AMR-WB+: A NEW AUDIO CODING STANDARD FOR 3RD GENERATION MOBILE AUDIO SERVICES

Jari Mäkinen[1], Bruno Bessette[2,3], Stefan Bruhn[4], Pasi Ojala[1], Redwan Salami[2], Anisse Taleb[4]

[1] Multimedia Technologies Laboratory, Nokia Research Center, Finland.
[2] VoiceAge Corp., Montreal, Qc, Canada. [3] University of Sherbrooke, Qc, Canada.
[4] Multimedia Technologies, Ericsson Research, Sweden.

ABSTRACT

Highly efficient low-rate audio coding methods are required for new compelling and commercially interesting applications of streaming, messaging and broadcasting services using audio media in 3rd generation mobile communication systems. After an audio codec selection phase 3GPP has standardized the extended AMR-WB (AMR-WB+) codec that provides unique performance at very low bit rates from below 10 kbps up to 24 kbps. This paper discusses the requirements imposed by mobile audio services and gives a technology overview of AMR-WB+ as a codec matching these requirements while providing outstanding audio quality.

1. INTRODUCTION

3GPP is specifying multimedia services for 3rd generation mobile networks. During the year 2004, an extensive codec evaluation process, testing different coding algorithms, was conducted for the Release 6 multimedia service specifications. After defining the design constraints and performance requirements, a test plan has been laid out and the selection process consisted of subjective listening tests in order to analyze the candidate codecs performance in different operation conditions.

Based on selection criteria and the results of the listening tests, 3GPP selected two codecs for Release 6 services. Both AMR-WB+ [1] and Enhanced aacPlus [2] codecs are recommended for audio coding.

The paper is organized as follows. First, a discussion of the service requirements for mobile audio is provided in section 2, where conclusions are derived regarding the expected relevant audio content types and the corresponding Transmission bit rate restrictions. Section 3 provides a technology overview of the AMR-WB+ codec. Novel coding techniques leading to the outstanding AMR-WB+ distortion-rate performance are highlighted. In particular, the hybrid codec structure combining ACELP technology from the AMR-WB speech codec with transform coding in the perceptually weighted signal domain according to the TCX paradigm are presented. Furthermore, new techniques for bandwidth extension, high-efficiency stereo coding and flexible codec control are presented. Formal subjective quality evaluation results are illustrated in section 4, where the properties of both recommended 3GPP Release 6 audio codecs at bit rates below 24kbps are compared. A summary in section 5 concludes the paper.

2. SERVICE REQUIREMENTS FOR MOBILE AUDIO

Audio coding for mobile applications has to cope with hard requirements due to the nature of mobile wireless transmission. The transmission resource allocated for audio impacts the total radio capacity of the communication system and is thus limited both due to technical and economical reasons. Hence, in order to use the available resources as efficiently as possible it is necessary to tailor the audio codec to the specific applications. This section discusses the requirements on the audio codec imposed by various relevant mobile audio use cases, which are defined for 3GPP mobile communications using GPRS or UTRAN radio access technology (RAT). Typical audio content and available bit rates depending on the use case and transport mechanism will be outlined.

2.1. Relevant Audio Content

Table 1 lists the relevant mobile audio/audio-visual media distribution use cases covered by the service requirement specifications for both transparent end-to-end packet-switched streaming service [3], Multimedia Broadcast/Multicast Service (MBMS) user services [4], and Multimedia Messaging Service (MMS) [5] in 3GPP systems. The table provides information on the envisioned content for the different use cases and indicates cases for which a given transport mechanism would not be applicable.

As can be seen, most cases are dominated by speech and mixed content. Music content distribution is an important exception. Furthermore, there are certain personalized use cases, which are not applicable for MBMS transport. High-quality music distribution with individually purchased tunes is also a personalized service for which PSS or MMS transport mechanisms are better suited. . All listed cases may comprise audio-only or audio-visual content.

Table 1. Audio/audio visual media distribution use cases and content types by transport mechanism.

