NOKIA

3GPP Enhanced Voice Services (EVS) codec

The Enhanced Voice Services (EVS) codec brings unprecedented quality to voice and generic audio, including music. Developed by the 3rd Generation Partnership Project (3GPP), the EVS codec supports the low latency required for real-time communication. Compared to earlier voice codecs, it can provide much better quality at similar bitrates, unparalleled quality at higher bitrates, or greatly increased network capacity while maintaining the same quality level. The EVS codec is fully interoperable with HD voice and forms a continuous evolution to higher-quality, more natural communication.

This paper discusses the EVS codec, its benefits, and the ways in which it will revolutionize voice communications. The paper also discusses 3GPP voice and audio codec evolution, and the role that Nokia has played in developing the EVS codec.

NOKIA

Contents

Introduction

Voice quality plays an important role in the evolution of mobile systems. Users are becoming accustomed to computer-based voice over IP (VoIP) services that provide extended audio bandwidth. They also expect high quality from mobile voice services.

High-definition (HD) voice is being rolled out globally, bringing substantial improvement to voice quality in mobile communications [1]. It allows people to communicate more clearly and easily by using the wider audio bandwidth (up to 8 kHz) provided by the Adaptive Multi-Rate – Wideband (AMR-WB) codec to make speech more intelligible and natural, create a feeling of transparent communication, and make speaker recognition easier [2]. HD voice also doubles the audio bandwidth compared to traditional telephony. It is therefore not surprising that HD voice makes one-to-one voice calls much more intimate and conference calls more efficient. The improved intelligibility and natural sound enables clear calls even in noisy environments.

Although HD voice is still being introduced, the next improvement in voice quality has now emerged [3]. The 3rd Generation Partnership Project (3GPP) recently standardized a new Enhanced Voice Services (EVS) codec that will take voice quality to a level "beyond HD." By extending audio bandwidth up to 20 kHz, the EVS codec covers the full range of human hearing. It is the first 3GPP conversational codec (i.e. with latency suitable for communication) that provides equally high quality for voice, generic audio such as music, and content that mixes voice and music.

The EVS codec enables people to use their mobile devices to share music or other content from live events with quality that creates a realistic impression of the live experience. Developed specifically for IP-based communications, the EVS codec provides high robustness to delay jitter and packet losses. It reaches the limits of human audio perception and can cope with network impairments.

Evolution of 3GPP mobile voice/audio codecs

With the new EVS codec, Nokia continues its strong contribution to the development of mobile voice and audio codecs for GSM and 3GPP systems (3G WCDMA and LTE). Nokia has collaborated with other companies to develop the following codecs:

- GSM Enhanced Full-Rate (EFR) codec [4], which is used in the 2G GSM system and supports the highest quality rate in the multi-rate AMR-NB codec
- AMR Narrowband (AMR-NB) codec [5], the default codec for all 3GPP voice services in 3G and beyond (WCDMA and LTE)
- AMR Wideband (AMR-WB) codec [2], a wideband audio evolution of the AMR-NB codec that provides the basis for HD voice.
- Extended AMR-WB (AMR-WB+) codec [6], which provides modes for coding stereo signals and supports generic audio for non-conversational applications such as music streaming
- EVS codec, the next step in 3GPP codec evolution.

Table 1 summarizes the evolution of 3GPP codecs.

	GSM/GERAN EFR	AMR-NB	AMR-WB	AMR-WB+	EVS
Year standardized	1996	1999	2001	2004	2014
Audio bandwidth	Narrowband	Narrowband	Wideband	Fullband	Fullband Super wideband Wideband Narrowband
Use in 3GPP	Used in the GSM system.	Default codec for voice in 3G and beyond (WCDMA, LTE)	Codec for HD voice. Default codec for wideband voice in 3G and beyond (WCDMA, LTE).	Recommended codec for generic audio in 3G and beyond (WCDMA, LTE)	Evolved HD voice codec for LTE
Bitrate(s)	12.2 kbit/s	4.75 – 12.2 kbit/s	6.6 – 23.85 kbit/s	6 – 48 kbit/s	5.9 – 128 kbit/s

Table 1: 3GPP codecs developed by Nokia in collaboration with other companies

Beyond 3GPP, Nokia has made significant contributions to the following speech codec standards:

- PCS1900 EFR (1995)
- US-TDMA EFR (1996)
- CDMA EVRC (1996)
- ITU-T G.722.2 (2001)
- CDMA VMR-WB (2004)
- ITU-T G.718 (2007) and its super wideband extension (2009).

The evolution of GSM/3GPP voice codecs has significantly improved voice quality and system efficiency, and played an important role in the success of modern mobile communication. It has continuously extended audio bandwidth from narrowband voice to the whole range of human hearing, bringing high quality for speech and generic audio (including music). Robustness to transmission impairments and system efficiency have improved at the same time.

The new EVS codec substantially improves voice quality, error resilience, and coding efficiency for narrowband (NB) and wideband (WB) audio bandwidths. The introduction of super wideband (SWB) and fullband (FB) audio brings further quality enhancements. Compared to HD voice, the EVS codec can provide substantially improved quality at similar bitrates and unprecedented quality at higher bitrates. It can also deliver significantly improved network capacity while maintaining the same quality as HD voice.

Figure 1 shows theoretical bandwidths for NB, WB, SWB and FB audio. The actual bandwidths in use may be somewhat narrower. For example, the bandwidth is typically 300–3400 Hz for NB audio.
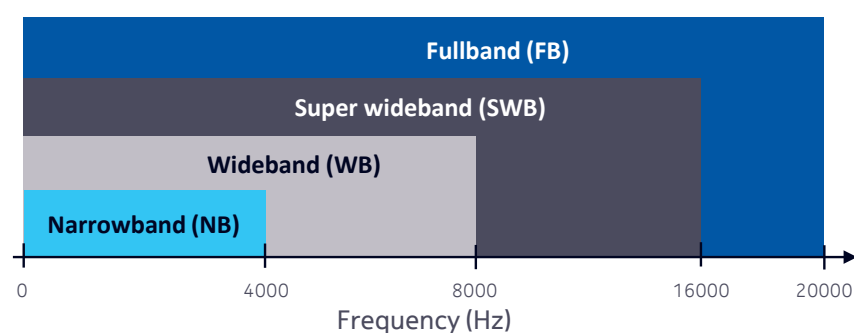

Figure 1: Theoretical bandwidths for NB, WB, SWB and FB audio

One important goal of the EVS codec developers was to enable interaction with the HD voice service. The EVS codec is fully interoperable with the AMR-WB codec used for HD voice. It represents a continuous evolution of HD voice to even higher quality and more natural communication. The EVS codec complements HD voice and provides a solution for achieving higher levels of voice and audio quality.

