A White Paper on SMOS

Renshou Dai

November 15, 2001

1 Introduction

SMOS stands for Sage (Instruments) Mean Opinion Score. SMOS provides an automated, fast, convenient and accurate end-to-end assessment of voice quality for any VoX applications. SMOS follows the automated responder test format and employs robust in-band telemetry and synchronization with true real-time processing. The test can be conveniently used in both laboratory environment as well as across a real telephone network with literally the simple press of one button.

SMOS was developed largely out of Sage's prior experience with PSQM [1] and PVIT [2]. SMOS provides an accurate MOS score that truly matches human perception even in a live VoP network where certain impairments such as voice jitters (sudden delay variations or frame slips) and attenuation distortion may render other voice quality tests such as PSQM [1] inapplicable. SMOS contains a reliable Bark-domain [5] partial equalization along with asymmetric masking to properly account for attenuation distortion, and a robust de-jittering algorithm to remove and simultaneously measure any voice jitters (sudden delay variations) [2]. The psychoacoustic core is based on the work of *Wang et al* [4], *Zwicker et al* [5] and Sage's own internal research.

Besides the MOS number, SMOS also provides a set of other "orthogonal" measurements that are vitally important in determining the overall voice quality of a network, or trouble-shooting the configuration and traffic engineering of a problematic network. These measurements are orthogonal to MOS because they are not properly reflected in the MOS number, yet they are also important indicators of the overall QoS of the network. These other measurements are round-trip delay, codec type, effective bandwidth, voice-band gain, silence noise level, total amount of compressive jitters (positive frame slips or shortening of delays) and the total amount of expansive jitters/frame slips (lengthening of delays). In following sections, we describe each measurement in detail.

2 Summary of Measurements and Specifications

Table 1 summarizes the measurements performed by SMOS.

As shown in Table 1, the only other important measurement that is left out in this SMOS test is echo delay and echo level. Fortunately, Sage's Echo Sounder [3] measures echo delay and echo level from both 2-wire POTS and 4-wire digital interfaces. Specific packet network impairments such as packet loss, voice jitters (frame slips) and voice clippings can be measured through Sage's PVIT [2].

3 MOS measurement

SMOS measures an objective Mean-Opinion-Score between 1 and 5. 5 means perfect and 1 means the worst. For all practical measurements, the upper limit of MOS will be between 4.5 and 4.6.

Measurement type	Range	Precision
MOS	[1.00, 4.60]	±0.05
Delay	[0.0, 5000.0] ms	±0.125 ms
Codec type	[2.4,64]kbps codecs	PCM,ADPCMs,VCDs
Compressive jitter	[3,2000] ms	±1 ms
Expansive jitter	[-2000,-3] ms	±1 ms
Effective bandwidth	[0.00,1.00]	±0.01
Gain	[-80,20] dB	±1 dB
Silence noise	[0,90] dBrnC	±1 dB

Table 1: SMOS measurement types, ranges and precisions

A MOS number above 4.0 is considered to be toll quality. A MOS number between 3.0 to 4.0 is considered to be communication quality (intelligible but unnatural, or could be annoying and lack of speaker recognition etc). A MOS number below 3.0 is unacceptable for voice communication.

In a typical VoP (Voice-over-Packet) network, the measured MOS number largely reflects speech degradation caused by the following likely impairments:

1. Lossy voice coder compression.
2. Packet loss and voice clipping.
3. Voice jitters in active voice period.
4. Interference signal and noise.
5. Excessive attenuation distortion.

3.1 MOS numbers of known codecs

The psychoacoustic core used by SMOS is very accurate in quantifying the voice quality degradation caused by compressive codecs. By "accurate" we mean the measured MOS number is highly correlated with actual human perception. Table 2 lists the MOS numbers (measured by SMOS) of a set of commonly used codecs that are available to us. Notice that the MOS numbers in Table 2 were obtained when there were no other impairments except the lossy codec compression. In an actual VoX network, other impairments such as packet loss and analog distortion etc will cause the MOS numbers to be lower than those in Table 2 for any given codec type.

3.2 MOS numbers at various packet loss rate

Unlike lossy codec compression, which is a static voice degradation that can be accurately measured by the psychoacoustic core of SMOS, the packet loss effect on voice quality is more complex for the following reasons:

- Packet loss are dynamic and occurs "randomly" in a real network. Not every lost packet is equal. The packet losses that occur within the active period of the voice signal are far more perceptible than those occurring in silence period. So there is no simple relation between the percentage packet loss and the MOS number. The packet loss distribution, size and the activity of the speech signal itself all determine the actual human perception.

Codec type	MOS by SMOS test
G.711 PCM@64kbps	4.52
G.711 PCM robbed-bits	4.49
G.711 PCM@56kbps	4.44
G.726 ADPCM@40kbps	4.34
G.726 ADPCM@32kbps	4.20
G.726 ADPCM@24kbps	3.98
G.726 ADPCM@16kbps	3.38
G.729E@12kbps	4.19
G.729@8kbps	4.07
G.723.1@6.3kbps	4.00
G.723.1@5.3kbps	3.93
Cell-phone VSELP@8kbps	3.85
Cell-phone EFR-ACELP@7.4kbps	4.05
Cell-phone QCELP@13kbps	4.10
Cell-phone QCELP@8kbps	3.90

Table 2: MOS readings measured by SMOS test for some commonly used codecs.

