

STANDARDIZATION OF THE NEW 3GPP EVS CODEC

S. Bruhn¹, H. Pobloth¹, M. Schnell², B. Grill², J. Gibbs³, L. Miao³, K. Järvinen⁴, L. Laaksonen⁴, N. Harada⁵,
N. Naka⁶, S. Rago⁷, S. Proust⁷, T. Sanda⁸, I. Varga⁹, C. Greer¹⁰, M. Jelínek¹¹, M. Xie¹², P. Usai¹³

¹Ericsson AB, ²Fraunhofer IIS, ³Huawei, ⁴Nokia, ⁵NTT, ⁶NTT DOCOMO, ⁷Orange, ⁸Panasonic,
⁹Qualcomm Technologies, Inc., ¹⁰Samsung, ¹¹VoiceAge, ¹²ZTE Corporation, ¹³ETSI

ABSTRACT

A new codec for Enhanced Voice Services (EVS), the successor of the current mobile HD voice codec AMR-WB, was standardized by the 3rd Generation Partnership Project (3GPP) in September 2014. The EVS codec addresses 3GPP's needs for cutting-edge technology enabling operation of 3GPP mobile communication systems in the most competitive means in terms of communication quality and efficiency. This paper provides an in-depth insight into 3GPP's rigorous and transparent processes that made it possible for the mobile industry, with its many competing players, to successfully develop and standardize a codec in an open, fair and constructive process. This paper also enables an understanding of this achievement by providing an overview of the EVS codec technology, the standard specifications, and the performance of the codec that will elevate HD voice services to the next quality level.

Index Terms—EVS, HD voice, speech/audio coding

1. INTRODUCTION

In 2010, 3GPP finalized a study [1] on how 3GPP could maintain the high value and competitiveness of its voice services and whether the new Evolved Packet System with LTE (Long Term Evolution) access could open up new opportunities for major service improvements in terms of voice and audio quality. The study considered mobile use cases pertinent to LTE access and use cases that may benefit from improved audio quality. The study also examined the need for enhanced codecs beyond AMR-WB (Adaptive Multi-Rate Wideband), the codec now deployed to deliver HD voice in 3GPP mobile systems. An important further aspect of the study examined how new enhanced voice services could interact with existing HD voice services.

Envisioned use cases for enhanced voice service include improved classical telco-grade telephony (typically realized as IMS Multimedia Telephony [2], [3]), high-quality multi-party conferencing or audio-visual communication, offering a 'being-there' quality of experience. Even streaming voice and audio as well as offline voice and audio delivery were considered possible application scenarios.

Based on the conclusions of the study, 3GPP immediately launched a work item (see work item description (WID) in [4]) targeting the standardization of a new voice codec for Enhanced Voice Services, EVS. The goal of the work item was to provide clear benefits in terms of overall service quality, efficiency and interoperability in 3GPP LTE networks. As a result, it is anticipated that enhanced voice services based on the new EVS codec will become the dominant voice service in 3GPP LTE networks. It is further envisioned that enhanced voice services will extend beyond 3GPP LTE system scope, ranging from deployments in circuit switched 3G mobile systems [5] to other

mobile and wireless (WiFi) networks, fixed networks and the Internet, including using WebRTC. In that context not only the performance of the EVS codec, in comparison to existing 3GPP and ITU-T codecs, is of interest but even to other state-of-the-art codecs like Opus [6].

This paper describes the process of the successful 3GPP standardization of the EVS codec. It presents how, starting from the goal defined in the EVS WID, over the formulation of clear terms of reference the application of a rigorous standardization process in the end resulted in a codec meeting and in many aspects surpassing the expectations.

2. THE 3RD GENERATION PARTNERSHIP PROJECT

The EVS codec was standardized in the 3rd Generation Partnership Project (3GPP). 3GPP unites six telecommunications standard development organizations (ARIB, ATIS, CCSA, ETSI, TTA, TTC) known as the 3GPP "Organizational Partners".

3GPP has four Technical Specifications Groups (TSG): Radio Access Networks (RAN), Service & Systems Aspects (SA), Core Network & Terminals (CT) and GSM EDGE Radio Access Networks (GERAN).

SA Working Group 4 "Codec" (SA4) deals with speech, audio, video, and multimedia codecs, in both circuit-switched and packet-switched environments. Other topics within the mandate of SA4 are: quality evaluation, end-to-end performance, and interoperability aspects with existing mobile and fixed networks (from codec point of view). In conducting its work, SA4 strives to specify the best possible technical solutions considering at the same time the planned global use of the codecs with flexibility needs imposed by different regional requirements and preferences, including differences in end-to-end Quality of Service (QoS) / Quality of Experience (QoE) / capacity trade-offs.

