A White Paper on Sage's PVIT

Renshou Dai*

December 29, 2000

1 Introduction

PVIT stands for Packet-Voice-Impairments-Test or Packet-Voice-Integrity-Test. It is one of a series of VQT (Voice-Quality-Test) tests designed by Sage to specifically address the next generation VoP (Voice-over-Packet) applications. Other related VQT tests available from Sage are PSQM [1] and Echo Sounder [2].

By VoP we include all the three dominant technologies: Voice over Frame Relay (VoFR), Voice over IP (VoIP) and Voice over ATM (VoATM). One may hear other "buzz" words such as voice over DSL, voice over Cable, and voice over LAN Ethernet etc. But DSL, Cable and Ethernet are simply different physical access or transmission media. The underlying technologies remain to be Frame Relay, IP or ATM.

The beauty of VoP lies on its efficient use of bandwidth. Unlike circuit-switched PSTN network, VoP employs packet or cell switching that performs statistical multiplexing. Bandwidth is dynamically allocated to various links based on their transmission activity. Along with bandwidth-saving techniques such as voice compression, silence suppression and jitter buffering, VoP offers more capacity that is normally impossible for PSTN.

However, a packet network was designed to carry data, not voice. Data traffic are typically non-real-time applications (file transfer and email, for example). The traffic tends to be bursty but in-sensitive to delay. Voice traffic, on the other hand, is a real-time "interactive" application that requires low delay, low jitter, and constant, albeit small, bandwidth. In data traffic, a lost packet can be re-transmitted. In voice traffic, a long-delayed packet has to be dropped, resulting in packet loss.

So, to some extent, VoP also stands for "Voice over Problematic Network". The in-compatibility between voice traffic and packet network means VoP application has to resolve serious Quality-of-Service (QoS) issues. QoS includes two broad categories: service quality and voice quality. Service quality covers such issues as the availability and reliability of the service, call connection time, call drop and advanced service features etc. The end-terminal voice quality is affected by delay, echoes, voice compression and transcoding, packet losses, voice clipping, silence noise, jitter (delay variation) and loudness level etc. The "art" of maintaining a high-quality VoP service is to maintain the delicate balance between the conflicting needs for short delay and low packet loss, the conflicting needs for high-fidelity (non-compressed) voice and band-width saving, and the conflicting needs for high-capacity heterogeneous service and the properly-prioritized traffic management.

Sage offers a series of advanced DSP-intensive tests that help quantify the QoS issues of a VoP application. More specifically, Sage's PSQM [1] and SMOS (Sage MOS) tests measure the overall voice quality in terms of Mean-Opinion-Score (MOS), round-trip delay and voice level change

*Sage Instruments, 240 Airport Blvd., Freedom, CA 95019

(loudness level). Sage's Echo Sounder [2], when combined with future Echo Generator, provides comprehensive information on the network echoes (their delays and levels) and the performance of echo cancelers. This PVIT test provides detailed diagnostic information on the packet network impairments such as packet loss, jitter, voice clipping and excessive comfort noise that impact voice quality. The primary challenge of maintaining a VoP service is to reduce the dynamic impairments such as packet/cell/frame loss. With PVIT, one can optimize the provisioning, configuration and setting of the network, gateway and IAD devices in terms of jitter buffering depth, packet size, traffic priority and choice of vocoders, VADs and Echo Cancelers etc. With PVIT's long-term monitoring capability, one can collect valuable statistical information on the network's dynamics.

2 Why PVIT?

PVIT continues Sage's tradition on telephony transmission impairment measurements. For the PSTN network, the network impairments are commonly measured as:

- Attenuation distortion;
- Envelop-delay distortion;
- Noise or Signal-to-Noise-Ratio;
- Nonlinear distortion measured as IMD (Intermodulation-Distortion);
- Delay;
- Echoes.

For a VoP application, the transmission impairments now become:

- Packet loss;
- Jitter/delay variation/packet slip;
- Incorrect comfort noise level;
- Voice compression and transcoding degradation;
- Long delay and improper loudness level;
- Long-delayed echoes.

PVIT measures such packet network impairments as packet loss, voice clipping, jitter and comfort noise level. Unlike other impairments such as delay, echoes and lossy voice compression/transcoding which are static in nature and are not necessarily unique to VoP (PSTN also has these problems), the packet loss, voice clipping and jitter type of impairments are unique to VoP and are dynamic in nature. By using PVIT to objectively monitor these dynamic impairments at a regular basis, one can assure the consistence of QoS (Quality-of-Service).

Unlike most other packet network impairment testers, Sage's PVIT measures packet network impairments from a telephony audio port (2-wire or 4-wire analog loop or T1/E1). By measuring from the telephony port, PVIT has the following advantages:

Technology independence: no matter what technology is used to transport voice (VoIP, VoATM or VoFR etc), PVIT works without the need to change one line of firmware.

