement method. The ODG values should range from 0 to −4, where 0 corresponds to an imperceptible impairment and −4 to impairment judged as very annoying. The improvement up to 0.1 is usually perceptually audible. The PEAQ has been widely used to measure the compression technique due to the capability to detect perceptual difference sensible by human hearing systems. Following experiments are based on this PEAQ system [24]. The twelve test tracks recommended by MPEG are shown in Table 1. These tracks include the critical music balancing on the percussion, string, wind instruments, and human vocal.

Table 1 The twelve test tracks recommended by MPEG.

Track		Signal Description			
		Signal	Mode	Time(sec)	Remark
1	es01	Vocal(Suzan Vega)	Stereo	10	(c)
2	es02	German speech	Stereo	8	(c)
3	es03	English speech	Stereo	7	(c)
4	sc01	Trumpet solo and orchestra	Stereo	10	(d)
5	sc02	Orchestral piece	Stereo	12	(d)
6	sc03	Contemporary pop music	Stereo	11	(d)
7	si01	Harpsichord	Stereo	7	(b)
8	si02	Castanets	Stereo	7	(a)
9	si03	pitch pipe	Stereo	27	(b)
10	sm01	Bagpipes	Stereo	11	(b)
11	sm02	Glockenspiel	Stereo	10	(a) (b)
12	sm03	Plucked strings	Stereo	13	(a) (b)

Remark:

(a) Transients: pre-echo sensitive, smearing of noise in temporal domain.

(b) Tonal/Harmonic structure: noise sensitive, roughness.

(c) Natural vocal (critical combination of tonal parts and attacks): distortion sensitive, smearing of attacks.

(d) Complex sound: stresses the Device Under Test.

To verify our algorithm has effect on the PS coding, we illustrate the difference between PS coding with always 20 stereo bands and one region in each frame and PS coding with proposed time-frequency algorithm at bit-rate of 24kbps, 36kbps, and 48kbps. All the experiments are done on NCTU-HE-AAC[25]. The results are shown in Figure 7 ~ Figure 9. In each figure, “Fix” means PS coding with always 20 stereo bands and one region in each frame. “Time-Frequency Algorithm” means PS coding with the time-frequency algorithm.

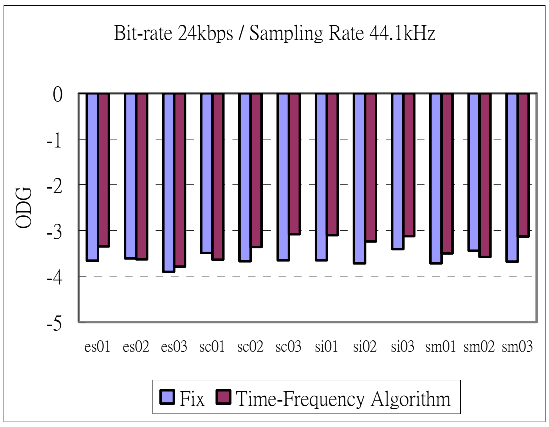

Figure 7 Objective measurements through the ODGs for two kinds of different time-frequency strategies in PS coding. The test is under 24kbps and the sampling rate at 44.1kHz.

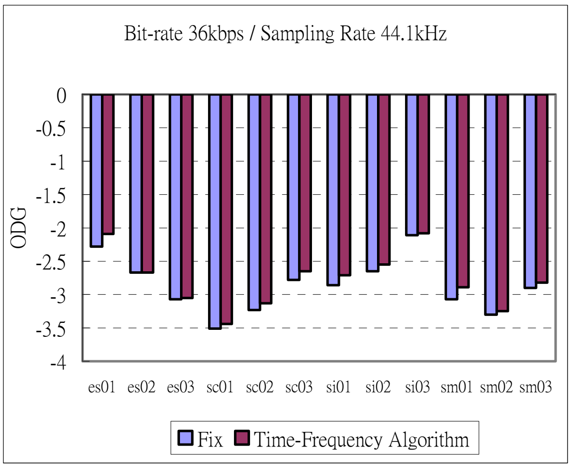

Figure 8 Objective measurements through the ODGs for two kinds of different time-frequency strategies in PS coding. The test is under 36kbps and the sampling rate at 44.1kHz.

