Voice delay on a telephony network is the time required for voice to be transmitted from a speaker’s mouth to a listener’s ear. Voice delay is one of the key factors in a customer’s perception of call quality. Excessive voice delay negatively impacts conversational quality by introducing disruptive pauses in a conversation. The International Telecommunications Union (ITU) recommendation G.114 includes results of subjective testing of the impacts of voice delay on a person’s perception of call quality. Mean Opinion Scores (MOS), on a scale from 1 to 5 and as established in ITU recommendations P.800 and P.830, are used to rate the quality of calls with varying degrees of voice delay. The results published in G.114 are shown in Figure 1.

![Figure 1: Mean Opinion Scores (MOS) for four delay conditions (source: ITU G.114.](image)

As shown in Figure 1, the rated quality of a call decreases significantly with increasing one-way delay.

It is important to note that voice delay impacts conversational quality, which is a round-trip phenomenon, rather than listening quality, which is a one-way phenomenon. Hence, a telephone user perceives round-trip voice delay, not one-way voice delay. That is, the delay for a speaker’s voice to reach a listener’s ear is perceptible to neither speaker nor listener. However, the delay between a speaker saying something, and then hearing the other person’s response, is perceptible.
An IP telephony network comprises many of the delays inherent in a circuit-switched network, such as the Public Switched Telephone Network (PSTN). These include physical transmission delays, as well as small delays related to PCM encoding, multiplexing, switching, and cross-connects. But an IP telephony network also comprises many processes that contribute to voice delay and are not present in a pure circuit-switched network. These processes include voice transcoding and decoding, packet capture, packet routing and queuing, and jitter buffers. As a result, an IP telephony network inherits most or all of the voice delay of a circuit-switched network, and adds additional processing delay.

However, customers’ thresholds for tolerating delay do not increase for an IP telephony network. To a customer, what is in-between the end points of a call does not matter; only the quality of the call matters. It is also important to note that, although the physical distance of a transmission path will impact delay, a customer’s threshold for tolerating delay remains unchanged. That is, a customer is no more likely to tolerate disruptive conversational delays on an international call than on a local call, despite knowing that the transmission distance is greater.

The excessive voice delays inherent in IP telephony networks, and the disruptive impact that such delay has on service quality, requires new levels of testing that were not required for the PSTN. Due to the disruptions in service that voice delay can introduce, and especially in consideration of the fiercely competitive environment of today’s voice service market, it is important to test IP telephony networks for delay in every phase of a network’s lifecycle. These include: designing networks in a lab, deploying and certifying them in the field, maintaining them and troubleshooting service problems, and optimizing them for both service quality and operational efficiency.

Testing voice delay is a three-part process:

1. Setting objectives for end user delay and establishing a network delay budget.
2. Testing end user delay to determine if a problem exists and the problem’s magnitude.
3. Troubleshooting a delay problem to determine its source so it can be fixed.

It is especially important to have an effective means for troubleshooting voice delay problems to quickly resolve customer-impacting problems in VoIP service. Network segmentation using the Agilent Voice Quality Tester (VQT), as described here, provides such a means.
Network Delay Budget

In setting objectives for end user delay, it is useful to refer to extensive work already performed by the International Telecommunications Union (ITU) in assessing the impact of various amounts of delay on user tolerance and perception of quality. According to ITU Recommendation G.114, a maximum of 150 milliseconds one-way delay (or 300 milliseconds round-trip delay) is acceptable for most user applications on a voice telephony network. This recommended maximum delay value is independent of transport technology or transmission distance (although greater values are allowed for satellite networks).

Based on the G.114 recommendation, a network operator can reasonably set 300 milliseconds as an objective for maximum round-trip voice delay. Then a network operator must determine a network delay budget that will allocate delay values to individual systems within the network. One example of a network delay budget is shown in Table 1.

