VoP Voice Quality

Infineon Technologies AG
August, 2005

David Frankel

revised by
Stephan PrueckImayer

Table of Contents

VoP Voice Quality

Designing & Verifying Quality in Voice-over-Packet Systems

1 INTRODUCTION

Infineon has a long history of delivering building blocks allowing telephony systems of the highest quality. That quality is measured not only in the reliability of the system, but in the voice quality enjoyed by its end-users.

The introduction of packet technology into next-generation telephony systems brings a number of significant challenges which, if not dealt with appropriately, can result in significant compromises to voice quality. Infineon is committed to continuing our tradition of delivering the highest level of voice quality even when leveraging the significant advantages of packet technology. Infineon subsystems allow next-generation packet telephony systems to achieve the same level of voice quality as traditional wireline systems.

Infineon's lengthy history in the nuances of high-quality voice has been brought to bear on these considerations. In the sections that follow, we present an explanation of each item and discuss how it is addressed in subsystem and system design. We establish criteria which, when met, will demonstrate proper implementation of the design. Then we present test scenarios that allow verification of subsystem and system operation.

From a voice quality perspective, packet technology has a great impact on several key subsystems, but it also ripples through the entire system with cumulative effects. In order to manage this significant challenge, we have organized our discussion hierarchically. We start with subsystems that can be isolated – echo cancellers, vocoders and tone detectors -- that can be readily specified and measured. Then, we look at factors that cross subsystem boundaries – latency, packetization, and packet loss – and highlight design criteria and measurement techniques for these factors. Figures 1 and 2 below may be helpful to identify how the various components of a Voice-over-Packet network interact.

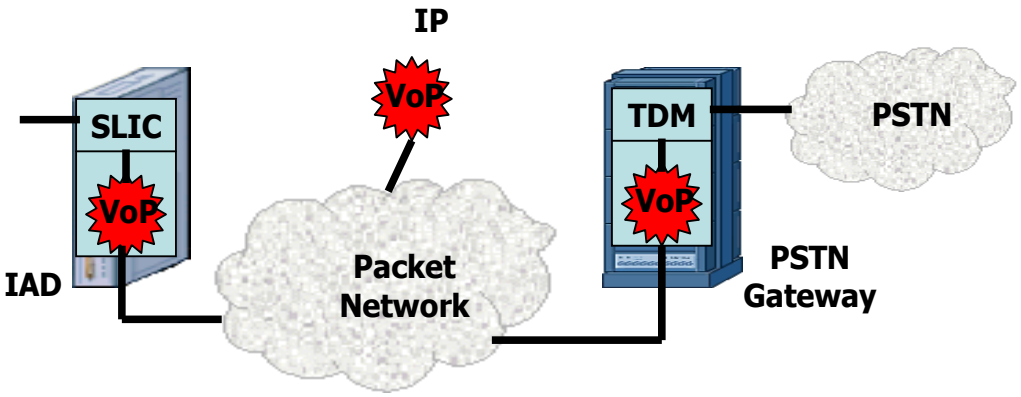

FIGURE 1. **Example of a Voice-over-Packet deployment**

Circuit Interface	**To SLIC (analog phone), handset, or TDM trunk**
Echo Canceller	**Removes locally-generated echo**
Tone Detector Tone Generator	**Identifies and generates tones (e.g. DTMF, Fax, Dial-Tone)**
Vocoder	**Encodes, compresses & decompresses audio signal**
Packet Processor	**Encapsulates audio samples into packets**
Packet Interface	**Connects to ATM or IP network; buffers packets**

