# Audio Coding Algorithm for One-Segment Broadcasting

● Masanao Suzuki ● Yasuji Ota ● Takashi Itoh

*(Manuscript received November 29, 2007)*

**With the recent progress in coding technologies, a more efficient compression of video and audio data has become available, resulting in popularization of new services such as one-segment broadcasting and content delivery. In 2003, ISO/IEC developed the latest MPEG audio standard called High-Efficiency Advanced Audio Coding (HE-AAC). Based on the conventional Advanced Audio Coding (AAC) algorithm, HE-AAC features a higher compression ratio than that of AAC. Because the total bit rate of one-segment broadcasting is as low as about 400 kb/s including video, audio, and other data, the bit rate for audio must be reduced as much as reasonably possible in order to increase the bit rate for video. Fujitsu has developed a new HE-AAC encoding algorithm that provides high sound quality at a lower bit rate of 32 kb/s while the bit rate of the current one-segment broadcasting is 48 kb/s. This paper describes the technology to improve the sound quality of the HE-AAC encoder and demonstrates its advantages based on the results of subjective listening tests.**

## 1. Introduction

One-segment broadcasting services for cell phones and contents delivery services on IP networks have become popular in recent years. This trend is attributable mainly to the advance of audio coding technologies and video coding technologies, the increase of the capacity of data storage media, and acceleration of the network speed.

The Moving Picture Experts Group (MPEG), a multimedia expert group in the International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) is an organization involved in the standardization of audio coding algorithms (**Figure 1**). In 1993, MPEG established the MPEG-1 audio standard, with which CD-quality sound can be realized at 192 kb/s. In 1997, MPEG-2 Advanced Audio Coding (AAC)[1] was standardized, which can provide transparent sound quality at 128 kb/s. Further, in 2003 High-Efficiency AAC (HE-AAC) was established as a standard with an enhanced compression ratio through the expansion of AAC.[2] HE-AAC has been adopted in the Japanese one-segment broadcasting standard[3] and the multi-media service standard[4,5] for the 3rd Generation Partnership Project (3GPP) for mobile applications.

While the MPEG standard defines the bitstream format and decoding algorithm, it does not define the encoding algorithm. This leads to varying sound quality depending on the encoding method. Because high sound quality is requested at a bit rate lower than 48 kb/s in one-segment broadcasting, it is essential to develop a new technology to improve the encoded sound quality.

In this paper, we explain an outline of the HE-AAC algorithm and the technology to improve the sound quality developed by Fujitsu. We also describe the results of a subjective listening test on the encoded sound quality.

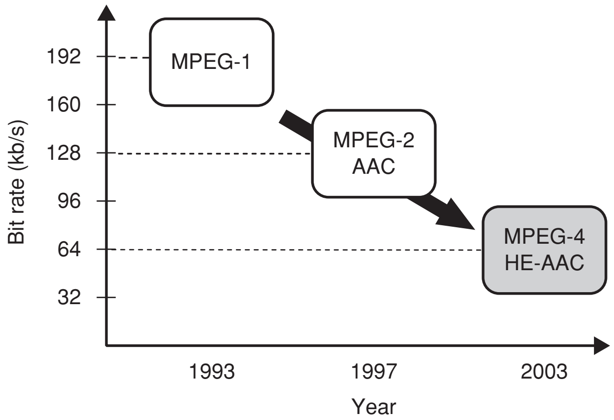



Figure 1
History of MPEG audio standardization.

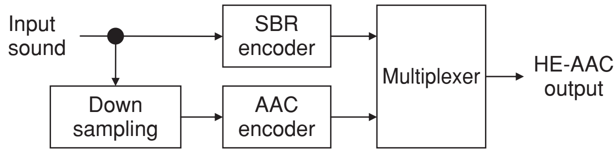



Figure 2
Block diagram of HE-AAC encoder.