<table>
<thead>
<tr>
<th>Process</th>
<th>Total Delay (msec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmission (4800 km) (^{(1)})</td>
<td>24</td>
</tr>
<tr>
<td>Circuit-switched network processing delays (^{(2)})</td>
<td>6</td>
</tr>
<tr>
<td>G.729a Encoding (^{(3)})</td>
<td>25</td>
</tr>
<tr>
<td>Packet processing (^{(4)})</td>
<td>3 – 30</td>
</tr>
<tr>
<td>Jitter buffer</td>
<td>50</td>
</tr>
<tr>
<td><strong>Total one-way delay</strong></td>
<td><strong>98 – 125</strong></td>
</tr>
<tr>
<td><strong>Total round-trip delay</strong></td>
<td><strong>196 – 250</strong></td>
</tr>
</tbody>
</table>

\(^{(1)}\) For transmission delay, one can generally apply 0.005 milliseconds per kilometer.

\(^{(2)}\) Includes PCM encoding, multiplexing, switching, and cross-connects; approximately 6 msec on a typical network connection, per ITU G.114

\(^{(3)}\) Varies depending on frames per packet; 25 msec for a two-frame per packet rate

\(^{(4)}\) Packet processing includes packet capture, routing, and queuing. This value can vary greatly, and depends on number of network devices, and network congestion.

Network delay budgets will vary due to differences in individual systems, such as codecs and jitter buffers. For example, different speech codecs introduce different amounts of delay. Table 2 shows the different types of delays attributed to different codecs. Even for a single codec, delay values can vary depending on the number of frames per packet. For example, G.729a allows one, two, or three frames per packet.

Table 1: Example of a network delay budget.
<table>
<thead>
<tr>
<th>Codec</th>
<th>Packet Rates (msec)</th>
<th>Frame size (msec)</th>
<th>Look-Ahead (msec)</th>
<th>Algorithmic Delay (msec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G.711</td>
<td>10, 20, 30</td>
<td>-</td>
<td>&lt;1</td>
<td>&lt;1</td>
</tr>
<tr>
<td>G.723.1</td>
<td>30</td>
<td>30</td>
<td>7.5</td>
<td>37.5</td>
</tr>
<tr>
<td>G.726</td>
<td>10, 20, 30</td>
<td>7.5</td>
<td>3</td>
<td>37.5</td>
</tr>
<tr>
<td>G.727</td>
<td>10, 20, 30</td>
<td>7.5</td>
<td>3</td>
<td>37.5</td>
</tr>
<tr>
<td>G.728</td>
<td>2.5</td>
<td>7.5</td>
<td>&lt;1</td>
<td>3 – 5</td>
</tr>
<tr>
<td>G.729a</td>
<td>10, 20, 30</td>
<td>10</td>
<td>5</td>
<td>15 – 35</td>
</tr>
</tbody>
</table>

Table 2: Typical Codec Delays

G.114 has more complete listings of delay values attributable to different network systems, and is a useful tool in establishing a network delay budget.

Testing End User Voice Delay

Measuring voice delay from an end user’s perspective is the only way to assess the impact that network delay will have on service quality. Much of the delay in an IP telephony call is introduced by VoIP gateway process such as codecs and jitter buffers. Therefore, it is necessary to measure delay in the voice domain to know what the true end user delay is on a network. Measurements only of IP packet transmission delays will most often omit over half of the end user delay.

Round-trip voice delay measurements are valuable because they more accurately characterize a user’s experience with regard to delay; that is, round-trip voice delay is the primary cause of disruptive conversational pauses. For example, if you add round-trip voice delay to the delay of the other speaker’s verbal response, you can assess the “pauses” in conversational exchanges. Also, round-trip delay measurements do not rely on a distributed time synchronization method, such as GPS or NTP, and are therefore more accurate.

Measuring voice delay requires an active measurement; there is no method available for passively measuring voice delay. The Agilent VQT uses a unique method known as normalized signal cross-correlation to provide the most accurate and robust delay measurement available. The VQT can be equipped with analog, T1, E1, Ethernet, IP phone handset, and PC soundcard telephony interfaces, enabling delay measurements to be made from most any end user or network access point.
Figure 2 illustrates one example of measuring end user delay with the VQT. This example is for a hybrid IP telephony network in which end users call from PSTN phones. A round-trip voice delay measurement between two distant sites on a network can be easily performed using two VQTs (VQT A and VQT Z), each of which can be remotely controlled from the same PC.

