G.719: The First ITU-T Standard for Full-Band Audio

April 2009

POLYCOM®

INTRODUCTION

Audio codecs for use in telecommunications face more severe constraints than general-purpose media codecs. Much of this comes from the need for standardized, interoperable algorithms that deliver high sound quality at low latency, while operating with low computational and memory loads to facilitate incorporation in communication devices that span the range from extremely portable, low-cost devices to high-end immersive room systems. In addition, they must have proven performance, and be supported by an international system that assures that they will continue to be openly available worldwide.

Audio codecs that are optimized for the special needs of telecommunications have traditionally been introduced and proven starting at the low end of the audio spectrum. However, as media demands increase in telecommunications, the International Telecommunication Union (ITU-T) has identified the need for a telecommunications codec that supports full human auditory bandwidth, that is, all sounds that a human can hear. This has led to the development and standardization of the G.719 audio codec.[1]

ITU-T RATIONALE FOR G.719

In its decision to standardize G.719 in June 2008, ITU-T cited the strong and increasing demand for audio coding providing the full human auditory bandwidth, and referred to several new applications that are driving this requirement.

Conferencing systems are increasingly used for more elaborate presentations, often including music and sound effects. While speech remains the primary means for communication, content sharing is becoming more important and now includes presentation slides with embedded music and video files. In today's multimedia presentations, playback of high-quality audio (and video) from DVDs and PCs is becoming a common practice; therefore, both the encoder and decoder must be able to handle this input, transmit the audio across the network, and play it back in sound quality that is true to the original.

New communications and telepresence systems provide High Definition (HD) video and audio quality to the user, and require a corresponding quality of media delivery to fully create the immersive experience. While most people focus on the improved video quality, telepresence experts and users point out that the superior audio is what makes the interaction smooth and natural. In fact, picture quality degradation has much lower impact on the user experience than degradation of the audio. Since telepresence rooms can seat several dozens of people—for example, the largest Polycom® RealPresence™ Experience (RPX™) telepresence solution has 28 seats—advanced fidelity and multichannel capabilities are required that allow users to acoustically locate the speaker in the remote room. Unlike conventional teleconference settings, even side conversations and noises have to be transmitted accurately to assure interactivity and a fully immersive experience.

Going beyond the technology aspects, ITU-T also stressed in its decision the importance of extending the quality of remote meetings to encourage travel reduction, which in turn reduces greenhouse gas emission and limits climate change.

THE HISTORY OF THE G.719 STANDARD

Polycom thrives through collaboration with partners, and has delivered joint solutions with Microsoft, IBM, Avaya, Cisco, Nortel, Broadsoft, and numerous other companies. The story of G.719 is also a story of partnership with another major player in the communication industry—Ericsson.

In June 2006, Polycom proposed to ITU-T the characteristics for a new 20 kHz bandwidth audio codec based on the Polycom Siren ™ 22 technology. In November, Polycom followed up with a proposal to standardize this technology as 'ITU-T G.722.1 Full-band extension for video- and teleconferencing applications.' The thinking then was that the new codec would be an annex to the existing G.722.1 standard.

In March 2007, ITU-T WP3/SG16 approved the Terms of Reference (ToR) and Timetable for G.722.1 Full-band extension, and two candidates were submitted in April. Polycom submitted a candidate using fixed point integer arithmetic while Ericsson submitted a candidate using floating point. On a technical level, both candidates shared the same high-level building blocks; however, orthogonal technologies were identified that, when combined together, would dramatically improve the performance of the codec. In July 2007, ITU-T WP3/SG16 approved the qualification test report for G.722.1 Full-band extension, and both candidates were qualified.

This led to the collaboration between Ericsson and Polycom that started in September 2007. The first joint draft for the standard became available in October. By February 2008, the collaboration led to a joint submission of the fixed point executables for the

Characterization Test. The codec met all requirements and passed the Characterization Test in April 2008.