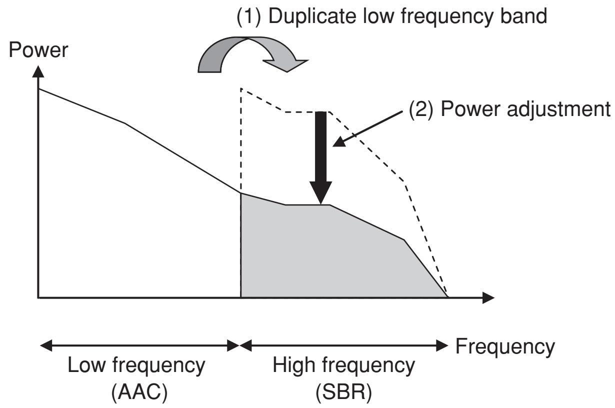



Figure 3
Basic concept of SBR algorithm.

## 2. Outline of HE-AAC algorithm

The configuration of the HE-AAC encoder is shown in **Figure 2**. In HE-AAC, low frequency components of the input sound are coded by the conventional AAC and high frequency components are coded by Spectral Band Replication (SBR). SBR can realize high sound quality at a lower bit rate than AAC by use of the high correlation between high frequency components and low frequency components. As shown in **Figure 3**, SBR duplicates the low frequency components that have a high correlation with the high frequency components and then finely adjusts the power of the duplicated high frequency components. The portion of the high frequency components that cannot be duplicated from the low frequency components is coded as additional information. As a consequence, SBR enables coding at a lower bit rate than AAC. The results of a listening test which was carried out by ISO/IEC revealed that HE-AAC can reduce the bit rate by about 25% compared with that of AAC.[6] The details of the HE-AAC encoder are explained below.

1) AAC encoder section

The HE-AAC encoder performs coding of the low frequency components of the input sound with the conventional AAC encoder. The configuration of the AAC encoder is shown in **Figure 4**. The AAC encoder transforms the input sound into a frequency spectrum by using a Modified Discrete Cosine Transform (MDCT) followed by quantization. Also, based on a psychoacoustic analysis, the perceptual importance of each frequency is determined so that a larger number of quantization bits can be allocated to more important frequencies and the number of quantization bits for less important frequencies can be reduced to enhance the coding efficiency. Further, to improve the coding efficiency, optional coding tools are available such as a block switching tool, a Temporal Noise Shaping (TNS) tool, and a Mid/Side (MS) stereo tool.

2) SBR encoder section

In the HE-AAC encoder, the high frequency components of the input sound are coded with an SBR encoder. The configuration of the SBR encoder is shown in **Figure 5**. SBR transforms the input sound into a time-frequency spectrum via the analysis filterbank. A grid generator controls the quantizer so that the time resolu-

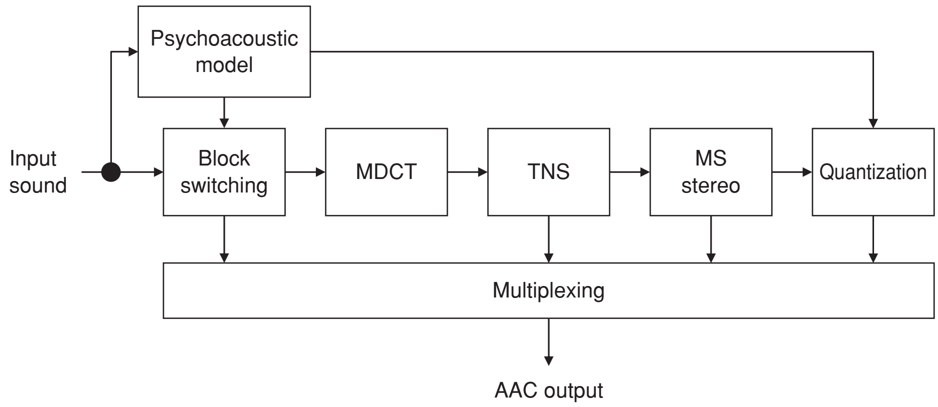



Figure 4
Block diagram of AAC encoder.