Measure round-trip delays between VQT A to VQT Z as follows:

1. Establish a telephone call between VQT A and VQT Z
2. Apply a port loopback on VQT Z
3. Measure round-trip delay from VQT A
4. Then apply a port loopback on VQT A
5. Measure round-trip delay from VQT Z
Important Note on Configuring a Round-trip Delay Measurement Between Two VQTs

When measuring round-trip delay between two VQTs (from VQT A to VQT Z), it is necessary to set the following two parameters to identical values:

- VQT Z: Loop Back tool: Advanced Configuration: “Port Loop Back Delay”

When measuring round-trip delay from a T1, E1, four-wire analog E&M, or 10/100 Ethernet port on a VQT, set both of these values to the minimum: 5 milliseconds for analog, and 0 milliseconds for T1/E1 and 10/100 (there is no “Port Loop Back Delay” parameter on T1/E1 and 10/100 interfaces).

When measuring round-trip delay from a two-wire analog FXO port on a VQT, the “Loopback Delay / Blind Window” parameter is used to bypass detection of the immediate echo that will be received from a hybrid wire junction (and would otherwise be mistaken for the round-trip test signal). Set this parameter to a value that will certainly exceed any immediate echo; 20 milliseconds in most cases. Then set the “Port Loop Back Delay” on the other VQT’s Loop Back tool to the same value.

When measuring round-trip delay from a two-wire analog FXO port to a T1/E1 port, use the “Network Simulator” tool on the T1/E1 VQT to apply a loopback, instead of the “Loop Back” tool. On the analog VQT from which delay will be measured, set the “Loopback Delay / Blind Window” parameter to a value that will exceed any immediate echo; 20 milliseconds in most cases. On the T1/E1 VQT, use the “Network Simulator” tool, select the same port and channel for “Audio Input” and “Audio Output”, and set “delay” to the same value as the “Loopback Delay / Blind Window” parameter.

When measuring round-trip delay from a two-wire analog FXO port to a 10/100 Ethernet port loopback, since no “Port Loop Back Delay” parameter exists on the 10/100 port, set the “Loopback Delay / Blind Window” parameter to a value that will exceed any immediate echo; 20 milliseconds in most cases. Then add this value to the VQT delay measurement result to get the actual round-trip delay.
Because the path-dependent processes (e.g., jitter buffers and encoding) of both directions will be included in each round-trip measurement, the results of round-trip measurements in each direction should be similar (this is not the case when measuring round-trip delay between a VQT analog/T1/E1 interface and a VQT 10/100 Ethernet interface).

Perform delay measurement trending to determine any temporal-based dependencies (voice delay may be greater during busy traffic periods), and to determine the consistency of voice delay in the network (a voice delay problem may be persistent, or it may be sporadic). Collecting delay trending data over a period of an entire day, or even several days, is helpful in baselining the performance of a network.

**Figure 3** illustrates another example of measuring end user voice delay. This example uses the VQT Phone Adapter to measure delay from the perspective of IP phone end users. Many of the techniques described in this application note can be used for the example in **Figure 3**, in which the VQT Phone Adapter is used to measure delay from at least one IP phone end point.

**Troubleshooting Networks for Voice Delay**

When a series of end user delay measurements indicate a problem, such as exceeding a delay budget, troubleshooting must commence with the objective of identifying the system(s) responsible for excessive delay. This can be done by segmenting the end-to-end call path, and performing segmented voice delay measurements. Segmentation is accomplished using the various test interfaces on the VQT.

The type of network segmentation that is effective depends much on the architecture of the network-under-test. Some, many, or all of the techniques described here may be useful depending on the network architecture.

