Solving QoS in VoIP: A Formula for Explosive Growth?

What users and their organizations expect from VoIP is essentially PSTN quality and objective verification that they are receiving it. VoIP service providers must meet these expectations if they want to profit from faster market penetration.

by: Luis F. Ortiz, Director of Software Architecture
Brooktrout Technology

The absence of solutions ensuring Quality of Service (QoS) has been a deterrent to the widespread adoption of Voice over Internet Protocol (VoIP). Potential users often think that speech quality won't be as good as what they are accustomed to—the familiar public switched telephone network (PSTN).

Solutions that improve absolute voice quality and that enable objective quality measurements can easily be incorporated into Service Level Agreements (SLAs). With these solutions in hand, the reluctance of potential users can begin to evaporate, and the adoption of VoIP will stand poised for a period of explosive growth.

Voice Quality and MOS Ratings

Voice quality is subjective because it's a measure of the intelligibility and clarity of speech as perceived by the listener. However, perceptions drive decisions. VoIP service providers must be extremely sensitive to the perceptions of their customers, because a decision to change service can be precipitated from such negative perceptions as

- When a user perceives unacceptable instantaneous quality, the user is likely to terminate the call prematurely.
- If a user perceives overall poor quality after completing a call, there is likely to be a harboring of residual dissatisfaction.
- If service providers achieve quality by overprovisioning their networks, the resulting high costs undermine the user's perception of value, despite excellent voice quality.

However, since a perception is of little use to a service provider unless it can be measured in some way, the industry has developed a numerical representation of voice quality called Mean Opinion Score (MOS). MOS ratings on a scale of 1 (bad) to 5 (excellent) are derived by soliciting perceptions from groups of real people to test messages.

MOS ratings are meaningful because they accurately reflect the dynamics of how perceived quality is affected by transmission-channel changes. If channel quality changes abruptly during a call, for example, the MOS rating recovers more slowly from a step improvement in quality (time constant about 15 seconds) than it does from a step degradation (time constant about 5 seconds).

Another example is the so-called recency effect. If a period of bad channel quality comes near the end of a call, it lowers the MOS rating for overall call quality much more than if it had appeared earlier in the call. While MOS ratings are

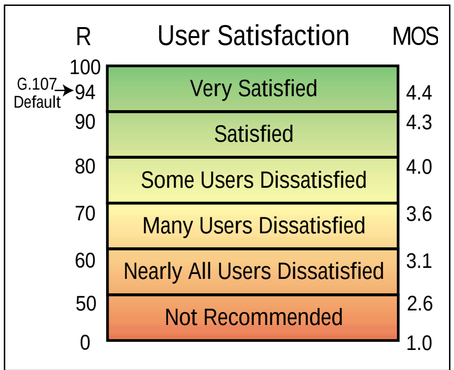

Figure 1 Comparison of R-Value, MOS and user satisfaction.

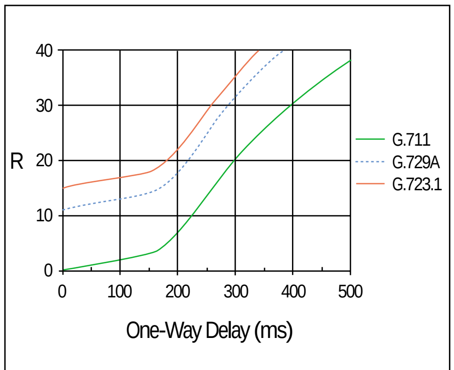

Figure 2 Delay degradation for three different codecs.

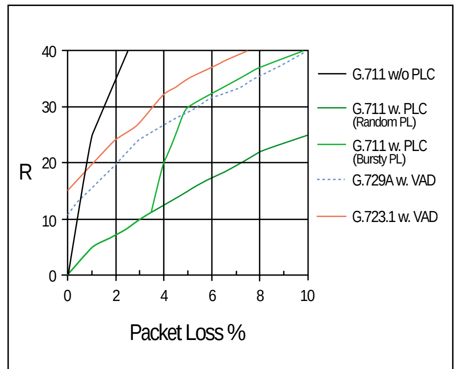

Figure 3 Packet loss degradation for three codecs showing the effect of PLC on VoIP quality.

meaningful for statistically characterizing user reactions to test messages, they don't offer a real-time, non-intrusive tool that's suitable for IP Telephony.