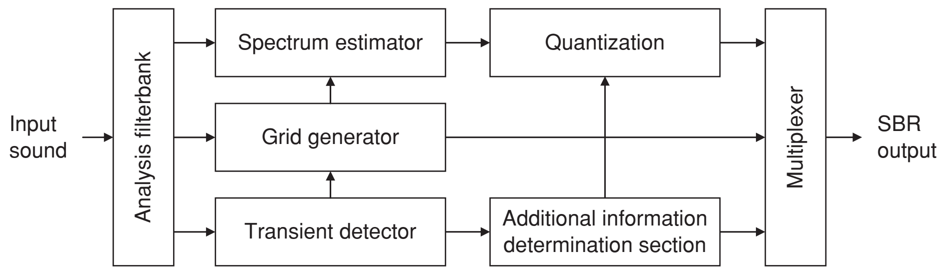



Figure 5
Block diagram of SBR encoder.

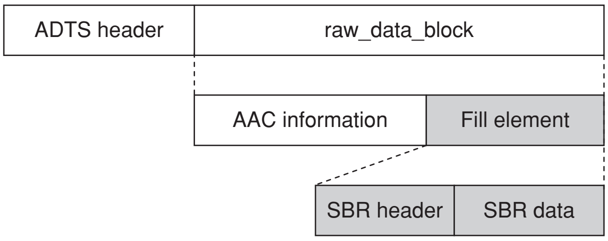
tion is higher when the input sound is transient, while the frequency resolution is higher when it is stationary. The property of the input sound is judged in the transient detector. While quantizing the power level which is output from the spectrum estimator, the quantizer duplicates the high frequency components from the low frequency components, and outputs only a small amount of control information to adjust the power. For components that cannot be covered by duplication from the low frequency components, a small amount of additional information is output from the additional information determination section. Based on this method, SBR can carry out the coding with a lower bit rate than AAC which quantizes the whole frequency.

3) Multiplexer

The HE-AAC data format is shown in **Figure 6**. HE-AAC adopts the same data format as AAC, where the information for the low frequency components (AAC output) is accommodated in raw_data_block. The information for SBR output (SBR header and data) which represents the high frequency components is accommodated in the field for containing extended information (Fill element). Because HE-AAC is a forward-compatible type of AAC, an HE-AAC decoder can decode not only the HE-AAC bitstreams but also the AAC bitstreams. While a conventional AAC decoder can decode HE-AAC bitstreams, the scope of the decoding is limited to the low frequency components (AAC information).

## 3. Challenges in commercialization

While the ISO/IEC HE-AAC standard defines the data format and decoder algorithm,




ADTS: Audio Data Transport Stream

Figure 6
Bitstream format of HE-AAC.

it does not define the encoder algorithm. For this reason, an improvement in the sound quality is essential for the commercialization of the HE-AAC encoder.

Because high frequency components are coded by duplicating the low frequency components coded by AAC in HE-AAC, the sound quality of AAC is dominant in the sound quality of HE-AAC as a whole. Accordingly, it is necessary to improve the sound quality of both SBR and AAC in order to improve the HE-AAC sound quality.

One-segment broadcasting service adopts HE-AAC with a bit rate of 48 kb/s as its audio coding algorithm. Since the total bit rate of one-segment broadcasting is as low as about 400 kb/s including video, audio, and other data, the bit rate of audio should be reduced to lower than 48 kb/s (e.g. 32 kb/s) to increase the bit rate of video in order to enhance the video quality. However, if the bit rate of HE-AAC is reduced to 32 kb/s, the sound quality, especially of human speech, often deteriorates because of quantization noise. Because it is used mainly for speech programs such as news, drama, and live sports broadcasting, degraded speech quality may tire the audience by making it hard for them to listen to the broadcast.