EVS standardization

Developers began working on the 3GPP EVS codec in 2007 with a pre-study to develop the concept of the codec. Most EVS specifications were approved in September 2014. Standardization work was completed in December 2014, for 3GPP Release 12. The EVS codec specifications can be found in [7-18].

The EVS standardization process included three phases:

- Qualification: Pre-selecting the best candidate codecs
- Selection: Choosing the codec that could meet all requirements
- Characterization: Determining the full performance of the new codec.

Codec selection was based on a rigorous process that compared codec performance and other characteristics against LTE voice codec requirements.

The EVS codec was developed for all-IP 3GPP LTE but will enhance any VoIP or circuit-switched (CS) system. In 3GPP Release 12, the EVS codec is recommended for voice in packet-switched (PS) multimedia telephony over LTE and WCDMA. The AMR-WB interoperable mode of EVS is defined as an alternative implementation of the AMR-WB codec (when the EVS codec is supported), and EVS is defined as the default codec for SWB and FB speech. Release 13 extends the use of EVS to CS telephony over WCDMA.

The EVS codec was developed jointly by 12 companies: Ericsson, Fraunhofer IIS, Huawei, Nokia, NTT, NTT DOCOMO, Orange, Panasonic, Qualcomm, Samsung, VoiceAge, and ZTE Corporation.

EVS features

The EVS codec enhances audio quality and improves coding efficiency for NB and WB audio bandwidths using a wide range of bitrates, starting from 7.2 kbit/s. In addition to fixed-rate coding rates, the codec supports a source-controlled variable bitrate (SC-VBR) mode at an average bitrate of 5.9 kbit/s for NB and WB audio. The EVS codec also provides a significant step in voice quality with SWB and FB operation, starting from 9.6 and 16.4 kbit/s, respectively. It supports a maximum bitrate of 24.4 kbit/s for NB and 128 kbit/s for all other audio bandwidths.

The EVS codec offers input and output sampling at 8, 16, 32, and 48 kHz. To optimize coding quality, an integrated bandwidth detector automatically adapts to the actual bandwidth of the input signal, which may be lower than the bandwidth indicated to the codec. The codec's ability to switch the bitrate at every 20 ms frame allows it to easily adapt to changes in channel capacity.

The codec features discontinuous transmission (DTX) with algorithms for voice and sound activity detection (VAD) and comfort noise generation (CNG). In the SC-VBR coding mode, DTX and CNG are always used for inactive speech coding. An advanced error concealment mechanism mitigates the quality impact of channel errors that result in lost packets. The codec also contains a system for jitter buffer management (JBM) to tackle the jitter or variation in the delay of received packets. There is also a special channel-aware mode that can increase robustness in particularly adverse channel conditions. The channel-aware coding mode operates at 13.2 kbit/s for both WB and SWB audio.

In addition to the EVS Primary modes described above, the EVS codec supports backward compatibility with the AMR-WB codec through an interoperable (IO) mode. This IO mode enhances quality compared to AMR-WB for all nine bitrates between 6.6 kbit/s and 23.85 kbit/s with encoder and decoder improvements, such as optimized pitch prediction and enhanced post-processing. The EVS codec supports seamless switching between the AMR-WB IO and EVS Primary modes.

The feature set of the EVS codec results in a highly flexible, dynamically configurable codec that spans quality ranges from the highest compression rates all the way up to transparent coding. The AMR-WB IO mode can be used for voice over LTE as an alternative implementation of AMR-WB in terminals and gateways that support the EVS codec.

EVS technology

The EVS codec uses content-dependent, on-the-fly switching between speech and audio compression to provide good quality for speech and music signals. It is the first codec to provide core switching technology at the low algorithmic delay of 32 ms.

Figure 2 presents a high-level view of the structure of the EVS codec.
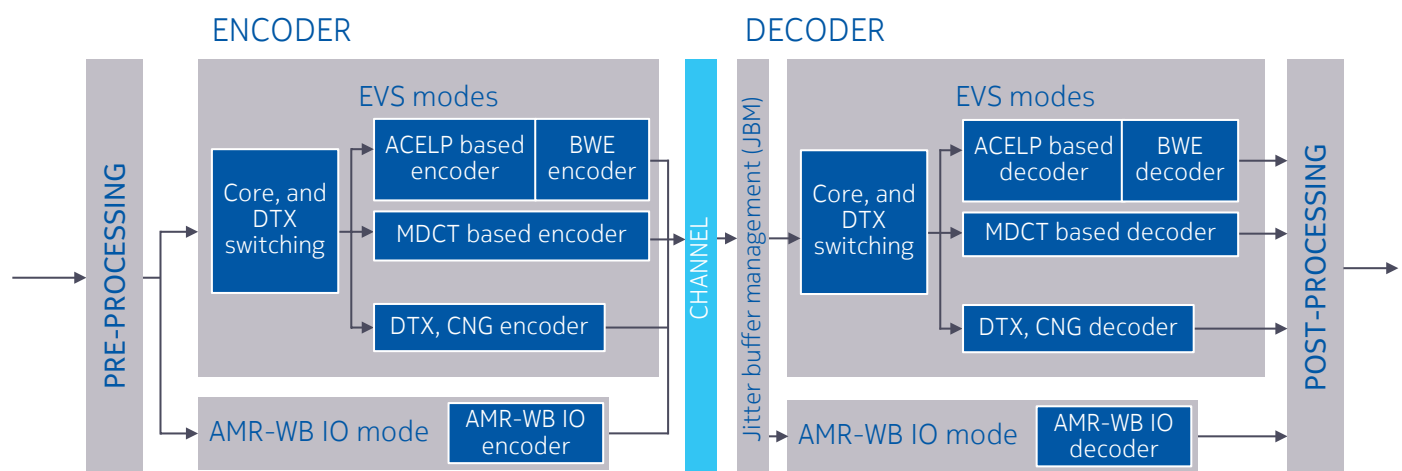

Figure 2: High-level diagram of the EVS codec

Pre-processing includes high-pass filtering at 20 Hz, resampling, signal activity detection, noise update and estimation, bandwidth detection, and signal analysis and classification. The pre-processing unit's analysis and processing capabilities enable a fine-grained classification of the signal type. It applies a targeted coding method to each case.

Fine tuning of input signal characteristics enables the EVS codec to work with a wide range of content types. Although signal bandwidth is signaled as an input parameter, the bandwidth detector identifies lower-than-signaled bandwidth. These capabilities help prevent inefficient coding for band-limited content. Based on the bitrate and input bandwidth, the EVS codec uses two different internal sampling rates for the encoding process, thereby allowing the encoded signal to achieve greater fidelity to the original input signal. Within the EVS modes, core and DTX switching is performed based on information extracted during pre-processing, as well as on the overall input parameters (bitrate, use of discontinuous transmission, and input signal bandwidth). If AMR-WB IO mode is selected, encoding is performed according to the interoperable AMR-WB encoder. The codec can switch between cores and modes at each 20 ms frame boundary.