By the spring of 2008, ITU-T members realized the importance of the new codec and decided to assign a new number (G.719) instead of creating another annex to G.722.1 (which already had several annexes). The new G.719 number made the codec more visible and highlighted its unique capabilities and place within the ITU-T portfolio.

At the ITU-T Study Group 16 meeting in May 2008, G.719—spanning data rates up to 128 kbps—was officially consented for Alternative Approval Process (AAP). This allowed speedy standardization completion by June 2008 when the AAP was finalized and ITU-T G.719 full-band codec was adopted.

THE PLACE OF G.719 IN THE ITU-T CODEC PORTFOLIO

The standard for voice transmission quality was set about 120 years ago with the invention of the telephone. Based on technical capabilities at the time, it was decided that transmitting the acoustic frequencies from 300Hz to 3300Hz is sufficient for a regular conversation. Even today, basic narrow-band voice encoders, such as ITU-T G.711, work in this frequency range, and are therefore referred to as 3.3 kHz voice codecs. Another important characteristic for a voice codec is the bit rate. For example, G.711 has a bit rate of 64 kbps; that is, transmitting voice in G.711 format requires a network bandwidth of 64 kbps (plus network protocol overhead). All ITU-T audio codecs belong to the G series of standards. Other popular ITU-T narrow-band codecs are G.729A (3.3 kHz, 8 kbps) and G.723.1 (50 Hz-3.3 kHz, 5.3-6.3 kbps).

Advanced audio coding went far beyond basic voice quality with the development of wide-band codecs supporting 7 kHz, 14 kHz, and most recently 20 – 22 kHz audio. ITU-T G.722 was the first practical wideband codec, providing 7 kHz audio bandwidth at 48 to 64 kbps. Polycom entered the forefront of audio codec development with Siren 7 (7 kHz audio at 16 to 32 kbps), standardized by ITU-T as G.722.1 (7 kHz, 16/24/32 kbps). Figure 1 provides an overview of the most popular codecs.

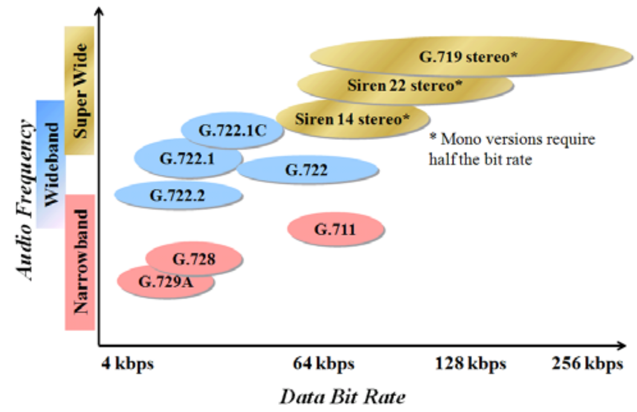

Figure 1: Popular Audio Codecs

Polycom later enhanced and extended Siren 7 into two further wide-band codecs. Siren 14 (14 kHz, 48/64/96 kbps) supports stereo and mono audio; the mono version then became ITU-T G.722.1C (14 kHz, 24/32/48 kbps). Polycom also developed Siren 22 (22 kHz, 64/96/128 kbps stereo and 32/48/64 kbps mono). Siren codecs are supported in communication products (for example, telephones, video endpoints, soft clients, conferencing servers, and recording servers) from Polycom and other vendors.

G.719 is the latest addition to the ITU-T portfolio, and operates at bit rates from 32 kbps to 128 kbps per channel, that is, 64 kbps – 256 kbps for stereo. Its quality is higher than any other ITU-T codec and is expected to be even better than the existing Polycom super-wideband codecs Siren 14 and Siren 22. The initial test results are presented later in this paper.

