

RESEARCH

Open Access

New aliasing cancelation algorithm for the transition between non-aliased and TDAC-based coding modes

Jeongook Song and Hong-Goo Kang*

Abstract

This paper proposes a new aliasing cancelation algorithm for the transition between non-aliased coding and transform coding with time domain aliasing cancelation (TDAC). It is effectively utilized for unified speech and audio coding (USAC) that was recently standardized by the Moving Picture Experts Group (MPEG). Since the USAC combines two coding methods with totally different structures, a special processing called forward aliasing cancelation (FAC) is needed at the transition region. Unlike the FAC algorithm embedded in the current standard, the proposed algorithm does not require additional bits to encode aliasing cancelation terms because it appropriately utilizes adjacent decoded samples. Consequently, around 5% of total bits are saved at 16- and 24-kbps operating modes in speech-like signals. The proposed algorithm is sophisticatedly integrated on the decoding module of the USAC common encoder (JAME) for performance verification, which follows the standard process exactly. Both objective and subjective experimental results confirm the feasibility of the proposed algorithm, especially for contents that require a high percentage of mode switching.

1 Introduction

Unified speech and audio coding (USAC; ISO/IEC 23003-3) standardized in early 2012 shows the best performance for speech, music, and mixed type of input signals [1]. Verification tests confirmed the superiority of quality, especially at low bit rates [2]. In an initial stage of designing the coding structure, it was not possible to acquire high-quality output to all input contents because only a single type of traditional audio or speech coding structure was adopted. The best result could be obtained by simultaneously running two types of codecs: Adaptive Multi-rate Wideband plus (AMR-WB+ [3]) for speech signals and high-efficiency advanced audio coding (HE-AAC [4]) for audio signals. In case of encoding signals with mixed characteristics, one of two coding modes is chosen depending on the characteristic of input contents. Although this approach improves the quality of all types of contents, many problems occur at transition frames where mode switching is needed between entirely

different types of codecs. For example, the segment of perceptually weighted signal encoded by speech codec needs to be smoothly combined with that of the signal encoded by audio codec. Since the characteristic of speech and audio codec is different, however, the overlapped segment between two codecs must not be similar to the input signal. How to determine the encoding mode for the various types of input signal is also important. The problems are mostly solved by adopting novel technologies such as a signal classifier, frequency domain noise shaping (FDNS), and forward aliasing cancelation (FAC) technique [5].

The FAC algorithm is one of the key technologies in USAC, which enables the successful combination of two different types of codecs, especially at transition frames. To remove the aliasing terms caused by cascading different types of codecs, FAC additionally generates the aliasing cancellation signals, and then they are quantized and transmitted to the decoder. In the earlier version of USAC that had not introduced the FAC technique, the frame boundary of transition frame was variable; thus, a special windowing operation was needed for compensating the aliased signal in the overlap region. Although FAC

*Correspondence: hgkang@yonsei.ac.kr
School of Electrical and Electronic Engineering, Yonsei University,
Seoul 120-749, South Korea

somewhat solves the problem, it still requires additional bits.

This paper proposes a new aliasing cancellation algorithm that does not need any additional bits, which uses the decoded signal of the adjacent frames. At first, the algorithm generates the relevant aliasing cancellation part by considering the error that occurred by the encoding mode switching. Then, the output signals are reconstructed by adding the generated aliasing cancellation part to the decoded signal and by normalizing the weight caused by the encoding window. In the overall process, the most important thing is how to obtain the aliasing cancellation part by properly utilizing the adjacent signal.

The aliasing cancellation process of the proposed algorithm is conceptually similar to that of the block switching compensation scheme proposed for low delay advanced audio coding (AAC-LD [6,7]). In the literature, the scheme introduced time domain weightings applicable as a post processing in the decoder in order to remove a look-ahead delay inevitable for a window transition from the long window to the short window. This is similarly considered as an aliasing cancellation signal described in this paper. However, its application and the resulting aliasing form are different.

A new aliasing cancellation algorithm is sophisticatedly integrated in the decoding module of the USAC common encoder (JAME) [8], which has been designed by

our team as an open source paradigm. Objective and subjective test results show that the proposed method has comparable quality to the FAC algorithm while saving the bits for encoding the aliasing signal component in the FAC algorithm.

Section 2 describes the overview of USAC techniques and FAC algorithm. In Section 3, the proposed algorithm is explained in detail. In Section 4, experiments and evaluation results are also described.

2 USAC overview and FAC algorithm

2.1 Overview

USAC, recently standardized codec by MPEG, provides high quality for speech, audio and mixed signals even in very low bit rates [2]. Figure 1 shows a block diagram of the encoding process that consists of frequency domain (FD) and time domain (TD) coding modules. At first, the encoding mode is determined by analyzing the spectral information of input signal in the signal classifier block [9]. The FD coder transforms the time domain input signal into frequency spectrum by taking the modified discrete cosine transform (MDCT) [10], then calculates the perceptual entropy of each frequency band using a psychoacoustic model [11,12]. The number of allocated bits to each band is determined by considering the distribution of perceptual entropy. In the TD coding module, an input signal is encoded by either algebraic code-excited linear

