Noise and Voice Quality in VoIP Environments

White Paper

Dennis Hardman
Technical Marketing Engineer
Agilent Technologies
Colorado Springs, Colorado, USA

Agilent Technologies

Introduction

In many ways, noise reduction techniques in a VoIP (Voice over IP) environment mirror those used in traditional voice transmission systems. In other words, as long as PSTN (public switched telephone network) technologies are used in tandem with VoIP technologies – and in almost all cases, they are – PSTN-like noise reduction is required. However, when voice signals are encoded, packetized, and transmitted (even for part of the voice path) across a VoIP network, other network behaviors and impairments come into play that might or might not be adequately handled by traditional telephony noise reduction and cancellation techniques. Voice signals are encoded in new ways across a VoIP network, and are transported from point-to-point across networks designed for non-real-time traffic. In addition, VoIP networks are often not subject to historical and, until relatively recently (due to the deregulation of the late 1990's), regulatory standards and constraints[1]. As a result, an interesting and challenging host of noise sources emerge. Noise reduction in VoIP networks must take these new sources into account.

Noise is *any* interfering sound. In the context of VoIP, a broader definition is perhaps required. Noise can be more generally defined as *distortion*. In other words, noise can be thought of as any undesirable characteristic that degrades the signal of interest. Given this definition, in a VoIP environment there are two types of distortion: additive and subtractive . And along with the signal distortion described here (which clearly affects *sound quality*), VoIP network behavior can impact *conversational quality* in ways not often seen in most PSTNs.

Strictly speaking, VoIP does *not* introduce any "new" sources of noise or distortion that do not already exist in one form or another on other communications networks. For example, IP networks have always exhibited packet loss and jitter (delay variation). PSTNs produce analog channel noise and echo, and always have. Quantization distortion, attenuation/level problems, low bit rate codec distortion, and so on have all existed for some time. It is the relatively new and unique combination of real-time voice with non-real-time data network behavior, *and* the interworking of VoIP with traditional PSTNs that create the new challenges of voice and conversational quality. Because of this, noise *avoidance* in a VoIP environment is as important as noise *reduction*.

This paper briefly describes VoIP technologies and deployments, introduces in more detail the signal distortion and conversational quality impairments that VoIP exhibits, and discusses some of the techniques being used to ameliorate these impairments. Finally, the paper provides an examination of measurement techniques that target the unique voice-over-IP environment. Note that this chapter approaches noise reduction (and avoidance) from a *system* point of view. Detailed descriptions of network components, processes, or noise reduction techniques can be found in other Agilent white papers and signal processing text books.

VoIP Overview

VoIP, or voice-over-IP, refers to an expanding family of voice processing and transport technologies that seek to take advantage of existing data network infrastructures. VoIP networks promise to reduce the cost of local and long distance telephone calls for individuals and businesses alike, and have the potential to provide unique new services and hasten computer-telephony integration. Relative to traditional telephone networks and data communications networks, VoIP is still in its infancy. But as voice and data service providers look for new ways to improve service offerings while increasing profits and reducing costs, VoIP stands a good chance of becoming one of the most important voice processing and transport technologies in the communications industry. To be widely accepted and deployed, however, VoIP must address several significant challenges. One of these challenges is matching the signal and conversational quality that is consistently delivered by PSTNs (public switched telephone networks) and to which telephone customers have become accustomed. Related to the challenge of achieving acceptable sound and conversational quality is the technical challenge of integrating and interworking VoIP with existing voice networks[2].

With regard to voice signal quality, one of the primary differences between the PSTN and the VoIP network is that the PSTN provides a dedicated voice channel of consistent bandwidth for each voice call, whereas a VoIP network provides best-effort voice packet delivery consistent with IP network behavior. Another way of looking at this difference is that PSTN voice channels are designed with the voice signal in mind (i.e., they have just the right amount of bandwidth and the right frequency response to minimally support a conversational quality voice signal). IP networks, on the other hand, were never really designed for real-time, dedicated bandwidth applications like voice. This difference affects virtually all aspects of noise and distortion avoidance for VoIP implementations and VoIP/PSTN integration. Another interesting difference is the fact that PSTNs provide call setup and management intelligence in the core of the network (via SS7 signaling and central office processing) whereas VoIP networks have pushed this intelligence to the edge of the network where it resides in VoIP endpoints such as personal computers or IP/Ethernet telephones. This can also impact voice quality because the network core is no longer as tightly controlled or regulated.

This paper provides a basic overview of VoIP implementations and technologies, showing where noise and distortion issues can arise. Douskalis[2] and Minolli and Minolli[3], provide more detailed information about the technologies, implementations, and measurement techniques for voice over IP.

VoIP Implementations

In its most basic and generic form, a VoIP network consists of user endpoints (telephone, fax, modem, VoIP computer terminal, etc.) connected to media gateways which, in turn, are connected to the IP signaling and media transport network. This basic architecture is shown in Figure 1.

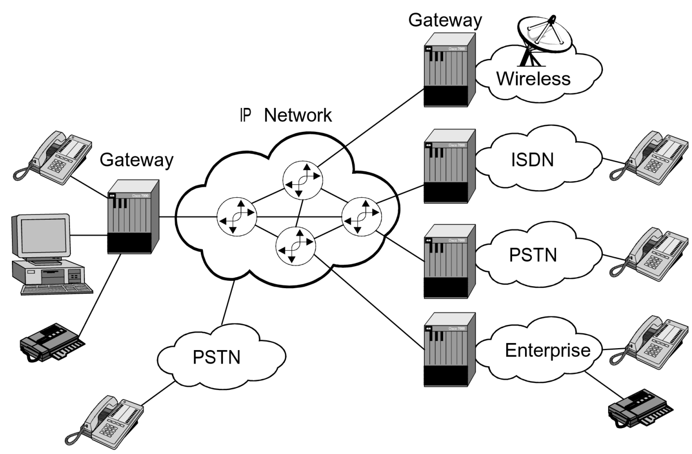

Figure 1: Basic VoIP implementation

Because a VoIP network must provide ubiquitous call access, it is almost certainly connected to and integrated with various other voice transport networks including cellular, ISDN (Integrated Services Digital Network), PSTN, and proprietary enterprise data and voice networks. Depending on the VoIP protocols and equipment used, other devices can be deployed and implementations can become quite complex. Please note that in Figure 1 and in the remainder of this paper, “PSTN” generally refers to any analog voice circuit ranging from an analog telephone connected to the analog FXS (Foreign Exchange Station) port of a VoIP gateway or router, to an analog telephone connected to a service provider’s local loop and central office.

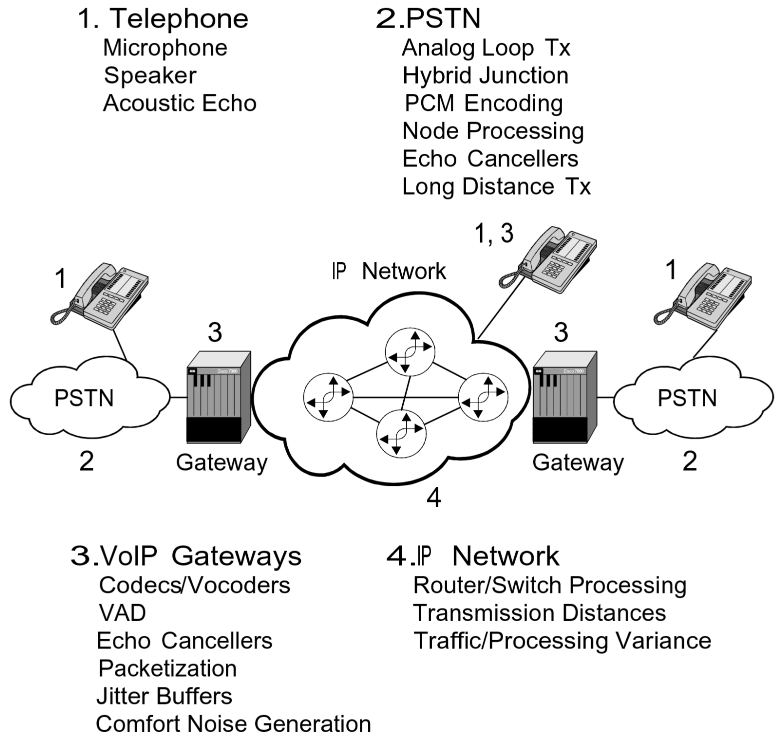

Figure 2: Sources of Voice Quality Impairments (Copyright 2001, Agilent Technologies, Inc. Reproduced with Permission)

There are various places in a typical VoIP and/or VoIP/PSTN implementation that can cause, or make worse, noise and signal distortion. Figure 2 identifies the main sources of voice signal impairment: IP network behavior and processing, VoIP network processing, and PSTN/VoIP integration. This chapter focuses primarily on these three sources of distortion. Remember, however, that PSTN-specific impairments can and do affect VoIP signal quality when VoIP networks and PSTNs are involved in the same voice signal path.

