

VoIP: Better than PSTN?

by Jan Linden, PhD, Vice President of Engineering,
Global IP Sound, Inc.

Many factors in the past have slowed the anticipated growth of Voice over IP (VoIP). Now, VoIP solutions that achieve quality and reliability, close to what we are used to from the Public Switched Telephony Network (PSTN), are emerging as the market is quickly growing. However, as will be shown in this article, there is no reason to limit the expectations to achieve only the same level of quality as in PSTN. It is quite well known that by deploying wideband voice codecs much better quality can be achieved. However, a little known fact is that there are ways to achieve better quality than a standard PSTN solution, even when using narrowband codecs. For example, the full available spectral bandwidth is not typically used in traditional PSTN solutions, something that can easily be done in a VoIP system. But implementing a wideband codec or expanding the bandwidth of narrowband codecs does not automatically guarantee great quality. There are many potential pitfalls when deploying VoIP. In this article we also discuss implementation issues related to VoIP that will impact the final voice quality.

Speech Signals and Speech Coding

Sampled digital signals can contain frequency content up to half the sampling frequency. Typically, a young adult has a hearing span from about 20 to 20,000 Hz. Consequently, the sampling frequency of CD audio was chosen to be 44.1 kHz, which is more than double that of the highest frequency perceivable by most humans.

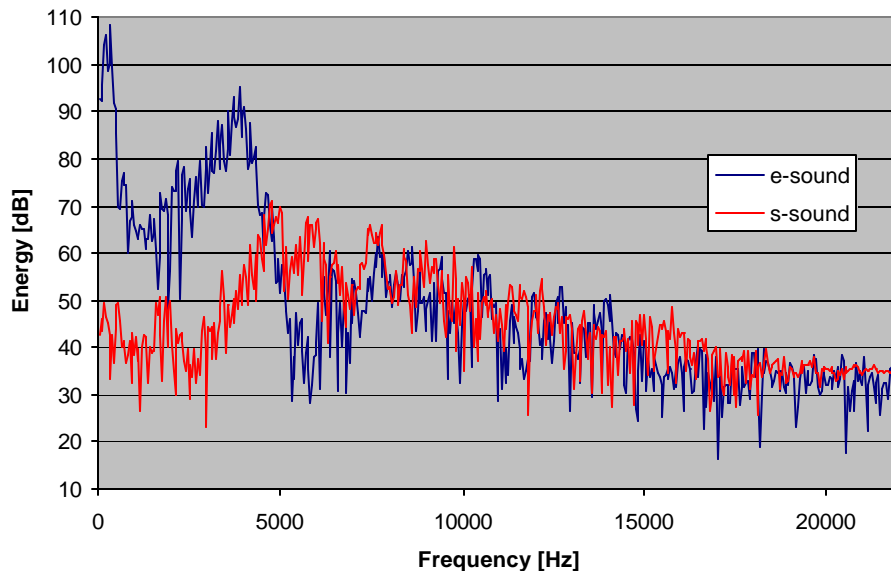


Fig. 1: Energy Spectrum Of Two Speech Sounds

Vowel 'e' has most of its energy in the lower frequency range while fricative 's' has its energy more evenly distributed with most energy in the mid frequency range

In the spectral contents of two typical speech sounds (see Fig. 1) the vowel 'e' and the fricative 's' the energy of speech signals is concentrated at fairly low frequencies. In fact studies have shown that a speech signal can be band limited to 10 kHz (20 kHz sampling) without affecting its perception 0. Consequently, bandwidth is one of the most important parameters to affect the quality of a telephone conversation.

[Sound](#) Sample 1: First: Speech sampled at 44.1 kHz. Second: Speech sampled at 20 kHz

In order to achieve a low transmission rate a speech compression technique, referred to as speech coding, is usually deployed. The ultimate goal in speech codec design is to achieve the best possible quality at the lowest possible bit rate, with constraints on complexity and delay. One obvious way of lowering the bit rate is by choosing a lower sampling frequency. By far the two most popular choices of sampling frequency are 8 and 16 kHz, referred to as narrowband and wideband codecs, respectively.