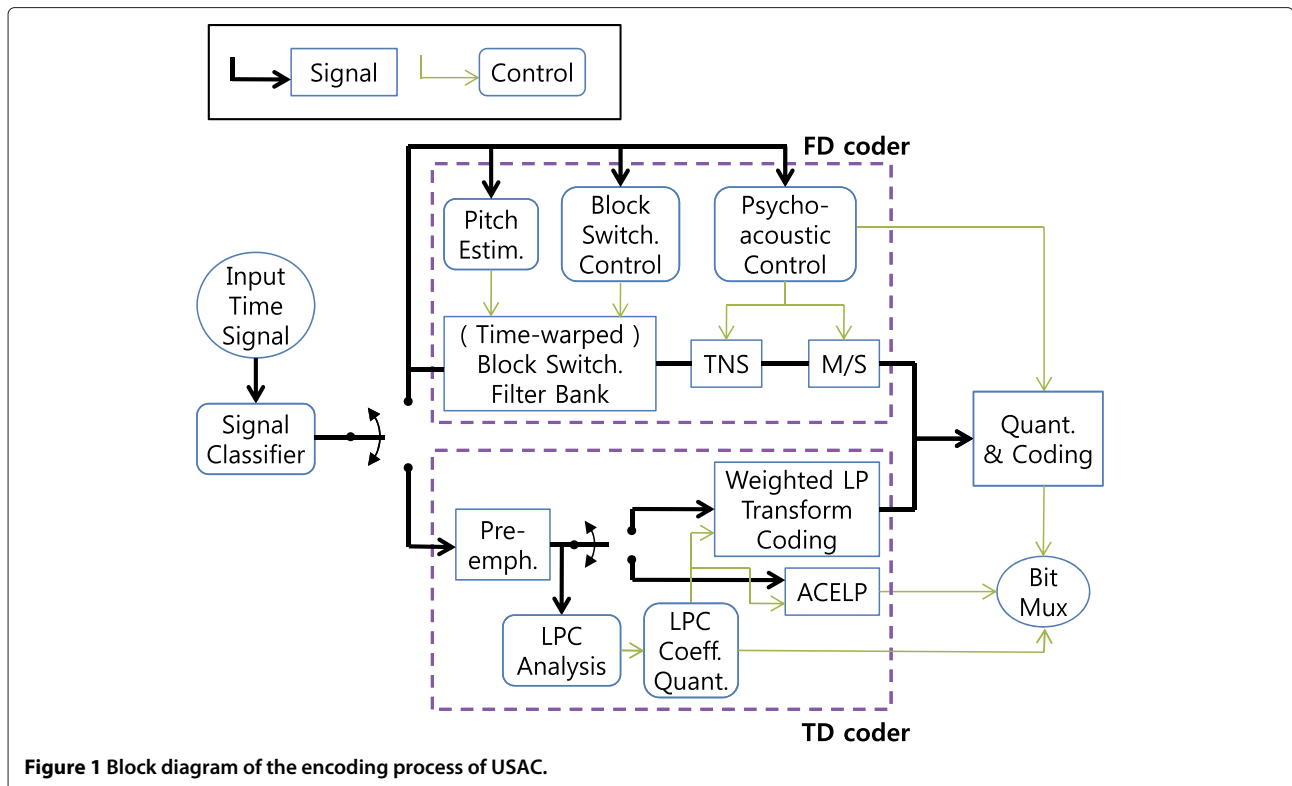


Figure 1 Block diagram of the encoding process of USAC.

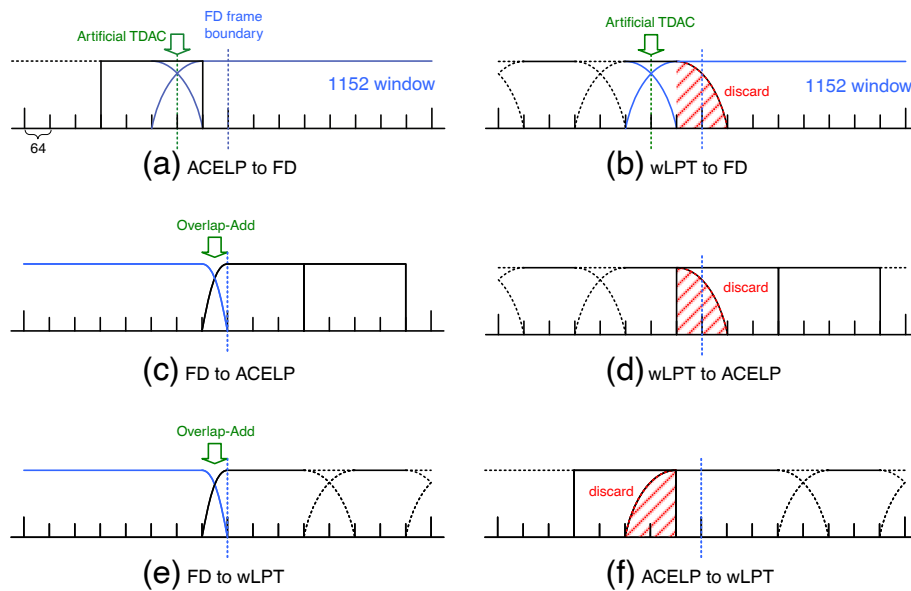


Figure 2 Synthesis process in the transition frame of an initial version of USAC. (a) ACELP to FD, (b) wLPT to FD, (c) FD to ACELP, (d) wLPT to ACELP, (e) FD to wLPT, and (f) ACELP to wLPT.

prediction (ACELP) or weighted linear prediction transform coding (wLPT) similar to the AMR-WB+ codec. The wLPT is a modified version of transform coded excitation (TCX) mode that the residuals of LPC filter are encoded in the frequency band using the MDCT method [13]. Note that its quantizer is the same as the one used for the FD coder to keep compatibility and efficiency. Finally, the quantized spectrum is encoded by context adaptive

arithmetic coding (CAAC), which has a higher coding efficiency than the Huffman coding [14].

2.2 Forward aliasing cancelation algorithm

Since the USAC consists of two different types of coding methods, it is very important to handle the transition frame where the encoding mode is switched from FD codec to TD codec or vice versa. Note that the MDCT

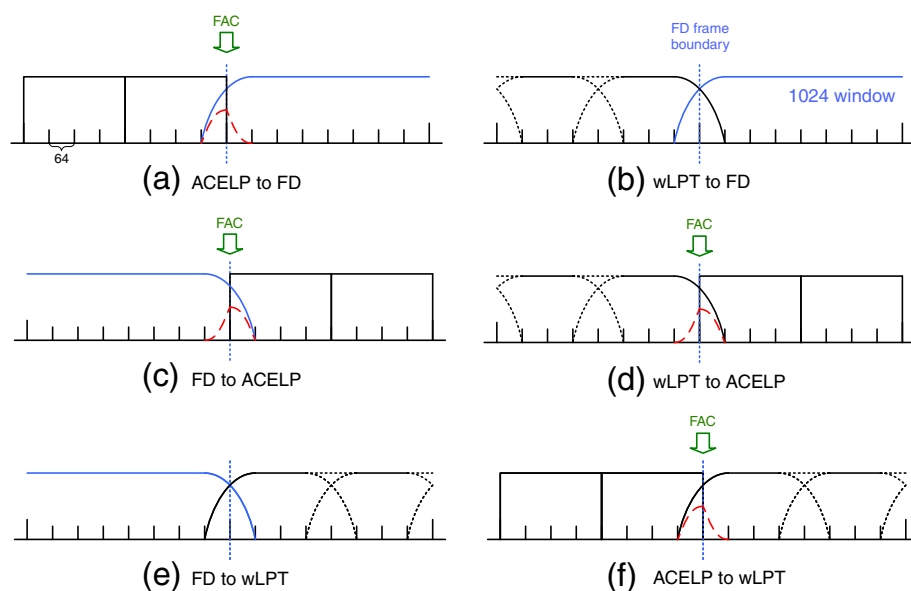


Figure 3 Synthesis process in the transition frame using FAC algorithm. (a) ACELP to FD, (b) wLPT to FD, (c) FD to ACELP, (d) wLPT to ACELP, (e) FD to wLPT, and (f) ACELP to wLPT.

removes the aliasing part of the current frame by combining the signal decoded at the following frame. However, if the encoding mode of the next frame is TD codec, the aliasing term must not be generally canceled. In an initial version of USAC, this problem was solved by discarding the aliased signal and using inconsistent frame length. When the frame length of TD codec is decreased due to aliased signal, the following frame length is increased for synchronizing the starting position of FD codec [15].

Figure 2 describes the synthesis process in the transition frame of an initial version of USAC. The synthesized signals in the overlapped region between wLPT and other coding methods are discarded as given in Figure 2b,d,f. In case the encoding mode is changed from FD codec to TD codec, the signals decoded by ACELP are windowed

to perform an overlap-add processing with the FD output. Since the frame encoded by TD codec starts at the front position of the frame boundary, the starting point of the long frame of FD mode needs to be compensated by decreasing the length, which allows the early start of the TD codec mode. Since the frame size is inconsistent, therefore, a new type of window should be designed [15].