Protocol independence: no matter what protocols are used for call set up and signaling (SIP, H.323 and MGCP etc.) and no matter what protocol is used for transporting the voice data (RTP or UDP etc.), PVIT works in the same way.

Direct correlation to voice quality: PVIT measures such impairments as directly perceived by human listeners. The packet network itself is always "jittery" and "chaotic". The packet network impairments may or may not directly propagate into the final audio signal side, depending on how these impairments are handled by the voice gateway or IAD. Therefore, an impairment (jitter, for example) observed at the packet network side bears little relevance to the final voice quality. By measuring from the voice terminal, PVIT measures only those impairments that actually affect voice quality.

The packet loss measured by PVIT is directly related to speech intelligibility. More packet loss means less intelligibility. The packet slip (delay variation or jitter) measured by PVIT quantifies the "jerking" or "gapping" effect due to the deletion or insertion of packet. Voice clipping measures the leading edge signal clipping caused by VADs (Voice-Activity-Detector). Excessive voice clipping means the first syllable of each sentence or word is muted or degraded. This can be as annoying as plain packet loss, if not more. The silence noise measurement provides level information (through C-message or Psophemetric filter) on the silence comfort noise generated by the CNG (Comfort-Noise-Generator). If the noise level is too high, the circuit will sound too noisy. If the noise level is too low, the circuit will sound "dead". More detailed definition of each of the measurement is provided in the next section.

PVIT is designed to measure only those impairments associated with the packet network activities (i.e., the corruption, replacing, deletion and insertion of packets). The PVIT test signal and receiver algorithm have robust discriminative capability against the non-packet-network impairments such as severe voice compression by vocoders and severe analog-type of distortions (attenuation distortion, nonlinear distortions and noise etc), so that these non-packet-network impairments will never trigger false alarms. The PVIT results reporting also takes a conservative approach, meaning that when PVIT reports a packet loss or a packet slip (jitter), then, a packet of signal must indeed be corrupted/replaced or deleted/inserted at the packet network side. PVIT may miss some minor impairments or those that occur during the silence period.

The PVIT test signal was designed to meet the following requirements:

Voice-like characteristics: the test signal must have a spectral and envelop characteristics resembling those of a real speech signal with proper "silence" period so that the signal presents a fair stress to vocoders, VADs and even echo cancelers.

Robustness: the test signal must be robust enough to withstand the potentially severe distortions inflicted by voice compressions (vocoders) and analog transmissions.

Intelligence: the test signal must have built-in intelligence so that it can record the packet network impairments and let the receiver algorithm detect and measure those impairments.

Non-repetitiveness: the details of each frame of signal must be different from the adjacent frames to maintain "sharp" discrimination capability.

3 Measurement Definitions and Specifications

The measurements reported by PVIT can be classified into two categories: the primary information and the secondary derivative information. The primary information are those associated with each impairment event. Each impairment event is defined (and reported) through its following attributes:

Nature of the event: There are five different impairment events being reported. These events are: packet loss, signal-lost, packet-slip, voice clipping and noise hit.

Size of the event: Except for noise-hit, the rest of the impairment events are quantified through their durations (sizes) in milli-seconds. Noise hit is quantified by the noise level in dBrnC or dBm.

Time stamp: The time when the event actually occurred. The time stamp is relative to the moment when the test was actually started.

The secondary information are derived from the above primary sources. More specifically, the following derivative information are reported by PVIT:

Counter of each impairment: Associated the five impairments, there are five counters, each indicating how many times a specific impairment has occurred.

Accumulated results: The accumulated positive packet slip (positive jitter caused by packet deletion) and the accumulated negative packet slip (negative jitter caused by packet insertion) are reported. The "algebraic" sum of the two accumulated slips is reported as "NET PACKET SLIP". The accumulated packet slips indicate the total amount of "jerking" and "gapping" that have affected voice quality. The "NET PACKET SLIP" indicates the final net delay change between the start of the test and the end of the test.

Averaged results: The average size (duration) of voice clipping and the average comfort noise level are both reported. The average voice clipping size indicates the performance of a VAD, whereas the average comfort noise level indicates the performance of a CNG. A high averaged voice clipping indicates that the VAD is causing serious voice quality degradation. An improper average noise level (too high or too low) indicates the mal-function of a CNG.

Normalized percentage packet loss: The accumulated packet losses and voice clippins is normalized by the total test duration and expressed as a percentage. This percentage packet loss (5%, for example) is a close estimate of the percentage packet loss that actually occurred in the packet network. This single derivative information might be the most useful one for a gross estimation of QoS.

Table 1 is a summary of the measurement specifications.

In the following sub-sections, we describe in detail each of the primary measurement. In the descriptions, we will use a real speech signal to illustrate the definition of each impairment. But keep in mind that the PVIT test itself does not use real speech signal. It uses its own unique "intelligent" signal that has similar spectral and energy-envelop characteristics as real speech signal.

