Extended AMR-WB for High-Quality Audio on Mobile Devices

Redwan Salami and Roch Lefebvre, VoiceAge
Ari Lakaniemi and Kalervo Kontola, Nokia
Stefan Bruhn and Anisse Taleb, Ericsson

Extended AMR-WB for High-Quality Audio on Mobile Devices

Redwan Salami and Roch Lefebvre, VoiceAge
Ari Lakaniemi and Kalervo Kontola, Nokia
Stefan Bruhn and Anisse Taleb, Ericsson

ABSTRACT

This article presents the architecture, performance, and application scenarios of the AMR-WB+ (Extended AMR-WB) audio codec, which provides high quality at exceptionally low rates, and consistent quality over all audio types. This codec was recently selected by 3GPP and DVB to support low-bit-rate audio and audiovisual applications on mobile networks.

INTRODUCTION

Recent advances in mobile networking and device technologies have enabled the introduction of innovative multimedia services delivered to wireless devices. Market evidence from around the world clearly demonstrates the growing popularity of mobile multimedia services such as streaming, downloading, and uploading of audio and audiovisual content.

Deploying mobile multimedia services imposes tough requirements on the network to ensure a consistent and satisfactory end-user experience. The Third-Generation Partnership Project (3GPP) has defined new mobile multimedia services over 2.5G (GPRS) and 3G (UMTS) networks, which exploit the enhanced data capacity of these evolved networks. The underlying standards specify packet-switched streaming service (PSS) [3GPP TS 22.233], multimedia messaging service (MMS) [3GPP TS 22.140], multimedia broadcast/multicast service (MBMS) [3GPP TS 22.246], and IP multimedia subsystem (IMS) messaging [3GPP TS 22.340] and presence service [3GPP TS 22.141].[1] Between these services are also common elements such as media types and media formats. Content can be a combination of different media, including audio and video clips, graphics, images and text.

MMS enables sending and receiving multimedia messages between mobile terminals as well as from a content server to a mobile terminal and vice versa. MMS is showing growing momentum as adoption of multimedia-capable terminals increases and network interoperability improves. PSS is a framework for point-to-point multimedia streaming services. It uses Real-time Transport Protocol (RTP) to transport media streams such as audio. MBMS is a point-to-multipoint service in which the same data is transmitted from a single source entity to multiple recipients. Two delivery methods have been specified for MBMS: RTP-based streaming and downloading that uses the File Delivery over Unidirectional Transport (FLUTE) protocol with an associated post-delivery method.

Audio content relevant for these services includes music, speech, and speech mixed with diverse audio types as in sportscasts, news, and movies. Further, the bit rates available for these services (especially on 2.5G networks) may be as low as 10 kb/s. Therefore, efficient audio coding at very low bit rates with consistent performance over different content types is of prime importance for the success of these services. Consequently, 3GPP has standardized audio codecs in order to increase interoperability between networks and mobile terminals.

When 3GPP Release 5 was specified, no single codec existed that performed suitably well at the same time for speech and music at the required low bit rates. The reason is that traditional speech and audio codecs belong to different families of encoding algorithms, each with well-known strengths and weaknesses. Therefore, in 3GPP Release 5, the AMR [3GPP TS 26.071] and AMR-WB [3GPP TS 26.171] codecs were selected to be mandatory for narrowband and wideband speech services, respectively, and low-complexity advanced audio coding (MPEG-4 AAC-LC) [1] was recommended for general audio. AMR-WB delivers high quality for 7 kHz speech signals at bit rates as low as 12.65 kb/s, but does not perform well on audio signals. On the other hand, AAC-LC provides good audio quality at bit rates above 48 kb/s; however, the quality degrades significantly at lower rates, in particular for speech signals. Thus, there was still a need to fill the gap between those codecs.

In December 2002, a new 3GPP Release 6 work item was approved to extend the AMR-WB speech coder with the goal of delivering perceptually high quality for speech, audio, and mixed content (referred to as extended AMR-

[1] *3GPP technical specifications are available for download at www.3gpp.org. We simply provide specification numbers for easy consultation.*

0163-6804/06/$20.00 © 2006 IEEE

WB or AMR-WB+). At the same time, aacPlus (aka MPEG-4 HE AAC) had just been standardized in MPEG and it was suggested for adoption by 3GPP for low-bit-rate audio services replacing AAC-LC. Hence, a process was launched in 3GPP for testing and selecting audio codecs for Release 6 multimedia services. A main requirement was to perform better than Release 5 codecs (AMR-WB and AAC LC) for any content type. The testing was divided in two categories: "low bit rate" for rates at 24 kb/s and below, and "high bit rate" for rates at 32 and 48 kb/s. The candidates for the high-rate selection included aacPlus and Enhanced aacPlus (Eaac+) (aka aacPlus v2). The candidates for the low-rate selection included those two aacPlus versions and AMR-WB+. In both testing categories, the test material used included speech, music, and speech mixed with different audio content to reproduce the possible application scenarios for the targeted bit rates and applications.

