

# Quality of Service Testing in the VoIP Environment

Whitepaper

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 about to be eclipsed 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.

Why is VoIP attracting so much attention? The advantages to a company adapting 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 bandwidth, which fills the available capacity much more effectively. Speech compression technologies used in preparing voice signals for transport on a packet network eliminate "silent space" and save 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. The advantage currently attracting the most attention is the ability to bypass the PSTN and its toll charges, and applications made possible by merging voice with the Internet will drive the level of interest even higher.

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 widespread acceptance of VoIP as a mainstream business tool.

In the brief history of IP telephony, voice connections over the Internet have gotten 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 nuance, inflection 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 or service provider for whom guaranteeing solid voice quality performance is a critical issue.

## VoIP Call Elements

A VoIP call can consist of several elements: endpoints, gateways, some type of packet switched network and sometimes a circuit switched network. Which of these elements are present in a particular call depends on what type of endpoints are being used. An endpoint in a VoIP scenario is either a PC with an Internet Telephone type application installed, an Internet Telephone itself or a regular telephone. In a conference situation all types of endpoints could be participating.



The difference in what elements 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 PC's or Internet Telephones, the speech encoding and packetizing functions are incorporated, 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 voice network and the packet network. IP Telephony Gateways are equipped with standard interfaces to the PSTN (analog, T1/E1) as well as interfaces to the packet network. The necessary encoding/decoding, compression/decompression and packetizing/depacketizing 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 algorithms used for the vocoder functions can differ from manufacturer to manufacturer, a possible performance variable.

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 interface to the PSTN is an impor-

tant 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.

**Latency:** Voice signals need to be processed for transport over a packet switched network. The necessary compression and packetizing (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 the device (DSP's, compression algorithms) 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 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 intranet, there is no control over the number of router-to-router legs, or hops, a packet takes. 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 tuned so that the process itself does not create excessive delay. 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. 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 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 can be done? Using the Hammer VoIP Test System as an example, we will examine how the requirements for testing can be addressed and reliable tests and measures 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 interfaces capable of sending both analog (the type of traffic typically offered by home telephones) and digital (T1, E1 and ISDN are typically presented by calls from switches) signals. The amount of load needed to exercise an 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.

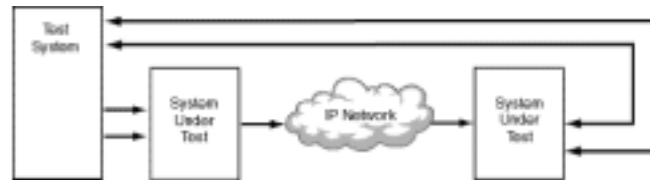


Figure 3

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 pre-defined test templates can be filled in with appropriate variables, such as calls/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 passed.

## DTMF Testing

Because users may be using VoIP services to access systems that require DTMF inputs (IVR's, 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.

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. Recognizing the necessity of incorporating the human factor but realizing the impracticality of always having a group available to evaluate transmissions, the ITU-T (International Telecommunication Union Standardization Sector) has developed methodology to automate the process. This methodology is presented in two documents, the P.800 and P.861 Recommendations, and the methodology they describe is the most well-established and fully realized currently existing in this area.

## Mean Opinion Scores

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 the range of sounds they encompass. The sentences are spoken by male and female adult voices, and child's voices, so that a wide assortment of human sounds are represented. The clips are then played to the judges in specially designed listening room, where noise and other environmental factors are controlled. From their ratings on the sound quality of the speech clips, MOS scores of one to five are derived, with five being highest quality.

## Perceptual Speech Quality Measurement

The P.861 Recommendation 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 Speech Quality Measurement (PSQM), in which the physical signals that constitute the original source speech clips and encoded speech (speech that has passed through a vocoder) are mapped onto psychophysical speech representations, in other words, how speech is perceived by the human 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) seeming 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. It is then possible to algorithmically correlate these results to benchmark MOS scores. These methods have been incorporated into the Hammer VoIP Test System, and 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 way. Having discussed how to stress a VoIP system with telephony traffic, what is this type of traffics' 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 worsened, the effect on voice quality would increase, with expected results as in Figure 4.

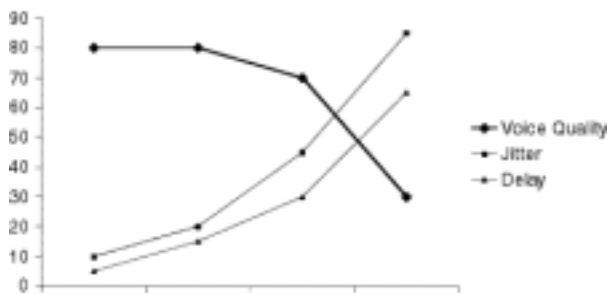


Figure 4

It would make sense to incorporate some thresholds-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 create 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. A company that is serious about providing test systems to the VoIP industry should have these capabilities in their plans.

## A Word About H.323

Another issue that has until recently presented a challenge to the expansion of the VoIP industry is lack of agreed upon standards. Different implementations by equipment manufacturers created interoperability problems. Now it appears that most have accepted the ITU's H.323 standard for multimedia communication between devices over IP networks.

H.323 will be the cornerstone for LAN based consumer, business entertainment and professional applications. It is in the plans of most VoIP equipment providers to support H.323 protocols. As this development proceeds, a VoIP testing system must keep pace by incorporating the ability to effectively test these protocols. For example, it would be useful during the testing of VoIP applications to monitor the H.323 frames to correlate telephony events with H.323 protocol events. Monitoring certain statistics in

the real time audio stream would improve performance analysis when used in conjunction with the audio quality testing discussed previously.

Another important test system function would be the ability to actually emulate an H.323 terminal and present that type of traffic to a gateway, along with the other types of traffic. This would complete the test environment picture by representing the PC-to-PC type transmission.

## Conclusion

This paper has examined the devices involved in VoIP transmissions, 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 that 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 under discussion by VoIP equipment and service providers for improving quality of service and even providing customers with QoS guarantees. If implemented, these methods should help in improving how conversations in the VoIP environment sound, adding some consistency to quality performance. This is necessary before general business acceptance of outside enterprise intranets will occur. In the final analysis, the success of the industry hinges on the positive perception of human beings using telephones.

**Empirix**

205 Lowell Street  
Wilmington, MA 01887  
United States  
T: 978-661-4800  
F: 978-988-0148

**Empirix Europe**

Building A, Trinity Court  
Wokingham Road  
Bracknell, Berkshire  
RG42 1PL  
United Kingdom  
P: 44-1344-66-80-80  
F: 44-1344-66-81-72

**Empirix Japan**

Yamate-K Building  
2-10-15 Nakameguro Meguro-ku  
Tokyo, Japan  
153-0061  
P: 81-3-3791-2336  
F: 81-3-3791-5353

**Empirix Singapore**

87 Science Park Drive #03-01  
The Science Hub  
Singapore Science Park 1  
Singapore, 118260  
P: 65-872-7239  
F: 65-872-5239

**Internet**

[www.empirix.com](http://www.empirix.com)

**E-mail**

[info@empirix.com](mailto:info@empirix.com)

All descriptions and specifications are intended for general information only and are subject to change without notice. Some mentioned features are optional. All names, products, services, trademarks or registered trademarks are the property of their respective organizations.



[www.empirix.com](http://www.empirix.com)