

AMR-WB+: Low Bit Rate Audio Coding for Mobile Multimedia

**Kalervo Kontola,
Jari M Makinen**

Nokia, Finland

kalervo.kontola@nokia.com,

jari.m.makinen@nokia.com

**Anisse Taleb,
Stefan Bruhn**

Ericsson, Sweden

anisse.taleb@ericsson.com,

stefan.bruhn@ericsson.com

**Bruno Bessette,
Redwan Salami**

VoiceAge, Canada

bruno.bessette@USherbrooke.ca

redwan.salami@voiceage.com

Abstract

Various new mobile multimedia broadcasting systems are being deployed on the market. These systems are different for example in their radio part, transport methods and media formats. Harmonization of components between different systems is beneficial only when the selected system component is optimal for all systems.

Audio content is and will be the main content type in broadcasting, with or without other content types. A good audio codec should provide consistent high quality for different audio types like speech, music, and speech mixed with music and maintain this high quality consistency even at low bit rates. That is essential when audio is delivered with some other content or data consuming most of the channel capacity.

The Extended AMR-WB (AMR-WB+) audio codec meets these requirements, providing high quality at exceptionally low rates, and in the same time providing consistent quality over all audio types. It has been designed and standardized for 3GPP audio services in Packet-switched Streaming Service (PSS), Multimedia Broadcast/Multicast Service (MBMS), and Multimedia Messaging Service (MMS), and also selected in DVB, especially in its new IP Datacast (IPDC) over DVB-H service. AMR-WB+ contains modes designed to speech, music and mixed content. AMR-WB+ encodes to bitrates from 6 kb/s to 48 kb/s input signal that are either monophonic or stereophonic and with sampling frequency in range from 8 to 48 kHz. AMR-WB+ is very scalable audio codec.

Keywords

Audio, AMR-WB+, 3GPP, DVB, broadcasting

INTRODUCTION

Multimedia content can be delivered to and from mobile devices by various ways like streaming, downloading and uploading. They can be also divided into broadcasting (and its subset multicasting) and unicasting services. For example, 3GPP and DVB broadcasting support streaming and downloading and 3GPP allows also various point-to-point connections.

Even if those transfer methods are different there are commonalities what kind of content users want to receive. For example, when one is walking, it is easier to listen to audio than watch only a video or graphics on a display. On the other hand, many services can be enhanced by using a few different media components. It can be expected that most of the users want to get at least audio but of course would like to have other media components too.

For example, delivery of news, sports, audio books, mobile TV or music videos could include both audio and video or other visual data. In such case a typical audio content includes at least speech or music, often also various combinations of them. In order to prevent switching of codecs, one audio codec should be used in various cases. Therefore an optimal audio codec for such would provide consistent high quality for all kind of audio types and maintain it even at low bit rates. This requirement for low bit rates becomes especially true in mobile multimedia applications when only the part of the channel is available for audio. Examples are when the channel is shared with video content or when the transport mechanism headers and error correction methods consume part of the channel capacity thus reducing the overall available bit rate. For these reasons, the requirement for low bit rate high quality audio codec is valid for mobile multimedia applications.

The Extended AMR-WB (AMR-WB+) audio codec meets these requirements, providing high quality at exceptionally low rates, and in the same time providing consistent quality over all audio types. Designed and standardized for 3GPP audio services in Packet-switched Streaming Service (PSS), Multimedia Broadcast/Multicast Service (MBMS), and Multimedia Messaging Service (MMS), AMR-WB+ has been also selected in Digital Video Broadcasting (DVB), especially in its new IP Datacast (IPDC) over DVB-H (Handheld) service.

This article presents first short overview of multimedia broadcasting to mobile devices and especially in 3GPP and DVB-H services. Then are described main requirements for audio codecs used in those standards and other mobile multimedia applications. The next chapters explain the AMR-WB+ audio codec, its complexity and performance and finally conclusion.

MOBILE MULTIMEDIA BROADCASTING

Many mobile devices support widely used FM radios to listen normal radio programs. However, other content types cannot be combined to those services directly but have to be delivered via other means. For example, interactive low bit rate visual content can be synchronized to broadcasted FM radio and sent over GPRS (General Packet Radio Service) like is done in Visual Radio application.

The current trend is to have similar or better user experience by broadcasting combined multimedia content. Due to many technical and other reasons, these systems are different in many ways, for example in their radio part, service signaling, transport methods and media formats. Harmonization of components between different systems is beneficial since it lowers implementation costs of devices supporting the many broadcasting systems and provides also better interoperability between applications. However, harmonization is beneficial only when the selected system component is optimal for all content delivery methods. Natural examples of those common system components are media codecs. Content can be a combination of different media types like audio, video, graphics, images and text. Many application specifications provide also methods for scene description, multimedia integration and synchronization schemes.