The forward aliasing cancellation algorithm is proposed to solve the awkward frame structure mentioned above. Figure 3 shows the FAC algorithm [5]. All transitions are made at the same position in each frame boundary. Note that the FAC is needed for the ACELP transition frames given in Figure 3a,c,d,f. Since the decoded output of ACELP mode cannot cancel out the aliased outputs

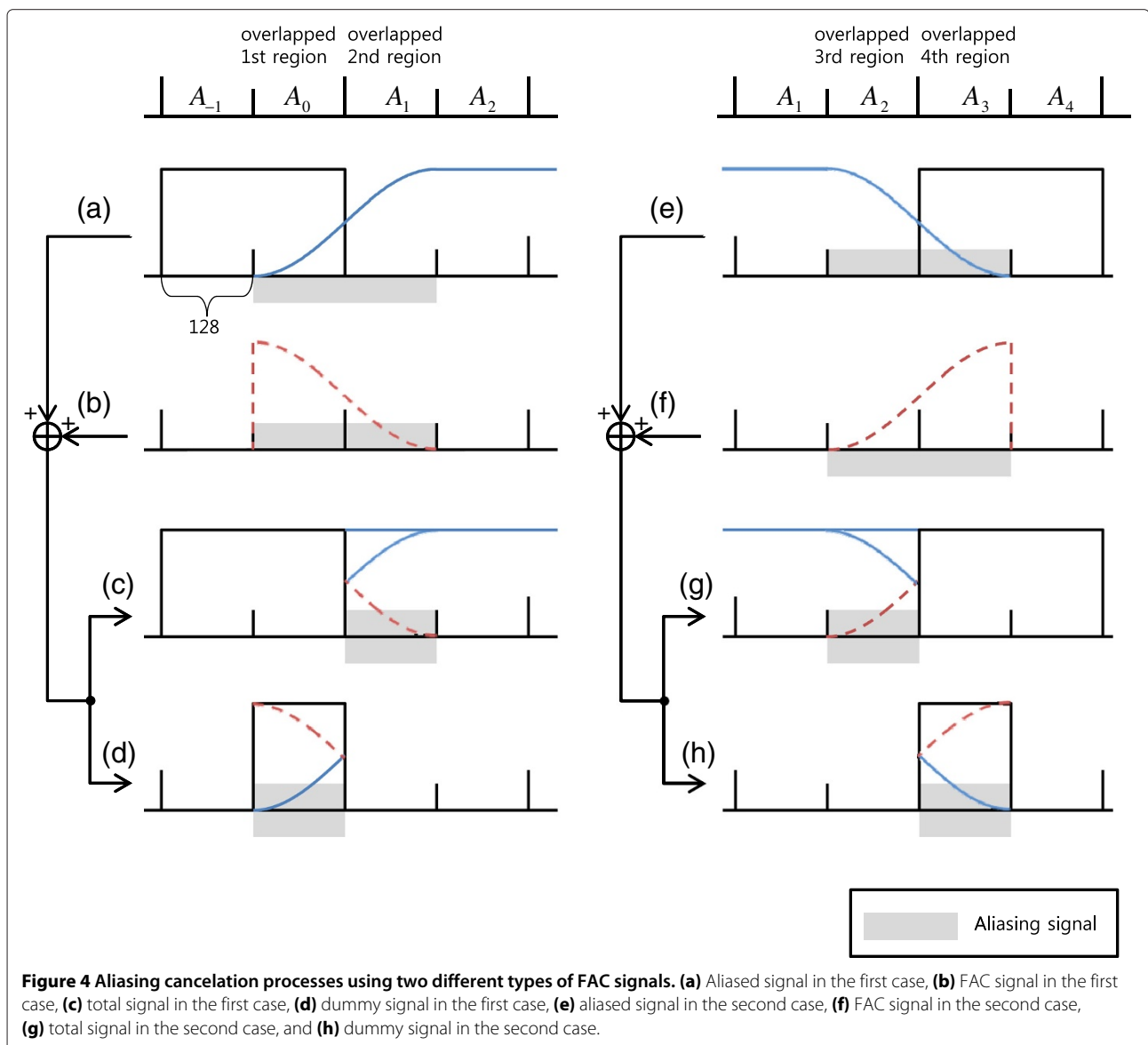


Figure 4 Aliasing cancellation processes using two different types of FAC signals. (a) Aliased signal in the first case, (b) FAC signal in the first case, (c) total signal in the first case, (d) dummy signal in the first case, (e) aliased signal in the second case, (f) FAC signal in the second case, (g) total signal in the second case, and (h) dummy signal in the second case.

decoded by FD or wLPT codec modes, the FAC algorithm artificially generates the additional signals for canceling the aliasing component. The generated signals are mixed with the quantization error portion of wLPT or FD coder, and then they are quantized by the adaptive vector quantization (AVQ) tool [9]. The AVQ tool consists of three parts: FAC gain, two codebook indices, and 16 Voronoi extension indices for AVQ refinement. Seven bits are allocated to the FAC gain, and the bits for other indices are variable due to adopting a unary coding. For example at 24 kpbs, around 130 bits per one frame is used to encode the FAC parameters. It corresponds to the 11% of the average frame bits.

3 Proposed aliasing cancelation algorithm

As is shown in Figure 4, the FAC algorithm happens to be applied in two different cases depending on the order of coding modules, i.e., whether the transition is made from ACELP to other coding modes (wLPT or FD) or vice versa. The first case given in Figure 4a,b,c,d describes the way of removing aliasing signals from ACELP to other coding modes. The second case given in Figure 4e,f,g,h does the reverse direction. The aliased signals given in Figure 4a,e are compensated by adding the FAC signal. The FAC

signal given in Figure 4b,f consists of an aliasing cancelation component and a symmetric windowed signal. Note that the aliasing cancelation term in the FAC signal plays a key role in designing the proposed algorithm later. The dummy signal is simply generated by adding the FAC signals and aliased signals in the decoding stage. Since the aliasing signal depicted in Figure 4d is canceled out, the sum of the remained signals becomes the output signal marked with the black rectangular shape. From now on, it is called 'dummy signal'. Assuming that there is no quantization error, dummy signals are equivalent to ACELP signals in the same position in the time domain. Similarly, dummy signals in Figure 4h are also equivalent to the first 128 samples of ACELP signals. Since those ACELP signals are available in the decoder, the dummy signals do not need to be sent as it does in the FAC algorithm. In other words, the region located on dummy signals can be directly decoded by the synthesized signal obtained from the ACELP scheme, i.e., it is regarded as a non-aliased part. Please also note that the method requires additional bits to quantize FAC signals. This paper proposes a new aliasing cancelation algorithm that does not need any additional bits while successfully removing the aliasing parts.