Use case \ Transport	PSS	MMS	MBMS streaming	MBMS download
Information (News, sports, stock quotes, traffic, weather)	Dominant 'speech', 'mixed'	Dominant 'speech', 'mixed'	Dominant 'speech', 'mixed'	Dominant 'speech', 'mixed'
Travel Guide	Dominant 'speech', 'mixed'	Dominant 'speech', 'mixed'		
M-Commerce (Online shopping, Advertisements)	Dominant 'speech', 'mixed'	Dominant 'speech', 'mixed'		
Edutainment (Learning, How-to) Corporate (Instructions)	Dominant 'speech', 'mixed'	Dominant 'speech', 'mixed'		
TV, Movies	'speech', 'music', 'mixed'	'speech', 'music', 'mixed'	'speech', 'music', 'mixed'	'speech', 'music', 'mixed'
Person-to-person MMS		Dominant 'speech', 'mixed'		
Audio Content Distribution – Music	'music'	'music'	'music'	'music'
Audio Content Distribution - Audio books	Dominant 'speech', 'mixed'	Dominant 'speech', 'mixed'	Dominant 'speech', 'mixed'	Dominant 'speech', 'mixed'

2.2. Available bit rates depending on RAT

Table 2 provides examples of the bit rates available for audio/audio-visual media distribution based on 3GPP mobile bearer realizations. These bit rates are realistic given the impact on the radio capacity. Depicted are the total bit rates offered by the bearer and the available effective net bit rates for the media composition excluding the additional transport overhead. While UTRAN can offer net bit rates for audio-only of up to 48 kbps, GPRS service using 3 time slots provides a maximum bit rate of approximately 24 kbps for PSS and MBMS streaming. For the MBMS streaming service, where FEC protection mechanisms are likely to be used, the available net bit rates may even be as low as 18 kbps. For the MMS and MBMS download cases, assuming reasonable message sizes of 100 or 300 kByte, then the available bit rates for audio-only content will be 14 to 24 kbps if the length in time in the order of 0.5 to 3 minutes.

For audio-visual content the bit rates available for audio are further reduced by the bit rate required for the video. Although the required video bit rate is highly dependent on the content, a reasonable assumption for best-possible audio-visual quality is that video requires about 75% of the available bit rate while audio consumes the remaining 25%. Such an assumption leads to the audio net rates given in table 2 for audio–visual content. It is easily apparent that merely very low bit rates of 10 to 16 kbps or, in case FEC is used, even lower rates are available. The highest possible rate of 24 kbps could be achieved only when using MBMS streaming with a 128 kbps bearer.

2.3. Conclusion

In summary, it appears quite clearly that the most relevant bit rate range for mobile audio application is from about 10 to 24 kbps. Therefore, compression ratios of 64 to more than 150 are required when comparing to a 16 bit stereo PCM signal sampled at 48 kHz. Additionally, unlike traditional audio, the envisioned mobile use cases imply dominant speech content together with mixed and music content.

Another important aspect is error resilience since at least in MBMS streaming there is a high likelihood for packet losses on the wireless link.

Furthermore, low complexity especially of the decoder is crucial considering that simultaneous video and FEC decoding must be manageable on mobile terminals with limited computational resources.

Table 2. Available audio bit rates for audio and audio-visual media distribution depending on service and radio access technology

		Audio		Audio-visual	
Transport	RAT	Total	Audio (net rates)	Total	Audio (net rates)
PSS	UTRAN	64 kbps	48 kbps	64 kbps	~14 kbps
	GPRS	36 Kbps	24 kbps	36 Kbps	<~ 10 kbps
MBMS Streaming	UTRAN	64 kbps	48 kbps	64/(128) kbps	12-16 kbps/ (24 kbps)
	GPRS	36 Kbps	24 kbps	36 Kbps	<~ 10 kbps
MMS	UTRAN	100 kB	0.5 min * 24 kbps	75 kB (video) + 25 kB (audio)	20 sec * 10 kbps
	GPRS		1 min * 14 kbps		
MBMS download	UTRAN	300 kB	1.5 min * 24 kbps	225 kB (video) + 75 kB (audio)	60 sec * 10 kbps
	GPRS		3 min * 14 kbps		

3. AMR-WB+ TECHNOLOGY OVERVIEW

The AMR-WB+ coder is based on a hybrid ACELP/TCX model, this allows switching between LP-based and transform-based coding depending on the signal characteristics. The input signal can be mono or stereo with sampling frequencies ranging from 16 up to 48 kHz.