## 4. Challenges in sound quality

The main challenges of HE-AAC at a lower bit rate can be addressed by solving the three types of sound quality deterioration as outlined below:

1) Sound quality deterioration in stereo signal

When coding stereo signals, the sound quality between the right and left channels are inconsistent if the level of quantization noise is different between both channels. The discrepancy in the sound quality between both channels is large, particularly at low bit rates, because of the increased quantization noise. If there is a difference in sound quality between the right and left channels, one person's voice is heard as the voices of two people when listening to news or dramas, since the sounds of both channels are almost the same in these types of programs.

2) Sound quality deterioration in speech

The SBR frequency band is comprised of two types: i.e. high resolution and low resolution. The low resolution has a wider frequency range than the high resolution. Accordingly, as in the case of speech, if signals with decreasing power and increasing frequency are coded at low resolution, the power for the high frequency band tends to be larger than that before the coding. Especially, when coding the consonants such as /s/, /th/, /z/ sounds, the power tends to concentrate on the high frequency range. Because of this, an unnatural emphasis of the high frequency band occurs when consonants are coded with low resolution.

3) Sound quality deterioration at insufficient bit rate

Although the average bit rate of SBR is about 3 to 4 kb/s, it may go up to 10 kb/s or more instantaneously. As a consequence, AAC needs to be coded at around 20 kb/s when HE-AAC is used with the bit rate of 32 kb/s, resulting in a deterioration of the sound quality in the low frequency band because of an insufficient bit rate. This also leads to compromised sound quality in the high frequency band that is duplicated by SBR.

## 5. Development of technology

By addressing the above-mentioned challenges, we have developed a new HE-AAC

Table 1
HE-AAC encoder specification.

| | |
|---|---|
| Coding algorithm | ISO/IEC 13818-7 (MPEG-2 HE-AAC) |
| Profile | LC profile |
| Input parameter | 48 kHz, stereo |
| Bit rate | 32 kb/s |
| Audio bandwidth | 15 kHz |

encoding algorithm that provides high sound quality equivalent to that of FM broadcasting (15 kHz bandwidth) at a lower bit rate of 32 kb/s. The specification of the developed algorithm is shown in **Table 1**. The output data format of this algorithm complies with the MPEG-2 HE-AAC standard. The technologies to improve the sound quality of the HE-AAC encoder will be described in the following paragraphs:

1) Sound quality improvement in stereo signals

We have developed technology to allocate quantization bits so that the quantization noise between the right and left channels becomes equal for perceptually important frequencies by analyzing the property of the input sound at the AAC encoder. Further, by compensating the bit allocation and signal power based on the correlation between both channels, the sound deterioration problem that frequently occurs in stereo speech can be eliminated.

2) Sound quality improvement in speech

Our second technology divides the frequency band to be coded by SBR into a low frequency band and a high frequency band and then compensate the power for each band. If the power difference between the low frequency band and high frequency band exceeds a threshold, the compensation is carried out so as to reduce the power of all the bands coded by SBR. An example of the time-frequency characteristics of a Japanese consonant is shown in **Figure 7** (it represents the *su* sound of the phrase *shite imasu* in Japanese). The horizontal axis represents time and the vertical axis represents frequency. The darkness of the color indicates the level of power, where paler colors indicate larger power. Before

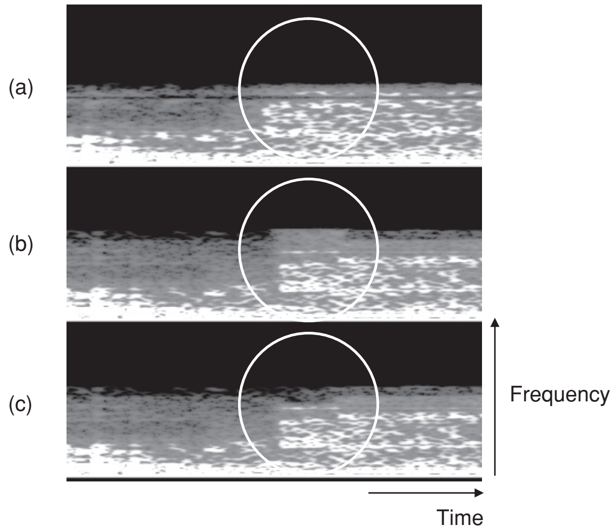



Figure 7
Comparison of spectrum between (a) original speech (not compressed), (b) coded speech without the proposed method, and (c) coded speech with the proposed method.

