

Wideband Speech Coding Standards and Applications

Abstract

Increasing the bandwidth of sound signals from the telephone bandwidth of 200-3400 Hz to the wider bandwidth of 50-7000 Hz results in increased intelligibility and naturalness of speech and gives a feeling of transparent communication. The emerging end-to-end digital communication systems enable the use of wideband speech coding in a wide area of applications. Recognizing the need of high quality wideband speech codecs, several standardization activities have been recently conducted, resulting in the selection of a new wideband speech codec, AMR-WB, at bit rates from 6.6 to 23.85 kbit/s by both 3GPP and ITU-T. The adoption of AMR-WB by the two bodies is of significant importance since for the first time the same codec is adopted for wireless as well as wireline services. This will eliminate the need for transcoding, and ease the implementation of wideband voice applications and services across a wide range of communication systems and platforms.

This document presents a summary of wideband speech coding standards for wideband telephony applications. The quality advantages and applications of wideband speech coding are first presented, and the issue of telephony over packet networks is discussed. Several wideband speech coding standards are discussed and a special emphasis is given to the AMR-WB standard recently selected by 3GPP and ITU-T.

1. Introduction

Most speech coding systems in use today are based on telephone-bandwidth narrowband speech, nominally limited to about 200-3400 Hz and sampled at a rate of 8 kHz. This limitation built into the conventional telephone system dates back to the first transcontinental telephone service established between New-York and San Francisco in 1915. The inherent bandwidth limitations in the conventional public switched telephone network (PSTN) impose a limit on the communication quality. The increasing penetration of the end-to-end digital networks, such as the second and third generation wireless systems, ISDN, and voice over packet networks, will permit the use of wider speech bandwidth that will offer a communication quality significantly surpassing that of PSTN and gives the sensation of face-to-face communication. Most of the energy in speech signals is present below 7 kHz although it may extend to higher frequencies particularly on unvoiced sounds. In wideband speech coding, the signal is sampled at 16 kHz and bandlimited to 50-7000 Hz, which results in a speech quality close to face-to-face communications quality.

Wideband speech coding results in major subjective improvements in speech quality. Compared to narrowband telephone speech, the low-frequency enhancement from 50 to 200 Hz contributes to increased naturalness, presence and comfort. The high-frequency extension from 3400 to 7000 Hz provides better fricative differentiation and therefore higher intelligibility. A bandwidth of 50 to 7000 Hz not only improves the intelligibility and naturalness of speech, but adds also a feeling of transparent communication and eases speaker recognition.

Figure 1 shows a typical energy spectrum of a narrowband and wideband voiced speech signal. A similar energy spectrum of an unvoiced speech signal is shown in Figure 2. In unvoiced speech periods, the important energy present above 4 kHz is filtered out in case of narrowband speech. This effects speech intelligibility such as differentiation between “s” and “f”. In voiced periods, most of the energy is present at low frequencies, and filtering out the energy below 200 Hz affects speech naturalness.

VoiceAge Corporation and/or its suppliers make no representations about the suitability, accuracy or completeness of the information contained in this document. This Document is provided AS IS without any warranty of any kind, either expressed or implied. In no event will VoiceAge Corporation and/or its suppliers be liable to you for any special, indirect or consequential damages or any damages whatsoever resulting from the loss of use, data, profits or savings, arising out of the use or inability to use the information contained in this document, without limiting the foregoing, VoiceAge Corporation and its suppliers disclaim any express or implied warranties of merchantability or fitness for any particular purpose, title and non-infringement. VoiceAge Corporation and/or its suppliers may make improvements and/or changes in the document at any time without notice.

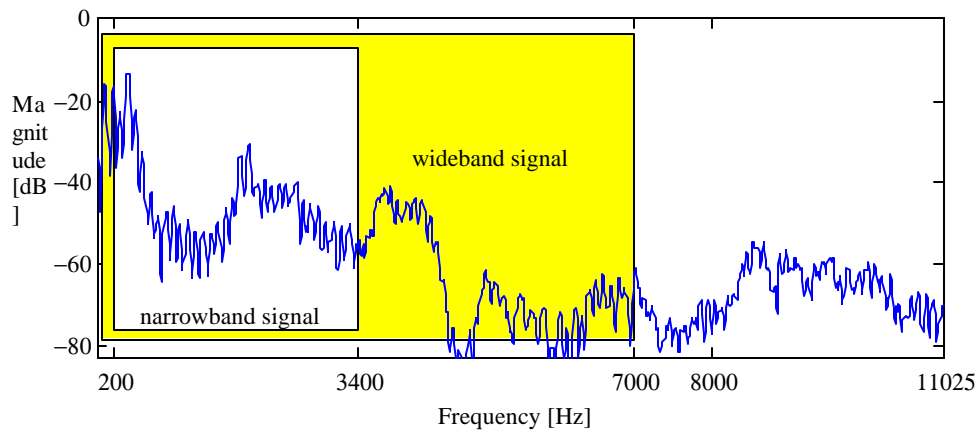


Figure 1: Example of the energy spectrum of a voiced speech segment.

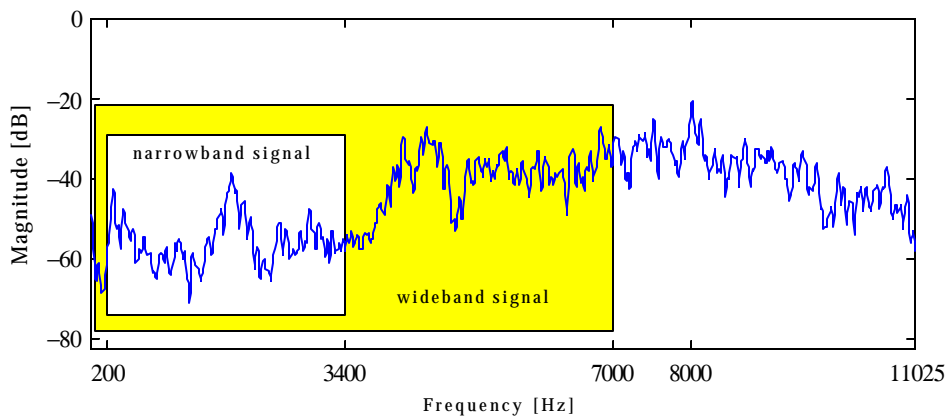


Figure 2: Example of the energy spectrum of an unvoiced speech segment.

