Introduction to POLQA v3, implementing the
3rd edition of ITU-T Rec. P.863 (2018)

- Technical White Paper -

Contents

Overview

POLQA, the third-generation perceptual speech quality test method standardized as P.863 in 2011, has been widely adopted as the state-of-the-art MOS benchmarking technology for mobile networks. Since its first release in 2011 and a major update in 2014, network technology and codecs have developed significantly. The imminent introduction of 5G will especially lead to numerous new challenges that also drove an update of POLQA.

In strong collaboration with ITU-T Study Group 12, the parties of the POLQA Coalition (OPTICOM GmbH, Rohde & Schwarz SwissQual AG and TNO) proposed POLQA|3 as an evolved version to ITU-T Study Group 12, which was approved as Rec. P.863 Edition 3 in March 2018.

The new POLQA|3 supersedes all earlier versions. Main changes are summarized in table 1 and outlined in the following.

POLQA	ITU-T P.863	Audio-Bandwidth	Main Changes
1.1	Ed. 1	4 kHz (NB) + 14 kHz (SWB)	Initial release
2.4	Ed. 2	4 kHz (NB) + 14 kHz (SWB)	Update for increased accuracy,
3	Ed. 3	4 kHz (NB) + 24 kHz (FB)	Redesign for Full-band audio, validated for VoLTE, EVS and OTT, ready for 5G telephony

Table 1, POLQA generations

This white paper outlines the objectives for the update, compares these to the achieved improvements, and highlights the most important changes for users. It will be demonstrated that this update marks a significant step towards even higher measurement accuracy and a broader range of applications for POLQA|3, while maximum backward compatibility of the measured scores is maintained.

Objectives for POLQA|3

POLQA 2.4 is known to be highly accurate and extremely reliable over a vast range of telephony applications. Amazingly, it not only performed nicely for conditions outside of its scope, but also for newly developed codecs like e.g. EVS. However, there were a few, well documented shortcomings under some circumstances which had to be taken into account. Usually, these were not observed in the typical mainstream applications, but they will become more prevalent with the advent of 5G networks and therefore updating POLQA to its third generation seemed necessary and appropriate.

The challenges 5G brings with regard to speech quality testing are the expected significant increase in OTT and LTE (or in this case better named “5G voice telephony”) calls which will include a variety of new codecs, wider audio bandwidth, smarter time warping and signal processing to render these delay variations inaudible. While the performance of POLQA 2.4 was quite good under these circumstances, it was clearly less accurate than for those conditions it was actually designed for. Consequently, a major design goal for POLQA|3 was to bring its accuracy up to the same extremely high level as for more traditional use cases. Another, equally important design goal was to keep the scoring of already correctly scored test conditions virtually unchanged, so as to make the transition to POLQA|3 as seamless as possible, particularly in the sense of backward compatibility with test and benchmark results, achieved with earlier versions.

Changes Compared to POLQA 2.4

One important factor to reach these objectives was to extend POLQA's *super-wideband* mode (SWB), currently with an upper audio frequency response limited to 14 kHz, into a *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 is now sensitive to distortions across the entire audio spectrum. This extension makes it fit for the assessment of codecs which are designed to operate in full-band mode as well, e.g. EVS, OPUS, AAC or LC3 since distortions above the former frequency limit of 14 kHz are now correctly taken into account.

Apart from extending the application space and scope, this also leads to a more appropriate scoring of the EVS codec as compared to the open source audio codec Opus (which is used in many OTT applications).

Additionally, the discriminative power of POLQA at the upper, high quality range of the MOS scale could also be increased by this extension.

All this comes without sacrificing backward compatibility with POLQA 2.4, since 14 kHz bandwidth limited reference signals can still be used with the new POLQA full-band mode.

With today's OTT voice services, VoIP transmissions and also VoLTE / "5G voice telephony", it was observed that smart and highly dynamic delay-jitter compensation leads to variations of the duration of very short pauses during speech. Such variations of 'micropauses' are barely audible to listeners, yet POLQA 2.4 detected these signal modifications and reacted too harsh. Fixing this in POLQA|3 was very important for its application in future networks and could be fully achieved by detecting these variations in the temporal alignment of POLQA and taking them into account within the perceptual model.