COMPARING G.719 WITH OTHER SUPER-WIDEBAND CODECS

While it is clear that G.719 outperforms other ITU-T codecs, let's see how it stacks up against super-wideband codecs defined outside ITU-T. Table 1 compares G.719 with its close relative Polycom Siren 22 and with two MPEG codecs: MP3 (officially MPEG-1/2 Layer-3), the popular codec used in portable music players, and MPEG-4 AAC-LD, a new MPEG codec that was developed for interactive communication applications.

	Bandwidth	Bit rate	Delay	Complexity	Standard	Algorithm
SIREN22	22 kHz	32-64 kbps mono 64-128 kbps stereo	40 ms	18.4 MIPS	N/A	Transform Coding
G.719	20 kHz	32-128 kbps mono 64-256 kbps stereo	40 ms	17.7 MIPS	ITU	Transform Coding
MP3	< 20 kHz	64-384 kbps	≥ 54 ms	> 100 MIPS	ISO	Transform Coding
MPEG-4 AAC-LD	< 20 kHz	24-192 kbps (TAA: 64 kbps mono 128 kbps stereo)	≥ 20 ms	> 130 MIPS	ISO	Transform Coding

Table 1: Audio Codecs Comparison

Table 1 highlights the close relationship between G.719 and Siren 22. Bandwidth and delay are similar but, most importantly, G.719 inherits the low complexity of Siren 22 and is therefore a great choice for implementation on telephones and mobile devices.

In comparison, MPEG codecs have much higher complexity and require more powerful and expensive digital signal processors (DSPs). The G.719 complexity value in Table 1 is based on the TriMedia DSP implementation for the floating-point version (G.719 Annex A) mono, and is the maximum value for the worst case scenario (complexity for stereo is twice this number). The TriMedia DSP implementation measures complexity in Million Instructions Per Second (MIPS), a standard measure of complexity.

In order to take audio quality to the next level, the G.719 maximum bit rate was increased to 256 kbps (stereo)—compared to Siren 22 technology's maximum bit rate of 128 kbps (stereo). This doubling of the maximum allowable bit rate allows completely transparent coding of music material. At the same time, the acoustic bandwidth was set to 0Hz – 20 kHz for full compatibility with most high-fidelity and HD audio sources.

Ericsson contributed important algorithms to G.719 that further improve the codec for music transmission, for example, the adaptive time-frequency transform, discussed below, is of great benefit for percussive music sounds. It is therefore expected that G.719 will handle music better than Siren 22.

G.719 ENCODER

The G.719 standard defines both the encoder and the decoder functions. This adheres to the ITU-T approach to communication standards, and allows for more consistent control of audio quality. In comparison, MPEG is primarily concerned with media distribution, and MPEG audio-video standards only define the decoder. Consequently, the encoder can do anything as long as the decoder can understand and play the received audio-video content. The MPEG approach leads to more flexibility in the encoder implementation but less control of the quality, resulting in substantial variation in performance among implementations, a serious disadvantage in communication applications.

Let's look at the G.719 encoder. Once the analogue audio signal is converted into digital, it is fed into the G.719 encoder to compress the signal (reduce the number of bits), so that the audio information can be efficiently transported over the network. Figure 2 depicts the most important functions in the G.719 encoder.

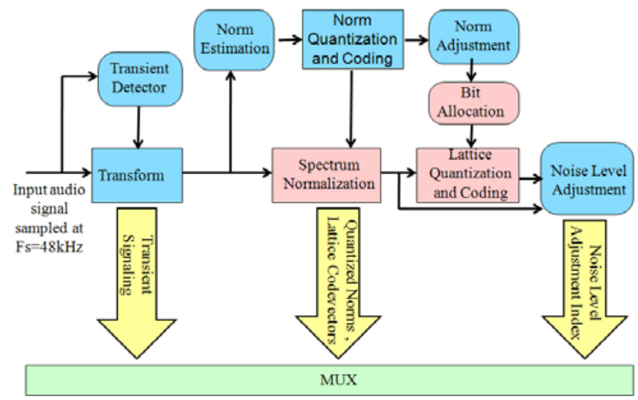

Figure 2: G.719 Encoder

The signal is sampled at 48 kHz because with a Nyquist limit of 24 kHz, this provides plenty of bandwidth headroom. In addition, it fits well with 20 ms frames and is an integral multiple of both 8 kHz and 16 kHz rates which are commonly used in communications. The lower sample rate of 44.1 kHz, which is used in CDs, is considerably more difficult to convert accurately, and so is not well suited to audio communications.

