Call Quality Is More Than Voice Quality

Gary Audin and Dr. Fiona Lodge

Users also care about time to dial tone, service accessibility and other factors that make up the calling experience.

We can no longer presume the reliability, stability or predictability of our telephony voice services. The performance we took for granted in traditional telephony can only be achieved in voice over IP (VOIP)/IP-telephony deployments through monitoring, measurement and management. And what to monitor, measure and manage is now the responsibility of both the voice and the data technicians.

A caller's dialing experience and perception of voice quality are the biggest factors contributing to his or her satisfaction. The dialing experience can be measured, while the caller's judgment of voice quality is subjective.

A common misunderstanding centers around what constitutes call quality. It's essential that enterprises understand that voice quality and call quality are two very different beasts—each tamed and measured in different ways.

Enterprises can't afford the risk of providing poor service levels or failure to meet compliance standards. Obviously, voice quality is a big piece of the puzzle, but if you only monitor and manage this small piece, you're missing the bigger picture—and may in fact have a system that provides unacceptable "experience" levels—even though voice quality is fine.

Gary Audin (Delphi-inc@att.net) is president of Delphi, Inc. consultancy. He is a member of the VoiceCon Program Committee, a frequent BCR *author, and a teacher for several* BCR *seminars. Dr. Fiona Lodge is product manager for Prognosis and can be reached at fiona.lodge@prognosis.com.*

The Parts Of A Successful Call

A call begins when the user picks up the phone and ends when the call disconnects. The calling experience should take into consideration everything in between. This is what calling experience means: the quality experienced by the user for the entire call, including but not limited to voice quality. Voice quality alone does not account for other elements—such as delay to dial tone, connection success and service availability—that make up a satisfactory call.

The caller's experience includes four key factors: voice quality, call quality and service quality, which are experienced on every call; plus usability of supplementary services that may also be employed.

- **Voice quality—**Can the caller hear the other party? Can the speaker be recognized? Are the words garbled? Is there noise on the call? Parameters for voice quality should be determined, and objectives should be set and then measured for their delivery.
- **Call quality—**Does the caller have a dial tone? Does the PBX or PSTN set up the call? Again, parameters should be determined, and objectives should be set and then measured for their delivery.
- **Service quality—**Is the endpoint, or network, busy? Is the call lost? Can 800/900 numbers be accessed? There is a very long list of possibilities for service quality.
- **Usability of supplementary services—**Does the interactive voice response (IVR) application work properly, and does the voice mailbox have adequate storage for the calls?

Defining and monitoring all these factors will help to show that voice quality, though important, is not the entire picture when considering calling experience.

Analyzing Voice Quality

A caller's dialing experiences and perception of voice quality are the foremost factors in user satisfaction. Users can and do assign descriptive adjectives (good, OK, poor, terrible) to voice quality. The user's judgment of voice quality is subjective. Think of the cellular calls in your experience. If you know the caller well and understand their speech patterns, even a poor connection can carry a comprehensible conversation. Now imagine that you have the same poor-quality connection with someone you don't know, someone with an unfamiliar accent, who speaks quickly or whose native language is unfamiliar. Comprehension becomes much less likely.

Voice quality testing can be automated, using either passive or active systems

The primary description of an acceptable voice call can be summed up in one word: clarity. Clarity is the voice signal (speech) clearness, fidelity, intelligibility and lack of distortion. The following five components define the elements of sound quality for one direction of a call, and represent part of the overall voice quality:

- **Loudness—**The sound volume (signal) level cannot be too low or too high.
- **Distortion—**All speech is distorted, even in the PSTN conversion of analog to digital speech. But is the distortion perceptible to the listener? That is the important question. And, if so, does the distortion reduce overall voice quality?
- **Noise—**In all calls there exists background noise in the form of static and hum. This is known as noise level. The noise may, however, be at a level low enough that the listener does not notice it at all. In fact, "comfort noise" is injected into VOIP calls during silence suppression/voice activity detection (VAD), so that the listener has the perception that the call is still connected. An absolutely silent call (no noise whatsoever) is commonly assumed to have been disconnected.
- **Fading—**The signal level may change, increasing or decreasing during the call.
- **Crosstalk—**This is the condition where another party's conversation on a separate call can be heard on the user's call.