Wideband speech signals are sampled at 16000 Hz. When each sample is represented by 16-bit integers, the resulting raw bit rate becomes 256 kbit/s. Thus speech coding, or speech compression, becomes of significant importance for wideband speech communications in a viable application. Different activities have been recently conducted for standardization of wideband speech codecs around 16 kbit/s. These activities resulted in the selection of the adaptive multi-rate wideband codec (AMR-WB) by both 3GPP/ETSI and ITU-T. The adoption of new speech coding standards, enhanced terminal acoustics, and meeting the quality of service (QoS) challenge in voice over packet networks are helping to prepare the ground for wideband telephony.

In this document, we give an overview of wideband speech coding and wideband telephony including wideband speech coding applications and standards. The AMR-WB standard will be treated with more details given its potential importance in emerging applications.

2. Wideband speech coding applications

The naturalness of wideband speech coding is a significant feature in high fidelity telephony and for extended telecommunications processes such as audio teleconferencing and program broadcasting. Several application areas of wideband speech are briefly discussed below. The issue of wideband telephony over IP-based packet networks is also discussed.

Third generation mobile communication systems:

Delivering multimedia services is one of the main goals of 3G wireless communications. This implies the use of high quality audio and speech in multimedia content. Even in voice telephony applications in 3G, wideband speech is an important step for wireless service providers to deliver speech quality surpassing that of traditional PSTN. The importance of wideband speech has been recognized by 3GPP which had recently selected the AMR-WB codec that will be discussed below in more details.

High fidelity telephony over broadband packet networks and ISDN:

Broadband packet networks and ISDN are end-to-end digital networks that enables the delivery of high fidelity wideband telephony, offering service providers a leading edge over traditional narrowband PSTN. These packet networks include xDSL, Packet Cable, ATM, Frame Relay, and broadband ISDN. IP-based protocols used for transporting data can be used for transporting real-time voice signals. Some QoS issues related to voice over IP (VoIP) will be discussed below.

Audio and video teleconferencing:

Wideband speech coding gives audio and video teleconferencing improved sound quality and presence of speakers, and a more realistic rendering of the actual sound scene over ISDN or packet-based networks.

Internet applications:

Wideband telephony can enhance several Internet applications such as: broadcasting and streaming, chat and virtual reality immersion environments, multimedia real-time collaboration tools, archiving and distribution of narrative content, and network based languages learning applications.

Digital radio broadcasting:

Wideband speech can be used in digital AM radio broadcasting, audio and video broadcasts of news, TV programs, and closed circuit lectures.

Wideband telephony over packet networks

In the early deployment of the second generation mobile telephone systems, speech quality degradation was tolerated since mobility was the most important issue. As these systems were turning into mass-market with hundreds of millions of users worldwide, insuring a speech quality equivalent to that delivered by PSTN became of significant importance. For this reason, enhanced quality coders were adopted to replace the initial coders. In the year 2002, the third generation wireless systems are expected to be launched, with a main goal to deliver high speed packetized data for mobile multimedia applications,



with number of users expected to exceed a billion in next few years. A speech quality exceeding that of PSTN becomes an increasing prospective. Here, the introduction of wideband telephony with a signal bandwidth of 50–7000 Hz compared to the narrowband telephone bandwidth of 300–3400 Hz will bring a new user experience approaching the face-to-face communication quality.

Voice over packet network (VoPN) applications are also gaining much interest. In the near future, most voice traffic will be transported through packet networks such as IP, ATM, Frame Relay, DSL, and Packet Cable. Most of these networks were initially designed for transporting data. It was initially thought that voice can be transported over these networks as any other form of data. However, it was realized that transporting real-time voice over packet networks was not an easy task due to the lack of procedures that guarantee the performance in the network and lack of specific characteristics for real-time voice needs. In circuit-switched networks, several procedures have been put in place for guaranteeing high quality of service for voice communications. These procedures can be found in ITU-T G-series Recommendations which cover all network aspects such as transmission plan (G.101), loudness (G.111), impairments (G.113), transmission delay (G.114), and talker echo (G.131, G.163). Further, ITU-T SG-12, with its speech quality expert group (SQEG), plays an important role in drafting recommendations for measurements of the network parameters and characteristics, and can be found in the P-series Recommendations. Of special interest for voice quality evaluation are the subjective and objective quality testing measures found in Recommendations P.800 and P.862, respectively.

In VoPN, the parameters that affect the quality of service (QoS) are the delay, jitter, packet loss, and quality degradation due to voice compression. Thus, evaluation of the QoS in VoPN is more complex than that of circuit-switched networks. In circuit-switched networks, companded-PCM at 64 kbit/s is mainly used for digitizing speech giving almost a transparent quality. Further, the channel impairment, delay and jitter are properly controlled, mainly because of the deterministic management due to the reservation of a dedicated circuit compared with the statistical management of most packet networks. Another parameter which has a significant impact on the speech quality, and has been often overlooked, is the input signal bandwidth. With VoPN, which are end-to-end digital networks, wideband telephony can be easily introduced offering a leading edge over traditional narrowband telephone networks.

In packet switched networks, the issue of codec robustness to packet network environment is of significant importance. The data networks world still sees codecs as “black boxes”. For a few years, we have been promised that VoIP is coming, but this still has to be seen. Service providers are still reluctant to embrace VoIP mainly because of quality of service issues. In order for VoIP to take off, it must be cheaper and have a quality equivalent or better than PSTN. Concurrently PSTN is getting cheaper and VoIP is still far from offering equivalent quality. Network experts are trying to solve the problem by introducing intelligence to the routers and gateways (e.g., DiffServ, RSVP) which will make the solution more expensive and statistical resource management less efficient.

The solution is to bring the intelligence for guaranteeing a good QoS to the codec “black box” mainly solving the packet loss and jitter problem. By enabling also wideband speech services in packet networks, the quality of PSTN can be exceeded.

Wideband speech coding with efficient error concealment and jitter management will be a major competitive factor for VoIP services. With the deregulation in the telecommunications market, new service providers (e.g., voice over xDSL, Packet Cable, wireless local loop) will need this solution for competing with the established PSTN services. Service providers for these networks will be able to provide wideband telephony with high-speed data services at low costs.

