

Review Article

AVS-M Audio: Algorithm and Implementation

Tao Zhang, Chang-Tao Liu, and Hao-Jun Quan

School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China

Correspondence should be addressed to Tao Zhang, zhangtao@tju.edu.cn

Received 15 September 2010; Revised 5 November 2010; Accepted 6 January 2011

Academic Editor: Vesa Valimäki

Copyright © 2011 Tao Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, AVS-M audio standard targeting at wireless network and mobile multimedia applications has been developed by China Audio and Video Coding Standard Workgroup. AVS-M demonstrates a similar framework with AMR-WB+. This paper analyses the whole framework and the core algorithms of AVS-M with an emphasis on the implementation of the real-time encoder and decoder on DSP platform. A comparison between the performances of AVS-M and AMR-WB+ is also given.

1. Introduction

With the expanding of wireless network bandwidth, the wireless network has been documented to support not only the traditional voice services (bandwidth of 3.4 kHz), but also music with bandwidths of 12 kHz, 24 kHz, 48 kHz, and so forth. This advancement promotes the growth of various audio services, such as mobile music, mobile audio conference, and audio broadcasting. However, the current wireless network is unable to support some popular audio formats (e.g., MP3 and AC3) attributed to the bandwidth limitation. To solve this problem, many audio standards for mobile applications have been proposed, such as G.729 series standard (ITU-T), AMR series standard (3GPP), and AVS-M audio standard (AVS workgroup, China) [1, 2].

ITU-T proposed a series of audio coding algorithm standards, including G.711/721/722/723, and so forth. In 1995, ITU-T released a new audio coding standard, G.729, which adopted Conjugate Structure Algebraic Code Excited Linear Prediction (CS-ACELP). G.729 employs only 8 kbps bandwidth to provide almost the same quality of Adaptive Differential Pulse Code Modulation (ADPCM) with 32 kbps bandwidth. Therefore, it is now widely used in IP-phone technology.

The audio coding standards-Adaptive Multirate (AMR), Adaptive Multirate Wideband (AMR-WB), and Extended Adaptive Multirate Wideband (AMR-WB+) proposed by Third Generation Partnership Project (3GPP) have been

widely employed. With Algebraic Code Excited Linear Prediction (ACELP) technology, AMR is mainly used for speech coding. As the extension of AMR, AMR-WB+ is a wideband speech coding standard, which integrates ACELP, Transform Coded excitation (TCX), High-Frequency Coding and Stereo Coding. AMR-WB+ supports the stereo signal and high sampling rate thus, it is mainly used for high-quality audio contents.

Audio and Video coding Standard for Mobile (AVS-M, submitted as AVS Part 10) is a low-bit rate audio coding standard proposed for the next generation mobile communication system. This standard supports mono and stereo pulse code modulation signals with the sampling frequency of 8 kHz, 16 kHz, 24 kHz, 48 kHz, 11.025 kHz, and 44.1 kHz [3] for 16-bit word length.

In this paper, we mentioned the framework and core algorithms of AVS-M and compared the performances of AVS-M and AMR-WB+. The two modules contributed by Tianjin University, sampling rate conversion filter and gain quantizer, are introduced in detail in Section 4.

2. AVS-M Encoder and Decoder System

The functional diagrams of the AVS-M encoder and decoder are shown in Figures 1 and 2, respectively, [4–6].

The mono or stereo input signal is 16-bit sampled PCM data. The AVS-M encoder first separates the input

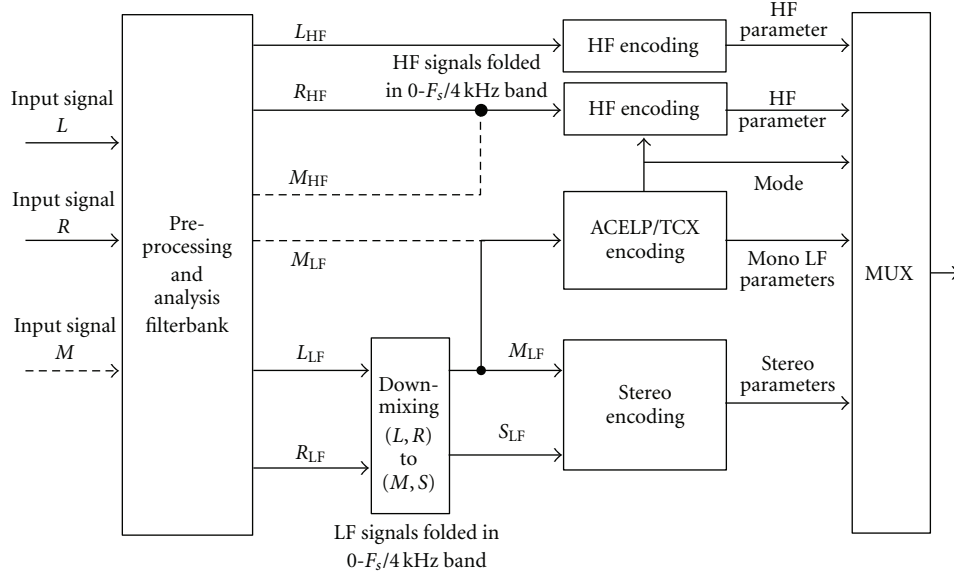


FIGURE 1: Structure of AVS-M audio encoder.

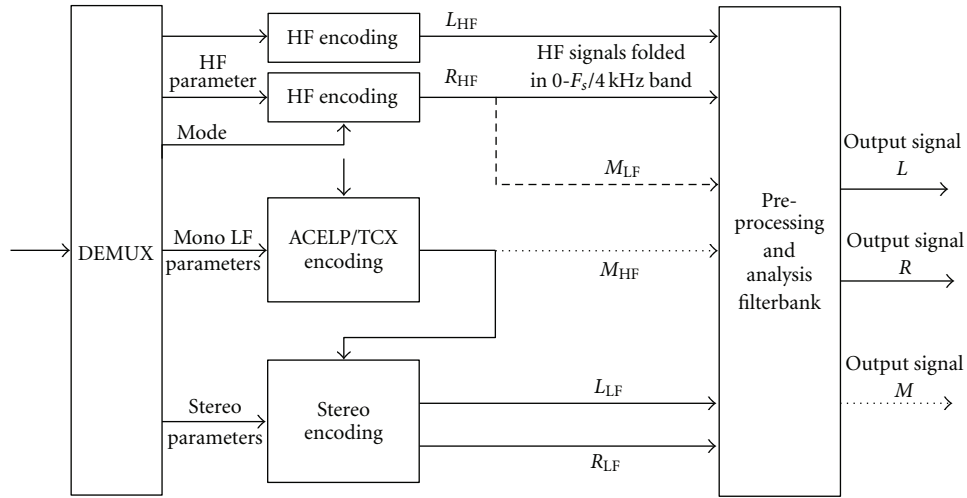


FIGURE 2: Structure of AVS-M audio decoder.

signal into two bands: low-frequency (LF) signal and high-frequency (HF) signal. Both of them are critically sampled at the frequency of $F_s/2$. The mono LF signal goes through ACELP/TCX, and HF signal goes through Band Width Extension (BWE) module. For stereo mode, the encoder downmixes the LF part of the left channel and right channel signal to main channel and side channel (M/S). The main channel is encoded by ACELP/TCX module. The stereo encoding module processes the M/S channel and produces the stereo parameters. The HF part of the left channel and right channel is encoded by BWE module to produce the HF parameters which are sent to the decoder together with the LF parameters and stereo parameters.

After being decoded separately, the LF and HF bands are combined by a synthesis filterbank. If the output is restricted to mono only, the stereo parameters are omitted and the decoder works in the monomode.

3. Key Technologies in AVS-M Audio Standard

3.1. Input Signal Processing. The preprocessing module for the input signal consists of sampling rate converter, high-pass filter, and stereo signal downmixer.

In order to maintain the consistency of the follow-up encoding process, the sampling frequency of the input signal needs to be converted into an internal sampling frequency F_s . In detail, the signal should go through upsampling, low-pass filtering and downsampling. Finally, the F_s ranges from 12.8 kHz to 38.4 kHz (typically 25.6 kHz).

Through the linear filtering, the residual signals of the M signal and the LF part of the right channel were isolated, respectively, which are then divided into two bands, very low band ($0 \sim F_s * (5/128)$ kHz) and middle band ($F_s * (5/128) \sim F_s/4$ kHz). The addition and subtraction of these middle band signals produce the middle band signals of the left and

right channels, respectively, which are encoded according to the stereo parameters. The very low band signal is encoded by TVC in stereo mode.

3.2. ACELP/TCX Mixed Encoding Module. ACELP mode, based on time-domain linear prediction, is suitable for encoding speech signals and transient signals, whereas TCX mode based on transform domain coding is suitable for encoding typical music signals.

The input signal of ACELP/TCX encoding module is monosignal with $F_s/2$ sampling frequency. The superframe for encode processing consists 1024 continuous samples. Several coding methods including ACELP256, TCX256, TCX512, and TCX1024 can be applied to one superframe. Figure 3 shows how to arrange the timing of all possible modes within one superframe.

There are 26 different mode combinations of ACELP and TCX for each superframe. The mode could be selected using the closed-loop search algorithm. In detail, all modes are tested for each superframe and the one with the maximum average segment Signal Noise Ratio (SNR) is selected. Obviously, this method is comparably complicated. The other choice is the open-loop search algorithm, in which the mode is determined by the characteristics of the signal. This method is relatively simple.