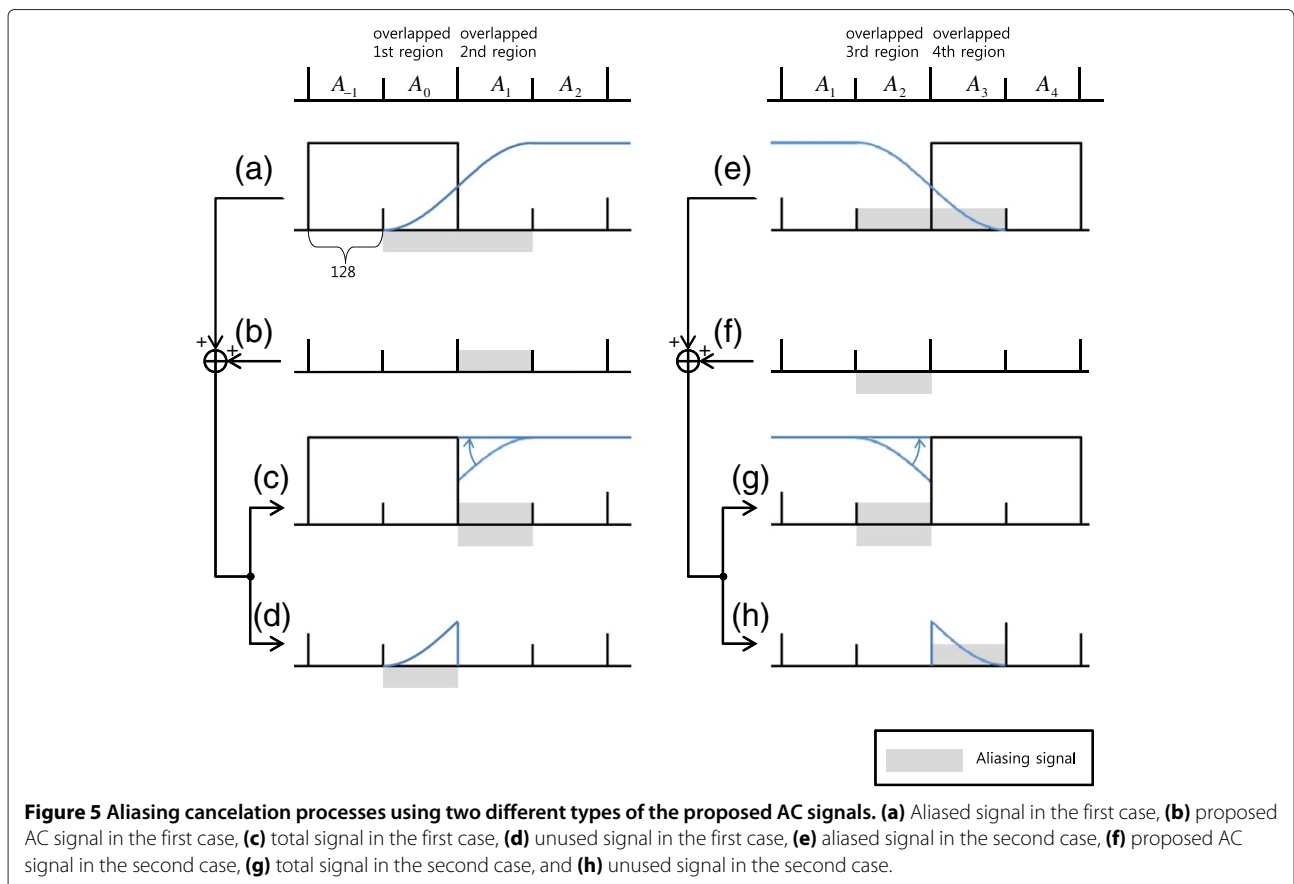


Figure 5 Aliasing cancelation processes using two different types of the proposed AC signals. (a) Aliased signal in the first case, (b) proposed AC signal in the first case, (c) total signal in the first case, (d) unused signal in the first case, (e) aliased signal in the second case, (f) proposed AC signal in the second case, (g) total signal in the second case, and (h) unused signal in the second case.

Table 1 Test items for the evaluation of the proposed algorithm

Item number	Class	Label	Item
1	Music	m1	salvation
2	Music	m2	te15
3	Music	m3	Music_1
4	Music	m4	Music_3
5	Music	m5	Phi7
6	Speech	s6	Es01
7	Speech	s7	louis_raquin_15
8	Speech	s8	Wedding_speech
9	Speech	s9	te1_mg54_speech
10	Speech	s10	Arirang_speech
11	Mixed	×11	twinkle_ff51
12	Mixed	×12	SpeechOverMusic_1
13	Mixed	×13	SpeechOverMusic_4
14	Mixed	×14	HarryPotter
15	Mixed	×15	Lion

Figure 5 shows the schematic diagram of the proposed algorithm. As is shown in Figure 5b,f, the proposed algorithm generates signals for canceling the aliasing components. After aliasing cancelation (AC) signal is added to the aliased output of the decoder, the combined signal becomes unaliased as given in Figure 5c,g. The signals

given in Figure 5d,h are simply disregarded because the region can be reconstructed by the ACELP output only as already described in Figure 4.

Hereinafter, we further derive the relationship of the FAC signals (Figure 4b,f) and the aliasing cancelation signals (Figure 5b,f) by utilizing the specific relation between the formulae of MDCT and DCT-IV. Note that the MDCT is a modified form of DCT-IV that is suitable for saving the bits. MDCT spectral coefficient, $X_M(k)$, and DCT-IV spectral coefficient, $X_D(k)$, are respectively defined as follows [10]:

$$X_M(k) = \sum_{n=0}^{2N-1} x(n) \Phi_k \left(n + \frac{N}{2} \right), \quad (1)$$

$$X_D(k) = \sum_{n=0}^{N-1} x(n) \Phi_k(n), \quad (2)$$

$$\Phi_k(n) = \cos \left(\frac{\pi}{N} \left(n + \frac{1}{2} \right) \left(k + \frac{1}{2} \right) \right),$$

where N is the frame length.

$$\begin{aligned} \Phi_k(2N + n) &= \cos((2k + 1)\pi + \alpha) = -\Phi_k(n), \\ \Phi_k(2N - 1 - n) &= \cos((2k + 1)\pi - \alpha) = -\Phi_k(n), \\ \alpha &= \frac{\pi}{2N} \left(n + \frac{1}{2} \right) (2k + 1). \end{aligned} \quad (3)$$