The next four elements complete the list of factors to be considered for voice quality:

- **Echo—**The sound of the speaker's voice returning and being heard by the speaker. Think of echo as a problem of long round-trip delay. The listener may not perceive short delay echoes. The longer the echo delay, the more difficult it is for the speaker to ignore.
- **End-to-end delay—**The time it takes for speech to travel from the speaker's mouthpiece to the listener's earpiece. This is a major problem when a call travels over a satellite link. The PSTN within the U.S. usually has a delay of 30 ms or less. The end-to-end delay goal is to have a one-way delay of 100 ms or less in VOIP calls, with an upper limit of 150 ms. Very long delays will cause the speakers to pause because they are not sure when the other speaker has finished, or they may barge in on each other's conversation, or simply hang up.
- **Silence suppression/voice activity detection performance (VAD)—**When these technologies are used, the beginning and the end of words tend to be clipped off, especially the "T" and "S" sounds at the end of a word.
- **Echo canceller performance—**The longer the end-to-end delay, the more the echo needs to be eliminated. Echoes may occur in only one direction or in both directions. The echo cancellers may not work or they may not be able to effectively compensate when there is significant jitter during the connection, such as occurs during VOIP calls.

FIGURE 1 Voice Performance Measurement Standards

	Analog signal	VOIP specific
Passive test of live calls (All calls)	ITU P.563	VQmon ITU P.VQT
Active test with test calls (samples)	ITU P.862 (PESQ)	VQmon ITU G.107

These nine elements, when combined, will contribute to the clarity (or lack thereof) of a voice call.

The Meaning Of Mean Opinion Score

Mean opinion score (MOS) is the PSTN standard numeric value used to measure voice quality. The MOS varies from a maximum of 5, which is considered to be essentially the same as speaking directly into the person's ear, to a value of 1, which is an unacceptable voice quality to all users.

A MOS of 4.4 to 4.5 is considered equivalent to a toll quality call as experienced on the PSTN; users will be very satisfied. A MOS of 4.0 is still considered acceptable to the vast majority of users. When the MOS decreases to 3.5, some users may find the voice quality unacceptable.

Most cellular calls have a MOS rating of 3.8, where speaker and word recognition may be impaired.

Typically, users will be dissatisfied and hang up when the MOS falls below 3.5. Below 2.6 is considered to be an awful call and the user will need to find an alternative network for this call, such as when a wireless call is terminated and the speaker moves to the PSTN.

The International Telecommunication Union (ITU) ITU-T standard P.800 for a MOS measuring technique was last updated in the mid-1990s and has always been a subjective exercise. About 30 or more people are asked to listen to 8–10 seconds of speech under controlled conditions. The listeners are asked to rate the calls from very satisfied to awful, scoring the calls from 5 to 1. MOS does not include what has been defined as the call experience, only the sound or voice quality. The industry started to move to objective machine measurement of voice quality several years ago with the advent of cellular phone networks.

There are two forms of testing voice quality: passive and active. Passive testing analyzes the received speech experienced during the call. This can be done on each call since the measurement adds no extra network traffic. The active approach sends a known speech file and compares the received file to an original copy of the speech file stored at the receiving location. This generates traffic and is done periodically, rather than for every call.

Active testing is better suited to testing a network before it goes live with VOIP traffic, and is also useful for troubleshooting. Continuous passive testing (monitoring) is better suited to a live VOIP environment, especially since the performance of the IP network varies over time due to changing traffic and performance conditions, and is therefore unpredictable.

Passive monitoring requires testing many live calls to develop a good performance measure.

There are now multiple methods and standards for measuring voice quality and determining a MOS rating (Figure 1):

- **P.563—**Formerly called P.SEAM, a passive system for the analog call.
- **Perceptual Speech Quality Measure (PSQM)—**ITU-T standard P.861 was designed to measure analog-to-digital conversion and is not good for testing VOIP because it does not account for jitter.
- **Perceptual Evaluation of Speech Quality (PESQ)—**ITU-T standard P.862, which was approved in December 2003.
- **Perceptual Analysis Measurement System (PAMS)—**This is a technique like PSQM, developed by British Telecom and now supported by Psytechnics.
- **G.107—**Also called the E-model, this is a mathematical formula for calculating/predicting the MOS. It is more of a network-planning and troubleshooting tool.
- **VQmon—**A proprietary technique, developed by Telchemy and based on G.107 plus ETSI TS 101 329-5 Annex E.

How good are these machine methods for determining a MOS rating when compared to a human MOS rating? Tests have been performed comparing the human measurements to different machine measurement techniques. The results are as follows:

- P.563 has a correlation to the human MOS scoring of 0.85 to 0.90.
- PESQ has a correlation of 0.95 and is therefore a more accurate technique.

A correlation of 1.0 is an exact match and indicates that the people-generated and machine-calculated MOS are exactly equal. Correlation numbers of less than 1.0 mean that the machine technique is a less accurate measurement of the voice quality.

