infovista

KNOW YOUR NETWORK™

A New Approach for Testing Voice Quality

sQLEAR Q&A

Content

Why sQLEAR?

To meet the needs of today's evolving mobile networks, there is a growing need for flexible, real-time, automated QoE-centric service evaluation, troubleshooting, and optimization. This has been driven by a number of factors. First, the volume of 4G subscribers has grown dramatically. Second, the range of 4G services has also increased. Finally, 5G network rollout is underway, bringing significantly increased network complexity, an even greater number and a larger variety of devices as well as more service diversity.

It is well known that mobile video consumption has exploded. Today, video services represent about 70% of mobile data usage - and are expected to grow further with 5G. Despite the outstanding success of mobile video services, it's often overlooked that voice services still deliver approximately 70% of Mobile Network Operator (MNO) revenue. However, this valuable revenue is challenged by two main intertwined trends, the sustained OTT voice services' expansion and the fast increase of VoLTE subscriptions number along with VoLTE support on new type of devices such as smartwatches, Cat-M1-capable Internet of Things (IoT) chipsets. In addition, VoLTE technology will enable the 5G voice solution, while representing the base for interoperable consumer and enterprise communication services on different devices across LTE, Wi-Fi and 5G.

The continued strong competition from OTT voice applications, is predicted to grow exponentially with 5G deployments, presenting a continuously persistent threat to MNOs' voice service revenue. The powerful MNOs' counter to this threat is the expansion of the carriers VoLTE services.   The GSA report, April 2018, states that VoLTE has now been launched in more than 145 networks in over 70 countries across all regions. The findings presented in the Ericsson Mobility Report, June 2018, show that at the end of 2017, VoLTE subscriptions exceeded 610 million. The Ericsson's findings also project that the number of VoLTE subscriptions will reach 5.4 billion by the end of 2023, accounting for around 80 percent of combined LTE and 5G subscriptions. **Figure 1** shows VoLTE subscriptions prediction per region.

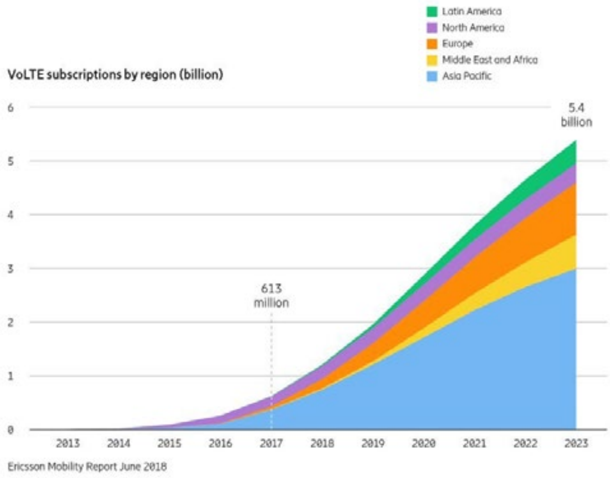

Figure 1. Prediction on VoLTE subscriptions expansion

As a result, MNOs must not only support the expected growth in video traffic, but they must also protect their existing voice revenues and eventually increase these through VoLTE services expansion to consumer and enterprise communications supported on a variety of new devices. The performance and quality of voice services as experienced by users is a significant factor in ensuring that they continue to use MNO voice services rather than OTT alternatives.

Maintaining the voice service Quality of Experience (QoE), reducing subscriber churn to OTT providers, and growing voice revenue through VoLTE expansion, while optimizing CAPEX/OPEX are therefore key concerns for MNOs. Obtaining accurate, easy to implement and controlled voice QoE predictors, as well as securing the ability to act on these in real-time, are thus crucial for enabling cost efficient, optimized network operations that will meet and maintain customer expectations and demands.