Table 2 Actual achieved bit rates by each item in the operating mode

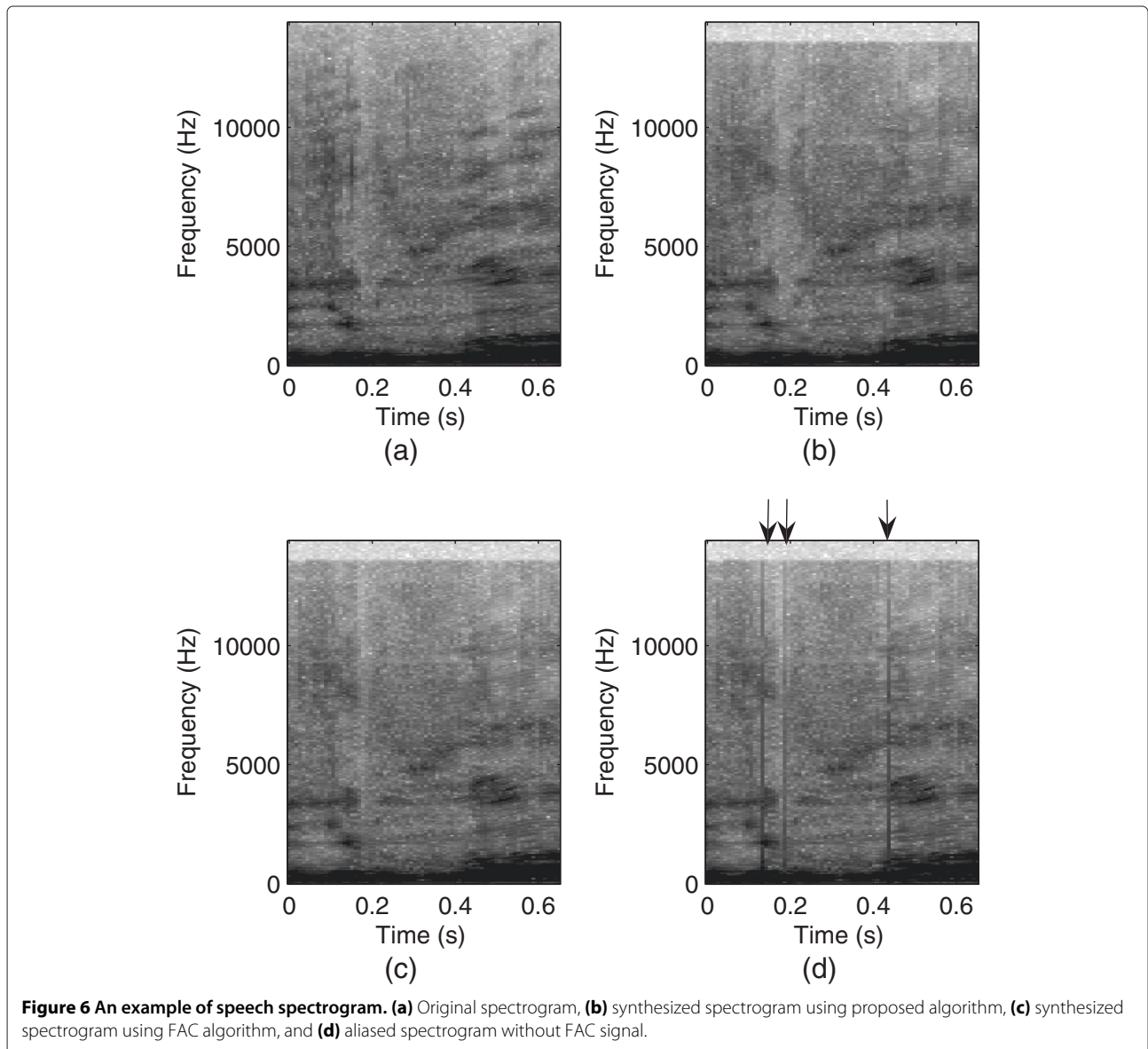
Item	12-kbps mode		16-kbps mode		24-kbps mode	
	Conv. (kbps)	Prop.-B (kbps)	Conv. (kbps)	Prop.-B (kbps)	Conv. (kbps)	Prop.-B (kbps)
m1	12.09	12.09	16.18	16.18	24.71	24.71
m2	12.15	12.13	16.22	16.19	24.75	24.73
m3	12.20	11.94	16.19	15.96	24.69	24.69
m4	12.14	12.11	16.24	16.20	24.76	24.58
m5	11.85	11.83	15.94	15.91	24.52	24.50
s6	11.83	11.64	15.74	14.58	24.29	22.96
s7	12.20	11.99	16.19	15.24	24.70	23.34
s8	11.95	11.90	15.50	14.76	24.08	22.53
s9	11.60	11.45	15.37	14.65	23.93	22.52
s10	11.76	11.61	15.41	14.76	23.94	22.71
×11	12.12	12.11	16.20	16.18	24.73	24.73
×12	11.81	11.76	15.94	15.91	24.49	24.41
×13	12.20	11.58	16.27	15.30	24.77	23.67
×14	12.21	12.08	16.19	15.65	24.71	23.47
×15	11.89	11.68	15.76	14.85	24.30	23.06
Music	12.08	12.02	16.15	16.09	24.69	24.64
Speech	11.87	11.72	15.64	14.80	24.19	22.81
Mixed	12.04	11.84	16.07	15.58	24.60	23.87
Total	12.00	11.86	15.96	15.49	24.49	23.77

By utilizing the cosine property given in Equation 3, the MDCT spectral coefficient can be represented by the form of DCT-IV:

$$\begin{aligned}
 X_M(k) &= \sum_{n=0}^{N/2-1} (-x(\frac{3}{2}N - 1 - n) - x(\frac{3}{2}N + n))\Phi_k(n) \\
 &+ \sum_{n=N/2}^{N-1} (x(n - \frac{N}{2}) - x(\frac{3}{2}N - 1 - n))\Phi_k(n) \\
 &= \sum_{n=0}^{N-1} c(n)\Phi_k(n).
 \end{aligned}
 \tag{4}$$

Equation 4 informs us that the MDCT spectral values transformed by $2N$ consecutive inputs are exactly equivalent to DCT-IV spectral values transformed by N inputs, which are folded at the $\frac{1}{2}N$ position and $\frac{3}{2}N$ position. Since the DCT-IV should be invertible, we know that the folded signals are the aliased parts generated by taking an inverse MDCT [16]. Two parts of the folded signals are

$$\begin{cases}
 -x\left(\frac{3}{2}N - 1 - n\right) - x\left(\frac{3}{2}N + n\right) & , 0 \leq n < \frac{N}{2} \\
 x\left(n - \frac{N}{2}\right) - x\left(\frac{3}{2}N - 1 - n\right) & , \frac{N}{2} \leq n < N.
 \end{cases}
 \tag{5}$$



Let $x(n)$ be input samples and A_m be the $(1 \times \frac{N}{2})$ vector of input samples:

$$A_m = [x(m\frac{N}{2}) \ x(m\frac{N}{2} + 1) \ \cdots \ x((m+1)\frac{N}{2} - 1)]. \quad (6)$$

Equation 5 is reformulated as

$$\begin{cases} S(A_0, A_1) = A_0 - A_1 R, \\ U(A_2, A_3) = -A_2 R - A_3, \end{cases} \quad (7)$$

where R depicts an $(\frac{N}{2} \times \frac{N}{2})$ reverse identity matrix:

$$R = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ & & 1 & 0 \\ & \ddots & & \vdots \\ 1 & & & 0 \end{bmatrix}. \quad (8)$$