Because round-trip delay measurements do not require GPS- or NTP-based time synchronization, the VQT can be quickly connected to different access points on the network, and delay measurements performed, with minimal configuration required.
General Techniques

Apply these general techniques when performing each network segmentation technique described shortly:

- Configure the round-trip delay measurement between two VQTs according to the note in the previous section.
- In each step, perform delay measurement trending to determine any temporal-based dependencies, and to determine if a delay problem is persistent or if it is sporadic.
- Record results for each delay measurement in a table for easy comparison and analysis.
- For delay measurements made at each network segment, determine if measurement results agree with what is attributed to the network segment in a network delay budget.

Network Segmentation Techniques

*Figure 4* continues with our example from *Figure 2*. Delay is first measured between Network Site A and Network Site Z using VQT A and VQT Z. Then an intermediate site, Network Site B, is identified such that a call from A to B shares some of the same IP backbone as a call from A to Z.

Connect VQT B to an access point at Network Site B, and measure round-trip delays between VQT A and VQT B as follows:

1. Establish a telephone call between VQT A and VQT B
2. Apply a port loopback on VQT B
3. Measure round-trip delay from VQT A.
4. Then apply a port loopback on VQT A.
5. Measure round-trip delay from VQT B.
An excessive delay between two specific sites (e.g., Network Sites A and Z) may be indicative of an underlying problem that affects other sites. The technique shown in Figure 4 will determine the extent of delay problems on the network. This also illustrates a first level of network segmentation, in which the network is segmented into geographic end points. This will help in localizing the source(s) of excessive delay so that proceeding steps of segmentation can be more precisely directed.

While VQT B is connected to an access point at Network Site B, measure also delay from VQT B to VQT Z, as shown in Figure 5. This technique illustrated in figure 4 and 5 geographically segments the network into two parts to help localize the source of excessive delay.

Perform such geographic segmentation for different sites on the network as makes sense, which depends on the IP backbone network architecture.
Before we proceed with further segmentation, refer to **Figure 6** for an alternative method for measuring end user delay. The VQT can be connected to the Media Gateways directly, or via digital cross-connect systems. This enables delay measurements from the perspective of customers that are other carriers, and isolates the end-point delay measurements to the IP telephony network, bypassing any front-end circuit-switched networks.

All proceeding techniques for network segmentation can be performed by connecting the VQT directly to the Media Gateway as illustrated in **Figure 6**, or by connecting the VQT to a circuit-switched access network, as is generally illustrated.

A second level of network segmentation begins with the illustration in **Figure 7**. With this technique, the delay attributed to any front-end networks, such as the PSTN used to access the IP telephony network, can be measured.

Place a call from VQT A to itself, using two analog lines or two T1/E1 channels into the circuit-switched network and the Media Gateway. At the Media Gateway, the call is “switched” using a time-division multiplex (TDM) hairpin at the DS0 level, so that the call is routed back to the front-end circuit-switched network with no VoIP processing, and is answered by VQT A. A delay measurement will then include only the front-end circuit-switched portion of the network.
Repeat this procedure at Network Site Z, using VQT Z and a TDM hairpin route at a Media Gateway at Network Site Z.

This method can be considered the complement of the method illustrated in Figure 6. In Figure 6, the IP telephony portion of the call is tested. In Figure 7, the front-end and tail-circuit portions of the call are tested. Together, they add up to the complete call illustrated in Figure 2.

In Figure 8, use VQT B to segment the network to the Media Gateway at Network Site A. VQT B terminates a call from VQT A at the first IP network component, typically an Ethernet switch. Much of the delay of an IP telephony call is due to the Media Gateway processes. The technique illustrated in Figure 8 is valuable for segmenting a delay measurement to a single Media Gateway. Delays introduced by the front-end switched network and the Ethernet switch will be small.