VoIP Protocols

An increasing number of voice over IP protocols provide the signaling, call services, audio/video stream transport, and, in some cases, Quality of Service needed to successfully place and answer VoIP telephone calls. Two of the more commonly implemented protocols are the complex and feature rich ITU-T Recommendation H.323 and the more simple, Session Initiation Protocol (SIP). These protocols share a common basic protocol stack as shown in Figure 3.

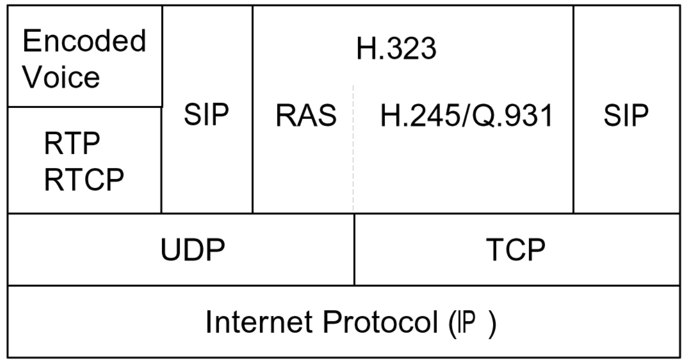

Figure 3: VoIP Protocol Stack(s)

Whether the VoIP protocol is H.323 or SIP, the protocol stack has some common characteristics. IP is carried over the physical and network transport layers. UDP (User Datagram Protocol) and TCP (Transport Control Protocol) are encapsulated into IP packets. VoIP-specific protocol packets are encapsulated into UDP or TCP depending on the particular signaling function. Digitized and packetized voice handled by RTP (Real-time Transport Protocol) and RTCP (Real-time Transport Control Protocol) is encapsulated into UDP datagrams. In the context of voice signal distortion, it is the IP/UDP/RTP portion of the stack that is the most interesting. Aspects of the VoIP signaling stack (H.323, SIP) also affect voice signal quality to some extent because characteristics of the voice channel are often defined by the signaling process when calls are setup. Finally, RTCP plays a role in maintaining the quality of a VoIP call because it can be used to gather information about delay, jitter, and packet loss.

General Noise/Distortion Issues in Voice Over IP

Before exploring VoIP-specific distortion issues and how they are dealt with, a few basic concepts should be introduced. While these concepts are not necessarily limited to voice over IP applications, they do, however, affect VoIP signal quality. In fact, early VoIP network designers have all too often failed to meet existing standards for voice quality that apply to any voice network regardless of the underlying technology[4].

General Telephony Impairments

As described in the previous section, VoIP networks almost always interface with some aspect of the public switched telephone network (PSTN). This means that most PSTN impairments can impact voice and conversation quality on interconnected VoIP networks. For example[5]:

- Signal level is arguably *the* most important factor affecting perceived voice quality. Clearly, if signal levels are too low, users cannot understand what is said, and if levels are too high, clipping (distortion) can occur.
- Circuit noise and background noise have many sources from both the analog and digital portions of a telephony network. Since much of this noise is outside the voice band, it can cause some problems for VoIP vocoders if not eliminated via adaptive noise filters or other techniques[6].
- Sidetone is in fact a form of intentional echo that occurs at the telephone set. It is designed into telephone sets so that users can regulate their own voice levels and receive the necessary feedback that the circuit over which they are speaking is still “alive”. A similar phenomenon is addressed in VoIP networks in which voice activity detectors (a.k.a. silence suppressors) are used. In this case, artificial background noise is actually injected into the voice circuit during silent periods between speech utterances to provide feedback that the circuit is still active.

- Attenuation and group delay distortion are impairments that are dependent on the frequency characteristics of a particular voice channel. Similar to analog circuit noise, attenuation and group delay distortion can cause unpredictable effects when coupled with low bit rate perceptual codecs used in VoIP[6].
- Absolute delay is the time it takes for a voice signal to travel from talker to listener, and delay values typical of PSTNs (10s of milliseconds) have little effect on perceived voice quality if there is no echo or if echo is adequately controlled. However, due to signal processing, VoIP networks introduce unavoidable delays of 50 milliseconds and above which can expose echo (as described below) and affect conversational quality.
- Talker and listener echo can be problematic in traditional PSTNs and have been around for many years. In most situations, this echo is not perceptible because it returns to the talker/listener too quickly to be distinguished from regular speech. However, when larger end-to-end delays are introduced by VoIP processing, existing PSTN echo can become a real problem.
- Quantizing and non-linear distortion occurs in digital systems when an analog signal is encoded into a digital bit stream. The difference between the original analog signal and that which is recovered after quantizing is called quantizing distortion or quantizing noise. High quality PCM encoders used in PSTNs exhibit a predictable level of quantization noise and can, therefore, be dealt with in a relatively straightforward way. However, this assumption cannot be carried into the VoIP domain because voice-band codecs (vocoders) operate on a different premise and produce non-linear distortion. Thus, in VoIP environments, quantization noise cannot always be measured or eliminated in the same way[6].

Because such PSTN impairments as described above can have an unpredictable effect on voice signals processed and transported across VoIP networks, aggressive noise reduction on circuits known to interface with VoIP networks should probably be employed.

Additive vs. Subtractive Distortion

All voice transmission systems are subject to the effects of both additive distortion (circuit noise, background noise, etc) and subtractive distortion (transient signal loss, severe attenuation, etc). For VoIP systems, however, these types of distortion are even more significant. Because perceptual codecs play such an important role in VoIP applications, noise added to the voice signal prior to encoding can have unpredictable effects depending on whether the noise has frequency components within the voice band or not and depending on the type of encoding used. In VoIP, traditional subtractive distortion such as excessive attenuation is now accompanied by the effects of packet loss where discrete portions of the encoded voice signal simply disappear. Again, due to the use of low-bit-rate codecs to preserve network bandwidth, this packet loss can be particularly disruptive. An equally interesting and related source of distortion is error concealment in which subtractive distortion such as packet loss is actually compensated for by *intentional* additive distortion in the form of predictive packet insertion [7].

Non-linearity and Time-variance

Two of the primary differences between a PSTN or PSTN-like voice channel and a VoIP voice channel are the conditions of time variance and linearity. For the most part, a PSTN voice channel is LTI or Linear and Time-Invariant. (A voice channel is more or less linear if the voice waveform that enters the system is reproduced at the receiving end. A voice channel is time invariant if, once it is set up, its transmission characteristics normally do not change over time.) A VoIP voice channel, on the other hand, is often non-linear and time-variant, a condition that makes noise reduction in a VoIP environment particularly challenging. For example, the end-to-end delay of the digital encoding/decoding scheme of a voice-over-IP channel can change during a single telephone call (time variance), resulting in changes in sound and conversational quality. Modern VoIP codecs encode and decode voice signals in non-linear ways because they strive primarily to preserve the subjective sound quality of a given voice signal rather than the objective audio waveform. Depending on how these codecs are implemented (and depending on other network conditions such as packet loss), significant levels of distortion can be introduced to the voice signal.

