G.719: The First ITU-T Standard for High-Quality Conversational Fullband Audio Coding

Anisse Taleb, Ericsson

Stefan Karapetkov, Polycom

ABSTRACT

This article presents an overview of the recently standardized ITU-T G.719 codec, its key technologies, and their impact on audio quality. These technologies, while leading to exceptionally low complexity and small memory footprint, result in high fullband audio quality, making the codec a great choice for any kind of communication devices, from large telepresence systems to small low-power devices for mobile communication.

INTRODUCTION

Speech codecs have been specifically optimized for the needs of telecommunications systems. Traditionally, speech codecs have been designed for narrowband telephony, with an audio bandwidth of 300–3400 Hz and sampled at 8 kHz. This historical restriction dates back to the first telephone service established between San Francisco and New York in 1915. A number of narrowband codecs have been adopted by different standardization fora, notably the International Telecommunication Union — Telecommunication Standardization Sector (ITU-T) G.711, ITU-T G.729, and the Third Generation Partnership Program (3GPP) Adaptive Multirate Narrowband (AMR-NB), to name a few.

With increased penetration of end-to-end digital communications systems, such as second- and third-generation wireless systems, integrated services digital network (ISDN), and voice over IP (VoIP), wideband telephony became possible. The ever increasing bit rates available for coding speech allow the design of wideband codecs offering a larger audio bandwidth, from 50 Hz to 7 kHz. The latter enables higher intelligibility, better differentiation between certain sounds (e.g., between (s) and (f)), and overall a more natural voice communication service.

Several standardization bodies recognized the benefit of and need for a wideband speech coding standard. In 1988 the ITU-T took the lead and produced G.722 as the first international wideband speech coding standard. This was later followed by G.722.1 and 3GPP Adaptive Multirate Wideband (AMR-WB)/ITU-T G.722.2 which offer wideband speech coding at lower rates.

Figure 1 illustrates how, for some signals, a large portion of energy is beyond the wideband frequency range. While the use of wideband speech codecs primarily addresses the requirement of intelligibility, the perceived naturalness and experienced quality of speech can be further enhanced by providing a larger acoustic bandwidth. This is especially true in applications such as teleconferencing where a high-fidelity representation of both speech and natural sounds enables a much higher degree of naturalness and spontaneity. The logical step toward the sense of being there is the coding and rendering of super-wideband signals with an acoustic bandwidth of 14 kHz.

The response of ITU-T to this increased need for naturalness was standardization of the G.722.1 Annex C extension in 2005. More recently, this has also led ITU-T to start work on extensions of the G.718, G.729.1, G.722, and G.711.1 codecs to provide super-wideband telephony as extension layers to these wideband core codecs.

While speech remains the primary means of communication, conferencing systems are increasingly used for more elaborate presentations, often including music, sound effects, and video. This in turns puts new requirements on today's communication codecs, which are expected to deliver excellent quality on a much wider range of input signals. One could even argue that the success of the new wideband and super-wideband codecs has triggered an increased demand for teleconferencing applications that are able to render true high-quality sound. To answer this strong demand for low-latency audio coding providing the full human auditory bandwidth, ITU-T launched the standardization of a new codec with the basic requirement of delivering fullband audio at 20 kHz audio bandwidth.

In March 2007 ITU-T WP3/SG16 approved the Terms of Reference (ToR) and timetable for the "G.722.1 fullband extension." In fact, prior to the adoption of the G.719 name, it was intended that an extension to G.722.1 — to cover a wider bandwidth — would be sufficient. In July 2007 ITU-T WP3/SG16 approved the qualifica-

 0163-6804/09/$25.00

tion test report for G.722.1 fullband extension, and two candidates, submitted by Ericsson and Polycom, were qualified. At first glance, the high-level descriptions of the qualified candidate codecs presented similar structures. However, after further discussion, complementary technologies were identified, which encouraged a joint collaboration between the candidate proponents. In May 2008 the resulting joint effort was completed, leading to a codec spanning rates from 32 kb/s up to 128 kb/s and fulfilling the requirements set for high-quality audio on speech, music and mixed audio material. Finally, in June 2008 the Alternative Approval Process (AAP) was finalized and the codec adopted under the new number, ITU-T G.719 [1].

This article analyzes the requirements and design constraints for high-quality conversational audio, and explains how these are met by the new ITU-T Recommendation G.719. It takes a deeper look at the algorithms that are part of G.719 — focusing on adaptive time-frequency transform and fast lattice vector quantization — and argues that these algorithms led to the low complexity, small footprint, and high performance of G.719.