FIGURE 2. **Elements of the Voice-over-Packet (VoP) functionality**

2 ECHO CANCELLATION

Echo cancellers have been part of the voice network for decades, but they take on an expanded role in voice-over-packet (VoP) systems. In the traditional voice network, a talker's voice could be reflected back toward the talker by a number of elements in the wired network. Most common was "hybrid echo" where two-to-four-wire conversion performed in central offices and telephone sets introduced an impedance mismatch that resulted in a reflection of part of the signal. The result was (and is) that the talker hears his own voice. However, in the traditional network, on local calls, the delay was so brief (one or two milliseconds) between the time the words were spoken and when they returned to the speaker's ear that the echo was blended with the sidetone produced purposefully by the telephone set and was ignored. Only in long-distance connections, where speed-of-light considerations meant that the round-trip delay could approach 20 milliseconds or more, resulting in a perceptible echo that became

increasingly annoying as the delay grew and the energy level increased. (It is a combination of these two factors – delay and energy – that determine the impact of echo. Early echo control mechanisms involved simple attenuation of the entire signal path, ideally reducing the weaker echo signal to the point that it would go unnoticed while still allowing the stronger signal from the far-end talker to be heard.)

Packet networks, especially when extended close to an end-user, almost always introduce enough additional latency into the connection that echo cancellation is required on every call, even when the connection is across town or within the building. (A thorough discussion of latency sources appears in a later section.) This means that the echo canceller becomes an important consideration for voice quality whenever the phone is in use – not just on selected calls.

In most cases, each call requires two echo cancellers (often each called “half cancellers”). One canceller is “pointed” towards the “local” user (that is, the one directly connected to the packet-voice system). This canceller eliminates any reflections generated by hybrid or other effects, benefiting the talker at the far end (that is, it prevents the far-end talker’s voice from being echoed back). The other canceller is pointed towards the remote user (who is typically connected to the PSTN); it cancels any echoes generated in the traditional network.

Ideally, the canceller is located as close to the echo source(s) as possible, prior to any vocoding. This allows the algorithm within the canceller to most accurately capture the originating signal on the way towards the echo point, and to best identify and eliminate the weaker reflected signal when it returns. The best spots for the local echo canceller are in the CODEC/SLIC or in the telephone itself (in the case of a digital telephone). When located this way, the “tail length” of the canceller (that is, the round-trip echo time within which the canceller will eliminate echo) can be relatively short. (Tail length requirements for directly-connected analog telephones are less than a millisecond, but since some existing office telephone systems introduce additional delay, local cancellers interfacing to these systems should have a tail length of at least 8 milliseconds.)

If the system design calls for a “gateway” interface to the PSTN, the “remote” canceller can be located in that device. This is analogous to how cancellers are used in the long-distance network, where they are typically installed with the tandem switches that interface to the long-haul trunks. With a gateway installation, the packet network sits between the two half-cancellers and is outside the echo path of both cancellers. This means that the remote canceller does not “see” any of the latency associated with the packet network or the packetization process, and any compression or packet loss that might occur in that network will not impact canceller performance.

It is acceptable to position the remote canceller such that the packet network is within its echo path. However, in this case, optimum performance of the canceller is only guaranteed when no or minimal compression is used, and when the packet network is operating without any packet loss. The tail length of this canceller must also incorporate the round-trip delay through the packet network.

The tail length of the remote canceller should be at least 48 milliseconds (plus packet network delay, if included), although such lengthy delays will probably never be encountered without other cancellers being invoked elsewhere in the connection. (For example, when calls are placed to and from mobile networks, which have their own lengthy delays, the mobile network includes its own canceller functionality.)

When telephone calls are made within a pure-packet environment, no remote canceller is required, since each end of the connection cancels its own local echo.

While echo cancellation is critical to quality voice transmissions, it can interfere with voice-band data transmissions used for modems and faxes. For this reason, all cancellers must monitor for fax and modem tones. The most common one is the “disable tone” issued in either

direction – a 2100 Hertz tone which is generated with most fax machines and modems. However due to the multitude of such systems already deployed, also a multitude of other tones, e.g. the DIS modulation tone, has to be detected to ensure a 100% distinction between a voice call, a fax call and a modem call.

Depending on the type of call the echo canceller has to be set accordingly - for voice calls in adaptive mode, for fax calls still enabled, but with frozen coefficients, and for modem calls completly switched off.