the compensation [Figure 7 (b)], the power of the high frequency band is larger in comparison with the original speech [Figure 7 (a)], causing an unnatural quality of certain sounds. On the other hand, the figure after the compensation [Figure 7 (c)] indicates a smaller power in the high frequency band versus the situation before the compensation, resulting in more natural sounds. Based on this method, the unnatural emphasis of consonants was reduced and a natural sound quality was achieved.

3) Method to limit SBR bit rate

The third technology we have developed controls the maximum bit rate of SBR for the purpose of ensuring the minimum bit rate allocation to AAC. Because SBR performs differential coding of the power level for the time-frequency spectrum, the bit rate tends to be lower as the difference of power between the adjacent bands (time direction/frequency direction) becomes smaller. Using this property, the maximum bit rate for SBR is controlled to ensure the minimum bit rate assignments to AAC by compensating the power, in a way that the differential power at high frequency spectrum becomes smaller if the bit rate of SBR exceeds a predetermined upper limit.

## 6. Results of sound quality evaluation

The sound quality of the HE-AAC encoder we developed was evaluated using the MUlti Stimulus test with Hidden Reference and Anchor[7] (MUSHRA) method. The MUSHRA method is a subjective listening test in which sound quality is evaluated based on a 101-point scale by listening and comparing the uncompressed original signals and the multiple coded audio signals. In this system, a rating from 0 to 100 is given to the sound quality of each coded audio signal, where 0 means "bad" and 100 means "excellent". Ten types of sound sources are used for the evaluation. They are castanets, harpsichord, pitch pipe, glockenspiel, speech, singing, accordion, and three types of pop music. The sampling frequency was 48 kHz (stereo). The sound quality was evaluated by 24 music experts (music college students).

The bit rate of the developed HE-AAC encoder is 32 kb/s. For comparison purposes, the sound quality of a conventional AAC encoder with 56 kb/s and 64 kb/s was also evaluated. The coding algorithm of the AAC encoder for comparison is the same as the algorithm of the AAC encoder section used in the HE-AAC encoder which we developed. The only difference between them is the maximum frequency, i.e. 5.6 kHz in the AAC encoder for HE-AAC, 13 kHz in the conventional AAC encoder for comparison.

The evaluation results are shown in **Figure 8**. The bar chart in Figure 8 indicates the average evaluation points, where the error bar indicates the 95% confidence interval. If any part of the confidence interval of these two conditions overlaps, it means that with 95% probability there is no difference in the sound quality between them. These evaluation results revealed that the 32 kb/s HE-AAC encoder we developed has a higher sound quality than the

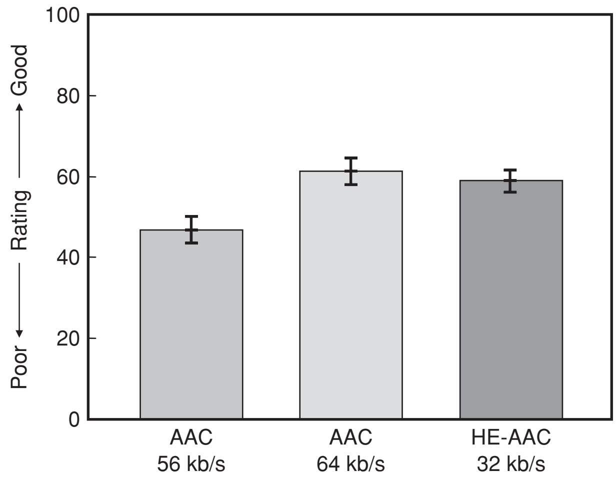



Figure 8
Listening test result for HE-AAC encoder compared with AAC encoder.