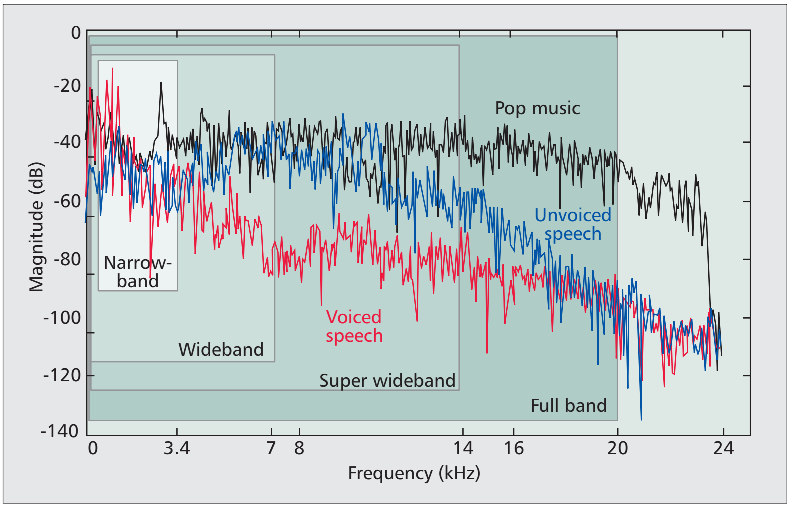

Figure 1. *Observed spectrum of different sounds, voiced speech,[1] unvoiced speech, and pop music on different audio bandwidths.[2]*

Requirements and Design Constraints for High-Quality Conversational Audio

Audio codecs for use in telecommunications face tougher requirements than general-purpose media codecs. Much of this comes from the need for standardized interoperable algorithms that deliver high-quality sound at low latency and low bit rate. In addition, low computational and memory costs are extremely important, since they determine which hardware can support the codec. A low-complexity codec can be incorporated in communication devices that span the range from portable, battery powered, low-cost devices to high-end immersive room systems.

New communications and telepresence systems [2] provide high-definition video and high-quality audio to the user, and require a corresponding quality of media delivery to fully create the immersive experience. While most of the focus has been on improved video quality, superior audio quality contributes as much to making the interaction smooth and natural.

In standardizing the G.719 codec, the following features were considered to be the most relevant:

- Input and output audio signals had to have a bandwidth of 20 kHz at a sampling rate of 48 kHz.
- Low computational complexity was an important objective.
- Primary signals of interest were open-microphone speech with office and conference room background noise, with and without multiple talkers. Music, natural sounds, and clean speech were of secondary interest but had to be rendered adequately.

The initial requirements included support for bit rates of 32 kb/s, 48 kb/s, and 64 kb/s. The lowest rate of 32 kb/s was required for flexibility (i.e., to adjust to video bit rate requirements or to the needs of forward error correction [FEC]). Since ITU-T H.221 limits the usable rate to 48 kb/s, this rate was needed for codec use in ISDN videoconferencing, according to ITU-T H.320. Finally, the 64 kb/s rate was intended for high-quality audio when the application had a sufficient bit rate for audio stream encoding. Not only does G.719 fulfill the requirements regarding bit rates, it exceeds them by providing a larger set of bit rates from 32 kb/s up to 128 kb/s at virtually no additional cost. Overall, a fine granularity of 4 kb/s, at low bit rates, allows a flexible selection of bit rates ranging from 32 kb/s and up to 96 kb/s; the granularity increases to 8 kb/s up to the maximum rate of 128 kb/s for higher-quality audio.

As ITU-T audio codecs are designed for live interactive traffic, they are fully specified using a normative source code for both encoder and decoder. On the other hand, MPEG being primarily concerned with media distribution, MPEG audio coding standards define a bitstream format and a decoder specification, and do not impose any constraints on the encoders as long as the generated bitstream follows the specification and is understandable by the decoder. Driven by the needs of the broadcasting industry, this approach allows more flexibility for encoder implementations and enables further performance enhancements long after the standard has been adopted. The drawback of this approach is that it results in substantial performance variations among different encoder implementations of the same standard. Since strict and consistent control of quality is required in communication applications, ITU-T Recommendation G.719 specifies, in fixed-point, both the encoder and the decoder [1]. In addition, an alternative floating-point implementation is also provided in G.719 Annex B, approved in November 2008.

One of the challenges in performing the subjective quality tests of G.719 is access to a challenging and, more important, open reference codec. The reference codec is usually a state-of-the-art codec to which the relative performance

[1] Voiced speech is produced when the vocal cords vibrate during the pronunciation of a phoneme. Unvoiced speech, by contrast, does not entail the use of the vocal cords.