ACELP/TCX windowing structure instead of MDCT is adopted in the AVS-M audio standard. The main reason is that the MDCT-based audio standards (such as AAC, HE-AAC) show a high perceptual quality at low bit rate for music, but not speech, whereas the audio standards (such as AMR-WB+), that based on the structure of ACELP/TCX, can perform a high quality for speech at low bit rate and a good quality for music [7].

3.3. ACELP/TCX Mixed Encoding. The Multirate Algebraic Code Excited Linear Prediction (MP-ACELP) based on CELP is adopted in ACELP module. As CELP can produce the voice signal using the characteristic parameters and waveform parameters carried in the input signal. The schematic picture of ACELP encoding module is shown in Figure 4 [8–10].

As illustrated in Figure 4, the speech input signal is firstly filtered through a high-pass filter (part of the Pre-processing) to remove redundant LF components. Then, a linear prediction coding (LPC) is used for each frame, where Levinson-Durbin algorithm is used to solve the LP coefficients [11]. For easy quantization and interpolation, the LP coefficients are converted to Immittance Spectral Frequencies (ISF) coefficients.

3.3.1. ISF Quantization. In each frame, the ISF vector, which comprises 16 ISF coefficients, generates a 16-dimensional residual vector (marked as VQ_1) by subtracting the average of ISF coefficients in current frame and the contribution of the previous frame for the current frame. The (16-dimensional) residual ISF vector will be quantified and transferred by the encoder. After the interleaved grouping and intra-frame prediction, the residual ISF vector is quantized based on the Combination of Split Vector Quantization Multistage Vector

Quantization, as shown in Figure 5. The 16-dimensional residual ISF vector is quantified with 46 bits totally [12, 13].

After the quantization and interpolation, the un-quantified ISP coefficients will be converted to LP coefficients and processed by formant perceptual weighting. The signal is filtered in the perceptual weighting domain. The basis of formant perceptual weighting is to produce the spectrum flattening signal by selecting the corresponding filter according to the energy difference between high- and low-frequency signal.

Following perceptual weighting, the signal is down-sampled by a fourth-order FIR filter [14]. And, then, open-loop pitch search is used to calculate an accurate pitch period to reduce the complexity of the closed-loop pitch search.

3.3.2. Adaptive Codebook Excitation Search. The subframe represents the unit for codebook search, which includes the closed-loop pitch search and the calculation and processing of the adaptive codebook. According to the minimum mean square weighted error between the original and reconstructed signal, the adaptive codebook excitation $v(n)$ is obtained during the closed-loop pitch search. In wideband audio, the periodicities of surd and transition tone are relatively less strong and they may not interrupt the HF band. A wideband adaptive codebook excitation search algorithm is proposed to simulate the harmonics characteristic of the audio spectrum, which improves the performance of the encoder [15].

First, the adaptive code vector $v(n)$ passes through a low-pass filter, which separates the signal into low band and high-band. Then, correlation coefficient of the high band signal and the quantized LP residual is calculated. At last, based on the comparison of the correlation coefficient and a given threshold, the target signal for adaptive codebook search can be determined. The gain can also be generated in this process.

3.3.3. Algebraic Codebook Search. Comparing with CELP, the greatest advantage of ACELP speech encoding algorithm is the fixation of the codebook. The fixed codebook is an algebraic codebook with conjugate algebraic structure. The codebook improves the quality of synthesis speech greatly due to its interleaved single-pulse permutation (ISPP) structure.

The 64 sample locations for each subframe are divided into four tracks, each of which includes 16 positions. The number of algebraic codebooks on each track is determined by the corresponding bit rate. For example, for 12 kbps mode, the random code vector has six pulses and the amplitude of each pulse is +1 or -1. Among the 4 tracks, both track 0 and track 1 contain two pulses and each other track contains only one pulse. The search procedure works in such a way that all pulses in one track are found at a time [16].

The algebraic codebook is used to indicate the residual signal, which is generated by the short-time filtering of the original speech signal. The algebraic codebook contains a huge number of code vectors, which provides an accurate error compensation for the synthesis speech signal; so, it greatly improves the quality of synthesis speech generated

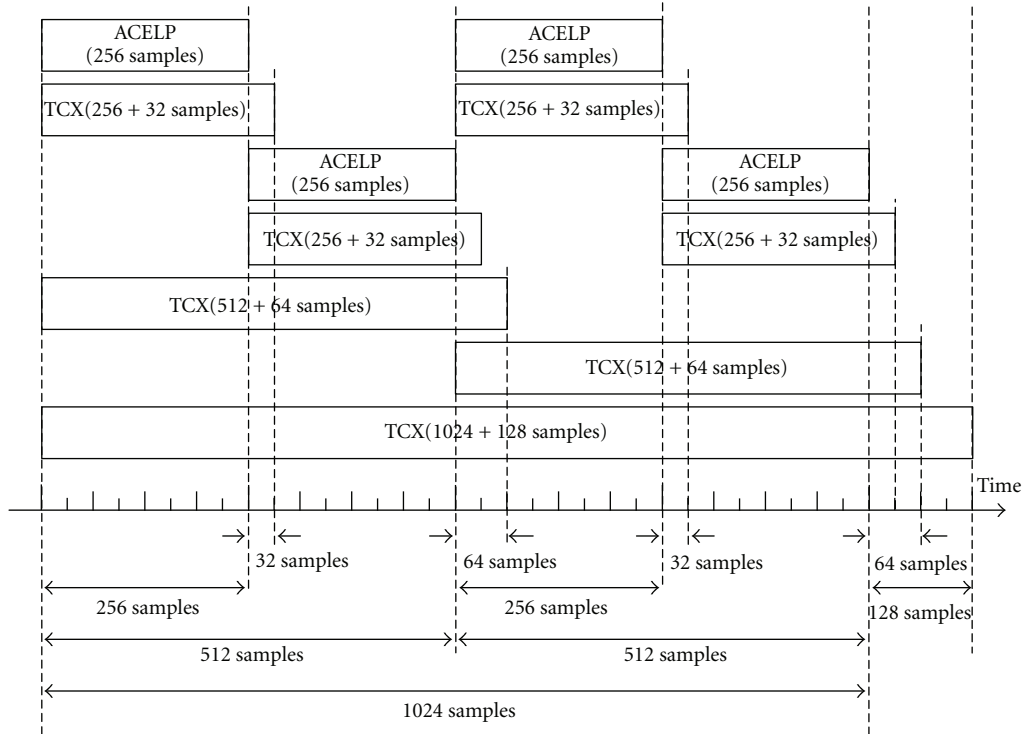


FIGURE 3: Each superframe encoding mode.

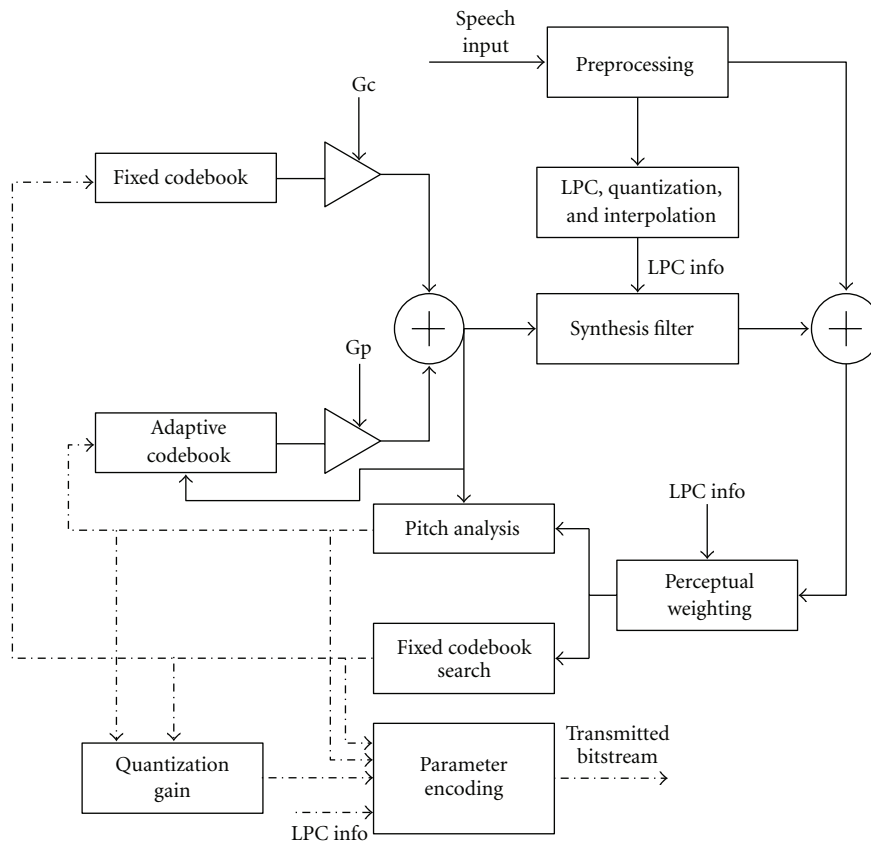


FIGURE 4: ACELP encoding module.