In addition to the above modifications, which significantly extend the scope of POLQA, it was also possible to improve a few other, minor nuisances of POLQA 2.4:

- The variation of POLQA scores caused by small differences in the leading silence of the degraded signal could be reduced to non-significance by increasing the overlap of FFT windows to 75%.
- POLQA|3 reacts much less sensitive to linear frequency distortions than POLQA 2.4 did. This makes measurements less dependent on the frequency characteristics of headsets.
- POLQA 2.4 included the calculation of an indicator for 'reverb', which influenced the MOS score. In fact, taken in combination with the increased accuracy of the perceptual model of POLQA|3, it was found to be no longer necessary.
- Under the hood, the perceptual model was significantly improved and streamlined. Many cases which required special treatment in previous POLQA versions could be consolidated and the whole model is far more accurate and consistent.

Backward Compatibility

How well the design goal of maintaining backward compatibility matches with valid POLQA 2.4 use cases can be seen in Figure 1, where both POLQA versions were used to process a very large set of reference files encoded with various different standard codecs. It can be seen that the average MOS for the vast majority of conditions remains virtually the same. It is also visible from this chart that some other design goals were met, too, namely the better discrimination at the upper end of the scale (conditions 'REF' and 'SWB') as well as slightly increasing the difference between EVS-FB 24.4 kbit/s and Opus-FB 24.4 kbit/s. It has to be noted though that none of the test conditions used to create this chart exhibited any significant degradations in the frequency range above 16 kHz.

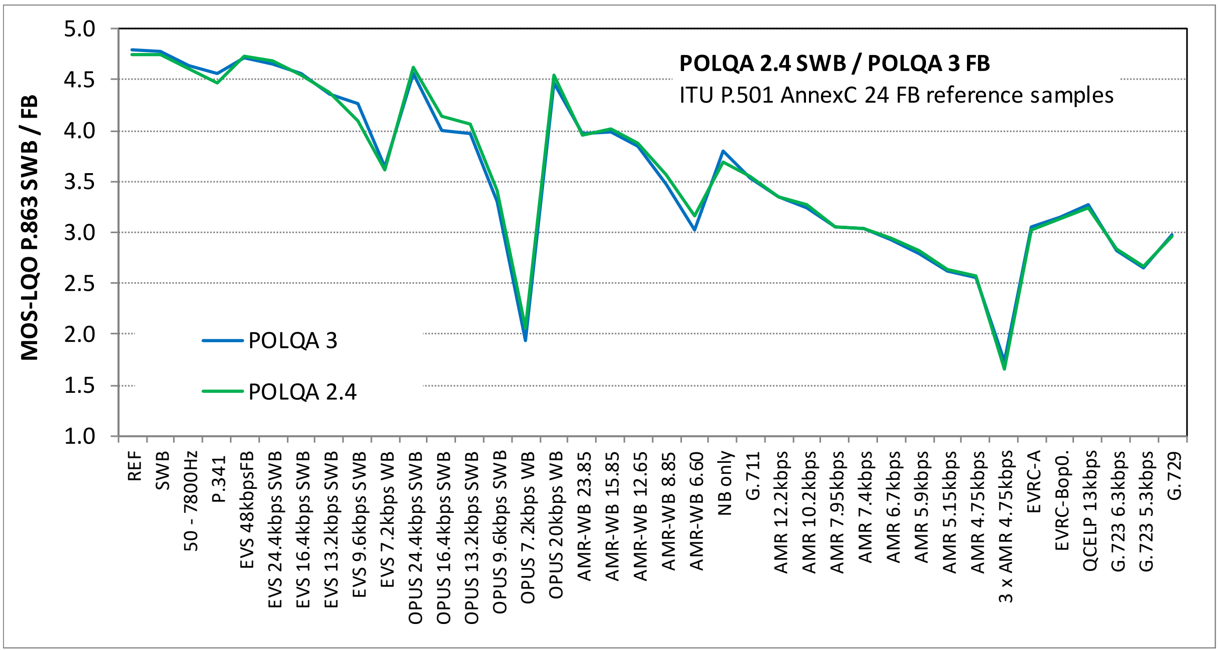

Figure 1: Prediction of common codec conditions by POLQA 2.4 compared to POLQA 3 in full-band/super-wideband mode

Real Field Measurements

Figure 2 presents the results of a field validation of POLQA|3. For this validation, real-field measurements (71% of the conditions in the test) combined with some simulated conditions were measured with POLQA|3 and the results were compared to results from subjective tests.