Several new mobile multimedia broadcasting systems are being deployed on the market. This article presents in more details 3GPP MBMS [5] and recently specified IPDC over DVB-H service [13] because those are going to be used in various mobile broadcasting services.

Broadcasting via 3GPP MBMS

3GPP MBMS is a point-to-multipoint service in which data is transmitted from a single source entity to multiple recipients. The MBMS bearer service offers broadcast and multicast modes. The current delivery methods for the MBMS user services are download and streaming [5]. MBMS file download delivery method uses the FLUTE protocol [9] and the RTP [8] is the transport protocol for MBMS streaming that delivers continuous multimedia data.

3GPP mobile communications can use either GPRS or UMTS (Universal Mobile Telecommunications System) radio access technology. The available total bit rates depend on the fixed bearers. With UMTS at the moment a typical channel bandwidth is 64 or 128 kb/s (later even 256 kb/s) for streaming, but transport mechanism headers and possible use of forward error correction (FEC) could consume some part of that. Therefore net rate available for streaming content can be assumed to be 48 kb/s in the case of a 64 kb/s bearer. The values for GPRS would be 36 and 24 kb/s, respectively. Typical message size in MBMS downloading is 300 kB.

Broadcasting via IPDC over DVB-H

IP Datacast over DVB-H is an end-to-end broadcast system for delivery of any types of digital content and services using IP-based mechanisms optimized for devices with limitations on computational resources and battery. An inherent part of the IPDC system is that it comprises a unidirectional DVB broadcast path that may be combined with a bi-directional mobile/cellular interactivity path. [13]

The DVB-H system is based on the existing DVB-T standard for fixed and in-car reception of digital TV [11]. The main additional elements of the DVB-H in the link layer are time slicing and additional FEC coding. In the time slicing data is sent in bursts and to receiver's radio is signaled when it would get those data bursts and otherwise it could be switched off. The lower the content bit rate, the longer time a receiver can be off and better power saving is achieved. Time slicing reduces the average power in the receiver front-end significantly - up to about 90%–95% [11]. In addition time slicing enables smooth and seamless frequency handover. Use of time slicing is mandatory in DVB-H. In order to protect against transmission errors, optional multiprotocol encapsulation–FEC (MPE-FEC) can be used.

IPDC support both streaming and file delivery. The former uses RTP [8] and the latter uses FLUTE [9]. Thus these protocols have been harmonized with 3GPP MBMS. An IPDC service provider decides how many services are transmitted within one burst. Again, low content bit rate is beneficial because in that way more services can be included in one burst.

REQUIREMENTS FOR AUDIO CODECS IN MOBILE MULTIMEDIA

These 3GPP and DVB-H multimedia broadcasting systems allow several use cases like delivery of news, sports, commercials, traffic or weather information, audio books, music, various edutainment, mobile TV and music videos. All of these examples use audio as a primary or secondary content. It is obvious that such content comprises a wide variety of audio types such as speech, music and speech mixed with music, both in mono and stereo.

Like presented above, low bit rate audio coding is useful in those standardized broadcasting scenarios. That is especially true when the channel is shared with video in which case typically only about 25% from the available content bit rate is allocated to audio. Audio books may include many hours of audio content and therefore very high audio compression is essential in order to shorten the downloading time and to reduce the requirements for content storage space. Podcasting type of services is a new example in which both music and speech content is sent and the length of the program may be even many hours. Therefore use of low bit rate audio coding saves channel

capacity no matter if it run over wired or wireless networks and no matter if the service is a streaming or a download service. Users could also generate audio content by themselves and then deliver that e.g., via podcasting. When such data is encoded e.g. in live concert then it is essential to use low bit rate audio coding to reduce delivery costs. Sometimes even mono encoding would be reasonable in such cases. There could be also other broadcasting services that provide some content as free (e.g. preview content) and lower bit rate could be used to decrease delivery cost.

In conclusion, there is a high demand for low bit rate high quality audio codec especially for the various conceivable mobile multimedia applications. In addition of those above mentioned advantages of well done harmonization of components, from audio codec point of view it impacts also to licensing costs.

Using AMR-WB+ in Mobile Applications

The AMR-WB+ audio codec [2] was designed based on main service requirements of 3GPP mobile communications defined for GPRS or UMTS radio access technology. These requirements were set such that unlike 3GPP Release 5 audio codecs they would lead to new codecs enabling high quality multimedia services at very low bit rates and in the same time providing consistent quality over all audio types.