3.1 Packet loss

Inside a packet network, a packet loss can occur due to the following reasons:

Packet arrives too late: For whatever reason (most likely, traffic congestion), if a packet arrives too late (later than a jitter buffer can tolerate, including the scenario of out-of-order-packets), this packet will be treated as a lost packet.

Packet misrouted: If a packet is mis-routed to nowhere and never arrives at its destination, it will of course be treated as a lost packet.

Measurement	Range	Accuracy & Precision
Packet Slip (jitter)	3 to 150 ms	± 1 ms
Packet Loss	6 to 642 ms	± 3 ms
Voice Clipping	6 to 22 ms	± 3 ms
Noise level	-80 to -10 dBm	± 1 dB
Percentage Packet Loss	0% to 50%	90% to 100% of actual packet loss

Table 1: Summary of key measurement specifications. The measurement range means the maximum range that can be measured within each 1-second measurement aperture. The total accumulated ranges are practically unbounded (only limited by the UI display on the actual test instrument). Not shown in the table is the time stamp specification for each primary impairment event. The time stamp for packet loss and voice clipping events is accurate to within ± 4 ms. The time stamp for packet slip (jitter) event is accurate to within ± 500 ms. It's practically impossible to pin-point the exact location of jitter when it occurs in the silence period. The counter of each event and the averaged results are of course 100% accurate.

Errored-packet: Under certain harsh transmission environment (wireless or xDSL, for example), if the bit errors exceed the correctable amount inside a packet, this packet will be treated as a lost packet.

In all of the above scenarios, the audio playout unit inside a gateway or IAD will either replay the previous packet (to minimize perceptible speech degradation) or play out a packet of "nonsense" signal (comfort noise or silence, for example) in place of the lost packet. Either way, at the telephony audio signal side, this packet replacement (due to packet loss) will cause a contiguous block of signal being corrupted. PVIT detects this type of signal corruption and measures it as a packet loss event. Figure 1 shows the scenarios of packet loss. Excessive packet loss degrades voice intelligibility.

3.2 Voice Clipping

The intention of voice clipping measurement is to quantify the voice-quality-degradation caused by VADs (Voice-Activity-Detectors). VADs help reduce bandwidth requirement through the silence suppression scheme. An overly aggressive VAD, however, can cause the leading or trailing edges of an active signal burst being clipped. Figure 2 shows the scenarios of voice clipping. As shown in Figure 2, if the leading edge of an active signal burst is clipped for less than 22 ms, this corruption will be measured by PVIT as a voice clipping event. But if the clipping lasts longer than 22 ms, it will be reported as a packet loss event, instead. The ambiguity between voice clipping and packet loss is impossible to resolve at the audio signal side. The 20 ms threshold is arbitrary, and subject to improvement in future based on field application feedback on the usefulness of voice clipping measurement. Also notice that, because of the ambiguity between voice clipping and packet loss, the percentage packet loss calculation also includes the accumulated amount of voice clipping, as these voice clippings may simply be caused by packet loss, not VADs. Some in-house testings also indicate that certain low-bit-rate (< 8 Kbps) vocoders also tend to cause some minor voice clippings (< 10ms) that can also be measured by PVIT. This measurement could be useful for some people, but may also be a nuisance to others. We are looking for feedbacks on this voice clipping measurement.

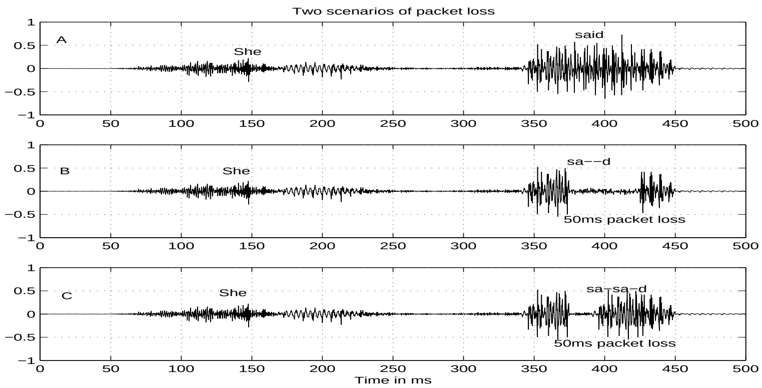

Figure 1: Hypothetical Scenarios of packet loss. A: A snapshot of male speech saying "She said". B: A 40 ms block of speech signal centered at 400 ms time mark (horizontal axis) is replaced by "comfort" noise due to packet loss. C: A 40 ms block of speech signal centered at 400 ms time mark is replaced by its previous 40 ms block of signal due to packet loss. Both cases of B and C will be correctly measured by PVIT as a packet loss event with duration (size) of 40 ms and time stamp around 400 ms.

