ITU-T G.729.1 Scalable Codec for New Wideband Services

Imre Varga, Qualcomm Inc.
Stéphane Proust, France Telecom
Hervé Taddei, Huawei Technologies

ABSTRACT

G.729.1 is a scalable codec for narrowband and wideband conversational applications standardized by ITU-T Study Group 16. The motivation for the standardization work was to meet the new challenges of VoIP in terms of quality of service and efficiency in networks, in particular regarding the strategic rollout of wideband service. G.729.1 was designed to allow smooth transition from narrowband (300–3400 Hz) PSTN to high-quality wideband (50–7000 Hz) telephony by preserving backward compatibility with the widely deployed G.729 codec. The scalable structure allows gradual quality increase with bit rate. A low-delay mode makes the coder especially suitable for high-quality speech communication. The article presents the standardization goals and process, an overview of the coding algorithm, and the codec performance in various conditions.

INTRODUCTION

The International Telecommunication Union — Telecommunication Standardization Sector (ITU-T) has conducted a substantial amount of work in standardization of speech and audio codecs. In recent years, ITU-T has focused on flexibility enhancement in which ITU-T pioneered scalable coding. Scalable coding is a highly flexible coding technology involving a core layer with multiple *embedded* layers [1]. The core layer provides the minimum bits needed for the decoder to resynthesize the speech signal with a minimum (core) quality. Additional layers aim at improving the quality.

G.729.1 [2] is the first speech codec with an embedded scalable structure built as an extension of an already existing standard. It offers full backward bitstream interoperability at 8 kb/s with the much used G.729 standard [3] in voice over IP (VoIP) infrastructures. G.729.1 is one of the best for wideband speech quality, and its quality is preserved regardless of the access modes and device capabilities thanks to strong robustness to IP packet losses. G.729.1 scalable structure allows for dynamic instant bit rate selection by simple truncation of the bitstream at any component of the communication chain such as gateways or other devices combining multiple data streams. This feature provides high flexibility by easy adaptation to various service requirements and interconnected networks and terminals. With such bit rate adaptation, optimum speech quality is provided according to service and network constraints, and packet droppings that severely impair the overall quality are limited. Thus, the G.729.1 scalable codec minimizes transcoding and ensures smooth introduction of new features without breaking existing services.

This article summarizes the G.729.1 standardization process, the main targeted applications, and related requirements and constraints. It presents the most important characteristics and the performance of G.729.1. For more algorithmic details, the reader may refer to [2, 4, 5].

The work presented in this article was performed for Siemens Enterprise Communications. I. Varga currently works for Qualcomm Inc.

The work presented in this article was performed for Siemens AG. H. Taddei currently works for Huawei Technologies.

STANDARDIZATION PROCESS

G.729.1 standardization was launched in January 2004 under the nickname G.729EV. In July 2005 the qualification test results showed that two candidate codecs (from France Telecom and from a consortium of three companies: Siemens, Matsushita, and Mindspeed) passed all the qualification phase requirements. Two more candidates (from ETRI and VoiceAge) had very few marginal failures and qualified as well. Following the SG16 chairman's advice, the qualified candidate companies agreed to work together on a single best possible candidate. The performance of the new combined algorithm was thoroughly examined during the characterization phase against an agreed set of requirements and objectives. G.729EV completed electronic balloting (Alternative Approval Process [AAP]) in May 2006. The approved standard is published as ITU-T Recommendation G.729.1 as well as G.729 Annex J.

After approval, several improvements were added [2]:

- Annex A (January 2007) contains the Real-Time Transport Protocol (RTP) payload format, capability identifiers, and parameters for the signaling of G.729.1 capabilities using H.245.

 0163-6804/09/$25.00

As G.729.1 is primarily dedicated to speech signals for conversational services, the performance for music signals was not the major criterion. Yet, since applications like Music on Hold and Sharing music ("Hear what I hear") require a good quality level at 32 kb/s, the committee agreed to set appropriate requirements.

- Annex B (February 2007) defines an alternative implementation of the G.729.1 algorithm using floating point arithmetic.
- Annex C (June 2008) specifies discontinuous transmission (DTX) and comfort noise generation (CNG) schemes in fixed point. Annex D (November 2008) is the corresponding floating point specification.
- The low-delay mode functionality originally used for the narrowband layers (8–12 kb/s) was added to the first wideband layer at 14 kb/s of G.729.1 (August 2007). The motivation was to address applications such as VoIP in enterprise networks where low end-to-end delay is crucial.