Practically, windowing is introduced to remove the side-lobe artifacts. By introducing the windowing to Equation 7, the first and second folded signals can be expressed as

$$\begin{cases} S(A_0 \circ W_0, A_1 \circ W_1) = A_0 \circ W_0 - A_1 \circ W_1 R, \\ U(A_2 \circ W_2, A_3 \circ W_3) = -A_2 \circ W_2 R - A_3 \circ W_3, \end{cases} \quad (9)$$

where the operator ' \circ ' denotes the Hadamard product [17] and W_k depicts a $(1 \times \frac{N}{2})$ window matrix:

$$W_k = \left[w\left(\frac{\pi}{4}k\right) \ w\left(\frac{\pi}{4}k + \frac{\pi}{2N}\right) \ \cdots \ w\left(\frac{\pi}{4}k + \frac{\pi}{2N}\left(\frac{N}{2} - 1\right)\right) \right]. \quad (10)$$

The window matrix, W_k , must be symmetric and satisfy the Princen-Bradley condition for perfect reconstruction [10]:

$$\begin{aligned} W_0 R &= W_3, \\ W_1 R &= W_2, \end{aligned} \quad (11)$$

$$W_k \circ W_k + W_{k+2} \circ W_{k+2} = [1 \ 1 \ \cdots \ 1].$$

Note that the aliased signals in the overlap region of Figures 4 and 5 are equivalent to the folded signals through the analysis of MDCT transform as is given in Equation 9. Therefore, the time signals in the overlapped regions can be synthesized perfectly using the aliasing cancelation terms and windowing property. FAC signals in Figure 4b,f are respectively defined as

$$\begin{cases} U(A_0 \circ W_2, A_1 \circ W_3) = -A_0 \circ W_2 R - A_1 \circ W_3, \\ S(A_2 \circ W_0, A_3 \circ W_1) = A_2 \circ W_0 - A_3 \circ W_1 R. \end{cases} \quad (12)$$

We can obtain the dummy signals in Figure 4d,h by

$$\begin{cases} S(A_0 \circ W_0, A_1 \circ W_1) \circ W_0 - U(A_0 \circ W_2, A_1 \circ W_3) R \circ W_2 = A_0, \\ -U(A_2 \circ W_2, A_3 \circ W_3) \circ W_3 - S(A_2 \circ W_0, A_3 \circ W_1) R \circ W_1 = A_3. \end{cases} \quad (13)$$

Note that there is no difference between dummy signals and adjacent ACELP signals if they have the same quantization error or do not have any quantization error. The synthesized signals in Figure 4c,g are calculated as follows:

$$\begin{cases} -S(A_0 \circ W_0, A_1 \circ W_1) R \circ W_1 - U(A_0 \circ W_2, A_1 \circ W_3) \circ W_3 = A_1, \\ -U(A_2 \circ W_2, A_3 \circ W_3) R \circ W_2 + S(A_2 \circ W_0, A_3 \circ W_1) \circ W_0 = A_2. \end{cases} \quad (14)$$

Actually, the aliasing parts in Equation 9 are $-A_0 \circ W_0$ and $-A_3 \circ W_3$. As previously mentioned in Figure 5, it is clear that outputs are perfectly synthesized if these terms are removed. The new algorithm generates the aliasing cancelation terms from the adjacent ACELP signals such as

$$\begin{cases} (-S(A_0 \circ W_0, A_1 \circ W_1) R + A_0 \circ W_0 R) \circ W_1^{-1} = A_1, \\ (-U(A_2 \circ W_2, A_3 \circ W_3) R - A_3 \circ W_3 R) \circ W_2^{-1} = A_2. \end{cases} \quad (15)$$

Theoretically, if there is no quantization error, the FAC algorithm and new aliasing cancelation algorithm are able to perfectly reconstruct the original signal in the transition frame. Practically, since the quantization error is generated by several passes of non-linear filters in the time and frequency domain, it is very difficult to mathematically model the impact of the error. However, it is clear that the FAC method has a quantization error in the frequency domain, while the proposed algorithm includes

Table 3 SNR at 12-, 16-, and 24-kbps operating modes

Category	Mode	Prop.-B (dB)	Conv. (dB)
Music	12 kbps	10.662	10.682
	16 kbps	12.452	12.544
	24 kbps	14.530	14.531
Speech	12 kbps	10.794	10.840
	16 kbps	12.033	12.026
	24 kbps	12.934	12.842
Mixed	12 kbps	9.831	9.846
	16 kbps	11.678	11.703
	24 kbps	13.456	13.479
Total	12 kbps	10.429	10.456
	16 kbps	12.054	12.091
	24 kbps	13.640	13.617

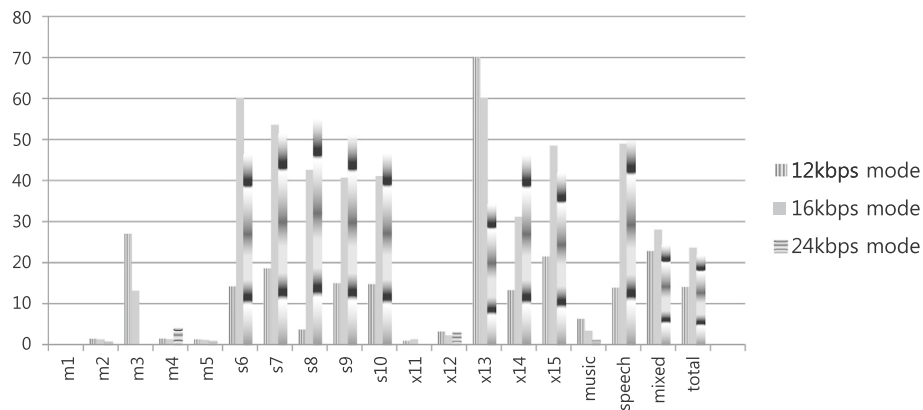


Figure 7 FAC frame rate of each test item and each category.

the error caused by ACELP encoding and inverse windowing. Accordingly, the amount of quantization error can be evaluated and compared by measuring signal-to-noise ratio (SNR) values. As will be shown from the experimental results given in the next section, there is no difference between the proposed and the conventional FAC algorithm. Subjective listening test also confirms the result.

4 Performance evaluation

4.1 Simulation setup and implementation

To verify the performance of the proposed algorithm, the USAC common encoder (JAME) is used as a baseline. The JAME developed by ourselves is officially released by MPEG as an open source [8], and its decoder module generates the bit-exact output set by the standardization process. In the recent verification test [18], the JAME encoder shows significantly better quality than the reference model encoder (RME) and comparable quality to the state-of-the-art reference quality encoder (RQE). Since the RQE is not publicly available, the JAME is a good baseline

system for implementing the proposed algorithm. Table 1 summarizes the 15 test items used for the USAC standardization process, which are selected for performance evaluation in this paper. Both objective and subjective tests are performed to evaluate the performance of the proposed algorithm.