AMR-WB+ has been standardized for 3GPP Release 6 audio services in MBMS [5], Packet-switched Streaming Service (PSS) [7], and Multimedia Messaging Service (MMS) [6]. PSS and MMS allow sending more customized audio and other data and also person-to-person MMS is possible. Otherwise the application scenarios for MMS and PSS are similar to MBMS. Therefore the media formats and content types are harmonized between these three services. AMR-WB+ has been also selected in DVB, especially in its new IPDC over DVB-H service [12][13]. The RTP payload of the AMR-WB+ has been specified in IETF 4352 [10] and is used in MBMS, PSS and IPDC streaming applications.

Many current podcasting services send audio at bit rates from 30 to 60 kb/s that is too low for the used mp3 audio codec. AMR-WB+ performs substantially better at those bit rates. It would also be a premium choice for audio book applications comprising various audio content types.

AMR-WB+ CODEC OVERVIEW

The codec input signal can be monophonic or stereophonic with sampling frequencies ranging from 8 kHz up to 48 kHz. 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 codec contains also the standard AMR-WB speech codec modes ranging from 6.60 kb/s to 23.85 kb/s, as well as

AMR-WB voice activity detection (VAD) and discontinuous transmission (DTX) functionality [1].

AMR-WB+ is an audio codec utilizing novel hybrid technology that combines the strengths of both audio and speech coding techniques [2]. The hybrid technology includes ACELP (Adaptive Code Excited Linear Prediction) coding for optimal compression of speech kind of signals and transform-based coding (TCX) to achieve high quality compression for richer sounds like music. The AMR-WB+ encoder selects the most optimal coding model (ACELP or transform) based on source audio signal, thereby providing high-quality audio across a wide range of sounds. In addition, AMR-WB+ utilizes a parametric stereo model to convey stereo image of audio signal at very low bit rates and therefore to enhance the end-user audio perception.

AMR-WB+ is very scalable in terms of the total bit rate as well as the bit rate allocation between mono and stereo coding. Scalable internal sampling frequency enables both audio bandwidth and bit rate adjustments thus increasing the flexibility of the codec even further by facilitating operation over a wide variety of applications with different audio quality requirements. Switching between codec configurations with different overall bit rates but also mono or stereo operation as well as different internal sampling frequencies is possible even during the encoding process enabling flexibility for real time services. AMR-WB+ encoder and decoder structures are shown in Figure 1 and 2. The codec processes the input signal by decomposing it into two frequency bands: The low frequency band up to 6.4 kHz is generated by downsampling the input signal to 12.8 kHz internal sampling rate and then applying the hybrid ACELP/TCX encoding model. The high-frequency band encoded with bandwidth extension approach contains all frequencies above 6.4 kHz. It should be noted, however, that the audio bandwidth can be changed with scaling the internal sampling frequency with a factor ranging from 0.5 to 1.5. This enables efficient use of the available service bandwidth.

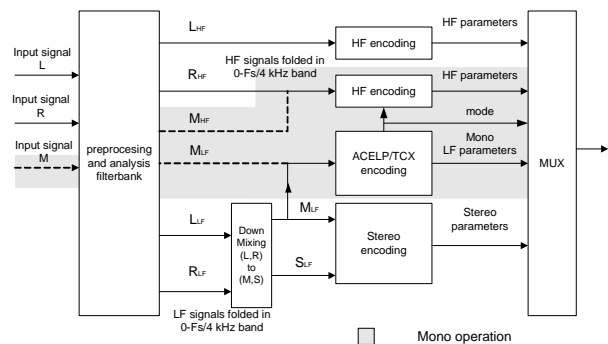


Figure 1. AMR-WB+ Encoder Architecture

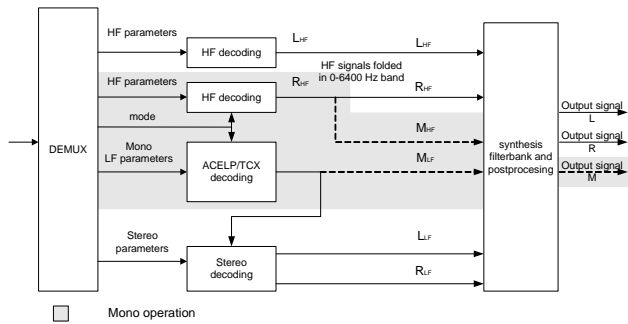


Figure 2. AMR-WB+ Decoder Architecture

Other key elements of AMR-WB+ technology are its built-in robustness to typical network impairments such as packet losses and its seamless adaptability to varying available bit rates. Detailed codec description can be found in [2].

