Abstract

The Third Generation Partnership Project (3GPP) and European Telecommunications Standards Institute (ETSI) have carried out development and standardisation of a wideband speech codec for GSM and the third generation mobile communication WCDMA system since 1999. The Adaptive Multi-Rate Wideband (AMR-WB) codec algorithm was selected in December 2000, and the corresponding specifications were approved in March 2001. The AMR-WB codec was jointly developed by Nokia and VoiceAge. AMR-WB extends the audio bandwidth from 3.4 kHz to 7 kHz and gives superior speech quality and voice naturalness compared to existing 2nd and 3rd generation mobile communication systems. The wideband speech service provided by the AMR-WB codec will give mobile communication speech quality that even exceeds (narrowband) wireline quality.

1. Introduction

The current 2nd and 3rd generation mobile communication systems operate with narrow audio bandwidth limited below about 3.4 kHz. Although they give good performance for narrowband speech, the introduction of a wider audio bandwidth of 7 kHz provides substantially improved speech quality and voice naturalness. As Figure 1 shows, speech contains plenty of information above 3.4 kHz which can be reconstructed only by using wideband coding.

Recent advances in speech coding have made wideband coding feasible in the bit-rates applicable for mobile communication. Since 1999 3GPP and ETSI have carried out development and standardisation of a wideband speech codec for the WCDMA 3G and GSM systems. A pre-study phase of the feasibility of wideband coding preceded the launch of standardisation. After almost two years of intense development and two competitive codec selection phases, the 3GPP/ETSI wideband codec algorithm was selected in December 2000. The speech codec specifications were finalised and approved in March 2001. The 3GPP/ETSI wideband codec is an adaptive codec capable of operating with a multitude of speech coding bit-rates. The codec is referred to as Adaptive Multi-Rate Wideband (AMR-WB) codec. The codec was jointly developed by Nokia and VoiceAge.

The AMR-WB codec includes a set of fixed rate speech and channel codec modes, a Voice Activity Detector (VAD), Discontinuous Transmission (DTX) functionality in GSM and Source Controlled Rate (SCR) functionality in 3G, in-band signalling for codec mode transmission, and link adaptation to control the mode selection.

AMR-WB adapts the bit-rate allocation between speech and channel coding, thereby optimising speech quality to prevailing radio channel conditions. While providing superior voice quality over the existing narrowband standards, AMR-WB is also very robust against transmission errors due to the multi-rate operation and adaptation. The adaptation is based on similar efficient principles as in the previously standardised 3GPP/ETSI AMR codec (referred to also as the AMR narrowband codec, AMR-NB).

The AMR-WB codec has been developed for use in several applications: in the GSM full-rate channel, in GSM EDGE Radio Access Network (GERAN) 8-Phase Shift Keying (8-PSK) Circuit Switched channels, in the 3G Universal Terrestrial Radio Access Network (UTRAN) channel, and also in packet based VoIP (voice over IP) applications.

Figure 1: Spectrogram of a speech sentence: “Everyone looked extremely confused about the news”

2. Evolution of GSM and WCDMA 3G speech coding

The 13 kbit/s Full-Rate (FR) codec was the first voice codec defined for GSM. The codec was standardised in 1989 and it is used in the GSM full-rate channel (gross bit-rate of 22.8 kbit/s). The 5.6 kbit/s Half-Rate (HR) codec was standardised in 1995 to provide channel capacity savings through operation in the half-rate channel (gross bit-rate of 11.4 kbit/s).

The 12.2 kbit/s Enhanced Full-Rate (EFR) codec was the first GSM codec to provide voice quality equivalent to that of a wireline telephony [1]. The EFR codec brought substantial quality improvement over the two previous GSM codecs. EFR was standardised first for the GSM-based PCS 1900...
system in the US during 1995 and was adopted to GSM in 1996. The codec was jointly developed by Nokia and the University of Sherbrooke.

A further development in GSM voice quality was the standardisation of the AMR narrowband codec in 1999 [2]. The AMR-NB codec offers major improvement over EFR in error robustness in the full-rate channel by adapting speech and channel coding depending on prevailing channel conditions. By switching to operate in the half-rate channel during good channel conditions, AMR also gives channel capacity gain over the EFR codec. AMR contains eight speech coding bit-rates between 4.75 and 12.2 kbit/s. The AMR codec was adopted in 1999 by 3GPP as the mandatory speech codec for the WCDMA 3G system. The AMR codec was jointly developed by Ericsson, Nokia and Siemens.

The AMR wideband codec is the most recent voice codec standardised for GSM and WCDMA 3G systems. While all previous codecs in mobile communication operate on narrow audio bandwidth limited below 3.4 kHz, AMR-WB extends the audio bandwidth to 7 kHz bringing as a result substantial quality improvement. The AMR-WB codec, like the AMR-NB codec, is an adaptive codec consisting of several modes. The codec operates on the speech coding bit-rates between 6.6 and 23.85 kbit/s. Like the AMR-NB codec, AMR-WB has also a low bit-rate source dependent mode for coding background noise. The AMR-WB codec was jointly developed by Nokia and VoiceAge.