GOALS AND APPLICATION SCENARIOS

The primary applications for G.729.1 are packetized speech over wireline networks (VoIP, voice over asynchronous transfer mode [VoATM], IP phones, etc.) like G.729, but also high-quality audio/videoconferencing. G.729.1 is designed to offer a single coding format that can adapt to the demands of all integrated services over IP.

Improvement in sound quality is achieved by extending the audio bandwidth. *Bandwidth scalability* is very useful since it allows the flexibility to switch audio bandwidth fully dynamically when interoperating with existing infrastructure. *Bit rate scalability* allows simple adjustment of the bit rate to network or terminal capabilities and multiplexing in gateways. Simultaneous scalability in bit rate and audio bandwidth allows the codec to cope with multiple access technologies for heterogeneous networks and terminals.

In order to obtain a codec with the desired capabilities, some design constraints were set during the standardization process. They are described in the next section.

DESIGN CONSTRAINTS

Besides the usual requirements for a speech coder, specific constraints were identified: interoperability with G.729 / G.729 Annex A including Voice Activity Detector (VAD)/DTX, audio bandwidth scalability from narrowband to wideband, and bit rate scalability covering the range from 8 to 32 kb/s. The frame size was set to 20 ms, which corresponds to the usual packet size in VoIP applications.

Limiting the complexity was highlighted as a key factor: complexity must not exceed 40 weighted million operations per second (WMOPS), and the memory size must be below 30 kwords (16-bit words) for the RAM and below 64 kwords for the ROM.

Although significantly higher than the 15 ms delay of legacy G.729, the delay requirement was set to be below 60 ms with an objective of being lower than 45 ms. This was considered acceptable for many VoIP applications.

The level of bit rate granularity (enhancement layer size) was carefully considered: on one hand, a highly flexible solution offering fine-grain bit rate scalability (FGS) would reduce the impact of wideband (WB) VoIP on network sizing and allow better bit rate optimization and network efficiency. On the other hand, FGS could put too many constraints on possible technologies. As a compromise, experts agreed to set byte-level granularity as an important objective and to not exceed 2 kb/s rate intervals above 14 kb/s.

For the targeted wireline applications, setting the minimum bit rate for wideband capability at around 14 kb/s was found appropriate in the committee. Introduction of an additional narrowband layer at 12 kb/s to offer quality improvement (close to G.711 public switched telephone network [PSTN] quality) serves users with narrowband G.729-based devices.

As usual for ITU-T speech and audio coding Recommendations, it was required that the specification be written in modular ANSI-C code using the 16–32 bit fixed-point basic operators set provided in the ITU-T Software Tool Library [6].

Once the design constraints were set, the desired quality requirements could be set, and those are described next.

QUALITY REQUIREMENTS

Quality requirements were set to a subset of five bit rates: two in narrowband (8 and 12 kb/s) and three in wideband (14, 24, and 32 kb/s). Additionally, gradual quality increase was required from 14 to 32 kb/s in 2 kb/s intervals.

Given all the constraints of scalability, interoperability, and complexity limitations, the quality requirements were set in comparison with already deployed wideband coders (e.g., G.722) at bit rates above 14 kb/s. For lower bit rates, the requirements were set such that the codec would offer significant improvement over narrowband services for VoIP customers.

As G.729.1 is primarily dedicated to speech signals for conversational services, the performance for music signals was not the major criterion. Yet, since applications like music on hold and sharing music ("hear what I hear") require a good quality level at 32 kb/s, the committee agreed to set appropriate requirements. Table 1 introduces some quality requirements/objectives set for the G.729.1 standardization process in the five tested bit rates for some conditions. A more complete set is given in [7, 8].

TECHNICAL OVERVIEW OF G.729.1

ITU-T G.729.1 is an 8–32 kb/s scalable narrowband/wideband coder. The codec operates on 20 ms frames (called superframes), although the core layer and the first enhancement layer at 12 kb/s operate on 10 ms frames similar to G.729. A major novelty was that G.729.1 provides bit rate and bandwidth scalability at the same time. Bit rates at 8 and 12 kb/s are narrowband. The wideband rates range from 14 to 32 kb/s at 2 kb/s intervals.

The G.729.1 encoder is assumed to operate at the bit rate of 32 kb/s providing the highest quality. However, the G.729.1 RTP payload format provides a mechanism to specify a lower encoding bit rate through the maximum bit rate supported (MBS) field [9]. Therefore, the instantaneous encoding bit rate of G.729.1 can also be adapted for every 20 ms frame.