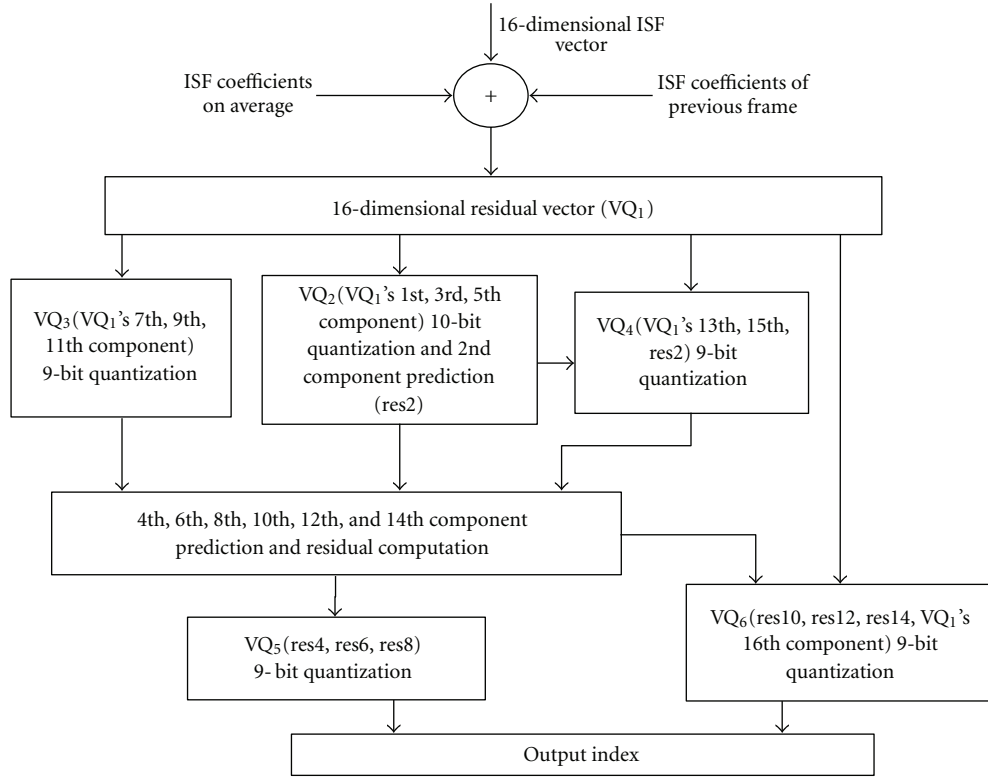


FIGURE 5: AVS-M audio ISF vector quantization.

by ACELP speech encoding algorithm. The parameter of algebraic codebook includes the optimum algebraic code vectors and the optimum gain of each frame. When searching the optimum algebraic code vector for each subframe, the optimum pitch delayed code vector is fixed and, then, the code vector with the optimum pitch delayed code vector is added upon. After passing through the LP synthesis filter, the optimum algebraic code vector and gain can be fixed through synthetic analysis.

The input of the decoder includes ISP vectors, adaptive codebook, and the parameters of algebraic codebook, which could be got from the received stream. The line spectrum parameters of ISP are transformed into the current prediction filter coefficients. Then, according to the interpolation of current prediction coefficients, the synthetic filter coefficients of each subframe can be generated. Excitation vectors can be obtained according to the gain weighting of adaptive codebook and algebraic codebook. Then, the noise and pitch are enhanced. Finally, the enhanced excitation vectors go through the synthesis filter to reconstruct the speech signal.

3.3.4. TCX Mode Encoding. TCX excitation encoding is a hybrid encoding technology. It is based on time domain linear prediction and frequency domain transform encoding. The input signal goes through a time-varying perceptual weighting filter to produce a perceptual weighting signal. An adaptive window is applied before the FFT transform. Consequently, the signal is transformed into the frequency domain. Scalar quantization based on split table is applied

to the spectrum signal. TCX encoding diagram is shown in Figure 6 [3, 17].

In TCX, to smooth the transition and reduce the block effect, nonrectangular overlapping window is used to transform the weighting signal. In contrast, ACELP applies a non-overlapping rectangular window. So, adaptive window switching is a critical issue for ACELP/TCX switching.

If the previous frame is encoded in ACELP mode and the current frame is encoded in TCX mode, the length of overlapping part should be determined by the TCX mode. This means that some (16/32/64) data at the tail of previous frame and some data at the beginning of current frame are encoded together in TCX mode. The input audio frame structure is shown in Figure 7.

In Figure 7, L_{frame} stands for the length of current TCX frame. L_1 stands for the length of overlapping data of previous frame. L_2 is the number of overlapping data for the next frame. L is the total length of current frame. The relationships between L_1 , L_2 , and L are as follows:

When the $L_{\text{frame}} = 256$, $L_1 = 16$, $L_2 = 16$, and $L = 288$,

When the $L_{\text{frame}} = 512$, $L_1 = 32$, $L_2 = 32$, and $L = 576$,

When the $L_{\text{frame}} = 1024$, $L_1 = 64$, $L_2 = 64$, and $L = 1152$.

We see that the value of L_1 , L_2 , and L should change adaptively, according to the TCX mode (or frame length).

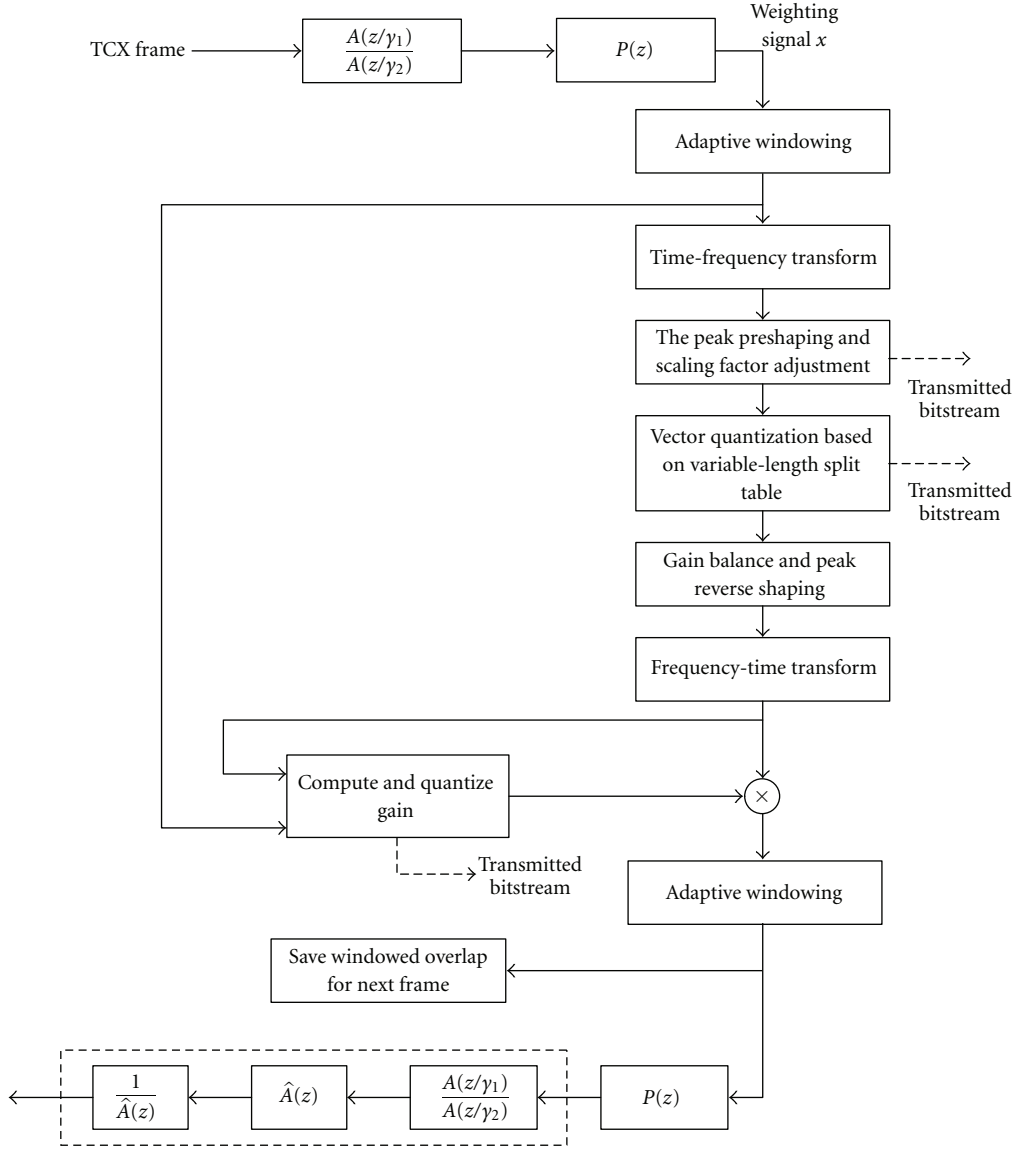


FIGURE 6: TCX encoding mode.

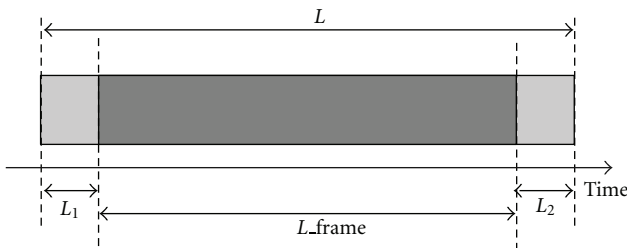


FIGURE 7: The TCX input audio frame structure.

After the perceptual weighting filter, the signal goes through the adaptive windowing module. The adaptive window is shown in Figure 8.

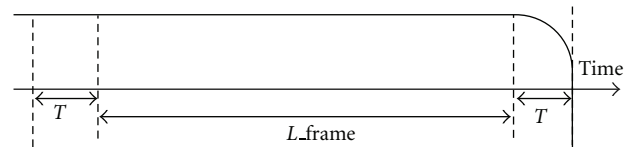


FIGURE 8: Adaptive window.