The Transient Detector adapts the frame size, a key parameter of the encoding process, to the character of the incoming sound. This is a critical element of the codec, and is discussed in more detail later in this paper.

The Transform, a Discrete Cosine Transform (DCT), converts the signal from the time domain into the frequency domain. Transforms produce high coding gain (great compression efficiency) for speech because speech concentrates its energy in relatively few frequencies. For the converse reason, there is less coding gain from transforming broadband sounds such as white noise because they are spread across the entire frequency spectrum. Fortunately, coefficient accuracy is less critical for white noise, which helps recover coding efficiency. Music and natural sounds are somewhere in between these two in terms of coding efficiency.

The coefficients generated by the DCT are sent to the Spectrum Normalization function, which divides the spectrum into multiple frequency bands—that is, low, medium, high, and very high frequencies—and finds the average energy level (norm) for each.

The Norm Quantization and Coding algorithm assigns values to each norm. To increase efficiency, the algorithm encodes the difference between the norm in one frequency band and the norm of the next frequency band; this results in smaller values than if the values themselves were encoded, leading to transmitting less information over the network.

The output (encoded norms) is used by the Spectrum Normalization algorithm to normalize the spectrum coefficients coming from the Transform. Energy in all frequency bands usually goes up and down together to a large extent. G.719 takes advantage of this and divides the spectrum coefficient values by the norm, thus reducing the coefficient values (again, reducing the amount of data generated and so boosting coding efficiency).

Bit allocation is an algorithm that allocates more bits to encode larger values and fewer bits for encoding smaller values. This is also done to optimize the use of bits by G.719.

In this context, a Lattice is a grid of numbers, or vectors. The Lattice Quantization and Coding function uses Polycom's Fast Lattice Vector Quantization (FLVQ) algorithm to efficiently encode the data. Spectral data must be normalized before applying FLVQ, to bring the coefficients into the range of the FLVQ lattice. The FLVQ algorithm is a key element in reducing the complexity and memory footprint of G.719, and is discussed in more detail below.

The final function in the G.719 encoder—Noise Level Adjustment—is necessary because we usually run out of bits for encoding, that is to say there are not enough bits for background noise in addition to all the other components of the sound. There are three options in this case: send zeros (dead silence), send an estimated comfort noise, or implement a method to intelligently apply lower frequency coefficients to high frequency ranges. G.719 achieves excellent results by using the latter approach.

G.719 DECODER

The decoder side of a codec takes the incoming bit stream (as received from the communication network), and recreates its best estimate of the original signal from it. Figure 3 describes the G.719 decoder.

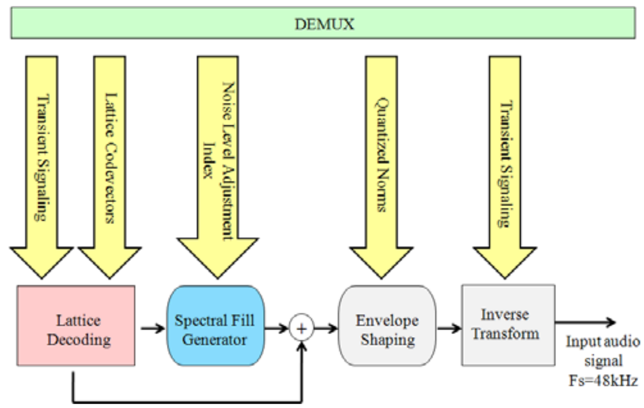

Figure 3: G.719 Decoder

The G.719 decoding algorithm is a logical consequence of the encoder described above: lattice decoding (reverse FLVQ) is followed by spectral fill. However, the coefficients are then fed into something called Envelope Shaping.