3. EVS CODEC STANDARDIZATION PROCESS

3.1. Reference codecs for standardization

To standardize an EVS codec with very high quality and evaluate all features of EVS under the same or equivalent conditions in subjective and objective testing, state-of-the-art codecs standardized by 3GPP and ITU-T, including AMR, AMR-WB, AMR-WB+, G.711, G.711.1, G.718, G.718B, G.719, G.722, G.722.1, and G.722.1C, were identified as potential reference codecs in the EVS-10 document [7] for EVS standardization. These reference codecs can provide good quality for speech and music, and many of them have been standardized in recent years.

3.2. Terms of Reference

3.2.1. Design constraints

The design constraints specified in the EVS-4 document [8] set the framework for the EVS codec in terms of capability and resource usage. As such they list functionalities that are divided into

mandatory, recommended and optional features to be provided by EVS codec candidates. Mandatory features are: support for input-output sampling at 8, 16, 32, 48 kHz independent of coded audio bandwidth; support of narrowband (NB), wideband (WB), and super-wideband (SWB) coded bandwidths; support for bit-rates of 7.2, 8, 9.6, 13.2, 16.4, 24.4, 32, 48, 64, 96, and 128kbps for the *EVS primary modes*; support of the 9 AMR-WB bit-rates for the *EVS AMR-WB interoperable modes*; presence of a jitter buffer management (JBM) solution conforming to TS 26.114 [3]; rate switching at arbitrary frame boundaries; packet loss concealment; and discontinuous transmission (DTX) operation for rates up to 24.4kbps.

EVS-4 also sets constraints on maximum algorithmic delay (32ms); frame length (20ms); maximum computational complexity (88 WMOPS); memory limits; and limit of the output gain. As a recommended feature, 5.9kbps operation with source controlled variable bit-rate (VBR) is included. Further constraints are set for optional features such as e.g. fullband (FB) operation.

3.2.2. Performance requirements

The minimum performance of the EVS codec was defined across relevant operating points in the EVS-3 document [9]. This document reflects the performance required for an enhanced voice service, following the recommendations specified in TR 22.813 [1]. The EVS-3 document lists subjective performance requirements in the form of statistical tests (e.g. not worse than, better than), as well as objective performance requirements on VAD, background noise attenuation, and JBM. The subjective requirements cover operating points in clean speech, noisy speech (car, street, office noise), music and mixed content, including clean and noisy channel (0%, 3%, 6% FER) and input levels (in dBov), for mandatory and recommended features:

- EVS Primary mode in NB, WB and SWB (EVS-NB, EVS-WB, EVS-WB) and in delay/loss conditions (JBM performance)
- AMR-WB IO modes in 3 configurations including interworking scenarios with legacy AMR-WB terminals:
 - Case A: AMR-WB IO encoding-AMR-WB decoding
 - Case B: AMR-WB encoding-AMR-WB IO decoding
 - Case C: AMR-WB IO encoding/decoding

In addition, performance requirements were also defined for bit rate switching and interworking, and for optional operation modes.

3.3. Phases of Standardization

Traditionally, 3GPP speech codec standardization is organized in three phases: Qualification (Pre-selection of the candidates most likely to meet the performance requirements), Selection (The best codec(s) is chosen) and Characterization (Additional testing to determine the new codec's performance).

3.3.1. Qualification phase

Thirteen companies declared their intention to submit codecs to the Qualification. Each codec was evaluated in 12 subjective experiments, each conducted twice; once in the candidates' own laboratory and once in a laboratory selected at random from the other 12. Tests were blinded with all of the processing being conducted by a dedicated Host laboratory (Dynastat).

Each of the candidates was evaluated against the requirements by an independent Global Analysis Laboratory (Dynastat).

In March 2013, the top five candidates were judged to have qualified although all 13 codecs had passed more than 95% of the 296 requirements tested.

After Qualification it was announced that several of the qualified candidates had been jointly developed and that there were

five consortia. At this stage one organization (Motorola) withdrew from the process.

3.3.2. Decision to submit single codec

As a result of examining the codec high level descriptions provided by each candidate at the Qualification meeting, it became clear to the various consortia that all of the qualified candidates were based upon very similar coding principles.