- Codec type dependency. The perceived speech degradation caused by packet loss also depends on the specific codec type. The G.711 PCM codec, for example, has no sample-to-sample or packet-to-packet dependency. A 10 ms packet loss causes exactly 10ms speech degradation at the audio signal side. But for a G.729 vocoder, for example, there is high frame-to-frame dependency. A loss of 10ms packet not only causes the degradation on this voice frame *per ce*, it will also affect the adjacent voice frames so that the actual audio signal degradation could be as large as 20 ms. This is also the reason why we recommend using Sage's PVIT [2] to measure the actual audio signal side packet loss, instead of the packet side statistics. In short, a 5% packet loss in G.711 codec is not equal to 5% packet loss in G.729 codec in terms of MOS number degradation.

- Packet loss concealment. Naturally, the perceived voice quality degradation due to packet loss also depends on how the lost packets are "concealed" or handled. A lost packet, for example, can be replaced by a frame of silence, or by the previous voice frame, or replaced by the interpolation between adjacent frames, or even "ignored" (tossed out) to cause a frame slip (voice jitter or sudden delay variation). In each case, the perceived or measured MOS numbers will be different for the same amount of packet loss with the same codec type (5% packet loss with G.711 coding, for instance).

- Recency issue. Human brain perceives the same packet loss (or any other impairments) differently depending on when they occur during a call. A packet loss that occurred 10 minutes ago seems less "annoying" than a packet loss that just occurred a few seconds ago. A plain psychoacoustic model, of course, can not take this phenomenon into account. There is no real scientific value either in taking advantage of this recency phenomenon when designing traffic engineering.

Despite the complexity, we still list the measured MOS numbers with various amount of packet loss as a reference. Notice that we did not control the packet to occur only in the active period

or silence period. The tests were conducted through *Telogy's Golden Gateway* that can simulate packet network impairments. A lost packet is replaced by its previous packet.

Table 3 lists the MOS numbers (measured through SMOS) of G.711 codec with various amount of packet loss. All test durations are 16 seconds. In real network, in order to "average" out the dynamics of packet loss, we recommend using longer test duration if the intent is to obtain a stable MOS reading in a highly impaired network.

Packet loss	MOS by SMOS test
0%	4.48
3%	4.38
6%	4.27
9%	4.14
12%	4.00
15%	3.93
18%	3.84
21%	3.75
24%	3.59
27%	3.51
30%	3.45

Table 3: MOS readings measured by SMOS test for G.711 codec with various percentage packet loss.

Table 4 lists the MOS numbers at different packet loss for G.729 vocoder.

Packet loss	MOS by SMOS test
0%	4.07
3%	3.98
6%	3.87
9%	3.75
12%	3.63
15%	3.51

Table 4: MOS readings measured by SMOS test for G729 vocoder with various percentage packet loss.

One should keep in mind that, when testing SMOS through a network with packet loss, the MOS readings contain high level of variance due to the dynamic nature of packet loss (one cannot control exactly where the packet loss will occur. Some occurs during silence and some occur during active period which will cause different amount of degradation). Therefore, the MOS readings on each direction may be different (even if the nominal packet loss is the same). Also, the readings may not be repeatable at each test, again because the actual packet loss occurrence cannot be exactly controlled. For this reason, all the numbers shown in Table 3 and Table 4 should be understood to have a "natural" variance of about ± 0.05.

3.3 Do not infer packet loss from MOS

The numbers in Table 3 and Table 4 may give some readers a false impression that one can "accurately" infer the exact amount of packet loss from certain "fuzzy" MOS type of measurement. For the reasons described above, such claim is false and such attempt should be avoided.

Any psychoacoustic-model-based MOS measurement is aimed at quantizing the speech degradation in the same way as real human listeners do. If a human listener can not tell the exact amount of packet loss and voice clipping and jitter, neither can the psychoacoustic MOS measurement. There is no such simple and monotonical (that is, may not even be proportional depending on how the lost packet is handled and what codec is used) relation between packet loss and MOS degradation. Any attempt to "correlate" packet loss with MOS number is a gross simplification of a complex scientific problem. Sage does not recommend our customers to do this. The exact amount of packet loss should be measured with Sage's PVIT [2].

3.4 MOS with voice clippings

To further save bandwidth, silence suppression scheme is often used to more efficiently encode and decode the "silence" period. However, aggressive silence suppression through "sloppy" VAD (Voice-Activity- Detector) can lead to annoying leading-edge voice clipping and the "suppression" of non-silent low-level non-voice segment of the test signal. Like packet loss, such impairments will be measured by SMOS as degradation of MOS reading, although the exact amount of such impairments can not be determined.

3.5 MOS with voice jitters/frame slips

Although deemed as "non-perceptible" to human ears, voice jitters/frame slips are the most detrimental to some voice processing devices such as network/embedded echo cancellers and voice-band modems (fax). In fact, voice jitter is also the primary reason why ITU-T P.861 PSQM was declared as "in-applicable" for end-to-end VoX voice quality measurement, hence the development of Sage's SMOS.

In SMOS test, the jitters that occur strictly in the silence period do not affect MOS reading. SMOS test contains a robust jitter-removal algorithm that can completely remove the voice jitters so that they do not affect the performance of the psychoacoustic core.

However, if the jitters occur in-discriminately in active period, then the MOS reading will be degraded even more severely than plain packet loss. This is consistent with human perceptions, of course. For example, a 30 ms jitter in active voice segment not only implies a 30 ms packet loss, it also causes an unnatural 30 ms "jerking" effect. Even if the voice jitters strictly occur in the silence periods, they do affect the speech naturalness (causing gapping and jerking effects), although they do not affect the speech clarity measured as MOS.

Considering the significance of jitters, SMOS also measures the total amount of jitters that occur during the test. More details on jitters and jitter measurement will be discussed in section 5.

