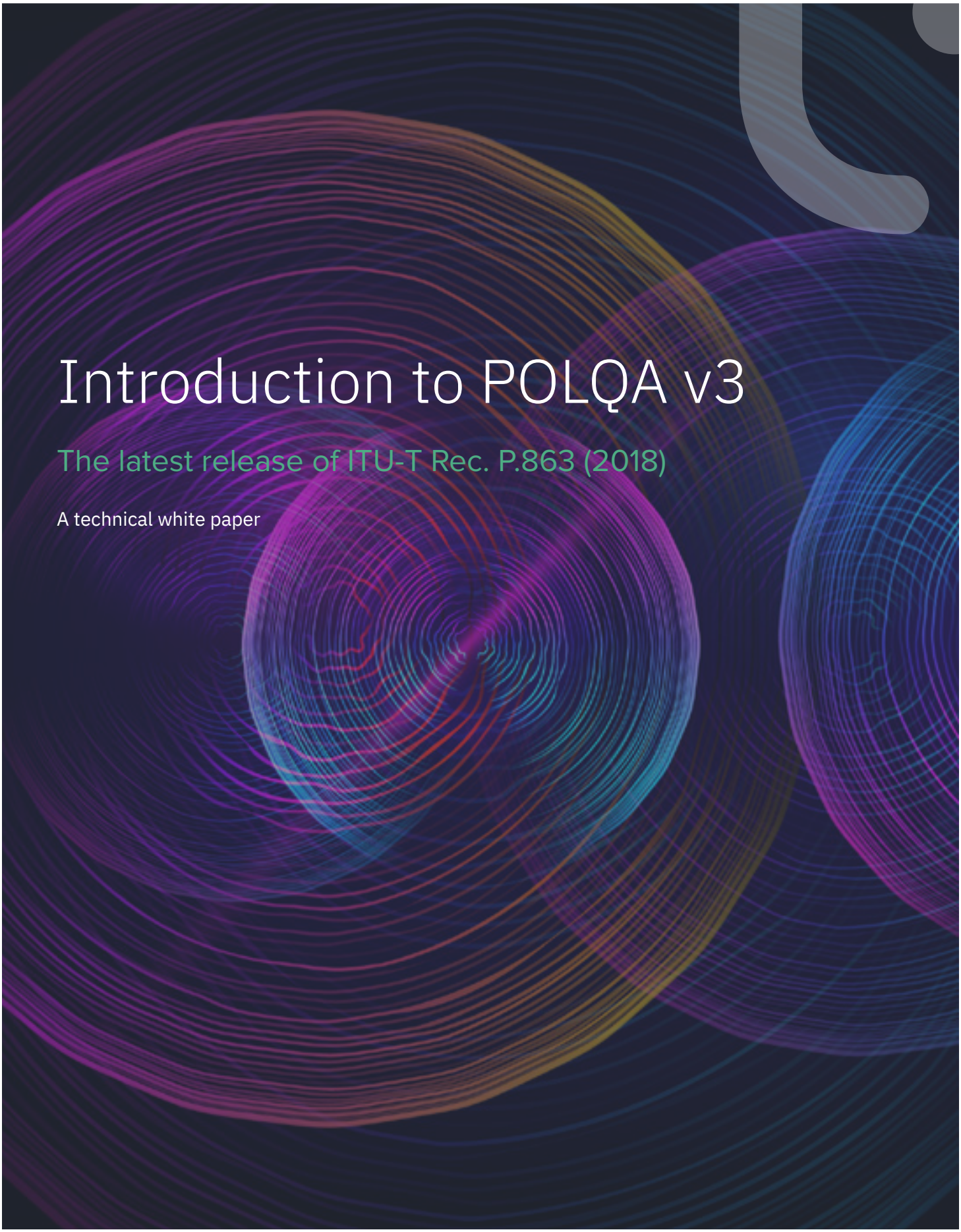
infovista

KNOW YOUR NETWORK™



Introduction to POLQA v3

The latest release of ITU-T Rec. P.863 (2018)

A technical white paper

Overview

POLQA is an objective perceptual QoE metric that predicts the overall listening speech quality of narrowband (NB) to super-wideband (SWB) telecommunication scenario as perceived by users. It was initially standardized as ITU-T P.863 in 2011 and went through a couple of revisions, the latest one POLQA v2.4 in 2014. POLQA has become widely adopted as the preferred speech quality test method for predicting voice MOS in mobile networks.