ITU specification G.168 is the authoritative document for echo cancellers. Because this function is so critical, and has been around for some time, there is a wealth of knowledge regarding proper operation and performance testing. G.168, revised in 2002 and 2004, provides a comprehensive set of test procedures to insure proper canceller operation, including response to disable tones in addition to performance during various speech scenarios.

Some of the G.168 tests require controlling particular aspects of the canceller algorithm, and also require a complex test set-up to generate and analyze echoes. So it can be challenging to run all the tests when the canceller is installed inside a larger system. However, Infineon has performed all these tests in its laboratories, and has a complete conformance report available for its partners.

The Infineon echo canceller is fully compliant with G.168. When designed into a system, external canceller test systems (such as offered by Sage Instruments) can be used to verify basic operations. If the system design provides for control of the canceller's "non-linear processor" and "H register" then additional tests as specified in G.168 can also be performed.

Subsystem	Echo Canceller
Specification	ITU-T G.165, G.168, G.168-2002, G.168-2004
Test Conditions	Standalone: EC algorithm is fully tested and characterized in Infineon's lab. In System: Necessary EC controls and interfaces must be provided to permit testing per G.168. No compression, transcoding, packet loss, or other impairments between EC and echo source.
Pass Criteria	Per G.168-2002 Section 6.4.2
Notes	Complete test report available from Infineon upon request.

3 VOCODER

The vocoder performs conversion between the analog interface to a conventional telephone set, and the digital network. There are numerous vocoder algorithms available; many have been standardized and several are supported in the Infineon product set and are discussed here. The vocoder consists of an "encoder" and a "decoder"; the encoder converts the incoming audio signal to a specially-coded digital pattern, and the decoder does the conversion in the opposite direction. The encoder and decoder work in pairs. In a pure VoP system, a packet-enabled telephone set would have an encoder that would send encoded audio to a

telephone set at the other end of the connection, which would then use its decoder to convert it back; another encoder/decoder pair would handle the audio signal in the opposite direction.

The traditional PSTN uses a Pulse Code Modulation (PCM) vocoder. One variant, called mu-law, is used in North America, and another, called A-law, is used in the rest of the world. Both variants are standardized by the ITU-T in specification G.711. These PCM vocoders sample the analog signal 8000 times per second, and each sample is coded in 8 bits. Thus, PCM uses 64K bits of data per second per channel of speech, in each direction.

Most other vocoders for telephony applications also use 8000 samples per second. However, the algorithms may code the individual samples using fewer bits, or they may process the samples as blocks, with the blocks consuming, in aggregate, fewer than 8 bits per sample. The most popular versions of ADPCM (Adaptive Differential PCM), standardized in G.726, use four or two bits per sample, resulting in data rates of 32K or 16K bits per second. G.728 also consumes 16Kbps; G.729A consumes 8Kbps and G.723.1 goes as low as 5.3Kbps.

Note that clock synchronization is an important factor in accurate audio reproduction. If the input signal is being sampled 8000 times per second, those samples must be played out at that same rate. A clock difference between the sampling and playout points will introduce a frequency shift in the audio. It can also result in underflow or overflow of buffers, as packets will be arriving too slowly or too quickly compared to the playout rate. The PSTN uses an extremely sophisticated and accurate set of clock synchronization techniques, and any VoP design must include consideration of network clocking.

Selection and implementation of vocoders can impact voice quality and other attributes of a telephony service, so it must be done carefully. Several of the potential issues associated with non-G.711 vocoders are discussed below.

Voice quality: Vocoder algorithms may reduce fidelity or introduce artifacts that can be noticed even by untrained listeners. Some algorithms are “tuned” to particular attributes of speech, and may perform better or worse depending on characteristics of the speaker (such as gender, pitch, or even language).

Tone transmission: Dual-Tone Multi-Frequency tones (DTMF tones, or “touch tone”) and other signals are important for voice-mail, interactive voice response, and other services. Most of the ITU standardized vocoders can transmit these tones without issue. If the tones cannot be passed reliably through the vocoder, they must be “relayed” – that is, detected at one end, and then signaled (via a mechanism defined specifically for this purpose) to the other end, where they are accurately regenerated. The ITU has standardized DTMF in Q.23. And IETF has standardized the relay function in RFC 2833.