Infovista's sQLEAR voice QoE predictor is a new and unique solution, which is specifically designed to answer these concerns and goals, benefiting thus MNOs and regulators alike. It provides cost-effective and accurate evaluation of voice quality trends and enables predictions of future QoE delivery. In addition, it can be used to perform monitoring, benchmarking, and troubleshooting of MNO voice services. sQLEAR ensures operational efficiency for MNOs through effective troubleshooting of IP networking and the underlying transport layers.

What is sQLEAR?

ITU-T standardized and currently available voice QoE evaluation algorithms can be perceptual and parametric. Perceptual algorithms are based on human perception and cognition models, which are either intrusive or non-intrusive. Perceptual intrusive algorithms (ITU-T P.863) use test (or reference) speech samples sent over the network as well as the resulting degraded speech sample in order to provide a QoE score, which represents the estimator of the subjectively perceived voice quality as defined by MOS (Mean Opinion Score, ITU-T P.800, P.800.1, P.800.2). Perceptual non-intrusive algorithms (ITU-T P.563) do not require a reference speech sample, but they rather use only the resulting degraded speech sample in order to estimate MOS. The parametric voice QoE algorithms (ITU-T P.564) use only network parameters to estimate MOS, and therefore these solutions are non-intrusive.

In addition to the different models and approaches used by these algorithms, it is important to mention that only ITU-T P.863 can be used for evaluating VoLTE scenarios; neither P.563, nor P.564 has been designed for VoLTE.

sQLEAR has been coined as an abbreviation of Speech Quality (by machine) LEARning. It is an algorithm which predicts the impact on the perceived voice quality that results from IP transport and underlying transport (packet-based radio and core network parameters), as well as the codec and jitter buffer in the end-user voice client (with consideration of codec/client parameters). In addition, sQLEAR algorithm is machine learning based and it uses as input speech reference sample(s), but not the resulting degraded speech samples. Therefore, sQLEAR uniquely becomes the first parametric intrusive voice QoE evaluation algorithm. The parametric intrusive approach innovatively ensures sQLEAR with superior accuracy when compared to existing parametric non-intrusive solutions (ITU-T P.564) and/or perceptual non-intrusive (ITU-T P.563) as well as with competitive performance against intrusive perceptual solutions (ITU-T P.863).

sQLEAR does not use the degraded speech sample, and therefore does not use the audio path for QoE prediction. This confers the advantage of avoiding device specific degradations and characteristics, such as background noise, automatic gain control, voice enhancement techniques, and frequency response. Therefore, sQLEAR enables operators to cost efficiently focus on network rather than on device specific performance.

sQLEAR output is represented in terms of MOS (as defined by ITU-T P.800.1, MOS-LQO) and it represents the first outcome from ongoing activities in the P.VSQMTF (Voice Service Quality Monitoring and Troubleshooting Framework) work item from ITU-T Study Group 12.

Designed for current and future voice services, sQLEAR can be used for the evaluation of VoLTE services that employ the High Definition (HD) Enhanced Voice Service (EVS) codec and client, including the channel aware (CA) and Inter- Operability (IO) modes. The IO mode ensures backwards compatibility with AMR codec.

What techniques does sQLEAR use?

sQLEAR is based on several key factors. These include the transmitted speech reference; transport protocol information, such as jitter and packet loss; and codec information, which includes rate and channel-aware mode (EVS codec case). The prediction algorithm uses deep packet inspection (DPI) to obtain relevant information. This means that the impact of the network on voice QoE can be determined without the necessity of recording actual speech content. The time characteristics of the reference signal are used to identify the importance of individual sections of the bitstream in regard to speech quality. This offers the advantage of being able to take into account the real voice signal after the jitter buffer.