Assuming stereo input, a sum signal and a difference or side signal are first computed. The sum signal is further decomposed in two bands: a low frequency signal s_L, downsampled to 12.8 kHz, the nominal internal frequency of AMR-WB, and a high-frequency signal s_H, containing all frequencies above 6.4 kHz. The hybrid ACELP/TCX encoding model is applied to s_L while a bandwidth extension approach is used to encode s_H.. The side signal is encoded using a low-rate semi-parametric approach, which preserves the stereo image.

3.1. Encoding of the low-frequency mono signal

The low-frequency mono signal s_L is encoded using hybrid ACELP/TCX. AMR-WB [8] is used in ACELP mode while TCX with algebraic VQ [9] is used in transform- coding mode. The signal is processed in 1024-sample super-frames in which frames of 256, 512 or 1024 samples can be used. Any 256-sample frame can be encoded using either AMR-WB or TCX, while a 512-sample frame, which can be formed at the beginning or at the end of the super-frame, and a 1024-sample frame are encoded in TCX. There are thus 26 different mode combinations within a super-frame.

Mode selection can be performed either in closed-loop or in open-loop, which allows a control the complexity of the encoder. Since each 256-sample frame can be in either one of 4 modes (AMR-WB or TCX within the frame, or part of a 512 or 1024 sample TCX frame), all 26 mode combinations within a super-frame can be tried and compared in closed-loop by subjecting each 256-sample frame to only 4 encodings, as described in [1]. In order to

save encoding complexity, instead of a fully closed-loop search, the mode combination can be determined in open-loop fashion. I.e., the input signal of the encoder is analyzed and the mode combination is selected based on audio signal characteristics.

Since TCX is a transform-based coding mode, applied to the target or weighted signal, non-rectangular overlapping windows improves the coding gain. On the other hand, ACELP uses an implicit rectangular window on the target. Hence, windowing and mode switching is an important issue in this hybrid structure. For this purpose, the window used in TCX mode has the following characteristics. The window is flat in the middle part, covering most of the TCX frame up to the end of the frame. Then, the window extends in the next frame in a decreasing half-cosine shape to form a look-ahead and overlap part. The length of the look-ahead increases with the TCX frame length. Finally, the window at the beginning of the frame can have two shapes: it is flat if the previous frame was ACELP, otherwise it is the complementary half-cosine shape at the end of the previous TCX window.

In a transition from a TCX frame to another TCX frame, the window overlap manages the frame transition. In a transition from ACELP to TCX, however, the transition must be managed otherwise. Specifically, the zero-input response (ZIR) of the weighting filter ($W(z)$) is computed and truncated, and then subtracted from the beginning of the TCX frame. This ensures that the target signal smoothly tends towards zero at the beginning of the frame, since the ZIR is a good model of the first few samples of the weighted signal. This reduces the framing effects in spectral encoding of the windowed signal. The input signal is mapped to the frequency domain using an FFT.

After the FFT operation, the signal is quantized using a lattice VQ approach described in [9]. At the decoder, the truncated ZIR response will be added back to the inverse FFT of the quantized spectral coefficients.

3.2. Encoding of the high-frequencies

The high-frequency signal s_H, with frequency content above 6.4 kHz, is encoded using a bandwidth extension (BWE) approach. The approach consists of extracting a parametric representation, namely the spectral envelope and the gains, which is quantized and sent to the decoder. The fine structure of the high frequency signal is extrapolated at the decoder by using the excitation signal in signal s_I, which is available at the decoder.

The spectral envelope is modeled by an 8-th order LP filter, calculated on the downsampled version of s_H. Hence, the LP filter models the envelope of the spectrally folded high frequency content of the signal. The LP coefficients are transmitted once per frame. The update rate of this LP filter then depends on the mode selection and frame lengths within the super-frame. Gain corrections are computed and transmitted for each sub-frame; these ensure continuity at the 6.4 kHz junction between the lower band and the higher band. Since only a few parameters are transmitted, the total bit rate used for the BWE is as low as 0.8kbps.