The speech core used in the EVS codec is based on the principles of Algebraic Code-Excited Linear Prediction (ACELP), inherited from the AMR-WB standard [2]. ACELP relies on the modelling of speech using linear prediction within an analysis-by-synthesis method. The linear prediction (LP) and excitation parameters are encoded and most of the bit budget of this model is allotted to the LP parameters. To use the fine content description of the coder from the preprocessing stage, the codec applies different coding representations for each signal type. This differentiation requires high memory consumption. The large range of bitrates that the codec supports further increases the need for differentiation and, implicitly, memory.

Nokia's multi-stage structured quantizer technology [19] is based on multiple-scale lattices. It supports high encoding efficiency that accommodates all signal types, bandwidths, bitrates, and internal sampling rates while keeping the encoding complexity and read-only memory (ROM) tables within practical limits. In addition to the flexibility brought by the quantizer structure, controlled alternation between predictive and non-predictive modes for encoding the LP parameters provides good resilience to frame loss errors. For speech signals, the part of the bandwidth not covered by the ACELP model for SWB signals is encoded using a time-domain bandwidth extension (BWE) technique. Multi-bandwidth listening test results show a significant quality improvement for SWB compared to WB for all supported operating points.

Encoding based on the modified discrete cosine transform (MDCT) is best suited for music and various background-type signals. Compared with other music-oriented content distribution codecs such as AAC [20], the EVS codec offers high-quality compression of music signals at low delay and low bitrates. It delivers these enhancements by using different MDCT-based modes for different content types and operating modes.

DTX within the EVS modes is important for optimizing battery life in mobile communications. In DTX mode, CNG replaces transmission of noise in the decoder. Improved VAD helps the codec distinguish between active speech, active music, and inactive periods (recording noise, background noise), and estimate the level of the background noise. Based on these decisions, the EVS codec implements two versions of CNG, one based on LP and the other a frequency domain CNG.

Post-processing tools – such as music enhancer, inactive signal post processing, bass-boost filter, and formant post filter – help to ensure the high fidelity of the decoded signal. Compared with the AMR-WB codec, the most notable improvements due to post-processing are audible in noisy channel conditions and for mixed content.

EVS performance

The EVS codec's subjective speech and audio quality have been evaluated extensively by means of listening tests. Several rounds of listening tests were conducted during the 3GPP standardization process. The most recent rounds were the selection-phase [21] and characterization-phase listening tests [22]. Nokia has conducted several additional listening tests to better characterize the codec's subjective speech and audio quality [23]. 3GPP has published a technical report about the performance characteristics of the EVS codec [18].

The main attributes that affect subjective speech and audio quality are audio bandwidth, signal type, channel conditions, and available bitrate.

- Wider audio bandwidth provides additional naturalness to voice quality and enables high quality for generic audio. Traditionally, NB has been used for telephony. Recently, WB audio using the AMR-WB codec (HD voice) has gained traction. EVS takes quality to a new level with SWB and FB bandwidths. Table 2 summarizes the sampling rates and audio bandwidths supported by the EVS codec. The EVS encoder has a global high-pass filter set at about 20 Hz for all sampling rates. This removes very low-frequency noise outside the range of human hearing and improves coding.

	Sampling rate	Audio bandwidth
Narrowband (NB)	8 kHz	up to 4 kHz
Wideband (WB)	16 kHz	up to 8 kHz
Super wideband (SWB)	32 kHz	up to 16 kHz
Fullband (FB)	48 kHz	up to 20 kHz

Table 2: Sampling rates and audio bandwidths in the EVS codec

- Signal type has a significant effect on coding, especially if the codec is not optimized for speech and music. For example, the older 3GPP codecs AMR-NB and AMR-WB are optimized primarily for speech and provide limited audio quality with music signals, even at the highest available bitrates. By contrast, EVS is designed to support speech and music, and to be signal independent. It provides reasonable quality for generic audio even at low bitrates. Despite its name, EVS is much more than a voice codec.

- Channel conditions refer to the real-time transmission channel (e.g. LTE radio network) where, depending on the radio channel characteristics and network conditions (e.g. congestion), some audio packets may be lost, corrupted or received too late for real-time decoding. In these cases, the audio codec has to perform frame error concealment (FEC). Instead of muting or producing disturbing sounds, the decoder uses special algorithms to generate an artificial but pleasant-sounding signal to replace the lost signal frames. Speech codecs can typically handle a small number of lost frames (up to about 3 percent) without severe artefacts. The EVS codec improves these handling capabilities. Even with frame loss of more than 10 percent, EVS provides understandable voice output to temporarily cope with such extreme conditions without severe artefacts.

- The bitrate influences network and radio resource consumption. The most typical bitrates in 3GPP voice telephony are around 12–13 kbit/s. AMR-NB supports bitrates from 4.75–12.2 kbit/s. AMR-WB supports bitrates from 6.6–23.85 kbit/s. EVS supports a wide range of bitrates from 5.9–128 kbit/s. Subjective quality should improve with increasing bitrates. Depending on which modes are used, EVS can provide significantly improved quality over AMR-NB and AMR-WB at similar bitrates, quality raised to an unprecedented level at increased bitrates, or significantly improved network capacity while maintaining the same quality.

Figures 3–8 contain results from subjective listening tests conducted by Nokia. These results are represented on a mean opinion score (MOS) scale. The MOS is calculated based on how non-expert listeners cast votes for different test signal segments containing speech sentences, noisy speech excerpts, or music sequences. In a typical single listening test, 24–32 different listeners hear each condition. Each listener casts votes for 4–8 different signal segments with the same conditions. The individual MOS is typically an average of 96–256 casted votes.

The quality scale in the standardized P.800 test is from 1–5 [31]. The Nokia tests extended this scale so that end points could be scored from 1 (poor) to 9 (excellent). The range extension gives better resolution in the results. The tests used Absolute Category Rating (ACR) with a nine-point scale, also known as the ACR9 MOS methodology.

The tests covered several other standardized codecs and reference conditions to provide insight on the level of improvement delivered by the EVS codec. The reference conditions and codecs included:

- Direct reference conditions with limited audio bandwidth but no coding. The tests used the following low-pass cut-off frequencies: 20 kHz for FB, 16 kHz for SWB, 8 kHz for WB, and 4 kHz for NB.
- AMR-NB codec, also known as AMR [25]
- AMR-WB wideband codec [26]

- ITU-T G.722.1 Annex C [27]. This low-complexity SWB voice codec has an audio bandwidth of 14 kHz. It is used in teleconferencing systems.
- ITU-T G.718 Annex B [28]. This is an embedded (8–64 kbit/s) speech codec for NB, WB and SWB audio. The tests used SWB audio.
- ITU-T G.719 [29], a FB voice and audio codec used in teleconferencing systems
- IETF Opus, version opus-tools-0.1.9-win32.zip [30]. The tests used the constant bitrate (CBR) configuration.