3. Wideband speech coding standards

The ITU-T was the first standard body to conduct early work on speech coding. The first ITU-T recommendation on speech coding, G.711, was completed in 1972. This recommendation covers μ -law or A-law pulse code modulation (PCM) at 64 kbit/s and is widely in use in PSTN telephony. The ITU-T Recommendation G.726 followed in 1984 for speech coding at 32 kbit/s using adaptive differential PCM, ADPCM [1] (it was initially G.721). Later ITU-T narrowband speech coding recommendations are G.728 at 16 kbit/s (1992) [2], G.729 at 8 kbit/s (1995) [3, 4] and G.723.1 at 6.3 and 5.3 kbit/s (1995) [5]. ITU-T recommendations are generally developed for a wide variety of applications. In practice, however, regional standardization bodies such as ETSI and TTA have been leading the development of speech coding standards for mobile communication systems.

Regional digital cellular standards have played a crucial role in the advancement of wireless second generation communication systems. For the GSM mobile communication system developed within ETSI, several coders were standardized including the full-rate (FR) codec in 1987, the enhanced full-rate codec (EFR) in 1996 [6] and the adaptive multi-rate (AMR) codec in 1999 [7]. Also in TTA, several codecs have been developed for North-American digital cellular standards. The latest standards are IS-641 at 7.4 kbit/s for NA-TDMA, IS-127 at 8.5, 4.0 and 0.8 kbit/s for NA-CDMA and the recent selectable mode vocoder (SMV) at 8.5, 4.0, 2.0 and 0.8 kbit/s for NA-CDMA and CDMA2000.

Speech coding standardization is usually conducted in several phases and may span over a few years period. It starts with setting “terms of reference” which define several requirements and objectives in relation with the designated applications. These requirements include the bit rate, delay, complexity, quality under different operating conditions (clean, random errors, frame erasures, background noise, tandeming, etc.) A qualification phase is first conducted where candidate codecs are individually tested with a limited set of conditions to show their potential to meet the requirements. The qualified candidates are then tested in a selection phase for conditions reflecting the requirements and some objectives. The candidate that better meets the requirements and objectives is selected. A characterization phase is later conducted in which other practical operating conditions not present in the selection phase are tested (objectives, transcoding with other standards, etc.) The tests in the different phases are performed using formal subjective test procedures with different languages (e.g. ITU-T Recommendation P.800 [8]).

Although most of the effort on speech coding focused on narrowband speech, several organizations recognized the quality difference available by allowing the input speech to cover a larger bandwidth. The ITU-T established the first wideband speech coding standard, G.722, in 1988 [9]. G.722 is simple to implement and achieves good performance at rates of 48, 56, and 64 kbit/s. In 1995, ITU-T started another activity for wideband coding which results in 1999 in the standardization of G.722.1 at 24 and 32 kbits [10]. More recently, new wideband speech coding activities have been undertaken in ITU-T and ETSI/3GPP for coders at bit rates around 16 kbit/s. A new codec, known as AMR-WB (adaptive multi-rate wideband) codec was selected in 2000 by 3GPP for wideband coding in GSM and 3G wireless systems. The AMR-WB codec participated in the ITU-T activity and was selected in 2001 as the winning candidate in the ongoing standardizing process in ITU-T for a wideband coder at rates of 13 to 24 kbit/s.

In this section, these wideband speech coding standards, which are summarized in Table 1, will be briefly described with more emphasis on the new AMR-WB standard.

Table 1: Summary of wideband speech coding standards

Standard	G.722	G.722.1	AMR-WB (likely G.722.2)
<i>Date</i>	1988	1999	2000
<i>Bit rate(kbit/s)</i>	48, 56, 64 (embedded)	24, 32	23.85, 23.05, 19.85, 18.25, 15.85, 14.25, 12.65, 8.85, 6.6
<i>Type</i>	Sub-Band ADPCM	Transform Coding	Algebraic Code Excited Linear Prediction (ACELP)
<i>Delay</i>			
<i>Frame size</i>	0.125 ms	20 ms	20 ms
<i>Lookahead</i>	1.5 ms	20 ms	5 ms
<i>Quality</i>	Commentary (at 64 kbit/s)	Poor speech performance in some operating conditions; scope of standard limited to hands-free and low packet loss rates Good music performance	Good speech performance at rates 12.65 kbit/s and higher 15.85 \geq G.722 @ 56 23.05 \geq G.722 @ 64 Poor music performance
<i>Complexity</i>	10 MIPS	< 15 MIPS	38 WMOPS
<i>RAM</i>	1 K	2 K	5.3 K
<i>Fixed-point</i>	Bit-exact	Bit-exact C	Bit-exact C
<i>Floating-point</i>	None	Exists in Annex B	In preparation
<i>VAD/DTX/CNG</i>	None	None	Exist
<i>Principle</i>	ISDN;	Same + VoPN	3G wireless;
<i>Applications</i>	Video Conferencing		+ Same as G.722.1

3.1. ITU-T G.722 STANDARD

G.722 is the 64 kbit/s ITU-T standard for wideband applications at a 16 kHz sampling rate which was recommended in 1988. It is essentially a two-subband coder with ADPCM coding of each subband signal utilizing the techniques similar to the G.726 narrowband standard. At the encoder, the speech signal is sampled at a rate of 16 kHz and decomposed into two subbands of equal bandwidth. Each subband signal is downsampled by a factor of two prior to encoding. The subband decomposition is performed using Quadrature Mirror Filters (QMF) with finite impulse response. For the G.722 standard, the QMF have 24 coefficients. This configuration introduces a total delay of 3 ms, and an overall distortion of less than 1 dB over the whole frequency range. At the decoder, the quantized subband signals are upsampled by a factor of two using the same filters than at the encoder to decompose the signal into subbands. With QMF, this ensures that the aliasing due to overlapping subbands cancels out (neglecting the quantization stage). The reconstructed subband signals are then added together to form the synthesis signal.

It was found that the bit allocation which yields the best results is 6 bits/sample for the lower band and 2 bits/sample for the higher band. To allow for transmission of for data (e.g., fax, modem) and speech on the same channel, the lower band resolution is variable, and can take the values 6, 5 or 4 bits/sample. The higher band resolution is fixed at 2 bits/sample. Each reduction of 1 bit/sample in the lower band frees 8 kbit/s for data transmission. Hence, the coder can operate in three different modes: 64 kbit/s audio transmission with no data transmission; 56 kbit/s audio transmission with 8 kbit/s data transmission; and 48 kbit/s audio transmission with 16 kbit/s data transmission.