Fax and modem transmission: Most vocoders running at rates below 64Kbps cannot reliably transfer full-speed fax transmissions and will be challenged supporting modems (particularly those running at speeds above 4800bps). Some VoP systems support fax and modem transmissions by reverting to a PCM (64Kbps, or “clear channel”) vocoder when such a call is detected. (Detection is usually accomplished by sensing the 2100-Hz echo canceller disable tone transmitted prior to modem negotiation.) Others use a “relay” mechanism whereby the fax or modem signal is demodulated at the telephone interface, transferred over the packet network in a native (without modulation) format, and then re-modulated when placed back on the PSTN or at the terminal point. T.38 is the ITU standard for fax-relay and V.150 for modem-relay.

Other transmissions: Because some vocoders are specifically optimized for speech, they may produce noticeable impairments to other types of audio, such as music or background noise.

Delay: Some vocoder algorithms, by design, accumulate a number of samples prior to encoding. And some algorithms are particularly processor-intensive. Both of these considerations

can impact the end-to-end delay through a VoP system; delay is discussed in greater detail in a later section.

Transcoding: Most voice-over-packet implementations include an interface to the PSTN. This dictates that, ultimately, the signal must be coded in G.711. When G.711 coding is used in the VoP segment(s) of the network, the samples are simply moved, without modification, between the packet and traditional networks. However, if another coding scheme is used in the VoP segment(s), then “transcoding” is required to convert between that scheme and G.711. While transcoding to and from G.711 is relatively benign, transcoding to and from other vocoders can result in additional voice quality degradation. The degree of impairment associated with each transcoding varies depending on the specific vocoders involved. It may be useful to highlight a few transcoding examples, assuming a VoP user employing G.729:

Call to PSTN: One transcoding, between G.729 and G.711.

Call to another VoP user employing G.729, via PSTN: Two transcodings (from G.729 to G.711, and then back to G.729).

Call to another VoP user employing G.729, not via PSTN (direct connection via packet network): No transcodings.

Call to a mobile customer: Two transcodings (from G.729 to G.711, and then to the coding scheme used by the mobile system).

When assessing the voice quality associated with a vocoder, Infineon specifications assume no transcodings (or, in the case of a non-PCM vocoder, at most one transcoding to and from PCM).

Silence suppression: Some vocoders include this feature either as an integral part of the vocoder or as an option (G.729 Annex B, for example). Silence suppression consists of two parts. The “voice activity detector” (VAD) located in the encoder determines whether a speaker is active or not. If a party is not speaking, the encoder sends a message to that effect to the decoder, and ceases continuous transmission until the party starts speaking again. In the decoder, a “comfort noise generator” (CNG) substitutes a locally-generated background audio signal so that the listener still “hears” a connection.

Sometimes silence suppression introduces audio “clipping” – the VAD may take a noticeable amount of time to recognize that a party has resumed speaking and restart the coded transmission. The result can be that the first syllable of speech is missed. Also, the background audio produced by the CNG may not accurately mimic the actual sound at the source, so that a distracting shift occurs when it activates and deactivates. Some vocoders specify that the encoder periodically update several parameters about the background noise, allowing the CNG to better match reality.

Silence suppression can save significant amounts of bandwidth – 50% or even more. A call using an 8kbps vocoder might consume, with overhead, 10kbps in each direction during active conversation. But on average, with silence suppression, it might use only 5kbps. How many calls can be carried on a 64kbps link? If we try to put more than 6 calls on that link, there will be occasional periods when all parties are talking (not often, but periodically) resulting in the link capacity being exceeded, packets being dropped, and voice quality suffering. This is discussed further under the “packet loss” heading. Regardless of whether more calls are packed into a link, silence suppression can free up bandwidth for other types of packet traffic.

A further advantage of silence suppression is that it provides the perfect intervals in which automatic adjustments can be made to the jitter buffer. This is discussed in further detail in the “latency” section.