Clean speech results

The clean speech (i.e. no background noise) results in Figure 3 show that EVS delivers significantly better quality than AMR or AMR-WB at all operating points.
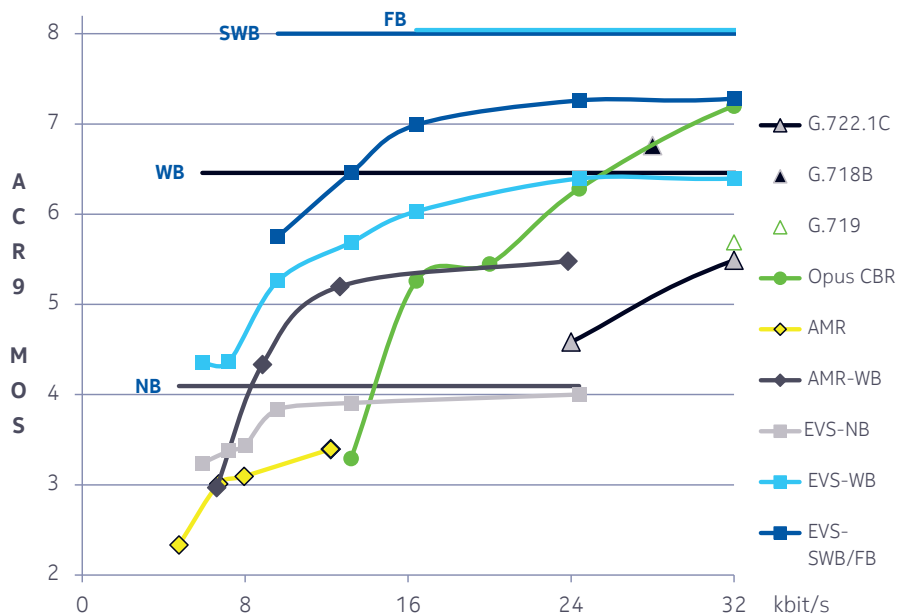

Figure 3: Clean speech MOS values

In the SWB and FB modes, EVS delivers significantly better quality than G.718B, G.722.1C, or G.719. At bitrates below 32 kbit/s, EVS in the SWB and FB modes delivers better quality than Opus CBR, with the quality difference becoming substantial below 24.4 kbit/s. Notably, EVS-SWB 9.6 kbit/s is superior to AMR-WB 23.85 kbit/s and Opus CBR 20 kbit/s, providing better voice quality at less than half the bitrate.

Noisy speech results

The noisy speech (i.e. speech with background noise) results in Figure 4 are very similar to the clean speech results. The exception is that G.722.1C and G.719 perform somewhat better in noisy speech conditions than the other codecs in the test. This phenomenon is well known from previous listening tests [24]. Nevertheless, EVS in the SWB and FB modes delivers significantly better quality than either of these other two codecs. The test used office, street, car, and cafeteria noise types.
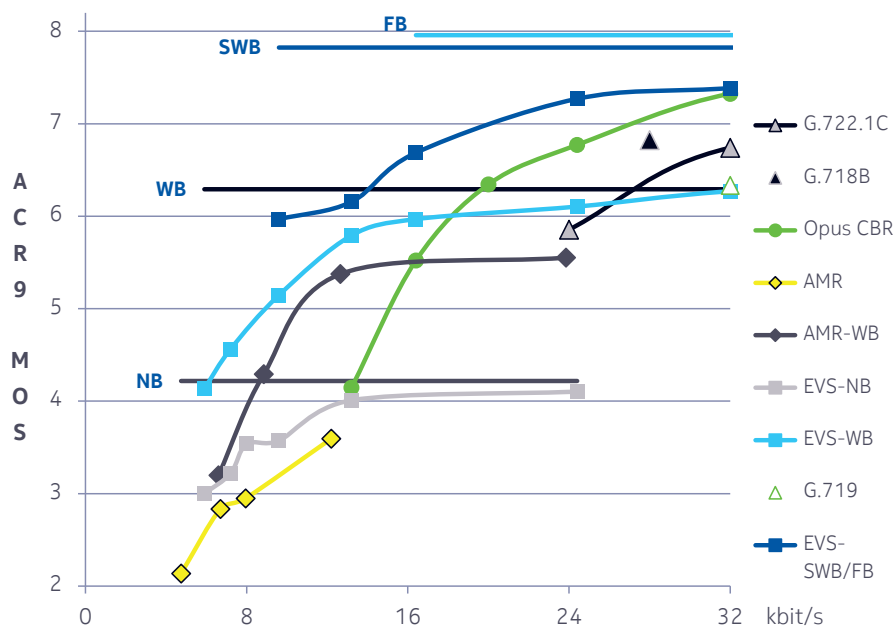
Figure 4: Noisy speech MOS values

Music and mixed content results

The music and mixed content results in Figure 5 are quite similar to the clean speech results. For Opus, however, the lack of quality improvement at increased bitrates is more visible at 20 kbit/s. With music and mixed content, EVS shows excellent performance compared to Opus, AMR, and AMR-WB. At each bitrate in the range 5.9–24.4 kbit/s, EVS achieves an MOS up to two points better than any other codec. The reason for this good performance at low bitrates is the switched ACELP/transform coding cores used in EVS. The content used in the test intermixed speech and music signals (e.g. radio advertisements and telephone on-hold messages/music).
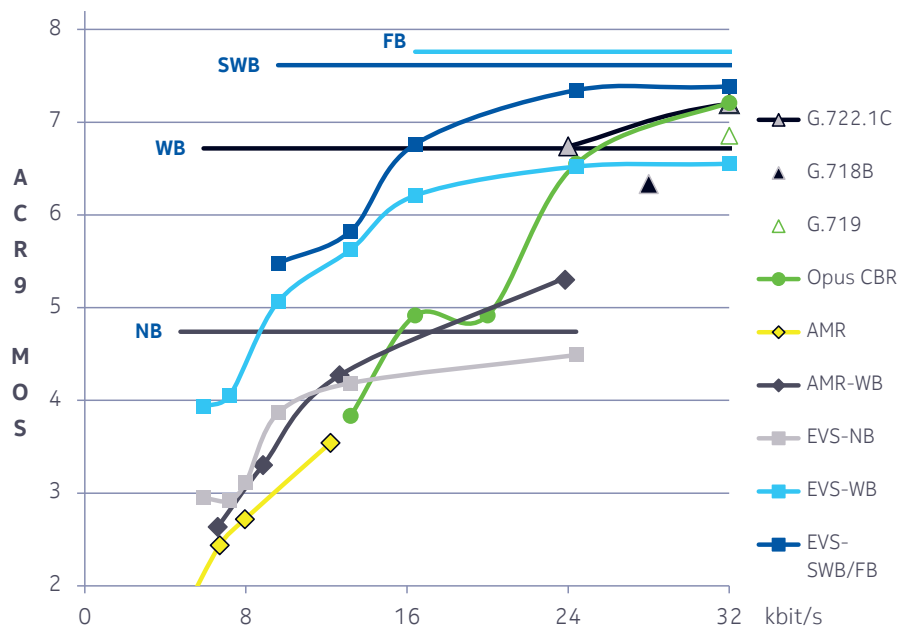

