

# Quality of Service Testing in the VoIP Environment

March 2005



In recent years, the business world has reaped tremendous benefits from the many exciting products and applications made possible by the marriage of data and voice technologies. Now, the efficiencies and enhanced services resulting from that revolution are being enhanced by the magnitude of change possible as data networks become the transport for voice. IP Telephony, or Voice over IP (VoIP), is the exploding new technology enabling voice to be carried over IP-based, packet-switched local and wide area networks.

Now the efficiencies and enhanced services resulting from that revolution are being eclipsed by the magnitude of change possible as data networks become the transport for voice. The open standards approach adopted within the voice over IP world has meant that public and private network operators are now able to remove their dependence on buying everything from one vendor that was a fact of life in the closed/proprietary nature of the traditional circuit-switched world. Other advantages that have attracted a lot of attention include the ability to bypass the PSTN and its toll charges, applications made possible by merging voice with Internet data, lower operating expenses, and network service presence..

### **Why is VoIP Attracting so Much Attention?**

The advantages to a company adopting the technology are significant. In traditional circuit-switched networks, when a connection is established, a channel is dedicated end-to-end for the duration of the communication. This means that any unused bandwidth (around 60 percent of speech is silence) remains unavailable until the call is released. In the packet switched world, many types of communication share the total bandwidth, which makes much more effective use of the available capacity. Speech compression technologies, used in preparing voice signals for transport over packet networks, require less bandwidth for the “voice” signals and eliminate “silent space” thereby saving still more bandwidth. In addition to this economy of scale, combining all traffic onto a single network represents an opportunity for major savings in the physical plant. Other advantages that have attracted a lot of attention include the ability to bypass the PSTN and its toll charges, and enabling applications made possible by merging voice with Internet data.

Because of this enormous potential, the market for VoIP solutions is ramping up quickly. However, there remain significant quality issues that must be addressed before we see more widespread acceptance of VoIP as the mainstream telecom service and business tool.

In the brief history of IP telephony, voice connections over the Internet have received a bad reputation because of poor quality. Due to its real-time nature, an effective voice conversation requires a reasonable level of continuity. That continuity can be negatively impacted by large numbers of packets (representing other types of data) competing with voice packets for network bandwidth, a situation that never existed in the circuit-switched world. The voice quality issue is made more complex because of its subjective nature. Equipment responsible for processing voice for transport over an IP network must be able to retain all the nuances, inflections, and pauses that comprise effective human communication; not always an easy task given the challenges mentioned previously, and one whose capability must be verified using methods that take human perceptual subjectivity into account.

It is important that service and equipment providers build into their VoIP solutions the ability to test, measure, and evaluate the performance of the various elements needed to create a VoIP transmission. This paper will identify those elements and suggest some strategies for testing that can help ensure the level of quality required to make VoIP a viable service offering. This paper is intended for any manufacturer, system integrator, service provider, or enterprise for whom guaranteeing solid voice quality performance is a critical issue.

## VoIP Call Elements

A VoIP call can consist of several elements: endpoints, gateways, call servers, a packet-switched network, and sometimes also a circuit-switched network. Which of these elements are present in a particular call depends on what type of endpoints are being used and what type of call is being made. An endpoint in a VoIP scenario can be an IP telephone (a PC with a “softphone” installed or an IP phone itself), a mobile phone (TDM or IP based), or a regular POTS telephone. In a teleconference situation, all types of endpoints will likely be participating, as well as media servers.

The difference between each of the elements that are present is determined by where the necessary voice signal processing is done to package the voice for transport over the packet network. If the endpoints are IP telephones, the speech encoding and packetizing functions are incorporated within the IP phone, and a call would look like Figure 1.



Figure 1.

When a standard telephone is at one or both ends of the connection, an interface must be provided between the circuit-switched and packet-switched networks. IP Telephony gateways are equipped with standard interfaces to the PSTN (analog, T1/E1) as well as interfaces to the packet network (Fast/Gigabit Ethernet). The necessary signaling conversion, media transcoding (encoding/decoding, compression/decompression), and packetizing/de-packetizing is done in-between. See Figure 2.



Figure 2.

The processing of a voice signal into the format necessary for transport over a packet network is performed in all cases by an encoding/decoding subsystem called a vocoder. These systems encode, compress (usually), and packetize the signal. When the signal reaches its destination, the process must be reversed by the vocoder on the destination end, either in a gateway or the endpoint itself. The implementations of vocoder algorithms and packetization can differ from manufacturer to manufacturer, and the effects can vary greatly. Compressing/decompressing audio signals can introduce latency and reduce the quality of the signal, whilst packing/unpacking data can introduce latency and jitter.