R-Value Correlates to MOS

An emerging standards effort within the International Telecommunication Union (ITU) is currently creating an "E-Model" for estimating the voice quality of IP telephony. The output of the E-Model is a scalar "Transmission Rating Factor" called the R-Value (or simply "R"). The importance of R is that it's repeatable and can be calculated in real time from measurable channel and equipment characteristics. With appropriate modeling, R can be correlated to MOS with remarkable accuracy. This gives service providers a mechanism for measuring the quality of their service (R) and accurately relating it to the perceptions of their users (MOS).

The idea is to start with a perfect score (R of 100) and then to quantify "R-Degradations" that are modeled from equipment and channel characteristics. The sum of the R-Degradations is subtracted from 100 to yield an overall R. The lower the resultant R, the lower the quality.

Figure 1 illustrates the relationships between user satisfaction and MOS or R-Values. With standard narrowband (300 to 3,400 Hz) telephony, the E-Model generates a maximum attainable R of 94.15 corresponding to a MOS of ~4.5. An R of 80 corresponds to a MOS of ~4.0, which is the nominal PSTN quality most users perceive as satisfactory. R-Values below the mid-60s result in widespread user dissatisfaction.

Quality Impairments and R-Degradations

There are many factors that can impair voice quality, and each of these can be associated with an R-Degradation. Technical Services Bulletin TIA/TSB-116 of the Telecommunications Industry Association (TIA) contains a discussion of all such factors, three of which will be discussed here: coding technique, delay and packet loss.

The coding technique has a profound effect on the response to channel impairments and the ability to employ error mitigation. In addition, a non-reducible R-Degradation is introduced by different codecs, which effectively lowers the starting point for the overall R-Value.

Codec	R-Degradation	Max. R-Value
G.711	0	94
G.729A	11	83
G.723.1	15	79

Figure 2 shows how one of the key factors in packetized voice—one-way delay—correlates to R-Degradation for different speech compression techniques. The knee in each curve illustrates how R-Degradation increases at a much faster rate after reaching a one-way delay of around 175 ms.

Packet loss occurs as a result of random channel errors, excessive congestion delays or from re-routing in the IP network. Lost packets typically occur in bursts, and two or more missing packets in a row are much more detrimental to speech quality than single missing packets distributed randomly.

Endpoint Strategies for Improving Voice Quality

Voice quality can be optimized by making trade-offs aimed at producing the lowest possible sum of R-Degradations. We concentrate here on endpoint strategies while recognizing that network optimization techniques, such as bandwidth reservation or the use of private vs. public resources, can help by improving the underlying quality of the transmission channel. Endpoint techniques for mitigating the effects of packet errors fall into four classes—correction, distribution, containment and concealment.

Packet error correction or error distribution mechanisms can be employed at the cost of additional delay and/or bandwidth. Error correction can be achieved using duplicate streams, Forward Error Correction (FEC) or jitter buffers to capture or regenerate delayed, lost or corrupted packets. Error distribution involves spreading out errors over a larger period of time so that multiple consecutive packet losses appear more randomly distributed.

These mechanisms must be judiciously employed, because using extra bandwidth increases transmission cost and adding extra delay introduces a greater risk of passing the knee of the Figure 2 delay curve where the quality effects of delay are exaggerated.

Error containment is largely coupled with the choice of codec. High-compression voice codecs such as G.729 and G.723 omit the transmission of redundant information and transmit only changes from one sample to the next. This means that if a frame is lost, the error persists into the following frames because the reference state of the decoder has been corrupted. codecs that don't depend on history, such as G.711, have a higher base R-Value and are more amenable to error concealment.

Packet Loss Concealment (PLC), wherein lost packets are replaced with substitutes, is the most productive error-mitigation technique for VoIP. There are four classes of PLC:

- 0th Order: substitute a constant packet
- 1st Order: substitute the last packet
- 2nd Order: substitute an interpolated packet
- 3rd Order: substitute a modeled packet

The 0th- and 1st-Order PLC methods aren't effective against the most important quality impairment, which is consecutive packet losses. They're therefore considered primitive by modern standards.