[2] Spectrum above 20 kHz is not perceivable by the human auditory system.

The adaptation of the transform resolution to the character of the incoming sound is a key parameter of the encoding process.

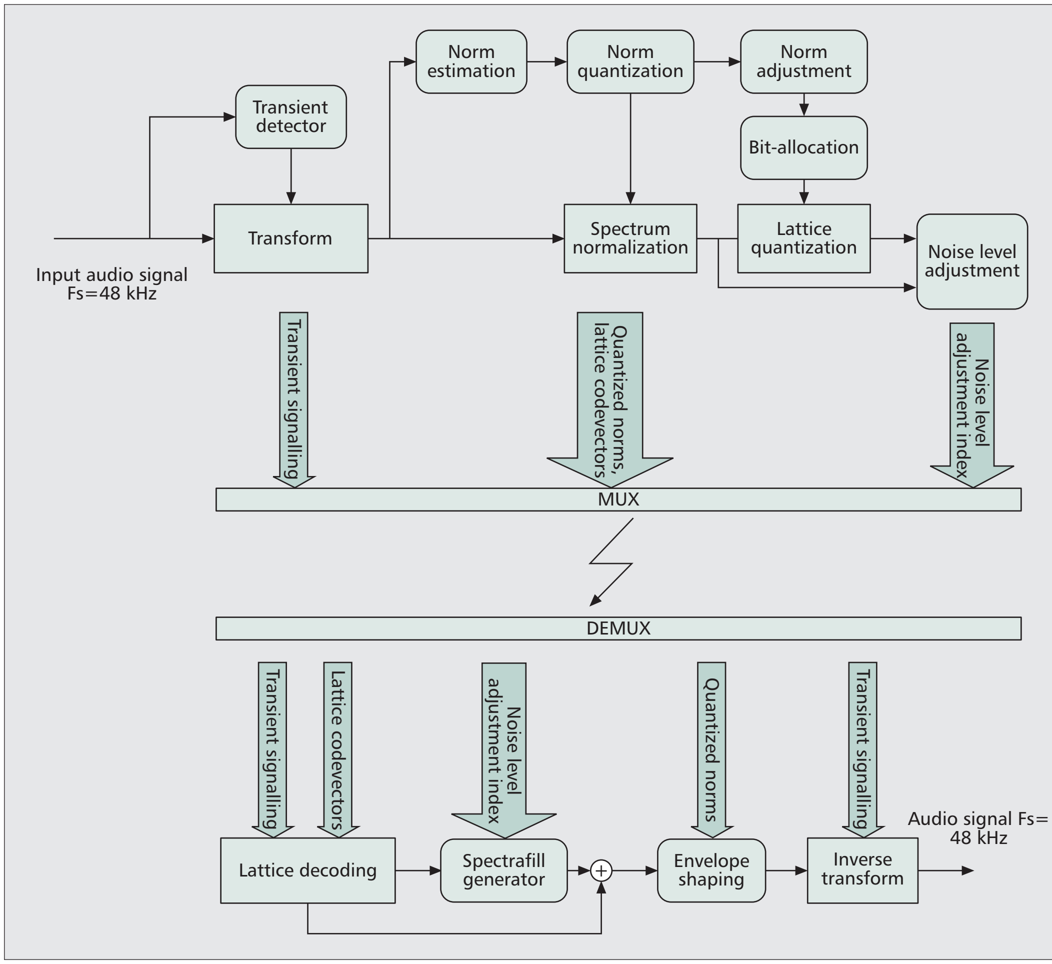

Figure 2. *G.719 codec high-level block diagram.*

of the codec under test (in this case G.719) is measured. As discussed above, MPEG does not specify any encoder; therefore, the LAME MP3 [3] — one of the best available open and highly optimized MP3 encoders — was used for the comparative tests. One could even argue that such comparison was unfair as it disadvantaged G.719 since the LAME MP3 was running at much higher delay and complexity.

TECHNICAL OVERVIEW OF ITU-T G.719

OVERVIEW OF THE CODEC

Figure 2 shows a high-level[3] block diagram of the G.719 codec. At the encoder, the input signal sampled at 48 kHz is first processed through a transient detector. The codec processes the signal in frames of 20 ms. Depending on the detection of a transient, a high-frequency resolution or low-frequency (and higher temporal) resolution transform is applied on the input signal overlapped frames. The adaptation of the transform resolution to the character of the incoming sound is a key parameter of the encoding process. This critical element of the codec is detailed later in this article.