The layered structure, as illustrated in Fig. 1, includes three coding stages with 12 embedded layers. The wideband input signal is first split in two parts using a quadrature mirror filter pro-

Bit rate	Condition	Requirement	Objective
8 kb/s	Clean speech	Same as or better than G.729A	Better than G.729A
	Clean speech with frame erasures	Not worse than G.729A	Better than G.729A
	Noisy speech	Not worse than G.729A	Better than G.729A
12 kb/s	Clean speech	Not worse than G.729E	Better than G.729E
	Clean speech with frame erasures	Not worse than G.729A	Not worse than G.729E
	Noisy speech	Better than G.729A	Better than G.729E
14 kb/s	Clean speech	Better than G.729A and not worse than G.722.2 at 8.85 kb/s	Not worse than G.722.2 at 12.65 kb/s
24 kb/s	Clean speech	Not worse than G.722 at 48 kb/s	Better than G.722 at 48 kb/s
	Music		Not worse than G.722 at 48 kb/s
32 kb/s	Clean speech	Not worse than G.722 at 56 kb/s	Better than G.722 at 56 kb/s
	Music	Not worse than G.722 at 56 kb/s	Not worse than G.722 at 64 kb/s

Table 1. *Summary of some quality requirements for G.729.1 standardization (G.729A and G.729E denote G.729 Annexes A and E, respectively).*

Layer 2 (12 kb/s) focuses on improving the lower band residual signal after Layer 1 encoding has been conducted. For that purpose, an additional FCB search is conducted. This additional codebook is optimized to encode the difference between the original signal and the locally decoded signal at 8 kb/s.

ducing a 0–4000 Hz bandwidth signal and a 4000–8000 Hz bandwidth signal.

The lower band input signal obtained after decimation is processed by layers 1 and 2. These layers use embedded code-excited linear prediction (CELP) to code the lower band (50–4000 Hz) at 8 and 12 kb/s. Layer 1 is bitstream-compatible with G.729/G.729 Annex A (low-complexity version of G.729 aka G.729A). For quality reasons, G.729A was not used as the core; a reduced complexity version of G.729 is the core codec. To reduce the complexity, the G.729 fixed codebook (FCB) search is replaced by a fast codebook search using a global pulse replacement method orthogonalized with the adaptive codebook search. The FCB structure is the same as in G.729. The G.729 open-loop pitch estimation procedure is changed to smooth the pitch tracking to improve frame erasure concealment (FEC).

Layer 2 (12 kb/s) focuses on improving the lower band residual signal after layer 1 encoding has been conducted. For that purpose, an additional FCB search is conducted. This additional codebook is optimized to encode the difference between the original signal and the locally decoded signal at 8 kb/s. Layer 2 puts more emphasis on improving the encoding of the high frequency area (i.e., above 2 kHz). For this purpose a tri-pulse pattern codebook is used with a central pulse having amplitude of +1 and two side pulses with lower magnitudes and opposite signs α. Each fixed code vector is obtained by adding four occurrences of this pattern scaled by a sign factor ± 1. The centers of the pattern occurrences occupy the same sets of positions as in the layer 1 fixed codebook. The factor α is related to the amount of voicing and ranges from 0 for purely unvoiced segments to a value of 0.34 for purely voiced segments. Therefore, this layer has more high-frequency content in voiced signals and less for unvoiced signals. The 12 kb/s fixed codebook is searched using a modified perceptual filter that stresses high frequencies. The perceptual filter is modified by applying a high-pass filter. This modification of the perceptual filter is introduced by modifying the target vector and the weighting filter impulse response. The 12 kb/s fixed-codebook gain is scalar quantized relative to the 8 kb/s quantized fixed codebook gain. The number of bits depends on the subframe index. For the first subframe of each 10 ms frame, 3 bits are required; for the second subframe of each 10 ms frame, only 2 bits are used.

The second stage, in layer 3, performs parametric coding of the higher band (4000–7000 Hz) using time-domain bandwidth extension (TDBWE) and provides wideband quality at 14 kb/s [10]. TDBWE is a very low-complexity bandwidth extension (about 3 WMOPS) that analyzes the higher band signal at the encoder side in terms of energy in the time and frequency domains. On the decoder side, a random signal is generated and shaped in the time and frequency domains using the transmitted side information.

The third stage, layers 4–12, enhances the full wideband signal (50–7000 Hz) by a predictive transform technique referred to as time-domain aliasing cancellation (TDAC) at 16–32 kb/s. Two sets of frequency coefficients are computed. The first one is composed of the modified discrete cosine transform (MDCT) coefficients of the difference signal between the lower band original signal and the locally decoded layer 2 signal processed by a perceptual weighting filter. The second set (higher frequency band) represents the

In order to improve the resilience and recovery of the decoder in frame erasures, parameters based on available lower band information and consisting of signal class , phase, and energy information are transmitted in layers 2 and 3. These parameters are used for FEC.