Substituting an interpolated packet (2nd Order) can be useful for both single and double packet losses. Logically, this involves computing the average of the neighboring packets and using this as an estimate for the missing packet(s).

Third-order techniques are the most advanced and are useful even in the face of triple packet losses. A 3rd-Order PLC engine monitors a stream of packets and builds a dynamic model of the embedded speech. When one or more packet losses occur, the output of the model is inserted to replace the lost packets. These PLC systems are good at tracking speech characteristics to which humans are sensitive, such as pitch or amplitude profiles.

Figure 3 illustrates the R-Degradation effects of packet loss using different codecs, with and without PLC. This model shows that PLC can have a dramatic impact on VoIP quality in the face of impairments.

Some PLC schemes, such as those available from Global IP Sound, achieve even better results than illustrated in Figure 3 (while maintaining compatibility with existing codecs at the transmitting end). Codecs using enhanced techniques can sustain reasonable voice quality even through significantly impaired channels, which use a combination of error distribution, error containment and 3rd-Order error concealment.

Dynamic Measurement of Voice Quality

The quality monitoring tools in widespread use today don't satisfy the needs of VoIP service providers. Some are based on long-term averages of network statistics for jitter, packet loss and delay, which don't relate directly to how users perceive quality on individual calls. Others are intrusive snapshots, which at best reflect averages rather than experiences of individual users. Examples of the latter are Perceptual Speech Quality Measure (PSQM) and Perceptual Analysis Measurement System (PAMS) test calls, which actually impair quality by using valuable network bandwidth.

The limitations of today's quality monitoring systems are overcome with new tools based on emerging standards that extend the E-Model to include packet-loss distribution and recency effects. A prime example is VQmon, a product from Telchemy, which models the effects of time-varying network impairments and human short-term auditory memory. It's a non-intrusive, real-time monitoring technique that is implemented at the network endpoints, typically at the egress from a VoIP service provider.

In a recent subjective study comparing VQmon with the widely used PSQM and PAMS measurement algorithms, VQmon was shown to be more accurate in predicting the quality rank (MOS) given by human listeners. This is particularly impressive when considering that PSQM and PAMS operate by comparing the original unimpaired audio file with the impaired file (strictly an off-line function), whereas VQmon operates in real time by analyzing the pattern of transmission channel events.

VQmon computes the R quality metric at the conclusion of each call (and on an instantaneous basis, if desired), taking into account dynamic user-perception effects such as recency and gradual recognition of quality changes. Presented at the conclusion of every call is a detailed record of user-perceived voice quality for that call:

- R_{end} represents user's memory of call quality, incorporating recency effect.
- R_{av} represents the average quality for the specific call.
- R_{max} represents the maximum quality level experienced during the call.
- R_{min} represents the minimum quality level experienced during the call.

VoIP service providers can easily integrate this information with their network or service management systems. They can also include it into SLAs, because it documents the real-time quality experiences of every call. VQmon is being integrated into commercially available products such as endpoint processing and interface modules from Brooktrout.

Recent technology advances have made possible a comprehensive framework for addressing QoS in voice-over-packet services. Metrics based on the standards-based E-Model can be used in two key ways to overcome previous obstacles to VoIP market acceptance: 1) to apply trade-offs that improve absolute quality and 2) to generate accurate estimates of user-quality perception on a real-time, call-by-call basis.

This allows VoIP service providers to offer SLAs with the kind of meaningful voice-quality guarantees that current and potential users have been demanding. Once users can secure the benefits of VoIP without concerns over quality, widespread market acceptance will be the likely outcome. Now is the time. ◢

For more information please contact:

Brooktrout Technology

Needham, MA

Tel: (781) 449-4100 • Fax: (781) 449-9009

info@brooktrout.com

www.brooktrout.com

Brooktrout Technology Europe, Lid.

Hoeilaart Office Part

Vandanmmestraat 5, Box 2

1560 hoeilaart, Belgium

Tel: 32-2-658-0170 • Fax: 32-2-658-0180