The adaptive transform is based on a modified discrete cosine transform (MDCT) in case of stationary frames and a variant of the former for non-stationary frames. The MDCT is a real valued Fourier-related transform. It is based on the type-IV discrete cosine transform (DCT-IV), with the additional property of being lapped, that is, subsequent signal blocks are overlapped: the last half of one block coincides with the first half of the next block. Because of the overlapped frame processing, the codec overall delay is twice the frame length (i.e., 40 ms). In addition to the energy-compaction properties of the DCT-IV, the MDCT has proven especially attractive for signal compression applications, since the lapped property helps avoid artifacts stemming from the block boundaries.

The obtained spectral coefficients are grouped into bands of unequal lengths. This unequal division into frequency bands, small at the low frequencies and increasingly large for the high frequencies, is based on the human ear auditory filters' bandwidths. This allows an approximate but efficient implementation of the psycho-acoustical principles with a better allocation of resources to the most psycho-acoustically important signal components.

The norm of each band is estimated, and the resulting spectral envelope consisting of the norms of all bands is quantized and encoded. The algorithm efficiently encodes the difference between the norms of two adjacent frequency bands; this results in smaller values than if the values themselves were encoded, leading to

[3] *For the sake of clarity some low-level decoder functions are omitted.*

much fewer bits. To further increase efficiency, a Huffman codebook is used to encode the norm differential indices, if it requires less bits than a uniform bit representation.

The quantized norms are used to normalize the spectrum. The quantized norms are further adjusted based on adaptive spectral weighting. This weighting implements a low-complexity psycho-acoustical weighting, increasing the norm of psycho-acoustically important frequency bands and decreasing the norm of masked and irrelevant frequency bands.

Given a certain amount of bits, the bit allocation allocates more bits to the frequency bands whose modified norm is large while assigning fewer bits to the low norm frequency bands. This allocation efficiently optimizes the use of the available bits to maximize the perceived quality at the decoder for a given bit rate.

The normalized spectral coefficients are quantized and encoded based on the allocated bits for each frequency band. Each spectral band is subdivided into smaller vectors of eight dimensions and then quantized by a Fast Lattice Vector Quantization (FLVQ) scheme. The FLVQ algorithm is a key element in reducing the complexity and memory footprint of G.719, and is detailed later.

The final element in the G.719 encoder is the comfort noise level adjustment. In fact, the bit allocation algorithm sometimes allocates zero bits to certain frequency bands because there are not enough bits to encode the entire spectrum; this is frequently so at low bit rates. The level of the non-coded spectral coefficients is transmitted to the decoder to adjust the level of the non-coded frequency bands artificially regenerated.

The G.719 decoder (Fig. 2) performs the inverse operations of the encoder. The transient flag is first decoded which indicates the frame configuration (i.e., stationary or transient). The spectral envelope is decoded, and the same bit-exact norm adjustments and bit allocation algorithms are used at the decoder to recompute the bit allocation. Then the quantization indices of the normalized transform coefficients are decoded. After dequantization, an extension technique intelligently reuses the lower frequency coefficients to non-coded frequency bands [1, 4]. Low-frequency non-coded spectral coefficients (allocated zero bits) are regenerated by using a spectral-fill codebook built from the received spectral coefficients (spectral coefficients with non-zero bit allocation). High-frequency non-coded spectral coefficients are regenerated using bandwidth extension. A noise level adjustment index is used to adjust the level of the regenerated coefficients.

The decoded spectral envelope is applied to obtain the decoded full band spectrum. Finally, the time-domain decoded signal is recovered by applying the inverse transform of the transform performed at the encoder.

Adaptive Time-Frequency Transform

While for stationary signals MDCT is an efficient coding approach, for non-stationary frames it fails to deliver a compact representation of the signal. In fact, when quantizing the transform coefficients in such a block-based algorithm, the inverse transform at the decoder spreads the quantization noise distortion evenly in time. This results in unmasked distortion in a low-energy region preceding a signal attack. Temporal pre-masking may mask this distortion; however, this is only possible when the transform block size is sufficiently small.

A well-known solution to this problem is block switching [4]. This technique is based on the idea of changing the time resolution of the transform upon transient detection. Typically the analysis block length is changed from a long duration during stationary signals to a short duration when transients are detected. However, to satisfy the perfect reconstruction constraints, window switching with lapped transforms such as the MDCT requires insertion of transition windows between short and long blocks. This generates increased delay as this switching cannot be done instantaneously; hence, a long look-ahead for transient detection is needed.

In contrast, G.719 uses a novel technique to *instantaneously* obtain higher temporal resolution transform without requiring additional delay and with very little overhead in complexity. The resulting non-stationary frames have a temporal resolution equivalent to that of 5 ms frames. The codec operates normally on 20 ms long frames; however, as soon as a transient is detected in the current frame, the codec can instantaneously switch to a higher time resolution by decomposing the frame into smaller 5 ms equivalent frames.