Human Perception's Role

It is very difficult to separate the quantification of voice quality (that is, the evaluation or measurement of noise and distortion) from the subjective experience of the human talker and listener. Voice quality can really only be judged relative to the situation being assessed and the human experience of it [8]. Voice circuit designers know that the physiology of the human ear and the psychology of human perception must be taken into account when designing voice processing and transmission systems, and therefore, when detecting and avoiding distortion. DSP (Digital Signal Processing) and voice processing design efforts increasingly concern themselves with only those parts of the voice signal likely to be perceived [9]. This selective processing ultimately reduces transmission bandwidth requirements, benefiting those who must implement voice over IP systems in bandwidth limited situations. Therefore, noise reduction and avoidance in a VoIP environment often concerns itself only with the perceptually important aspects of noise and distortion.

Obviously, the human ear can detect only those auditory signals within a finite frequency and loudness range. However, cognitive aspects of human perception play an important role in network design. For example, humans adapt to very brief auditory drop-outs without losing the meaning or content of a spoken phrase. Human listeners will perceive a particular voice sample as having worse quality if a burst of distortion occurs at the end of the sample as opposed to at the beginning of the sample [8, 10]. In addition, a listener's expectation and mood can also affect her/his assessment of voice quality. These, and other aspects of human perception play a role in noise reduction in VoIP.

Listening Quality vs. Conversational Quality

As mentioned, two of the biggest challenges facing voice over IP systems are listening/sound quality and conversational quality. These two types of quality are related because end-users often do not make a conscious distinction between them. However, the distinction between the two should be preserved. Clearly, listening/sound quality is directly impacted by noise or other types of distortion. It is also clear that a distorted voice signal will negatively impact a telephone conversation. But several telephony phenomena, further exacerbated by VoIP processing, affect the character of voice conversations without really affecting sound quality at all. These phenomena include end-to-end and round-trip network delay, delay variance (jitter), and echo. Delay and echo will be covered, along with sound quality (a.k.a. clarity) in the next section.

Primary VoIP Quality Metrics

In VoIP environments, three elements (shown in Figure 4) emerge as the primary factors affecting voice listening/sound and conversation quality.

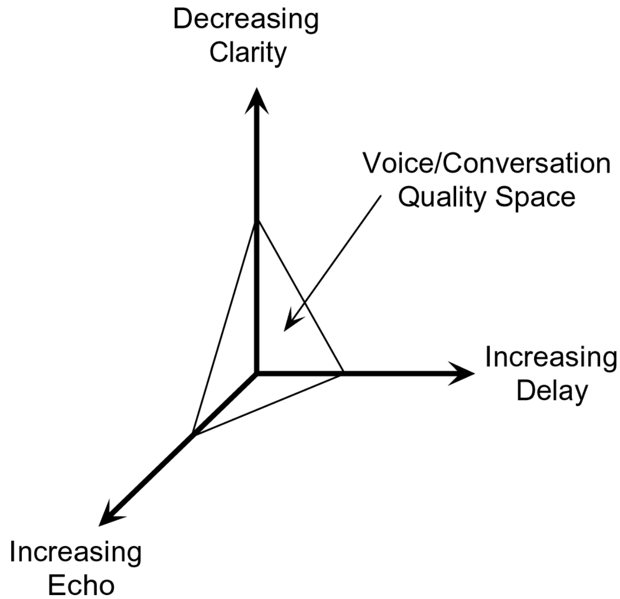

Figure 4: Voice Quality Metrics (Copyright 2000, Agilent Technologies, Inc. Reproduced with Permission)

Clarity and delay can be thought of as orthogonal in that they normally do not directly affect each other. Echo, on the other hand, affects perceived clarity and, in many cases, can be made more perceptible (and annoying) by increasing delay. While Figure 4 shows a rough relationship between clarity, delay, and echo, a strict mathematical relationship does not exist. Suffice it to say, however, when clarity is good, delay is short, and echo is reduced, overall voice quality is improved. Often, tradeoffs must be made between these parameters. For example, to decrease delay, VoIP designers can use less complex encoding schemes, but the clarity of the voice signal can suffer (i.e. coding distortion can increase).

Clarity

Clarity generically refers to a voice signal's fidelity, clearness, lack of distortion, and intelligibility. This is primarily a sound quality metric where the presence of noise and distortion plays the most significant role. Clarity is a very subjective metric and is challenging to measure, particularly in voice-over-IP applications. Traditionally, the clarity of a voice signal or voice channel has been measured subjectively according to ITU-T Recommendation P.800 resulting in a mean opinion score (MOS). MOS values can range from 1 to 5 with 5 being the best possible score. MOS, and other, more economical and objective measurement techniques that take into account human perception and physiology are described later. In a VoIP environment, clarity problems are often caused by packet loss, uncontrolled jitter, and analog circuit noise. Clarity is also significantly impacted by the codecs used on the voice channel.

Delay and Delay Variance (Jitter)

End-to-end delay is the time it takes a voice signal to travel from talker to listener. This voice signal delay is the additive result of VoIP/IP network processing and packet transport. Delay affects the quality of a conversation without affecting the actual sound of the voice signal – delay does not introduce noise or distortion into the voice channel. When end-to-end delay reaches about 250 milliseconds, participants in a telephone conversation begin to notice its effects. For example, conversation seems "cold" and participants start to compensate. Between 300 to 500 milliseconds, normal conversation is difficult. End-to-end delay above 500 milliseconds can make normal conversations impossible. In PSTNs, end-to-end delay is typically under 10 milliseconds. In VoIP networks, however, an unavoidable lower limit on end-to-end delay can be as much as 50 to 100 milliseconds because of codec operations such as packetization and compression. There is, however, one aspect of delay that has the potential to cause voice signal distortion, and that is delay variance (or jitter). Jitter is the variation in individual voice packet arrival times at voice gateways. For data networks, jitter is less of a problem because arriving packets can be buffered for longer periods of time. For real-time applications such as voice, however, some jitter can be tolerated, but more stringent upper limits must be imposed. When packets arrive outside this upper limit, the packets are discarded or ignored causing what amounts to packet loss. Packet loss directly affects voice signal distortion , and must be controlled or managed in VoIP systems to reduce its negative effect.

Echo

Echo is the sound of the talker's voice returning to the talker's ear. Echo, like delay, influences conversational quality more than it does sound quality. However, echo can significantly affect a talker's *perception* of sound quality in much the same way an interrupting burst of noise affects a listener's perception of sound quality. In the context of VoIP, echo (which often already exists on the PSTN but is rarely noticed) is made more noticeable by the unavoidable delay caused by VoIP processing. The causes and solutions to VoIP-exposed echo will be covered later.

Specific Noise/Distortion Issues in Voice Over IP

This section describes some of the sources of (and solutions to) noise and distortion that are either created or made significantly worse by VoIP technologies and implementations.

Packet Transmission

When voice is introduced into networks not originally designed for real-time audio transmission, normal network behaviors can suddenly become the source of significant voice quality impairments. Clearly, the idea of voice carried over data networks is not new. ATM (asynchronous transfer mode) networks provide services, protocols, and QoS (Quality of Service) processes designed specifically for this application. Frame Relay networks have come a long way with regard to QoS and VoFR (voice over frame relay) services.

However, IP packets can be encapsulated into a broad range of WAN (Wide Area Network) data network protocols, not all employing robust QoS and voice handling capabilities. A typical voice path could involve a number of these protocols (carrying VoIP packets). And while there have been significant advances in IP QoS, the very fact that IP was designed as a data network protocol implies that there will be voice quality problems associated with otherwise normal data network behavior.

This section begins with a brief description of physical layer bit errors and data link layer frame/cell loss, and then describes the two primary packet-based causes of distortion on a VoIP network: packet loss and jitter. A brief description of IP QoS follows.