In September 2013, the remaining 12 companies (Ericsson, Fraunhofer IIS, Huawei, Nokia, NTT, NTT DOCOMO, Orange, Panasonic, Qualcomm, Samsung, VoiceAge and ZTE Corporation) that confirmed their intent to submit a codec in selection declared their intention to work together and to develop a single jointly-developed candidate for the Selection Phase by merging the best elements of the codecs from each of the different consortia.

3.3.3. Selection phase

Even though only a single codec entered the Selection Phase the strict 3GPP process for codec selection was maintained.

The subjective Selection testing comprised 24 experiments, each conducted in two languages. Independent Host Lab (Dynastat), Cross-check Lab (Delta-Sense Lab), Listening Labs (Dynastat, Delta, Mesaqin) and Global Analysis Lab (Dynastat) were used. This testing allowed the codec to be evaluated in 389 requirements. Remarkably the codec exhibited only two systematic failures (in both languages) at the 95% confidence level. One of these failures was subsequently addressed as it was found to be the results of a software bug.

Objective testing was also performed and several organizations volunteered to verify that the code supplied to 3GPP conformed to the requirements in a verification phase.

3.3.4. Characterization phase

In order to evaluate the selected codec in the broadest possible way a further set of 18 subjective experiments have been designed. Six of these experiments have been conducted in two different languages, for a total of 24 tests.

The aim of these additional experiments, and other objective evaluations, was to evaluate features of the codec which remained untested or to highlight areas of interest to 3GPP such as tandeming cases, fullband cases, and multi-bandwidth comparisons. The same listening laboratories used for selection were again employed in characterization.

3.4. Selection process

Codec selection in 3GPP follows pre-defined procedures: Proponents are obliged to provide certain information about their candidate to facilitate selection, and strict rules are set prior to selection to provide guidance on selecting the candidate to be standardized. Verification serves the purpose of cross-check and provision of additional (technical) information.

3.4.1. Selection Deliverables

Proponents were required to provide the following information about their candidate for selection (named selection deliverables):

- High-level description and draft codec specifications
- Report of compliance to Design Constraints
- Funding payment (proponents paid for selection testing)
- IPR declaration
- Objective evaluation results
- Candidate codec fixed-point source code

3.4.2. Selection rules

The strict 3GPP selection process involved the following rules (which were agreed before selection) to determine the candidate to be standardized:

- Provision of a full set of selection phase deliverables
- Compliance with design constraints
- Fulfilment of objective performance requirements
- Codec performance analyzed in sets according to EVS WID:
 - Enhanced quality and coding efficiency for NB and WB speech services
 - Enhanced quality by SWB speech
 - Enhanced conversational music quality
 - Robustness to packet loss and delay jitter
 - Backward interoperability to AMR-WB

3.4.3. Verification

The selected codec was verified by independent entities in order to cross-check important basic parameters, including

- Verification of bit-exactness of fixed-point C-code and executable
- Cross-check of objective requirements
- Sample rate support and frequency responses
- Performance with DTMF tones
- Performance with special input signals
- Idle channel behavior
- Language dependency
- Special input voices and talker dependency

3.5. Evaluation methodology and plans

3.5.1. Test plan

The selection test plan defined 24 P.800 experiments consisting of 7 ACR and 17 DCR tests and containing 389 conditions for the codec under test. A total of 6 talkers/language (3 male + 3 female) and 6 categories (e.g. classical, modern, movie trailer, ...) were used for the speech experiments and the music experiments, respectively. Each experiment was conducted twice (i.e. by 2 different listening laboratories in different languages). In total, 48 listening tests were performed with 10 languages. Each test involved the use of 32 naïve listeners. The 778 ToR (terms of reference) conditions were tested against performance requirements by the dependent groups Students T-test with 95 % confidence interval. Additional evaluation against performance objectives was performed using the independent groups t-test wherever available.

The selection test plan also defined numerous objective evaluations, including gain, JBM compliance, active frame ratio, attenuation in inactive region, bit rate and complexity.

3.5.2. Processing plan

To evaluate the EVS codec under well-defined and reproducible conditions, SA4 developed a selection processing plan [12] defining processing steps for subjective and objective tests. Most methods are based on well-established procedures already used in other standardization efforts such as AMR-WB. Additional methods address novel features of the EVS codec, e.g. evaluation of the jitter buffer manager.

The processing methods were implemented and crosschecked by two independent entities, ensuring that the audio material was processed error-free for the subjective evaluations.