3.3. Stereo encoding

For AMR-WB+ stereo coding the same band decomposition as in the mono case is used. The low-band stereo signal coding is done according to a novel semi-parametric technique. The two channels are down-mixed to form a mono signal that is encoded by the AMR-WB+ core codec described in Section 3.1. Additionally, stereo image information is encoded by further decomposing the low-band into two bands (0-1.0 kHz) and (1.0 –6.4 kHz). For the very-low-frequency band a stereo balance factor is derived representing the level ratio between mono and side signal. In order to provide 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. Frame-length candidates are chosen from the total length of one super-frame or subdivisions of length equal to ¼-th, ½-th of the total length of the super-frame.

The high frequency part of the low-band signal is encoded according to a novel shape-gain constrained time-domain filter approach that resembles an inter-channel predictive technique. The new approach overcomes the problems of inter-channel prediction by providing a stable stereo image and leads to a highly efficient representation of the stereo information in the band from 1.0-6.4 kHz. The high-band part (above 6.4 kHz) is encoded by using parametric BWE on the two stereo channels as in Section 3.2.

3.4. Scalability of AMR-WB+

The use of algebraic VQ both in the TCX part of the mono codec and the perceptually most relevant very-low-frequency band of the stereo encoding makes AMR-WB+ highly scalable in terms of the total bit rate and the bit rate distribution between mono and stereo coding.

Allowing scaling of the nominal internal sampling frequency of 12.8 kHz with factors in a range from 0.5 - 1.5 increases the scalability of the codec even further. Scaling of the internal sampling frequency is equivalent to scaling both the total codec bit rate and the coded audio bandwidth. This allows for very-low-rate AMR-WB+ operation at limited bandwidth as well as for high-rate operation (up to 48 kbps) with an audio bandwidth of up to about 20 kHz.

3.5. Complexity of AMR-WB+

At bit rates below 24kbps, the complexity of AMR-WB+ encoder is estimated to 38.3 wMOPS for stereo content creation. This figure is very close to that of the complexity of AMR-WB speech codec (36.6 wMOPS). The stereo decoder complexity at 24 kbps is merely 15.5 wMOPS enabling full audio services with low-cost terminals capable for wideband telephony.

4. AMR-WB+ QUALITY EVALUATION

The quality of AMR-WB+ was evaluated in a range from 14 to 48 kbps with a multitude of experiments according to MUSHRA methodology. For low rates up to 24 kbps a detailed performance comparison with Enhanced AAC+ is given below. For higher rates a monotonous quality increase is maintained up to a level that is close to transparency.

4.1. Test layout

The conducted tests followed the MUSHRA methodology according to the 3GPP audio codec selection plan [7]. Ericsson and Nokia listening laboratories executed the tests independently using the same material. The partial results from both laboratories were combined to form the final results.

The material used for testing originated from the 3GPP low-rate audio codec selection. In accordance with the envisioned audio content of typical wireless audio applications, this set comprises 24 test items in total, containing 8 music, 8 speech, 4 speech-between-music and 4 speech-over-music items. All items were represented as stereo sampled at 48 kHz.

The test conditions comprise 3 stereo codec conditions of the respective codecs operated at 48 kHz output sampling rate, where AMR-WB+ was used at 14 kbps, 18 kbps and 24 kbps and E-AAC+ at 16.1 kbps (minimum stereo rate), 18 kbps and 24 kbps. AMR-WB+ at 24 kbps with output sampling rate of 24 kHz was included as a reference corresponding to the AMR-WB+ operation in the official 3GPP selection tests. Two low-pass filtered (3.5 kHz and 7 kHz) anchor conditions with reduced stereo image (6 dB) were included. Furthermore, the original signal was provided both as open and hidden references. In total 46 experienced listeners were used (Ericsson 20, Nokia 26) to which the test items were presented in random order.