In 2004, after extensive selection tests involving eight listening laboratories, 3GPP selected both AMR-WB+ [3GPP TS 26.290] and Eaac+ [3GPP TS 26.401] as recommended audio codecs for Release 6 multimedia services. Both codecs have their merits in certain application scenarios. Specifically, Eaac+ showed good performance for music at high rates, while at low rates AMR-WB+ performed well for music and exhibited better performance for speech and mixed content. AMR-WB+ was later selected by Digital Video Broadcasting (DVB) as an optional codec for both the generic codec toolbox for DVB-compliant delivery in RTP packets over IP networks [2] and in IP datacast (IPDC) over DVB-H service [3].

This article first discusses the service requirements for mobile multimedia services including PSS, MMS, and MBMS. The 3GPP AMR-WB+ audio coding standard recommended for high-quality audio compression at low bit rates is introduced as a necessary ingredient in order to realize these services. The article is organized as follows. The next section describes the service requirements and available audio bit rates for various radio access technologies. The main features of the AMR-WB+ hybrid coding model are presented. The results from subjective listening tests illustrating the performance and strengths of the codec are given. Finally, the last section gives the conclusions.

Service Requirements for Audio in Mobile Multimedia Applications

This section discusses the requirements imposed on the audio codec by various relevant mobile audio use cases, which are defined for 3GPP mobile communications using GPRS or UMTS radio access technology.

The main service requirements are as follows:

- Transmission capacity remains an important constraint when delivering mobile audio content.
- The codecs should support compression of a typical 16-bit stereo PCM audio sampled at 48 kHz to bit rates as low as 10 to 24 kb/s.
- Since the envisioned mobile multimedia services include speech together with other audio content, the codecs should work equally well for them.
- The codecs have to be robust to wireless packet-switched data-link impairments such as frame erasures.
- The decoders in mobile devices must be low complexity, particularly when having simultaneous video and forward error correction (FEC) decoding.

In the following, the aspects of typical audio contents and available bit rates depending on the use case and transport mechanism are outlined.

Relevant Audio Content

Table 1 lists both the relevant and not-applicable use cases for mobile audio/audiovisual media distribution covered by the service requirement specifications for end-to-end PSS, MBMS, and MMS user services, and MMS in 3GPP systems. The table provides also information on the envisioned content for the different use cases and shows that most cases are dominated by speech and speech mixed with other content (e.g., movie soundtrack). Music content distribution is an important exception. All listed cases may comprise audio-only or audiovisual content.

Available Bit Rates Depending on Radio Access Technology

Although the data-carrying capacity of wireless mobile networks is constantly increasing, bandwidth under real-world operating conditions remains a limited resource. Therefore, efficient compression techniques are viable for increasing the number of users that can be simultaneously served in a given network and it also reduces the requirements for content storage space. Table 2 provides examples of the available bit rates for various 3GPP bearers. Depicted are the total bit rates supported by the bearer and the available effective net bit rates for the media composition, excluding the additional overhead for IP transport and protocols.

For PSS and MBMS streaming of audio-only content, GPRS using three time slots provides a maximum bit rate of approximately 24 kb/s, while UMTS can offer net bit rates of up to about 48 kb/s (using a 64 kb/s bearer). However, application-layer FEC mechanisms have been standardized for MBMS streaming, so the available net bit rates may in fact be as low as 18 kb/s (assuming GPRS and about 10 to 15 percent of the total bit rate to be used for FEC). An MMS message with 100 kbytes of audio-only content could include 0.5 min at 24 kb/s or 1 min at 14 kb/s. Correspondingly, a 300-kbyte MBMS message could have 1.5 min of audio at 24 kb/s.

With audiovisual content, the bit rates required for video further reduce the bit rates available for audio. Although the required video bit rate is highly dependent on the content, a reasonable assumption for best-possible audiovisual quality is that video requires about 75 percent of the available bit rate, leaving the remaining 25 percent for audio. Such an assump-

Efficient compression techniques are viable for increasing the number of users that can be simultaneously served in a given network and it also reduces the requirements for content storage space.

AMR-WB+ is an audio codec which contains AMR-WB as a possible coding mode, as well as a novel hybrid technology that combines the strengths of both audio and speech codecs.