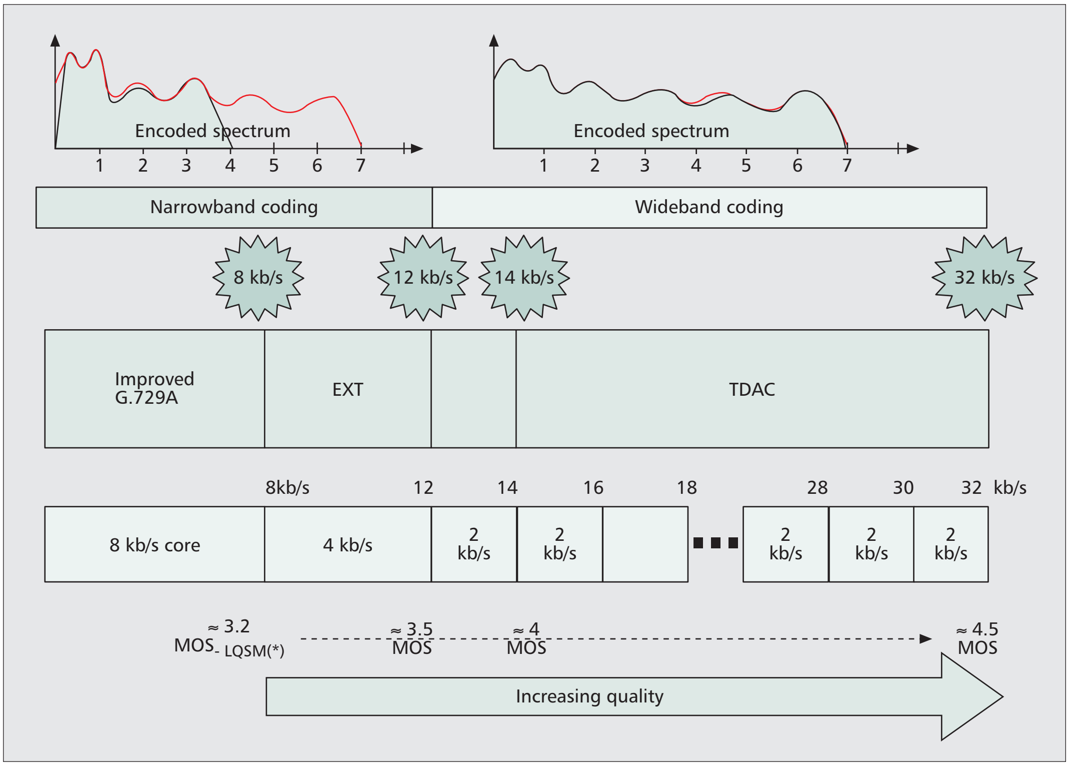

Figure 1. *Illustration of G.729.1 quality variation vs. bit rates and bandwidth scalability. (*) MOS values are MOS-LQSM resulting from ITU-T P.800 subjective tests mixing wideband and narrow band.*

coefficients of the original signal between 4000 and 7000 Hz. The MDCT coefficients in the 0–7000 Hz band are split into 18 subbands (16 coefficients per subband for the first 17 subbands and 8 coefficients for the last one). The spectral envelope is then computed and transmitted. Based on this envelope, the perceptual importance of each subband can be estimated, and the transmission order of the subbands is decided accordingly, as well as the bit allocation per subband. Finally, the frequency coefficients are encoded in each subband using an embedded spherical vector quantization. This operation is divided into two steps: search for the best codevector and indexing of the selected codevector. The quantized and received frequency coefficients gradually replace the TDBWE coefficients while the bit rate increases.

In order to improve the resilience and recovery of the decoder in frame erasures, parameters based on available lower band information and consisting of signal class (voiced, unvoiced, onset, or voiced/unvoiced transition), phase, and energy information are transmitted in layers 2 and 3. These parameters are used for FEC.

As already mentioned, extensions were added or are in the process of being added to G.729.1. These extensions enhance the codec in areas not covered by the G.729.1 main body approved in May 2006 and are described in the next section.

Codec Extensions

By default, G.729.1 operates on input/output signals sampled at 16 kHz (wideband). However, several optional modes have been included. Special attention should be drawn to the low-delay mode of G.729.1. Initially, G.729.1 comprised a low-delay mode at 8 and 12 kb/s that could only operate with input and/or output signals of 8 kHz sampling frequency. This mode was designed for narrowband speech communication in applications where end-to-end communication delay reduction is important. In 2007 this low-delay mode was extended to the first wideband bit rate (14 kb/s). The algorithmic delay is reduced by avoiding the overlap-add operation of the TDAC coding stage. Indeed, delay is an important performance parameter, and transmitting speech with low end-to-end delay is also required in several applications making use of wideband signals (sampled at 16 kHz). Besides reducing G.729.1 delay by 20 ms, complexity is also significantly reduced in low-delay mode operation (Table 2).