The final element is the packet-switched network itself — the “cloud” that provides the data transport between the other elements. The network, consisting of various physical media, network protocols, and the routers and switches controlling the flow of traffic, is the most problematic of the connection elements, as will be seen in the next section.

**Note:** while the PSTN interface is an important component, and should be tested, it does not generally impact voice quality and is not discussed here.

## What Needs to be Tested, Measured, and Evaluated

In the previous section, we identified the elements that comprise a VoIP transmission. This section will discuss the potential problems these elements can introduce, usually when trying to perform under heavy traffic demands: connection failure, latency (delay), jitter (variable delays) and dropped packets.

**Connection Failure.** The endpoint applications and devices discussed above need to be able to place and receive calls, so this capability needs to be verified. A gateway needs to be able to receive and send circuit-switched traffic on one side and packet-switched traffic on the other, and this basic functionality needs to be verified as well. Call control signaling is the key for establishing and maintaining successful connections, as well as tearing them down again.

**Latency.** Voice signals need to be processed for transport over a packet-switched network. The necessary compression and packetization (and the reverse of these processes) is done either by the intelligent endpoints or a gateway. Execution of these functions requires a small amount of time, which can vary depending on the architecture of each device (DSPs, compression algorithms, distributed signaling and media) and the amount of traffic to be processed. This processing time introduces delay, which is called latency. The human ear, being a subjective evaluator, can tolerate some latency, usually up to around 250ms, before perceiving a drop in the quality of a connection. So, knowing how much latency an endpoint, network, or gateway introduces, especially when traffic load is high, is important to test in order to ensure the 250ms threshold is not exceeded. As it happens, the major portion of the delays are introduced after the packets leave the endpoint or gateway. Depending on how busy each successive router in the network is, it can introduce another few milliseconds or more into the cumulative latency. Outside of a carefully managed network, there is no control over the number of router-to-router legs or hops a packet has to take. Therefore, monitoring the total end-to-end latency that packets are experiencing is necessary in maintaining a good quality VoIP transmission.

**Jitter.** Not only is it impossible to predict or control (using current networks) how many hops packets from a VoIP call will traverse, packets from the same call can be assigned different routes, with varying numbers of hops and different traffic volumes along the way. Because of this, packets from the same conversation can experience different amounts of delay on their way to their destination. These variable delays produce a condition called jitter, where packets arrive at their destination at different intervals. Most gateways have buffers to collect packets and return acceptable continuity to the data, and these must be suitably tuned so that the process itself does not introduce excessive delay. The human ear can tolerate some jitter, usually up to around 75ms, before perceiving an unacceptable drop in quality. So, another area of testing would involve monitoring jitter to make sure it is being dealt with effectively.

**Dropped Packets.** When a router becomes overloaded with traffic, it may intentionally drop packets to relieve the congestion. Routers or gateways may often drop packets when the packets in question arrive out of order with an excessive amount of delay. Too much jitter can also result in overflow of the jitter buffers and thereby loss of packets. With traditional data traffic, for which these networks are optimized, there are error-checking methods built into the protocols to address these situations and maintain data integrity. These methods require some overhead which is not conducive to real-time traffic, and were not implemented for voice transport. Again, a certain number of missing packets (generally between 1% and 3%, depending on the data represented) can be forgiven by the human ear. Beyond this, the call quality can degrade to unacceptable levels, so it is important to monitor and test for dropped packets.

## How to Test, Measure and Evaluate

We have identified several conditions which, if they occur, can negatively impact a user's perception of the quality of the VoIP transmission: connection failure, latency, jitter and dropped packets. The failure of a call to connect is an obvious and easily measured call control problem, but the effect the other conditions have on voice quality is more difficult to quantify; how an audio signal is perceived by humans is very subjective. Because of this, it is important to closely simulate “real world” conditions so that testing is done on what humans are actually hearing.

How then do you create a test environment in which an effective evaluation of the performance of a VoIP device or network can be done? Using a Hammer test solution as an example, we will examine how the requirements for testing can be addressed and reliable tests and measurements provided. As stated previously, heavy load conditions are most likely to be the cause of performance degradation. Unless the testing being done is on a system already in service in the real world, that load must be simulated. The Hammer system appears to the System Under Test (SUT) as users, by generating the same type of traffic that actual users would. In order to do this, the Hammer incorporates telephony protocols and interfaces capable of sending signals over analog (the type of traffic typically offered by home telephones), TDM (E1 or T1 using CAS, ISDN, or SS7 — typically used in calls from switches), and IP (Ethernet using e.g. SIP for call control and RTP for media). The amount of load needed to exercise a SUT would determine how many of these interfaces are necessary.