AMR-WB+ CODEC PERFORMANCE

In mobile multimedia broadcasting applications, limited computational resources of battery-driven devices set very strict requirements for memory consumption and complexity. In most mobile terminal applications, since only the decoder is needed, the encoder complexity is not a significant issue. Therefore only performance figures of the AMR-WB+ decoder are given with more details.

Decoder Memory Consumption and Complexity Evaluation

AMR-WB+ specifications include also both floating-point and fixed-point reference C-code. The fixed-point implementation [4] applies a set of basic operators simulating generic DSP instructions. Each basic operator is assigned a weight which reflects the number of DSP cycles corresponding to that operator resulting in complexity estimates called WMOPS (weighted million operations per second). The complexity on a DSP is measured in MIPS (million instructions per second). The ratio between the estimated WMOPS and MIPS depends on the DSP used and the level of optimization.

The estimated average and worst case decoder complexities are shown in Table 1. As can be seen, AMR-WB+ decoding is a very low complex process for both mono and stereo signals. 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 [3]. This is lower than the complexity at 48 kb/s due to the use of lower internal sampling frequency at 24 kb/s. In fact, an important feature of AMR-WB+ is that the complexity can be scaled by means of changing the internal sampling frequency. This makes it attractive for operation on

services that are used mainly by devices having limited computational resources.

Table 1. Decoder complexity estimates in WMOPS based on the reference fixed point C-code [4]

Condition bit rate, content	Average/WorstCase Complexity (wMOPS)
14 kb/s, mono	8.4 / 8.8
24 kb/s, mono	10.4 / 11.0
36 kb/s, mono	12.5 / 13.2
18 kb/s, stereo	16.0 / 16.6
24 kb/s, stereo	17.7 / 18.3
32 kb/s, stereo	20.3 / 21.0
48 kb/s, stereo	23.0 / 23.9

Memory consumption of AMR-WB+ decoder below 24 kb/s is 13 Kwords for Random Access Memory (RAM) and 20 Kwords for Read Only Memory (ROM).

Audio Quality Evaluation

3GPP conducted characterisation tests for 3GPP audio codecs to assess its subjective performance in different application scenarios with various content types. Both AMR-WB+ and E-AAC+ codec were tested, since both audio codecs were selected for 3GPP multimedia services (and later also to IPDC over DVB-H service). Extensive 3GPP characterisation testing was collectively done by several listening test laboratories. The used test methodology was the MUSHRA methodology. The audio quality at low bit rates were evaluated in the range of 14-28 kb/s, in both mono and stereo operation. These tests included a variety of content types (speech, music and other). A detailed description of these extensive tests, including the results and conclusions, can be found in a recent 3GPP technical report [3].

These recent characterization tests represent the actual performance of the 3GPP standardized audio codecs. In addition, compared to earlier assessments of the 3GPP Release-6 audio codecs the characterization tests encompass a wider range of codec operating points and conditions thus providing a better picture as means for codecs comparisons. Figure 3 summarizes the subjective performance of AMR-WB+ and E-AAC+ at rates between 9 to 20 kb/s, on mono operation. At an equivalent bit rate, AMR-WB+ turns out to be always better than E-AAC+. This is especially true for speech and mixed signals but also for music.

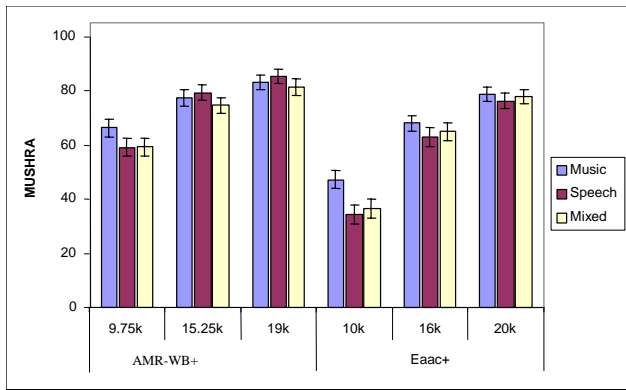


Figure 3. Subjective evaluation at low rates in mono operation.

Figure 4 summarizes the subjective performance of AMR-WB+ and E-AAC+ stereo operations at rates between 14 to 28 kb/s. Again, at an equivalent bit rate, performance in speech and mixed signals is better for AMR-WB+ than E-AAC+. The situation is reversed for music at the higher rates (around 27-28 kb/s), but not at the lower rates.