In recent wideband coding standardization activities, the G.722 standard is used as a reference to judge the quality of new lower bit rate wideband coders.

3.2. ITU-T G.722.1 STANDARD

G.722.1 was standardized in 1999 at bit rates of 24 and 32 kbit/s. The activity started in 1995 aiming initially at speech and audio coding at bit rates of 16, 24, and 32 kbit/s. The requirements were set so that the quality of the new standard at these three bit rates would be equivalent or better than G.722 at 48, 56, and 64 kbit/s, respectively, in a large variety of operating conditions. The operating conditions covered for instance clean speech, speech with background noise, music, tandeming, frame erasures and level variation. It is important to note that the requirements were set for both speech and audio signals. Initially the standard was planned to have two modes, Mode A with low delay (10 ms frames) and Mode B with low complexity (less than 15 MIPS with 20 ms frames and 20 ms lookahead). Later, both modes were merged in Mode B while relaxing the complexity constraint. During this exercise, it became apparent that no single technology was able to fulfill the requirements for both speech and audio signals. Candidate coders using linear prediction based technology showed the potential to meet the requirements for speech while exhibited poor performance for audio signals since they relied on speech source modeling. On the other hand, transform coding based techniques delivered good audio performance but were incompetent on speech signals.

After a few rounds of qualification and selection tests, and recognizing the difficulty to meet all requirements for both speech and audio, the requirements were adjusted. The bit rate of 16 kbit/s was also dropped. The new requirements were set so that the quality of the new standard at 24 and 32 kbit/s would be equivalent or better than G.722 at 48 and 56 kbit/s, respectively. Two candidate codecs were competing at the end, NTT's candidate based on code-excited linear prediction (CELP) coding and PictureTel's based on transform coding. NTT's proposal performed well on speech signal while giving a somewhat compromised quality on general audio signals. The codec had, however, a complexity close to 200 MIPS rendering it hardly viable. On the other hand PictureTel's proposal met all the requirements only on audio signals while having a complexity below 15 MIPS. Although not all the requirements were met by either candidate, the pressure from the industry encouraged to complete the standard, and PictureTel's proposal was finally selected. The complexity was a major factor in the selection. However, the scope of the recommendation was limited to reflect the fact that the requirements were not met in all operating conditions. The standard is entitled "*Coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss*". This reflects the fact that it met the requirements for background noise conditions and failed the 3% frame erasure conditions and some clean speech and tandeming conditions.

G.722.1 uses 20 ms frames with a 20 ms lookahead. A modulated lapped transform (MLT) with 40 ms overlapping windows is used. The transform coefficients are grouped in 500 Hz bands and the gains in the bands are quantized and Huffman coded. The scaled coefficients are quantized using a perceptual categorization procedure and then Huffman coded.

The standard is described in C code with fixed point arithmetic using a set of fixed-point 32-bit basic operators. G.722.1 Annex B contains a floating point version of the standard.

It should be noted that no characterization tests have been performed for G.722.1. Characterization tests are used to quantify the performance in typical practical operating conditions including tandeming with other standardized codecs. The characterization test usually extend the earlier tests conducted in the selection phase. As another limitation of the G.722.1 recommendation, no work has been done in the ITU-T for developing a VAD/DTX/CNG algorithm for it.

For a demonstration of G.722.1 quality, a subset of the selection test results is shown below. Figure 3 shows some results from Experiment 1a (British English) for the codec at 24 kbit/s in clean, tandem, and frame erasure conditions.

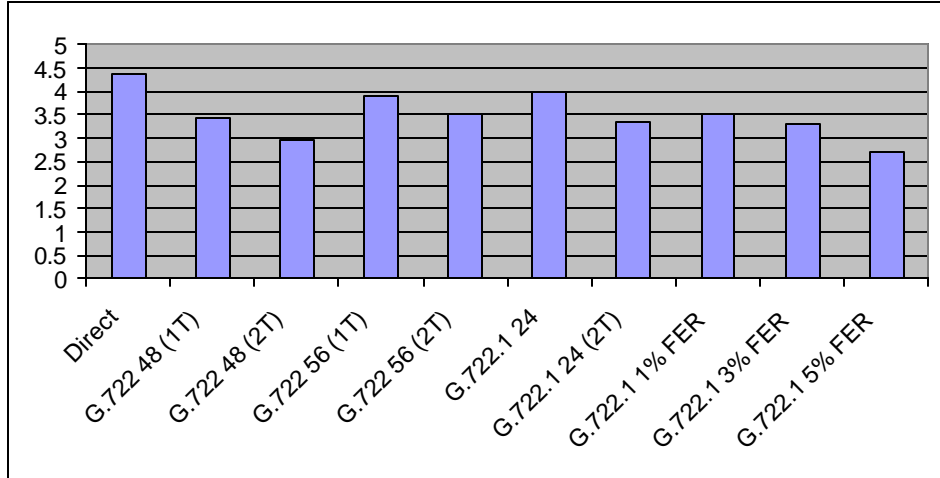


Figure 3: G.722.1 performance. Extracted from ITU-T Selection test (Experiment 1a with British English).

Figure 4 shows some results from Experiment 1b (Canadian English) for the codec at 32 kbit/s in clean, tandem, and frame erasure conditions. In Figures 3 and 4, “Direct” refers to the 16 bit/sample uncoded signal, “1T” and “2T” refer to single coding and two tandeming, “FER” refers to frame erasure rate.

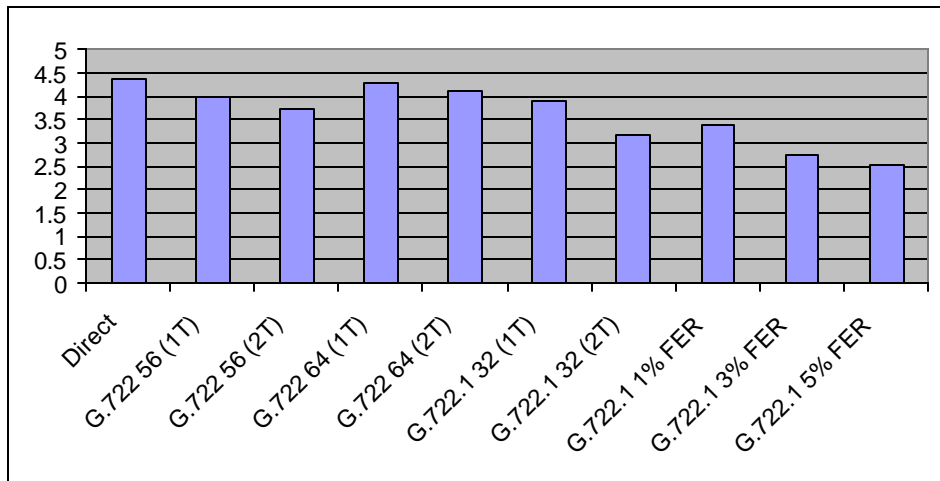


Figure 4: G.722.1 performance. Extracted from ITU-T Selection test (Experiment 1b with Canadian English).