3.6 MOS with analog-type of distortion

By analog-type of distortion, we mean attenuation distortion and additive noise.

Such analog type of distortion are unlikely to occur in a true all-digital VoX environment. But if the tests are performed through the several-mile-long analog POTS loop, the 2-wire loop attenuation distortion and cross-talk noise may cause measurable degradation on MOS reading.

SMOS contains a Bark-domain equalization algorithm that removes the attenuation distortion because human ears are not so sensitive to such distortion. But the equalization is only partial because attenuation distortion does affect speech quality (certain intelligibility is lost when one tries to spell a name across a phone call with "severe" attenuation distortion).

In short, unlike in PSQM where a slight attenuation distortion can result in significant drop of MOS reading, SMOS test maintains a steady MOS reading that degrades more "gracefully" with the increase of analog attenuation distortion and is consistent with human perception.

To be more specific, Table 5 lists the MOS numbers when there is different amount of attenuation distortion with G.711 coding. The attenuation distortion is simulated by a C-message digital filter specified in IEEE743 [6]. The frequency response curve of the C-message filter is shown in Figure 3 of section 9. The 4 cases represent the following 4 scenarios of a call connection:

- Case 0: No attenuation distortion. Digital connection.
- Case 1: Calling from the analog POTS phone to a digital phone. The test signal is passed through the C-message filter (simulating the POTS loop attenuation distortion), and then going through the G.711 PCM companding.
- Case 2: Calling from the digital phone (IP phone etc) to the analog POTS phone. The test signal is G.711 PCM companded first, and then passes through the C-message filter.
- Case 3: Calling from one POTS phone to another POTS phone. The test signal is passed through a C-message filter, G.711 companded and then passed through a second C-message filter.

Attenuation type	MOS by SMOS test
Case 0, PCM only	4.50
Case 1, Filter then PCM	4.39
Case 2, PCM then filter	4.39
Case 3, Filter, PCM, Filter	4.25

Table 5: MOS readings measured by SMOS test for G.711 encoding with various combinations of attenuation distortions. The attenuation distortion is simulated by a C-message filter.

In comparison, if no equalization is performed, then the MOS reading in case 3 of Table 5 will be as low as 3.70 (as in the case of PSQM test).

Table 6 lists the MOS numbers when there is additive noise. The noise is added after G.711 PCM companding. The noise is a uniformly-distributed and unfiltered white noise (within 0 to 4000Hz). Notice that, the artificial voice signal used by SMOS has an active speech level of -20 dBm.

In SMOS test, the attenuation distortion is measured as effective bandwidth, which will be further discussed in section 7. The additive noise is reflected as the silence noise level measurement, which will be discussed in section 9.

4 Round-trip Delay Measurement

The MOS reading given by SMOS indicates the speech transmission quality (the so-called listening clarity). No matter how long the delay is, delay does not change the MOS reading.

Additive noise level	MOS by SMOS test
No noise	4.50
-55 dBm	4.45
-45 dBm	4.28
-35 dBm	3.93
-25 dBm	3.41
-20 dBm	3.14

Table 6: MOS readings measured by SMOS test for G.711 encoding with various levels of additive noise. The noise is a uniformly-distributed white noise (within 0 to 4000Hz). Notice that at -20 dBm, the additive noise is in fact equal in level to the artificial voice test signal [7].

But in a real phone conversation, delay matters a lot. In fact, delay is probably the most important, yet often ignored, parameter in judging the QoS of a VoX application. If delay were not an issue, the VoX quality would be perfect. There would be no need to compress voice to save bandwidth. There would be no packet loss, because the receiving gateway can wait "forever" for the late-arrived packets, or using TCP retransmission (instead of UDP with RTP encapsulation) to guarantee the packet delivery. But in reality, the need for short delay for real-time interactive audio (telephone) and video applications is the single most difficult issue facing a packet-switched network [8].

Long delay causes two problems in a practical phone conversation:

1. Long delay affects the natural conversation interactivity, and causes hesitation and over-talk. A caller starts noticing delay when the round-trip delay exceeds 150 ms. ITU-T G.114 [9] specifies the maximum desired round-trip delay as 300 ms. A delay over 500 ms will make phone conversation impractical.

2. Long delay exacerbates echo problems. An echo with level of -30 dB would not be "audible" if the delay is less 30 ms. But if the delay is over 300 ms, even a -50 dB echo is audible. ITU-T G.131 [10] specifies the echo delay and echo level requirements.

Generally speaking, the three parameters that determine voice quality are clarity (indicated by MOS), delay and echo [8]. This is particularly true for VoX application, as the most notable difference between a PSTN call and a VoX call (especially for local calls) is the huge delay difference. A coast-to-coast long distance call over PSTN network has round-trip delay about 45 to 100 ms. But a VoDSL application can have a delay as high as 250 ms even for local calls.

SMOS test measures round-trip delay through accurate time-domain cross correlation with spread spectrum signal preceded by a series of pilot signals. The measurement algorithm can withstand a vocoder compression as low as 2kbps, and can tolerate packet loss up to 40% (with G.711 encoding). The measurement range is 0 to 5000 ms, and accuracy is ± 0.125 ms.

In SMOS test, delay is measured at the beginning of the call (after the TPT (test-progress-tone), see Figure 4 of section 10). For VoX applications, the delay may be different at the end of the call due to voice jitters/frame slips. Also for VoX applications, the delays could be "naturally" different from call to call.

5 Voice jitter/Frame slip measurement

By frame slip (or voice jitter), we mean a sudden delay variation at the audio signal side. Figure 1 shows how voice jitters distort the audio signal.