Layer 1 Bit Stream Errors / Layer 2 Frame or Cell Loss

Bits and bytes can be errored or lost at the physical layer of the OSI (Open System Interconnection) data communications stack. Bit error rates, if below those expected of normally operating T1, E1, DS3, or 10/100 Base – T Ethernet networks will not affect the sound of a voice signal in any significant way (although a single errored sample can produce an audible click or pop). In fact, if bit error rates become high enough to be truly disruptive, chances are the integrity of the call itself is at risk. ITU-T recommendation G.821 defines levels of bit error rates for specific media and distance specifications. It is beyond the scope of this paper to describe the details of bit errors and bit error rates. However, with regard to the effect bit errors can have on VoIP applications, the following can be said[11]:

- In telephony applications, bit errors generally come in bursts and are usually caused by clock synchronization problems, electrical disturbances, and physical layer processing problems.
- Intuitively, one might conclude that evenly distributed low bit error rates would have little effect on overall voice quality. However, voice applications may routinely discard any IP packet that has even one error, particularly TCP packets. If packet sizes are large, the resulting packet loss can be debilitating.
- UDP, the portion of the VoIP stack that contains encoded voice, can be configured to tolerate bit errors. This characteristic is configured in the operating system and can reduce the packet loss associated with small numbers of bit error.

Frame or cell errors or loss at the OSI data link layer can also have a significant impact on the clarity of voice traffic carried by protocols higher in the stack. Frame relay summarily discards errored frames and relies on transport layer processes for retransmission, thus increasing jitter and ultimately increasing packet loss. ATM discards cells when QoS or traffic shaping processes are triggered to maintain agreed upon traffic levels, relying on upper level protocols to recover or retransmit lost data. Typically, if cell or frame loss at layer 2 is a problem, mere signal quality at the VoIP application layer will be the least of a VoIP implementer's worries. Call and channel reliability is the more significant issue. The good news is that, for the most part, layer 2 data protocols often provide error correction and run over very robust physical layers.

IP Packet Loss

IP, by its very nature, is an unreliable networking protocol. In its most basic (and ubiquitous) form, IP makes no delivery, reliability, flow control, or error recovery guarantees and can, as a result, lose or duplicate packets, or deliver them out of order [3]. IP assumes that higher layer protocols or applications will detect and handle any of these problems. Obviously, this kind of network behavior can be problematic for real-time VoIP. When an IP packet carrying digitized voice is lost, the voice signal will be distorted. Before describing the kinds of distortion packet loss can create, it is useful to briefly describe the causes of packet loss [11]:

- Packet Damage - Many applications will discard incoming packets, when presented with one that has been damaged. An example of packet damage is bit errors due to circuit noise or equipment malfunction.
- Network Congestion, Buffer Overflow, and IP Routing - Perhaps the largest cause of packet loss is packet discard due to network congestion. When a particular network component receives too many packets at one time, its receive buffers overflow causing packets to be discarded. IP networks also deal with network congestion by rerouting traffic to less congested network paths, but this can increase delay and jitter.

Typically, when packets are intentionally discarded due to damage or congestion, networking applications will retransmit the data. This can cause duplicate packets to be sent, can result in packets arriving too late to be used, or can cause packets to be received in the incorrect order. For non-real-time applications, this kind of network behavior is not catastrophic – in fact, it is expected. However, late or misordered packets can have, from a VoIP standpoint, the same effect as lost packets.

Determining the effect packet loss has on voice signal distortion is a complex task and depends on several variables. Fundamentally, lost packets mean lost voice information resulting in audible dropouts, pops, and clicks. Generally speaking, more packet loss means more distortion. However, the location in the packet stream at which packet loss occurs, the type of codec used (and its bit rate, packet size, compression algorithms, and error concealment methods), and the amount of jitter on the network all contribute to just how much (and how perceptible) the distortion will be. Later, codec type, packet loss rates, and jitter will be related to specific distortion measures. However, a few general thoughts are presented here:

- With regard to human perception, there is a difference between a steady state and widely distributed packet loss rate, and bursty packet loss. One might expect a steady state of annoyingly perceptible distortion would be more disruptive than an occasional burst of distortion. In addition, the location of the burst affects perceived voice signal quality as well. For example, in a sixty second call, packet loss bursts towards the end of the call are perceived to be more disruptive than those that occur near the beginning of the call.
  Low bit-rate, perceptual codecs exhibit more distortion for a given packet loss percentage than waveform codecs. For example, G.711, a waveform preserving, linear codec, encodes the most voice information (no compression, maximum number of bits for each voice sample) as compared to most other codecs. Therefore, when a G.711 codec is being used, packet loss has less effect on perceived quality than with other codecs. On the other hand, perceptual codecs (G.729, G.723, G.721) encode and decode based on perceptual relevance using compression to reduce the number of bits needed. Experimental evidence shows that lost packets can have a larger impact on the voice signal in this case.

Jitter (Varying Packet Delay)

One of the primary causes of practical packet loss is varying packet delay (jitter) that is not accounted for by network components such as VoIP gateways. In an ideal network, each voice packet arrives at its destination with the same end-to-end delay. This would allow the receiving gateway to assemble and play out the voice packets once they started arriving. As long as the end-to-end delay does not exceed about 120-180 milliseconds, end users will report no conversational impairment. However, real IP networks can and do deliver voice packets with varying end-to-end delay due to multiplexer and switch operations, queues, routing changes, congestion, and other network behavior [11]. For example, when a series of voice packets arrive at the destination 50, 58, 43, 89, 104, and 66 milliseconds respectively after each was sent, the receiving device can have problems reassembling and playing out the voice signal unless a process is in place to account for this jitter. To account for jitter, jitter buffers are implemented in voice gateways.

The key point to remember is that *jitter does not sound like anything to the end user* unless it is bad enough that packets arrive too late to be used. This late arrival time results in a situation that for all practical purposes is the same as packet loss.

Solutions to Packet Loss and Jitter - QoS

There are various ways that the negative effects of packet loss and jitter can be avoided or even eliminated. Since many VoIP calls will span wide area networks (WANs) as well as local area networks (LANs), the Quality of Service (QoS) methods mentioned next involve aspects of both WAN and LAN networking technologies. Other solutions to packet loss and jitter involve specific VoIP processing. Examples of QoS solutions include [12]:

Over provisioning involves making sure that the network has much more bandwidth capacity than it needs, thus ensuring that voice over IP traffic is never subject to congestion or other causes of packet loss and jitter. This, however, is not practical for large telephony service providers.

ATM and Frame Relay both provide QoS support, with ATM having the most robust and extensive capabilities (and often the most expensive) particularly with regard to cell/packet loss and jitter.

IP Type of Service (TOS) and filtering provides basic QoS and is built into the IP protocol. However, this method requires specific router configurations and may be unsuitable for larger networks.

Integrated Services and Resource Reservation Protocol (RSVP) permit a terminal or voice gateway to request a specific IP quality of service. However, limited packet loss and jitter control is offered.

Differential Services (including Multi-Protocol Label Switching or MPLS) is a relatively new technology that offers both packet loss control and jitter control.

In addition to the more general QoS methods mentioned above, packet loss and jitter can be dealt with by VoIP processing such as codec error concealment in which lost packets are replaced, optimum codec packet size, and intelligent jitter buffer configuration.

VoIP Processing

In addition to the voice sound quality problems caused by voice packet delivery and processing (i.e. IP network performance) discussed in the last section, voice quality is impacted by processes that are very specific to VoIP gateways and other VoIP equipment. While some of these processes are used to solve quality issues of one sort or another, they themselves can introduce voice distortion or conversation impairments.

Codec Characteristics and Performance

Perhaps the most important factor with regard to voice signal quality in a VoIP environment is the voice codec (coder/decoder) implemented in VoIP gateways/routers, IP telephones, and other VoIP terminals. In fact, it is the voice codec (along with the initial quality of the signal being encoded/decoded) that defines the best possible voice quality that can be delivered. In other words, the quality of a voice signal will never be better than what a particular codec can deliver under optimum conditions, although it can certainly be worse due to conditions such as background noise or packet loss[4].