Use case	Transport				
	Content	PSS	MMS	MBMS	Download services
Information: news, sports, stock quotes, traffic, weather	Speech, mixed	Yes	Yes	Yes	Yes
Travel guides	Speech, mixed	Yes	Yes	N/A	N/A
M-commerce: online shopping, commercials	Speech, mixed	Yes	Yes	N/A	N/A
Edutainment: training, instructional, corporate presentations	Speech, mixed	Yes	Yes	Yes	Yes
TV, movies	Speech, music, mixed	Yes	Yes	Yes	Yes
Person-to-person MMS	Speech, mixed	N/A	Yes	N/A	N/A
Audio content distribution: audio books	Speech, mixed	Yes	Yes	Yes	Yes
Audio content distribution: music	Music	Yes	Yes	Yes	Yes

Table 1. *Audio/audiovisual media distribution use cases and content types by transport mechanism for mobile media services ("mixed" refers to speech mixed with other content; N/A means not applicable).*

tion leads to the audio net rates for audiovisual content given in Table 2. Clearly, only very low bit rates of 10 to 16 kb/s or (when FEC is used) even lower rates are available for audio in this case. The highest possible rate of about 24 kb/s could be achieved only when using a more expensive 128 kb/s bearer.

Overview of AMR-WB+

The AMR-WB+ audio codec addresses the requirements described above, by using a hybrid coding technology to deliver consistent quality for both speech and music signals at bit rates from 6 to 48 kb/s.

AMR-WB+ is an audio codec which contains AMR-WB as a possible coding mode, as well as a novel hybrid technology that combines the strengths of both audio and speech codecs. A detailed description of the algorithm can be found in 3GPP TS 26.290 [4]. The hybrid technology includes adaptive code excited linear prediction (ACELP) coding to optimally handle speech signals and transform-based coding to effectively represent richer sounds like music. The AMR-WB+ encoder selects the best coding mode (ACELP or transform) on a per-frame basis, thereby providing high-quality across a wide range of sounds with very efficient use of the available service bandwidth. In addition, AMR-WB+ integrates a parametric stereo model to enhance the end-user perception of high-fidelity sound reproduction at remarkably low bit rates. Other key considerations behind AMR-WB+ technology are its built-in robustness to typical network impairments such as packet losses and its adaptability to varying available bit rates.

The AMR-WB+ encoder can accept both mono and stereo signals. The decoder produces a synthesis signal with the same number of channels as the encoder input, although mono output from a decoded stereo signal is also supported. Numerous sampling frequencies are supported by the encoder, from 8 up to 48 kHz. The sampling frequency, and thus the supported audio bandwidth, increases with bit rate. In mono, bit rates from 6 to 36 kb/s are supported. In stereo, the bit rate can range from 8 up to 48 kb/s.

The AMR-WB+ encoder operates at a nominal internal sampling frequency of 25.6 kHz. The audio input is first resampled at the internal sampling frequency, and then split into two equal bands which are critically downsampled to 12.8 kHz. This allows efficient integration of the AMR-WB speech encoder, which operates at a sampling frequency of 12.8 kHz.

Gradual bit rate and bandwidth scaling is realized by varying the internal sampling frequency from 0.5 to 1.5 times the nominal frequency of 25.6 kHz. Hence, the internal sampling frequency of AMR-WB+ is in the range 12.8 to 38.4 kHz. The corresponding audio bandwidth thus ranges from 6.4 kHz (at the lowest bit rates) to 19.2 kHz (at the highest bit rates). Since the frame length (in samples) is kept constant, varying the internal sampling frequency (in kHz) of the encoder changes the frame duration (in ms) in an inversely proportional manner. For example, if the internal sampling frequency is doubled, then the frame duration is reduced by a factor of two.

The encoder processes the input signal in blocks of 2048 samples, independently of the internal sampling frequency. After band splitting and critical downsampling, the lower band and the higher band are thus processed in blocks of 1024 samples. In AMR-WB+, a block of 1024 samples is called a superframe. The superframe

			Audio content		Audiovisual content	
Transport		Radio access technology	Channel bandwidth or message size	Audio (net rate) or content length	Channel bandwidth or message size	Audio (net rate) or content length
Service	PSS	GPRS	36 kb/s	24 kb/s	36 kb/s	< ~ 10 kb/s
		UMTS	64 kb/s	48 kb/s	64 kb/s (128 kb/s)	~ 14 kb/s (~24 kb/s)
	MBMS streaming	GPRS	36 kb/s	< 24 kb/s	36 kb/s	< ~ 10 kb/s
		UMTS	64 kb/s	< 48 kb/s	64 kb/s (128 kb/s)	12-16 kb/s (~24 kb/s)
	MMS	GPRS/UMTS	100 kbytes (audio)	0.5 min @ 24 kb/s or 1 min @ 14 kb/s	75 kbytes (video) + 25 kbytes (audio)	20 s @ 10 kb/s
	MBMS download	GPRS/UMTS	300 kbytes (audio)	1.5 min @ 24 kb/s or 3 min @ 14 kb/s	225 kbytes (video) + 75 kbytes (audio)	60 s @ 10 kb/s

Table 2. *Available audio bit rates or required download time for audio and audiovisual media distribution depending on service and radio access technology.*

in the low band (nominal bandwidth of 0 to 6.4 kHz) is encoded using the hybrid ACELP/TCX (transform coded excitation) model described below. The superframe in the high band (nominal bandwidth of 6.4 to12.8 kHz) is encoded with only 64 bits (per 1024 samples), using a bandwidth extension (BWE) method where only the energy and spectral envelope are transmitted.