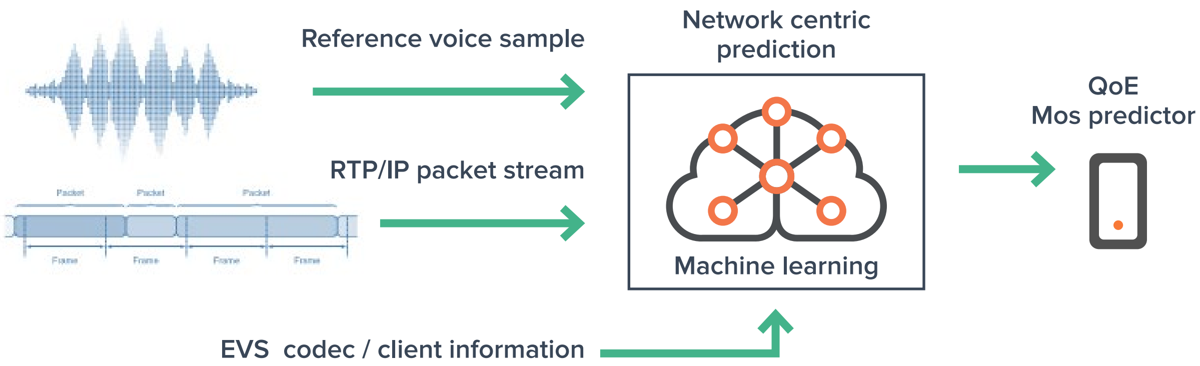

Figure 2. sQLEAR concept

sQLEAR uses state-of-the-art machine learning to build a model that describes the speech quality perceived by users based on all of these information resources (**Figure 2**)

The unique machine learning techniques offer three clear advantages. First, the complexity of the inter-dependencies between all network/codec/client parameters, as well as their significance in impacting the speech quality, is better described and processed by machine learning algorithms than by the multi-dimensional optimization techniques required for the estimation of new coefficients of multi-variable non-linear functions, which are generally used for parametric voice QoE evaluation algorithms. Second, any time changes that emerge from the introduction of new codecs/clients need to be accounted for, machine learning techniques are more flexible and quicker to tune. This provides a significant advantage from the perspective of implementing the algorithm and ensuring operational efficiency. Third, there is no need for additional calibration to the MOS scale using first or third order polynomials, because the machine learning based algorithm "learns" the precise MOS scale that it needs to predict.

What is the sQLEAR learning and evaluation process?

sQLEAR has been developed using a simulator that is based on 3GPP standardized code for both the EVS VoIP client and the EVS voice codec. The simulations are performed at the IP level, and are based on knowledge obtained from the analysis of real-life field data, collected during a significant number - and broad diversity - of drive tests in different locations, conditions, and in a number of MNO networks. The standardized EVS VoIP client ensures that all devices with embedded EVS exhibit the same behavior. As a result, sQLEAR is completely transparent to and independent of the devices used in testing.

Figure 3 depicts the simulation chain. A reference audio file is injected into the simulation process, and coded with different EVS codec settings for bandwidth, codec rate and channel-aware mode. The resulting VoIP file output includes audio packets coded together, with an ideal arrival time increasing by 20ms for each packet. Network errors, in the form of jitter and packet loss patterns, are applied to the coded audio to simulate degradations that may occur in an all-IP network.

To simulate the jitter and packet loss behavior of a radio and core network, jitter files are created by using a combination of simulations and drive test data. A large set of databases, spanning approximately 120,000 samples and covering a broad range of conditions that generate voice degradations for the entire voice quality range, have been generated. These conditions include:

- "Live (drive test) data modulated with simulations", to broaden the range of conditions (e.g. randomizing the live degradation's position and amplitude);
- "Gilbert burst packet loss and burst jitter up to 30%", to mimic error cases that are observed in live drive test data. For example, during handover, during which packets are buffered and then suddenly released;

- “Gilbert severe burst jitter to 70%”, to improve the learning and testing of large jitter cases, which are the most difficult to predict;
- “Random packet loss and random jitter”, to handle reordering of packets; and
- “Manually designed packet loss”, to simulate mobile devices that move in and out of coverage, which results in long and short consecutive packet loss.