The G.719 adaptive transform switches windows in a block-based fashion. A time aliasing operation is applied to each overlapped frame. If no transient is detected, a DCT-IV application leads to the MDCT of the input frame. However, if a transient is detected, the time aliased signal is reordered in the time domain and further split into smaller 5 ms subframes. These are further post-windowed and transformed to the frequency domain. Since all operations are performed in the same time frame, neither transition windows nor additional look-ahead in the transient detection are needed. Thanks to its adaptive transform, G.719 better represents the rapid changes in the input signal spectral characteristics and provides improved transient encoding.

FLVQ

It is well known that structurally unconstrained vector quantization (UVQ) leads to optimal performance for a given bit rate. However, UVQ suffers from large codebook storage requirements and the complexity of the nearest neighbor search.

Lattice vector quantization (LVQ) is a constrained vector quantization where the codebook entries are points of a lattice [5]. The actual lattice codebook does not need to be explicitly stored since all lattice points can be generated from a so-called generator matrix.

The G.719 Fast LVQ algorithm consists of two subquantizers: an RE_8-based lower-rate lattice vector quantizer named LVQ1 and a D_8-based higher-rate lattice vector quantizer named LVQ2 [5]. When a low rate of a 1 b/sample is allocated, LVQ1 is used; otherwise, for higher bit rates, LVQ2 is used. The codebooks for the

The G.719 Fast LVQ algorithm consists of two subquantizers: an RE_8-based lower-rate lattice vector quantizer named LVQ1 and a D_8-based higher-rate lattice vector quantizer named LVQ2.

Bit rate (kb/s)	Encoder only		Decoder only		Encoder plus decoder	
	Average	Maximum	Average	Maximum	Average	Maximum
32	6.7	8.0	6.9	7.4	13.5	15.4
48	7.4	9.1	7.2	7.8	14.7	16.9
64	7.9	9.9	7.5	8.2	15.5	18.1
80	8.3	10.6	7.9	8.5	16.2	19.0
96	8.6	10.8	8.1	8.8	16.7	19.5
112	8.8	10.9	8.4	9.0	17.2	19.9
128	9.0	11.8	8.6	9.2	17.6	21.0

Table 1. *Complexity of the G.719 codec in WMOPS.*

Memory type	Encoder	Decoder	Codec (encoder plus decoder)
Static RAM (kwords)	1.0	3.9	4.9
Scratch RAM (kwords)	12.2	12.2	24.4
Data ROM (kwords)	8.3	8.9	10.7
Program ROM (1000s of basic ops)	1.2	1.2	1.8

Table 2. *Memory usage of the G.719 codec.*

quantizer are constructed from a finite region of the lattices and match the probability density function of the normalized spectral coefficients. Algebraic search and indexing procedures are less complex with LVQ2 (D_8) than with LVQ1 (RE_8).

Finding the nearest neighbor and indexing consist of simple algebraic and rounding operations that contribute greatly to the low complexity of G.719. In addition, Huffman coding is used on the resulting indices to further improve the coding efficiency.

Complexity and Memory Requirements

Complexity is a paramount parameter for a codec. Complex codecs require more powerful and more expensive digital signal processors (DSPs) to run on; this increases the product cost and power consumption, which limits the codec usability. This is especially true in telephones and more so for mobile devices in which complexity translates directly to battery lifetime. Similarly, the memory footprint of a codec has direct impact on the product cost and codecs with large memory footprints are not appropriate for the majority of communication devices.

The fixed-point C-code implementation of G.719, which is an integral part of the Recommendation, is based on a set of instructions that mimics a generic DSP instruction set. These instructions, the so-called basic operators, are defined in ITU-T Recommendation G.191 [6]. Each basic operator is assigned a weight which reflects the number of cycles corresponding to that operator, resulting in complexity estimates called weighted million operations per second (WMOPS). The complexity on a DSP is usually measured in million instructions per second (MIPS). The ratio between the estimated WMOPS and MIPS depends on the DSP used and the level of optimization used. The deviation between WMOPS and MIPS has been found in practice to be on the order of ±20 percent.

The complexity of the G.719 codec in WMOPS is reported in Table 1 and clearly shows that the codec complexity is low. To be more precise, these absolute WMOPS figures should be compared to the complexity of well-known codecs. For instance, AMR-WB/G.722.2 has a total complexity of around 34 WMOPS in the 12.65 kb/s mode, which is significantly higher than that of G.719 at any rate. The table also shows that the incremental complexity for encoding at higher bit rates is small. It only requires approximately two additional WMOPS to double the bit rate from 64 kb/s to 128 kb/s. The highest value encountered is 21 WMOPS, which is still very low.