A test scenario would be incomplete without creating end-to-end connectivity in approximation of real world circumstances, which would involve either the same or another test system receiving the traffic and recording information on the sound quality and timely receipt of the calls. This test environment is shown in Figure 3.

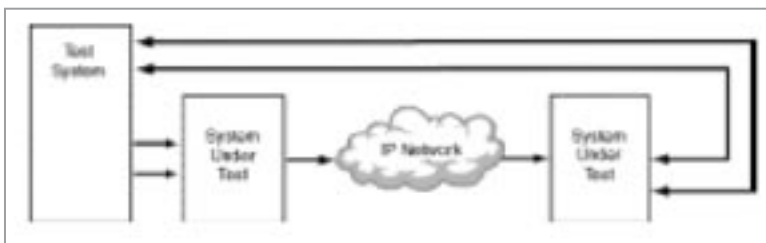


Figure 3.

With the physical resources in place to create the traffic load, there must be an easy way for the tester to control it. In the Hammer, this is provided by a single screen, Windows-based user interface from which a series of predefined test templates can be filled in with appropriate variables, such as calls per hour and call duration. Also, from this point, the tests can be assigned physical resources and either started immediately, scheduled, or saved for future scheduling. These tests should include:

### Connection Testing

Calls are generated, the connections between the originating and receiving ends are verified, and the calls are torn down. Tones may optionally be sent to make sure data can be successfully passed.

### DTMF Testing

Because users may be using VoIP services to access systems that require DTMF inputs (IVRs, for example), and because DTMF tones are handled differently than speech when processed for IP transport, there should be a test dedicated specifically to checking their integrity. Several DTMF tones should be played sequentially into the SUT using many telephony interface channels at once and their integrity verified at the receiving end.

## **Telephony Load Testing**

Varying levels of telephony traffic load should be generated, with the amount of load being “dialable”, perhaps in increasing increments or randomly, on many different channels. Audio quality should be continuously checked to determine if there is a load level where it begins to decline.

## **Mean Opinion Scores of Speech Quality (Subjective)**

But what is the most meaningful method of assessing audio quality? As previously stated, the true final arbiter of what is acceptable is the human judge. The ITU-T (International Telecommunication Union, Telecommunication Standardization Sector) developed methodology to standardize the process. This methodology is presented in the P.800 document.

The P.800 Recommendation describes the steps necessary to arrive at Mean Opinion Scores, which represent the benchmark assessment by a group of human judges of the sound quality of speech clips recorded and listened to under specifically controlled parameters. The clips are sentences chosen not for their meaning, which is somewhat nonsensical, but rather for the range of sounds they encompass. The sentences are spoken by male and female, adult and child voices, so that a wide assortment of human sounds are represented. The clips are then played to the judges in specially designed listening rooms, where noise and other environmental factors are controlled. From their ratings on the sound quality of the speech clips, MOS scores of 1 to 5 are derived, with 5 being the highest quality.

## **Perceptual Evaluation of Speech Quality (Objective)**

Recognizing the necessity of incorporating the human factor but realizing the impracticality of always having a group available to evaluate transmissions, the ITU-T developed methodology to automate the process. This is presented today in the P.862 document, and the methodology it describes is the most well-established and fully realized currently existing in this area.

The P.862 Recommendation, which replaces the earlier P.861 document, introduces algorithms that automate the evaluation of sound transmission quality using repeatable, objective calculations that incorporate the necessary subjectivity of the human factor. This method of analysis is called Perceptual Evaluation of Speech Quality (PESQ), in which real voice prompt signals that constitute the original source speech clips and encoded speech (speech that has passed through a vocoder) are aligned in time, and are then mapped onto psychophysical speech (psychoacoustic) representations or, in other words, how speech is perceived by the human ear and brain. Taken into account, in these representations, are weighted factors to allow for the subjectivity of human perception. An example of this would be background noise in a transmission (hiss, static) that seems worse during a silent pause than while someone is speaking. Once the mapping of signals is complete, a “cognitive subtraction” is performed and a quantitative measurement of the results is produced. The results are then algorithmically correlated to benchmark MOS scores, where a maximum score of 4.5 is representative of the average top score achieved in a judging group adhering to P.800. This method has been incorporated into the Hammer IP test systems. Any test system should incorporate an industry-defined and accepted procedure for measuring speech quality that includes recognition of the human factor, in a quantifiable and repeatable manner.

Having discussed how to stress a VoIP system with telephony traffic, what is this type of traffic's relationship to the conditions described previously?