Note that USAC is designed to have a capability of dynamic bit allocation in each frame. Therefore, the achieved average bit rate of each test item in each operating mode needs to be measured. Two methods are implemented for evaluation. First is the conventional method using FAC algorithm (Conv.), and second is the proposed method using new aliasing cancelation algorithm (Prop. -B). Table 2 shows the actual achieved bit rates of two methods in operating modes of 12, 16, and 24 kbps. The bit rates of the proposed algorithm (Prop.-B) are much less than those of the conventional algorithm (Conv.) because it does not need bits for encoding FAC signal. As shown in Table 2, we attached the symbol (-B) into the label of the proposed output (Prop.-B) for emphasizing not to use additional bits.

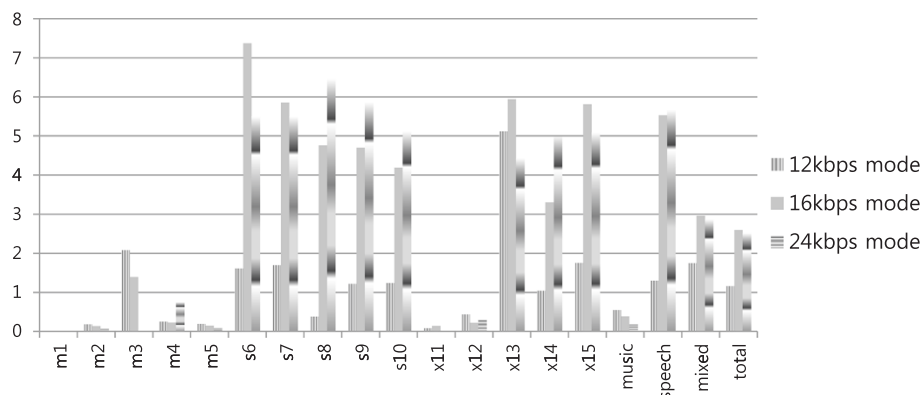


Figure 8 FAC bit ratio of each test item and each category.

Table 4 Subjective test environment

Feature	Description
Methodology	MUSHRA
Number of subjects	8
Headphones	Sennheiser HD600
Systems under the test	ref: Hidden reference lp35: 3.5 kHz Low-pass anchor Conv.: JAME with FAC Prop.-B: JAME with New AC
Modes	12, 16, and 24 kbps mono

4.2 Objective test

Figure 6 shows an example of speech spectrogram that includes mode transition frames. If there is no aliasing cancelation algorithm, the output has severe distortion as shown in Figure 6d. Since the distortion is spread out to all frequency bands, it is heard as strong click noise. These perceptually annoying noises exist more frequently in speech and mixed signals because more transition frames occur in the contents. To clarify the effectiveness of the proposed algorithm, the signal-to-noise ratio is measured at 12-, 16-, and 24-kbps operating modes.

Table 3 summarizes the results. The SNR of the proposed algorithm (Prop.-B) is similar to that of the FAC algorithm (Conv.). Note that the proposed method does not need any additional bits compared to FAC algorithm;

thus, the transmitted bits for encoding FAC frames can be saved. To measure the number of bits to be saved, the FAC frame rate is computed in each test item and each category. The FAC frame rate, α , is calculated as

$$\alpha (\%) = \frac{N_{\text{fac}}}{N} \times 100, \tag{16}$$

where N_{fac} is the number of FAC frames, and N is the number of total frames.

The FAC bit ratio, β , is obtained as

$$\beta (\%) = \frac{1}{\bar{B}} \sum_{i=1}^N B_{i,\text{fac}} \times 100, \tag{17}$$

where $B_{i,\text{fac}}$ is the FAC bits for i th frame and \bar{B} is the number of total bits.

Figures 7 and 8 depict the results. The FAC frame rate at the 12-kbps operating mode is lower than those at the 16- and 24-kbps operating modes because the allocated bits for the FAC frame are insufficient. Since music contents generally do not use the ACELP coding mode, it hardly has any FAC frame. On the contrary, FAC rates of speech at 16- and 24-kbps operating modes are around 50%. In case of mixed signal, the speech-dominant content has many FAC frames. The FAC bit ratio of the speech-like signals at 16- and 24-kbps operating modes are over 5%. The rate at the 12-kbps operating mode is lower than others due to the insufficient amount of available bits.

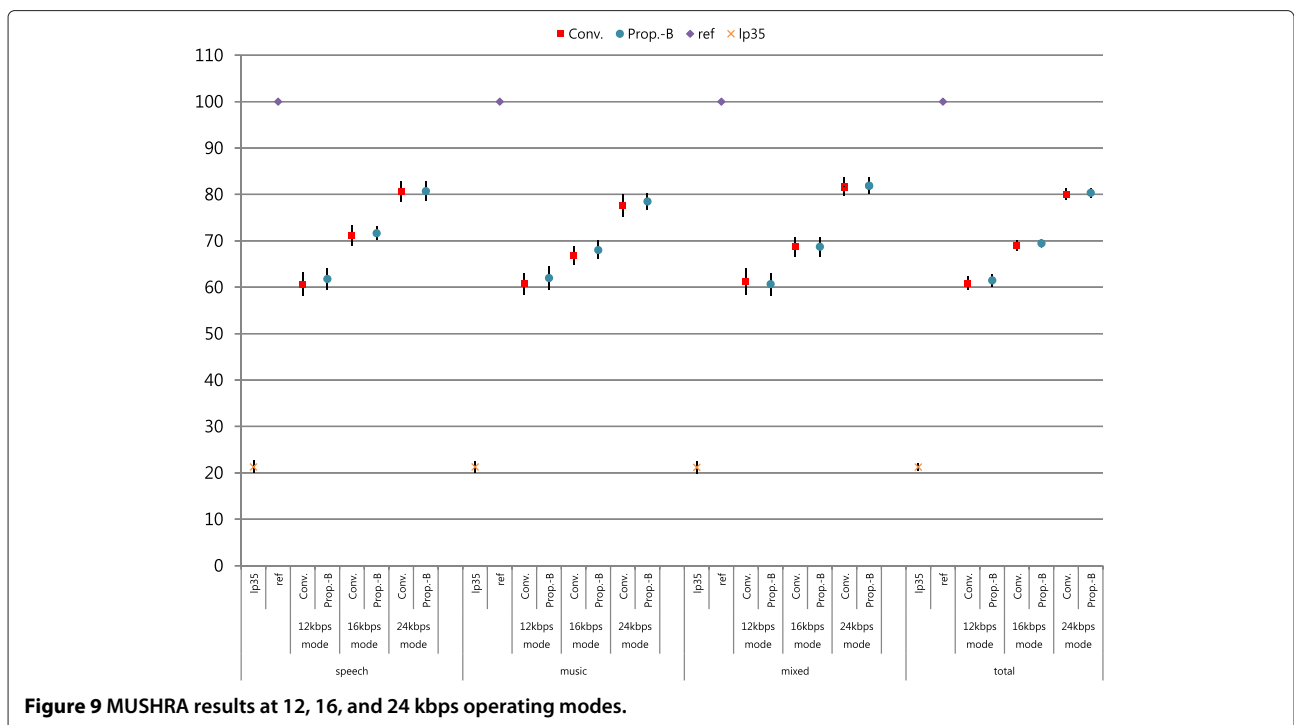


Figure 9 MUSHRA results at 12, 16, and 24 kbps operating modes.