These excellent numbers for a full-audio-bandwidth high-performance codec essentially mean that the G.719 codec is very well suited to run on an inexpensive DSP. It also means that the complexity of G.719 is such that it frees DSP resources for other processing aspects of the application (e.g., state-of-the-art acoustical front-end processing).

G.719 storage requirements, such as static RAM, scratch RAM, and data ROM — measured in 16-bit kwords — are given in Table 2. However, the program ROM is measured as the number of basic operators (and function calls) in the fixed-point ANSI-C source code of the G.719 codec.

It is important to point out that since the primary application of the codec is a communication scenario, the encoder and decoder can share tables and code. The memory savings are reflected in the last column, where the ROM figures are less than the sum of the ROM needed for the encoder and decoder separately.

Transport and File Format of ITU-T G.719

For VoIP applications, G.719 coded frames are transported over Real-Time Transport Protocol (RTP) [7] using the G.719 RTP payload format [8]. This format supports encapsulation of one or multiple G.719 frames per packet, beneficial to packetization overhead and also multichannel audio sessions.

A multirate encoding capability to vary the encoding rate on a per frame basis is also supported, enablingthe G.719 codec to operate as source controlled variable bit rate as well as to react to network bandwidth variations by adapting the encoder bit rate.

The G.719 RTP payload provides a means of redundancy transmission and frame interleaving to improve robustness against possible packet loss. It enables generic FEC functionality as well as a G.719-specific form of audio redundancy coding. Conceptually, previously transmitted transport frames are aggregated together with new ones. A sliding window can be used to group the frames sent in each payload.

ITU-T G.719 compressed audio can be stored in a file using the ISO-based container file and is specified in Annex A to G.719 [1]. Note that the ISO base media file format structure is the basic building block of several application-derived file formats, such as 3GP and MP4 file formats. This allows storage of G.719 along with many other multimedia formats, as well as synchronized playback of G.719 and other audiovisual media.

Quality Performance of ITU-T G.719

The content used for testing the codec quality consisted of music, mixed content, and speech. Extensive listening tests have been performed on the G.719 codec by several international listening laboratories using the BS.1116 methodology [9]. Two experiments were conducted. Experiment 1 assessed clean speech with and without reverberation, reverberant speech with office noise and interfering talker, and also included 3 percent frame erasure conditions, whereas music and mixed content were tested in experiment 2. For music, several types were used, including pop and classical music but also vocals in different languages. Each experiment was run twice using different audio materials and, for each run, 24 native speakers of the language. Experiment 1 was run in American English and French, experiment 2 in American English and Spanish. Figure 3 shows the codec performance in these two experiments (Fig. 3a for experiment 1, Figs. 3b and 3c for experiment 2; the error bars indicate the 95 percent confidence interval).

In experiment 1 the G.719 codec passed all nine of the *Requirement* ToR tests in American English. Those tests specified that the G.719 should be "not worse than" the reference coder, LAME MP3. In fact, in all nine Requirement ToR tests, the G.719 outperformed the codec and was significantly "better than" the reference codec. The results in French were similar.