4. HIGH LEVEL CODEC OVERVIEW

4.1. Technology

The EVS codec combines Linear Predictive (LP) coding technology traditionally used for speech coding with modified discrete cosine transform (MDCT)-based frequency-domain coding from audio coding area. The input signal is first passed through a common pre-processing comprising a filter-bank analysis and

sampling rate conversion. Subsequent spectral analysis, LP analysis and pitch analysis are used in bandwidth detection, noise estimation and signal classification. The signal classification first decides about current frame activity. In DTX mode, inactive frames are represented with comfort noise generation (CNG). If the frame is active, it is further classified to be processed with the Algebraic Code-Excited Linear Prediction (ACELP)-based coding for speech dominated content or with the MDCT-based coding for generic audio content. Both ACELP and MDCT coding use further classification to optimize the coding for specific inputs and different bandwidth extension technologies to complement the core coding. At the decoder, post-processing techniques are applied including comfort noise addition, formant enhancement and low frequency postfiltering. Special care has been taken for frame errors handling with transmitted side information at higher bitrates. The architecture of the EVS codec is described in [13].

4.2. Features

The EVS codec enhances coding efficiency and quality for NB and WB for a large bit rate range, starting from 5.9 kbps VBR. It further provides a significant step in quality over these traditional telephony bandwidths with SWB and FB operation starting from 9.6 and 16.4 kbps, respectively. Maximum bit rate is 128 kbps. The ability to switch the bit rate at every 20-ms frame allows the codec to easily adapt to changes in channel capacity.

The codec features discontinuous transmission (DTX) with algorithms for voice/sound activity detection (VAD) and comfort noise generation (CNG). An error concealment mechanism mitigates the quality impact of channel errors resulting in lost packets. The codec also contains a system for jitter buffer management (JBM). Furthermore, it features a special channel-aware mode achieving increased robustness in particularly adverse channel conditions. Enhanced interoperation with AMR-WB is provided over all nine bit rates between 6.6 kbps and 23.85 kbps.

5. SPECIFICATION OVERVIEW

5.1. EVS core specifications

5.1.1. Technical Specifications

The EVS codec overview is provided in TS 26.441 [14], the detailed algorithmic description in TS 26.445 [10], ANSI C code in fixed-point and floating-point representations are in TS 26.442 [15] and TS 26.443 [16]. Test sequences are specified in TS 26.444 [17], AMR-WB backward compatible functions are in TS 26.446 [18], error concealment of lost packets is in TS 26.447 [19] and Jitter Buffer Management is in TS 26.448 [20], Comfort Noise Generation (CNG) is in TS 26.449 [21], Discontinuous Transmission (DTX) is in TS 26.450 [22] and Voice Activity Detection (VAD) is in TS 26.451 [23].

5.1.2. Maintenance of the specifications

The EVS core specifications are maintained by 3GPP through the Change Request (CR) procedure [24]. As a result, any bugs or ambiguity discovered in the standard specifications can be corrected.

5.2. Service related specifications

The EVS RTP Payload Format is specified in Annex A of TS 26.445 [10]. The Media Type Parameters descriptions and handlings are also specified in this Annex of TS 26.445 [10] and in TS 26.114 [3]. The EVS RTP Payload Format includes a Compact format and a Header-Full format to support flexible operation under various network environments and also to fulfill the requirements of the Design Constraints [8]. The payload format for

AMR-WB [11] is also supported for interoperability with UEs not supporting EVS but supporting AMR-WB. The use of the Compact and the Header-Full formats as well as general parameters such as bitrates, bandwidths, DTX on/off, Codec Mode Request on/off and etc., can be negotiated by using Media Type Parameters. In addition to the RTP Payload Format and the Media Type Parameters, other guidelines and operations necessary for the handling of the EVS codec such as QoS handling, codec control, and interworking with other networks are specified in TS 26.114 [3].

6. PERFORMANCE EVALUATION RESULTS¹

The fixed-point EVS codec was rigorously tested using the ITU-T P.800 [25] methodology with naïve listeners, demonstrating fulfillment of all testable EVS WID objectives. The extensive Selection and Characterization testing (full results available in 3GPP TR 26.952 [26]) required a budget exceeding €1.1M. The first EVS WID objective was to provide improvements for NB and WB services, and Figure 6.1 illustrates the NB improvement over AMR at 12.2 kbps and the viability of the EVS codec at even lower bit rates. Similarly, Figure 6.2 shows the WB improvements over AMR-WB that EVS offers, including equivalence to Direct at 24.4 kbps.