By applying the jitter files and simulating network degradations, EVS frames are removed when there is packet loss and the arrival time of the frames are changed relative to the jitter file. The new Jittered VoIP file is then submitted to the EVS jitter buffer. It is decoded and time scaled, which produces a degraded audio file. Finally, the degraded audio file is graded using ITU-T P.863 and compared it to the original reference, resulting in a MOS score.

As a result, each simulated jitter file that describes a network condition has a corresponding degraded audio file and an associated MOS score. The 120,000 samples represent the databases used for sQLEAR learning and evaluation, with a 50%-50% split, as recommended by current academic research in machine learning (see “Handbook of Statistical Analysis and Data Mining Applications”, by Elsevier Publisher, 2009)

sQLEAR uses a combination of bagged decision trees and SVM, (support vector machine) machine learning algorithm categories. These proved to provide the best performance (for correlation and prediction error) when compared to ITU-T P.863, a point addressed under the question of accuracy later in this paper.

Which parameters does sQLEAR use?

As a machine learning-based technique, the inputs that sQLEAR uses are defined by features which are aggregated from basic network parameters. sQLEAR uses machine learning both for the creation and selection of features, as well as for the QoE prediction performed which, in turn, is based on the selected features.

There are two sources of features. First, the information derived from the RTP stream generated by the simulated jitter buffer implementation. Second, statistical measures built from the RTP stream, which proved through extensive testing to have a significant impact on the accuracy of the algorithm in comparison to ITU-T P.863.

There are a number of factors that play an important role in the MOS estimation process. These include speech content and frequency; and the duration of silence, as well as its distribution within voice samples. These have significant impact on the performance of sQLEAR and, as a result, the accuracy of QoE prediction. Therefore, to improve sQLEAR performance further, audio reference-based features are also used.

These features are weighted based on the location of the feature (“position-based feature”) in the reference voice sample. The position is described either by a feature giving the position of, for example, a dip in packet loss or by weighting the number of frame erasures. The most successful weighting function proved to be the rms (root mean square) of each 20ms voice frame in the reference voice. It should be noted that information on the reference voice is used, and not the recording. In addition,

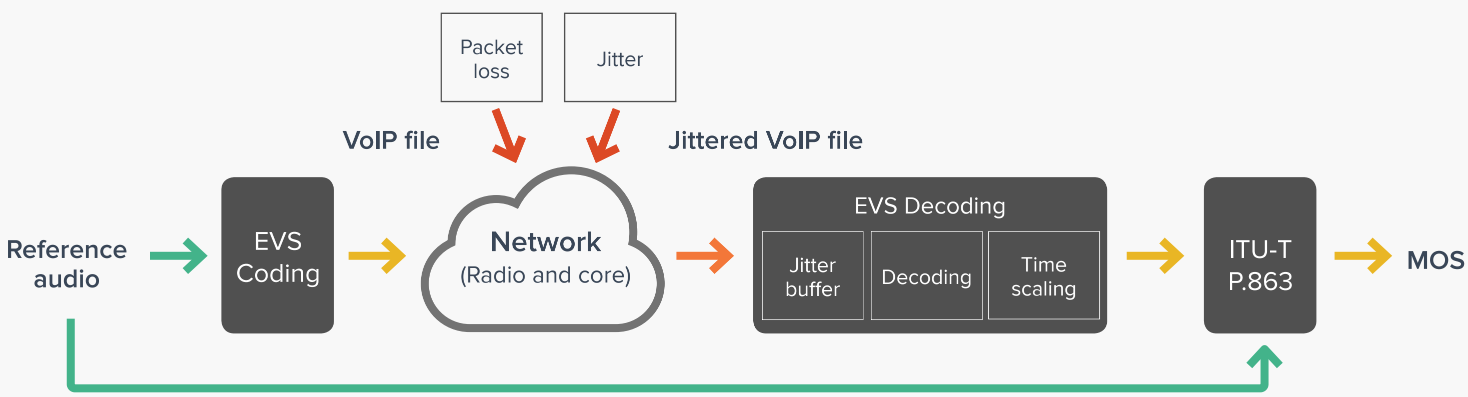

Figure 3. Simulation chain

the weighting process does not involve any perceptual processing of the voice sample. Consequently, sQLEAR does not depend on the audio path.