Multiple standards exist for automated measurement of voice quality

Voice Packets Over A Data Network

Voice quality is heavily affected by the underlying IP network infrastructure. A VOIP call starts as an analog signal that is digitized and possibly even compressed in a codec. A piece of the spoken word, about 10, 20 or 30 ms of digitized sound, is placed inside an IP packet. As the packet travels through the IP network, it encounters network impairments.

The receiving VOIP device compensates for delay variation (jitter), replaces lost packets, removes the digitized portion of a word from the packet, then reconstructs the entire word and, finally, converts the digitized word back into an analog sound. This process, illustrated in Figure 2, determines the delivered voice quality.

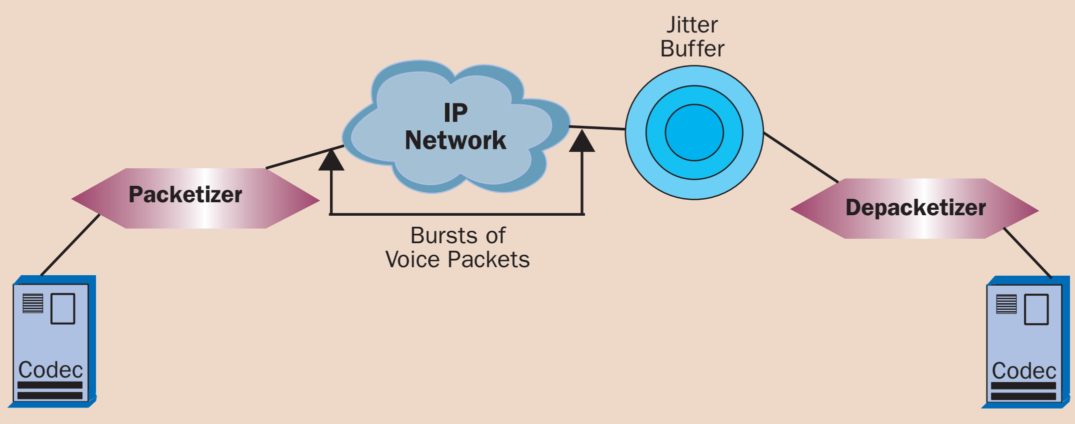
In VOIP/IP-telephony implementations there will be network impairments, such as extended latency, packet loss and jitter, that are not usually perceived by the data network user, but cause considerable degradation when voice packets are carried through the same data network.

Latency (end-to-end delay) can be frustrating to a caller if the delay is too long. A voice user would typically disconnect a call if delay were greater than half a second one way, or one second round trip. For this reason, connections with very long round trip delays do not qualify for an acceptable voice call unless there is no other alternative connection.

IP network performance issues that cause voice quality degradation are:

- **Latency—**The one-way delay, in milliseconds (ms), from the network entry point to the exit point. For PSTN calls, latency is specified/recommended by the ITU standards.

FIGURE 2 VOIP Packet Path



The ITU standards are far more lenient than end users will actually accept

TABLE 1 Typical Time For Voice Functions		
Function	Typical PSTN	VoIP experience
Time to dial tone	0.5 seconds	0.5 to 5 seconds
Ring Alert	0.75 to 2 seconds	1 to 10 seconds
Disconnect time	1 to 2 seconds	1 to 5 seconds
Recovery time to next dial tone	1 to 2 seconds	1 to 5 seconds

TABLE 2 Network Availability		
Function	Typical PSTN experience	VOIP experience
Network busy	.01%	.01% to 2%
Probability of lost call	.001%	.001% to 2%

- **Jitter—**The variation in the delay as the packet moves across the IP network, measured in milliseconds. Some packets arrive on time while others incur extra delay. This is not mentioned in the ITU PSTN phone standards because the PSTN jitter is so short that it does not produce voice quality problems.
- **Packet loss—**Since VOIP packets contain speech, a lost packet is equivalent to losing part of a word. Packets may be lost randomly. This is not difficult to fix at the receiving location if packet loss is less than 1 percent. However, the packet loss problem becomes more difficult to resolve when a burst of packets is lost, for example four packets lost at one time. This leads to garbled speech because the receiving location cannot fix the problem. This is not mentioned in the ITU PSTN phone standards because the PSTN does not carry speech in packets.

What Is Call Quality?

Over and above the issue of voice quality, the elements that comprise call quality include:

- Delay time in hearing the dial tone after the phone is off hook. In many VOIP phones, the dial tone is generated by the phone, not the call server, and should therefore be instantaneous. When the dial tone is generated by the call server (as with Cisco CallManager), there can be noticeable delay that is influenced by the network performance between the endpoint and the call server. Dial tone time may also be influenced by the IP phone design. Dial tone generation may take longer in a softphone (i.e., a voice client running on a PC).
- Time to be alerted that the phone at the remote end is ringing.
- Time to disconnect the remote end connection at the end of a call.
- Recovery time to get the next dial tone.