There is no windowing for the overlapping data of previous frame. But for the overlapping data of next frame, a cosine window $w(n)$, ($w(n) = \sin(2\pi n/4L_2)$, $n = L_2, L_2 + 1, \dots, 2L_2 - 1$) is applied. Because of the overlapping part of the previous frame, if the next frame will be encoded in TCX mode, the length of the window for the header of the next frame should equal to L_2 .

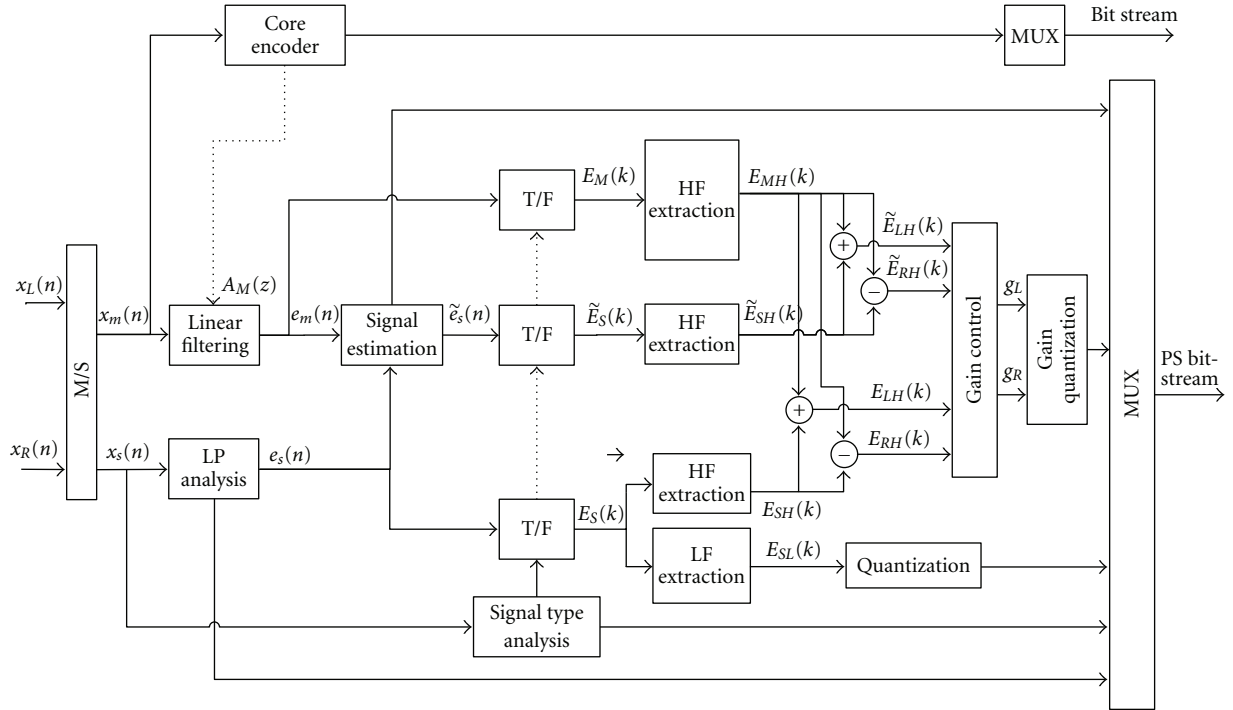


FIGURE 9: Stereo signal encoding module.

The input TCX frame is filtered by a perceptual filter to obtain the weighted signal x . Once the Fourier spectrum X (FFT) of x is computed, a spectrum preshaping is applied to smooth X . The coefficients are grouped in blocks with 8 data in one block, which can be taken as an 8-dimensional vector. To quantize the preshaped spectrum X in TCX mode, a method based on lattice quantizer is used. Specifically, the spectrum is quantized in 8-dimensional blocks using vector codebooks composed of subsets of the Gusset lattice, called RE8 lattice. In AVS-M, there are four basic codebooks (Q_0 , Q_2 , Q_3 , and Q_4) constructed with different signal statistical distribution. In lattice quantization, finding the nearest neighbor y of the input vector x among all codebook locations is needed. If y is in the base codebook, its index should be computed and transmitted. If not, y should be mapped to a basic code and an extension index, which are then encoded and transmitted.

Because different spectrum samples use different scale factors, the effect of different scale factors should be reduced when recovering the original signal. This is called gain balance. At last, the minimum mean square error can be calculated using the signal recovered from the bitstream. This can be achieved by utilizing the peak preshaping and global gain technologies.

The decode procedure of TCX module is just the reverse of encode procedure.

3.4. Monosignal High-Band Encoding (BWE). In AVS-M audio codec, the HF signal is encoded using BWE method [18]. The HF signal is composed of the frequency components above $F_s/4$ kHz in the input signal. In BWE, energy

information is sent to the decoder in the form of spectral envelop and gain. But, the fine structure of the signal is extrapolated at the decoder from the decoded excitation signal in the LF signal. Simultaneously, in order to keep the continuity of the signal spectrum at the $F_s/4$, the HF gain needs to be adjusted according to the correlation between the HF and LF gain in each frame. The bandwidth extension algorithm only needs a small amount of parameters. So, 16 bits are enough.

At the decoder side, 9-bit high frequency spectral envelopes are separated from the received bitstream and inverse quantified to ISF coefficients, based on which the LPC coefficients and HF synthesis filter can be obtained. Then, the filter impulse response is transformed to frequency domain and normalized by the maximum FFT coefficients. The base signal is recovered by multiplying the normalized FFT coefficients with the FFT coefficients of LF excitation. Simultaneously, 7-bit gain factor can be separated from the received bitstream and inverse quantified to produce four subband energy gain factors in the frequency domain. These gain factors can be used to modulate the HF base signal and reconstruct HF signal.

3.5. Stereo Signal Encoding and Decoding Module. A high-effective configurable parametric stereo coding scheme in the frequency domain is adopted in AVS-M, which provides a flexible and extensible codec structure with coding efficiency similar to that of AMR-WB+. Figure 9 shows the functional diagram of the stereo encoder [19].

Firstly, the low-band signals $x_L(n)$ and $x_R(n)$ are converted into the main channel and side channel (M/S for short)

TABLE 1: The core module comparison of AVS-M and AMR-WB+.

Modules	Improvements	Performance comparison of AVS-M and AMR-WB+
Sampling rate conversion filter	Adopting a new window	With the same order and cut-off frequency with the filter of AMR-WB+, the filter of AVS-M reduces the transition band width and the minimum stop-band attenuation greatly (about 9 dB). Therefore better filtering effect is obtained than that of AMR-WB+
Parametric stereo coding	(1) According to bit rate, the low-frequency bandwidth can be controlled flexibly on accurate coding (2) Using gain control in the frequency domain for the high frequency part (3) Using the time-frequency transform for the channels after sum/difference processing, to avoid the time delay caused by re-sampling	Compared with AMR-WB+, AVS-M has flexible coding structure with lower complexity, does not require resampling, and gives greater coding gain is and higher frequency resolution
ACELP	An efficient wideband adaptive codebook excitation search algorithm is supported	With lower complexity, AVS-M gives similar performance with AMR-WB+
ISF quantization	(1) Line spectral frequency (LSF) vector quantization based on interlace grouping and intra-prediction is used (2) Based on the correlation of LSF coefficients of intra and inter frame, AVS-M uses the same amount of bits to quantify the LSF coefficients with AMR-WB+	Compared with AMR-WB+, the average quantization error is reduced and the voice quality is improved slightly
Perceptual weighting	Voice quality is improved by reducing the significance of formant frequency domain	AVS-M has the similar performance with AMR-WB+
Algebraic codebook search	(1) Based on priority of tracks (2) Multirate encoding is supported, and the number of pulses can be arbitrarily extended	With low computation complexity, AVS-M has better voice quality than AMR-WB+ at low bit rate, and the performance at high bit rate is similar to AMR-WB+
The ISF replacement method for error concealment of frames	(1) The last number of consecutive error frames is counted. When consecutive error frames occur, the correlation degree of current error frame and last good frame is reduced (2) When frame error occurs and the ISF parameters need to be replaced, the ISF of last good frame is used instead of other frames	Experiment shows that we can get better sound quality under the same bit rate and frame error rate with AMR-WB+. The computational complexity and memory requirement of the AVS-M decoder are reduced

signal $x_m(n)$ and $x_s(n)$, which then go through the linear filter to produce the residual signals of M/S signals $e_m(n)$ and $e_s(n)$. A Wiener signal estimator produces the residual estimated signal $\tilde{e}_s(n)$ based on $x_m(n)$. Then, $e_m(n)$, $e_s(n)$, and $\tilde{e}_s(n)$ are windowed as a whole to reduce the block effect of quantization afterward. The window length is determined according to the signal type. For stationary signals, a long window will be applied to improve the coding gain, while short windows are used for transient signals. Following the windowing process, a time-to-frequency transform will be applied, after which the signals are partitioned into high-frequency part and low-frequency part. The LF part is further decomposed into two bands, the very low frequency (VLF) and relatively high-frequency part (Midband). For the VLF part of $e_s(n)$, a quantization method called Split Multirate Lattice vector quantization is performed, which is the same as that in AMR-WB+. Because the human hearing is not sensitive to the details of the HF part, just the envelope is encoded using the parameter encoding method. The high-frequency signal is partitioned into several subbands.

For stationary signal, it will be divided into eight uniform subbands; and for transient signal, it will be divided into two uniform subbands. Each subband contains two gain control coefficients. Finally, vector quantization will be used to the coefficients of Wiener filter, as well as the gain coefficients g_L and g_R .