Another important functionality essential for efficient speech communication systems is DTX. DTX operation allows encoding of speech at a lower average rate and hence is very useful to decrease power consumption of the terminals by taking speech inactivity into account. During inactive periods, only parametric descriptions of the background noise are transmitted. ITU-T Recommendation G.729.1 Annexes C and D offer a silence compression scheme with full interoperability with the G.729 Annex B silence compression scheme. This scheme can support all G.729.1 bit rate modes and provides an embedded silence insertion description (SID) structure that consists of a lower band core layer bit-stream interoperable with G.729 Annex B SID frame, a lower band enhancement layer, and a higher band layer.

A superwideband extension for encoding signals with 32 kHz sampling frequency is under development with expected completion in the fourth quarter of 2009. The goal is to encode signals with an audio bandwidth of 50–14,000 Hz. This extension will be made on top of G.729.1 (and of G.718), and will include both mono and stereo capabilities. The bit rate will rise up to 64 kb/s.

PERFORMANCE

G.729.1 performance has been thoroughly assessed during its standardization process. This section describes some of the conducted tests and their results. The performance assessment includes the quality assessment for the conditions listed in the Terms of Reference (ToR), the evaluation of the complexity, and the algorithmic delay.

QUALITY

G.729.1 quality was characterized in two steps. For step 1, five formal subjective experiments were designed. Each experiment was run twice in two different languages, each with 32 naive native listeners using monaural headphones. Three kinds of input signals were considered: clean speech, music, and noisy speech with four types of background noise at various signal-to-noise ratios (SNRs). For clean speech and music, Absolute Category Rating (ACR) methodology was used, whereas Degradation Category Rating (DCR) methodology was used for noisy speech. The experiments were allocated to six listening laboratories (BenQ, Dynastat, France Telecom, NTT-AT, TTA, and VoiceAge).

Figures 2 and 3 show the performance of G.729.1 and the reference coders in narrowband speech and wideband speech, respectively. The Y-axis indicates the MOS/DMOS scores depending on whether the test is an ACR or a DCR experiment.

In addition, another experiment was performed to assess the quality for wideband clean speech signals at 10 bit rates from 14 to 32 kb/s with 2 kb/s intervals using an objective methodology (P.862.2, dubbed WB-PESQ) with five languages. Results are shown in Fig. 4. A last experiment informally tested frequent/infrequent switching across the different bandwidths/bit rates. The slow switching case corresponds to real application scenarios such as multicall on ATM virtual channel (ATM VC) or change of wireless LAN access point. The fast switching case was designed to test the fine bit rate granularity with a bit rate randomly selected among the 12 available bit rates at each frame.

Step 2 characterization tests comprised one experiment in narrowband and two others in wideband to further assess G.729.1 quality, especially in error conditions in clean speech, and to compare the quality of four wideband ITU-T codecs (G.722, G.722.1, G.722.2, G.729.1) in clean speech and music conditions. The quality of the low-delay mode was also formally tested under error conditions, in particular to check the robustness against IP packet losses causing voice codec frame erasures at various frame error rates. Figure 5 shows the performance of these ITU-T WB codecs under various frame erasure

Computational complexity at 32 kb/s	35.79 WMOPS
Computational complexity at 32 kb/s in Low-delay mode	23.57 WMOPS
Static RAM	5 kwords
Scratch RAM	3.7 kwords
Data ROM	8.5 kwords
Program ROM	32 kwords
Algorithmic delay	48.9375 ms
Low-delay mode (narrowband), 8–12 kb/s	25 ms
Low-delay mode, 14 kb/s	28.94 ms