3.3 Packet Slip/Jitter

The packet slip or jitter measured by PVIT at the telephony audio port is a measure of sudden delay variation. Figure 3 shows two scenarios of packet slip or jitter that can be measured by PVIT. As shown in Figure 3, in case B, a block of 50 ms signal is deleted, causing the relative delay between adjacent words be shortened by 50 ms. This will cause the "jerking" effect perceived by human listeners. PVIT will measure this effect as a positive jitter or packet slip of 50 ms.

In case C, a block of 50 ms signal is forcefully inserted, causing the relative delay between adjacent words be suddenly increased by 50 ms. This will cause the "gapping" effect perceived by human listeners. PVIT will measure this effect as a negative jitter or packet slip of 50 ms.

Notice that the packet slip or jitter measured by PVIT is not the packet network jitter. A packet network, by nature, is always jittery. This means the packets will not arrive at the receiving side at uniform rate. They may even arrive out of order. But at the telephony audio port, the voice signal must be played out at uniform constant rate. To resolve the conflict between a jittery packet network and a smooth PSTN network, a voice gateway typically uses a jitter buffer (or de-jitter buffer). A jitter buffer is commonly several times larger than the packet size, so that it can tolerate jittered packets or even the out-of-ordered packets.

So, theoretically, by increasing the jitter buffer size, the packet loss can be minimized to near zero. But a large jitter buffer will increase the delay, an "impairment" that a real-time interactive voice conversation can only tolerate within a limit (one-way delay less than 125 ms, for example).

To balance the conflicting need for short delay and minimal packet loss, typical voice gateways employ adaptive algorithms to dynamically adjust the jitter buffer size based on the packet network jitter severity. A sudden resizing of jitter buffer will cause either a packet be deleted (jitter buffer becomes smaller) or a packet be inserted (jitter buffer becomes bigger). This jitter buffer resizing will cause a sudden delay variation at the audio signal port, and PVIT will measure this delay variation as positive or negative packet slips.

Of course, if no jitter buffer is used, then the packet network jitters will directly propagate into the audio signal side and be measured by PVIT. But this scenario is less likely to happen in a real

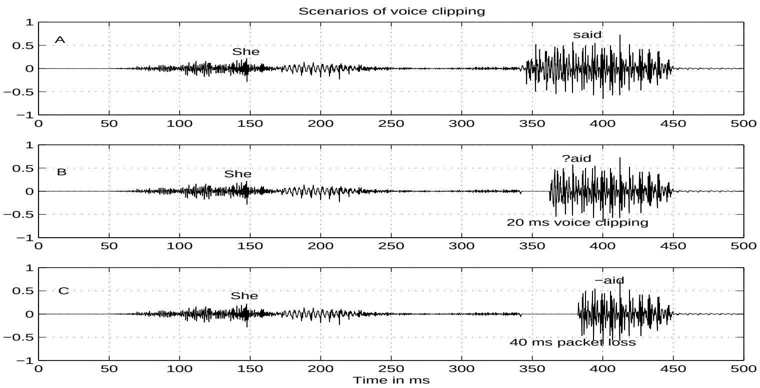

Figure 2: Hypothetical Scenarios of voice clipping. A: A snapshot of male speech saying "She said". B: A 20 ms block of speech signal at the leading edge of the active voice "said" centered at 350 ms time mark is clipped (replaced by silence or comfort noise). This will be measured by PVIT as a voice clipping event of duration (size) of 20 ms and time stamp around 350 ms. C: A 40 ms block of speech signal at the leading edge of the active voice "said" centered at 350 ms time mark is clipped (replaced by silence or comfort noise). This will be measured by PVIT as a packet loss event of duration (size) of 40 ms and time stamp around 350 ms. This shows the subtle ambiguity between packet loss and voice clipping. Inside the software, if the leading-edge clipping lasts longer than 22 ms, it will be reported as packet loss, instead of voice clipping.

VoP application.

Another likely cause of packet slip or jitter measured by PVIT is related to the sampling clock difference and the asynchronous nature of the VoP application. In the PSTN network, the codecs (where analog to PCM and PCM to analog conversions occur) inside a class-5 switch derive the 8000Hz sampling clock from the synchronous T1/E1 network. This automatically eliminates the possibility of having different conversion rates at two different class-5 switches. But now let's hypothesize an extreme VoP application scenario. Assume one customer makes a call through VoDSL to another customer connected to a PSTN network. Assume the IAD (Integrated-Access-Device at the VoDSL side) has highly skewed clock, and it samples the incoming voice signal at 8100Hz, instead of the exact 8000Hz. When the ATM cells (most VoDSL employs ATM switching) arrive at a voice gateway, the gateway will play out the cells into the PSTN network at the normal 8000Hz rate. Now there is a problem. For every second, the IAD sends in 8100 samples of signal, but the gateway can only finish playing 8000 of them. As time goes by, the left-over samples (100 for every second) will keep accumulating. Sooner or later, the gateway has to delete a few blocks of samples to keep up with the IAD. This will cause some surprising packet slips measured by PVIT. The 100Hz difference is of course an exaggeration. The real clock difference in a real world should be less than 1 Hz. But in a long-term testing (over than 1 hour, for example), this minute clock difference may still cause packet slip. This phenomenon shows another potential problem that a VoP application has to overcome.