Why is envelope shaping necessary? This is because the signal coming out of Lattice Decoding and the Spectral Fill Generator has the same energy in all frequencies. We have to apply the norms, i.e., reverse the spectrum normalization described earlier, to arrive at the correct energy levels in each frequency band. Envelope Shaping is the algorithm that performs this function.

As the final step, the Inverse Transform function converts the signal from the frequency domain back into the time domain. It is, in effect, a reverse DCT.

TRANSIENT DETECTION AND ADAPTIVE TIME-FREQUENCY TRANSFORM

A "transient" is any rapid change in audio signal energy and spectral distribution of energy, for example, percussive sounds, tapping on table, and hand clapping. These sounds are very challenging to previous audio codecs, which have typically either tried to code them efficiently, which results in a "blurring" and muffling of the sound, or have tried to code them accurately, which results in more data output and lower efficiency. G.719 achieves both accuracy and efficiency with its Transient Detection function that identifies such sounds, coupled with

Adaptive Time-Frequency Transform (ATFT) function that instantly switches the time resolution between "transient" and "non transient" (or normal) modes. Figure 4 shows this algorithm.

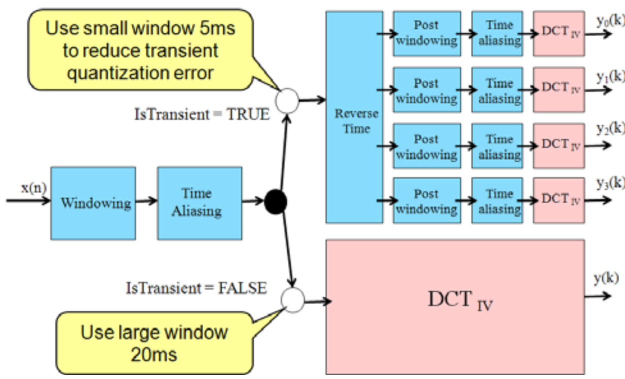

Figure 4: Transient Detection and Adaptive Time-Frequency Transform

The G.719's DCT normally uses a 20 ms window which delivers great quality for normal sounds. However, transient sounds that occur over shorter times lead to quantization errors, which sound like brief puffs of noise. To prevent this, the ATFT algorithm tracks these transients by switching to a much shorter 5 ms window when instructed to do so by the Transient Detector. Using short windows is not without drawbacks: while they reduce noise in the time domain, they cause more smear in the frequency domain. The best compromise, then, is to use the shorter window only for sharp (transient) noises, and longer windows for normal audio.

Post windowing is used in order to divide the 20 ms time aliased frames into 4 smaller (5 ms) frames. Time aliasing after this post-windowing is the same as the time aliasing after windowing, just at a smaller frame size. It is followed by a correspondingly smaller DCT.

Combining the 3 operations (post windowing, time aliasing, and DCT) amounts to a 4 x DCT on the segmented "time aliased signal" before the switch. The four outputs (Y0-Y3) are encoded jointly, and the spectrums are interleaved before the next step (FLVQ).

QUANTIZATION METHODS

In digital signal processing, quantization approximates a large set of possible values by a relatively small set of discrete symbols or integer values. Quantization methods typically attempt to achieve an optimal trade-off between quality and complexity (Figure 5).