4.2. Results

The overall listening test results are shown in Table 3. The performance of AMR-WB+ and E-AAC+ are graphically introduced in Figure 1, where the results are presented with the 95% confidence intervals. According to the statistical comparison (T-test), AMR-WB+ is statistically better than E-AAC+ in every condition at the same bit rates.

The results show that AMR-WB+ outperforms EAAC+ in the low bit rate range, and show a large superiority margin at rates 14 and 18 kbps. In addition, by examining the performance variation over the audio content categories it appears clearly that the AMR-WB+ provides a consistent quality over all audio content types.

5. CONCLUSION

The mobile environment set strict requirements for multimedia codec bit rates and complexity while the quality expectations for the services remain high. The new 3GPP AMR-WB+ audio codec standard is proven to meet the requirements providing high quality over all audio content types at very low bit rates.

Table 3. The MUSHRA listening test results presented in numerical format.

Condition	Music	Speech over music	Speech	Speech between music	All
Hidden ref.	98.93	98.77	99.28	98.68	98.98
7.0 kHz anchor	50.54	53.39	55.36	56.17	53.56
3.5 kHz Anchor	27.03	28.91	29.91	31.85	29.11
EAAC+ 16kbs	57.65	55.23	48.62	51.25	53.17
EAAC+ 18kbs	63.26	63.09	53.72	55.72	58.79
EAAC+ 24kbs	82.37	77.27	67.95	68.57	74.41
AMR-WB+ 14kbs	53.27	54.63	56.66	65.38	56.64
AMR-WB+ 18kbs	64.35	66.60	68.65	74.10	67.78
AMR-WB+ 24kbs	73.62	77.75	79.13	78.67	76.99
AMR-WB+ 24kbs @24kHz	64.63	70.98	74.52	77.78	71.18

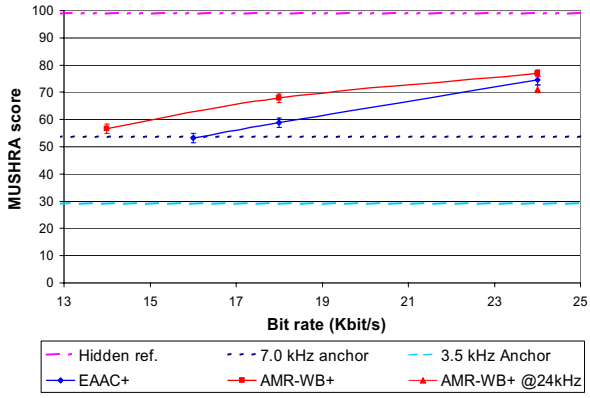

Figure 1. MUSHRA test result for low bit rate stereo

6. ACKNOWLEDGMENT

The authors wish to thank the other 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, Stephane Ragot, Vesa Ruoppila, Patrik Sandgren, Joachim Thiemann, Henri Toukomaa, Tommy Vaillancourt and Janne Vainio.

7. REFERENCES

[1] 3GPP TS 26.290; Extended AMR Wideband codec; Transcoding functions

[2] 3GPP TS 26.401; Enhanced aacPlus General Audio Codec; General Description

[3] 3GPP TS 22.233, "Transparent end-to-end packet-switched streaming service; Stage 1", v. 6.3.0

[4] 3GPP TS 22.246, "Multimedia Broadcast/Multicast Service (MBMS) user services; Stage 1", v. 6.1.0

[5] 3GPP TS 22.140, "Multimedia Messaging Service (MMS); Stage 1", v. 6.6.0

[6] Recommendation ITU-R BS.1534, "Method for the subjective assessment of intermediate quality level of coding systems"

[7] 3GPP Tdoc S4-030824, "AMR-WB+ and PSS/MMS Low-Rate Audio Selection Test and Processing Plan"

[8] 3GPP TS 26.190, "AMR wideband speech codec; transcoding functions"

[9] S. Ragot, B. Bessette and R. Lefebvre, "Low-complexity multi-rate lattice vector quantization with application to wideband speech coding at 32 kbit/s", Proc. IEEE ICASSP-2004, pp. I-501 to I-504, May 2004.