Through above analysis, it is clear that the parametric stereo coding algorithm successfully avoids the resamplings in the time domain; so, it reduces the complexity of encoder and decoder. The ability of flexible configuration for the low frequency bandwidth determined by the coding bit rate is also available, which makes it a high-effective stereo coding approach.

3.6. VAD and Comfortable Noise Mode. Voice activity detection (VAD) module is used to determine the category of each frame, such as speech music, noise, and quiet [20]. In order to save the network resource and keep the quality of service, long period of silence can be identified and eliminated from the audio signal. When the audio signal is being transmitted,

the background noise that transmitted with speech signal will disappear when the speech signal is inactive. This causes the discontinuity of background noise. If this switch occurs fast, it will cause a serious degradation of voice quality. In fact, when a long period of silence happens, the receiver has to activate some background noise to make the users feel comfortable. At the decoder, comfortable noise mode will generate the background noise in the same way with that of encoder. So, at the encoder side, when the speech signal is inactive, the background parameters (ISF and energy parameters) will be computed. These parameters will be encoded as a silence indicator (SID) frame and transmitted to the decoder. When the decoder receives this SID frame, a comfortable noise will be generated. The comfortable noise is changed according to the received parameters.

3.7. Summary. The framework of AVS-M audio is similar to that of AMR-WB+, an advanced tool for wideband voice coding standard released by 3GPP in 2005. Preliminary test results show that the performance of AVS-M is not worse than that of AMR-WB+ on average. The performance comparison and technical improvements of the core modules are summarized in Table 1 [13, 15, 16, 19, 21].

4. The Analysis of Two Mandatory Technical Proposals

4.1. Sampling Rate Conversion Filter. In AMR-WB+, sampling rates of 8, 16, 32, 48, 11, 22, and 44.1 kHz are supported. Three FIR filters are used for anti-overlap filtering: filter_lp12, and filter_lp165, filter_lp180. The filter coefficients are generated by Hanning window [4, 5].

AVS-M employs a new window function for the sampling rate conversion in the preprocessing stage. This new window is deduced from the classic Hamming window. The detail derivation of the modifying window is given in [22].

The signal $f = e^{-|n|}$ is two side even exponential, and its Fourier Transform is $F(e^{j\omega}) = 2/(1 + \omega^2)$. As ω increases from 0 to infinite, $F(e^{j\omega})$ decreases more and more rapidly. The modifying window $e(n)$ is given as the convolution of f and r , where r is in the form of

$$r(n) = \begin{cases} 1, & 0 \leq n \leq N-1, \\ 0, & \text{other.} \end{cases} \quad (1)$$

Here, N is the length of the window. In the time domain, $e(n)$ can be expressed as

$$e(n) = \frac{1 + e^{-1} - e^{-(N-n)} - e^{-(n+1)}}{1 + e^{-1} - e^{-(N/2)} - e^{-((N+3)/2)}}, \quad N \text{ is odd,} \quad (2)$$

$$e(n) = \frac{1 + e^{-1} - e^{-(N-n)} - e^{-(n+1)}}{1 + e^{-1} - 2 * e^{-((N+1)/2)}}, \quad N \text{ is even.}$$

In the frequency domain, $E(e^{j\omega})$ can be expressed as

$$E(e^{j\omega}) = e^{j((N-1)/2)\omega} \cdot \left[1 + 2 \cdot \sum_{n=0}^{(N-3)/2} \cos(n\omega) \right], \quad N \text{ is odd,} \quad (3)$$

$$E(e^{j\omega}) = e^{j((N-1)/2)\omega} \cdot 2 \cdot \sum_{n=0}^{N/2-1} \cos(n\omega), \quad N \text{ is even.}$$

By multiplying the modifying window $e(n)$ with the classical Hamming window, a new window function $w(n)$ can be generated. Because the Hamming window is

$$w_h(n) = 0.54 - 0.46 \cos\left(2\pi \frac{n}{N-1}\right), \quad (4)$$

$$n = 0, 1, 2, \dots, N-1.$$

The new window function $\omega(n) = e(n) \cdot w_h(n)$ can be expanded as

$$\omega(n) = \frac{1 + e^{-1} - e^{-(N-n)} - e^{-(n+1)}}{1 + e^{-1} - 2 * e^{-((N+1)/2)}} \cdot \left[0.54 - 0.46 \cdot \cos\left(\frac{2\pi n}{N-1}\right) \right], \quad \text{when } N \text{ is odd,}$$

$$\omega(n) = \frac{1 + e^{-1} - e^{-(N-n)} - e^{-(n+1)}}{1 + e^{-1} - e^{-(N/2)} - e^{-((N+3)/2)}} \cdot \left[0.54 - 0.46 \cdot \cos\left(\frac{2\pi n}{N-1}\right) \right], \quad \text{when } N \text{ is even.} \quad (5)$$

The Fourier transformation of $\omega(n)$ is

$$W(e^{j\omega}) = e^{j((N-1)/2)\omega} \cdot \left[1 + 2 \cdot \sum_{n=0}^{(N-3)/2} \omega(n) \cos(n\omega) \right], \quad N \text{ is odd,}$$

$$W(e^{j\omega}) = e^{j((N-1)/2)\omega} \cdot 2 \cdot \sum_{n=0}^{N/2-1} \omega(n) \cos(n\omega), \quad N \text{ is even.} \quad (6)$$

Table 2 compares the parameters of Hamming window and the new window $w(n)$.

On the peak ripple value, the new window $w(n)$ has a 3 dB improvement, and on the decay rate of side-lobe envelope, it makes a 2 dB/oct improvement. In Figure 10, the broken lines are for the new window $w(n)$ and the real lines are for the Hamming window.

Using this new window to generate three new filters in place of the original ones in AMR-WB+, the filter parameter comparison is shown in Table 3.

TABLE 2: New window parameter improvement.

N (length of window)		41	51	61	289
Peak ripple value (dB)	Hamming	-41	-41	-41	-41
	New	-44.3242	-43.8429	-43.5240	-42.7144
Delay rate of envelop (dB/oct)	Hamming	-6	-6	-6	-6
	New	-8.0869	-8.8000	-7.9863	-8.6869

TABLE 3: New filter parameter improvement.

parameter	filter_lp12		filter_lp165		filter_lp180	
	New	WB+	new	WB+	new	WB+
least stop-band attenuation (dB)	-52.98	-43.95	-52.99	-43.95	-52.99	-43.95

As we can see from Table 3, the new filters have about 9 dB improvement comparing to the original filters of AMR WB+ on the least stop-band attenuation [1, 21].

4.2. Gain Quantization. AMR WB+ adopts vector quantization for codebook gains to get coding gain. A mixture of scalar and vector quantization is used for the quantization of codebook gains in AVS-M [1, 9].

For the first subframe (there are 4 subframes in one frame), it is necessary to compute the best adaptive gain and the fixed gain with the criteria of the minimum mean square error, which is given by (7)

$$e = \sum_{n=0}^{N-1} [x_0(n) - g_a x_{u'}(n) - g_s t_{j'}(n)]^2. \quad (7)$$

Then, the adaptive gain is scalar-quantized with 4 bits, ranging from 0.012445 to 1.296012, and the fixed gain is scalar-quantized with 5 bits, ranging from 15.848932 to 3349.654392.

For the second, third, and fourth subframe, the fixed gain of the first subframe is used to predict that of current frame. The current adaptive gains of subframes and the predicted fixed gain are quantized using 2-dimensional vector quantization with 7 bits.

Predictor of the fixed gain is defined as (8)

$$\frac{\text{Fixed gain of Current subframe}}{\text{Fixed gain of the 1st subframe}}. \quad (8)$$

Hence, totally $9 + 7 \times 3 = 30$ bits are used to quantize the adaptive gains and the fixed gain of each frame, so this new approach uses just the same bits as in AMR-WB+. Table 4 shows the PESQ results of the new algorithm comparing with that of AMR-WB+ at 12 kbps and 24 kbps bit rate.

5. AVS-M Real-Time Encoding and Decoding

A real-time codec of AVS-M is implemented on the TMS320C6416 platform. C6416 is a high-performance fixed-point DSP of C64x DSP family. It is an excellent choice for professional audio, high-end consumer audio, industrial, and medical applications.

The key features of C6416 DSP [23] include: (1) 600 MHz clock rate and 4800 MIPS processing capacity, (2) advanced Very Long Instruction Word (VLIW) architecture: the CPU consists of sixty four 32-bit general purpose registers and eight highly independent functional units, (3) L_1/L_2 cache architecture with 1056k-byte on-chip memory; (4) two External Memory Interfaces (EMIFs), one 64-bit EMIFA and one 64-bit EMIFB, glue less interface to asynchronous memories (SRAM and EPROM) and synchronous memories (SDRAM, SBSRAM, ZBTSRAM), and (5) Enhanced Direct-Memory Access (EDMA) controller (64 independent channels).

Because C6416 is a fixed-point DSP, AVS-M Codec source code (version 9.2) should be ported to fixed-point implementation at first.

5.1. Fixed-Point Implementation of the AVS-M Audio Codec. In fixed-point DSPs, the fixed-point data is used for computation and its operand is indicated integer. The range of an integer data relies on the word length restricted by the DSP chip. It is conceivable that the longer word gives greater range and higher accuracy. To make the DSP chip handle a variety of decimal number, the key is the location of the decimal point in the integer, which is the so-called number of calibration. There are two methods to show the calibration, Q notation and S notation, the former of which is adopted in this paper.