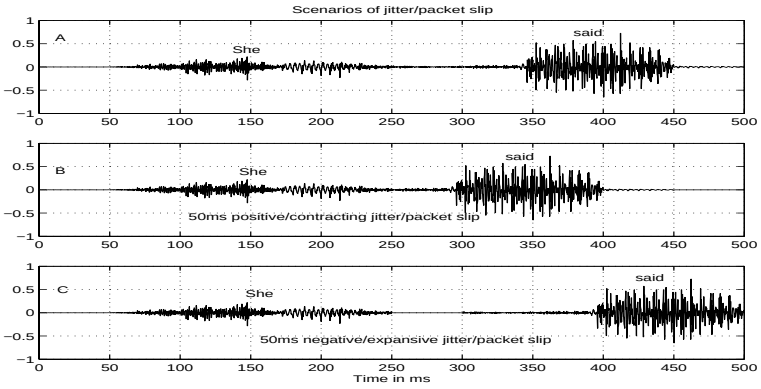

Figure 1: Hypothetical Scenarios of voice jitters/frame slips. A: A snapshot of male speech saying "She said". B: A 50 ms block of signal centered at 300 ms time mark is annihilated (deleted), causing the relative delay between the word "She" and the word "said" to be shrunk by 50 ms. This will be measured by SMOS as a positive frame slip (jitter) of 50 ms. This shortening of delay will cause a "jerking" effect in the audio signal. 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 SMOS as a negative frame slip of 50 ms with a time stamp around 275 ms. This lengthening of delay will cause a "gapping" effect in the audio signal.

Voice jitters are unique to packet-based switching that do not exist in a circuit-switched network. They are caused by the following packet-network-specific phenomena:

1. Jitter buffer resizing: A packet-switched network is inherently jittery (each packet will arrive asynchronously and even out of order), but the audio signal side (or the circuit-switched PSTN side) requires continuous and synchronous playout. To balance the dis-similarity, a jitter buffer is commonly used inside a voice Gateway or IAD at the receiving side. A large jitter buffer can minimize packet loss, but will induce longer delay. To balance the conflicting need for shorter delay and less packet loss, the jitter buffer may be dynamically re-sized depending on the network traffic situation. Whenever the jitter buffer re-sizes, the audio signal will experience a sudden delay variation (jitter or frame slip) in an amount (in ms) that matches the voice frame size (6, 10 or 30 ms etc).

2. Lack of jitter buffer: If no jitter buffering is used in a gateway or IAD, that is, the gateway/IAD just plays the packets as they arrive with no buffering, then the packet-network jitters will directly propagate into the audio signal side, causing large amounts of "chaotic" jitters in the audio signal side that can render almost all voice-testing inapplicable. This most likely happens on certain "low-cost" residential IADs and gateways etc.

3. Lack of packet/cell/frame sequence number: To remove the packet jitters and smooth out the out-of-order packets, the packets must be sequenced and numbered at the transmitting side

and examined by the receiving side. On VoIP, the RTP packet carries the packet sequence number. But for VoATM application, such sequence number may not be there because an ATM network is designed to be connection-oriented with guaranteed delivery (at least for AAL2), therefore, the cells never arrive out of order and there is no need for sequence number. However, for certain broadband local access technology that employs ATM transport (VoDSL and VoCable etc), the delivery of ATM cells may be questionable when there are other data traffic. In this scenario, the lack of sequence number may cause ATM cell loss to become voice jitters at audio signal side. For example, if cell 2 is lost, the receiving IAD/Gateway will not even know cell 2 is there when it receives cell 1 and cell 3 because they are not numbered. The IAD/Gateway will proceed to play cell 3 after cell 1, hence causing an abrupt delay change (frame slip or jitter) at the audio signal side.

For the above 3 scenarios, the second and third are most hideous because the jitter occurs indiscriminately at any period of the test signal. In the first scenario for VoIP (using RTP encapsulation), the jitter buffer resizing should only occur during the silence period, thus causing the least amount of speech degradation.

In SMOS test, the jitters that occur strictly in the silence period do not affect MOS reading. However, if the jitters occur in-discriminately in active period, then the MOS reading will be degraded even more significantly than plain packet loss. This is consistent with human perceptions, of course. A 30 ms jitter in active voice segment not only implies 30 ms packet loss, it also causes an unnatural "jerking" or "gapping" effect.

Although SMOS has a de-jittering algorithm that can properly factor out the jitters in silence periods, excessive amount of jitters does produce unnatural "jerking" and "gapping" effect. Furthermore, indiscriminate jitters that occur in active signal period will cause more severe speech quality degradation than plain packet losses. Jitter/frame slip is also highly detrimental to voice band modem (such as fax) and network echo cancellers. In short, voice jitters should not be considered as entirely "harmless" features of VoP networks. They should be minimzed as much as possible without sacrificing the delays and packet losses.

Considering the significance of jitters, SMOS measures such jitters or frame slips. During the period when artificial voice signal [7] is "played" (see signal sequence 5A and 5B in Figure 4 of section 10), any sudden delay variations larger than 2 ms will be detected, measured and accumulated. Two types of frames slips will be reported by SMOS:

1. Positive(+) frame slip: the total amount of compressive jitters (shortening of delays) that correspond to the down-sizing of jitter buffer or the deletion of packets/frames/cells etc.

2. Negative(−) frame slip: the total amount of expansive jitters (lengthening of delays) that correspond to the "up-sizing" of jitter buffer or the insertion of packets/frames/cells etc.

SMOS measures the total frame slips up to 2000 ms with a resolution of 1 ms. Notice that only the jitters inside the artificial voice signal period (sequence 5A and 5B in Figure 4 of section 10) are measured. The jitters that occur elsewhere (during the telemetry and TPT etc) are not measured.