Figure 1 shows high-level diagrams of the AMR-WB+ encoder and decoder. When processing a mono input, the encoder and decoder are restricted to the shaded area. In stereo operation, the signals flow through the whole diagrams, with the dashed lines removed. In Fig. 1, the left, right, and mid (or down-mixed from left and right in stereo) signals are labeled *L*, *R*, and *M*, respectively. The low-band signals are indexed *LF* and the high-band signals are indexed *HF*. In Fig. 1a, the preprocessing indicates conditioning filters and resampling to the internal sampling frequency. The analysis filterbank splits the input in the low and high bands. The downmixing produces mid and side channels from the left and right channels of a stereo input. The low band of the mid (or mono) signal is encoded using the *core* ACELP/TCX model. The high band of the mid (or mono) signal is encoded using BWE, which is denoted as HF encoding in Fig. 1a. In the specific case of stereo, a part of the low-band side signal (1–6.4 kHz nominal) is encoded using a parametric model briefly described below. The very low frequencies (0–1 kHz nominal) of the side signal are encoded using the core ACELP/TCX model. The high band of each channel (left and right) of a stereo input is encoded separately with the BWE approach. After quantization and encoding, the different parameters generate a bitstream which is multiplexed for transmission. The decoder, shown in Fig. 1b, demultiplexes the received bits, decodes them to obtain the quantized parameters, and then performs the inverse operations of the encoder to recover a synthesized audio signal.

In the ACELP/TCX core, each 1024-sample superframe is divided in four frames of 256 samples. Each frame can be encoded in one of four possible modes:

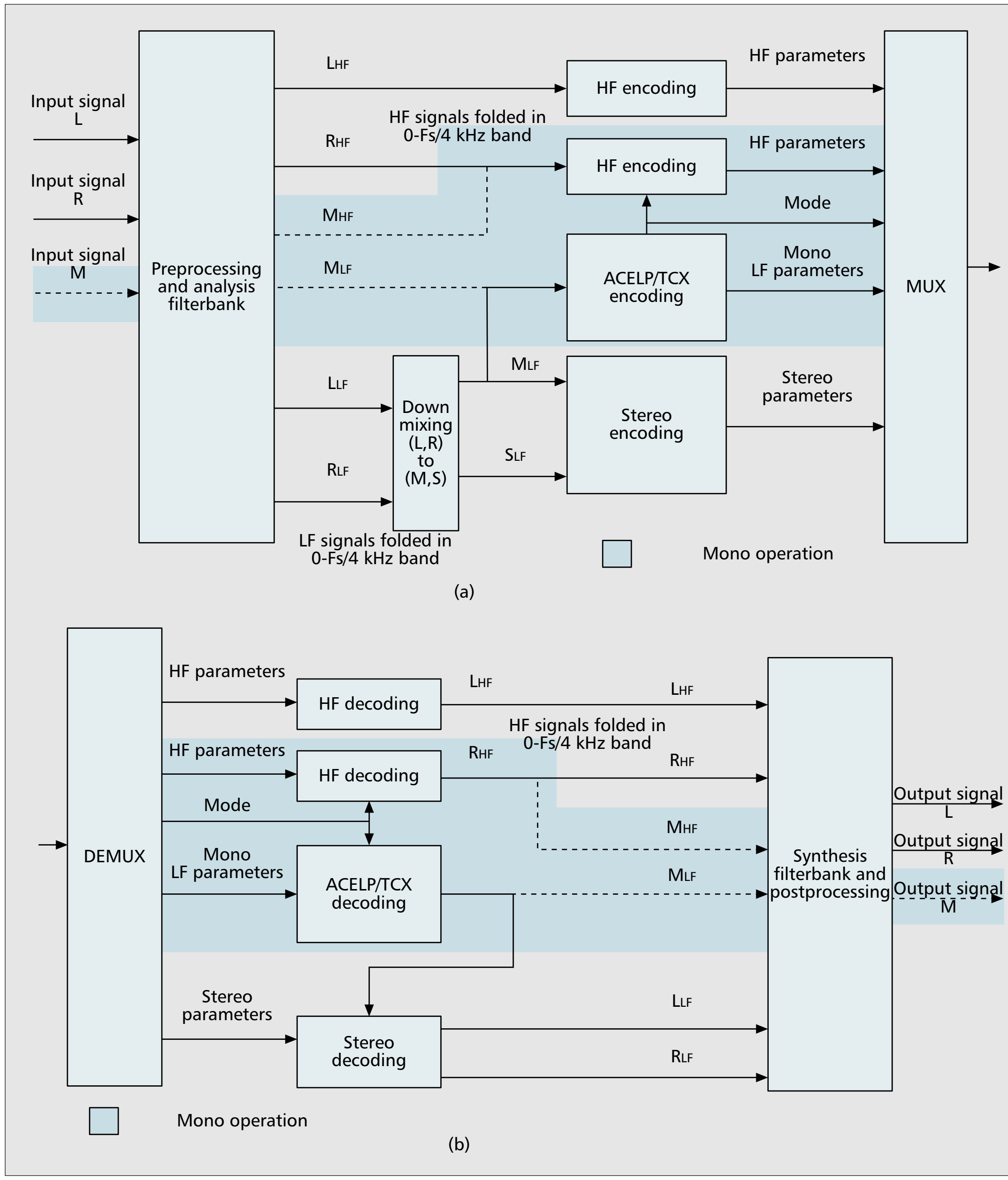
- As a 256-sample AMR-WB frame
- As a 256-sample TCX frame
- As part of a 512-sample TCX frame, in which case two consecutive frames are concatenated in a single larger frame before applying transform coding
- As part of a 1024-sample TCX frame, in which case the whole superframe is encoded in one single transform.