A heavy traffic load is the primary contributor to system and network delay, jitter and dropped packets. Therefore, an effective test system should be measuring these manifestations of performance degradation while the VoIP system is dealing with telephony load. In order to detect these conditions, the test system must be able to "sniff" packets on the network with audio content, and understand routing and control information embedded in the packet. For instance, out of place or missing sequence numbers would indicate a level of jitter, or that packets were missing. By time-stamping call events as they are generated and comparing the stamps to the synchronized clock upon receipt, end-to-end latency could be measured.

As these indications worsen, the effect on voice quality would increase, with expected results as in Figure 4.

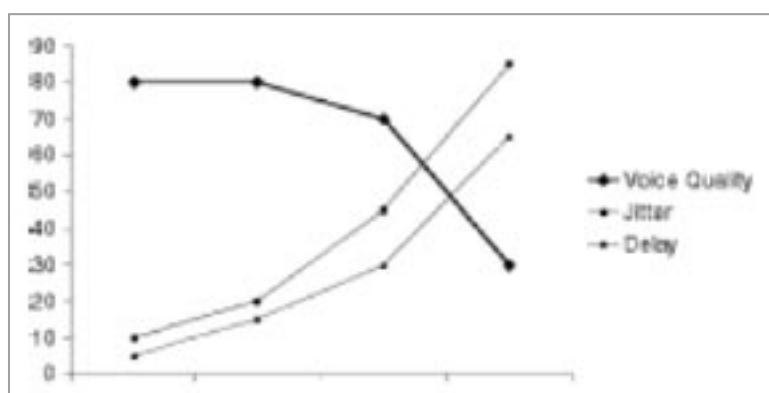


Figure 4.

It would make sense to incorporate some threshold-based analysis into the capability of the test system. In addition, at any time, a tester should be able to actually listen to an audio transmission, to gain an understanding of what a real-world user would be hearing under a variety of conditions.

What if a test system could actually create latency, jitter, packet loss and packet corruption, as well as emulate IP traffic load on top of the telephony load? Now the VoIP tester would be able to simulate a larger array of the potential real world circumstances, making the testing environment truly comprehensive. The Hammer test solution includes these network impairment generation capabilities. Any company that is serious about providing test systems to the VoIP industry should have these capabilities either supported or in their plans.

## A Word About SIP

Another issue that has, until relatively recently, presented a challenge to the expansion of the VoIP industry is a lack of agreed upon standards and interoperability. Different implementations by equipment manufacturers can create significant interoperability problems. Now it appears that most have accepted the IETF's Session Initiation Protocol (SIP) standard for multimedia communication between devices over IP networks.

SIP is quickly becoming the cornerstone for LAN and WAN based consumer, business entertainment, and professional applications. It is supported by the products of most VoIP equipment vendors and service providers. As this evolution continues, industry bodies (such as the SIP Forum) play an important role in ensuring the successful implementation and interoperability of SIP technology and networks. To help with this, a VoIP testing system must keep pace by incorporating the ability to effectively test SIP in tandem with the ability to emulate other protocols, emulate networks, generate media, and analyze quality of the media. Functional and load testing capabilities are also necessary for verifying both signaling and media handling. Monitoring certain statistics in the real time audio stream also improves performance analysis when used in conjunction with the audio quality testing discussed previously.

Another important test system function is the ability to actually emulate a SIP element and present that type of behavior to a gateway, along with the call itself. This would complete the test environment picture by representing the various element behaviors, which the Hammer test solution is able to provide.

## Summary

This paper has examined the devices involved in VoIP communications, the conditions that could affect their performance, and how to create a test environment that is comprehensive enough to reliably assess a VoIP system's performance under stress. In addition, we have discussed how the performance should be measured, using standards-based methodologies that recognize that the actual perception of users cannot be left out if a true evaluation is to be made.

There are many methods currently supported and under discussion by VoIP equipment and service providers for improving quality of service, and even providing customers with QoS guarantees.

These methods should help in improving how conversations in the VoIP environment sound, adding some consistency to quality performance. This is necessary to ensure the continued acceptance by enterprises and service providers of VoIP technology. In the final analysis, the success of the industry hinges on the positive perception of human beings using telephones.

To learn more about testing in the VoIP environment, contact your Empirix sales representative or your authorized reseller, call +1 781.266.3200 toll free in the U.S. at +1 866.EMPIRIX in the U.K. at +44 (0)1344 668080 Germany at +49 (0)8122 880 9790 Japan at +81 3 3791 2336 or email [info@empirix.com](mailto:info@empirix.com), visit us on the Web at [www.empirix.com](http://www.empirix.com)