3. AMR-WB codec standardisation

The AMR-WB codec standardisation was carried out as a competitive selection process consisting of two phases: Qualification Phase (spring 2000) and Selection Phase (June - October 2000). Seven AMR-WB candidate codecs were submitted for the Qualification Phase. The five best codecs proceeded into the Selection Phase.

In the Selection Phase, the codec candidates were tested in detail in six independent test laboratories. Testing was coordinated internationally and was carried out with multiple languages: Japanese, English, French, Mandarin Chinese, and Spanish. Each experiment in the tests was performed with two languages to avoid any bias due to a particular language. The tests covered speech with and without background noise, channel errors, mode adaptation and also source controlled rate operation. The candidate codecs were implemented in C-code with fixed-point arithmetic.

Based on the test results and technical details of the codec proposals, the Nokia/VoiceAge codec was selected as the AMR-WB codec in December 2000. Since then the speech codec specifications have been finalised and they were approved in March 2001.

Based on the good test results, ITU-T has approved the 3GPP/ETSI AMR-WB codec to participate in ITU-T low bit-rate wideband codec (around 16 kbit/s) selection tests as a candidate. The selection of the ITU-T codec will take place during 2001. This could lead to one harmonised wideband codec for GSM, 3G WCDMA and ITU-T.

4. AMR-WB speech codec

The AMR-WB speech codec utilises the ACELP (Algebraic Code Excitation Linear Prediction) technology which is employed also in the AMR-NB and EFR speech codecs [3]. The AMR-WB speech codec consists of nine speech codec modes with bit-rates of 23.85, 23.05, 19.85, 18.25, 15.85, 14.25, 12.65, 8.85 and 6.6 kbit/s [4]. AMR-WB includes also a background noise mode which is designed to be used in discontinuous transmission (DTX) operation in GSM and as a low bit-rate source dependent mode for coding background noise is other systems. In GSM the bit-rate of this mode is 1.75 kbit/s.

Figure 2 shows the possible operating scenarios for different bit-rates in different systems. The 12.65 kbit/s mode and the modes above it offer high quality wideband speech. The two lowest modes 8.85 and 6.6 kbit/s are intended to be used only temporarily during severe radio channel conditions or during network congestion.

The AMR-WB codec operates at a 16 kHz sampling rate. Coding is performed in the blocks of 20 ms. Two frequency bands, 50 – 6400 Hz and 6400 – 7000 Hz, are coded separately in order to decrease codec complexity and to focus the bit-allocation into the subjectively most important frequency range. Note that already the lower frequency band goes far above narrowband telephony.

The lower frequency band is coded with the ACELP algorithm. Linear Prediction (LP) analysis is performed once per 20 ms frame. A fixed codebook (ACELP-codebook) and an adaptive codebook are searched every 5 ms for optimal codec parameter values. The processing is carried out at a 12.8 kHz sampling rate.

The higher frequency band is reconstructed in the decoder using the parameters of the lower band and a random excitation. No information about the higher band is transmitted, except in the 23.85 kbit/s mode, where the higher band gain is transmitted. In other modes, the gain of the higher band is adjusted relative to the lower band using voicing information. The spectrum of the higher band is reconstructed by using a wideband LP filter generated from the lower band LP filter.

The AMR-WB codec is defined in fixed-point arithmetic using a set of basic operations defined by 3GPP/ETSI. The computational complexity of the AMR-WB speech codec is 35.4 WMOPS (Weighted Million Operations Per Second) which is around twice the complexity of the AMR-NB codec. The RAM memory requirement is slightly higher compared to AMR-NB. The ROM memory requirement is substantially lower than in AMR-NB due to higher code reuse between the modes. The complexity of the AMR-WB speech codec is show in Table 1. (For comparison, the complexity of AMR-NB speech codec is also shown in the table.)
5. AMR-WB channel codec

In all the GSM channel codecs, error protection is based on punctured Recursive Systematic Convolutional (RSC) codes. Each codec mode employs CRC (Cyclic Redundancy Check) for error detection. Speech encoded bits are divided into two different protection classes: Class 1a, where bits are protected by the convolutional code and CRC; and Class 1b, where bits are protected by the convolutional code alone. Class 1a contains bits that would cause substantial degradation to subjective speech quality if fed into speech decoder without bad frame substitution. Channel coding polynomials and interleaving were adopted from the previous GSM codecs in all the modes to maximise commonality with the existing GSM system.

For use in 3G UTRAN channels, a general channel coding toolbox of the 3G WCDMA system is used (as for the AMR-NB codec).

6. Adaptive operation in GSM channel

AMR-WB has high granularity of bit-rates between 6.6 and 23.85 kbit/s. For GSM channels, this makes it possible to maximise speech quality by adapting the codec bit-rate to increase robustness against transmission errors. For non-adaptive 3G UTRAN channels using fast power control, operators can select the suitable bit-rates to make an optimal trade-off between the speech quality and network capacity.

The link adaptation process bears responsibility for measuring the channel quality and selecting the most appropriate mode according to prevailing channel conditions [5]. Link adaptation also takes into account constraints for available bit-rate set by the network (e.g., network load).