To ensure proper interoperability among encoders and decoders, all implementations must conform to detailed specifications. The standardized vocoders have, in their specifications, the

precise methodology for the encoding and decoding algorithms, and often include exact test sequences to prove a conforming implementation. All of Infineon's vocoder implementations are in strict conformance with these specifications.

In some cases, the vocoder specification includes options, usually designated as Annexes or Appendices. To insure end-to-end interoperability, system designers must be exact in identifying which vocoders and associated options they support.

Vocoder performance, under ideal conditions, can be assessed using the Perceptual Speech Quality Measurement, as defined in ITU-T specification P.861. This approach measures distortion due to quantization noise, compression artifacts, and coding errors. However, it does not measure impairments due to delay or gain and was not specifically designed to address packet-based impairments. Thus, it is a good measure of the vocoder itself. Other end-to-end quality attributes will be addressed in later tests, and the newer ITU specification P.862 (Perceptual Evaluation of Speech Quality, or PESQ) will be introduced. PSQM produces a score in the range of 0 to 6.5, with 0 being best. PSQM can be converted to the more popular Mean Opinion Score (MOS), standardized in ITU-T P.800, using the formula: MOS = 5 – (PSQM * 4 / 6.5).

Subsystem	Vocoder
Specification	Algorithms: ITU-T G.711, G.723.1, G.726, G,728, G.729A
	Voice Quality Test: ITU-T P.861
	Artificial Voice: ITU-T P.50 Male
Test Conditions	Single encoder/decoder pair
	No packet loss
	No transcoding (other than to/from G.711)
	Zero gain/loss
	Standard speed
	Constant bit rate (no silence suppression)
Pass Criteria	G.711: PSQM < 0.6
	G.723.1: PSQM < 1.8
	G.726: PSQM < 1.6 (32Kbps)
	G.728: PSQM < 1.6
	G.729: PSQM < 1.6
Notes	

4 TONE DETECTION

In a VoIP environment a multitude of tones has to be detected simultaneously, ranging from the various fax and modem tones, to alert and dialtones, to DTMF tones.

Detection of DTMF tones is a critical element in many VoP systems. A tone detector monitors the audio path, alerting the application to the presence of a valid tone. In some situations, it may be required to mute the audio path as soon as the tone is detected, so that it is not heard by the far end. To avoid a double playout the supression of the DTMF sign has to be done

faster than minimum time for a DTMF tone to be recognized as valid. The function in VoIP systems executing this is called DTMF relay with early detection and is specified in RFC 2833 for the transmission of such events.

Tone detection algorithms must trade off sensitivity and selectivity. Some DTMF sources (such as very inexpensive telephone sets) do not generate the tones precisely as specified. So a very selective detector may miss tones that should be detected. On the other hand, if the detector is relaxed to the point that it detects tones that vary significantly from the standardized parameters, it will falsely detect tones during speech – a phenomenon called “talk-off.”

As noted earlier, accurate clocking must be provided to ensure that tones are not inadvertently frequency-shifted.

Infineon’s tone detection algorithms conform to the specifications and test requirements set forth in the published standards. ITU specification Q.23 specifies the transmission requirements for DTMF, while Q.24 gives detector criteria. EIA 464 also spells out criteria for tone transmission and detection.

Subsystem	Tone Detection
Specification	EIA 464, ITU Q.23
Test Conditions	Tone detector directly interfaced to audio source (without intervening compression, packet loss, or other impairments)
Pass Criteria	Detection: Per EI 464 Section 7.1.5 & ITU Q.24 Talk-off: No more than 100 detections in Telcordia Standard Speech Tape TR-TSY-000763
Notes	

5 LATENCY

As the end-to-end delay in a telephone conversation grows, it goes from undetectable to noticeable to annoying to intolerable. One-way delays of 100 milliseconds or more will start to compromise the quality of the connection.