Although "controlled" frame slips that occur only in silence periods do not affect MOS reading, excessive amounts of frame slips are not desirable either. Generally, we recommend a good system should maintain a total amount of jitter to less than 3% of the test duration. That is to say, for a 10 seconds test, the total amount of positive and negative slips measured by SMOS should be within [-300,300] ms. If SMOS measures higher amount of jitters, then the network should be re-configured for better traffic engineering and prioritization. On a network where there is no bandwidth contention (only one call through a 2Mbps VoDSL channel, for example), there should

not be any measurable frame slips. If there are, then the jitter buffer adaptation algorithm inside the IADs or gateways need to be further optimized.

6 Codec Type Detection

Codec type detection serves the following three purposes:

1. Establish a reference MOS number. Is MOS 4.15 a good number for a network under test? Without codec type information, one cannot answer such question. But if the codec type is known, one can then compare the measured MOS with the "theoretical" ideal number in Table 2. For example, if the codec is G.711 PCM, then MOS 4.15 is a very bad number, indicating there might be serious packet loss or coding problems. But if the codec is a typical 8kbps vocoder, then MOS 4.15 is just about right.

2. Verify Service-Level-Agreement (SLA) and network configuration. If a network is configured (based on SLA) to use G.711 PCM, not G.726 ADPCM, such SLA configuration can be verified by SMOS's codec type detection.

3. Trouble-shoot codec transcoding problem. For a "hybrid" long distance network, codec transcoding is not only a problem for voice quality, it also poses a dilemma for network management and SLA. More specifically, a call may start with G.726 ADPCM encoding. But in "middle" of the network, it may become G.723.1 vocoder. At the end, it becomes G.726 ADPCM again. The presence of G.723.1 transcoding in such case can be verified with SMOS's codec type detection. In this case, SMOS will report the G.723.1 vocoder, instead of the ADPCM.

Table 7 summarizes all the codec types that SMOS can detect and report:

Codec Type Report Symbol	Codec Type Description	reference MOS range
VCD4K	Sub-4kbs vocoders	[3.0,3.8)
VCD8K	5-8kbps vocoders	[3.8,4.2)
VCD16K	12-16kbps vocoders	[4.2,4.35)
ADPCM16	16kbps G.726 ADPCM	[3.4,3.6]
ADPCM24	24kbps G.726 ADPCM	[3.9,4.1]
ADPCM32	32kbps G.726 ADPCM	[4.2,4.3]
ADPCM40	40kbps G.726 ADPCM	[4.3,4.4]
ADPCM	G.726 ADPCM with unknown data rates	[3.5,4.3]
PCM	G.711 μ/A-law PCM or pure analog	[4.45,4.60]
UNSURE	Too much distortion, not sure	N/A

Table 7: Codec types that SMOS can detect and report

In reference to Table 7, the following points are worth noting:

- Despite the apparent connection between codec type and MOS number in Table 2 and Table 7, SMOS test does not use MOS number to infer the codec type. In SMOS test, the MOS number is measured through the P.50 [7] artificial voice signal. The codec type, on the other hand, is measured through a proprietary codec-detection signal with proprietary DSP algorithm (see signal sequence 2A and 2B in Figure 4 of section 10.

- VCD8K covers a large number of toll-telephony-quality vocoders ranging from 5kbps to 8kbps. Typical examples are G.723.1 (5.3 and 6.3 kbps), G.729, ACELP, VSELP and QCELP8 etc.
- VCD4K covers those sub-4kbps communication quality vocoders on either proprietary VoX systems or for secure voice applications.
- ADPCM: if the algorithm is not comfortable in further distinguishing among the variants (32k or 40k etc) of ADPCM because of too much signal distortion (by packet loss, jitter or noise etc), it will just simply report "ADPCM".
- Analog connection: the codec type detection does not distinguish between pure analog coding (transmission) and G.711 PCM codec. A pure analog coding will be reported as the same as G.711 PCM codec, because in any "modern" digital telephony network, there aren't any end-to-end pure analog transmission systems. Even a POTS to POTS local call will go through a class 5 switch with G.711 PCM coding.
- Algorithm reliability: the algorithm can correctly detect the codec type under as high as 25% packet loss impairment with G.711 PCM and VCD8K codings. For other codings (such as ADPCM and VCD16), the algorithm can withstand as high as 13% packet loss. If there are too much distortions, the reported codec type could be erroneous toward the lower data rate type. For instance, a PCM could be reported as ADPCM32, or an APCM40 could be reported as VCD16, etc. If there is much distortion (too much noise or packet loss etc), the algorithm may report "UNSURE".

7 Effective bandwidth measurement

This measures the attenuation distortion (measures how flat the frequency response of the system under test is within the 300 to 3400 Hz band). SMOS partially equalizes the attenuation distortion as human ears are not so sensitive to such distortion. But excessive attenuation distortion (low effective bandwidth) does indicate system design problems either due to poor analog circuitry design or unnecessarily excessive digital filtering.

Figure 2 shows two types of attenuation distortion.

Assume the frequency response (power amplitude) of a channel under test is $H(f)$ as shown in Figure 2, the effective bandwidth is calculated as:

$$BW = \frac{[\int_{300}^{3400} H(f)df]^2}{3100 \int_{300}^{3400} H^2(f)df}$$

By Schwarz's inequality, one can prove that BW is bounded between [0,1]. When $H(f)$ is completely flat between [300,3400]Hz, then $BW = 1$. Otherwise, $BW < 1$.