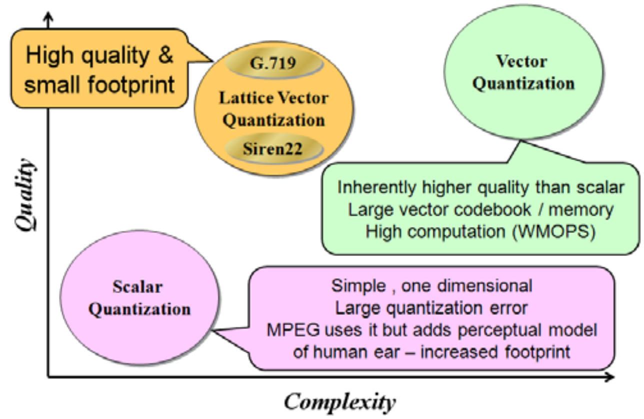

Figure 5: Quantization Methods

The two most common approaches have been Scalar Quantization and Vector Quantization. In Scalar Quantization, the values are represented by a fixed subset of representative values. For example, if you have a 16-bit value and can only send 8 bits, you'll probably send the 8 most significant bits and get an approximation of the original data at the expense of precision. In this example, the fixed subset is only those 16-bit numbers that can be divided by 256, so 255 of every 256 numbers are lost. Because Scalar Quantization is only one dimensional, this lost information leads to large quantization errors. MPEG codecs use this method but try to compensate for its inaccuracy by adding a perceptual model of the human ear, thus substantially increasing complexity. This explains the high complexity values for MP3 and AAC-LD in Table 1.

Instead of representing single individual values, the Vector Quantizer represents arrays of them, i.e. vectors. A vector quantizer uses a codebook; the larger the codebook the more precise the quantization, although at a cost of higher complexity (more details about vector quantization can be found in the book from Allen Gersho and Robert Gray, Vector Quantization and Signal Compression.[2])

When the vector quantizer's codebook is limited to a lattice (grid of values), we talk about Lattice Vector Quantization. This method combines the precision of the vector quantization with the reduced complexity of using a smaller codebook. Polycom's Fast Lattice Vector Quantization (FLVQ) algorithm further improves and optimizes this quantization method, and is an important source of G.719's high performance. FLVQ generates a constraint dictionary of vectors, these vectors being aligned with a lattice. This reduces the complexity (the processing load, as depicted in

Table 2) as well as the size of the codebooks (the memory footprint, as depicted in Table 3).

Note that the FLVQ algorithm is also used in Polycom Siren 22 codec; this explains the similar (and low) complexity of both G.719 and Siren 22, which sets the two codecs apart from the MPEG family (see Table 1).

G.719 COMPLEXITY

Complexity is an important codec parameter because more complex codecs require more powerful and more expensive Digital Signaling Processors (DSPs) to run on; this increases the product cost and power consumption, and can limit the usability of the codec, especially in telephones, low-power mobile devices, and products where cost or power consumption are of concern.

Complexity can be measured in different units. Weighted Million Operations Per Second (WMOPS) is one such unit defined by ITU-T. (Some operations take one machine cycle and others take multiple cycles, depending on the processor used, so the valued have to be "weighted.") ITU-T defines a standard testing environment to measure WMOPS using a fixed-point version of C source code, and Table 2 shows these values for the G.719 standard.

Bit rate (kbps)	Encoder only		Decoder only		Encoder plus Decoder	
	Average	Maximum	Average	Maximum	Average	Maximum
32	6.663	7.996	6.876	7.413	13.539	15.397
48	7.427	9.073	7.230	7.806	14.657	16.861
64	7.912	9.899	7.554	8.161	15.466	18.060
80	8.303	10.564	7.865	8.496	16.169	19.026
96	8.555	10.796	8.122	8.757	16.677	19.536
112	8.787	10.881	8.368	9.009	17.155	19.890
128	8.980	11.775	8.610	9.225	**17.590**	**21.000**

Table 2: G.719 Complexity

What can we learn from this table?

First, the encoder and decoder have similar complexity, which makes G.719 a universal codec that can be used in different applications. Symmetric distribution of complexity between encoder and decoder is beneficial for real-time communication since processing power can be evenly distributed in devices across the network.