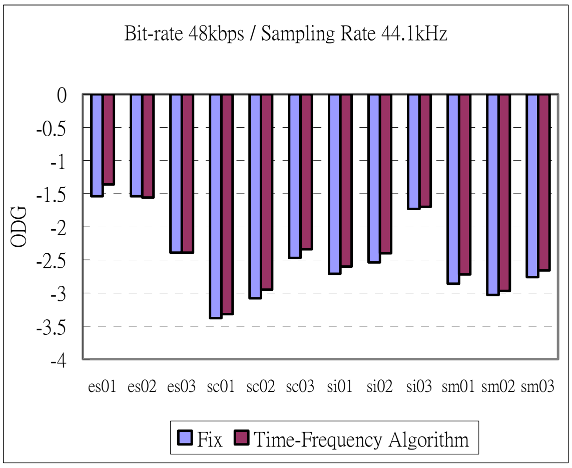

Figure 9 Objective measurements through the ODGs for two kinds of different time-frequency strategies in PS coding. The test is under 48kbps and the sampling rate at 44.1kHz.

From the results above, the track with the worst ODG at bit rate 24kbps is es03. The grade is −3.79 and −3.9 in Time-Frequency Algorithm and Fix method, respectively. The track with the worst grade at bit rate 36kbps is sc01, and the grades are −3.44 and −3.51 in Time-Frequency Algorithm and Fix method, respectively. The track with the worst grade at bit rate 48kbps is sc01, and the gradesare.32 and −3.38 in Time-Frequency Algorithm and Fix method, respectively. The best ODGs in Time-Frequency Algorithm and Fix method are −3.08(sc03) and −3.41(si03) at bit rate 24kbps, −2.08(si03) and −2.11(si03) at bit rate 36kbps, and −1.36(es01) and −1.54(es01 and es02) at bit ate 48kbps. The average ODGs of the twelve test tracks in Time-Frequency Algorithm and Fix method are −3.377 and −3.633 at bit rate 24kbps, −2.778 and −2.869 at bit rate 36kbps, and −2.414 and −2.503 at bit rate 48kbps. The result shows the Time-Frequency Algorithm has the better average ODG than the Fix method.

To verify HE-AAC with the PS coding has better quality than HE-AAC at low bit rate, we introduce the follow experiment. The experiment is on the twelve test tracks listed in Table 2. The tracks are encoded by NCTU-HE-AAC and NCTU-HE-AAC with PS coding tool and our Time/Frequency Algorithm at bit rates 24kbps, 36kbps, and 48kbps. The result illustrates the HE-AAC with PS coding tool improves the quality at low bit-rate in Figure 10. In Figure 10, the top arrow represents the worst ODG, the down diamond represents the best ODG, and the middle square represents average ODG among the twelve test tracks.

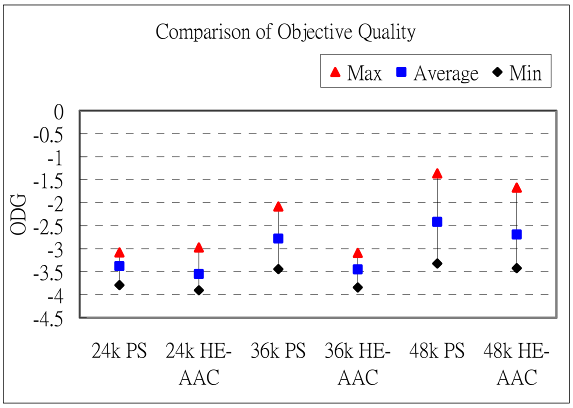

Figure 10 The ODG range comparison of NCTU-HE-AAC with and without PS coding tool at bit rates 24kbps, 36kbps, and 48kbps.