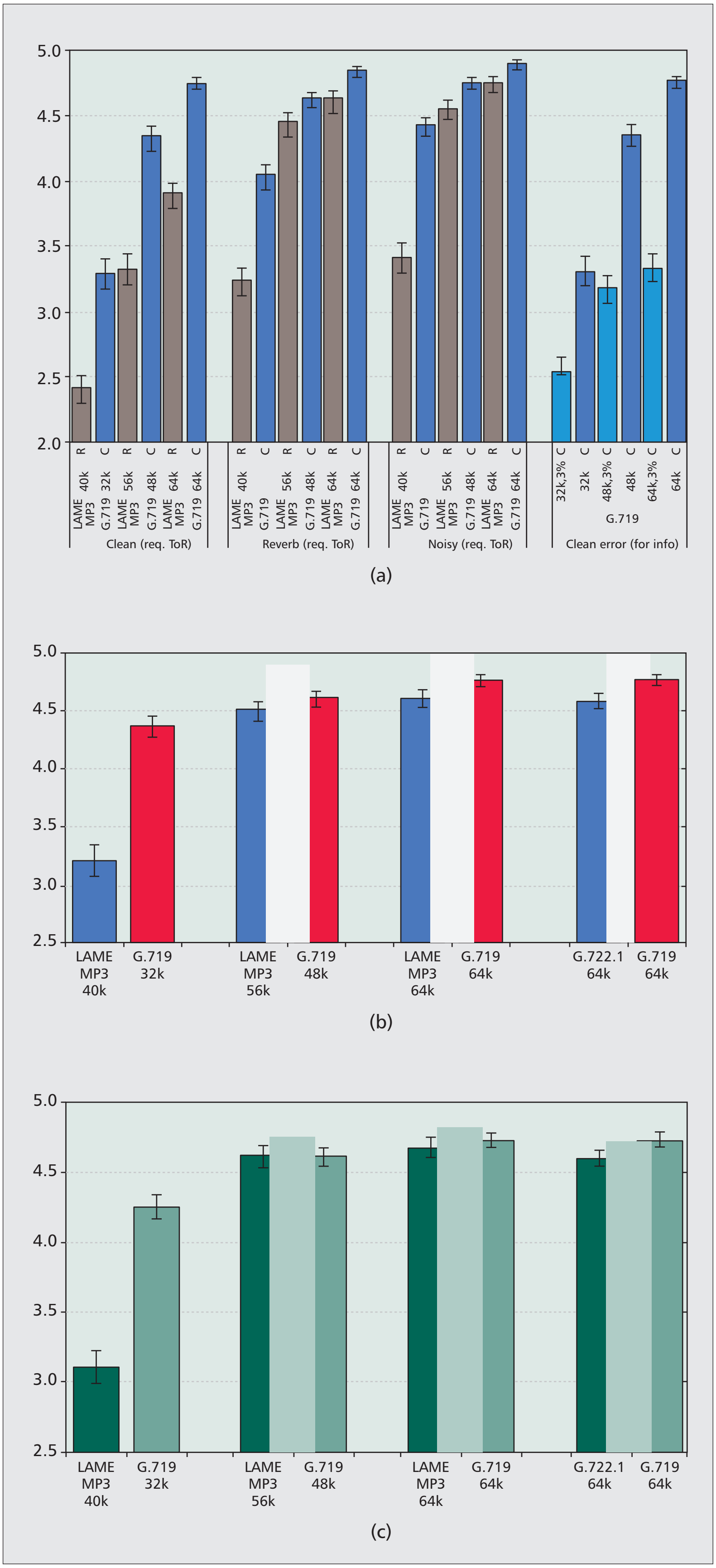

Figure 3. *ITU-T performance test results: average BS.1116 scores shown on the Y-axis: a) subjective evaluation on Speech content, clean, reverberant and noisy with interfering talker — North American English; b) subjective evaluation on music and mixed content — North American English; c) subjective evaluation on music and mixed content — Spanish.*

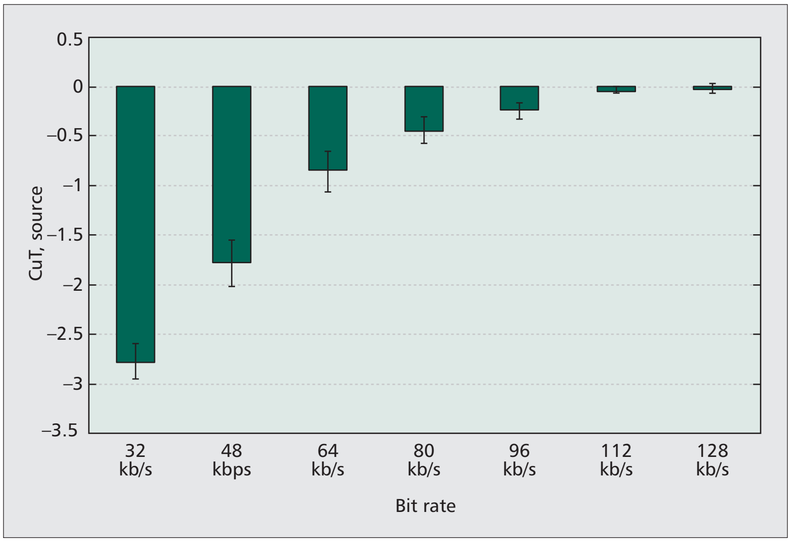

Figure 4. *Average of difference grades as function of the bit rate (difference between the BS.1116 scores of the source material and the coded material).*

For experiment 2, while the two language tests gave slightly different results, they were still very consistent, correlating almost perfectly. Comparisons across the two experiments showed that all requirements for music and mixed content were passed by the G.719 codec. It is remarkable that the quality of G.719 is better or equal to that of MP3 on music given that the latter has been optimized for music without delay and complexity constraints.