For PCM and ADPCM waveform coders, the effective bandwidth largely reflects the attenuation distortion caused by analog or digital filtering. These coders themselves do not introduce much attenuation distortions (like type A in Figure 2). Their effective bandwidths will be very close to 1 if there aren't any analog or digital filtering.

But for parameter-based low-bit-rate vocoders, their effective bandwidths will be much less than 1 even if there aren't any analog or digital filtering (type B in Figure 2).

As a reference, Table 8 lists the "theoretical" effective bandwidths of various codecs themselves (pure compression effect, no other analog or digital filtering effects). Notice that the effective bandwidth is not necessarily proportional to the codec bit rate, especially for the low-bit-rate vocoders.

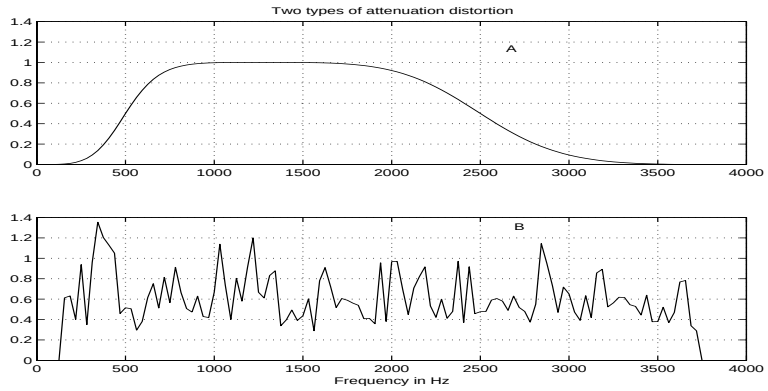

Figure 2: Two types of attenuation distortion. The frequency response curves are shown in linear scale. A: analog-type of attenuation caused by filtering, resulting in uneven but smooth frequency response. B: digital-type of attenuation distortion caused by low-bit-rate voice compression (G.723.1 6.3kbps vocoder, in this case), resulting in "flat" but uneven (non-smooth) frequency response. When measured by SMOS, type A will have an effective bandwidth of 0.74, whereas type B will have an effective bandwidth of 0.88.

If a system under test uses waveform coders (G.711 PCM and G.726 ADPCM), its measured effective bandwidth should be higher than 0.9. Anything below 0.85 signifies either excessive loop attenuation distortion (or poor analog circuitry design if testing through analog connection) or excessive band-limiting digital filtering.

If a system under test uses non-waveform low-bit-rate vocoders, then one must be careful in interpreting the value as shown in Table 8. Generally, the effective bandwidth should be maintained above 0.7. Anything below 0.65 indicates poor analog circuitry design (if testing through analog connection) or excessive digital filtering.

With analog interface, Sage's test equipment itself may report a "residual" effective bandwidth as low as 0.98 (not perfect 1.00) when testing unit-to-unit due to slight internal hardware roll-off (anti-aliasing filter before A/D converter the low-pass interpolation filter (sinc-filter) after D/A) at frequencies above 3300Hz. This "imperfection" has no effect on other measurements.

Codec type	Effective Bandwidth by SMOS test
G.711 PCM@64kbps	1.00
G.711 PCM robbed-bits	1.00
G.711 PCM@56kbps	1.00
G.726 ADPCM@40kbps	1.00
G.726 ADPCM@32kbps	0.99
G.726 ADPCM@24kbps	0.97
G.729@8kbps	0.77
G.723.1@6.3kbps	0.88
Cell-phone VSELP@8kbps	0.85
Cell-phone EFR-ACELP@7.4kbps	0.87

Table 8: Effective bandwidth readings measured by SMOS test for some commonly used codecs

8 Voice Band Gain Measurement

This measures the overall voice band (300 to 3400 Hz) signal level change (attenuation or gain). Flat gain change is not reflected in the MOS reading. But excessive level change (too loud or too faint) does affect human perception. A "clean" VoX system should maintain a proper voice level change (gain) in the range of [-10,-3] dB.

Notice that the gain is measured and integrated across the whole 300 to 3400 Hz band. This gain value may be different from those obtained using a single-tone signal (sending tone and measuring tone, for example) if the channel is not flat (that is, effective bandwidth less than 1). They will be equal only if the channel is completely flat (effective bandwidth equals 1).

9 Silence Noise Level Measurement

The silence noise level is measured to reflect the following two problems:

1. On a POTS to POTS PSTN call, the silence noise level indicates the cross-talk or any other interference noise level present in an analog loop line.

2. For a VoX system that employs silence suppression to save bandwidth and uses Comfort-Noise-Generator (CNG) to reproduce the "silence", this silence noise level measurement indicates the comfort noise level. The level should not too high (sounds too noisy) nor too low (sounds like a dead line).

Notice that the silence noise is only measured in a specific 400 ms silence period during the test. If the noise is transient (impulsive), and occurs in-frequently, then SMOS may or may not "catch" it.

The silence noise level is measured through a C-message filter [6], whose frequency response is shown in Figure 3. The noise level is expressed in dBrnC. An ideal system should maintain a silence noise level between [10,30] dBrnC. Above 30dBrnC sounds too "noisy" and below 10dBrnC may sound too "quiet".

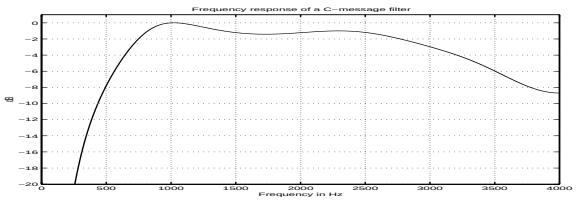

Figure 3: Frequency response of the C-message filter.