Table 2. *Complexity figures of the G.729.1 coder (encoder/decoder).*

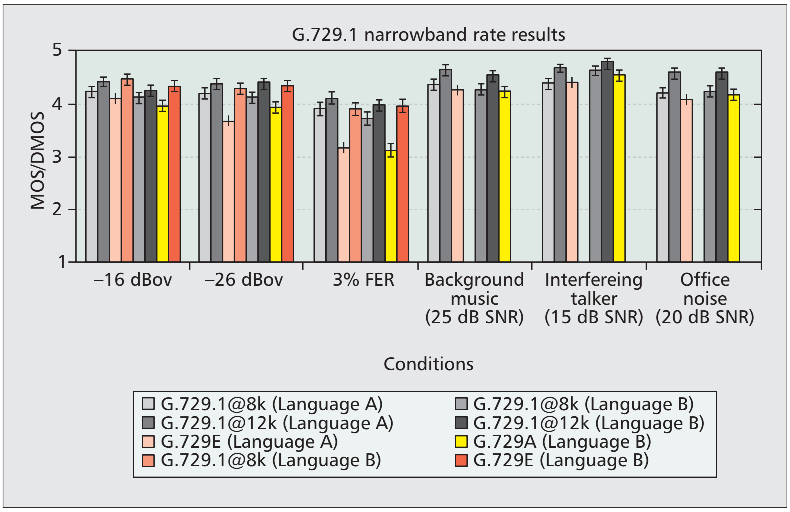

Figure 2. *Performance for narrowband speech. MOS in clean speech conditions; languages: A = French; B = American English. DMOS in noisy speech conditions; languages: A = Korean, B = German.*

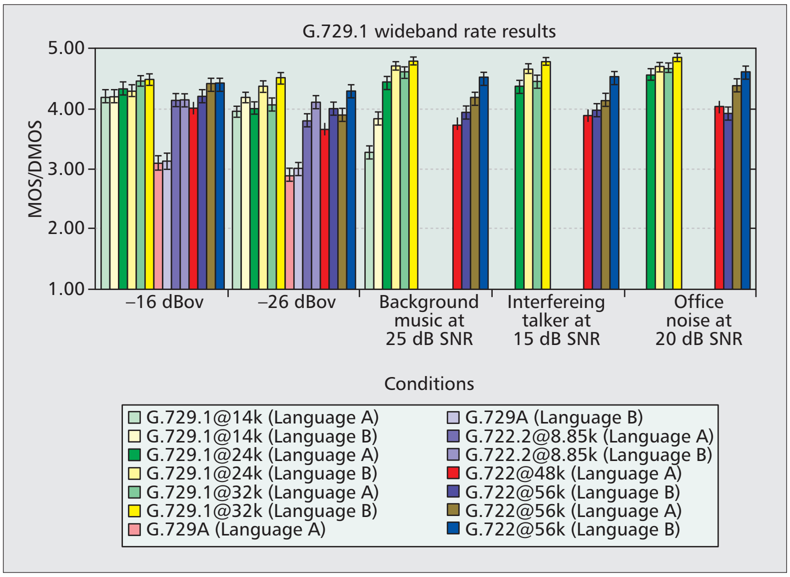

Figure 3. *Performance for wideband speech (step 1 experiments). MOS in clean speech conditions; languages: A = French, B = American English. DMOS in noisy speech conditions; languages: A = Japanese; B = English.*