Connect the VQT B 10/100 Ethernet port to a 10/100 Ethernet interface at Network Site A, and measure round-trip delays between VQT A and VQT B as follows:

1. Configure VQT B port setup for SIP or H.323, whichever is supported by the network.
2. Establish a telephone call between VQT A and VQT B
3. Apply a port loopback on VQT B
4. Measure round-trip delay from VQT A.
5. Then apply a port loopback on VQT A
6. Measure round-trip delay from VQT B

It is particularly valuable to measure round-trip delay (rather than one-way delay) with this technique. The VQT 10/100 Ethernet interface applies a loopback within the RTP/UDP/IP stack, with no internal gateway processing and associated processing delays injected by the VQT. Thus, VQT A’s delay measurement result reflects only that delay added by the network between the VQT ports, and particularly by the Media Gateway and Ethernet Switch.
It is important to note that, in Figure 8, a round-trip delay measurement from VQT B will not produce the same result as a round-trip delay measurement from VQT A. This is because the round-trip delay from VQT B will include two gateway encoding/decoding processes and two jitter buffers: one of each in the Media Gateway and one of each in VQT B. The round-trip delay from VQT A will include only one gateway encoding/decoding process and one jitter buffer: in the Media Gateway. VQT B will apply a loopback in the RTP/UDP/IP stack, without any gateway processing.

While VQT B is connected to an IP access point at Network Site A, measure also delay from VQT Z to VQT B, as shown in Figure 9. Unless network delay is sporadic in time, the delay measurement results from VQT A in Figure 8 and VQT Z in Figure 9 should add up to approximately the same as the delay measurement result from Figure 2. That is, round-trip delay from VQT A in Figure 8, added to round-trip delay from VQT Z in Figure 9, should be approximately the same as round-trip delay from VQT A in Figure 2. This may not be reflected in actual delay measurement results if network delay is sporadic and inconsistent.
Figure 10 illustrates an alternative technique to that shown in Figure 8, for segmenting a delay measurement to a Media Gateway at one site. This is useful if the VQT 10/100 Ethernet interface cannot be connected to the Ethernet switch, or if SIP or H.323 is not supported on the network. This is also useful for measuring any impacts on delay attributable to specific switching or routing functions of the data node (Ethernet switch) behind the Media Gateways, which may produce different results in Figure 10 from results in Figure 8.

In Figure 10, connect another analog, T1, or E1 port on VQT A to the front-end circuit-switched network accessing Network Site A. Place a call from VQT A to itself such that the call must be routed across one Media Gateway, onto the VoIP network, and then terminated to the same front-end circuit-switched network via the same or another Media Gateway.

When examining the call paths in Figure 8 and Figure 10, it is apparent that they are very similar. One-way delay measurements in Figure 10 should yield similar results to round-trip delay measurements from VQT A in Figure 8. The test setup in Figure 10 may be difficult to perform depending on how local call routing is configured in the IP telephony network.

Network segmentation proceeds in Figure 11. The 10/100 port of VQT B is now connected to the next data network component, such as an IP router. Measure round-trip delays between VQT A and VQT B using the same procedure as described in reference to Figure 8. Measure round-trip delays from both VQT A and VQT B. As described in reference to Figure 8, a round-trip delay measurement from VQT B will not produce the same result as a round-trip delay measurement from VQT A.

Any difference in delay results between Figure 8 and Figure 11 may be attributable to processing within the Ethernet switch or IP router, perhaps due to congestion.
While VQT B is connected to an IP router at Network Site A, also measure delay from VQT Z to VQT B, as shown in Figure 12. Unless network delay is sporadic in time, the delay measurement results from VQT A in Figure 11 and from VQT Z in Figure 12 should add up to approximately the same as the delay measurement result from Figure 2.

Next, the network segmentation illustrated in Figures 8-12 at Network Site A, is performed at Network Site Z. In Figure 13, use VQT B to segment the network to the Media Gateway at Network Site Z. VQT B terminates a call from VQT Z at the first IP network component, typically an Ethernet switch.