3.3. THE 3GPP/ETSI AND ITU-T AMR-WB STANDARD

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 standardization of a wideband speech codec for the WCDMA 3G and GSM systems. After almost two years of intense development and two competitive selection phases, the 3GPP/ETSI wideband codec algorithm was selected in December 2000. The speech codec specifications were finalized and approved in March 2001. The 3GPP/ETSI wideband codec is an adaptive codec capable of operating at a multitude of speech coding bit rates from 6.6 to 23.85 kbit/s. The codec is referred to as the Adaptive Multi-Rate Wideband (AMR-WB) codec [11].

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 [12,13,14], in-band signaling for codec mode transmission, and link adaptation to control the mode selection. The AMR-WB codec adapts the bit allocation between speech and channel coding, optimizing 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 principles as the previously standardized 3GPP/ETSI AMR narrowband codec, referred also to as AMR-NB.

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

3.3.1. Standardization of the AMR-WB Codec

The AMR wideband codec, jointly developed by Nokia and VoiceAge, was standardized for GSM and WCDMA 3G systems in 2001. While all the previous codecs in mobile communication operate on narrow audio bandwidth limited to 200–3400 Hz, AMR-WB extends the audio bandwidth to 50–7000 Hz bringing substantial quality improvement. The AMR-WB codec operates on nine speech coding bit-rates between 6.6 and 23.85 kbit/s. Like the other GSM and WCDMA 3G codecs, AMR-WB has also a low bit-rate source dependent mode for coding background noise.

The standardization of the AMR-WB codec was launched in mid-1999. Prior to that, a feasibility study phase had been carried out in ETSI during spring 1999 on the applicability of wideband coding for mobile communication. The results showed that the target is feasible and, consequently, the standardization was started. The work was carried out as a joint effort in ETSI and 3GPP targeting the development of a wideband speech service for both the second and third generation mobile communication systems.

After the launch of standardization, detailed speech quality requirements and design constraints covering for example implementation complexity and transmission delay were defined for the codec. The selection of the AMR-WB codec was then carried out as a competitive process consisting of two phases: the Qualification Phase in spring 2000 and the Selection Phase in June-October 2000. From altogether nine codec candidates, seven codecs were submitted to the Qualification Phase. The five best codecs proceeded into the Selection Phase.

In the Selection Phase, the codec candidates were tested thoroughly in six independent test laboratories. Testing was coordinated internationally and was conducted with five languages. Each experiment in the tests was performed in two languages to avoid any bias due to a particular language. The tests contained a wide set of operating conditions covering clean speech, background noise, channel errors, and mode adaptation conditions, and also source controlled rate operation. The candidate codecs were implemented in C-code with fixed-point arithmetic using the same basic operators used to define the previous ETSI and 3GPP codecs.

Based on the test results and technical details of the codec proposals, the Nokia/VoiceAge codec was selected as the 3GPP/ETSI AMR-WB codec in December 2000. Since then the speech codec specifications have been finalized and they were approved in March 2001.

After approval of the codec specifications, a Characterisation Phase took place. During this phase, the AMR-WB codec was subjected to extensive testing in various operating conditions. The results are summarized in 3GPP Technical Report.

3.3.2. Standardization of ITU-T codec around 16 kbit/s

Since a bit rate of 16 kbit/s was discarded during the ITU-T wideband standardization activity that led to the adoption of G.722.1, it was recognized that a new activity needs to be launched for a standard operating at around this bit rate. However, capitalizing on the past experience that showed the difficulty of a single coding technology to perform well for both speech and audio signals, it was decided to set the

requirements for speech signals only. Four bit rates were identified, around 13, 16, around 18, and 24 kbit/s. The requirements were set so that quality at 16 kbit/s be better than G.722 at 48 kbit/s, and quality at 24 kbit/s be equal or better than G.722 at 56 kbit/s. The delay is 20 ms speech frames with 10 ms lookahead. The work is now conducted in Question 7 of ITU-T Study Group 16 (Q.7/16).

The ITU-T wideband speech coding standardization was undertaken in parallel with the 3GPP/ETSI wideband coding activity described above. The ITU-T recognized the importance of the harmonization of their efforts with 3GPP for eliminating the quality degradation due to the need of transcoding in case a communication is transported over different networks with disparate speech codecs. Thus, in the ITU-T effort, it was agreed to allow the winning candidate of the 3GPP standardization to compete in the selection phase against ITU-T qualified candidates.

The problem of transcoding can be clarified by examining the multiplicity of existing speech coding standards in the telephone band of 200-3400 Hz. In the wireline side, ITU-T standards G.726, G.728, G.729 and G.723.1 are being used. A set of different codecs is used in wireless systems: EFR in GSM; AMR-NB in GSM and 3GPP; IS-641 in North American TDMA; IS-127 and IS-733 in North American CDMA; and the list can be continued. Most of these coders operate at similar bit rates and are based on similar technologies. However, when a call is established between two wireless networks or between a wireless and wired network then the transcoding between the two coders is needed. This results in a noticeable quality degradation.

The ITU-T wideband speech coding activity started in 1999, and qualification tests were conducted in summer 2000 for six submitted candidates. Based on the test results of the AMR-WB codec in 3GPP, ITU-T approved this codec to participate in the selection phase of the ITU-T standardization. The ITU-T selection tests were conducted in the spring of 2001 for the two remaining candidates and the results were presented in the July 2001 Rapporteurs meeting of Q.7/16. The AMR-WB codec showed better overall performance and was selected. It is likely to be decided in the November 2001 plenary meeting of WP3/SG16 as Recommendation G.722.2.