In-band signalling (400 bit/s in full-rate channel) transmits both the requested mode for the reverse link (based on channel quality measurements) and active codec mode of the forward link over the air interface to the receiver side. For a more detailed description of the adaptation see [6].

Figure 3 shows an example of how the codec mode adaptation works in the GSM full-rate channel when the Carrier to Interference ratio (C/I) varies between 22 and 2 dB. Based on estimated channel quality (e.g. C/I), one out of the activated codec modes (14.25, 12.65, 8.85, or 6.6 kbit/s) is chosen, resulting in high speech quality throughout the call.

7. AMR-WB VAD

The AMR-WB codec includes a Voice Activity Detection (VAD). VAD allows the codec to switch to a lower-rate mode for coding of background noise. This feature saves power in the mobile station and also reduces the overall interference level over the air interface.

The AMR-WB VAD is based on dividing the speech signal in each 20 ms speech block into sub-bands and carrying out analysis for each band. Background noise level is estimated in each frequency band. Intermediate VAD decision is calculated by comparing input SNR (ratio of the signal level and the background noise level) to an adaptive threshold. The adaptation of the threshold is based on noise and long term speech estimates. Final VAD decision is calculated by adding hangover period to the intermediate VAD decision.

VAD contains a tone detection function which indicates presence of a signalling tone, voiced speech, or other strongly periodic signal. The tone detection function is based on the normalised open-loop pitch gains which are calculated by open-loop pitch analysis of the speech encoder.

8. Speech quality of AMR-WB

During the selection tests, the AMR-WB codec was tested in a variety of test conditions in six independent test laboratories with five languages. The selection tests included 6 experiments and 19 sub-experiments. The experiments were:

1) Input level and tandeming of codecs
2) Clean speech performance with static errors
3) Car and street noise (15 dB SNR) performance for the GSM FR channel
4) Car and street noise (15 dB SNR) performance for higher-rate channels (EDGE, 3G WCDMA)
5) Performance in dynamic error conditions (codec mode adaptation on) in GSM
6) Operation of VAD/DTX in GSM

The testing covered the application of AMR-WB in GSM (including EDGE full-rate and half-rate channels) and in 3G UTRAN WCDMA. The AMR-WB codec showed very good performance in the selection tests: it met performance requirements in all of the laboratories throughout the tests.

The quality of the AMR-WB codec is described in the following sub-sections based on the selection test results [7].

8.1. 3G UTRAN and GSM EDGE channels

In these applications, the highest modes of AMR-WB can be used, and the codec is able to provide speech quality equal to the very high quality reference of ITU-T 64 kbit/s wideband codec (G.722-64k). This quality is obtained for clean speech as well as for speech within the presence of background noise.

Performance under transmission errors (for both clean speech and speech in background noise) is as follows:

- In the 3G UTRAN channel, under transmission errors at 1% Frame Error Rate (0.1% Residual Bit Error Rate), quality is still at least equal to G722-48k.
In the EDGE FR-channel, at 22 dB C/I and above quality is at least equal to error-free G.722-56k. At 16 dB C/I, quality is still at least equal to error-free G.722-48k. (In the EDGE HR-channel, the same performance is obtained at 3 dB higher C/I-ratios.)

Figure 4 presents an example from the selection test results showing the AMR-WB performance in 3G WCDMA channel (tests in Japanese).

![Figure 4: Speech quality of AMR-WB codec in 3G UTRAN channel (no restrictions for modes)](image)

8.2. GSM Full Rate channel

AMR-WB gives quality at least equal to G.722-56k. For erroneous channel at 13 dB C/I, quality is still at least equal to the quality of error-free G.722-48k. Below 13 dB C/I, smooth degradation is provided. This quality is obtained for clean speech as well as for speech in background noise.

Figure 5 shows an example of the performance of AMR-WB in GSM FR channel under various error conditions for clean speech (test in English).

Figure 6 shows an illustrative graph on speech quality in the GSM FR channel. AMR-WB is compared to narrow band codecs AMR-NB and EFR. In typical operating conditions (C/I > 10 dB), AMR-WB gives superior speech quality over all other GSM codecs. Even in the very poor radio channels (C/I ≤ 7 dB), it still offers comparable quality to AMR-NB and far exceeds the quality of the fixed rate codecs (like EFR).

![Figure 5: Speech quality of AMR-WB codec in GSM FR channel (max bit-rate 14.25 kbit/s used)](image)

9. Conclusions

AMR-WB extends the audio bandwidth to 7 kHz and gives superior speech quality and voice naturalness compared to existing codecs in 2nd and 3rd generation mobile communication systems. The introduction of AMR-WB brings a fundamental improvement of speech quality, raising it to a level never experienced in mobile communication systems before. It far exceeds the current high quality benchmarks for narrowband speech quality and changes the expectations of a high quality speech communication in mobile systems.

10. References

[5] 3GPP TS 45.009 “Adaptive Multi-Rate inband control and link adaptation”, 3GPP technical specification