Second, G.719's complexity is very low. The highest value is still low, 21 WMOPS for 128 kbps mono (42 WMOPS for stereo). This is an excellent number for a high-bandwidth, high-performance codec, and means that the G.719 codec is well suited to run on an inexpensive DSP.

Third, the incremental performance to process higher bit rates is small. It only requires approximately 2 additional WMOPS to increase the bit rate from 64 kbps to 128 kbps. This means that any possible future codec extensions with higher bit rates would run well on the same DSPs running G.719 today.

G.719 MEMORY FOOTPRINT

In the same way as complexity, the memory footprint of a codec has direct impact on the cost of the product using this codec. A larger memory footprint means higher cost for memory, and codecs with large memory footprints are not appropriate for the majority of communication devices. Table 3 contains the memory requirements listed in the G.719 standard.

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

Table 3: G.719 Memory Usage

Memory size is measured in 16-bit kWords for static, scratch, and data RAM. One 16-bit kWord is 1024 Words each consisting of 16 bits, or a total of 16,384 bits of information.

In real-time communication applications, such as conferencing and telepresence, the endpoint must run both the encoder and the decoder simultaneously; therefore, the last column of the table is important. Here is how to read it: G.719 needs 4.9 kWords of Static RAM and 24.4 kWords of Scratch RAM and 10.7 kWords of Data ROM.

The content in Static RAM cannot be erased, so Static RAM is used to keep the codec variables that will be needed every time the codec runs. This RAM has to be allocated only to G.719, and no other program can use it. Scratch RAM allows deleting its content when the codec instance is terminated, and other programs can use this memory when the codec is not running. Data ROM is a permanent storage, and is used to keep the codebook for FLVQ.

The Program ROM permanently stores the instructions of the G.719 algorithm itself and is therefore measured in 1000s of basic operations, not in kWords.

OFFICIAL TEST RESULTS

ITU-T codecs go through rigorous tests to document the codec quality. Extensive listening tests use speech, music and mixed content, and include frame erasure conditions.

Several G.719 lab tests have been performed independently. One of the challenges in such tests is to find good reference codecs to which the codec under test (CUT, or G.719 in this case) will be compared. As discussed above, while MPEG does not define the encoder (only the decoder), the LAME MP3 encoder is one of the best available MP3 encoders, and was therefore used for the first two tests.

Dynastat Listening Laboratory Report

This test included 24 experts listening to speech (in English and in Spanish), music, and mixed content. G.719 was compared to two reference codecs: LAME MP3 and G.722.1C (the version of Siren 14 standardized by ITU-T). In all nine Requirement Terms of Reference tests, the expert panel unanimously agreed that the CUT (G.719) is significantly better than each of the reference codecs.

Orange /France Telecom Test Report

This test also had 24 experienced listeners but used speech in French with G.719 and LAME MP3. The conclusion was that G.719 generated less audible degradation than the LAME MP3 at comparable bit rates.

Ericsson Lab Test Report

This listening laboratory test was conducted by Ericsson and included 12 expert listeners listening to critical music material, including solo musical instruments. The test used the standard ITU-R BS.1116) methodology[3] for testing high quality codecs, and the results are summarized in Figure 6.

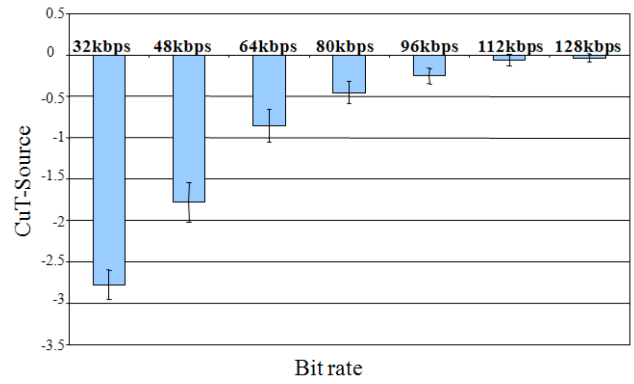

Figure 6: ITU-R BS.1116 Test Results for G.719

The test measures the difference between the score of the codec under test (G.719 in this case) and the score given to the source material, that is, original non-coded signal. The score is represented by a negative since the quality of the coded material is always lower than the quality of the original, especially at low bit rates. A figure close to zero means that the codec is indistinguishable from the original. These results clearly show the benefit of higher bit rates in improving the quality of the codec up to transparency.