In transform mode, the spectral coefficients are quantized in eight-dimensional blocks using a technique called scalable algebraic vector quantization (VQ) [5]. Note that in the TCX mode, the frame is extended with a lookahead which is needed for overlapping transform windows. The use of different modes and frame sizes achieves better encoding of different types of sounds (e.g., speech, sustained audio, attacks) by allowing a trade-off between time and frequency resolution. More details on this hybrid ACELP/TCX technique can be found in [6].

A key component of the encoder is mode selection — choosing the most suitable mode combination for each superframe. Mode selection can be performed in either closed-loop or open-loop, depending on the complexity constraints at the encoder. Closed-loop mode selection is more complex than open-loop mode selection since it involves encoding the input signal more than once before selecting the best mode. Low-complexity open-loop mode selection results in some quality degradation especially for nonspeech signals.

The effectiveness of AMR-WB+ in providing various encoding modes is illustrated by the following mode distributions measured for different content types. Typical stationary instrumental music: ACELP (1 percent), TCX-256 (4 percent), TCX-512 (15 percent), and TCX-1024 (80 percent). Pop music: ACELP (11 percent), TCX-

AMR-WB+ uses a novel highly efficient parametric stereo coding technique. Since the encoder uses both time-domain and frequency-domain coding, a time-domain interchannel prediction approach is used.



■ **Figure 1.** *a) Structure of the AMR-WB+ encoder; b) structure of the AMR-WB+ decoder.*

256 (17 percent), TCX-512 (36 percent), and TCX-1024 (36 percent). Speech from several male and female talkers: ACELP (48 percent), TCX-256 (25 percent), TCX-512 (17 percent), and TCX-1024 (10 percent). This last example also shows that the TCX mode can be useful even for speech signals (stationary segments).

Encoding the High Frequencies

Recall from Fig. 1a that the input audio signal (mono signal, or downmix from stereo) is divided in two bands. The first, 6.4 kHz sampled at 12.8 kHz, is encoded with the core ACELP/TCX model as described above. Then the frequencies with above content at 6.4 kHz (signal M_{HF} in Fig. 1a) are encoded using a BWE approach. The approach consists of extracting a parametric representation, namely, the spectral envelope and gains, which are quantized and sent to the decoder. The spectral envelope is calculated once for the whole frame while the gains are calculated at every 64 samples. The spectral fine structure of the high-frequency signal is extrapolated at the decoder from the low-band excitation signal (0–6400 Hz nominal range), which is available from the encoded low-frequency signal M_{LF}.

Stereo Encoding

AMR-WB+ uses a novel highly efficient parametric stereo coding technique. Since the encoder uses both time-domain and frequency-domain coding, a time-domain interchannel prediction approach is used. Further, perceptually important cues for sound localization are low-frequency interchannel time differences and high frequency interchannel level differences. This suggests a division of the full frequency band into at least a low band and a high band. A low band (0–1 kHz) is encoded according to a waveform coding technique. A stereo balance factor is firstly derived, representing the ratio between the mono and side signal levels. Subsequently, in order to provide the perceptually important time resolution of the low-band stereo image, a critically down-sampled representation of the normalized side signal is waveform encoded. The coding is done in the frequency domain using a closed-loop variable-frame-length technique and algebraic VQ, reapplying the TCX coding methods of the core ACELP/TCX algorithm. In addition, a supplementary TCX mode using time-domain envelope shaping is used in order to efficiently encode the transient signals. Correspondingly, frame-length candidates are chosen from the total length of one superframe or subdivisions of length equal to one-quarter or one-half of the total length of the superframe. For frequencies above 1 kHz, the coding merely aims to match a target spectral shape. Specifically, band decomposition as in the mono case is used where the band up to 6.4 kHz is encoded according to a shape/gain-constrained time-domain filter approach. For the remaining high-band part of the stereo signal (above 6.4 kHz), limited spectral resolution is sufficient. For this band, it was found to be adequate to do the coding according to parametric BWE as described above, but applied on each of the two stereo channels.