Lowering the bit rate by deploying powerful coding techniques will result in higher distortion but, by exploiting knowledge about the human auditory system, techniques that mask the distortion can achieve high perceptual quality at very low bit rates. Depending on the coding algorithm, the resulting distortion has very different characteristics. Several low bit rate codec standards such as ITU G.729 not only add noise but also distort the spectral characteristics of the signal. This is obvious if the coded signal is compared with the original signal. Without the original as a reference, however, quality is usually graded as acceptable.

[Sound](#) Sample 2: First: Narrowband speech. Second: Encoded with G.729.

In order to achieve high compression and hence low bit rate, most speech coding algorithms make the assumption that the input signal is pure speech. However, in many realistic scenarios, background sounds or other types of noise are added to the speech input, resulting in poor speech quality. There are two ways of handling this challenge: one is to make the coding technique more robust against different types of input signals, and the other is to try to remove as much of the background noise as possible before the input signal is fed to the speech coder. Inevitably, such noise canceling techniques will, in addition to suppressing the noise, also distort the speech signal itself.

Other types of input signals, such as music, usually will not sound well in a low bit rate codec. On the other hand, even with a very high quality audio codec, quality will not be very good in a narrowband scenario because music signals typically have a much wider audio bandwidth than speech signals.

[Sound](#) Sample 3: First: Music sampled at 44.1 kHz. Second: Music sampled at 8 kHz. Third: Music encoded with G.729.

Limitations of PSTN

Legacy telephony solutions are narrowband, which seriously limits the achievable quality. Wideband codecs could potentially be used in digital telephone systems, but this has never been practical enough to gain any real interest.

In fact, in traditional telephony applications, the speech bandwidth is restricted much more than the inherent limitations of narrowband coding. Typical telephony is band limited to 300 Hz to 3400 Hz. This bandwidth limitation explains why we are used to expect telephony speech to sound weak, unnatural, and lack crispness.

[Sound](#) Sample 4: First: Speech sampled at 44.1 kHz. Second: Narrowband speech. Third: Telephony band speech.

Most phone lines connected to a household are traditional two-wire copper cables. Pure digital connections are typically only found in enterprise environments. Due to poor connections or old wires, significant distortion is often generated in the analog part of the phone connection, a type of distortion that is entirely absent from VoIP implementations. The cordless phones so popular today also generate significant amounts of analog distortion due to radio interference and other implementation issues.

The Promise of VoIP: Better Than PSTN Narrowband

It is clear that there are some significant sources of quality degradation in today's PSTN and VoIP offers a way to avoid such distortion and even achieve much better quality than we have become accustomed to.

As we have mentioned previously, even without changing the sampling frequency, the bandwidth of the speech signal can be enhanced over telephony band speech and it is possible to extend the lower band down to about 50 Hz, improving the bass of the speech and having a major impact on the naturalness, presence and comfort of the conversation.

Extending the upper band to almost 4 kHz (a slight margin for sampling filter roll-off is necessary) improves the naturalness and "crispness" of the sound. All in all a fuller, more natural voice and higher intelligibility can be achieved just by extending the bandwidth within the limitations of narrowband speech. This is the first step toward "face-to-face" communication quality offered by wideband speech (Sound Sample 4, again).

In addition to the extended bandwidth, there are fewer sources of analog distortion in VoIP resulting in the possibility to offer significantly better than PSTN quality.

Even though this improvement is clearly distinguishable, far better quality can be achieved by taking the step to wideband coding.

Wideband Coding

One of the great advantages of VoIP is that there is no need to settle for narrowband speech. In principle CD quality would be a reasonable alternative, allowing for the best possible quality but a high sampling frequency will result in higher transmission bandwidth and puts tough requirements on hardware components such as microphones, loudspeakers, and ADCs. As previously mentioned, for speech a sampling frequency of 10 kHz suffices to offer excellent quality but 16 kHz has been chosen in the industry as the best trade-off between bit rate and speech quality for wideband speech coding.

By extending the upper band to 8 kHz significant improvements in intelligibility and quality can be achieved. Most notably, fricative sounds such as 's' and 'f,' which are very hard to distinguish in telephony band situations, sound very natural in wideband speech.