As discussed earlier in this paper, bit rates above 64 kbps were not initially intended to become part of the standard, but were added towards the end of the standardization process to make the codec more universally attractive.

More detailed test results will be published in the *IEEE Communications* magazine later this year.

MANHATTAN SCHOOL OF MUSIC – THE FIRST REAL WORLD TEST

The relation between Polycom and the Manhattan School of Music is described in the white paper “Music Performance and Instruction over High-Speed Networks."[4] This paper focuses on the use of the Polycom Siren 22 codec and other acoustic technologies for music instruction and performance at the Manhattan School of Music (MSM) in New York.

The world-renowned violinist and composer and MSM faculty member Pinchas Zukerman brought the concept of incorporating live, two-way audio-visual communication into the Zukerman Performance Program at the School. Within a year of adopting audio-visual technology from Polycom for music education, Zukerman was offering his students regular remote lessons from Albuquerque to New Zealand. Figure 7 shows this application in use.

Figure 7: Manhattan School of Music Using Polycom Audio-Visual Technology

MSM will be the ultimate performance test of the new G.719 technology. We expect this new codec to build on the excellent Siren 22 performance, actually performing better than Siren 22. The increased bit rate will improve the overall quality of music transmission, while the new Adaptive Time-Frequency Transform will bring improved crispness and transparency to the handling of musical transients.

CONCLUSION

Polycom's history is one of collaboration with partners in the communication industry. The successful technology collaboration with Ericsson on the G.719 codec again proves that bringing together the best audio expertise can result in great standards.

Major technical achievements of the G.719 codec are its high quality and low complexity that make it perfect for devices ranging from telephones and low-power mobile devices to soft clients and to high end video and telepresence systems. Indeed, G.719 has the potential to become the audio codec for Unified Communication.

The communications industry is beginning to adopt this new ITU-T standard and leverage its capabilities. We expect the first products with G.719 to appear in 2009/2010, and believe that wide acceptance of this new 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.

ABOUT THE AUTHOR

Stefan Karapetkov is Emerging Technologies Director at Polycom, Inc. where he focuses on visual communications market and technology. He holds an MBA from Santa Clara University (USA) and an MS degree in Engineering from the University of Chemnitz (Germany). He has spent more than 13 years in product management, new technology development, and product definition. Follow his blog at http://videonetworker.blogspot.com/

REFERENCES

1. International Telecommunication Union, 2008. *Series G: Transmission Systems and Media, Digital Systems and Networks. Digital terminal equipments – Coding of analogue signals by pulse code modulation: Recommendation G.719.* ITU-T, June 2008.

2. Gersho, A., Gray, R., 1992. *Vector Quantization and Signal Compression.* Kluwer Academic Publishers, January 1992.

3. International Telecommunication Union, 1997. *Recommendation ITU-R BS.1116 Methods for Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems.* ITU-R, 1997.

4. Karapetkov, S., Orto, C., 2008. *Music Performance and Instruction over High-Speed Networks.* White Paper, November 2008. http://www.polycom.com/common/documents/whitepapers/music_performance_and_instruction_over_highspeed_networks.pdf

ACKNOWLEDGEMENTS

I would like to thank my colleagues Peter Chu and Minjie Xie as well as Anisse Taleb and Manuel Briand from Ericsson for their contributions to this work. Special thanks to Jeff Rodman for his support.