The result shows that the average ODG grades are improved at bit rates 24kbps, 36kbps, and 48kbps. The average ODG grades are −3.377 in 24kbps PS, −3.551 in 24kbps HE-AAC, −2.778 in 36kbps PS, −3.447 in 36kbps HE-AAC, −2.414 in 48kbps PS, and −2.689 in 48kbps HE-AAC. It is shown that the PS coding has the best effect at 36kbps bit-rate.

conclusion

This paper has introduced the design of Time/Frequency stereo parameter sets. The DP algorithm in RD assists the PS coding tool to deliver the necessary stereo parameter sets. The experiments have shown that the Time-Frequency algorithm improves the coding efficiency. The latest released executive binary and the associated intensive subjective tests are conducted and illustrated at website of PSPLAB [12].

REFERENCES

[15] "Coding of Moving Pictures and Audio," Draft ISO/IEC 14496-3 (Audio 3rd Edition)

[16] ISO/IEC, "Text of ISO/IEC 14496-3:2001/FPDAM 1, Bandwidth extensions," ISO/IEC JTC1/SC29/WG11/N5203, October 2002, Shanghai, China.

[17] M. Dietz, L. Liljeryd, K. Kjörling, O. Kunz, "Spectral Band Replication, a novel approach in audio coding," 112nd AES Convention, Munich, Germany, May 2002, Preprint 5553.

[18] M. Wolters, K. Kjörling, D. Homm, H. Purnhagen, "A closer look into MPEG-4 High Efficiency AAC," 115th AES Convention, New York, USA, October 2003, Preprint 5871.

[19] H.W. Hsu, C.M. Liu, and W.C. Lee, "Audio Patch Method in MPEG-4 HE AAC Decoder," 117th AES Convention, San Francisco, USA, October 2004, Preprint 6221.

[20] 3GPP TS 26.410 V6.0.0 (2004-09), http://www.3gpp.org

[21] H.F. Silverman and D.P. Morgan, "The application of dynamic programming to connected speech recognition," IEEE Acoustics, Speech, and Signal Processing Magazine, vol. 7, pp. 6-25, July, 1990.

[22] R. Bellman, "On the theory of dynamic programming," Proceedings of the National Academy of Sciences, vol. 38, pp. 716-719.

[23] C.M. Liu, W. C. Lee, C. H. Yang, K. Y. Pang, T. Chiou, T. W. Chang, Y. H. Hsiao, H. W Hsu, C. T. Chien, "Design of MPEG-4 AAC Encoder," 117th AES Convention, San Francisco, USA, October 2004, Preprint 6201.

[24] ITU Radiocommunication Study Group 6, "Draft Revision to Recommendation ITU-R BS.1387-

Method for objective measurements of perceived audio quality”.

[25] C.M. Liu, L.W. Chen, H.W. Hsu, and W.C. Lee, “Bit Reservoir Design for HE AAC,” 118th AES Convention, Barcelona, Spain, May 2005, Preprint 6382.

[26] http://psplab.csie.nctu.edu.tw/projects/index.pl/nctu-aac.html.

Design for High Frequency Adjustment Module in MPEG-4 HE-AAC Encoder based on Linear Prediction Method

Han-Wen Hsu, Yung-Cheng Yang, Chi-Min Liu, Wen-Chieh Lee,

Abstract (60~120)

High frequency adjustment module is the kernel module of spectral band replication (SBR) in MPEG-4 HE-AAC. The objective of high frequency adjustment is to recover the tonality of reconstructed high frequency. There are two crucial issues, the accuracy measurement of tonality and the decision of shared control parameters. Control parameters, which are extracted according to accurate signal tonalities, will be used to determine gain control and energy level of additional components in decoder part. In other word, the quality of the reconstructed signal will dominantly depend on the high frequency adjustment module.