Any voice-over-packet implementation will introduce delay additive to that which would otherwise be present. In the traditional PSTN, the longest delays are associated with speed-of-light delay (it takes something less than twenty milliseconds for signals to travel between the two US coasts). But longer delays are associated with international calls, mobile networks, certain private office telephone systems, and other emerging subsystems. Since all of these delays accumulate, it is important that any new technology that becomes part of a telephone connection strive to keep any additional delay to a minimum.

A latency budget can be helpful in planning for the various delays that are contributed by the different elements in a VoP system, and for ensuring that the implementation matches the design expectations. Listed here are the primary contributors to latency in a VoP access system consisting of an Integrated Access Device (IAD) and a PSTN gateway connected via a

packet network. (Note that the budget should be assessed in both directions – from the Gateway towards the IAD, and from the IAD to the Gateway, as some parameters may differ. The discussion here traces the path from the IAD to the Gateway. The budgeting process should examine both the minimum and maximum delays through each element; this variability contributes to jitter, which must be addressed by the jitter buffer function.)

Echo Cancellation: The audio signal passes through the echo canceller, where any echo generated at the IAD end of the connection is removed. The delay associated with the echo canceller for Infineon systems is 0 milliseconds. For other systems, especially softcoders - systems where all VoIP functions are executed by the host processor - the delay for echo cancellation can be significant.

Encoding: The encoding delay varies depending on the vocoder. An encoder with a “frame size” of 32 samples will have to wait 4 milliseconds to accumulate those samples before it can begin its processing. Typical frame sizes are 10 ms (80 samples) for G.729A and 30ms (240 samples) for G.723.1, and a “look-ahead” feature of the algorithm may extend the delay. The other vocoders have smaller frame sizes.

Packetization: Encoded samples must be accumulated to fill the desired packet size. In ATM environments, for example, there are usually 40 to 48 bytes available for the encoded audio payload. With G.711 encoding, 40 to 48 samples would be required, resulting in a delay of 5 to 6 milliseconds. If an aggressive compression scheme were being used (such as G.729A at 8kbps), it would take 8 times longer to fill the cell payload – resulting in a delay of up to 48 milliseconds. IP packets often have even bigger payloads. Sometimes, system designers choose smaller packets or partially-filled payloads to reduce delay. Also, samples from several different calls can be multiplexed into a single packet payload, reducing delay but adding complexity.

Queuing: Typically, an IAD connects to several telephone ports as well as to one or more local area networks, and multiplexes these streams onto a single packet network link. A packet for a given voice call may have to wait until other packets, already queued ahead of it, have been dispatched. Even if voice packets are given priority and placed at the head of the queue, there may be other voice packets contending for the packet link. And if transmission of a large data packet has already started when the voice packet becomes ready, that data transmission will have to complete before the voice packet goes. All of this can result in significant variability, or jitter, in the delay associated with the voice packets. Data packets often vary in length, but in IP environments, they often range up to 1500 bytes. Over a data link operating at, say, 384kbps, a packet of that size would take over 30 milliseconds to transmit, delaying a waiting voice packet by that amount. Voice packets associated with other calls, while smaller, could result in additional delay. The minimum queuing delay would be 0, in the situation that no other packets were waiting for transmission. Thus, the delay, and the variability, can be significant, especially if the packet link is relatively slow and maximum packet sizes are relatively large. Some implementations “fragment” large data packets into smaller pieces to reduce this delay and the variability.

In ATM networks, a “Traffic Contract” specifies the attributes for each “Virtual Circuit” through the network. These contracts can be used to ensure that voice packets (or cells, in the ATM environment), are queued ahead of other traffic types. And since ATM breaks all transmissions into relatively small cells, queuing delays can be tightly controlled (and even specified as part of the Traffic Contract).

In IP networks, a number of prioritization mechanisms have been defined. Support for these mechanisms varies from network to network. “DiffServ” and “TOS” are prioritization schemes that allow packets to be marked so that they’ll be given higher priority as they move through the network. Networks that support Multi-protocol Label Switching (MPLS) permit a “flow” to be designated to receive priority treatment.