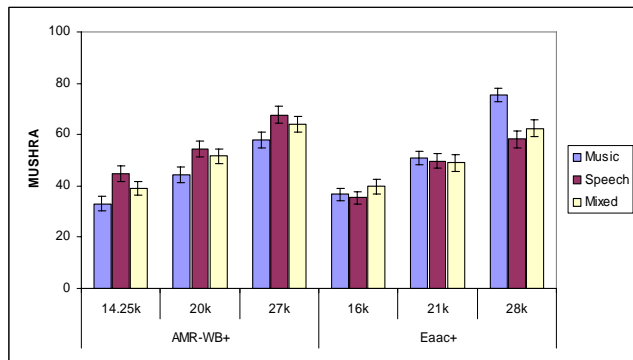


Figure 4. Subjective evaluation at low rates in stereo operation

Further evaluations of AMR-WB+ at Ericsson and Nokia test sites show a consistent quality increase with increasing bit rates. At the maximum stereo operation bit rate of 48 kb/s AMR-WB+ achieves quality grades corresponding to the highest MUSHRA category ‘excellent’, which is close to transparent. Overall, across content types, the performance at this rate is comparable to the quality achievable with E-AAC+. AMR-WB+ maintains its relative strength compared to E-AAC+ for speech and mixed signals, whereas for music the situation is reversed.

ACKNOWLEDGMENTS

The authors wish to thank all the contributors to the AMR-WB+ codec, without whose significant contribution, this achievement would not have been possible: Daniel Enström, Ingemar Johansson, Kari Järvinen, Ari Lakaniemi, Roch Lefebvre, Pasi Ojala, Stephane Ragot,

Vesa Ruoppila, Patrik Sandgren, Joachim Thiemann, Henri Toukoma, Tommy Vaillancourt and Janne Vainio.

CONCLUSION

Wireless multimedia applications convey generic audio content, which type may range from speech to music. Therefore an ideal audio codec should have consistent quality for any of these audio content types. In order to save network capacity or device battery, such codec should provide high quality at lowest possible bit rates, especially in mobile multimedia broadcasting like IPDC over DVB-B and 3GPP MBMS.

In this paper we show that AMR-WB+ is a codec matching this requirement. At low rates, the subjective quality of AMR-WB+ does not suffer as much from rate reduction compared to a pure transform coder as E-AAC+. Also AMR-WB+ has the best audio quality over all content types. Both is explained by the use of a hybrid model, the quality for speech-dominant content is significantly better than for traditional audio codecs, but the characterisation tests also indicated that AMR-WB+ provides consistently good audio quality for music.

The AMR-WB+ audio codec has been standardized for 3GPP Release 6 audio services in MBMS, PSS, and MMS. It has been also selected in DVB, especially in new IPDC over DVB-H service for mobile devices. It would also be a premium choice for audio book applications comprising various audio content types.

REFERENCES

- [1] 3GPP TS 26.190, “AMR Wideband speech codec; Transcoding functions”
- [2] 3GPP TS 26.290; “Extended AMR Wideband codec; Transcoding functions”
- [3] 3GPP TR 26.936, “Performance characterization of 3GPP audio codecs”
- [4] 3GPP TS 26.273, “ANSI-C code for the fixed-point Extended Adaptive Multi-Rate - Wideband (AMR-WB+) speech codec”
- [5] 3GPP TS 26.346: “Multimedia Broadcast/Multicast Service; Protocols and Codecs” (Release 6)
- [6] 3GPP TS 26.140: “Multimedia Messaging Service (MMS); Media formats and codes” (Release 6)
- [7] 3GPP TS 26.234: “Transparent end-to-end Packet-switched Streaming Service (PSS); Protocols and codecs”(Release 6)
- [8] IETF RFC 3550: “RTP: A Transport Protocol for Real-Time Applications”, Schulzrinne H. et al., July 2003.
- [9] IETF RFC 3926: “FLUTE – File Delivery over Unidirectional Transport”, Paila T. et al., October 2004.

- [10] IETF RFC 4352: "*RTP Payload Format for the Extended Adaptive Multi-Rate Wideband (AMR-WB+) Audio Codec*", Sjoberg J. et al., January 2006.
- [11] G. Faria et al., "*DVB-H: Digital Broadcast Services to Handheld Devices*", *Proc. IEEE*, vol. 94, Issue 1, pp. 194-209, Jan 2006.
- [12] Digital video broadcasting (DVB); "*Specification for the use of Video and Audio Coding in DVB Services Delivered Directly over IP*", ETSI TS 102 005 v1.2, European Telecommunications Standards Institute (DVB Document A084 Rev. 1, Nov 2005 available for download at www.dvb-h-online.org)
- [13] "*IP Datacast over DVB-H: Architecture*", DVB Document A098, Nov 2005 (available for download at www.dvb-h-online.org)