When performing unit-to-unit test using Sage's test equipment, with analog connection, the inherent silence noise level will be around 20 dBrnC because SMOS uses fixed ranger and 20 dBrnC is pretty much the noise floor of the A/D and D/A converters being used. With digital connection, if clear-channel is used (64kbps T1 or plain E1), the noise level will be reported as 0 dBrnC. With robbed-bits signaling, the silence noise level will be round 8 dBrnC.

10 Signal Flow Sequence

SMOS is an automated test that performs a series of measurements. The signal sequence is therefore also complex. Figure 4 shows the signal flow on both directions (from director (near) to responder

(far) and vice versa).

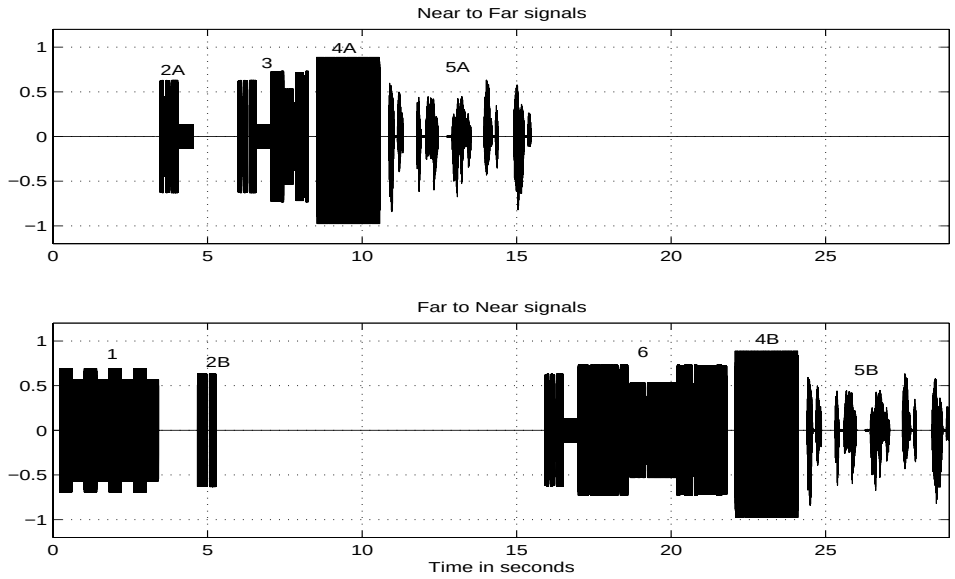

Figure 4: Signal flow sequence of SMOS test. Top: near to far. Bottom: far to near. Signal 1 is the TPT. Signal 2A and 2B are a sequence of signals used for delay measurement. Signal 3 is sequence of synchronization and command telemetry signal. Signal 4A and 4B are used for codec-type detection. Signal 5A and 5B are the P.50 artificial voice signal for MOS measurement. Signal 6 is the synchronization and results telemetry signal.

In reference to Figure 4, detailed descriptions of the signals are as follows:

1. TPT: this is a 625Hz, -7 dBm, and amplitude-modulated test progress tone. This tone is sent from the far-end responder to the near-end director after the far-end answers the call originated from the director.

2. Delay measurement signal: after detecting the far-end TPT, the near-end sends out a sequence of signals for delay measurement. This sequence of signal consists of a series of pilot tones and a frequency-hopped spread-spectrum signal. The far-end detects one of the pilot tones, synchronizes to the spread-spectrum signal, and loops back the spread-spectrum signal with precise delay and precedes the signal with regenerated and frequency-shifted pilot tones.

3. Synchronization and command telemetry: after delay measurement, the near-end sends a sequence of signal for rough synchronization and command telemetry (passing user-inputs from near-end to the far-end).

4. Codec-type detection: 4A and 4B are special white-noise-like -7 dBm signal used for codec-type detection. The effective bandwidth, voice level change (gain) are also measured using this signal. The silence noise level is measured in the silence period preceding this signal.

5. P.50 artificial voice signal: 5A and 5B are the artificial voice signal used for MOS measurement. The duration of this signal varies according to user input (how long you want the test to be). Only 5 seconds of test is shown in Figure 4. Voice jitters (frame slips) are also measured using this artificial voice signal.

6. Synchronization and results telemetry: after the near-to-far measurement, the far-end sends back the measurement results through this sequence of signal: synchronization signal and results telemetry signal.

11 User Option

The only user option is test duration (in seconds).

When testing through a "static" network where there are no dynamic impairments such as packet losses, a duration between [7,15] seconds are adequate.

When testing through a dynamic network where packet loss impairments abound, the test duration should be longer (> 16 s) to obtain a statistically more stable MOS reading.

Only MOS reading has to do with test duration. The other measurements have fixed measurement duration.

Upon each test, the artificial voice signal follows a random pattern (that is, different from test to test) to guarantee statistical "ergodicity" of the test.

12 Application Examples

The purpose of this section is to demonstrate how to use the plethora of information provided by SMOS to evaluate the true QoS of a system under test.

12.1 Case 1, delay is too long

Table 9 lists a set of hypothetical SMOS results when testing through a VoDSL/VoCable application. Based on the SMOS results in Table 9, the following conclusions can be drawn:

	MOS	Codec Type	Bandwidth	Gain	Noise	+slip	-slip	delay
N-F	4.23	ADPCM32	0.97	-8 dB	23 dBrnC	12 ms	-18 ms	
F-N	4.24	ADPCM32	0.98	-8 dB	23 dBrnC	12 ms	-6 ms	
2-way								400.0 ms

Table 9: Hypothetical SMOS results over a VoDSL application.