Transmission: Voice packets in an ATM environment are usually 1 cell (53 bytes), while in an IP environment, they may be two to four times that size. At 384kbps, it will take one to four milliseconds to stream such packets onto the link. Of course, with faster links (at 1 or 10 megabits per second, for example), the transmission time becomes much less significant for these packet sizes.

Speed-of-light: The speed of light doesn't distinguish between packets and other kinds of electrical transmission. In an access system, distances are usually short (<20 KM), so the time it takes to get from one endpoint to the other is under a millisecond. However, if an unusual network design dictates that packets are shipped across the country and back, significant speed-of-light delay can be incurred.

Packet Switching: Typically, packet access networks have a small number of individual segments, or hops, and any connections between switches are relatively high bandwidth (dozens of megabits or more). The switching delay within the network itself should be minimal. However, if there are many switches, and/or low-speed inter-switch links are employed, then this becomes a larger factor in the latency budget.

Jitter Buffer: As noted, the delay through the system will vary, and a buffer is required at the receiving end (the Gateway, in this case) to smooth out this variability. Establishing this buffer has the effect of imposing additional delay, which must be included in the budget. The buffer should not be larger than what is required to address the actual jitter.

An "adaptive" buffer, which automatically senses jitter and adjusts accordingly, is ideal. The adaptive algorithm monitors packet arrival times. If network conditions are such that jitter is excessive, the jitter buffer is extended to avoid "losing" packets that arrive too late. When jitter is lower, the adaptive algorithm shortens the buffer, reducing the end-to-end delay. Ideally, these adjustments are done during silent intervals and are imperceptible to the end-user. (On the other hand, fax or modem traffic may suffer with an adaptive buffer, thus requiring also a fixed jitter buffer)

In an access network (covering the last mile to the user), the network is usually carefully controlled and the jitter buffer will likely not need to exceed 50 milliseconds. On the other hand, if the packet voice traffic is traveling cross-country or intercontinental over an unmanaged network, the buffer may need to accommodate up to 200 milliseconds of jitter.

Decoder: Even with complex vocoders, the decoding process is usually very fast – less than a millisecond for DSP based system. For softcoder applications several milliseconds are required.

Echo Cancellation: The audio signal is passed through the echo canceller, which "remembers" it so that any echo can be removed from the audio traveling in the opposite direction. For DSP based systems there is no delay associated with this function..

Polling: In some implementations, there may be additional delay added due to the architecture of the subsystems within the IAD or Gateway. For example, the encoder might deposit the coded audio into a buffer, which is then polled periodically by a packet processor. The polling interval adds delay which must be included in the latency budget.

Once the system is up and running, measuring the end-to-end delay is fairly straightforward using commercial test equipment. However, if measured values are not consistent with the budget, isolating the source of the anomaly can be quite difficult. It is useful to include loopback functions at each subsystem to allow a granular assessment of delay. It may also be useful to design in other test and analysis hooks to ensure that delay can be adequately measured and monitored.

Subsystem	End-to-end System Delay
Specification	Per System Latency Budget
Test Conditions	Contributions of individual elements calculated using loopback and other instrumentation techniques.
Pass Criteria	Infineon Echo Canceller: Forward direction: 0 microseconds Reverse direction: 0 microseconds Infineon Encoders: G.711: < 250 microseconds G.723.1: <40 milliseconds (30ms frame size) G.726: < 250 microseconds G.728: < 2 milliseconds G.729A: < 16 milliseconds (10ms frame size) Infineon Decoders: G.711: < 250 microseconds G.723.1 < 1 millisecond G.726: < 250 microseconds G.728: < 1 millisecond G.729A: < 1 millisecond
Notes	Infineon's system meets and exceeds British Telecom requirements for 25 milliseconds end-to-end delay under the specified parameters

6 PACKET LOSS TOLERANCE

Packet loss is an important consideration in any VoP system. Some implementations are designed for zero packet loss during normal operation. This is accomplished by providing more capacity than could possibly be consumed even under maximum load, or by prioritizing voice traffic ahead of other traffic, or by denying new calls under certain load conditions, or through a combination of these techniques. Even when the design calls for zero packet loss, it is still important that the system recognize and log packet loss events, in order to quickly identify and rectify engineering, provisioning, or operational faults.