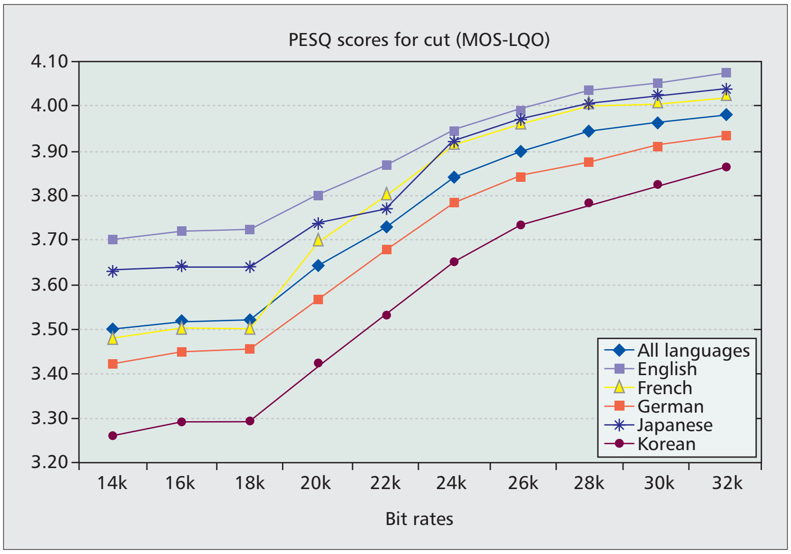

Figure 4. *P.862.2 scores vs. bit rate granularity for WB clean speech.*

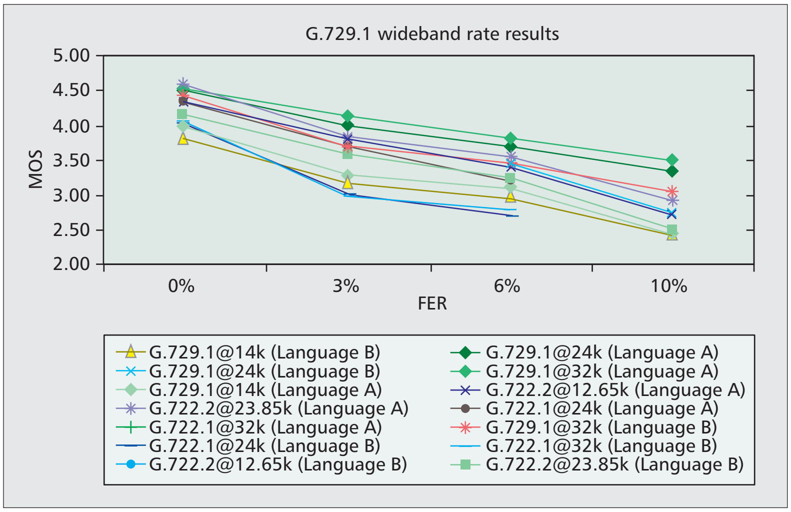

Figure 5. *Robustness to frame erasures (characterization step 2, experiment 2).*

conditions (up to 10 percent FER for G.722.2 and G.729.1 codecs).

The complete results can be found in [7, 8]. Their analysis showed that G.729.1 met all requirements and passed many objectives:

- In clean speech conditions, G.729.1 at 8 kb/s is better than G.729 Annex A and, at 12 kb/s, it is equivalent to G.729 Annex E. At 14 kb/s it is better than G.729A at 8 kb/s and G.722.2 at 8.85 kb/s.
- In noisy speech conditions, G.729.1 at 8 kb/s is equivalent to G.729A (it is even better for two types of background noise — office and babble), and at 12 kb/s, it is better than G.729A.
- At 24 and 32 kb/s, it is better than G.722 at 48 and 56 kit/s (respectively) for speech in various conditions.
- Furthermore, music quality at 32 kb/s is good for a conversational coder (equivalent to G.722 at 56 kb/s).
- G.729.1 robustness is superior to most current standards in the presence of frame erasures, and its quality remains higher than 3.5 mean opinion score (MOS) even at high rates of packet loss (FER 6–10 percent).

COMPLEXITY AND DELAY

The complexity is measured on 16-bit fixed-point ANSI C code, which uses the basic operators as defined in ITU-T G.191 Software Tool Library [6]. The basic operators reflect a model digital signal processor (DSP) and include bit-exact routines of typical functions used in signal processing. The observed worst case complexity of the G.729.1 coder (encoder plus decoder) is 35.79 WMOPS at 32 kb/s. The algorithmic delay is 48.94 ms; the low-delay mode reduces it to 25 ms at 8–12 kb/s, and to 28.94 ms at 14 kb/s. Table 2 summarizes the complexity and delay figures.

CONCLUSION

This article describes the new ITU-T G.729.1 scalable codec for new wideband services and its standardization process by ITU-T SG16. The main motivation for the work was to boost the evolution from narrowband to high-quality wideband VoIP by offering a single narrowband and wideband coding format that would ease deployment in networks and adapt to the demands of all integrated services over IP. This can then limit the proliferation of multiple coding formats and the related deployment costs and quality degradations due to transcoding between different compression schemes.

Wideband coding provides naturalness of speech by enhanced sound quality in conversational applications. The layered scalable structure provides high flexibility to smoothly and finely improve the quality by increasing the bit rate with the best possible network efficiency. It also ensures interoperability with legacy systems thanks to a core layer bitstream interoperable with the existing and successful G.729 standard that is widely used in VoIP today.

After the approval of the standard in ITU-T in May 2006, several steps were taken toward introduction of G.729.1 to the market. First field tests were reported to ITU-T in 2008.

Furthermore, G.729.1 was included by other standards development organizations: it has been adopted as a Korean national standard, the European Telecommunications Standards Institute (ETSI) has recommended G.729.1 as an optional codec for next-generation (NG)-DECT, and G.729.1 is on the list of the recommended codecs for next-generation networks (NGNs) in TISPAN.

ACKNOWLEDGMENTS

The ITU-T G.729.1 Recommendation is a result of the efforts of many contributing companies and testing laboratories. The authors would like to thank ITU-T SG16 and SG12 delegates for their contributions.