COMPLEXITY OF AMR-WB+

AMR-WB+ specifications within 3GPP provide both floating-point and fixed-point reference C codes. The latter is implemented using a set of basic operators which simulate generic DSP instructions. Each basic operator is assigned a weight which reflects the number of cycles corresponding to that operator resulting in complexity estimates called weighted million operations per second (WMOPS). The complexity on a DSP is measured in million instructions per second (MIPS). The ratio between the estimated WMOPS and MIPS depends on the DSP used and the level of optimization. For modern DSPs such as TI C55, WMOPS and MIPS are about the same.

The estimated worst-case decoder complexity (48 kb/s stereo operation) is 23.9 WMOPS (this resulted in 24 MIPS on a C55 DSP). The decoder complexity at 24 kb/s is 17 WMOPS for stereo and 11 WMOPS for mono [3GPP TR 26.936]. An important feature of AMR-WB+, making it attractive for operation on power-limited battery-driven devices, is that its complexity can be scaled by means of the internal sampling frequency.

In most applications, since only the decoder is needed on the mobile terminal, the encoder complexity is not a significant issue. In some applications, such as terminal-generated messaging, it can be assumed that the signal will be encoded in a terminal at bit rates up to 24 kb/s.

For 24 kb/s mono content creation (in low-complexity, open-loop-mode selection operation), the average complexity is about 38 WMOPS, which is similar the complexity of the AMR-WB codec. In normal (closed-loop mode selection) mono operation, the encoder complexity at 24 kb/s is about 60 WMOPS [3GPP TR 26.936].

TRANSPORT AND FILE FORMAT OF AMR-WB+

The RTP payload format for AMR-WB+, including parameters required for session setup, is defined in IETF RFC 4352 [7]. It supports encapsulation of multiple AMR-WB+ transport frames per packet, and provides means for redundancy transmission and frame interleaving to improve robustness against packet loss.

The AMR-WB+ audio can be stored into a file using the ISO-based 3GP file format defined in 3GPP TS 26.244, which has the media type "audio/3GPP." Note that the 3GP structure also supports the storage of many other multimedia formats, thereby allowing synchronized playback.

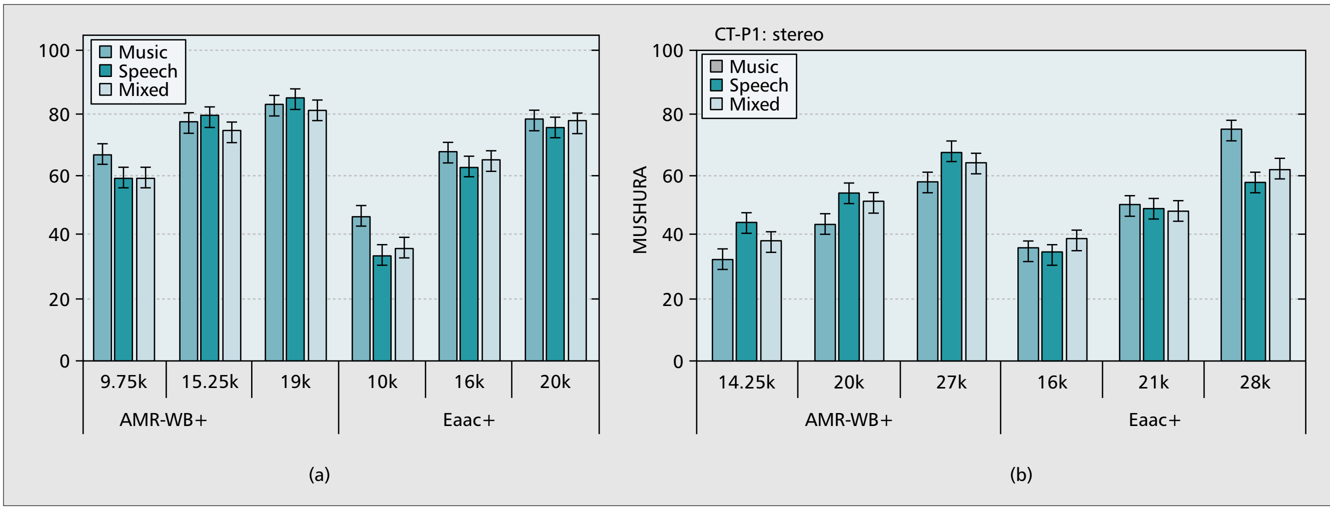
PERFORMANCE OF AMR-WB+

The AMR-WB+ audio codec has been extensively tested by numerous independent laboratories to assess its subjective performance in various application scenarios with relevant content types including speech, speech over music, and music. The test methodology mostly used was the MUSHRA methodology [8]. In the 2004 selection tests conducted by 3GPP, AMR-WB+ was evaluated in low-rate experiments in the 14–24 kb/s range, for both mono and stereo operation. The results showed that AMR-WB+ has the best audio quality at low rates when combining content types, compared to the competing algorithms aacPlus and Eaac+. As anticipated, the quality for speech-dominant content was significantly better than for audio coders, but the tests also showed that AMR-WB+ provided consistently good quality for music.