In other implementations, packet loss will occur by design. Some systems use packet loss as a congestion indicator, and may invoke bandwidth conserving techniques such as a lower bitrate vocoder. Sometimes, network design parameters allow packet loss up to a certain level.

Packet loss deserves special attention in a VoP system because its effects can be severe and difficult to diagnose. Usually, in a conventional telephony system, a channel is allocated to a call when the conversation begins, and that channel is dedicated for the duration of the call. If no channels are available, the call origination is denied. However, in a VoP system, it is possible for network congestion, and therefore packet loss, to occur at any time during a call. The nature and degree of packet loss can have varying effects.

Voice calls can tolerate some degree of packet loss. The human ear can “recover” some of the lost information, and certain vocoders include redundancy and packet loss recovery in their algorithms. It is important to note that “random” packet loss is much more tolerable than “burst” losses. For example, if a network is experiencing 5% packet loss, a VoP call will sound much better if one out of every twenty packets is lost, rather than a burst of 50 packets disappearing from a group of 1000. If each packet represented 10 milliseconds of speech, the latter would represent a half-second gap every ten seconds, which would be quite noticeable and annoying, and no vocoder algorithm could adequately conceal it. On the other hand, the human ear would likely not even notice a ten millisecond gap every half-second, especially with a “smart” vocoder that bridged the gap with a “best guess” audio output.

The situation with respect to fax and modem calls is much different. These transmissions are much more sensitive to packet loss, both random and burst. In addition to the loss of data, missing packets may also affect clock synchronization. Without special provisions, quality carriage of fax and modem calls will require virtually zero packet loss, and accurate network clocking.

There are several different techniques for concealing a lost packet. Some algorithms mathematically interpolate between the packet preceding the one lost and the one following (sometimes called “bad frame interpolation”). Other algorithms use “forward error correction” in which some information is duplicated from one packet to the next, allowing the receiver to reconstruct missing data when a packet is lost. Most of the vocoder standards include methods for dealing with lost packets, although there are also some proprietary packet loss concealment algorithms in use.

Assessment of voice quality in a VoP system can be accomplished using the “Perceptual Evaluation of Speech Quality” (PESQ) metric, standardized in ITU-T G.862. The PESQ algorithm is designed to factor packet loss and jitter into its assessment.

Because PESQ is a comprehensive assessment, the score it produces is influenced by virtually all attributes of the VoP implementation. Before embarking on PESQ, it is important that the individual elements of the system have been verified to meet design parameters.

When using PESQ (or performing other system-level tests), it is important that the packet network be carefully controlled and monitored. This section has focused on packet loss; the degree and type of loss must be according to test guidelines. In addition, other packet network anomalies such as jitter will influence the PESQ score and must be considered. Traditional factors in voice quality, including the use of compression algorithms and the introduction of noise, will also be reflected in the PESQ results.

Subsystem	Packet Loss Tolerance
Specification	ITU-T G.7XX ITU-T P.862.1
Test Conditions	Test Conditions as specified in ETSI Speech Event 2004 (C1 to C6)

Subsystem	Packet Loss Tolerance
Pass Criteria	Voice quality per P.862.1 (average values): G.711: 4,36 G.723.1 (5,3kbit/s): 3,73 G.723.1 (6,3kbit/s): 3,86 G.726 (16kbit/s): 3,06 G.726 (32kbit/s): 4,24 G.728: 4,13 G.729AB: 3,86 G.729E: 4,17
Notes	Infineon Voice Quality Test Report available on request

7 SUMMARY

While voice quality has always been a concern of telephony system designers, voice-over-packet technology introduces additional considerations. Infineon's VoP building blocks are designed and implemented to insure that voice quality is uncompromised. Carefully designed VoP systems, incorporating Infineon VoP technology, will meet or exceed voice quality expectations established by traditional telephony systems. Infineon's adherence to established industry standards provides the metrics by which our quality can be quantitatively measured.