In Q notation, the different value of Q indicates the different scope and different accuracy of the number. Larger Q gives smaller range and higher accuracy of the number. For example, the range of the Q_0 is from -32768 to 32767 and its accuracy is 1, while the range of the Q_{15} is from -1 to 0.9999695 and its accuracy is 0.00003051. Therefore, for the fixed-point algorithms, the numerical range and precision are contradictory [24]. The determination of Q is actually a tradeoff between dynamic range and precision.

5.2. The Complexity Analysis of AVS-M Fixed-Point Codec. In order to analyze the complexity of the AVS-M Codec, the AVS-M Fixed-point Codec is developed and the complexity is analyzed [25, 26]. The method of Weighted Million Operation Per Second (WMOPS) [27] approved by ITU is

TABLE 4: PESQ comparison at 12/24 kbps.

Sequence	WB+ (12 kbps)	New (12 kbps)	WB+ (24 kbps)	New (24 kbps)
CHaabF1.1.wav	3.922	3.999	4.162	4.181
CHaaeF4.1.wav	3.928	3.878	4.171	4.209
CHaafM1.1.wav	4.057	4.063	4.319	4.302
CHaaiM4.1.wav	4.017	4.064	4.285	4.264
F1S01_noise_snr10.wav	3.609	3.616	3.795	3.796
F2S01_noise_snr10.wav	3.289	3.286	3.503	3.489
M1S01_noise_snr10.wav	3.41	3.401	3.603	3.615
M2S01_noise_snr10.wav	3.331	3.345	3.547	3.535
som_ot_x_l_org_16K.wav	2.999	3.019	3.332	3.333
som_nt_x_l_org_16K.wav	3.232	3.211	3.569	3.585
som_fi_x_l_org_16K.wav	3.387	3.387	3.633	3.634
som_ad_x_l_org_16K.wav	3.246	3.264	3.591	3.685
sbm_sm_x_l_org_16K.wav	3.694	3.696	3.94	3.937
sbm_ms_x_l_org_16K.wav	3.712	3.711	4.007	4.015
sbm_js_x_l_org_16K.wav	3.76	3.754	4.068	4.067
sbm_fi_x_9_org_16K.wav	3.608	3.581	4.016	4.014
or08mv_16K.wav	3.65	3.65	3.88	3.88
or09mv_16K.wav	3.447	3.447	4.114	4.114
si03_16K.wav	3.9	3.913	4.114	4.102
sm02_16K.wav	3.299	3.296	3.579	3.625
Average	3.57485	3.57905	3.8614	3.8691

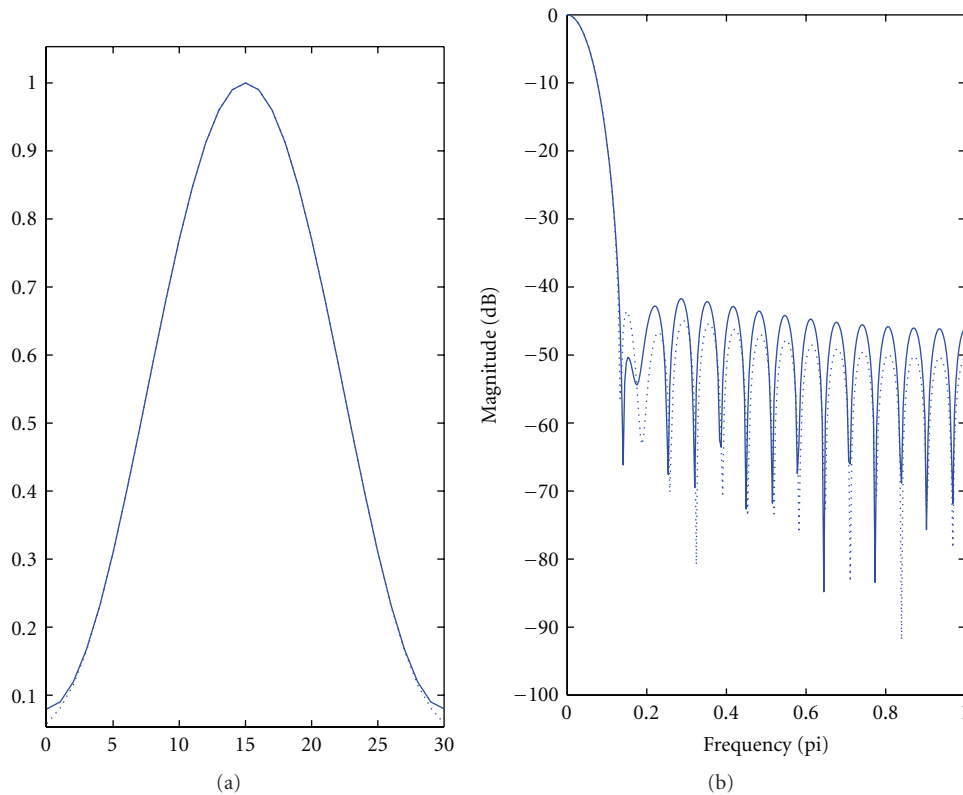
FIGURE 10: window shape and magnitude response of $w(n)$ and Hamming window.

TABLE 5: Complexity of AVS-M encoder.

Test condition	Command line parameters	Complexity (WMOPS)
12 kbps, mono	-rate 12-mono	avg = 56.318; worst = 58.009
24 kbps, mono	-rate 24-mono	avg = 79.998; worst = 80.055
12.4 kbps, stereo	-rate 12.4	avg = 72.389; worst = 73.118
24 kbps, stereo	-rate 24	avg = 83.138; worst = 83.183

adopted here to analyze the complexity of the AVS-M Codec. The analysis results are shown in Tables 5 and 6.

5.3. Porting AVS-M Fixed-Point Codec to C6416 Platform. By porting, we mean to rewrite the original implementation accurately and efficiently to match the requirements of the given platform. In order to successfully compile the code on the Code Composer Studio (CCS) [28, 29], the following procedures were needed.

5.3.1. Change the Data Type. Comparing with the Visual C platform, the CCS compiler is much stricter with the matching of the variable data type. Meanwhile, for the length of data type, different platform has different definition. For example, assigning a *const short* type constant to a *short* type variable is allowed on Visual C platform, but this generates a type mismatch error on the CCS platforms.

5.3.2. Reasonable Memory Allocation. The code and the data in the program require corresponding memory space; so, it is necessary to edit a cmd file to divide the memory space into some memory segmentations and allocate each *code* segment, *data* segment, and the *initial variable* segment into appropriate memory segmentations. For example, the malloc and calloc function would allocate memory in the heap segment, and some temporary variables and local variables would occupy the stack segment. Therefore, it is imperative to set these segments properly to prevent overflow.

5.3.3. Compiler Optimization. CCS compiler provides a number of options to influence and control the procedure of compilation and optimization. Consequently, proper compiler options can greatly improve the efficiency of the program. For example, the *-mt* option instructs the compiler to analysis and optimize the program throughout the project and improve the performance of the system. The *-o3* option instructs the compiler to perform file-level optimization, which is the highest level of optimization. When *-o3* is enabled, the compiler tries out a variety of loop optimization, such as loop unrolling, instruction parallel, data parallel and so on.

5.3.4. Assembly-Level Optimization. Although the above mentioned optimization was carried out, AVS-M encoder

TABLE 6: Complexity of AVS-M decoder.

Test condition	Command line parameters	Complexity (WMOPS)
12 kbps, mono	-mono	avg = 9.316; worst = 9.896
24 kbps, mono	-mono	avg = 13.368; worst = 13.981
12.4 kbps, stereo	None	avg = 16.996; worst = 17.603
24 kbps, stereo	None	avg = 18.698; worst = 19.103

still might not compress audio stream in real-time. Therefore, it is necessary to optimize the system further at a coding level. Here, we do assembly coding. At first, the profile tool is used to find out the key functions. Efficient sensitive functions are identified by analyzing the cycles that each function requires. Generally, the fixed-point functions with overflow protection, such as addition and subtraction, multiplication and shift, would take more CPU cycles. This is the main factor that influences the calculation speed. Consequently, the inline functions, which belong to C64 series assembly functions, are used to improve the efficiency. For example, the *L_add*, 32-bit integer addition, can be replaced by the inline function *int_sadd* (*int src1*, *int src2*).

5.4. Performance Analysis. After the assembly-level optimization, the encoder efficiency is greatly improved. The statistical results of AVS-M codec complexity are shown in Table 7.

Because the clock frequency of the C6416 is 600 MHz, it can thus be concluded that the AVS-M codec could be implemented in real-time after optimization on C6416 DSP platform.

6. The Perceived Quality Comparison between the AVS-M and AMR-WB+ [30]

Because of the framework similarity, we compare AVS-M and AMR-WB+. To confirm whether the perceptual quality of AVS-M standard is Better Than (BT), Not Worse Than (NWT), Equivalent (EQU), or Worse Than (WT) that of AMR-WB+, different test situations (bit rate, noise, etc.) are considered and *T*-test method is used to analyze the significance. Test methods comply with the ITU-T MOS test related standards. AVS-M is tested according to the AVS-P10 subjective quality testing specification [31]. The basic testing information is shown in Table 8.

ACR (Absolute Category Rating)—MOS; DCR (Degradation Category Rating)—DMOS.

The score category descriptions are given in Tables 9 and 10.

T-test threshold values are shown in Table 11.

Codec of AVS P10 (AVS-P10_RM20090630) and AMR WB+ (3GPP TS 26.304 version 6.4.0 Release 6) are selected as the test objects.

The reference conditions are as follows in Table 12.