In order to ensure that the jitter files are independent of codec and reference voice sample, and to simplify feature creation, sQLEAR performs two pre-processing operations: DTX cleaning and the addition of codec information (e.g. rate, mode). The output of the pre-processing is a new, “DTX cleaned”, jitter file, which also contains codec audio payload size and Channel Aware mode data (see **Figure 4**).

The DTX cleaning pre-processing handles the fact that DTX periods, which occur during silence, do not impact the perceived voice quality. Consequently, the pre-processing operation creates a new jitter file with no packet loss and no jitter during DTX periods, which greatly simplifies feature creation.

Add–on codec information represents the second pre-processing step. The codec information is added to the “DTX cleaned” jitter file. This consists of the audio payload size and Channel Aware mode indication. Packet payload size indirectly provides both codec rate and indicates if DTX was used. This information is given for every packet, since each can change.

To summarize, the pre-processing handles codec-specific operations, such as DTX, codec rate, and Channel Aware mode, and consequently ensures that the jitter files are both codec and reference voice sample independent.

How does sQLEAR work in the field?

As a QoE predictor of MOS, sQLEAR has to meet ITU-T requirements for test set-up, run-time, and/or measurement [P.863.1]. However, in order to ensure the best performance of the predictor against ITU-T P.863, some specifics need to be considered.

Reference voice samples. The machine learning-based QoE predictor has been trained using one or more reference files. Therefore, during run-time, when the algorithm is deployed, the same reference file(s) must be injected to the network under test. The measurement methodology and reference file(s) requirements follow ITU-T P.863/P.863.1 specifications.

Pre-processing during run-time. The pre-processing phase is performed in the same way as in simulations, as described above. In addition, during run-time, the pre-processing synchronizes the reference voice and the IP/RTP stream to ensure that the position-based features reflect the reference voice sample(s). This is performed by correlating the pattern of DTX and voice frames with the reference voice sample, which is stored at the receiving side. It should be noted that no recorded voice is needed.

Measurement procedure. sQLEAR uses the same test set-up as ITU-T P.863, but without the need for the recorded degraded voice sample as has already been explained. In addition, the recorded speech can be saved for further off-line analysis., if needed or desired.

The test set-up and run time / measurement scheme is illustrated in **Figure 5**.