The ITU standard measurements allow far longer delay for these functions than are commonly experienced in North America. For example, E.721 defines time to dial tone of .3 to 3 seconds for 95 per cent of calls, much longer than is typical (Table 1).

Using the standard measurements would lead to a very dissatisfied set of users. The VOIP experience should be compared to the PSTN experience, not the ITU standards.

Delivering Service Quality

Service quality covers a range of elements, all dealing with what is offered and how well the offered services work. Each of these features can be quantified. Some require only a Yes or No depiction. Others will have numeric values measured in time or frequency. All can be specified, measured and reported. Some can be determined at the time of implementation, while others must be monitored continually.

The types of services offered may include:

- Basic telephony service.
- 1-800 and 900 services.
- Voice mail.
- Call forwarding.
- Follow-me service.
- Caller name and number.
- Call blocking.
- Automated replies.
- Priority/precedence calling, plus as many as 400 more services.

Service quality of these offerings is vitally important to call centers, 911 Public Safety Answering Points (PSAP), help desks and any other call servicing organization. The quality of these services includes the following considerations:

- Is the service accessible?
- Is the service available at both ends of the call?
- How often does it fail?
- How long before service is restored?

These quality aspects are affected by the following types of issues (also see Table 2):

Telephony service availability:

- Frequency of network outages.
- Restoration time for the network outage.

Local-switch busy signals:

- Network busy signals.
- End-point busy signals.

Reliability of use:

- Lost (dropped) calls.
- Wrong number connected.
- Calls where the connection is not completed.

The ITU-T recommendation I.352 document, “Network Performance Objectives for Connection-Processing Delays in an ISDN,” specifies the requirements for the delay measurements. These are far more generous than would be acceptable for the North American PSTN.

Network availability— defined as, “Is the telephony service available to the user?”—is a ratio calculated by combining uptime and downtime into one number with a goal of 99.9 percent or better service (Table 2), and caluculated thus:

$$\text{Availability} = \frac{\textit{Uptime}}{\textit{Uptime} + \textit{Downtime}}$$

Uptime hours are equal to all the time a user can access the network for voice services. Downtime equals all the time the network is not accessible by the user. Vendors and service providers do not typically include natural disasters, power failures, maintenance, software upgrades and other customer-related suspension of operation in their definition of downtime.

TDM system vendors have been telling customers that the systems are 99.999 percent available. Rarely, however, has a customer actually measured availability to confirm this.

Usability Of Telephony Services

This fourth component of the calling experience is more qualitative than quantifiable. The goal of this aspect of the calling experience is the uniform delivery of services across the network and endpoints. The issues here include:

- Are services available to all users?
- Will there be version variations across the network?
- Are services like 911/E911 dependable?
- Are directories easy to use, consistent and accurate?
- Are dual tone multi-frequency (DTMF) signals not garbled and are they being read correctly by interactive voice response (IVR) applications?
- Does the network have call routing intelligence?
- Can the endpoints and network comply with regulations and auditing requirements?

The Call Experience: Measurement, Then Management

Data network management products initially dealt with fault and configuration management. These management systems were mostly concerned with the status and configuration of the network components.

As the move to real-time applications progressed, enterprises needed more insight into the ongoing operation, specifically knowledge of performance, accounting and security. Now the industry is taking the next direction: collecting more information faster and interpreting that information in a more useful form.

Alarms are not enough. Alerts to growing problems that will eventually lead to user dissatisfaction are now necessary. Assessment of network and endpoint performance is also now an integral part of management system responsibilities.

Today's VOIP/IP-telephony management systems can be applied throughout all phases for the migration to VOIP, from pre-implementation to operations after successful deployment. When deploying a management system in your organization, the following should be considered:

- To assess the data network's ability to support voice traffic, a management system should be used prior to VOIP deployment.
- Your management system should perform assessments (they will be needed) once the data network improvements have been made. The system should also be used to determine if the improvements met the designer's expectations.
- It is wise to pilot new VOIP products before full deployment. The management system should be used to measure the quality of call during the pilot to ensure the pilot meets expectations.
- Once your VOIP system and endpoints have been fully deployed, the management system will be in continuous use for diagnostics, troubleshooting, reporting and capacity planning.
- The management system should also be used to monitor and measure the impact of new applications and endpoints as they are added to the network.

Final Thoughts

Although they can be measured individually, voice quality and call quality can only be truly managed as a whole. Therefore, monitoring, measurement and management of voice quality and call quality are essential for any enterprise VOIP/IP-telephony deployment. No caller will settle for poor quality, regardless of the cause□

Few customers have attempted to verify vendors' claims about "five-nines" availability