TABLE 7: The AVS-M codec complexity comparison of before and after optimization.

codec	Channel type	Bit rate (kbps)	The total cycles (M/S) before optimization	the total cycles (M/S) after optimization
encoder	Mono	12	1968.138	362.538
decoder	Mono	12	639.366	81.256
encoder	Stereo	24	3631.839	513.689
Decoder	Stereo	24	869.398	86.163

TABLE 8: Basic testing information.

number	Method	Experimental content	Tested Codec @bit rate	Reference codec @bit rate
(1) 1a	ACR	Pure speech, mono, 16 kHz sampling	AVS-P10@10.4, 16.8, 24 kbps	AMR-WB+@10.4, 16.8, 24 kbps
(2) 2a, 2b	ACR	Pure audio, mono, 22.05 kHz sampling	AVS-P10@10.4, 16.8, 24 kbps	AMR-WB+@10.4, 16.8, 24 kbps
		Pure audio, stereo, 48 kHz sampling	AVS-P10@12.4, 24, 32 kbps	AMR-WB+@12.4, 24, 32 kbps
(3) 3a, 3b	DCR	Noised speech, mono, 16 kHz Sampling (office_noise, SNR = 20 dB)	AVS-P10@10.4, 16.8, 24 kbps	AMR-WB+@10.4, 16.8, 24 kbps
		Noised speech, mono, 16 kHz sampling (street_noise, SNR = 20 dB)		

TABLE 9: MOS score category description-ACR test.

MOS	5	4	3	2	1
Overall quality description	Excellent	Good	Common	Bad	Very bad

6.1. Test Result

6.1.1. MOS Test. In Figures 11, 12, and 13, the scoring of MNRU and direct factor trends are correct, which indicates that the results are reliable and effective. And based on Figures 11, 12 and 13, the conclusion could be drawn that, for the 16 kHz pure speech, 22.05 kHz mono audio, and 48 kHz stereo audio, AVS-M has comparable quality with AMR-WB+ at the three different bit rates. In other words, AVS-M is NWT AMR WB+.

6.1.2. DMOS Test. In Figures 14 and 15, the scoring of MNRU and direct factor trends correct, which suggests the results are true and effective. And from Figure 14, the conclusion could be drawn that, for the 16 kHz office_noise speech, the AVS-M has the fairly quality with AMR-WB+ (AVS-M NWT WB+) at 16.8 kbps and 24 kbps bit rate, but the quality of AVS-M is worse than that of AMR-WB+ at 10.4 kbps bit rate. From Figure 15, the conclusion could be drawn that, for the 16 kHz street noise samples, the AVS-M has the fairly quality with AMR-WB+ (AVS-M NWT WB+) at the three different bit rates. Especially at the bit rate of

24 kbps, the scoring of AVS-M is little better than that of AMR-WB+.

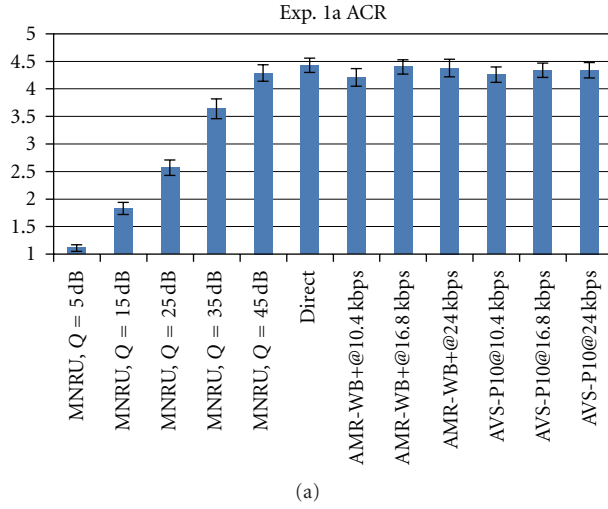
Based on the statistic analysis, AVS-M is slightly better than (or equivalent to) AMR-WB+ at high bit rate in each experiment. But at low bit rate, the AVS-M is slightly better for 1a and 2b, and AMR-WB+ is slightly better for 2a, 3a and 3b.

In terms of *T*-test, except 10.4 kbps condition, the performance of AVS-M is not worse than that of AMR-WB+ in all of the other tests.

7. Features and Applications

AVS-M mobile audio standard adopts the advanced ACELP/TCX hybrid coding framework, and the audio redundancy is removed by advanced digital signal processing technology. Therefore, high compression ratio together with high-quality sound could be achieved with the maximum system bandwidth savings.

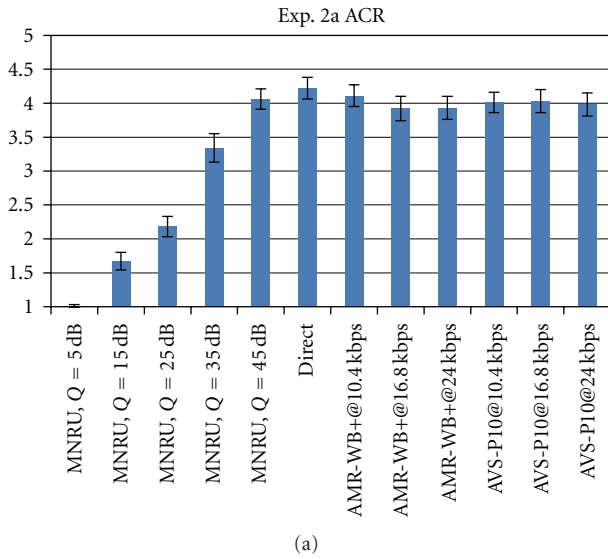
In AVS-M audio coding standard, the adaptive variable rate coding of the source signal is supported and the bit rate ranging from 8 kbps to 48 kbps could be adjusted continuously. As for differential acceptable error rates, the bit rate can be switched for each frame. By adjusting coding rate and acceptable error rates according to the current network traffic and the quality of communication channel, the best coding mode and the best channel mode could be chosen. So, the best combination of coding quality and system capacity could be achieved. Overall, AVS-M audio standard is with great flexibility and can support adaptive transmission of audio data in the network.



Bit rate	M-D	NWT T-V	Result
10.4 kbps	-0.052083333	0.156289216	Pass
16.8 kbps	0.052083333	0.123896549	Pass
24 kbps	0.052083333	0.186793279	Pass
Bite rate	M-D	BT T-V	Result
10.4 kbps	-0.052083333	-0.156289216	Fail
16.8 kbps	0.052083333	-0.123896549	Fail
24 kbps	0.03125	-0.146463032	Fail
Bite rate	M-D (abs)	EQU T-V	Result
10.4 kbps	0.052083333	0.135904281	Pass
16.8 kbps	0.052083333	0.148078308	Pass
24 kbps	0.03125	0.17504925	Pass

(b)

FIGURE 11: Experiment 1a MOS scores statistic analysis result and T -test result, M-D: mean difference; T-V: T -test value.



Bit rate	M-D	NWT t-v	result
10.4 kbps	0.104166667	0.178194471	Pass
16.8 kbps	-0.114583333	0.207076749	Pass
24 kbps	-0.052083333	0.203391541	pass
Bite rate	M-D	BT t-v	result
10.4 kbps	0.104166667	-0.178194471	fail
16.8 kbps	-0.114583333	-0.207076749	fail
24 kbps	-0.052083333	-0.203391541	fail
Bite rate	M-D (abs)	EQU t-v	result
10.4 kbps	0.104166667	0.212973936	pass
16.8 kbps	0.114583333	0.247493371	pass
24 kbps	0.052083333	0.243088896	pass

(b)

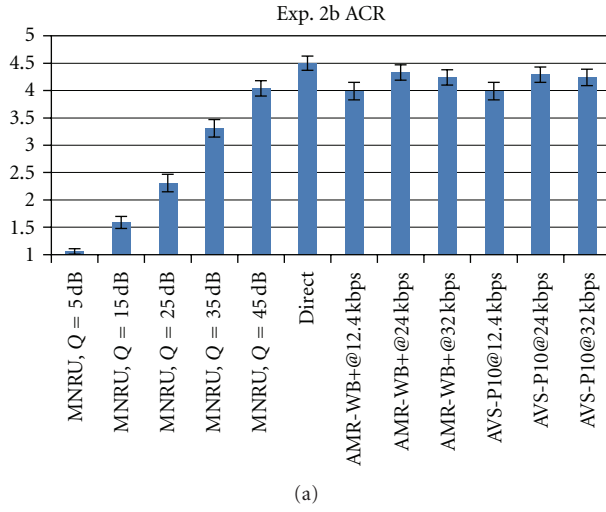
FIGURE 12: Experiment 2a MOS scores statistic analysis result and T -test result.

AVS-M audio standard adopts a powerful error protect technology. The error sensitivity of the compressed streams could be minimized by the optimization of robustness and error recovery technique. In AVS-M audio standard, the non-uniform distribution for the error protection information is supported and the key objects are protected more. So, the maximum error probability of key objects could also be curtailed in the case of poor network quality.

Because of high compression, flexible coding features, and the powerful error protection, the AVS-M audio coding standard could meet the demand of mobile multimedia services, such as Mobile TV [32, 33].