Figure 5: Music and mixed content MOS values

Channel condition results

Figures 6 and 7 show how EVS performs with varying amounts of channel errors. The frame error rate (FER) is shown as a percentage on the X-axis and the MOS is shown on the Y-axis. Direct NB, WB, and FB signals obtained MOSs of 4.91, 7.04, and 8.00 in this listening test. The tests used all signal types: clean speech, noisy speech, music, and mixed content. The channel-aware mode of the EVS codec was not used in these tests.

Figure 6 shows results for AMR, AMR-WB and EVS at bitrates of 12–13 kbit/s (13.2 kbit/s for EVS, 12.65 kbit/s for AMR-WB and 12.2 kbit/s for AMR). Comparisons of EVS-SWB and EVS-WB to AMR-WB and EVS-NB to AMR reveal the inherent robustness of the EVS codec. EVS shows significant improvement at all operating points. The performance improvement grows as the FER increases.
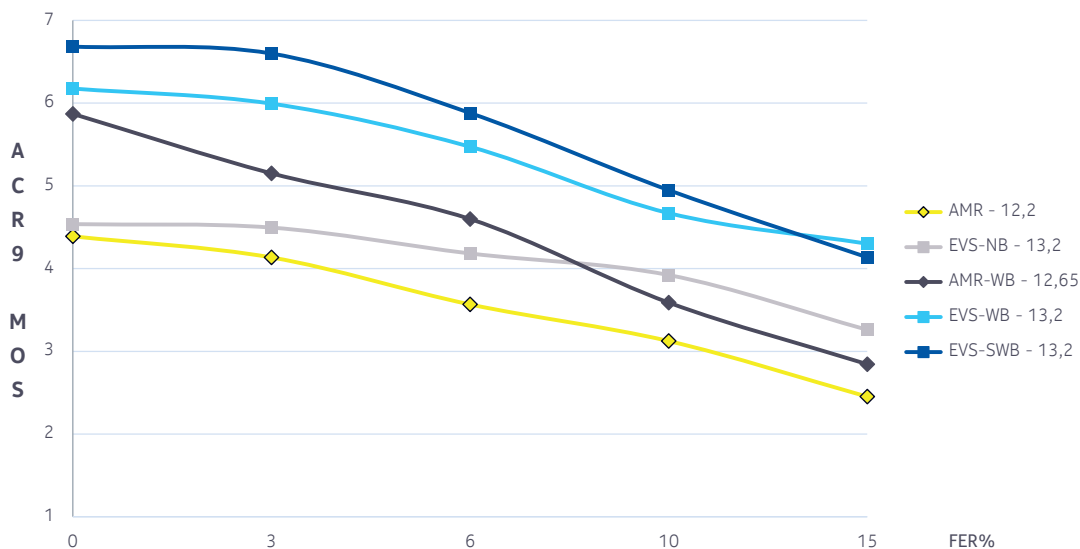

Figure 6: Audio quality when channel errors are present – EVS, AMR-WB, and AMR codecs

EVS-SWB, EVS-WB, and EVS-NB show little or no quality degradation at a 3 percent FER when compared to a clean channel. Even at a 4–6 percent FER, EVS-WB and EVS-SWB provide quality comparable to AMR-WB without errors, and EVS-NB provides quality comparable to AMR without errors. At a 10 percent FER, EVS-WB and EVS-SWB provide usable audio quality (similar to AMR-WB at a 4–5 percent FER) as a temporary measure to cope with errors, thus doubling the frame error robustness for similar quality. At the highest measured FER of 15 percent, EVS-WB provides better quality than AMR at a 3 percent FER. At high error rates above 8 percent, FER EVS-NB provides higher audio quality than AMR-WB, although it is limited to a narrower signal bandwidth.

Figure 7 shows how EVS performs compared to the highest bitrate of AMR-WB (23.85 kbit/s). The inherent robustness of EVS is evident from these results. EVS-WB and EVS-SWB at 13.2 kbit/s both perform significantly better than AMR-WB at 23.85 kbit/s for all FERs, even though the bitrate is almost halved. Comparing EVS-SWB at 32 kbit/s to AMR-WB at 23.85 kbit/s reveals an average improvement of 1.4 MOS points across all operating points. Even at a 6 percent FER, EVS-SWB at 32 kbit/s provides significantly better quality than AMR-WB at 23.85 kbit/s in clean channel conditions. EVS-SWB at 13.2 kbit/s with a 6 percent FER provides similar quality to AMR-WB at 23.85 kbit/s in clean channel conditions. EVS-FB at 48 kbit/s provides unprecedented quality for communications at all operating points, averaging a 1.8 MOS point improvement over AMR-WB at 23.85 kbit/s.
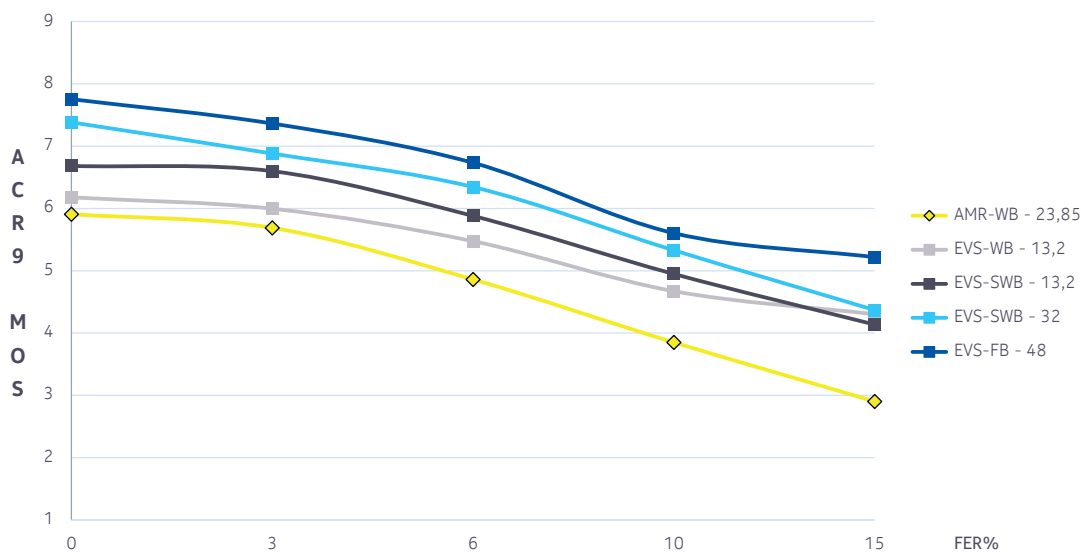
Figure 7: Audio quality when channel errors are present – EVS codec compared to the highest-bitrate mode of AMR-WB at 23.85 kbit/s