The adoption of AMR-WB by ITU-T is of significant importance since for the first time the same codec is adopted for wireless as well as wireline services. This will eliminate the need for transcoding, and ease the implementation of wideband voice applications and services across a wide range of communication systems and platforms.

3.3.3. Brief description of the 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 as well as ITU-T G.729 and G.723.1 at 5.3 kbit/s. The AMR-WB speech codec consists of nine 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. The codec includes also a background noise mode designed to be used in the discontinuous transmission (DTX) operation of GSM and as a low bit rate source dependent mode for coding background noise in other systems. In GSM the bit rate of this mode is 1.75 kbit/s.

The 12.65 kbit/s mode and the modes above it offer high quality wideband speech. The two lowest modes at 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 blocks of 20 ms. Two frequency bands, 50–6400 Hz and 6400–7000 Hz, are coded separately for decreasing complexity and focusing 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 using an ACELP algorithm. Several features have been added to obtain a high subjective quality at low bit-rates on wideband signals. Linear prediction (LP) analysis is performed once per 20 ms frame. Fixed and adaptive excitation codebooks 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. The gain of the higher band is adjusted relative to the lower band based on voicing information. The spectrum of the higher band is reconstructed by using an LP filter generated from the lower band LP filter.

The total computational complexity of the AMR-WB speech codec is 38.9 WMOPS (Weighted Million Operations Per Second). This figure corresponds to the theoretical worst case when the path through the codec giving the biggest complexity is assumed. The complexity estimate includes the VAD, DTX, and CNG functions. The complexity of the AMR-WB speech codec is shown in Table 2. Corresponding figures from AMR-NB are also included for comparison.

Table 2: Implementation complexity of AMR-WB and AMR-NB speech codecs.

	AMR-WB	AMR-NB
Computational complexity [WMOPS]		
<i>Speech Encoder</i>	31.1	14.2
<i>Speech Decoder</i>	7.8	2.6
<i>Total</i>	38.9	16.8
Data RAM [kWords, 16-bit]	6.5	5.3
Data ROM [kWords, 16-bit]	9.9	14.6
Program ROM [num. of ETSI basicops]	3889	4851

3.3.4. Speech quality of AMR-WB

AMR-WB provides high granularity of bit-rates making it suitable for many applications in 2G and 3G systems. The high speech quality makes the codec well suited also for wideband voice applications in wireline services as shown by its adoption by ITU-T.

Figure 5 shows an illustrative graph comparing AMR-WB to narrowband codecs AMR-NB and EFR in the GSM full-rate channel. In typical operating conditions ($C/I > 10$ dB), AMR-WB gives superior quality over all other GSM codecs. Even in poor radio channel conditions ($C/I < 7$ dB), AMR-WB still offers comparable quality to AMR-NB and far exceeds the quality of the fixed rate GSM codecs.

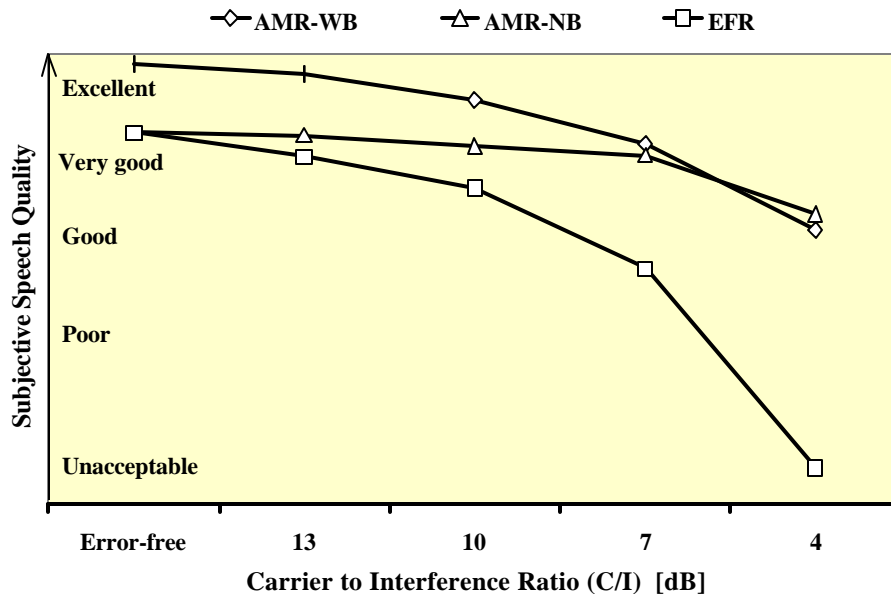


Figure 5: Speech quality of AMR-WB in the GSM FR channel compared to AMR-NB and EFR.

In the course of the AMR-WB standardization in 3GPP, the codec was extensively tested in the selection, verification and characterization phase. In spring 2000, the codec participated also in the Selection Phase of Q.7/16 in the ITU-T.

During the 3GPP selection phase, the AMR-WB codec was tested in six independent listening laboratories with five languages: English, French, Japanese, Mandarin Chinese, and Spanish. The testing covered different input levels, tandeming, background noise performance, and performance of VAD/DTX. In addition, the codec was tested under different error conditions in mobile communication channels both in GSM and WCDMA 3G. The AMR-WB codec showed consistently good performance. It met all performance requirements in all of the laboratories throughout the tests.

During the post-selection verification phase, the AMR-WB codec was tested in several additional conditions to verify its good performance. The tests included performance on DTMF tones and other special input signals, overload performance, muting behaviour, transmission delay, frequency response, detailed complexity analysis, and comfort noise generation. The codec showed good performance throughout the tests.

The characterization phase contained further testing to fully characterize all the nine modes of the chosen AMR-WB codec. Total of six different languages were used: English, Finnish, French, German, Japanese, and Spanish. The experiments included tests for input levels and self-tandeming, interoperability in real world wideband and narrowband scenarios, VAD/DTX, clean speech and speech in four types of background noise, channel errors in GSM and WCDMA 3G channels (for both clean speech and speech under background noise) and also testing for packet-switched applications [15].

Furthermore, the AMR-WB codec was tested in the ITU-T selection phase. The testing covered several input levels, tandeming, four types of background noise, frame erasure testing and testing with narrowband speech signals. The tests were conducted in several languages including English, French, German, and Japanese. In the ITU-T testing only a subset of modes were tested: 12.65, 15.85, 19.85 and 23.85 kbit/s.

The quality of the AMR-WB codec is described in the following subsections based on the different phases of the testing.