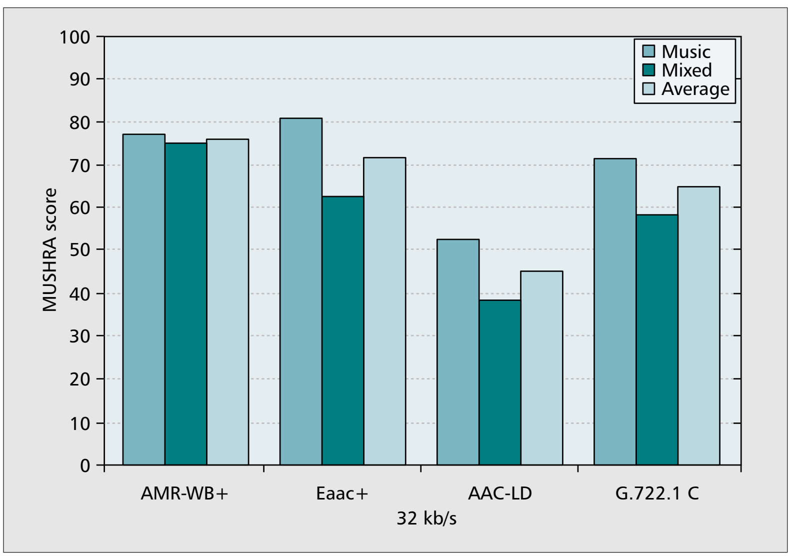
Following the selection of AMR-WB+ and Eaac+ in 2004, 3GPP conducted characterization tests using the fixed-point versions of the two codecs. These tests included a variety of content types (speech, music, and other) at different bit rates. These extensive tests, including the results and conclusions, can be found in a recent 3GPP technical report [3GPP TR 26.936]. The report also contains information about the complexity and delay analysis of both codecs. As noted in the report, the codecs used in the characterization test were different from the candidate codecs used in the earlier selection test (the improvements included bug fixes, optimized configurations, etc.). Thus, these recent characterization tests represent the actual performance of the 3GPP standardized audio codecs.

Figure 2 shows subjective performance of AMR-WB+ and Eaac+ from the 3GPP characterization tests [3GPP TR 26.936]. For the results in Fig. 2, AMR-WB+ was used in normal operation (closed-loop mode selection). Figure 2a summarizes the subjective performance of AMR-WB+ and Eaac+ at rates between about 10 to 20 kb/s, for mono operation. The advantage of using a hybrid model such as AMR-WB+ is obvious from this figure. At an equivalent bit rate, AMR-WB+ is always better than Eaac+. This is especially true for speech and mixed signals, but it is even the case for music. Figure 2a especially shows that when moving to very low rates, the subjective quality

The AMR-WB+ audio codec has been extensively tested by numerous independent laboratories to assess its subjective performance in various application scenarios with relevant content types including speech, speech over music, and music.



■ **Figure 2.** *a) Subjective evaluation at low rates in mono operation; b) subjective evaluation at low rates in stereo operation.*



■ **Figure 3.** *Summary of ITU-T characterization tests in mono at 32 kb/s.*

of AMR-WB+ does not suffer as much from rate reduction as a pure transform coder like Eaac+. At 9.75 kb/s, the performance of AMR-WB+ is close to that of Eaac+ at 16 kb/s.

The performance at low rates in stereo is shown in Fig. 2b. Again, at an equivalent bit rate, performance in speech and mixed signals is better for AMR-WB+ than Eaac+. For music, Eaac+ is better at the higher rates.

Further results from the characterization tests reported in 3GPP TR 26.936 show how AMR-WB+ performs at high rates, up to 32 kb/s. In Annex 1 of 3GPP TR 26.936, the results from an ITU-T characterization test using AMR-WB+ and Eaac+ as reference codecs are given. The test was performed in mono, using the MUSHRA methodology, and includes music and speech mixed with other audio content. In this test, the input signal was band limited to 14 kHz. Figure 3 summarizes the results obtained. In this test, the objective was to characterize Annex C of ITU-T recommendation G.722.1 for encoding 32 kHz sampled (14 kHz bandwidth) speech and audio at rates between 24 and 48 kb/s.

It can be clearly seen from Fig. 3 how AMR-WB+ delivers consistent quality across different content types.

Conclusion

Evolution in wireless communication systems has driven the deployment of innovative applications and services, including audio and video content. Even with the increased data capacity supported by 3G wireless systems, delivering multimedia content requires highly efficient use of network capacity. This is especially true for audio content, in order to allow room for video and other data in the wireless links. AMR-WB+ codec has the ability to consistently deliver high-quality audio across diverse content types, even at very low bit rates. In 2004, the 3GPP recommended AMR-WB+ as one of the key enabling technologies for the transmission and storage of audio content over wireless, thus helping new mobile multimedia services to deliver on their promise.