Regardless of the causes, the packet slips measured by PVIT reflect the sudden delay variations that will cause the unpleasant "jerking" and "gapping" effect on human listeners. This type of impairment only occurs in an asynchronous packet network. It rarely occurs in the PSTN network. Theoretically, this problem can also occur in calls ordinating from or terminating at a wireless

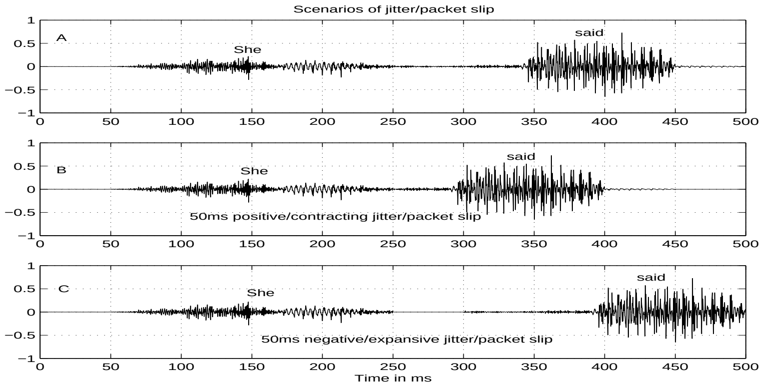

Figure 3: Hypothetical Scenarios of jitter or packet slip. A: A snapshot of male speech saying “She said”. B: A 50 ms block of signal centered at 300 ms time mark is annihilated (deleted), causing the relative delay between the word “She” and the word “said” to be shrunk by 50 ms. This will be measured by PVIT as a positive packet slip (jitter) of 50 ms with a time stamp around 300 ms. C: A 50 ms block of silence is “forcefully” inserted into the signal centered at 275 ms time mark. causing the relative delay between the word “She” and the word “said” to be increased by 50 ms. This will be measured by PVIT as a negative packet slip (jitter) of 50 ms with a time stamp around 275 ms.

mobile phone.

Ideally, in a VoP call, packet slip should only occur during the silent period to minimize the voice quality degradation. But in reality, of course, the packet slip may also occur in the middle of the active speech. No matter where it occurs, PVIT can measure them accurately and correctly. But notice that a packet slip occurring in the middle of the active speech also implies packet loss. To avoid redundant and sometimes even erroneous reporting, the packet slip (or jitter) event has the highest masking priority. This means, within any 1s measurement aperture, if a packet slip (jitter) is detected, then only this impairment is reported. The other likely impairments (such as packet loss and voice clipping) are ignored, as they might be a result of packet slipping.

3.4 Silence Noise and Noise Hit

The intention of silence noise measurement is to quantify the performance of CNG (Comfort-Noise-Generator) at the receiving side. The comfort noise level in the silence period should neither be too high (sounds too noisy) nor too low (sounds dead).

The PVIT test signal contains about 40% silence period. At the measuring side, these “silence” periods are located, and the comfort noise levels at those periods are then measured either through C-message (for North-America) or Psophemetric (the rest of the world) filters. If during any individual silence period, the noise level exceeds -45 dBm (45 dBrnC), a “Noise Hit” event will be reported by PVIT. PVIT does not report a “Too Quiet” event. But the continuously averaged noise level should indicate if the comfort noise level is too low. A proper comfort noise level should be around -65 dBm (25 dBrnC).

3.5 Signal Lost

The PVIT test signal contains about 60% active signal. The active signal level can be adjusted at the sending side (recommended to be between -15 dBm and -5 dBm, default is -10 dBm). At the measuring side, the active signal level within a 1s frame must be greater than -30 dBm to be declared as "active". Otherwise, this 1 second frame of signal will be declared as a "Signal Lost" event, and detailed impairment measurements will not be performed until the signal comes back again in the next frame. A "Signal Lost" event indicates either the sender has been turned off, or the call is disconnected (dropped) or the attenuation is too high that the sender level needs to be adjusted.

Notice that if the valid PVIT signal is replaced by a "garbage" signal with active level greater than -30 dBm, then PVIT will measure large packet loss, instead of declaring it as "Signal Lost".

4 How to Operate PVIT?

As implemented in Sage's 93x, PVIT is a manual test. This means the operator needs to "manually" establish a call connection between two 93x first (use the Dial/Ring function). After establishing the call, one 93x can be put in "SEND" mode, and the other in "MEASURE" mode.