Basic quality:

The clean speech quality provided by the six highest AMR-WB modes between 23.85 and 14.25 kbit/s is equal to or better than ITU-T wideband codec G.722 at 64 kbit/s. Results are consistent over all tested input levels and also in self-tandeming. The 12.65 kbit/s mode is at least equal to G.722 at 56 kbit/s. The 8.85 kbit/s mode gives still quality equal to G.722 at 48 kbit/s.

The clean speech quality of the AMR-WB codec is illustrated in Figure 6. This is an extract from the ITU-T selection tests. The figure shows codec performance for nominal signal level -26 dBov with clean speech in single coding (1T) and in self-tandeming (2T). The 12.65 and 23.85 kbit/s modes (AMR-WB 13 and AMR-WB 24) were included in this experiment carried out in the French language. ITU-T wideband speech codec G.722 with three bit rates of 48, 56 and 64 kbit/s was used as a reference codec. The results show that the performance of the 12.65 kbit/s AMR-WB mode already exceeds the performance of G.722 at 48 kbit/s and is comparative to G.722 at 56 kbit/s. The highest AMR-WB mode 23.85 kbit/s has performance equal to G.722 at 64 kbit/s. The above observations are valid both in single coding and in self-tandeming.

The background noise performance of the AMR-WB codec is shown in Figure 7. This is an extract from the ITU-T selection tests showing results for the English language. AMR-WB modes of 12.6, 15.85, 19.85 and 23.85 kbit/s were included in the test. Office and car noise were used at SNR 15 dB for both types of noise. For office noise, the lowest tested AMR-WB mode at 12.65 kbit/s has about equal performance to G.722 at 48 kbit/s while the other modes perform better or equal to G.722 at 64 kbit/s. For car noise, the performance of the two highest modes are about equal to G.722 at 56 kbit/s, and the 15.85 kbit/s mode is comparative to G.722 at 48 kbit/s.

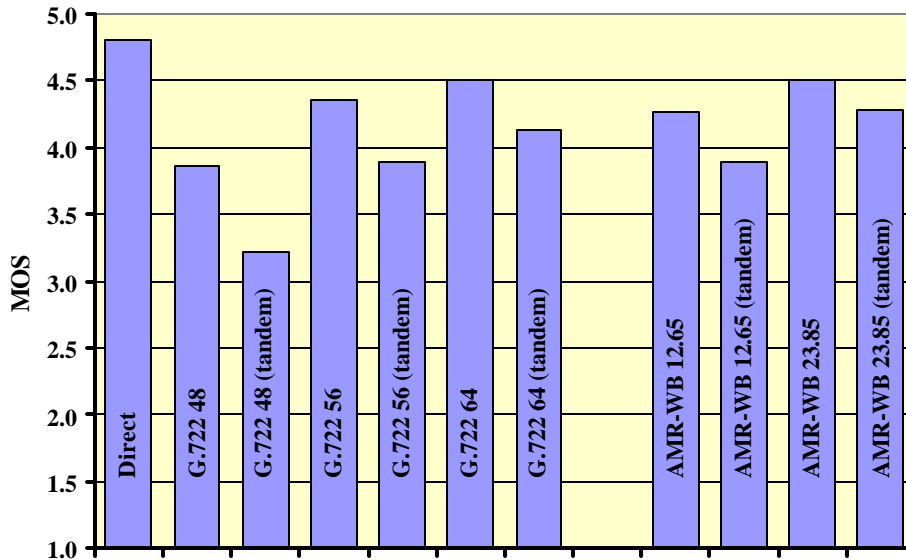


Figure 6: Quality with clean speech. From Experiment 1a of the ITU-T selection tests performed in the French language.

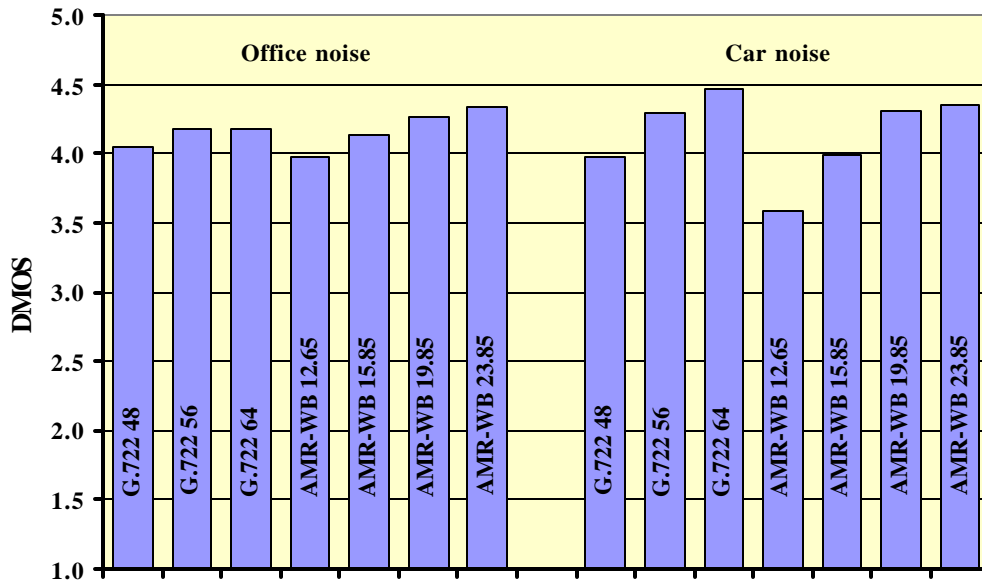


Figure 7: Quality in the presence of background noise (SNR 15 dB). From Experiment 3a of the ITU-T selection tests in American English.

Cellular environment:

The AMR-WB codec has been developed for use in mobile radio environment where typical usage conditions may include both channel errors and high level background noise. AMR-WB provides robust operation also in these conditions. In the GSM system, the channel condition is usually expressed in terms of carrier to interference ration (C/I). In good operating conditions, the C/I is usually above 13 dB. A C/I below 6 dB reflects bad operating conditions (such as at cell boundary) where high bit error rates are present in the radio channel.