The official ITU-T results focus on the rate operation points of paramount importance for the intended applications. To assess G.719 quality on other operating points, an additional subjective listening test was conducted on preselected critical material. Figure 4 shows the difference between the scores given to the codec under test (CuT), G.719 in this case, and to the original source material (i.e., non-coded signal). The score, represented by a negative figure, measures the subjective distance of the coded material with respect to the original material. A figure close to zero means that the codec is indistinguishable from the original. These results clearly show the benefit of the higher rates in improving the quality of the codec up to transparency, as seen in Fig. 4 where the quality gracefully improves until transparency at 128 kb/s.

Conclusion

ITU-T has a long history of developing and standardizing codecs targeting the telecommunications industry. A significant portfolio of codecs, the G.71x and G.72x series, has been built addressing various applications implying various constraints on the codec. The first ITU-T full band codec, G.719, delivers exceptional performance at low, medium, and high bit rates, and promises to become the one codec for unified communications, replacing many, if not all, of its predecessors.

The communications industry is beginning to adopt the new ITU-T G.719 standard and leverage its capabilities. It is expected that the first products with "G.719 inside" will appear in 2009. The wide acceptance of this new open standard will benefit users, vendors, and IT organizations by allowing audio-video equipment from all suppliers to communicate at new levels of audio clarity and efficiency.

Acknowledgments

The authors wish to thank all the contributors to the G.719 codec; without their significant contributions this achievement would not have been possible. They are Minjie Xie, Roni Even, Peter Chu, Manuel Briand, Gustaf Ullberg and Jonas Svedberg. Special thanks go to Jeff Rodman, Hans Hermansson, Claude Lamblin, and Simaõ Campos.

References

[1] ITU-T Rec. G.719, "Low-Complexity, Fullband Audio Coding for High-Quality, Conversational Applications," June 2008; http://www.itu.int/rec/T-REC-G.719/en

[2] ITU-T, "High-Performance Video-Conferencing," *Technology Watch Report 2*, Nov. 2007; http://www.itu.int/oth/T2301000002/en

[3] LAME MP3; http://lame.sourceforge.net/

[4] T. Painter and A. Spanias, "Perceptual Coding of Digital Audio," *Proc. IEEE*, vol. 88, no. 4, Apr. 2000, pp. 451–515.

[5] R. M. Gray, "Quantization," *IEEE Trans. Info. Theory*, vol. 44, no. 6, Oct. 1998, pp. 2325–83.

[6] ITU-T Rec. G.191, "Software Tools for Speech and Audio Coding Standardization," Sept. 2005; http://www.itu.int/rec/T-REC-G.191/en

[7] H. Schulzrinne *et al.*, "RTP: A Transport Protocol for Real-Time Applications," IETF RFC 3550, July 2003.

[8] M. Westerlund and I. Johansson, "RTP Payload Format for G.719," IETF RFC 5404, Jan. 2009.

[9] ITU-R Rec. BS.1116, "Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems," 1997; http://www.itu.int/rec/R-REC-BS.1116/en

Biographies

ANISSE TALEB (anisse.taleb@huawei.com) received his M.Sc. and Ph.D. degrees in electrical engineering from the Institut National Polytechnique de Grenoble (INPG), France. He was awarded the INPG Best Thesis Prize in 2000. After a post-doctoral research fellow position at the Australian Telecommunication Research Institute, he joined Ericsson in 2001, where he was a senior research engineer and project leader in the field of audio technology for over eight years. He currently holds the position of senior expert within the Multimedia Communication Laboratory at Huawei Technologies. He was involved in the development and standardization of the AMR-WB+ audio codec and the ITU-T G.719 codec for which he is the Recommendation editor. He is active in standardization bodies such as MPEG, ITU-T, and 3GPP. His research interests include speech and audio coding, robustness in packet networks, multichannel coding, and scalable coding. He has numerous publications and patents in the area of speech and audio coding.

STEFAN KARAPETKOV (stefan.karapetkov@polycom.com) received an M.Sc. degree in electrical engineering from the University of Chemnitz, Germany, in 1995, and an M.B.A. from Santa Clara University, California, in 2001. From 1995 to 2006 he worked in product management, new technology development, and enterprise product definition at Siemens Enterprise Networks in Germany and the United States. He developed mechanisms for advanced interworking between in-building wireless and public mobile networks (patents US6151506, EP987912A3, EP929986A2, and EP929984A2) and architecture for new connectionless services in data networks (patents US6205490, US6137799, US6081836, and US5974452). Since 2006 he has been emerging technologies director at Polycom where he focuses on visual communications. He has authored numerous white papers and articles about audio and video technology, and has been a frequent speaker at industry events around the world. He is fluent in four languages and blogs at http://videonetworker.blogspot.com/