

Wideband Audio and IP Telephony: Experience Higher-Quality Media

With the availability of high-definition television (HDT), high-definition digital video disk (HD DVD), and high-definition radio channels, the everyday consumers of audio and visual media are demanding better quality from all forms of media. Very soon, the ubiquitous analog home phone will no longer be immune to this expectation. Users of telephony devices expect more features, higher availability, and better quality from their devices.

The demand for a higher-quality audio experience is already apparent in the enterprise market space and will eventually reach the public phone service carriers, but this change will not happen overnight. The encoding of speech for the public switched telephone network (PSTN) has not changed in 50 years, and speech transport has been optimized to provide the standard “toll quality” to which phone users have become accustomed.

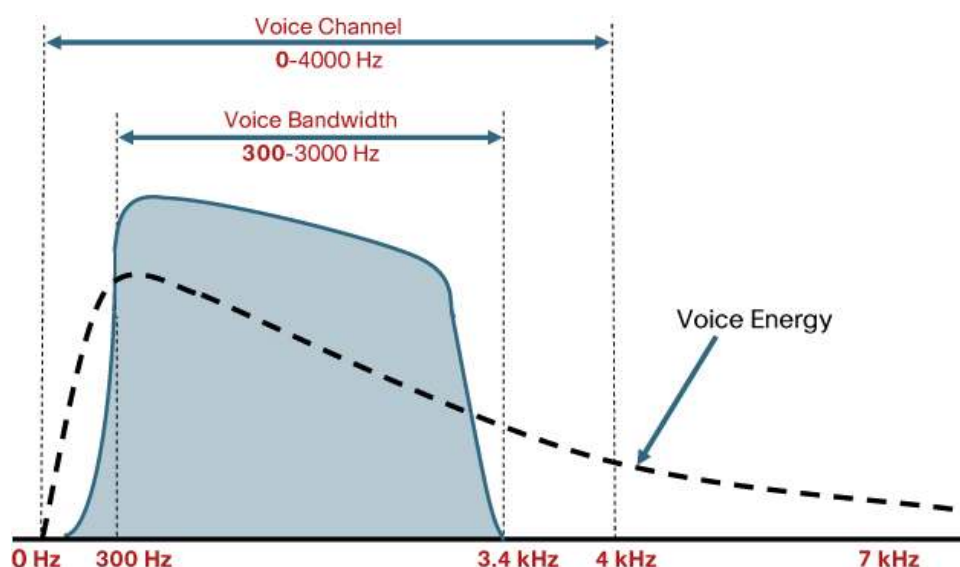
Many vendors are now implementing standards-based wideband codecs to improve the voice quality carried over IP networks. Wideband codecs provide clearer, more lifelike voice communications and markedly improved intelligibility because of the additional voice data included in the audio stream. They also double the voice signal range while using the same network bandwidth as narrowband codecs.

This white paper discusses current toll-quality voice limitations and shortcomings, and then details how wideband audio codecs and acoustics overcome these limitations to provide a better user audio experience.

The Toll-Quality Voice Standard

Every private branch exchange (PBX) manufacturer and IP telephony vendor wants its voice quality to match the toll quality of the PSTN. Consumers of voice services are very quick to notice when the voice quality does not meet this criterion. But if you look at the PSTN toll-quality standard, you quickly realize that the toll quality is mediocre at best when compared to other sources of audio transmission.

The frequency range that the human voice can produce ranges from 30 to 18,000 Hz. Although the lower frequencies are where most of the speech energy and voice richness is concentrated, much of the intelligibility of human speech occurs in the higher frequencies. When engineers originally designed telephone communications, they determined that a listener did not need to hear all the frequencies that make up the human voice to determine the words being spoken.. Because most of the energy necessary for intelligible speech is contained in a band of frequencies between 0 and 4000 Hz, this range was defined as the voice channel (Figure 1).

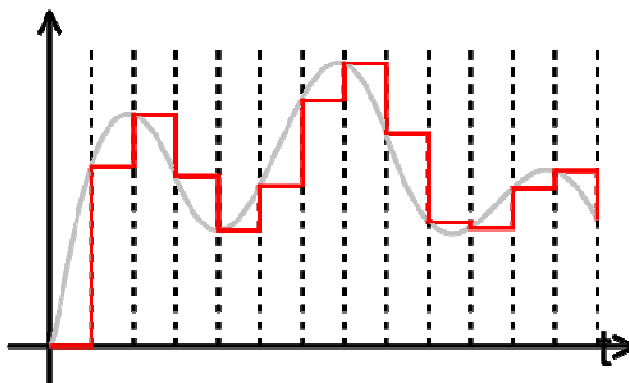
Figure 1. Frequency Range of the Voice Channel

Because telephone communications are carried over analog circuits, the introduction of noise is a problem. The further a voice signal has to travel, the more likely the signal will have “noise” introduced. To eliminate noise from the voice signal, therefore, the signal is sent through a band-pass filter to remove any frequency below 300 Hz and above 3400 Hz.¹ The band-pass filter allows voice carriers to reduce noise in a conversation because any signal outside the voice channel is discarded.

Analysis of the frequency spectrum of human speech shows that much of the differentiating content in certain sounds is found in the 4- to 18-kHz range. Certain consonants sound nearly identical when the higher-range frequency is removed. For example, many telephony users have a difficult time discriminating the sound of the letters “S and F” or “P and T” or “M and N”, making words such as “sailing and failing” or “Manny and Nanny” or “patter and matter” more prone to misinterpretation over a phone connection. Removing the higher frequency from the signal removes the information that allows listeners to clearly identify which sound is being spoken.