In the GSM full-rate channel with clean speech, AMR-WB provides quality better than or equal to G.722 at 64 kbit/s at about 11 dB C/I and above. A quality at least equal to G.722 at 56 kbit/s is obtained for error rates at about 10 dB C/I and above. Under background noise (15 dB Car Noise and 20 dB Office Noise), AMR-WB provides in the GSM full-rate channel quality equal to or better than G.722 at 64 kbit/s at C/I-ratios about 12 dB and above. AMR-WB gives quality equal to or better than G.722 at 56 kbit/s at C/I-ratios about 10 dB and above.

Figure 8 shows an extract of the performance of AMR-WB in GSM full-rate channel under channel errors and with 15 dB Car background noise. This experiment is taken from the 3GPP characterisation phase and it was carried out using the English language. The performance curves of each mode are shown. Note that the two highest AMR-WB bit-rates were not included in the test as they are not targeted for GSM full-rate use.

The VAD/DTX/CNG operation has been assessed as transparent to the listener in the 3GPP characterisation tests.

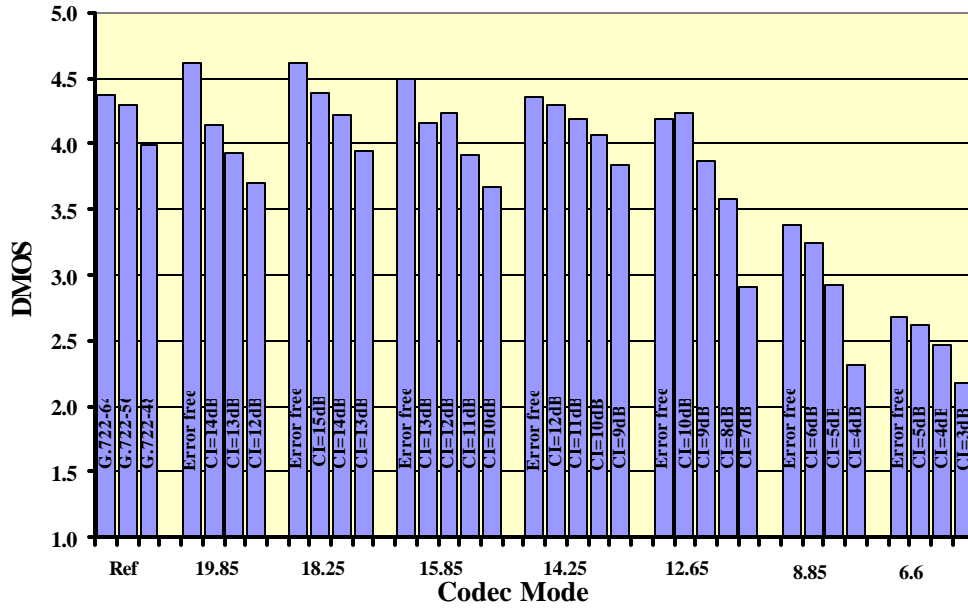


Figure 8: Performance in the GSM full-rate channel with 15 dB Car background noise and channel errors. From the 3GPP characterization with English language.

References:

- [1] ITU-T Recommendation G.726, “40, 32, 24, 16 kbit/s adaptive differential pulse code modulation (ADPCM)”.
- [2] ITU-T Recommendation G.728, “Coding of speech at 16 kbit/s using low-delay code excited linear prediction”.
- [3] ITU-T Recommendation G.729, “Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)”.
- [4] R. Salami, C. Laflamme, J.P. Adoul, A. Kataoka, S. Hayashi, T. Moriya, C. Lamblin, D. Massaloux, S. Proust, P. Kroon, and Y. Shoham, “Design and Description of CS-ACELP: A Toll Quality 8kb/s Speech Coder,” *IEEE Transactions on Speech and Audio Processing*, Vol. 6, No. 2, pp. 116 - 130, 1998.
- [5] ITU-T Recommendation G.723.1, “Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s”.
- [6] K. Järvinen, J. Vainio, P. Kapanen, T. Honkanen, P. Haavisto, R. Salami, C. Laflamme, J.-P. Adoul, “GSM Enhanced Full Rate Codec,” *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Munich, Germany, 20–24 April 1997, pp. 771–774.
- [7] K. Järvinen, “Standardisation of the Adaptive Multi-rate Codec,” *European Signal Processing Conference (EUSIPCO)*, Tampere, Finland, 4–8 Sept. 2000.
- [8] ITU-T Recommendation P.800, “Methods for subjective determination of transmission quality”.
- [9] ITU-T Recommendation G.722, “7 kHz audio-coding within 64 kbit/s”.
- [10] ITU-T Recommendation G.722.1, “Coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss”.
- [11] 3GPP TS 26.190 “Adaptive Multi-Rate wideband speech transcoding,” *3GPP Technical Specification*
- [12] 3GPP TS 26.194 “AMR Wideband speech codec; Voice Activity Detector (VAD),” *3GPP Technical Specification*.
- [13] 3GPP TS 26.193 “AMR Wideband speech codec; Source Controlled Rate operation,” *3GPP Technical Specification*.
- [14] 3GPP TS 26.192 “AMR Wideband speech codec; Comfort noise aspects,” *3GPP Technical Specification*.
- [15] 3GPP TR 26.976 “AMR-WB Speech Codec Performance Characterization” *3GPP Technical Specification*.

Glossary of Terms

3GPP:	Third Generation Partnership Project
ACELP:	Algebraic Code Excited Linear Prediction
ADPCM:	Adaptive Differential Pulse Code Modulation
AMR:	Adaptive Multi-Rate
AMR-WB	Adaptive Multi-Rate Wideband
CDMA:	Code Division Multiple Access
C/I:	Carrier to Interference Ratio
CNG:	Comfort Noise Generation
DTX:	Discontinuous Transmission
EFR:	Enhanced Full Rate
ETSI:	European Telecommunications Standard Institute
GSM:	Global System for Mobile Communications
IP:	Internet Protocol
ITU-T:	International Telecommunications Union – Telecommunication Standardization Sector
LP:	Linear Prediction
MIPS:	Million Instructions Per Second
NA-CDMA:	North American CDMA
NA-TDMA:	North American TDMA
PCM:	Pulse Code Modulation
PSTN:	Public Switched Telephone Network
QoS:	Quality of Service
SMV:	Selectable Mode Vocoder
TDMA:	Time Division Multiple Access
TIA:	Telecommunication Industry Association (North American)
VAD:	Voice Activity Detector
VoIP:	Voice over Internet Protocol
VoPN:	Voice over Packet Network
WMOPS:	Weighted Million Operations Per Second