To fulfill the SWB EVS WID objective, the EVS codec provides state-of-the-art SWB performance, both in benign conditions (Figure 6.3) and in more realistic conditions of background noise and frame erasures (Fig. 6.4). The WID objective for the robustness of EVS demonstrated in SWB was also confirmed in NB and WB testing. Addressing an objective for improved performance in mixed content and music, the EVS codec provides a significant improvement over legacy codecs, as shown by the results in Fig. 6.5 (WB) and Fig. 6.6 (SWB).

Figures 6.7-6.9 present the results of additional testing conducted by Nokia using a modified P.800 (ACR-9) test methodology [27]. The traditional voting scale was extended from five to nine with verbal descriptions given only for the extreme categories (1: “very bad”, 9: “excellent”). The experiments were multi-bandwidth, spanning a range from NB through FB. The EVS codec was compared with Opus [6]. With the exception of 9.6 kbps, VBR was compared with VBR and CBR was compared with CBR. It was however noted that requesting a 5.9 kbps bit rate from Opus resulted in actual bit rate between 7 kbps and 8 kbps depending on the signal type. For EVS, the best-performing bandwidth was selected at each bandwidth for comparison. More extensive results are available in [28].

7. CONCLUSION

A new codec for Enhanced Voice Services (EVS) has been standardized by 3GPP. EVS will be the successor to the present mobile HD voice codec AMR-WB, while retaining full interoperability. 3GPP’s rigorous and transparent standardization process involved the definition of demanding terms of reference (ToR’s). The EVS codec was tested against these ToR’s in three test phases and with extensive independent evaluations using an unprecedented budget exceeding €1.1M. The test campaign included 70 subjective tests performed in 10 languages, several input signal categories, and using independent test labs.

¹ Courtesy of 3GPP. Source: 3GPP TR 26.952 [26].

The results presented in this paragraph, figures 6.1-6.6, are a compilation of EVS codec selection test results from that source.

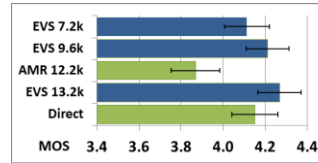


Fig. 6.1: NB clean speech

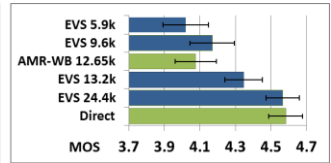


Fig. 6.2: WB clean speech

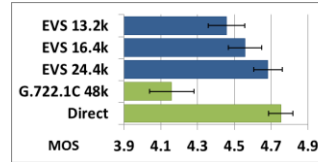


Fig. 6.3: SWB clean speech

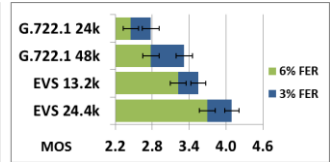


Fig. 6.4: SWB noise/FER

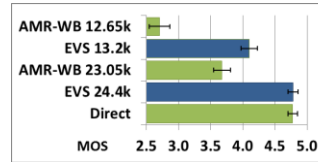


Fig. 6.5: WB Mixed/Music

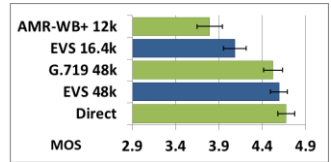


Fig. 6.6: SWB Mixed/Music

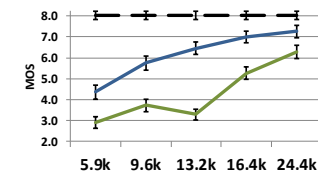


Fig. 6.7: Clean Speech

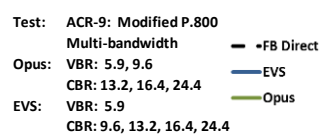


Fig. 6.8: Noisy Speech

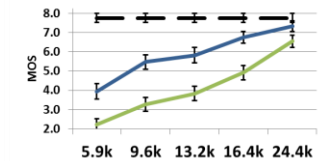


Fig. 6.9: Mixed and Music

The standardization successfully delivered the EVS codec standard with cutting-edge performance as compared to current codec standards from 3GPP, ITU-T and the IETF. The EVS codec excels especially at low bit rates of up to 24.4 kbps, a feature of utmost importance for the deployment of cost-effective mobile services, a cornerstone of mobile operator businesses.