Acknowledgment

The authors wish to thank all the contributors to the AMR-WB+ codec, without whose significant contribution this achievement would not have been possible: Bruno Bessette, Daniel Enström, Ingemar Johansson, Kari Järvinen, Jari Mäkinen, Pasi Ojala, Stephane Ragot, Vesa Ruoppila, Patrik Sandgren, Joachim Thiemann, Henri Toukomaa, Tommy Vaillancourt, and Janne Vainio

References

[1] ISO/IEC 14496-3:2001, "Information Technology — Coding of Audio-Visual Objects — Part 3: Audio."

[2] "Specification for the Use of Video and Audio Coding in DVB Services Delivered Directly Over IP," DVB Doc. A084 Rev. 1, Nov 2005 (available for download at www.dvb-h-online.org)

[3] "IP Datacast over DVB-H: Architecture," DVB Document A098, Nov 2005; http://www.dvb-h-online.org

[4] 3GPP TS 26.290, "Extended AMR Wideband Codec; Transcoding Functions."

[5] S. Ragot, B. Bessette, and R. Lefebvre, "Low-Complexity Multirate Lattice Vector Quantization with Application to Wideband Speech Coding at 32 kb/s," *Proc. ICASSP-2004*, May 2004, pp. I-501–I-504.

[6] B. Bessette, R. Lefebvre, and R. Salami, "Universal Speech/Audio Coding Using Hybrid ACELP/TCX Techniques," *Proc. ICASSP-2005*, Mar. 2005, pp. III-301–III-304.
[7] J. Sjoberg *et al.*, "RTP Payload Format for the Extended Adaptive Multi-Rate Wideband (AMR-WB+) Audio Codec," IETF RFC 4352, Jan. 2006.
[8] ITU-R Rec. BS.1534, "Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems."

BIOGRAPHIES

REDWAN SALAMI [SM] (redwan.salami@voiceage.com) received a B.Sc. degree in electrical engineering from Al-Fateh University, Libya, in 1984, and M.Sc. and Ph.D. degrees in electronics from the University of Southampton, United Kingdom, in 1987 and 1990, respectively. From 1990 till 2000 he was at Université de Sherbrooke as a research assistant and an adjunct professor. In 1999, he was a cofounder of VoiceAge Corporation where he currently holds the position of Vice-President, R&D. He contributed to several speech coding standards including ITU-T Recommendation G.729, EFR GSM codec, 3GPP AMR-WB codec (ITU-T Recommendation G.722.2), 3GPP2 VMR-WB codec, and 3GPP AMR-WB+ codec. He has numerous publications and patents in the area of speech coding.

ROCH LEFEBVRE (Roch.Lefebvre@USherbrooke.ca) received his Ph.D. in electrical engineering in 1995 from Université de Sherbrooke. Since 1996 he is a full-time professor at Université de Sherbrooke, where he also heads the Speech and Audio Processing Laboratory. He is one of the co-founders of VoiceAge corporation. His research interests include speech and audio coding, noise reduction, and robustness in packet networks.

ARI LAKANIEMI received his M.Sc. degree in signal processing from Tampere University of Technology, Finland, in 1995. He is currently working as a research engineer at Nokia Research Center, Helsinki, Finland.

KALERVO KONTOLA (kalervo.kontola@nokia.com) received his M.Sc. degree in signal processing from Tampere University of Technology, Finland, in 1993. He works at Nokia Research Center, Tampere, Finland. His current responsibilities include research and standardization work for many mobile-application-oriented standards from audio and other multimedia delivery points of view.

STEFAN BRUHN received his Ph.D. in electrical engineering from the Technical University of Berlin in 1995 and has been with Ericsson since then. He was involved in development and standardization of various 3GPP speech and audio codecs such as AMR, AMR-WB, and AMR-WB+. His current position is Senior Specialist in the area of speech coding within Ericsson Research. He holds more than 10 patents.

ANISSE TALEB received M.Sc. and Ph.D. degrees in electrical engineering from the Institut National Polytechnique de Grenoble (INPG), France, in 1996 and 1999, respectively. He was awarded the INPG Best Thesis Prize in 2000. After a short period (2000–2001) as a post-doctoral research fellow at the Australian Telecommunication Research Institute, Perth, in 2001 he joined Ericsson, where he is a senior research engineer in the field of audio technology. He was heavily involved in the development and standardization of the AMR-WB+ audio codec. He is currently involved in the M-PIPE IST 6th Framework European Project working on scalable audio coding, and is also active in standardization bodies such as MPEG and 3GPP. His current research interests include multichannel audio coding, scalable coding, and the unification of speech and audio coding.