The PVIT test is in option menu 36. Once in "SEND" mode, one can adjust the sending level of the test signal. The recommended level is -10 ± 5 dBm (default being -10 dBm). There is no need to change the level unless the insertion loss on a call is too high or the access device cannot accept high level of signal.

When in "MEASURE" mode, one can choose the test duration from 15 minutes, 1 hour, 24 hour to "CONTINUOUS" (forever). Of course, one can always manually stop the test anytime if you chose the "forever". After pressing the "START" button, the test now starts looking for the valid PVIT signal and synchronizes to it. During this brief period, the right corner will display "NO SIGNAL". Once valid signal has been detected and synchronization is finished, the right corner will start displaying a measurement clock (keeps ticking/incrementing in 1 second interval). There are several results screens:

The event screen: This screen displays the instantaneous event information (size or duration of the event and time stamp of the event) whenever an impairment event occurs. These events include packet loss, packet slip (jitter), voice clipping, signal lost and noise hit.

The counter screen: This screen displays the accumulated counters of all the impairment events.

Percentage packet loss screen: This screen displays the relative (normalized) percentage packet loss. This value is a close estimate of the percentage packet loss that actually occurred inside the packet network. Notice that the voice clipping effect is also included in this percentage packet loss calculation due to the ambiguity between voice clipping and packet loss.

Average voice clipping size screen: This screen displays the averaged voice clipping size (duration). The intention of this screen is to serve as an indicator of the "average" performance of VADs.

Average noise level screen: This screen displays the continuously averaged noise level. This information indicates the long-term performance of the CNG (Comfort-Noise-Generator).

Accumulated packet slips: the packet slips occupy a total of three screens. The total positive packet slip, the total negative packet slip and the algebraic sum of the positive and negative slips displayed as “NET of PACKET SLIP”. The total positive packet slip indicates the total “jerking” effect that has affected the voice quality. The total negative packet slip indicates the total “gapping” effect. The net of packet slip indicates the final delay difference between the start of the test and the end of the test. If the net is a positive value, it indicates the delay is shortened. If the value is negative, the delay is elongated.

Notice that the event screen only displays the most recently occurred impairment event. There is not a history or log screen that saves all the event information. To permanently save the individual event information, one must put the 93x in “PRINT” mode (under option 3). The individual event information will then be printed to a PC (with any serial communication software) through the back panel serial port.

When operating PVIT with 4-wire connection (T1/E1 or 4-wire analog), the “SEND” and “MEASURE” can be simultaneously performed on a single 93x unit. But of course, the other end that answers the call must provide a loop back.

PVIT can be tested through any networks with telephony audio ports. Each 93x needs to be connected to a telephony port. This telephony port can be a DS0 from a T1/E1, or a 2-wire or 4-wire analog loop. The analog loop can be a normal “wet” (with DC signaling) loop line, or a “dry” audio circuit. In the case of dry circuit, the noise level measurement could be offset by a constant due to the potential impedance mismatch. The ideal networks to be tested are those related to VoP (Voice-over-Packet) applications such as VoIP, VoATM, VoFR, VoDSL, VoLAN, VoCable etc. The wireless (cellular and PCS) network can also be tested with PVIT, as it involves dynamic frame erasure (packet loss) due to fading and poor RF coverage etc.

5 Where to Apply PVIT? PVIT Application Examples

There are practically unlimited number of ways of using PVIT to characterize the dynamic conditions of any voice networks. The following examples show a few typical application scenarios. But keep in mind that, these scenarios are just examples. They are by no means exhaustive or exclusive. You can easily find ways of applying PVIT that go far beyond the following examples.

5.1 Objective Measure of Voice Quality

All the information provided by PVIT (packet loss, packet slip (jitter), voice clipping and comfort noise level) have direct impact on the perceived voice quality. As mentioned before, excessive packet loss and voice clipping affect speech intelligibility; packet slip causes unnatural “jerking” or “gapping” effect (perceived sudden delay variations); improper silence noise level will either make the call sound too noisy or sound “dead” (too quiet).

Nowadays, there are quite a number of “popular” algorithms (PSQM, PAMS, PESQ and MNBs etc.) that claim to provide an objective MOS-equivalent (Mean-Opinion-Score) score for a voice call. But scientifically speaking, all these algorithms were designed to characterize the effect of static impairments introduced by low-bit-rate vocoders. These algorithms can not adequately account for the voice quality degradation caused by the dynamic impairments such as packet loss and jitter measured by PVIT. Other impairments such as long delay and echoes are not reflected by these algorithms at all.

A truly scientific approach to quantize the voice quality is to use PVIT to record all those dynamic impairments that affect voice quality, and find ways to minimize those impairments. The long delay and echoes can be measured through Sage’s Echo Sounder test [2].