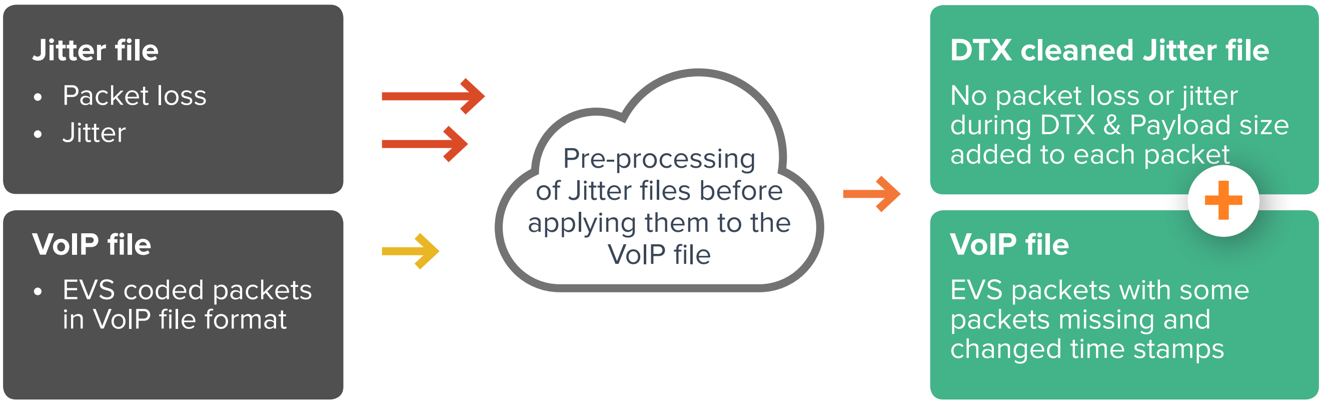

Figure 4. Jitter file pre-processing process

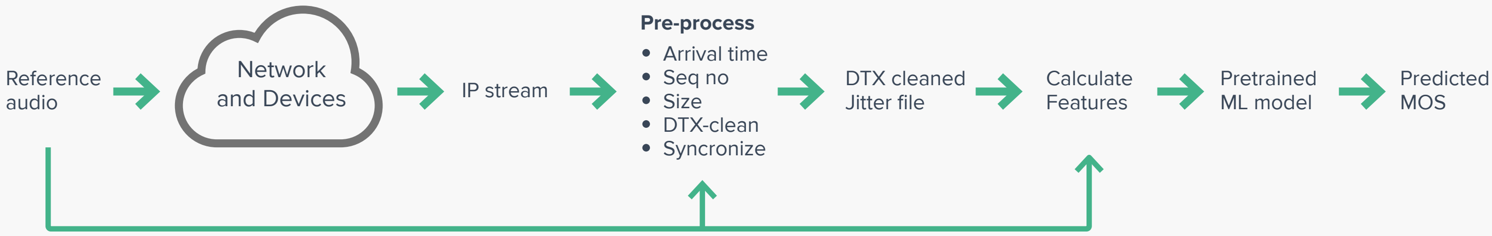

Figure 5. sQLEAR run-time scheme

Does sQLEAR work for different languages?

sQLEAR has been trained and validated on British- and American-English. However, as previously mentioned, one of the key strengths of the sQLEAR algorithm is that it can easily be trained and adapted. Therefore, it is a simple process to add different languages. This is a valuable feature when considering MNOs with multi-national footprints, which need to optimize costs at a group level.

sQLEAR does not use the audio path, but rather only the time structure of the reference signal for identifying the importance of individual sections of the bitstream in regard to speech quality, as well as for the creation of reference-based machine learning features. The addition of new languages requires only a brief temporal analysis of the reference voice file and the subsequent learning of the algorithm.

What is the accuracy of sQLEAR?

As the first delivery of ongoing work in the ITU-T P.VSQMTF "Voice service quality monitoring and troubleshooting framework for parametric intrusive voice QoE prediction" work item, sQLEAR meets ITU performance requirements. **Figure 6** shows the results of more than 96% correlation and prediction errors (rmse) lower than 0.26MOS across all evaluation databases (60,000 samples). These performance values are recognized to describe a high accuracy when considering the large amount of learning and evaluation data points; the difficulty created by the severity of the network conditions; and when compared with other existing ITU-T non-perceptual (e.g. P.564) and/or ITU-T perceptual non-intrusive (e.g. ITU-T P.563) voice QoE metrics.

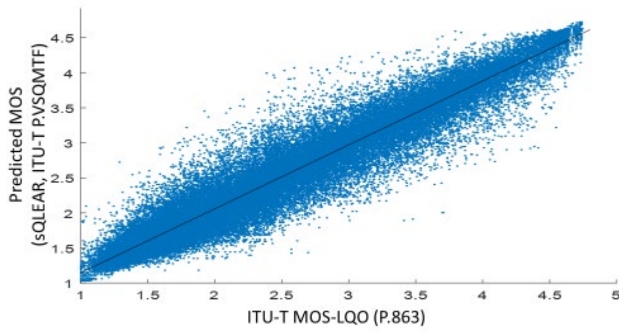

Figure 6. sQLEAR accuracy

How can sQLEAR be compared with other solutions?