The live, real-field conditions include state-of-the-art measurements of conditions such as

- VoLTE calls with EVS at 24.4 kbit/s SWB
- WhatsApp calls in LTE with Opus at 20 kbit/s WB
- 3G mobile to mobile calls with AMR-WB at 23.85 kbit/s and AMR-WB at 12.65 kbit/s 3G mobile to mobile calls with AMR-NB at 12.2 kbit/s
- 3G/2G mobile-to-mobile calls with transcoding from AMR-WB at 12.65 kbit/s to AMR-NB at 12.2 kbit/s.

Good, average and poor coverage network conditions are represented in the test.

The offline (simulated) coded conditions in the test aim to reproduce some additional common conditions found in the field (EVS 24.4 kbit/s SWB, EVS 13.2 kbit/s SWB, Opus 20 kbit/s WB, AMR-WB 23.85 kbit/s, AMR-WB 12.65 kbit/s and AMR (NB) 12.2 kbit/s).

The three full-band conditions in the test are a full-band reference and two anchors with packet loss. Four additional low-quality simulated conditions were obtained by either adding packet loss to a codec condition or by re-encoding the reference sample multiple times with the same settings.

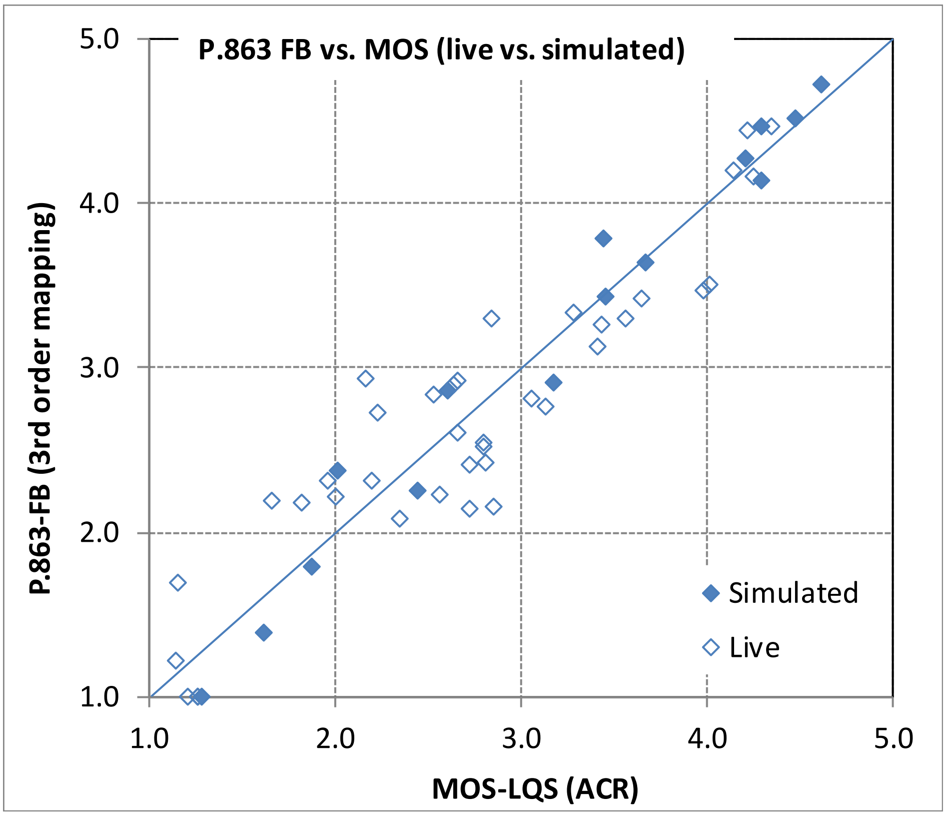

Figure 2: ACR LOT MOS scores and (mapped) POLQA 3 results for all conditions in the test