Summary of EVS performance

Figure 8 combines all the Nokia test results in a single graph. The graph also includes results for higher bitrates (up to 128 kbit/s). The combined results show that the quality delivered by EVS is better than, or statistically equivalent to, that of any other codec at all bitrates. EVS-NB at 24.4 kbit/s is statistically equivalent to NB direct. EVS-WB at 32 kbit/s is statistically equivalent to wideband direct. In addition, EVS-SWB and EVS-FB reach statistical equivalence to direct at 64 kbit/s and 96 kbit/s, respectively.
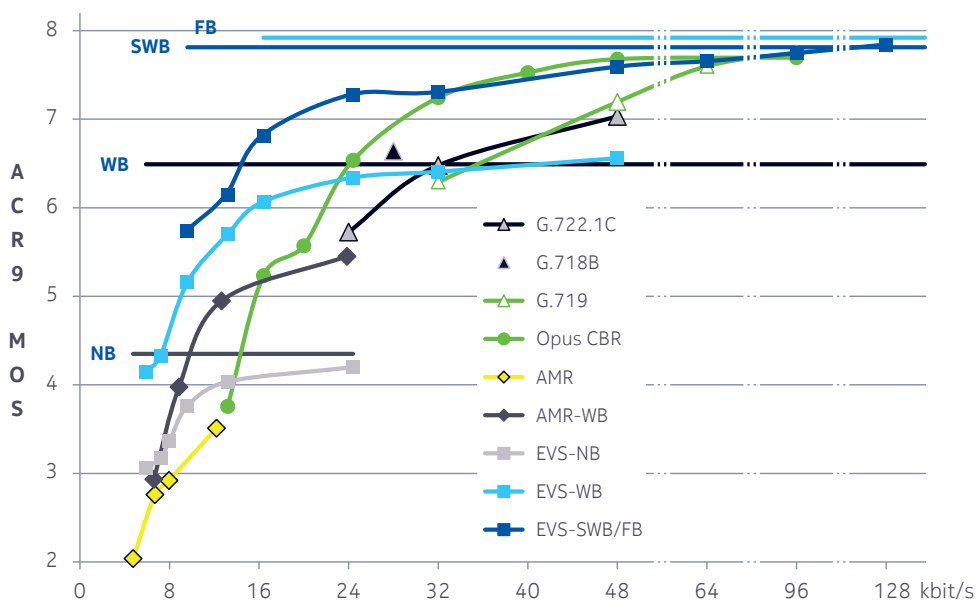

Figure 8: Combined MOS values

The results show that the 3GPP EVS codec produces cutting-edge voice and audio quality across all tested bitrates and bandwidths. The codec excels at bitrates up to 32 kbit/s, which are of utmost importance for the deployment of cost-effective mobile services. Compared to Opus and other tested codecs, EVS can provide the same or better quality at about half the bitrate. For example, EVS-SWB at 9.6 kbit/s delivers significantly better quality than Opus CBR at 20 kbit/s. EVS-SWB at 16.4 kbit/s provides the same quality as G.722.1C at 48 kbit/s.

EVS-SWB at 13.2 kbit/s provides an MOS of 6.15, AMR-WB at 12.65 kbit/s provides an MOS of 4.95 and AMR at 12.2 kbit/s provides an MOS of 3.51. This means that the improvement from AMR-WB to EVS-SWB (1.2 points) is almost as large as the improvement from AMR to AMR-WB (1.44 points). At slightly higher operating points – EVS-SWB at 24.4 kbit/s versus AMR-WB at 23.85 kbit/s – the improvement is almost 1.7 MOS points. This is a larger improvement than that provided between AMR at 12.2 kbit/s and AMR-WB at 12.65 kbit/s.

The SWB and FB modes of EVS at the same bitrates show similar performance. They are represented by a single curve in Figure 8. For audio with frequencies above 16 kHz (and for listeners capable of hearing such high frequencies), the FB mode provides significant quality improvement over the SWB mode.

EVS capacity benefits for LTE

Nokia has studied the voice capacity impact of using the EVS codec in a 3GPP LTE network. Based on the results of the subjective listening tests presented in the previous section, Nokia analyzed the EVS audio bandwidths and bitrates that produce subjective quality equal to or better than the most relevant use cases of the AMR and AMR-WB codecs. The most relevant reference use cases are:

- AMR 12.2 kbit/s. This codec bitrate is used for cases that require NB communications– for example, when interworking with a CS network that supports only NB voice.
- AMR-WB 12.65 kbit/s. This WB codec bitrate can be used when optimal interworking with HD voice towards a 3GPP 2G or 3G network is desired. This bitrate ensures optimal voice quality by enabling transcoder-free operation (TrFO) between LTE and CS networks.
- AMR-WB 23.85 kbit/s. This bitrate provides the best voice quality with the AMR-WB codec.

Lower AMR/AMR-WB bitrates can also be used if rate control is active on the CS side.

Table 3 shows the EVS audio bandwidths and bitrates that produce subjective quality equal to or better than that produced by AMR or AMR-WB for typical conversational voice scenarios, including clean speech and speech with background noise.

Reference	Equal bandwidth	Wider bandwidth
AMR 12.2 kbit/s	EVS-NB 8.0 kbit/s	EVS-WB 5.9 kbit/s
AMR-WB 12.65 kbit/s	EVS-WB 9.6 kbit/s	EVS-SWB 9.6 kbit/s
AMR-WB 23.85 kbit/s	EVS-WB 13.2 kbit/s	EVS-SWB 9.6 kbit/s

Table 3: EVS bandwidths and bitrates that equal or exceed subjective quality of reference scenarios

EVS voice capacity estimations in LTE are based on existing LTE voice capacity simulations for the AMR codec using 5.9 kbit/s and 12.2 kbit/s as reference points [32]. The main system simulation parameters were aligned with [33]. The reference AMR capacities were 410 and 240 users for 5.9 kbit/s and 12.2 kbit/s, respectively, at 5 MHz system bandwidth in the uplink direction and using a semi-persistent scheduling scheme. The uplink direction was chosen because LTE voice capacity is uplink limited and thus

defines the end-to-end system capacity. Better voice capacity can be achieved with an interface that supports semi-persistent scheduling over the air rather than dynamic scheduling.