In this paper, an efficient method based on Levinson-Durbin algorithm is proposed to measure tonality by linear prediction approach with adaptive orders to fit different subband contents. Furthermore, the artifact due to the sharing of control parameter is also concerned, and an optimal decision criterion of control parameter is proposed. Both the objective and the subjective tests are conducted to check the quality improvement.

Precis (500~750)

Spectral band replication (SBR) is a new audio coding enhancement tool which combined with conventional AAC encoder to improve audio quality at low bit rates. From the similarity between low frequency band and high frequency band, the main principle of SBR is to reconstruct high frequency band signal by replicating from low frequency band. More specific, besides applying proper energy shaping in both time domain and frequency domain, the mechanism of high frequency adjustment can maintain the accurate perceptual quality by the reconstruction for harmonics as well as for noise-like components. This paper is focused on the two design issues of high frequency adjustment module, including tonality measurement and control parameters extraction.

A. Tonality measurement

A simple linear predictor of second order is suggested to measure the tonality, that is defined as TNR (tone to noise ratio), of subband signals by the standard. However, in order to precisely measure the actual TNR, the poles of the linear prediction filter must match tones contained in the subband, and thus the prediction order should equal to the number of tones. It is obvious that the second order predictor would not be sufficient to capture all predictable components, if a subband contains more than two tones. On the other hand, for a subband containing less than two tones, some noise components will be captured excessively as predictable components by the second order predictor.

Some methods have been proposed to improve the accurateness of measurement without the necessary of known tone number. However, most of them need additional time-consuming calculation. For example, a direct estimation on frequency domain [*] is proposed to measure noise floor precisely, and an accurate method on time domain in [1] is given to capture all tone components. They are inefficient due to additional FFT and sinusoid analysis respectively. In this paper, an efficient approach based on Levinson-Durbin algorithm that can construct a lattice filter gradually, is proposed to decide accurate prediction order adaptively.

B. Control parameter extraction

After tonality measurement, the control parameter, which the gain control and the energy of additive components in SBR range are determined according to, should be extracted. From previous literature [1], the impact of control parameter extraction has been explored, and a decision criterion of optimal control parameter is proposed. However, the criterion is only considered for one individual subband, and the problem that the control parameter should be shared by multiple subbands is ignored. The artifact due to the sharing of control parameter will cause the erroneous reconstruction of TNR in some subbands and result in perceptual inconsistency to the original signal. In this paper, to minimize the artifact, a grouping method is proposed to choose a shared optimal control parameter for all subbands in a cluster.

C. Experiments

To confirm the accuracy of the proposed method for tonality measurement, both the constructed predictor with adaptive order and a second order predictor are applied to some artificial and natural tonal signals. The result shows that, under a practicable time consuming with the incensement of prediction order, the accuracy is improved largely. On the other hand, by taking the suggested method in the standard and the method proposed in [1] as reference modes respectively, the proposed grouping method for control parameters extraction has apparent improvement of quality as well as the accuracy of reconstructed spectrum. In the past few years, we have considered the design of AAC and HE-AAC encoders in AES Conventions 116-119. The resultant AAC encoder is referred to as the NCTU HE-AAC [2]. The grouping method is integrated in the NCTU HE-AAC and conducted by. both objective and subjective tests to check the quality improvement. The objective test measures used is the recommendation system by ITU-R Task Group 10/4.

References

[1] San-Uk Ryu, Keneth Roth "Enhanced Accuracy of the Tonality Measure and Control Parameter Extraction Modules in MPEG-4 HE-AAC," AES 119th Convention, New York, NY, USA, October 7-10, 2005.

[2] NCTU-AAC website http://psplab.csie.nctu.edu.tw/projects/nctu-aac.html.