Figure 1 shows all results of the subjective listening test as well as the P.863 FB predictions on a 'per-condition' basis (i.e. after mapping to the subjective scale), which means that the MOS values are averages of four individual recordings (one female speaker, one male speaker, two samples each) for the same test condition. It can be seen that all datapoints are very close to the ideal diagonal line. This holds for simulated as well as live conditions and underlines the superb performance of POLQA|3 in real-life use cases and field tests.

POLQA Derivatives by Machine Learning

Recently, there have been reported attempts for predicting voice QoE scores of networks based on machine learning, as an alternative to full-reference speech quality testing methods like POLQA. The idea behind this simplified approach is to make measurements independent from the terminal behavior, thus restricting the effects taken into account to artefacts resulting only from the network. Essentially, this is an attempt to predict a perceptual QoE score without analyzing the actual audio signals, based on entirely non-perceptual parameters.

Although POLQA scores of proven accuracy can be used to train machine learning algorithms, the results of such metrics will always remain a limited and weak simulation of POLQA. The concept is likely to completely fail for any condition it was not trained on. Consequently, using such derivative methods will bear a danger of optimizing and assessing networks without even noticing it was based on inaccurate and unreliable data. POLQA is in this respect far superior, since it assesses the true speech signals, as they are presented to the user, and will take into account whatever artefacts truly appear, analyzed by its highly accurate perceptual model.

Conclusions

POLQA as defined in ITU-T Rec. P.863 and provided under license of the POLQA coalition is applying a standardized algorithm that analyses the degraded wave file by means of one of the most advanced psycho-acoustic models of the human inner-ear and cognitive system for human voice perception. Because POLQA accurately models hearing and perception, it can be applied to quantify the perceptible distortions of any speech wave signals in a telecommunication context. Therefore, it can also be used with any speech stimulus of any language. POLQA neither needs knowledge about the codec type used, nor is it restricted to particular ones. The algorithm doesn't need input of any bitstream information, nor any other side information. It can assess the signal despite any transcoding, noise-floors or re-packaging. POLQA's accuracy is second to none and has been comprehensively validated in the course of International Standardization.

POLQA|3 marks a major improvement over previous POLQA versions and significantly extends its scope and applicability towards 5G and OTT codecs, while maintaining maximum backward compatibility. Upgrading to POLQA|3 is therefore highly recommended at any time soon to ensure compliance with the 3rd edition of ITU-T Rec. P.863 (2018).

The recent improvements emphasize POLQA's position as the unsurpassed gold standard for worldwide mobile network benchmarks and provide a solid foundation for the approaching 5G era.

Published by:

OPTICOM GmbH

Stintzingstrasse 24
D - 91052 Erlangen
GERMANY

Phone: +49 91 31 - 5 30 20 - 0
Fax: +49 91 31 - 5 30 20 - 20
info@opticom.de
www.opticom.de
www.polqa.info

VAT ID No. DE 194 631 268
Managing Directors:
Dipl.-Ing. Michael Keyhl, CEO
Dipl.-Ing. Christian Schmidmer, CTO
Register: AG Fürth (Bay.) HRB 7169

Rohde & Schwarz SwissQual AG

Allmendweg 8
CH-4528 Zuchwil
Switzerland

Phone: +41 32 686 65 65
Fax: +41 32 686 65 66

VAT ID no.: CHE-113.795.355
Director: Hanspeter Bobst, CEO
Registration no.: CHE-113.795.355

For inquiries on POLQA licensing please contact OPTICOM GmbH, or visit

www.polqa.info

for further details.

For an updated reference list of POLQA licensees, available POLQA products and solutions please refer to:

www.polqa.info/licensees.html