It is already well-known in the technical community that develops QoE algorithms (and equally well-defined by ITU-T), that a direct comparison between two different QoE metrics is not valid, especially when the two algorithms are designed based on different approaches and/or describe different voice quality aspects.

This is also the case for sQLEAR, which is based on a completely different approach from any other available voice QoE measurement technique, such as ITU-T P.863, ITU-T P.563, ITU-T P.564. It is not possible to compare directly an sQLEAR score, which is independent of the audio path and therefore transparent to the device's voice frequency characteristics, with an ITU-T P.863 (or an ITU-T P.563) score which works on the voice signal (audio path) – and thus embeds the device's performance. Neither can an sQLEAR score be directly compared with an ITU-T P.563 or ITU-T P.564 score since these two solutions have not been designed for VoLTE scenarios, much less so HD voice (EVS codec).

However, because both P.863 and sQLEAR support VoLTE service evaluation, quality trends provided by sQLEAR and P.863, and determined based on a large number of samples collected using the same voice references and identical network conditions, are expected to be the same with a high statistical significance confidence level. However, it should be noted that test devices which exhibit a particularly strong voice frequency characteristic could be significantly penalized or favored by perceptual models, such as ITU-T P.863. These effects are outside the network and therefore do not represent the network centric voice quality performance, which in turn sQLEAR is designed to predict. Consequently, in these special scenarios, depending on the strength of the device's voice frequency characteristic, expected and rightful differences can be detected between sQLEAR and ITU-T P.863.

What are the differences versus other voice QoE solutions?

By leveraging network/codec/client parameters and the reference speech sample, sQLEAR is neither solely speech-based, like perceptual intrusive and non-intrusive algorithms (e.g. ITU-T P.863, P.563), nor solely parametric based and non-intrusive such as ITU-T P.564. In addition, sQLEAR has been designed for the evaluation of VoLTE service's quality, while ITU-T P.563 and P.564 has not.

sQLEAR is uniquely based on machine learning, which ensures faster and easier adaptation to new environments, such as new codec/clients, and network parameters. This can be accomplished without new and costly subjective training sequences, which are required by perceptual intrusive models (e.g. ITU-T P.863).

sQLEAR avoids device specific degradations caused by the audio path of a mobile device and focuses on the packet-based radio and core network. In addition, it does not use recorded speech, since that will also reflect the specific device used as a measurement unit, instead of network performance. Therefore, sQLEAR predicts speech quality efficiently and effectively under the following circumstances and conditions:

- Independently of device acoustical characteristics (unlike P.863);
- Without the need of tuning and calibration for each device (unlike P.863), which eliminates costly implementation time; and
- From the network- and client-based error concealment perspective (one of the unique characteristics of sQLEAR), which is the most cost-efficient means to enable network optimization for high quality voice services since it renders speech quality scores comparable between different device models.

As previously mentioned, the sQLEAR measurement procedure is the same as for ITU-T P.863, in the sense that it sends a reference speech sample to the system under test and predicts voice QoE from the combination of output from the device and the sent reference sample. However, the output is different. In the case of sQLEAR, the output from the device is the RTP stream, while in case of ITU-T P.863, it is the recorded audio.

What are the similarities with other voice QoE solutions?

sQLEAR and ITU-T P.863 both belong to the class of intrusive voice quality evaluation algorithms, since they send test stimuli through the network under test.

Both ITU-T P.863 and sQLEAR support VoLTE services evaluation and are part of ITU work. sQLEAR is based on the ongoing ITU work item, ITU-T P.VSQMTF "Voice service quality monitoring and troubleshooting framework for intrusive parametric voice QoE prediction".

More questions?

For more questions contact us at:
www.infovista.com

and watch the space for our forthcoming "sQLEAR Implementation Guide" to find out more about sQLEAR learning, evaluation and performance, detailed database description and pre-processing procedures, machine learning based feature selection, and much more!

Thank you!

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.