Codec Description

Codecs digitize and packetize voice signals prior to their transmission across an IP network. Some codecs also compress the voice signal to preserve network bandwidth. Voice codecs are implemented in software and/or hardware and are often rated according to the following parameters[9]:

- Bit rate is a measure of the compression achieved by the codec.
- Delay is a measure of the amount of time a codec requires to process incoming speech signals. This processing delay is a portion of the overall end-to-end delay experienced by a voice packet.
- Complexity is an indication of a codec's cost and processing power.
- Quality is a measure of how speech ultimately sounds to a listener.

Clearly, tradeoffs must be considered when deciding which codec(s) to use in a given VoIP network or device. For example, in situations where bandwidth is at a premium, low bit rate codecs may be preferred at the expense of some signal quality. In other situations, voice quality must be preserved resulting in higher complexity, cost, and bandwidth requirements.

For telephony applications, there are three categories of codecs[7]:

- Waveform codecs are the most common type and are used ubiquitously in most PSTNs. These codecs seek to reproduce the analog signal waveform at the receiving end of the call and generally introduce the least amount of distortion and noise. They also require the highest amount of bandwidth. ITU-T's G.711 is the most common waveform codec.
- Vocoders (a.k.a. source codecs) do not seek to reproduce the analog signal waveform, but instead seek to reproduce the subjective sound of the voice signal. Vocoders are targeted strictly at voice signals, use less bits to encode the voice signal (thus, requiring less bandwidth), and are generally believed marginally suitable for telephony applications (although they have been and are used in some VoIP environments).
- Hybrid codecs are the most commonly used codecs in VoIP networks. Hybrid codecs meld the best characteristics of both waveform codecs and vocoders and also operate at very low bit rates.

Both vocoders and hybrid codecs seek, to a lesser or greater extent, to encode the perceptually relevant characteristics of a voice signal with the ultimate goal of producing good voice quality using less bandwidth than the waveform codec. Because of this, the analog voice waveform is not always reproduced. Traditionally, when the analog waveform is altered from its original shape, this is thought to represent either additive or subtractive distortion. All codecs introduce some level of distortion (e.g. quantization distortion). Whether this waveform distortion results in a degraded voice signal depends on the quality of the codec and other network conditions. It can also depend on whether the codec uses noise shaping techniques to reduce the amount of perceptual noise that is actually encoded, or error concealment to reduce the negative effects of packet loss.

Codecs, Bit-Rates, Packet Loss, Jitter, and Voice Quality

One generalization that can be made is that lower bit rate codecs introduce more perceptually relevant distortion (i.e. lower voice signal quality) than waveform codecs operating at higher bit rates [4, 13]. Figure 5 shows measurement results in which a MOS prediction algorithm – PAMS Listening Quality (Ylq) described in a later section - was used to evaluate the speech quality produced by four different codecs. As bit rates decrease, so too does voice quality (that is, perceptually relevant distortion increases).

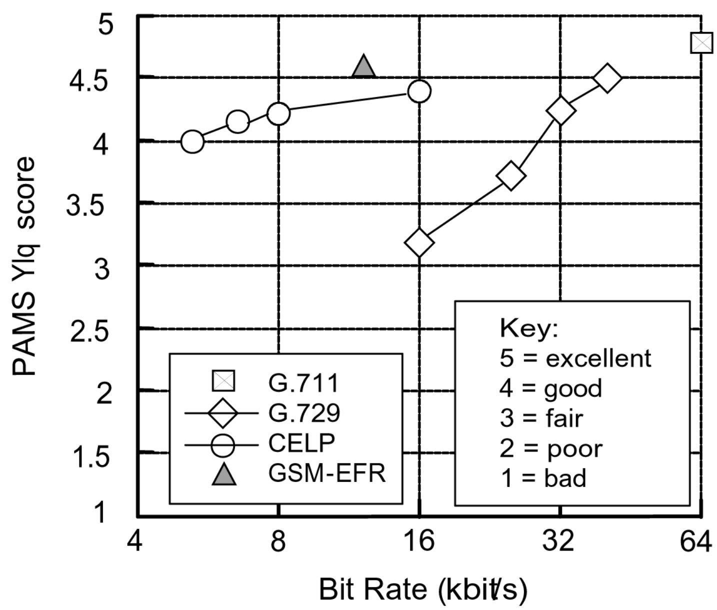

Figure 5: Listening Quality vs. Codec Bit Rate (courtesy of Psytechnics, Inc.)

Network conditions such as packet loss and jitter also affect the voice quality produced by a specific codec. It is very difficult to accurately quantify the effect packet loss and jitter will have on a particular voice signal passing through a specific codec. Predictably, as packet loss and/or jitter increases, so does signal distortion. But whether that distortion will have a significant impact on perceived quality depends on the type and location of the packet loss, whether jitter buffers are adequately compensating for varying packet arrival time, and on error concealment methods used by the codec.

Figure 6 shows the results of "distortion" measurements made on an ITU-T G.729 codec as packet loss percentages were increased [14]. The decrease of perceived signal quality as packet loss increases is consistent with other experimental results as well as with the experience of VoIP system implementers. Note, however, that at a specific packet loss percentage, the measured listening quality spans a broad range. These types of results are also shown to be true for other codec types under different experimental conditions [13].

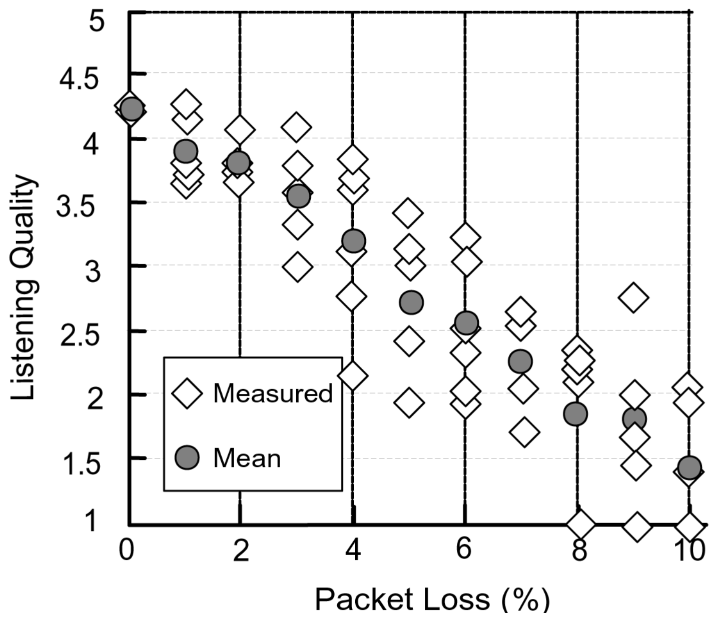

Figure 6: Listening Quality vs. Packet Loss (courtesy of Psytechnics, Inc.)

Jitter can also affect the signal quality produced by codecs. Figure 7 shows the same G.729 codec measured with increasing amounts of packet jitter. Again, the spread of measured listening quality is relatively broad, but the general trend is downward at higher jitter values and is consistent with other experimental results[13].