While voice quality is a complex function of many parameters, one can certainly pick the percentage packet loss measured by PVIT as an indicator of overall voice quality. Figure 4 shows a hypothetical inverse relation between MOS and the percentage packet loss. As shown in Figure 4,

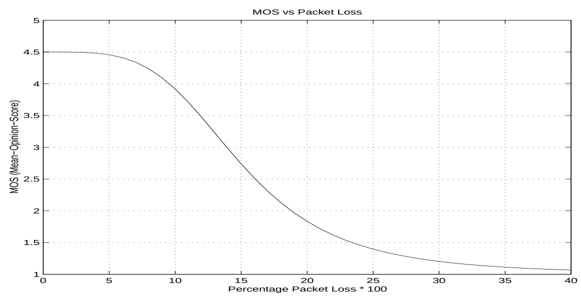

Figure 4: Hypothetical relation between MOS and percentage packet loss. Notice that the relation shown here is only hypothetical. The exact relation between MOS and packet loss requires extensive human listening tests. But there is no doubt that the exact relation will closely resemble the curve shown here. 64 Kbps PCM encoding is assumed for the curve.

in this hypothetical case, as packet loss exceeds 3%, the perceived MOS begins to decrease. As packet loss approaches 10%, the MOS drops below 4.0, the toll-quality threshold. As packet loss exceeds 18%, the MOS drops below the unacceptable threshold of 2.0. Although the curve shown in Figure 4 is hypothetical, it can certainly be used as a guidance to check the voice quality by measuring the percentage packet loss through PVIT.

5.2 Provision/Configuration of VoP Devices and Traffic Prioritization

A successful VoP application entails solid networking, optimal provisioning and configuration of voice and network devices, and proper traffic management (traffic prioritization). Take a voice gateway as an example, the following programmable parameters need to be optimally configured to achieve the best voice quality:

Jitter buffer depth: A small jitter buffer introduces shorter delay, but may cause more packet loss. Likewise, a large jitter buffer causes less packet loss, but will introduce longer delay. With PVIT, one can continuously monitor how jitter buffer size affects packet loss, and therefore, find the optimal (minimal) jitter buffer size that can satisfy the required packet loss. Figure 5 shows a hypothetical relation between packet loss and jitter buffer size. The optimal jitter buffer size in this hypothetical example is 60 ms if the packet loss needs to be less than 3%.

Dynamic vs static jitter buffer: Some voice gateways employ an adaptive algorithm to dynamically adjust the jitter buffer size to maintain both short delay and minimal packet loss. PVIT is a perfect tool to verify the performance of such algorithm. If the algorithm works perfectly, PVIT should observe very small amount of packet loss, but quite significant amount of packet slips (jitters caused by jitter buffer resizing). On the other hand, if the jitter buffer size is fixed (static), then PVIT should only observe packet loss, but no packet slips. Notice that excessive packet slips do cause unnatural “jerking” and “gapping” effect on the speech signal that could be more annoying than plain packet loss.

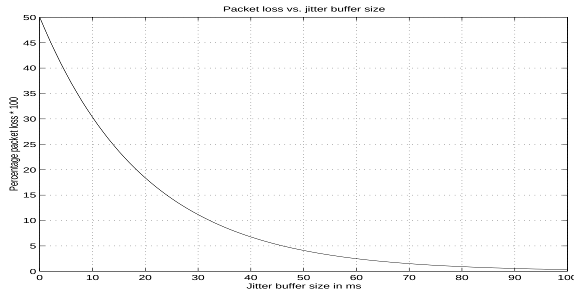

Figure 5: Hypothetical relation between packet loss and jitter buffer size. PVIT can be used to find such curve by continuously varying jitter buffer size and monitoring the resulting packet loss. With the curve shown here, for example, if the goal is to maintain a packet loss less than 3%, then the jitter buffer size needs to be set greater than 60 ms.

Packet size: the packet size has a more complex relation with the measured packet loss, which further justifies the use of PVIT. A small packet size is less efficient (it has more overhead) but introduces shorter delay. Small packet size can cause more traffic (requires more bandwidth), but may also cause less traffic collision (easier to be routed). A large packet, although efficient, introduces longer delay, and is more likely to collide with other traffic. Figure 6 shows a hypothetical relation between packet loss and packet size. If the goal is to minimize the packet loss, then apparently the optimal packet size is 30 ms.

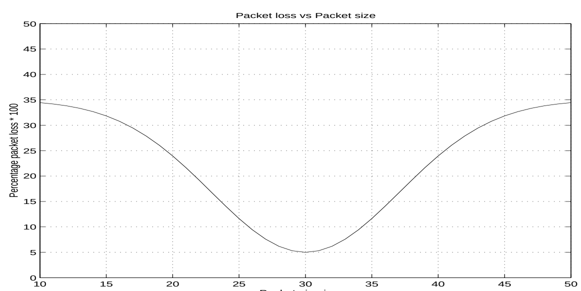

Figure 6: Hypothetical relation between packet loss and packet size. PVIT can be used to find such curve by continuously varying the packet size and monitoring the resulting packet loss. With the curve shown here, for example, if the goal is to achieve minimal packet loss, then the optimal packet size is apparently 30 ms.