4.3 Subjective test

Through the measurement of SNR and FAC bit ratio, it shows that the proposed algorithm has comparable performance to the USAC standard while it does not need any additional bits for FAC frames as given in Table 2. To verify the performance in terms of perceptual quality, listening tests are performed. Table 4 summarizes the test environment. Eight trained listeners participated in the Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) [19] test for the contents encoded and decoded at 12-, 16-, and 24-kbps operating modes. Results given in Figure 9 denote mean values and 95% confidence levels of test scores, and they used the same achieved bit rates given by Table 2.

The synthesized signal using the proposed algorithm (Prop.-B) has comparable performance to the FAC algorithm (Conv.). Note again that the proposed method does not need additional bits to remove the aliasing term as we have explained before.

5 Conclusions

Although the FAC algorithm solves the switching problem caused by combining two heterogeneous types of coders, i.e., time domain coder and frequency domain coder, it needs additional bits to cancel out the aliasing components at every transition frame. The proposed new aliasing cancelation algorithm does not need additional bits because it efficiently utilizes decoded signals in the adjacent frames. The proposed algorithm is sophisticatedly integrated into the recently released open-source platform. In case of speech-like signals, it saves over 5% of the total bits compared with the conventional FAC algorithm. Both subjective listening tests and objective tests confirmed that the proposed algorithm showed comparable quality to the conventional FAC algorithm, but it does not require any additional bits for FAC encoding.

Competing interests

The authors declare that they have no competing interests.

Authors' information

JS received his B.S. and M.S. degrees in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 2004 and 2008, respectively. He is currently pursuing his Ph.D. degree at Yonsei University. His research interests include speech coding, unified speech and audio coding, spatial audio coding, and 3D audio. HGK (M94) received his B.S., M.S., and Ph.D. degrees in electronic engineering from Yonsei University, Seoul, South Korea, in 1989, 1991, and 1995, respectively. He was a Senior Member of the Technical Staff at AT&T, Labs-Research, from 1996 to 2002. In 2002, he joined the Department of Electrical and Electronic Engineering, Yonsei University, where he is currently a professor. His research interests include speech signal processing, array signal processing, and pattern recognition.

Acknowledgements

The authors would like to thank the reviewers for their suggestions which have contributed a lot to the great improvement of the manuscript.

Received: 1 August 2013 Accepted: 7 January 2014
Published: 27 January 2014

References

1. M Neuendorf, M Multrus, N Rettelbach, G Fuchs, J Robilliard, J Lecomte, S Wilde, S Bayer, S Disch, C Helmrich, R Lefebvre, P Gournay, B Bessette, J Lapierre, K Kjörling, H Purnhagen, L Villemoes, W Oomen, E Schuijers, K Kikuri, T Chinen, T Norimatsu, CK Seng, E Oh, M Kim, S Quackenbush, B Grill, MPEG unified speech and audio coding - the ISO/MPEG standard for high-efficiency audio coding of all content types, in *130th AES Convention* (Budapest, 26–29 April 2012)
2. ISO/IEC JTC1/SC29/WG11, *Unified Speech and Audio Coding Verification Test Report N12232*. (ISO/IEC JTC 1, New York, 2011)
3. J Makinen, B Bessette, S Bruhn, P Ojala, R Salami, A Taleb, AMR-WB+: a new audio coding standard for 3RD generation mobile audio services. *IEEE Int. Conf. Acoustics Speech Signal Process.* (ICASSP '05), **2**, 1109–1112 (2005)
4. M Wolters, K Kjörling, D Homm, H Purnhagen, Closer look into MPEG-4 high efficiency AAC, in *115th AES Convention* (Jacob K Javits Convention Center New York, 10–13 October 2003)
5. ISO/IEC JTC1/SC29/WG11, *Proposal for Unification of USAC Windowing and Frame Transitions M17020*. (ISO/IEC JTC 1, New York, 2009)
6. D Virette, B Kövesi, P Philippe, Adaptive time-frequency resolution in modulated transform at reduced delay. *IEEE Int. Conf. Acoustics Speech Signal Process.* (ICASSP '08), **2**, 3781–3784 (2008)
7. ISO/IEC JTC1/SC29/WG11, *Proposed Core Experiment for Enhanced Low Delay AAC M14237*. (ISO/IEC JTC 1, New York, 2007)
8. ISO/IEC JTC1/SC29/WG11, *Unified Speech and Audio Coder Common Encoder Reference Software N12022*. (ISO/IEC JTC 1, New York, 2011)
9. ISO/IEC JTC1/SC29/WG11, *ISO/IEC 23003-3/FDIS, Unified Speech and Audio Coding N12231*. (ISO/IEC JTC 1, New York, 2011)
10. JP Princen, AB Bradley, Analysis/synthesis filter bank design based on time domain aliasing cancellation. *IEEE Trans. Acoustics Speech Signal Process.* **34**(5), 1153–1161 (1986)
11. K Brandenburg, M Bosi, Overview of MPEG audio: current and future standards for low-bit-rate audio coding. *J. Audio Eng. Soc.* **45**(1–2), 4–21 (1997)
12. JD Johnston, Estimation of perceptual entropy using noise masking criteria. *IEEE Int. Conf. Acoustics Speech Signal Process.* (ICASSP '98), **5**, 2524–2527 (1998)
13. G Fuchs, M Multrus, M Neuendorf, R Geiger, MDCT-based coder for highly adaptive speech and audio coding, in *European Signal Processing Conference (EUSIPCO 2009)* (Glasgow, August 2009), pp. 24–28
14. G Fuchs, V Subbaraman, M Multrus, Efficient context adaptive entropy coding for real-time applications, in *IEEE International Conference on Acoustics Speech Signal Process.* (ICASSP '11) (IEEE Piscataway, 2011), pp. 493–496
15. J Lecomte, P Gournay, R Geiger, B Bessette, M Neuendorf, Efficient cross-fade windows for transitions between LPC-based and non-LPC based audio coding, in *126th AES Convention* (Munich, 7–10 May 2009)
16. C-M Liu, W-C Lee, Unified fast algorithm for cosine modulated filter banks in current audio coding standards. *J. Audio Eng. Soc.* **47**(12), 1061–1075 (1999)
17. RA Horn, The Hadamard product. *Symp. Appl. Math.* **40**, 87–169 (1990)
18. ISO/IEC JTC1/SC29/WG11, *Verification Test Report on USAC Common Encoder, JAME N13215*. (ISO/IEC JTC 1, New York, 2012)
19. ITU, *Recommendation ITU-R BS.1534-1. Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems 2001–2003*. (International Telecommunication Union, Geneva, 2003)

doi:10.1186/1687-4722-2014-3

Cite this article as: Song and Kang: New aliasing cancelation algorithm for the transition between non-aliased and TDAC-based coding modes. *EURASIP Journal on Audio, Speech, and Music Processing* 2014 **2014**:3.