REFERENCES

[1] B. Geiser *et al.*, "Embedded Speech Coding: From G.711 to G.729.1," Ch. 8, *Advances in Digital Speech Transmission*, R. Martin, U. Heute, and C. Antweiler, Eds., Wiley, Jan. 2008, pp. 201–47.
[2] ITU-T Rec. G.729.1, "An 8–32 kb/s Scalable Wideband Coder Bit-Stream Interoperable with G.729," May 2006;

http://www.itu.int/rec/T-REC-G.729.1/en
[3] ITU-T Rec. G.729, "Coding of Speech at 8 kb/s Using Conjugate Structure Algebraic Code-Excited Linear Prediction (CS-ACELP)," Rel. 2.0, Dec. 2006; http://www.itu.int/rec/T-REC-G.729/en
[4] S. Ragot *et al*., "ITU-T G.729.1: An 8-32 kb/s Scalable Coder Interoperable with G.729 for Wideband Telephony and Voice over IP," *Proc. ICASSP*, Apr. 2007.
[5] P. Setiawan *et al*., "On the ITU-T G.729.1 Silence Compression Scheme," *EUSIPCO*, Lausanne, Switzerland, 2008.
[6] ITU-T G.191 STL-2005 Manual, "ITU-T Software Tool Library 2005 User's Manual," 2005; http://www.itu.int/rec/T-REC-G.191/en
[7] ITU-T TD 202 (GEN/16)-Attachment 1, "G.729EV Characterisation/Optimisation Step1: Summary of Results ," Apr. 2006 (source: Q7/12 Rapporteurs), available from ITU-T ties service.
[8] ITU-T TD 258 (GEN/16)-Attachment 2, "Executive Summary of G729.1 Characterisation Step 2 — Experiments 1, 2, & 3," Nov. 2006 (source: Q7/12 Rapporteurs), available from ITU-T ties service.
[9] IETF RFC 4749, "RTP Payload Format for the G.729.1 Audio Codec," 2006; http://www.ietf.org/rfc/rfc4749.txt
[10] B. Geiser *et al*., "Bandwidth Extension for Hierarchical Speech and Audio Coding in ITU-T Rec. G.729.1," *IEEE Trans. Audio, Speech, & Language Processing*, Nov. 2007, vol. 15, no. 8, pp 2496–2509.

BIOGRAPHIES

IMRE VARGA [SM] (imre.varga@ieee.org) received his M.Sc. and Ph.D. (summa cum laude) degrees from the Technical University of Budapest, Hungary, in 1982 and 1985, respectively. He served as a member of staff at the same institution between 1984 and 1990, working in the areas of network and filter theory, sigma-delta modulators, and adaptive filtering with applications for telecommunications. Between 1990 and 1994 he was responsible for the development of signal processing algorithms for professional audio at Barco-EMT, Lahr, Germany. He worked for Thomson Multimedia Corporate Research, Hannover, Germany between 1994 and 1997 as project leader and technical advisor for multimedia communication, speech coding, audio processing, command-and-control, and NLP-based dialog systems. He joined Siemens AG, Munich, Germany, in 1997 as R&D group leader of signal processing activities for mobile telephony; later he became department head for multimedia R&D, focused on speech and multimedia applications. Since early 2009 he has worked for Qualcomm Inc. He participates in standardization groups for speech coding and multimedia applications at ITU-T, 3GPP, and IMTC. He is the author of a number of technical papers.

STÉPHANE PROUST graduated from the Ecole Nationale Supérieure des Télécommunications de Bretagne in 1992. He joined France Telecom Research Center (CNET), and took an active part in the development and standardization process of speech coding algorithms within ITU-T and ETSI-GSM/3GPP. From 1998 to 2004 he took various responsibilities as project leader and team manager in the Network and IT Division of France Telecom, and participated in the deployment of the broadband network and operating support system for multimedia/VoIP services. In 2004 he was appointed as Voice and Video Media Standardization senior manager. Since 2004 he has been in charge of managing for France Telecom/Orange Group the standardization activities related to audio, video, and quality assessment areas, and participates actively in related ITU-T and ETSI/3GPP standardization activities.

HERVÉ TADDEI received his Dipl.-Ing. degree in electronics and computer science from the ENSSAT, Lannion, France, in 1995. He received a Ph.D. degree in signal processing and telecommunications in 1999 from the University of Rennes I, France. His research work was conducted at France Telecom R&D, Lannion, France, and focused on scalable speech and audio coding. In 2000 he worked on joint source and channel coding with Lucent Technologies Bell Labs, Murray Hill, New Jersey. From 2001 to 2008 he worked for Siemens, Munich, Germany. In 2008 he joined Huawei Technologies, Munich, Germany. His research interests cover the area of speech/audio coding and transmission, especially focusing on MPEG, 3GPP, and ITU-T standardization activities.