[Sound](#) Sample 5: First: Narrowband speech. Second: Wideband speech.

Many hardware factors in the design of VoIP devices also affect speech quality. Obvious examples are microphones, speakers, and ADCs converters. These issues are all very similar to challenges well known from designing devices for regular telephony, and as such are well understood. However, most regular phones sold today do not offer high quality audio due to cost saving designs. Hence, this is another area of potential improvement over the current PSTN experience.

Our focus here is on the pure speech quality difference between VoIP and PSTN. There are numerous other reasons why VoIP is rapidly replacing PSTN and extending the uses for voice communications. Cost and flexibility are two major reasons and, in addition, the convergence of voice, data, and other media presents a field of new possibilities. A great example is web collaboration, which combines application sharing, voice and video conferencing. Each of the components, transported over the same IP network, enhances the experience of the others.

VoIP Challenges

Thus far we have concentrated on showing the potential of VoIP solutions to provide better than PSTN quality. However, is there no truth in the common belief that VoIP quality is usually inferior to that of PSTN? It is true that quality will suffer if the challenges related to IP network transportation are not handled properly. The good news is that if these issues are properly handled, quality does not have to suffer because of the IP network characteristics.

Three major factors associated with packet networks have a significant impact on perceived speech quality: delay, jitter, and packet loss. All three factors stem from the nature of a packet network, in which there is no guarantee that a packet of speech data will arrive at the receiving end as expected, or even that it will arrive at all. These network effects are the most important factors distinguishing speech processing for VoIP

from traditional telephony. If the VoIP solution cannot cope with network degradation in a satisfactory manner the quality can never be acceptable. It is therefore of utmost importance that the characteristics of the IP network are taken into account in the design and implementation of VoIP products, as well as in the choice of components such as the speech codec.

An extremely important quality parameter is the transmission delay. If the latency is high, it can severely impact the quality and ease of conversation. The two main effects caused by high latency are annoying echo and talker overlap, both of which can significantly impact the perceived conversation quality.

In traditional telephony, long delays are basically only experienced for long-distance calls and calls to mobile phones. This is not necessarily true for VoIP.

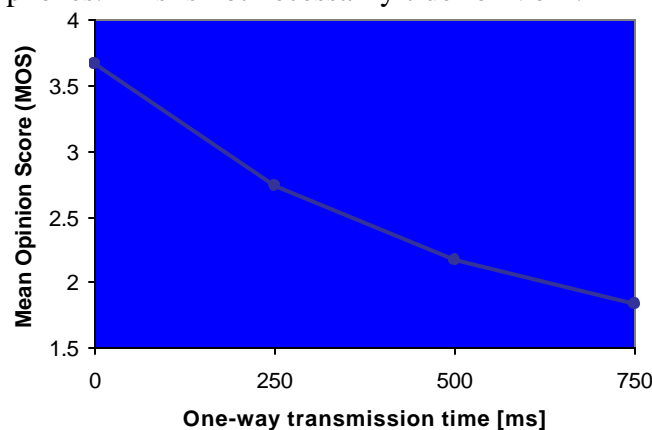


Fig. 2: Effect Of Delay On Conversational Quality From ITU-T G.114

The impact of latency on communication quality is not easily measured and varies significantly with the use. For example, long delays are more tolerable in a cell phone environment than in a regular wired phone because of the added value of mobility. The presence of echo also has a significant impact on our sensitivity to delay.

The ITU-T recommends, in standard G.114, that one-way delay should be kept lower than 150 ms for acceptable conversation quality (Fig. 2 **Fig. 2:** is from G.114 and shows the perceived effect on quality as a function of delay). Delays from 150 to 400 ms are acceptable provided that administrators are aware of the impact on quality, and latency larger than 400 ms is unacceptable.

Most packet losses occur in the routers, either due to high router load or to high link load. Packet losses also occur when there is a breakdown in a transmission link. When a packet is lost a mechanism for filling in the missing speech must be incorporated, since the requirement on low delay does not allow for retransmission of lost packets as is usually done for regular data traffic. For best performance such algorithms have to accurately predict the speech signal and make a smooth transition between the previous decoded speech and inserted segment.