TEMS has included support for POLQA in all its tools since 2011 and has sold over 20,000 POLQA v2.4 licenses to 530 companies in 130 Countries. We are proud to be one of the biggest suppliers of Audio Quality measurement equipment for Network Testing in the world.

Reasons for introducing POLQA v3

POLQA v2.4 is known to have a few well-documented shortcomings (ITU-T P.863.1) under some circumstances. These will become more prevalent with the advanced LTE (VoLTE) and 5G (VoNR) voice services, as well as the use of OTT voice applications.

The introduction of full-band (FB) voice service systems, of new codecs (modes, bit rates), as well as clients with smarter error concealment schemes (time scaling) and signal processing to render delay variations inaudible called for a significant update of ITU-T P.863.

Therefore, in March 2018, ITU-T SG 12 approved POLQA v3 after testing demonstrated that this update handles all these new conditions with high accuracy and without altering the scoring produced in traditional use cases.

Background

ITU-T P.863 Edition 3 (POLQA v3) addresses three categories of updates and improvements to the previous version:

1. **Introduction of the full band (FB) mode:**

   POLQA v2.4 SWB mode limits the upper audio frequency response to 14 kHz. POLQA v3 expands this to full-band (FB) mode, which is unlimited and comprises the full audio bandwidth of up to 24 kHz (at 48 kHz sampling frequency). This means that POLQA v3 is now sensitive to distortions across the entire audio spectrum when a FB reference speech sample is used.

   POLQA v3 comes with a single mode, FB mode, which covers all speech bandwidths (NB, WB, SWB, FB). Consequently, bandwidth and network degradations are reflected on a unique MOS scale, relative to the best FB voice quality.

2. **New codecs:**

   First, the new FB scale enables the assessment of codecs that are designed to operate in FB mode, such as EVS, OPUS, Advanced Audio Codec (AAC), or Low Complexity LC3 (Bluetooth), since distortions above the former frequency limit of 14 kHz are now correctly considered. Second, POLQA v3 has implemented a more appropriate scoring of the EVC codec (incl. inter-operability (IO) and channel aware (CA) modes and bit rates) as compared to the open-source audio codec OPUS, which is used in many OTT applications.

3. **Previous version's shortcomings and their fixes**

   Infovista has worked in ITU-T Study Group 12 along with the ITU experts to help identify and reduce shortcomings of POLQA v2.4. The addressed topics are as follows.

   - Artificial variability of scores with even small front-end delay ("leading silence" of the degraded signal) variations has been somewhat reduced by increasing the overlap of FFT (Fast Fourier Transformation) window to 75%. The best possible performance for this shortcoming can be achieved by running the algorithm with the additional High Accuracy (HA) mode, at the expense

of longer processing time. Infovista tests showed that the improvements are not statistically significant and remain dependent on the delay shift value. Therefore, since using HA mode affects the real-time scoring, Infovista's implementation is not using HA mode, but rather controls the synchronization of POLQA measurements to minimize any front- end delay variation.

- POLQA v3 comes with perceptual model modifications, which helps the algorithm to react less sensitive to linear frequency distortions; thus, the measurements are less dependent on the frequency characteristic of the devices, consequently more accurate. The model's modifications also enabled POLQA v3 to reduce previous artificial sensitivity to high speech frequencies and/or speakers (bias). Also, the POLQA v3 scores transparency described by the maximum score when a speech file is played against itself (max. MOS score) has been improved. However, as recommended by ITU-T, the Infovista implementation uses one single speech reference file (male-female combo) to be able to optimize all these improvements; sensitivity to devices' frequency response, too high speech frequencies, and speaker bias, maximum scores transparency.
- Advanced VoLTE, forthcoming VoNR voice services, as well as today's OTT voice applications are coming with smarter and highly dynamic delay-jitter compensation in the clients. These compensations result in very short, variable pauses during speech. Although barely audible to listeners, these 'micropauses' were detected by POLQA v2.4 as signal modifications and it reacted too harshly, consequently inaccurately. POLQA v3 detects these variations during the temporal alignment phase and is taking them into account within the perceptual model. Fixing this specific shortcoming extends the scope of POLQA to the large variety of forthcoming voice services.