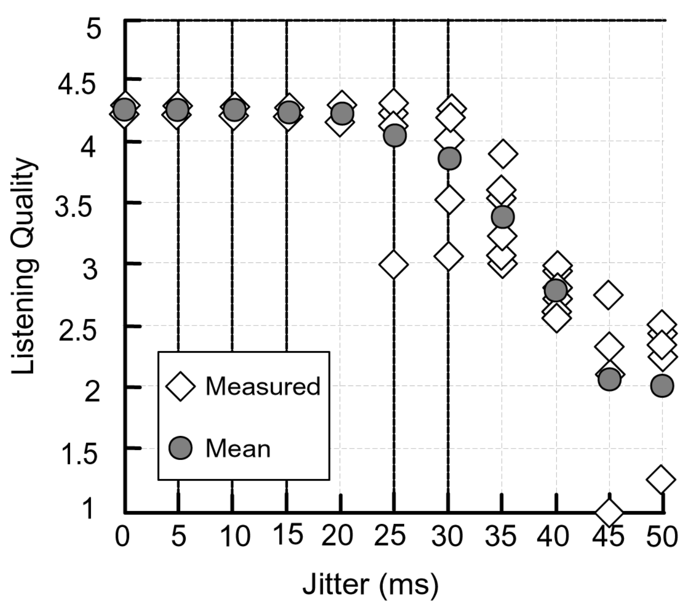

Figure 7: Listening Quality vs. Jitter (courtesy of Psytechnics, Inc.)

Jitter Buffers

As described, jitter (packet arrival variance) does not *sound* like anything as long as receiving VoIP equipment or processes can handle it. Receiving VoIP devices handle jitter by implementing a jitter buffer that can smooth out the packet delay variance so real-time applications can work properly. Basically, a jitter buffer delays the playout of individual arriving voice packets until enough of them have arrived to play out contiguous speech. This implies that jitter buffers add delay to the system.

There are two types of jitter buffers[4]:

- Static buffers provide a fixed length playout delay and any packet that arrives late is discarded. This playout delay is usually configurable but the underlying network must exhibit a predictable jitter for static buffers to be effective.
- Dynamic jitter buffers are more sophisticated in that they can adjust the playout delay based on the jitter exhibited by previous packets. This provides an automatic balancing act between avoiding lost packets and adding too much delay to incoming packets.

Jitter buffers can impact voice quality in a number of ways. For example, although most dynamic jitter buffers adjust their playout delay during periods of silence, they can and sometimes do adjust during speech utterances causing momentary distortion. Another example is simply a misconfigured static buffer that does not account for larger jitter values on a network or introduces too much delay. Again, delay and the potential for packet loss must be balanced.

Voice Activity Detection / Comfort Noise Generation

To more efficiently use bandwidth, voice-over-IP networks employ functionality referred to as silence suppression or voice activity detection. A voice activity detector (or VAD), is a component of a voice gateway or terminal that suppresses the packetization of voice signals between individual speech utterances (i.e. during the silent periods) in a voice conversation. VADs generally operate on the send side of a gateway, and can often adapt to varying levels of noise vs. voice. Since human conversations are essentially half-duplex in the long term, the use of a VAD can realize approximately 50% reduction in bandwidth requirements over an aggregation of channels.

While a VAD's performance does not affect voice signal quality directly, if it is not operating correctly it can certainly decrease the intelligibility of voice signals and overall conversation quality. Excessive front end clipping (FEC), for example, can make it difficult to understand what is said. Excessive hold-over time (HOT) can reduce network efficiency, and too little hold-over time can cause speech utterances to "feel" choppy and unconnected.

Complementary to the transmit-side VAD, a Comfort Noise Generator (CNG) is a receive-side device. During periods of transmit silence, when no packets are sent, the receiver has a choice of what to present to the listener. Muting the channel (playing absolutely nothing) gives the listener the unpleasant impression that the line has gone dead. A receive-side CNG generates a local noise signal that it presents to the listener during silent periods. The match between the generated noise and the "true" background noise determines the quality of the CNG.

VoIP / PSTN Hybrid Network Implementations

So far, noise and distortion sources have been discussed with regard to IP network behavior or VoIP-specific processing. Another important source of signal distortion and conversational quality degradation is the interoperation between a VoIP network and the PSTN.

Level / Loss Plans

PSTNs are designed with specific signal level, gain, and loss characteristics depending on where in the network the signal is measured and the type of equipment across which a signal passes[15]. VoIP networks, however, do not always adhere to specific loss plans. When voice signals pass from the PSTN to a VoIP network and back to the PSTN, they may have been attenuated and then amplified resulting in an increased noise floor. Other problems such as clipping can occur. Automatic gain control (AGC) is increasingly being used in VoIP gateways, but AGC can create noise problems of its own[4].

Transcoding / Multiple Encoding

In a pure VoIP network, a single codec type can be used at each end of a given voice call so the number of times a voice signal is encoded and decoded is limited to one and the coding and encoding schemes are compatible. This would limit the unavoidable codec and quantization noise introduced into the system and would keep end-to-end delay at a minimum. While this represents perhaps an optimum network design, it is not always practical.

Because a given voice call will likely traverse multiple VoIP systems or VoIP/PSTN hybrid networks, it is more common for multiple codecs to be used and for voice signals to be encoded and decoded multiple times[16]. In these situations, codec and quantization distortion accumulates (often in non-linear ways), attenuation distortion is multiplied, and idle channel noise is added to the signal at each coding stage. In addition to multiple codec processing, codecs of different bit rates might be used on a single voice signal. In general, when a voice signal is processed by multiple codecs (particularly of different types) along a single voice path, that voice signal's clarity cannot be better than that produced by the "worst" codec. The quality may, in fact, be noticeably worse if two or more low bit rate codecs are used. In addition, because many codecs distort speech in non-linear ways, the order in which they encode/decode speech will affect sound quality. Finally, end-to-end delay can increase significantly when more than one encoding and decoding process is in the voice path, resulting in increased echo perception and causing severe conversational quality problems.

Echo and Echo Cancellation

In most cases, echo is caused by an electrical mismatch between analog telephony devices and transmission media in a portion of the network called the tail circuit[15]. Specifically, this electrical mismatch occurs in a device called a hybrid that provides the junction between an analog four-wire E&M (ear and mouth) trunk line or digital transmission channel and an analog two-wire FXO (Foreign Exchange Office) line. The hybrid separates send-path and receive-path signals so they can be carried on separate pairs of wires or transmission channels. Because the methods used to separate send signals from receive signals are often not ideal, some of the received signal leaks onto the send-path and is perceived as echo. Another cause of echo can be acoustic coupling problems (called acoustic echo) between a telephone's speaker and microphone, for example, the hands-free set of a speaker telephone, PC terminal, or cellular telephone. Both types of echo are present on many PSTN networks, but because they are received at the talker's ear so quickly (under 30 milliseconds), they are perceived as sidetone or not perceived at all.

When a VoIP segment is introduced into the voice path, existing (and usually unnoticed) echo can suddenly become perceptible. It can be assumed that any echo generated from a near-end hybrid will still return to the talker too quickly to be perceived. However, far-end echo will be subjected to the unavoidable round-trip delay introduced into the voice path by voice-over-packet (VoP) network processing causing existing echo originating from the far-end analog tail circuit to become perceptible and even annoying to end users.

While not a source of voice signal distortion in the sense that the transmitted signal is degraded in some way, echo can definitely affect a talker's *perception* of call quality and disrupt *conversational quality*. It can even be argued that, as perceived by the talker, the sound of returning echo combined with the sound of the talker's voice constitutes a *distorted* voice signal.

To deal with unwanted echo, functional components known as "echo cancellers" are deployed in the local exchange, the VoIP-Gateway, or the VoIP terminal (PC, IP telephone, etc.), usually as close as possible to the tail circuit that generates the echo. Referring to Figure 8, an echo canceller next to the hybrid on User B's side of the network "faces out" at User B and cancels the echo of User A's voice that would otherwise be heard by User A.