8. ACKNOWLEDGMENT

The authors wish to thank Dynastat, Delta, Mesaqin.com, and ARL for their tremendous efforts in conducting a campaign of subjective testing and analysis of unprecedented size and in an extremely challenging time frame. Special thanks go to Alan Sharpley of Dynastat for his unique contribution through the global analysis and presentation of EVS codec subjective qualification, selection and characterization test results and for his assistance in the compilation of the results for this paper. The authors further wish to thank the many extremely skilled engineers among the 12 codec proponent companies that were essential to the success of the EVS codec standardization.

9. REFERENCES²

- [1] [3GPP TR 22.813](#), “Study of use cases and requirements for enhanced voice codecs for the Evolved Packet System (EPS)”.
- [2] [3GPP TS 22.173](#), “IP Multimedia Core Network Subsystem (IMS) Multimedia Telephony Service and supplementary services; Stage 1”.
- [3] [3GPP TS 26.114](#), “IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction”.
- [4] [SP-140151](#), 3GPP Work Item Description “Codec for Enhanced Voice Services”.
- [5] [SP-140485](#), 3GPP Work Item Description “Support of EVS in 3G Circuit-Switched Networks”.
- [6] Opus 1.1, <http://www.opus-codec.org/>
- [7] [3GPP Tdoc S4-100547](#), “EVS Permanent document (EVS-10): List of potential reference codecs”, Version 0.0.2.
- [8] [3GPP Tdoc S4-130778](#), “EVS Permanent Document (EVS-4): EVS design constraints”, Version 1.2.
- [9] [3GPP Tdoc S4-130522](#), “EVS Permanent Document (EVS-3): EVS performance requirements”, Version 1.4.
- [10] [3GPP TS 26.445](#), “Codec for Enhanced Voice Services (EVS); Detailed algorithmic description”.
- [11] [IETF RFC 4867](#) (2007), "RTP Payload Format and File Storage Format for the Adaptive Multi-Rate (AMR) and Adaptive Multi-Rate Wideband (AMR-WB) Audio Codecs", J. Sjöberg, M. Westerlund, A. Lakaniemi and Q. Xie.
- [12] [3GPP Tdoc S4-141026](#), “EVS Permanent Document EVS-7b: Processing functions for selection phase”, Version 1.4.
- [13] Martin Dietz et al., “Overview of the EVS Codec Architecture,” *Proc. ICASSP*, April 2015.
- [14] [3GPP TS 26.441](#), “Codec for Enhanced Voice Services (EVS); General overview”.
- [15] [3GPP TS 26.442](#), “Codec for Enhanced Voice Services (EVS); ANSI C code (fixed-point)”.
- [16] [3GPP TS 26.443](#), “Codec for Enhanced Voice Services (EVS); ANSI C code (floating-point)”.
- [17] [3GPP TS 26.444](#), “Codec for Enhanced Voice Services (EVS); Test sequences”.
- [18] [3GPP TS 26.446](#), “Codec for Enhanced Voice Services (EVS); Adaptive Multi-Rate - Wideband (AMR-WB) backward compatible functions”.
- [19] [3GPP TS 26.447](#), “Codec for Enhanced Voice Services (EVS); Error concealment of lost packets”.
- [20] [3GPP TS 26.448](#), “Codec for Enhanced Voice Services (EVS); Jitter buffer management”.
- [21] [3GPP TS 26.449](#), “Codec for Enhanced Voice Services (EVS); Comfort Noise Generation (CNG) aspects”.
- [22] [3GPP TS 26.450](#), “Codec for Enhanced Voice Services (EVS); Discontinuous Transmission (DTX)”.
- [23] [3GPP TS 26.451](#), “Codec for Enhanced Voice Services (EVS); Voice Activity Detection (VAD)”.
- [24] [3GPP Change Requests](#).
- [25] [ITU-T Recommendation P.800](#), “Methods for subjective determination of transmission quality”, Aug. 1996.
- [26] [3GPP TR 26.952](#), “Codec for Enhanced Voice Services (EVS); Performance characterization”.
- [27] A. Rämö, “Voice Quality Evaluation of Various Codecs”, *Proc. ICASSP*, March 2010.
- [28] Anssi Rämö, Henri Toukoma, “Subjective Quality Evaluation of the 3GPP EVS Codec,” *Proc. ICASSP*, April 2015.

² Note: When following specification hyperlinks it is advisable to open the respective most recent versions of the documents.