Choice of vocoders and VADs: Certain vocoders and VADs can cause more voice clippings and too high or too low silence comfort noise. PVIT can be used to sort out the optimal vocoders and VADs if the goal is to achieve the minimal voice clipping and ideal comfort noise level.

Traffic management and traffic prioritization: The beauty of a packet network is to allow service providers to be able to "seemingly" offer service to "infinite" number of customers on a given network. That, of course, is an illusion. A given network with finite bandwidth can only carry finite amount of data traffic. No matter how smart the switching technology is, the number of customers that can be serviced with descent voice quality is finite. Figure 7 shows a hypothetical relation between packet loss and the number of subscribers on a given VoP network. PVIT can be used to monitor the packet loss condition as a result of traffic situation,

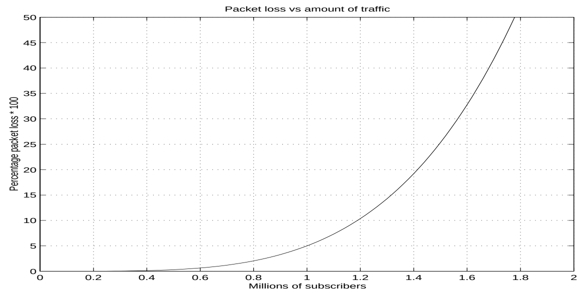

Figure 7: Hypothetical relation between packet loss and number of subscribers on a given VoP network. PVIT can be used to find such curve along with other bulk-call generator type of instruments. With the curve shown here, for example, if the goal is to maintain a packet loss less than 5%, then the number of subscribers should be limited to 1 million.

and obtain the type of curve as shown in Figure 7. If the network offers heterogeneous service to both voice and data, then the voice data need to be prioritized high enough such that the increase of data traffic does not cause packet loss on a voice call. PVIT, of course, can be used to verify if the prioritization scheme is working correctly.

5.3 Observe the Temporal Dynamics of a VoP Network

By the nature of packet network, the voice quality on a given voice channel depends on the overall traffic condition on the network. By the nature of human civilization, the overall traffic condition on a network depends on what time it is within a working day. If PVIT is set to perform a 24-hour long-term monitoring on a given voice call, the impairments measured by PVIT can reflect the network usage and traffic situation at various times of the day. Figure 8 shows a hypothetical packet loss situation at different times of a day that can be obtained through PVIT on a real network. With the kind of statistical information obtained through PVIT and shown in Figure 8,

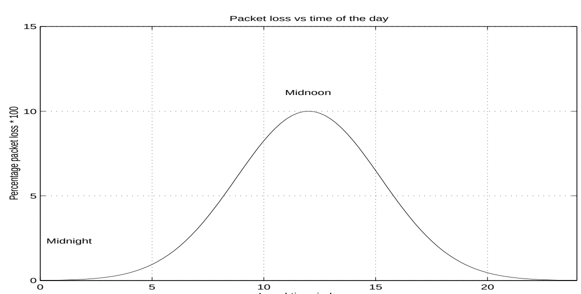

Figure 8: Hypothetical relation between packet loss and the time during a working day. PVIT can be used to find such curve as it can be set to do a long-term (24 hours, for example) test. The packet loss shown in the curve is the packet loss obtained within each 15 minute block. The curve shows that the packet loss peaks at mid-noon, as the network usage peaks at this time.

the service providers can then find ways to evenly distribute the traffic so as to maintain the best voice quality and highest capacity at any time of the day.

5.4 RF Coverage, Fading and Receiver Sensitivity

In the digital wireless telephony world (TDMA, CDMA and GSM etc), poor RF coverage and fading in certain areas can cause excessive bit errors on the receivers, which in turn will cause frame erasure on the audio signal, or even call drop. If the user can find ways to connect a Sage 93x or 923 to the audio port of a mobile phone, then PVIT can easily detect the frame erasure. Therefore, PVIT can be used to indirectly measure the RF coverage and fading situation. This, of course, not only applies to the outdoor wireless environment, it also applies to the indoor wireless communication.

When combined with a precise RF signal power meter and controller, PVIT can also be used to indirectly measure the sensitivity of a digital radio receiver. As the incoming signal power keeps dropping, the receiver may detect more bit errors, which in turn will cause more audio signal frame erasure that can be measured by PVIT.

References

[1] Renshou Dai, "A Technical White on Sage's PSQM Test," *Sage Document*, June 19, 1999.

[2] Renshou Dai, "A White Paper on Sage's Echo Sounder Test," *Sage Document*, April 14, 2000.