Connect the VQT B 10/100 Ethernet port to a 10/100 Ethernet interface at Network Site Z, and measure round-trip delays between VQT Z and VQT B as follows:

1. Configure VQT B port setup for SIP or H.323, whichever is supported by the network
2. Establish a telephone call between VQT Z and VQT B
3. Apply a port loopback on VQT B
4. Measure round-trip delay from VQT Z
5. Then apply a port loopback on VQT Z
6. Measure round-trip delay from VQT B
As described in reference to **Figure 8**, a round-trip delay measurement from VQT B in figure 13 will not produce the same result as a round-trip delay measurement from VQT Z.

While VQT B is connected to an IP access point at Network Site Z, measure also delay from VQT A to VQT B, as shown in **Figure 14**.

Network segmentation at Network Site Z proceeds in **Figure 15**. Connect the 10/100 Ethernet port of VQT B to the next data network component, such as an IP router. Measure round-trip delays between VQT Z and VQT B using the same procedure as described in reference to **Figure 13**. Measure round-trip delays from both VQT Z and VQT B. As described in reference to **Figure 8**, a round-trip delay measurement from VQT B will not produce the same result as a round-trip delay measurement from VQT Z.

Any difference in delay between **Figure 13** and **Figure 15** may be attributable to processing within the Ethernet switch or IP router, perhaps due to congestion.
While VQT B is connected to an IP router at Network Site Z, also measure delay from VQT A to VQT B, as shown in Figure 16.

Many other network segmentation techniques can be utilized. In Figure 17, for example, an IP network segment is tested using two VQTs with 10/100 Ethernet interfaces. The delay measurement result will not represent only the IP network delay, but will include gateway process delays (e.g., codecs and jitter buffers) that are embedded in each VQT. The benefit of this measurement technique is that a consistent and known delay value within the VQT will be applied to each measurement, and therefore the differences in measurement results (e.g., between results from Figure 17 and results from measuring into the IP routers at each network site) can be attributed to the IP network.
Another technique, known as “swap and test”, can be used to further isolate individual systems like codecs, for delay measurements and impact assessments. This is described as follows:

Perform several iterative voice delay measurements using the same VQT end points. For each measurement (or measurement trending run), change one system at a time and measure the impact each system has on delay. For example, measure voice delay using G.729 encoding. Then, with all other network systems the same, change to G.711 and measure voice delay to determine impact that G.729 voice compression has on delay. Then change back to the original codec (G.729), but reconfigure the jitter buffer on a media gateway to lower the maximum value. Measure delay with a smaller jitter buffer to determine its impact. Reconfigurations in the IP network routing can also help isolate the impacts that individual systems have on delay.

**Summary**

Once delay trending measurement results have been recorded for each segment tested, then comparisons can be made between the network delay budget and actual network delay. Although some individual systems (e.g., voice codec) cannot be completely isolated via segmentation for an exact delay measurement of that system and comparison with the budget, these techniques will provide good insight into node-level or segment-level delays and will help isolate the source of excessive delay.
By internet, phone or fax, get assistance with all your Test and Measurement needs.

Online assistance:
http://www.agilent.com/find/assist

United States:
(Tel) 1 800 452 4844

Canada:
(Tel) 1 877 894 4414
(Fax) (805) 282 6495

China:
(Tel) 800-810-0189
(Fax) 1-0800-650-0121

Europe:
(Tel) (31 20) 547 2323
(Fax) (31 20) 547 2390

Japan:
(Tel) (81) 426 56 7832
(Fax) (81) 426 56 7840

Korea:
(Tel) (82-2) 2004-5004
(Fax) (82-2) 2004-5115

Latin America:
(Tel) (305) 269 7500
(Fax) (305) 269 7599

Taiwan:
(Tel) 080-004-7866
(Fax) (886-2) 2545-6723

Other Asia Pacific Countries:
(Tel) (65) 375-8100
(Fax) (65) 836-0252

Product specifications and descriptions in this document subject to change without notice.

©Agilent Technologies, Inc. 2000-2002
Printed in U.S.A. April 25, 2002

Use this link to go directly to our network troubleshooting solutions:
http://www.agilent.com/comms/onenetworks