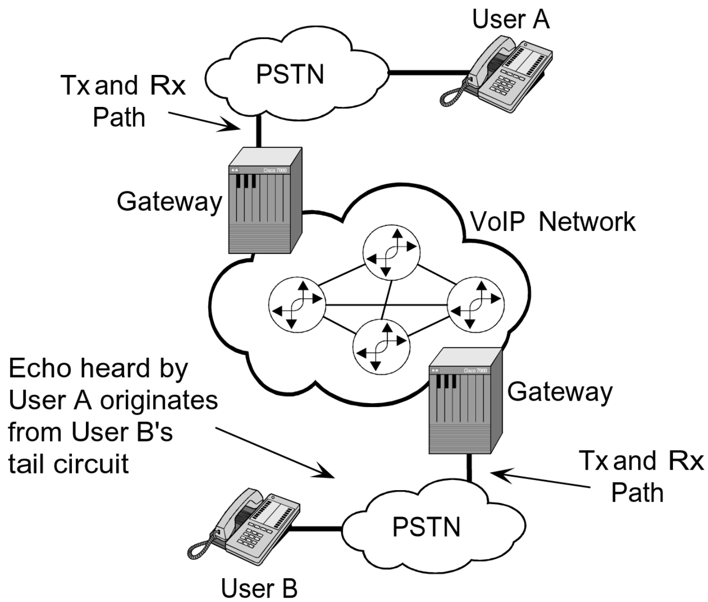

Figure 8: Perceived Echo's Origination Point (Copyright 2000, Agilent Technologies, Inc. Reproduced with Permission)

Modern echo cancellers form a mathematical model of the tail circuit they monitor, and then use this model (along with representations of the signal likely to be echoed - User A's voice, for example) to estimate the expected echo. This estimated echo is then subtracted from the speech originating on the tail circuit side of the echo canceller (User B's voice). Thus, normal speech is allowed to pass through the echo canceller, but echoes of received speech are removed. An interesting characteristic of most modern echo cancellers is their ability to "adapt" to signal and tail circuit conditions. In other words, at the start of a voice call, echo cancellers take some finite time to converge on the echo estimate that will be subtracted from far-end speech signals. For example, at the beginning of a VoIP telephone call that terminates through an analog tail circuit, echo may be perceptible but quickly diminishes as the echo canceller converges. Echo cancellers are often designed and configured to expect echo within a specific time window (echo delay) and within a specific level range (echo return loss). If the echo signal does not fit within these parameters, the echo canceller can contribute to perceived signal distortion by failing to remove the echo or by converging on inaccurate echo estimates. Voice circuits, particularly when VoIP components are used, must be intelligently designed to exhibit the correct echo return loss and echo delay.

An interesting point of failure (or poor performance) for many echo cancellers is when the talker at the far-end interrupts the near-end talker (a condition known as "double-talk"). Echo cancellers work with the assumption of a linear and time-invariant tail circuit. Double-talk, however, causes the tail circuit to appear to be non-linear, resulting in echo canceller divergence (in other words, its echo estimate becomes more *inaccurate*). In this case, the interrupting speech can become distorted.

Inconsistent QoS Implementation Across Networks

IP QoS can be an effective solution to packet loss and jitter, two important causes of noise and distortion in VoIP environments. However, because VoIP networks interoperate with PSTNs and other voice transport systems, QoS mechanisms must be defined on an end-to-end basis, requiring sufficient network resources to be provided *throughout* the voice path. This is not an overwhelming issue for an enterprise network or a single ISP environment where all resources can be administered through one network manager. But, it is almost impossible to administer when multiple ISPs or service providers are involved, as is the case in virtually every national or international long distance call. In addition, this fulfillment of QoS assumes that all equipment in the network is equally capable of identifying voice traffic and of providing the required network resources. While progress is being made on this front, end-to-end QoS is still the exception rather than the rule in today's IP networks because standards for many of these mechanisms have not been finalized and implemented by equipment manufacturers.

Measuring Noise and Distortion in a VoIP Environment

To reduce or avoid noise and distortion in a VoIP network (or any network for that matter), it is important to be able to characterize it or measure it in some way. Traditionally, voice signal quality testing techniques involved comparing waveforms on a screen, and measuring signal-to-noise ratio (SNR) and total harmonic distortion (THD) among others. These and other linear measurements are useful only in certain cases because they assume that changes to the voice waveform represent unwanted signal distortion. These testing methods also assume that telephony circuits are essentially linear and time-invariant. With VoIP and other voice-over-packet networks, particularly when they use low bit rate speech-codecs such as G.729 and G.723.1, neither waveform preservation nor circuit linearity can be assumed. Because of these conditions, specialized testing methods are often used.

VoIP Network Measurement Concepts

Before describing some of the more common measurement techniques used in voice over IP, general measurement concepts need to be covered. While some of these concepts apply equally well in other telephony environments, they are particularly important in VoIP.

Passive Monitoring vs. Active/Intrusive Testing

VoIP network testing is similar to other data and telecommunications testing in that it consists of passive monitoring and active/intrusive testing:

- Passive monitoring is a testing method in which the test device or process "listens" to some aspect of the voice traffic (digital or analog) to gather statistics and perform various types of analysis. Passive monitoring is non-intrusive and does not affect voice traffic or network behavior. It is often used in digital environments in which information encapsulated in frames, cells, or packets can be used to alert test personnel of a problem, or can be analyzed later to determine problem causes and identify traffic trends. Strictly speaking, subjective testing such as MOS (described later) can also be considered passive monitoring. Passive monitoring is often coordinated from 24x7 network operations centers (NOC), and is performed by those tasked with keeping an installed network up and running.
- Active/intrusive testing usually consists of injecting traffic of some type onto the voice channel and analyzing either the effect the traffic has on the channel or the effect the channel has on the traffic. This "energetic" approach to noise and distortion testing usually requires more sophisticated test equipment and software capable of emulating VoIP processes. Active testing is often performed by those responsible for new VoIP devices and software who do their work in research and development labs. Active testing is also useful when isolating the causes of noise and distortion.

Call Setup, Call Completion, and Services Testing

An important area of VoIP operations and performance that must be tested involves the signaling that occurs to establish, maintain, and disconnect VoIP telephone calls. Metrics include percentages of call success/completion, call services validation, call setup times, and so on. This aspect of VoIP operations has little direct effect on voice signal quality. However, "negotiations" occur during some call setup processes between VoIP entities which can result in a noise or distortion baseline. For example, SIP signaling protocols negotiate codecs and other channel characteristics. Protocol analyzers that can deliver data stream decodes are often used for this type of testing.

Packet Performance Testing

Given the impact packet loss and jitter have on a voice signal carried across a VoIP network, it is clear that packet delivery performance must be tested. Test methods can range from monitoring actual IP traffic to find evidence of packet loss and jitter, to injecting into the network under test specific packet streams with specific transmission and payload characteristics. Data communications test solutions that provide VoIP decodes, RTP and RTCP monitoring, and general IP traffic analysis capabilities represent perhaps the best ways to measure packet performance in a voice over IP environment.

Sound Quality, Distortion, Noise Testing

VoIP testing must also include a direct measure of sound quality, noise, and distortion. While VoIP signaling and packet performance are often measured at network interfaces within the VoIP network itself, sound quality measurements are performed from the perspective of the end-user of the telephony system. In other words, the quality of the signal received at the telephone set is what must be measured because this is what the user of the system will experience. Test devices that can transmit, receive, and analyze actual voice signals are preferred, although some testing methods use voice-like signals that emulate the frequency characteristics of voice.

Subjective Testing

Because of the subjective nature of voice signal quality, and because traditional audio measures are not always useful, new methods have been developed to evaluate voice clarity in a voice-over-packet environment. Early methods included MOS or mean opinion score, based on the ITU-T P.800 recommendation. This method requires that relatively large numbers of human listeners rate voice quality as part of a controlled and well-defined test process. The advantage of this method is that clarity evaluations are derived directly from the individuals who experience a voice call. Another advantage is the statistical validity provided by numerous evaluators. However, MOS evaluations can be very expensive, difficult to repeat when new telephony products need to be tested, and time consuming. Because of this, software or hardware based predictive methods have been developed to provide objective and repeatable measurement results.