Efficient Time-Frequency Grid Decision in HE-AAC Encoder through Dynamic Programming

J.Y, Tan, Han-Wen Hsu, Wen-Chieh Lee, and Chi-Min Liu

PSPLab, Computer Science and Information Engineering, National Chiao-Tung University, Hsin-Chu, 33050, Taiwan
cmliu@csie.nctu.edu.tw

Abstract

Spectral Band Replication (SBR) has been combined with MPEG AAC as bandwidth extension tool. The enhanced AAC by SBR is called High Efficient (HE) AAC. Time-frequency (T/F) grids deciding the replication unit in high frequency bands are the kernel module in SBR. This paper formulates the decision of the TF grid into a trellis-lattice search problem and proposes an efficient search algorithm to find the optimum path. Both subjective and objective tests are conducted to check the quality improvement over existing methods. The objective test measures used is the recommendation system by ITU-R Task Group 10/4.

Precis

Spectral Band Replication (SBR) has been proposed to compress high quality audio at low bit rates. With SBR module taking care of the high frequency contents, the conventional AAC encoder can compress the low frequency part using most of the available bits. The resulting scheme is referred to as the MPEG-4 High Efficient (HE) AAC or AACplus. From the similarity between the low and high bands, the basic principle of SBR is to reconstruct the high bands by replicating the low bands. In the past few years, we have considered the design of AAC and HE-AAC encoders in AES Conventions 116-119. The resultant AAC encoder is referred to as the NCTU HE-AAC [1]. This paper considers the design of T/F grid and integrates the method into the NCTU HE-AAC to show the improvement.

The HE-AAC encoders split the audio signals into subbands grouped as the low frequency part and the high frequency part. The low frequency part of the subband signals is encoded by AAC encoder while the high frequency part of the subbands are encoded through the SBR encoder. For the SBR range, the subband signals are segmented into the time/frequency (T/F) grids. The SBR coding is a replicating process that replicates the low frequency signals to high frequency signals. The signal in a grid is the basic reconstructed unit in the subsequent SBR coding process. The locations of time borders and the resolution of the T/F grids determine the accuracy of the replication and hence the audio quality.

A. Frequency Table Decision

The resolution of chosen frequency table determines the quality and the consumed bits of reconstructed audio. Furthermore, there are considerable issues as the frequency table changes. In the HE-AAC standard, frequency table affects low frequency part that will be duplicated for reconstructed high bands. Also, changing frequency table between frames may cause the frequency discontinuity of reconstructed high frequency spectrum. On the other hand, altering frequency tables increase the overhead of header bitstream. According to the above factors, this paper provides a method for selecting the appropriate frequency table taking account of artifacts, reconstructed audio quality and consumed bits.

B. Time Segments and Associated Frequency Resolution Decision

Based on the chosen frequency table, the low frequency band used to reconstruct high band can be be found. Accordingly, the reconstruction error can be evaluated. By comparing the reconstructed errors for all possible combinations that involve the location of time borders, the number of time envelopes, and the resolution of each envelope, we can find the optimal solution for distribution of time borders and associated resolution. However, it is very time-consuming to search the optimal solution from all the possible combinations. This paper shows that the problem can be formulated as a trellis-lattice search algorithm and a dynamic programming method is proposed to have the efficient search.. In the dynamic programming, we design and evaluate the cost functions to evaluate the possible combinations. Based on the cost functions, the dynamic programming break down the search problem into stages of decisions that follow a recurrence relation.

C. Frame Class Decision

There are four frame classes, FIXFIX, FIXVAR, VARFIX, VARVAR in the HE-AAC draft. The most suitable frame class should be decided according to the locations of the leading, the trailing SBR frame borders, the distribution of time borders, and the number of envelopes, This paper considers the decision of classes through a frame look ahead method

D. Experiments

Both subjective and objective tests are conducted on intensive audio tracks to check the quality improvement. The objective test measures used is the recommendation system by ITU-R Task Group 10/4. The grid method is incorporated into the NCTU HE-AAC [1] and exhibits a quality better than the reference HE-AAC codecs [2]-[4].

References

[1] NCTU-AAC website http://psplab.csie.nctu.edu.tw/projects/nctu-aac.html.

[2] 3GPP TS 26.410 V6.0.0 (2004-09), http://www.3gpp.org.

[3] Nero……

[4] Coding Technology