The expectation from POLQA v3 field measurements

POLQA v3 fixes the previous version's shortcomings and extends its applicability to today's and forthcoming voice services. To achieve this, the POLQA v3 perceptual model as well as its calibration to the MOS scale (introduction of the FB scale) has been modified. Therefore, although a very important criterion in developing POLQA v3 was to maintain backward compatibility with POLQA v2.4, a perfect score match (for same reference speech files and same conditions) between the two versions is not expected.

Clean encoding-decoding conditions, each calculated over a large number of reference files, showed statistically significant-close results of the two versions. At the same time, it has been determined that better discrimination at the upper end of the scale is achieved and it shows up as a slight increase of the difference between EVS-FB 24.4 kbit/s and Opus-FB 24.4 kbit/s.

However, when running POLQA in drive test tools, individual real-time scores for the same reference speech file and identical network conditions will be different for the two versions. Preliminary tests showed that individual POLQA v3 scores were slightly higher than POLQA v2.4 scores, in the good-quality/high-quality range, and slightly lower in the poor-quality range. It should be noted that POLQA v3 scores are more accurate than POLQA v2.4 scores. Hence, the average values over a large number of conditions of individual scores coming from the two versions are expected to be statistically significant close.

Infovista preliminary tests show that variations of the front-end delay can still show some small POLQA v3 score variability depending on the value of the delay shift. Therefore, it should be expected that, similar with POLQA v2.4, real live POLQA v3 scoring can still show some random occurrence of 'unexpected too low scores' and/or the random occurrence of score variability for the same field condition, mostly noticeable for good-quality/ high-quality scores.

Alternatives

Recently, Infovista has successfully implemented a new methodology (sQLEAR: Speech Quality by machine Learning) to predict voice QoE and it is ideally suited for monitoring, benchmarking, and troubleshooting VoLTE and OTT voice applications (e.g. WhatsApp) so far. The solution is available as an option in most new TEMS tools. We can now offer this as an alternative where the perceptual full-reference speech quality testing methods like POLQA may be uneconomic (i.e. when scaling to measure the audio quality of 1000s of users).

The idea behind our approach is to make measurements independent from the terminal behavior, thus restricting the effects they have on the resultant score and providing the network-centric view on the speech quality. Essentially, sQLEAR predicts a perceptual voice QoE score based on non-perceptual characteristics, respectively network, codec, and client parameters and information on DTX (discontinuous transmission) periods (silence) within the reference speech sample. It uses machine learning techniques trained and validated on POLQA v3 as reference values. sQLEAR has proved to be very successful and shows very similar results to POLQA v3.

sQLEAR has been developed in conformity to ITU-T P.565 framework and is currently under the ITU-T validation process.

You can find out more details about the sQLEAR solution here and details on its scope, applications, and how it positions itself against other voice QoE metrics here.

Conclusion

POLQA v3 marks a major improvement over the previous version and significantly extends its scope and applicability towards 5G voice services and OTT voice applications while maintaining maximum backward compatibility. In our opinion, with the introduction of this new version, POLQA continues to be the gold standard for predicting the user-centric view on voice quality in mobile networks and provides a solid foundation for the approaching 5G era. Infovista sQLEAR is tuned to match the accuracy of POLQA v3, yet designed to provide a network-centric view on voice quality in all IP mobile networks (VoLTE, VoNR, OTT voice applications).

So Infovista recommends the use of POLQA v3 for user-centric voice quality evaluation and benchmarking and sQLEAR for network-centric voice quality monitoring and in-depth troubleshooting and optimization.

Upgrading to POLQA v3 is therefore highly recommended to provide accurate measurement of newer mobile networks.

The POLQA Coalition has recently announced that they will stop supporting v2.4 shortly.

KNOW YOUR NETWORK™

About Infovista

Infovista, the leader in modern network performance, provides complete visibility and unprecedented control to deliver brilliant experiences and maximum value with your network and applications. At the core of our approach are data and analytics, to give you real-time insights and make critical business decisions. Infovista offers a comprehensive line of solutions from radio network to enterprise to device throughout the lifecycle of your network. No other provider has this completeness of vision. Network operators worldwide depend on Infovista to deliver on the potential of their networks and applications to exceed user expectations every day. Know your network with Infovista.