Objective / Predictive Testing

In recent years, algorithms have been developed that can predict MOS results, avoiding some of the disadvantages of full-blown MOS testing. To be successful, these algorithms must evaluate the quality of voice signals in much the same way that non-linear codecs encode and decode audio signals. That is, they evaluate whether a particular voice signal is distorted with regard to what a human listener would find annoying or distracting. Typically, these algorithms compare 'clean' test signals (either actual voice signals or special voice-like signals) to more or less distorted versions of the same signal (having passed through some communications system). Using complex weighting methods that take into account what is perceptually important, the physiology of the human ear, and cognitive factors related to what human listeners are likely to notice, these algorithms provide a qualitative score that often maps closely to MOS. Two very important clarity algorithms in use today are:

- Perceptual Evaluation of Speech Quality (PESQ) is based on the ITU-T P.862 standard that defines the algorithms used to compare reference speech samples with test samples to measure quality degradation due to distortion. PESQ replaces a previous perceptual quality algorithm called Perceptual Speech Quality Measure (PSQM) which was based on P.861. Perceptual Analysis Measurement System (PAMS) is an algorithm developed and licensed by Psytechnics, Incorporated that compares speech-like samples to obtain listening effort and listening quality scores [17].
- Both PESQ and PAMS produce MOS-like scores as well as high resolution disturbance values and error surfaces that allow testers to identify distinct distortion events including packet loss, transient noise spikes, and VoIP processing problems such as VAD front-end clipping. Figure 9 shows PESQ measurement results in an implementation produced by Agilent Technologies.

Figure 9: PESQ Measurement Results (Copyright 2000, Agilent Technologies, Inc. Reproduced with Permission)

Another approach to predicting perceived voice quality involves passively monitoring IP traffic to determine packet loss, jitter, and error burst characteristics. These metrics can then be analyzed mathematically in conjunction with known VoIP network characteristics such as delay and codec type, and human cognitive factors to ultimately produce a MOS estimation. Non-intrusive measurement techniques of this sort can be embedded into VoIP equipment or test equipment with relative ease, and can provide perceptually relevant distortion measures without producing additional network traffic [10].

Agilent Technologies IP Telephony Testing Solutions

Agilent Technologies offers complete IP telephony test solutions that enable network operators to install and maintain voice services with the highest level of quality on data networks. Agilent's IP telephone solutions (along with other Agilent test products) offer active and non-intrusive voice quality testing and signaling analysis to address the testing needs described in this paper. These solutions include:

- The Agilent Voice Quality Tester (VQT) is a comprehensive and objective voice quality test system that enables the design, deployment, and operation of voice services on next generation networks such a VoIP and Voice over ATM. The VQT provides robust and reliable measurements that go beyond giving test scores. It provides detailed scoring analysis that exposes the impairments to voice quality. The VQT offers testing and analysis for voice quality using PAMS, PSQM+, and the current ITU P.862 recommendation, PESQ. It also includes testing for voice delay, echo, silence suppression, DTMF, and signal loss.

- The Agilent Telephony Network Analyzer (TNA) simplifies and expedites the resolution of quality and signaling problems in IP telephony networks. It provides simple and precise diagnostics of VoIP Quality of Service (QoS) through non-intrusive measurements, including new non-intrusive voice quality measurement technology known as predictive MOS. It also provides simplified troubleshooting of call signaling and control through embedded expert analysis of VoIP protocols such as SIP and H.323. As an Agilent Network Analyzer solution, the TNA includes complete layer 1-7 testing over LAN, WAN, and ATM networks.

Summary

Reducing noise and distortion in a VoIP environment requires not only an understanding of traditional voice signal characteristics, processing, and transmission, it also requires an understanding of IP network behavior, VoIP-specific processing, and the interaction between emerging VoIP systems and existing telephony infrastructures (i.e. the PSTN). Noise reduction, therefore, involves making design and implementation decisions that balance desired voice quality with network capacity and cost.

To design and implement VoIP systems effectively, traditional voice and network testing must be augmented by specialized and targeted testing that takes into account the characteristics of voice-over-IP systems. Agilent's powerful IP telephony test products address these needs.

References

[1] McPhillips, E., *Voice Over IP Equipment and Service Markets*, HPL-IRI-1999-005, Hewlett-Packard, 1999, 3.

[2] Douskalis, B., *IP Telephony: The Integration of Robust VoIP Services*, Prentice Hall, Upper Saddle River, 2000.

[3] Minolli, D. and Minolli, E., *Delivering Voice Over IP Networks*, John Wiley & Sons, New York, 1998.

[4] Reynolds, R. J. B. and Rix, A. W., *Quality VoIP - an engineering challenge*, BT Technology Journal, Vol.19 No.2 April 2001.

[5] *Effect of Transmission Impairments*, ITU-T Recommendation P.11, March 1993.

[6] Bellemy, J. C., *Digital Telephony – 3rd Edition*, John Wiley and Sons, New York, 2000, chap. 3.

[7] Collins, D., *Carrier Grade Voice Over IP*, McGraw-Hill, San Francisco, 2001, chap. 3.

[8] Moller, S., *Assessment and Prediction of Speech Quality in Telecommunications*, Kluwer Academic Publishers, Boston/Dordrecht/London, 2000, 116-117.

[9] Madisetti, V. K. and Williams, D. B., Editors, *The Digital Signal Processing Handbook*, CRC Press, Boca Raton, 1998, chap 45.

[10] Clark, Alan, *Passive Monitoring for Voice Over IP Gateways*, Telecommunications Industry Association, TR41.4-01-02-068, February, 2001.

[11] Siegel, E. D., *Designing Quality of Service – Solutions for the Enterprise*, John Wiley & Sons, Inc., New York, 2000, chap. 4.

[12] Siegel, E. D., *Designing Quality of Service – Solutions for the Enterprise*, John Wiley & Sons, Inc., New York, 2000, chap. 7.

[13] Berger, J., Klaus, H., Kliche, I., Scheerbarth, T., Gierlich, H. W., and Kettler, F., *Report of 1st ETSI VoIP Speech Quality Test Event*, ETSI TC STC #10 (Tiphon #22), March, 2001.

[14] Reynolds, R., *IP, Voice, Quality, and Their Relationships*, informal Psytechnics Inc. report, June 2001.

[15] Green, J. H., *The Irwin Handbook of Telecommunications – 3rd Edition*, Irwin Professional Publishing, Chicago, 1997, chap. 2.

[16] *Subjective Performance Assessment of Telephone-Band and Wideband Digital Codecs*, ITU-T Recommendation P.830, February 1996.

[17] Rix, A. W., and Hollier, M. P., *The perceptual analysis measurement system for robust end-to-end speech quality assessment*, IEEE ICASSP, June 2000.

Disclaimer

Notes

www.agilent.com

Oneline assistance:
www.agilent.com/find/assist
By internet, phone or fax, get assistance with all your test and measurement needs.

Australia	1800 629 485
Austria	0820 87 44 11
Belgium	+32 (0) 2 404 9340
Brazil	+55 11 4197 3600
Canada	877 894 4414
China	800 810 0189
Denmark	+45 70 13 15 15
Finland	+358 (0) 10 855 2100
France	+33 (0) 825 010 700
Germany	+49 (0) 1805 24 6333
Hong Kong	800 930 871
India	1600 112 929
Ireland	+353 (0)1 890 924 204
Israel	+972 3 6892 500
Italy	+39 (0)2 9260 8484
Japan	0120 421 345
Luxembourg	+32 (0) 2 404 9340
Malaysia	1800 888 848
Mexico	+52 55 5081 9469
Netherlands	+31 (0) 20 547 2111
Philippines	1800 1651 0170
Russia	+7 095 797 3963
Singapore	1800 375 8100
South Korea	080 769 0800
Spain	+34 91 631 3300
Sweden	0200 88 22 55
Switzerland	Italian 0800 80 5353
Switzerland	German 0800 80 5353
Switzerland	French 0800 80 5353
Taiwan	0800 047 866
Thailand	1800 226 008
United Kingdom	+44 (0) 7004 666666
USA	800 452 4844

Product information subject to change without notice.

Printed in US, April 23, 2003

5988-9345EN

Together with Agilent, gain the Extreme Productivity Improvements that your business demands!

www.agilent.com/comms/XPI

Agilent Technologies