56 kb/s AAC. Further, it was demonstrated that there is no statistically significant difference in the sound quality between our HE-AAC encoder and a 64 kb/s AAC, indicating that they have almost equivalent sound quality.

## 7. Conclusion

Fujitsu has developed a new HE-AAC encoding algorithm that provides a high sound quality at a lower bit rate of 32 kb/s versus the bit rate of the current one-segment broadcasting (48 kb/s). This algorithm is comprised of an AAC encoder and an SBR encoder. By improving the AAC encoder for coding of low frequency bands, the issue of the sound deterioration problem that frequently occurred in stereo speech signals was eliminated. Further, by improving the SBR encoder for coding the high frequency band, the quality of speech (specifically consonants) was improved. Also, by controlling the maximum bit rate of SBR, the bit rate necessary for AAC was ensured, thus achieving a high sound quality for HE-AAC as a whole. Based on a subjective listening test, it was demonstrated that the new HE-AAC algorithm with a lower bit rate of 32 kb/s realizes an excellent sound quality which is equivalent to a 64 kb/s AAC.

## References


1) ISO/IEC 13818-7: Information technology - Generic coding of moving pictures and associated audio information - Part 7: Advanced Audio Coding (AAC). 2006.
2) ISO/IEC 14496-3: Information technology - Coding of audio-visual objects - Part 3: Audio. 2005.
3) ARIB STD-B32 version 2.1: Video coding, audio coding, and multiplexing specifications for digital broadcasting. 2007.
4) 3GPP TS 26.403 V6.0.0: Enhanced aacPlus general audio codec; Encoder specification AAC part. 2004.
5) 3GPP TS 26.404 V6.0.0: Enhanced aacPlus general audio codec; Enhanced aacPlus encoder SBR part. 2004.
6) ISO/IEC JTC 1/SC 29/WG 11, N6009: Report on the Verification Tests of MPEG-4 High Efficiency AAC. 2003.
7) ITU-R Recommendation BS.1534-1: Method for the subjective assessment of intermediate quality level of coding systems. 2001-2003.


**Masanao Suzuki**
*Fujitsu Laboratories Ltd.*
Mr. Suzuki received the B.S. and M.S. degrees in Electronic Engineering from Hokkaido University, Sapporo, Japan in 1991 and 1993, respectively. He joined Fujitsu Laboratories Ltd., Kawasaki, Japan in 1993, where he has been engaged in research and development of audio coding algorithms, speech coding algorithms and speech communication systems. He received the Kanto District Patent Award in 2006 and the OHM Technology Award from the Promotion Foundation for Electrical Science and Engineering in 2007.

**Yasuji Ota**
*Fujitsu Laboratories Ltd.*
Mr. Ota received the B.S. degree in Electronic Engineering from Niigata University, Niigata, Japan, in 1987. He joined Fujitsu Laboratories Ltd., Kawasaki, Japan in 1987 and has been engaged in research and development of speech coding algorithms and speech communication systems. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE), Acoustic Society of Japan (ASJ) and IEEE.

**Takashi Itoh**
*Fujitsu Laboratories Ltd.*
Mr. Itoh received the B.E. degree in Chemical Engineering and the M.E. degree in Information Engineering from the University of Tokyo, Tokyo, Japan in 1981 and 1983, respectively. He joined Fujitsu Laboratories Ltd., Kawasaki, Japan in 1983, where he has been engaged in research and development of algorithms, LSIs, and codecs for visual signal coding and the research and development of visual communication systems. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan, the Information Processing Society of Japan, and the Institute of Electrical Engineers of Japan.