1. The VoDSL system is using 32kbps ADPCM codec. The measured MOS number (4.23 and 4.24) is very close to the ideal theoretical number (4.27) in Table 2. So the voice clarity is good for such codec. There are no measurable packet losses.

2. The 0.97 and 0.98 effective bandwidth indicates that there is no attenuation distortion.

3. The voice gain (-8 dB) and noise level (23 dBrnC) all fall within the acceptable ranges.

4. The minute amount of frame slips are acceptable.

5. But, the round-trip delay of 400 ms is too long for voice telephony. The IAD or access gateway need to be optimized to shorten the delay to less than 150 ms. At this 400 ms, the echo level should also be measured through Sage's Echo Sounder [3].

12.2 Case 2, too much attenuation distortion

Table 10 lists a set of hypothetical SMOS results when testing through a POTS to POTS PSTN call. Based on the SMOS results in Table 10, the following conclusions can be drawn:

	MOS	Codec Type	Bandwidth	Gain	Noise	+slip	-slip	delay
N-F	4.30	PCM	0.75	-16 dB	23 dBrnC	0 ms	0 ms	
F-N	4.30	PCM	0.75	-16 dB	23 dBrnC	0ms	0ms	
2-way								43.0 ms

Table 10: Hypothetical SMOS results over a POTS to POTS PSTN call.

1. The PSTN network, of course, is using G.711 PCM codec. The measured MOS of 4.30 falls well below the theoretical 4.50 number.
2. The 0.75 effective bandwidth indicates that there is significant amount of attenuation distortion. This attenuation distortion is also causing the MOS number to be as low as 4.30. The loop lines of this PSTN call have less-than-desirable quality.
3. Other parameters all fall within acceptable ranges.

12.3 Case 3, too much packet losses

Table 11 lists a set of hypothetical SMOS results when testing through a true VoIP network. Based

	MOS	Codec Type	Bandwidth	Gain	Noise	+slip	-slip	delay
N-F	4.10	PCM	0.97	-6 dB	23 dBrnC	10 ms	-20 ms	
F-N	4.20	PCM	0.96	-6 dB	23 dBrnC	20 ms	-10 ms	
2-way								153.0 ms

Table 11: Hypothetical SMOS results over a VoIP call.

on the SMOS results in Table 11, the following conclusions can be drawn:

1. This VoIP call uses G.711 PCM codec. Its measured MOS number (4.1 and 4.2) falls well below its ideal number of 4.50.
2. There is no excessive amount of attenuation as the effective bandwidth (0.97 and 0.96) are high enough.
3. The voice jitters/frame slips are acceptably small.
4. The gain, noise level and delay are all acceptable.
5. The low MOS number must be caused by significant amount of packet losses.

	MOS	Codec Type	Bandwidth	Gain	Noise	+slip	-slip	delay
N-F	3.90	VCD8K	0.80	-9 dB	23 dBrnC	330 ms	-300 ms	
F-N	3.90	VCD8K	0.80	-9 dB	23 dBrnC	300 ms	-330 ms	
2-way								350.0 ms

Table 12: Hypothetical SMOS results over a VoIP call.

12.4 Case 4, too much frame slips

Table 12 lists a set of hypothetical SMOS results when testing through a true VoIP network. Based on the SMOS results in Table 12, the following conclusions can be drawn:

1. This VoIP call uses the toll-quality vocoders such as G.729 or G.723.1. Judging from delay and frame slips, G.723.1 is more likely than G.729, as G.723.1 has a voice frame size of 30 ms (therefore longer delay and potentially larger frame slips), whereas G.729 has a voice frame size of 10 ms. The measured MOS number 3.90 falls short of the ideal number between [4.0, 4.1].

2. The frame slips (+330 ms and -300 ms) are too much for a 10-second test. Some frame slips may have occurred in the active voice period, which explains the lower-than-theoretical MOS reading for G.723.1 vocoder. Of course, packet loss may also have occurred. PVIT [2] should be further used or identify the packet network impairments.

3. The round-trip delay of 350 ms is a bit too long, although not unusual for VoIP using G.723.1 vocoding. There are potential echo problems. Echo Sounder [3] should be used to "clear" the mind.

4. The effective bandwidth of 0.80, although low, is about normal for a G.723.1 vocoder.

5. Other parameters are within acceptable ranges.

References

[1] Renshou Dai, "A technical white paper on Sage's Perceptual-Speech-Quality-Measurement (PSQM) test," Sage Instruments white paper, May, 1999.

[2] Renshou Dai, "A white paper on Sage's Packet-Voice-Impairment-Test (PVIT)," Sage Instruments white paper, December, 2000.

[3] Renshou Dai, "White paper on Sage's Echo Sounder and Echo Generator," Sage Instruments white paper, March 2001.

[4] S. Wang, A. Sekey, A. Gersho, "An objective measure for predicting subjective quality of speech coders," *IEEE Journal on Selected Areas in Communications,* 10(5), 819-829, 1992.

[5] E. Zwicker, H. Fastl, *Psychoacoustics, facts and models,* Second updated edition, Spinger, 1990.

[6] "IEEE Standard Equipment Requirements and Measurement Techniques for Analog Transmission Parameters for Telecommunications," IEEE Std 743, 1995.

[7] "Artificial Voices," *ITU-T Recommendation P.50*, March 1993.

[8] Renshou Dai, "Multi-dimensional Approach to VoX Voice Quality Measurement," A seminar offered by Sage Instruments. 2001.

[9] "One-way transmission time," *ITU-T Recommendation G.114*, May, 2000.

[10] "Control of talker echo," *ITU-T Recommendation P.131*, Aug., 1996.