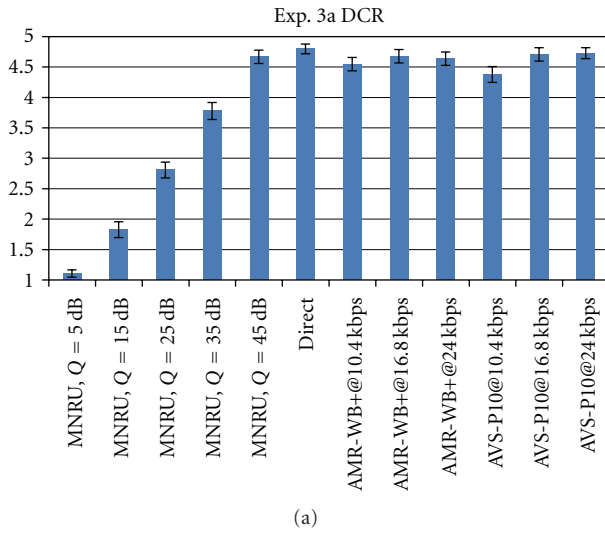
8. Conclusion

As the mobile audio coding standard developed by China independently, the central objective of AVS-M audio standard is to meet the requirements of new compelling and commercially interesting applications of streaming, messaging and broadcasting services using audio media in the third generation mobile communication systems. Another objective is to achieve a lower license cost that would provide equipment manufacturers more choices over technologies and lower burden of equipment cost [34]. AVS has been supported by the relevant state departments and AVS



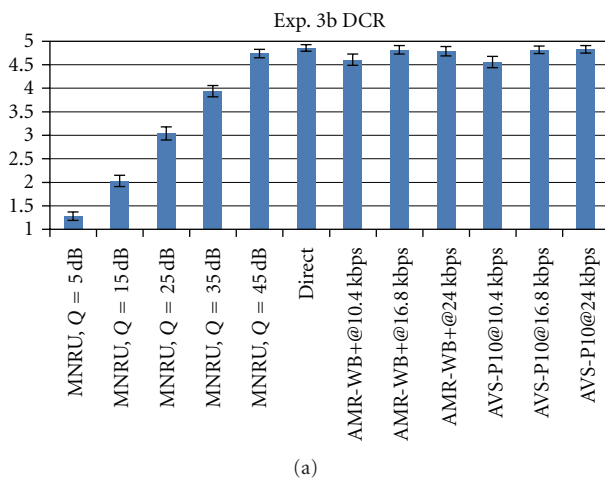
Bit rate	M-D	NWT t-v	result
12.4 kbps	0	0.198315856	pass
24 kbps	0.041666667	0.163010449	pass
32 kbps	0	0.172186382	pass
Bite rate	M-D	BT t-v	result
12.4 kbps	0	-0.198315856	fail
24 kbps	0.041666667	-0.163010449	fail
32 kbps	0	-0.172186382	fail
Bite rate	M-D (abs)	EQU t-v	result
12.4 kbps	0	0.237022553	pass
24 bps	0.041666667	0.194826342	pass
32 bps	0	0.205793206	pass

(b)

FIGURE 13: Experiment 2b MOS scores statistic analysis result and T -test result.

Bit rate	M-D	NWT t-v	result
10.4 kbps	0.177083333	0.120492288	fail
16.8 kbps	-0.03125	0.096696091	pass
24 kbps	-0.09375	0.104586939	pass
Bite rate	M-D	BT t-v	result
10.4 kbps	0.177083333	-0.120492288	fail
16.8 kbps	-0.03125	-0.096696091	fail
24 kbps	-0.09375	-0.104586939	fail
Bite rate	M-D (abs)	EQU t-v	result
10.4 kbps	0.177083333	0.144009613	fail
16.8 kbps	0.03125	0.115568946	pass
24 kbps	0.09375	0.124999906	pass

(b)

FIGURE 14: Experiment 3a DMOS scores statistic analysis result and T -test result.

Bit rate	M-D	NWT t-v	result
10.4 kbps	0.052083333	0.113710588	pass
16.8 kbps	0	0.088689547	pass
24 kbps	-0.041666667	0.101171492	pass
Bite rate	M-D	BT t-v	result
10.4 kbps	0.052083333	-0.113710588	fail
16.8 kbps	0	-0.088689547	fail
24 kbps	-0.041666667	-0.101171492	fail
Bite rate	M-D (abs)	EQU t-v	result
10.4 kbps	0.052083333	0.135904281	pass
16.8 kbps	0	0.105999708	pass
24 kbps	0.041666667	0.120917842	pass

(b)

FIGURE 15: Experiment 3b DMOS scores statistic analysis result and T -test result.

TABLE 10: DMOS score category description-DCR test.

DMOS	5	4	3	2	1
Description of the quality damage	Not aware	Just have noticed but not loathe	Noticed and little loathe	Obviously aware and loathe but tolerable	Intolerable

TABLE 11: T value.

	NWT test	BT test	EQU test
T -value	1.661051818	1.661051818	1.985250956

TABLE 12: Reference condition.

conditions	number	Description
Direct	1	Original 16 k Hz speech and 22.5 kHz of 48 k Hz audio
MNURU	5	$Q = 5, 15, 25, 35, 45$ dB

MNURU: modulated Noise Reference Unit.

industry alliance. Hence, it is foreseeable that AVS will greatly promote the industrialization of electronic information.

Acknowledgments

The authors thank AVS subgroup for their enormous contribution to the development of AVS-M audio standard. The authors thank Jialin He, Bo Dong, and Zhen Gao for their proofreading.

References

- [1] H. Zhang, *Research of AVS-M audio codec and implementation on DSP*, M.S. dissertation, Tianjin University, May 2008.
- [2] AVS, M 1740, "The Feasible Research Report of AVS-M Audio Standard," December 2005.
- [3] AVS, P 10, "Mobile Speech and Audio," July 2010.
- [4] 3GPP, TS 26.190, "Adaptive Multi-Rate-Wideband (AMR-WB) speech codec: transcoding functions," July 2005.
- [5] 3GPP, TS 26.290, "Extended Adaptive Multi-Rate-Wideband (AMR-WB+) Codec; Transcoding Functions," June 2006.
- [6] AVS, M 1751, "The Basic Framework of AVS-M Audio Standard," March 2006.
- [7] M. Neuendorf, P. Gournay, M. Multrus et al., "Unified speech and audio coding scheme for high quality at low bitrates," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '09)*, pp. 1–4, April 2009.
- [8] A. Gersho, "Advances in speech and audio compression," *Proceedings of the IEEE*, vol. 82, no. 6, pp. 900–918, 1994.
- [9] ITU-T, G.729, "Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP)," Geneva, Switzerland, March 1996.
- [10] W. S. Chung, S. W. Kang, H. S. Sung, J. W. Kim, and S. I. Choi, "Design of a variable rate algorithm for the 8 kb/s CS-ACELP coder," in *Proceedings of the 48th IEEE Vehicular Technology Conference (VTC '98)*, vol. 3, pp. 2378–2382, May 1998.
- [11] R. P. Ramachandran, M. M. Sondhi, N. Seshadri, and B. S. Atal, "A two codebook format for robust quantization of line spectral frequencies," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 3, pp. 157–168, 1995.
- [12] T. Zhang and W. Zhang, "Performance analysis and evaluation of AVS-M audio coding," in *Proceedings of International Conference on Audio, Language and Image Processing*, 2010.
- [13] AVS, M 1865, "An approach of vector quantization for linear spectral frequencies," September 2006.
- [14] AVS, M 2052, "An improvement of the perceptual weighting proposal," June 2007.
- [15] AVS, M 1922, "An approach of wideband adaptive codebook excitation search," December 2006.
- [16] AVS, M 1964, "An approach of track-first algebra codebook search," January 2007.
- [17] 3GPP, TS 26.304, "Extended Adaptive Multi-Rate-Wideband (AMR-WB+) Codec: Float-Point ANSI-C Code," March 2006.
- [18] J. Zhan, K. Choo, and E. Oh, "Bandwidth extension for China AVS-M standard," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '09)*, pp. 4149–4152, April 2009.
- [19] AVS, M 1918, "A high-effective configurable parametric stereo coding algorithm in frequency domain," December 2006.
- [20] AVS, M 1954, "The VAD requirement analysis in AVS-M Audio," December 2006.
- [21] AVS, M 1971, "Digital Audio Sampling Rate Conversion Filter," January 2007.
- [22] T. Zhang, H. Zhang, D. Yang, and J. He, "New window function," *Journal of Information and Computational Science*, vol. 5, no. 4, pp. 1923–1928, 2008.
- [23] TMS320C6416, "Fixed-Point Digital Signal Processors," 2005.
- [24] W. Yuan, "The realization of G.723.1 on TMS320VC5402," *Journal of University of Electronic Science and Technology of China*, pp. 50–53, 2002.
- [25] 3GPP, TS 26.273, "Fixed-point ANSI-C code," March 2006.
- [26] 3GPP, TR 26.936, "Performance characterization of 3GPP audio codec," March 2006.
- [27] 3GPP, TR 26.936, "Performance characterization of 3Gpp audio codec," March 2006.
- [28] TMS320C6000, "Optimizing Compiler User's Guide," 2004.
- [29] Texas Instrument, "TMS320C6000 programmer's guide," 2002.
- [30] AVS, "P 10 BIT-RICOT Test Report," January 2010.
- [31] AVS, M 2608, "AVS-P10 subjective quality testing program," September 2009.
- [32] C. Zhang and R. Hu, "A novel audio codec for mobile multimedia applications," in *Proceedings of International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '07)*, pp. 2873–2876, September 2007.
- [33] R. Hu and Y. Zhang, "Research and application on AVS-M audio standard," *Communication Electroacoustics*, vol. 31, no. 7, 2007.
- [34] AVS, M 1753, "The Analysis of AVS-M Mobile Audio Entering into 3G," March 2006.