Another approach to handle packet loss is to deploy a speech coding technique that has been specifically designed to handle packet loss. None of the current speech coding standards (ie ITU codecs) has been designed in such a manner and hence are all sensitive to packet loss. However, new robust codecs are being adopted outside of the traditional standards bodies for speech coding. For example, the Internet Engineering Task Force (IETF) has recently standardized the iLBC speech codec 0. There are also robust proprietary codecs available, such as 0.

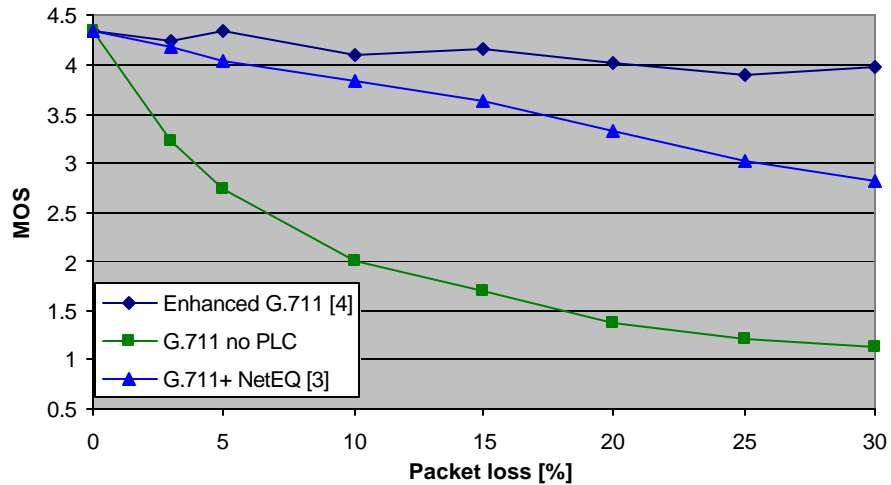


Fig. 3: Subjective Results For Approaches To Handling Packet Loss Concealment
{Source: Lockheed Martin Global Telecommunication (formerly COMSAT)}

VoIP transmission delay varies quickly and with significant amounts over time due to queuing effects in the IP network, causing a delay jitter. The jitter present in packet networks complicates the decoding process because the decoder needs to have packets of data readily available at regular intervals to produce smooth, continuous speech. A jitter buffer must be deployed to make sure that packets are available when needed. The objective of jitter buffer design is to keep the buffering delay as short as possible, while minimizing the number of packets that arrive too late to be used.

A new invention, which combines an advanced adaptive jitter-buffer control with error concealment, has recently been presented 0. The result is much quicker adaptation, and hence significantly lower delay, compared to a traditional jitter buffer that is limited in its adaptation resolution by the packet size. By deploying such a jitter buffer and packet loss concealment technique, combined with a robust speech codec, the challenges of IP networks can be overcome (Fig. 3).

Conclusions

We have shown here the improvement of communication quality achievable by extending the speech bandwidth beyond that currently experienced in the PSTN. We have also pointed out other opportunities for achieving better-than-PSTN quality in VoIP. In addition we have addressed how the major network challenges of VoIP can be overcome.

Combining these results an overall solution much better than experienced in today's phone systems can be obtained.

About The Author

Dr Jan Linden is Vice President of Engineering at Global IP Sound. He has been conducting R&D in speech processing and communications for the last 13 years. Prior to joining Global IP Sound he was with University of California, Santa Barbara (UCSB) and SignalCom, Inc. He holds a PhD and an MSc in Electrical Engineering from Chalmers University of Technology, has published more than 30 papers and has filed and been awarded several patents. Dr Linden can be contacted at jan.linden@globalipsound.com

References

- [1] ITU-T SG15, "Provisional terms of reference for wideband (7 kHz) speech coding," June 1994.
- [2] IETF RFC 3951, "Internet Low Bit Rate Codec (iLBC)," 2004.
- [3] Whitepaper, "GIPS NetEQ - A Combined Jitter Buffer Control/Error Concealment Algorithm for VoIP Gateways," available from [Global IP Sound](#).
- [4] GIPS Enhanced G.711, <http://www.globalipsound.com>.