The voice capacities of AMR-WB 23.85 kbit/s and EVS bitrates were interpolated or extrapolated from the capacities of the reference points. A regression analysis tool was used to perform the interpolation and extrapolation.

Figure 9 illustrates estimated EVS capacity gains compared to the AMR and AMR-WB reference cases shown in Table 3. The estimates assume that no other user plane or layer 1 control channel factor limits these gains.
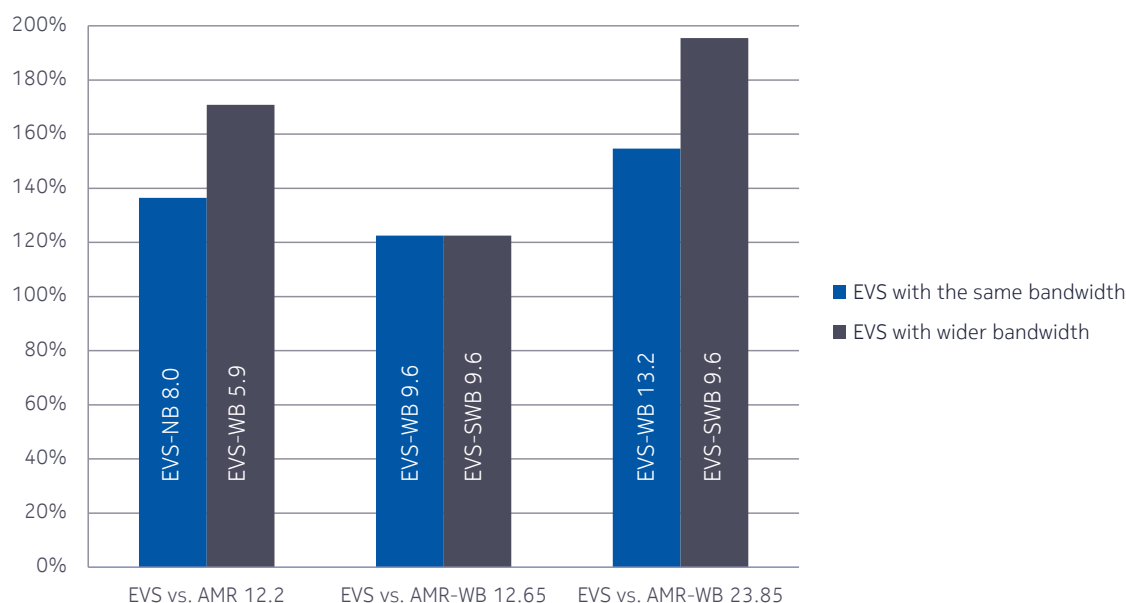

Figure 9: EVS voice capacity compared to AMR and AMR-WB.

The highest capacity gain is achieved when the EVS-SWB 9.6 kbit/s codec is used instead of the AMR-WB 23.85 kbit/s codec. A near doubling of capacity is possible. Significant capacity gains are also achievable against the NB reference scenario, where the EVS-WB 5.9 kbit/s codec provides a capacity improvement of about 70 percent over the AMR 12.2 kbit/s codec.

The lowest expected capacity gain is generated when the EVS codec is used instead of the AMR-WB 12.65 kbit/s codec. In this case, the EVS-WB 9.6 kbit/s and EVS-SWB 9.6 kbit/s codecs both provide a capacity gain of about 20 percent. The capacity gain is the same for the WB and SWB modes because 9.6 kbit/s is the lowest available bitrate for the EVS-SWB mode. The reduced resource consumption of VoLTE on the LTE data channel provides additional capacity for other services or applications.

At low bitrates, the EVS codec improves VoLTE uplink coverage in coverage-limited cells while maintaining the same voice quality at the cell edge. This improvement is created because the LTE terminal can focus its power on a smaller number of physical resource blocks. As shown in figures 6 and 7, the EVS codec can provide comparable audio quality at significantly higher FER levels than the AMR or AMR-WB codecs. It also improves uplink coverage. The EVS-WB/SWB 13.2 kbit/s codec with the channel-aware mode is well suited for use in large LTE cells. Its significantly higher error resilience decompensates sporadic packet losses, which typically arise as the terminal approaches end of coverage.

Beyond EVS

The desire to improve naturalness of voice communications and enhance robustness against transmission errors has been the main driver for new voice codecs. Natural voice quality is particularly important for multi-party conference calls and telepresence applications, which benefit from improved speaker recognition. It is also important for preventing listener fatigue during long sessions. The EVS codec currently supports coding of stereo signals, but only by means of coding two separate mono channels. Extending EVS in future 3GPP releases to stereo and multi-channel coding (with spatial localization of the participants) will be a natural evolution of EVS. It will bring further improvements to the user experience and to system efficiency.

Conclusion

3GPP has standardized a new codec for EVS. The EVS codec is the successor to the HD mobile voice codec AMR-WB. It provides full interoperability with HD voice. EVS offers cutting-edge performance that exceeds the capabilities of other codec standards from ITU-T, IETF and 3GPP. It excels with high audio quality for speech and music, outstanding compression efficiency, and superior robustness against transmission impairments. These characteristics are all essential for mobile operators seeking to deliver cost-effective mobile services and achieve long-term business success. Compared to HD voice, the EVS codec provides substantially improved quality at similar bitrates and unprecedented quality at higher bitrates. It also significantly improves network capacity while maintaining the same quality as HD voice.

Acronyms

2G	2nd Generation
3GPP	The 3rd Generation Partnership Project
AAC	Advanced Audio Coding
ACELP	Algebraic Code-Excited Linear Prediction
ACR	Absolute Category Rating
ACR9	Absolute Category Rating with 9-point scale
AMR	Adaptive Multi-Rate (same codec as AMR-NB)
AMR-NB	Adaptive Multi-Rate – Narrowband
AMR-WB	Adaptive Multi-Rate – Wideband
AMR-WB+	Extended AMR-WB
BWE	bandwidth extension
CBR	constant bitrate
CDMA	code division multiple access
CNG	comfort noise generation
CS	circuit-switched
DTX	discontinuous transmission
EDGE	Enhanced Data rates for GSM Evolution
EFR	Enhanced Full-Rate
EVRC	Enhanced Variable Rate Codec
EVS	Enhanced Voice Services

FB	fullband
FEC	frame error concealment
FER	frame error rate
GERAN	GSM EDGE Radio Access Network
GSM	Global System for Mobile communications
HD	high definition
IETF	Internet Engineering Task Force
IO	interoperable
ITU-T	International Telecommunication Union - Telecommunication standardization sector
JBM	jitter buffer management
LP	linear prediction
LTE	Long Term Evolution
MDCT	modified discrete cosine transform
MOS	mean opinion score
NB	narrowband
PCS	Personal Communications Service
PS	packet-switched
SC-VBR	source-controlled variable bitrate
SWB	super wideband
TrFO	transcoder-free operation
US-TDMA	United States – Time Division Multiple Access
VAD	voice/sound activity detection
VMR-WB	Variable-Rate Multimode Wideband
VoIP	voice over Internet Protocol
VoLTE	voice over LTE
WB	wideband
WCDMA	Wideband Code Division Multiple Access

References

[1]   http://www.gsacom.com/hdvoice/

[2]   B. Bessette et al., "The Adaptive Multi-Rate Wideband Speech Codec (AMR-WB)", IEEE Transactions on Speech and Audio Signal Processing, Vol. 10, No. 8, November 2002.

[3]   S. Bruhn et al., "Standardization of the new 3GPP EVS Codec", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2015, Brisbane, Australia, April 2015.

[4]   K. Järvinen et al, "GSM Enhanced Full Rate Codec", Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing, Munich, Germany, April 1997.

[5]   K. Järvinen, "Standardisation of the Adaptive Multi-Rate Codec", Proc. of X European Signal Processing Conference (EUSIPCO), Tampere, Finland, September 2000.

[6]   J. Mäkinen et al. "AMR-WB+: a new audio coding standard for 3rd generation mobile audio services", Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing, Philadelphia, USA, March 2005.

[7]   3GPP TS 26.441, "Codec for Enhanced Voice Services (EVS); General overview".
        Online: http://www.3gpp.org/DynaReport/26441.htm

[8]   3GPP TS 26.442, "Codec for Enhanced Voice Services (EVS); ANSI C code (fixed-point)".
        Online: http://www.3gpp.org/DynaReport/26442.htm

[9]   3GPP TS 26.443, "Codec for Enhanced Voice Services (EVS); ANSI C code (floating-point)".
        Online: http://www.3gpp.org/DynaReport/26443.htm

[10]   3GPP TS 26.444, "Codec for Enhanced Voice Services (EVS); Test sequences".
        Online: http://www.3gpp.org/DynaReport/26444.htm

[11]   3GPP TS 26.445, "Codec for Enhanced Voice Services (EVS); Detailed algorithmic description".
        Online: http://www.3gpp.org/DynaReport/26445.htm

[12]   3GPP TS 26.446, "Codec for Enhanced Voice Services (EVS); Adaptive Multi-Rate - Wideband
        (AMR-WB) backward compatible functions". Online: http://www.3gpp.org/DynaReport/26446.htm

[13]   3GPP TS 26.447, "Codec for Enhanced Voice Services (EVS); Error concealment of lost packets".
        Online: http://www.3gpp.org/DynaReport/26447.htm

[14]   3GPP TS 26.448, "Codec for Enhanced Voice Services (EVS); Jitter buffer management".
        Online: http://www.3gpp.org/DynaReport/26448.htm

[15]   3GPP TS 26.449, "Codec for Enhanced Voice Services (EVS); Comfort Noise Generation (CNG)
        aspects". Online: http://www.3gpp.org/DynaReport/26449.htm

[16]   3GPP TS 26.450, "Codec for Enhanced Voice Services (EVS); Discontinuous Transmission (DTX)".
        Online: http://www.3gpp.org/DynaReport/26450.htm

[17]   3GPP TS 26.451, "Codec for Enhanced Voice Services (EVS); Voice Activity Detection (VAD)".
        Online: http://www.3gpp.org/DynaReport/26451.htm

[18]   3GPP TR 26.952, "Codec for Enhanced Voice Services (EVS); Performance Characterization".
        Online: http://www.3gpp.org/DynaReport/26952.htm

[19]   A. Vasilache et al., "Flexible spectrum coding in the 3GPP EVS codec", IEEE International Conference
        on Acoustics, Speech and Signal Processing (ICASSP) 2015, Brisbane, Australia, April 2015.

[20]   M. Bosi et al., "ISO/IEC MPEG-2 advanced audio coding", Journal of the Audio engineering society
        45.10, pp. 789-814, 1997.

[21]   3GPP Tdoc S4-141065, "Report of the Global Analysis Lab for EVS Selection Phase", August 2014.
        Online: http://www.3gpp.org/ftp/tsg_sa/WG4_CODEC/TSGS4_80bis/ Docs/S4-141065.zip

[22]   3GPP Tdoc S4-141161, "Report of the Global Analysis Lab for the EVS Characterization Phase",
        3GPP, November 2014.
        Online: http://www.3gpp.org/ftp/tsg_sa/WG4_CODEC/TSGS4_81/Docs/ S4-141161.zip

[23]   A. Rämö and H. Toukomaa, "Subjective Quality Evaluation of the 3GPP EVS Codec", IEEE
        International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2015, Brisbane,
        Australia, April 2015.

[24]   A. Rämö and H. Toukomaa, "Voice quality characterization of IETF Opus codec", in Proc. Interspeech,
        pp. 2541–2544, Florence, Italy, August 2011.

[25]   3GPP TS 26.090, "Adaptive multi-rate (AMR) speech codec; Transcoding functions".
        Online: http://www.3gpp.org/DynaReport/26090.htm

[26]   3GPP TS 26.190, "Adaptive multi-rate wideband (AMR-WB) speech codec; Transcoding functions".
        Online: http://www.3gpp.org/DynaReport/26190.htm

[27]   ITU-T G.722.1, "Low-complexity coding at 24 and 32 kbit/s for handsfree operation in systems with
        low frame loss", ITU, May 2005. Online: http://www.itu.int/rec/T-REC-G.722.1/en

www.nokia.com

[28] ITU-T G.718, "Amendment 2, Frame error robust narrow-band and wideband embedded variable bit-rate coding of speech and audio from 8–32 kbit/s; Amendment 2: New Annex B on superwideband scalable extension for ITU-T G.718 and corrections to main body fixed-point C-code and description text", ITU, March 2010. Online: http://www.itu.int/rec/T-REC-G.718/en

[29] ITU-T G.719, "Low-complexity, full-band audio coding for high-quality, conversational applications", ITU, June 2008. Online: http://www.itu.int/rec/T-REC-G.719/en

[30] J.-M. Valin et al., "Definition of the Opus audio codec", IETF RFC 6716, September 2012. Online: http://opus-codec.org/

[31] ITU-T P.800, "Methods for subjective determination of transmission quality", ITU, August 1996. Online: https://www.itu.int/rec/T-REC-P.800-199608-I/en

[32] H. Holma and A. Toskala, "LTE for UMTS: Evolution to LTE-Advanced", Second Edition, John Wiley & Sons, 2011.

[33] NGMN Alliance, "Next Generation Mobile Networks Radio Access Performance Evaluation Methodology", January 2008. Online: http://www.ngmn.org/uploads/media/NGMN_Radio_Access_ Performance_Evaluation_Methodology.pdf

www.nokia.com