As the number of telephone endpoints expanded in the mid 20th century, phone carriers needed a better way to carry voice. So in the 1960s, telephone carriers began switching to a digital medium for voice. But they needed to be able to convert a continuous analog signal into a discrete digital signal. The voice digitization process required sampling of the data at regular intervals, and then quantifying the signal in digital form. The standard that was used to digitize voice is called pulse code modulation (PCM), which takes the signal level at the time of sampling and converts it into a digital value to represent the signal (refer to Figure 2).

¹ A band filter adversely affects adjacent frequencies, so the voice pass band is restricted to 300–3400 Hz.

Figure 2. Pulse Code Modulation

To ensure accurate signal regeneration (based on the Nyquist theorem), a voice signal must be sampled at twice the maximum frequency.² Because the maximum frequency of the voice channel is limited to 4000 Hz, based on the Nyquist theorem, the sampling rate was set to 8000 times/second. Simple math shows that 8000 samples/second x 8 bits/sample = 64,000 bps, or 64 kbps. Even today, a 64-kbps PCM encoded voice stream is the basis for toll quality that all voice service providers are measured against.

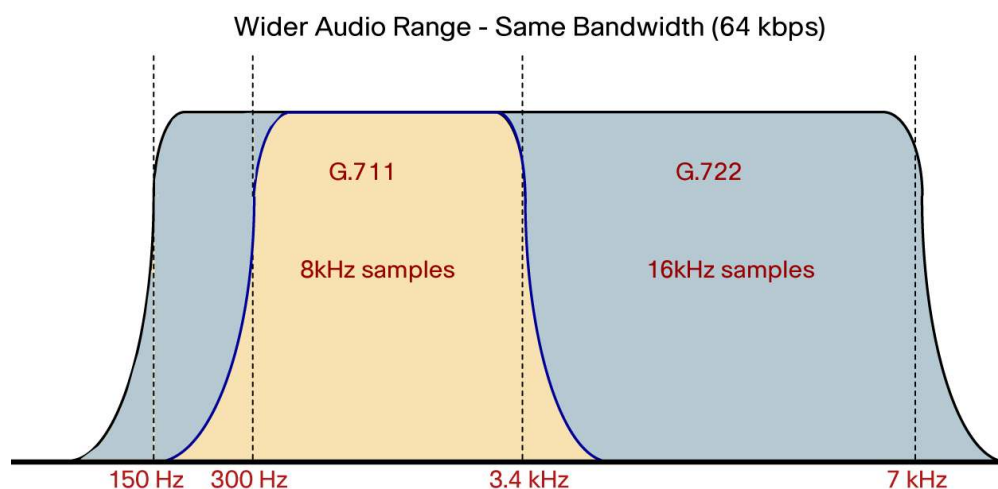
G.722 and Wideband Voice Quality

As the processing power of digital signal processor (DSP) chips increases, the ability for voice devices to perform an advanced voice-compression algorithm becomes easier and cheaper. Therefore, there has been a shift in the voice world to provide voice quality that is better than toll quality -- and the codec most commonly used to provide improved voice quality for voice over IP is the G.722 wideband codec.

The wideband audio codec is not a new standard. In fact, the first recommendation for G.722 was published in 1988. G.721, defined 7 years earlier, defined adaptive differential PCM (ADPCM) and introduced the ability to reduce the audio data stream from 64 to 16 kbps. The G.722 recommendation used the same ADPCM for voice compression, but instead of compressing 64 kbps into 16 kbps like G.721, G.722 maintained the same bit rate but doubled the audio content. Doubling the audio content meant the G.722 codec could support a wider voice frequency range within the same 64-kbps stream as toll-quality PCM.

The increased response range is apparent when you compare samples from the frequency response range of the G.722 wideband codec to those of a G.711 codec. The G.711 codec has an 8-kHz sampling rate, and any signal above 3.44 kHz is blocked. With G.722 wideband codec, the sampling rate is increased to 16 kHz, doubling the frequency range to 8 kHz. Figure 3 shows the increase in the frequency response from almost 3.4 kHz to almost 7 kHz.

² The Nyquist theorem states "Exact reconstruction of a continuous-time baseband signal from its samples is possible if the signal is band-limited and the sampling frequency is greater than twice the signal bandwidth".

Figure 3. G.711 and G.722 Frequency Response

G.722 works by having the inbound voice signal pass through a digital filter that separates the audio signal into 0 Hz-to-4 kHz and 4 kHz-to-8 kHz audio bands. These sub-bands are then encoded using ADPCM. As discussed earlier, most of the voice energy is concentrated in the lower half of the audio band (0–4 kHz), so 48 kbps of the bandwidth will be dedicated to the lower sub-band and the other 16 kbps will be allocated to the higher sub-band. By performing ADPCM encoding on each sub-band separately, G.722 can provide both low and high frequencies that will provide richer audio sound and better re-creation of the original signal.

Cisco Wideband Audio

To deliver the full benefits of wideband audio, the entire system, including the hardware and software, needs to be designed for a wideband experience. Cisco® has released a new set of phones with enhanced acoustics (speaker, microphones, handset, and housings), designed specifically to support the G.722 wideband codec and its expanded voice frequency range. With the release of these phones -- the Cisco Unified IP Phone 7975G, 7965G, 7945G, 7962G and 7942G models -- Cisco is enhancing the voice user's experience by supporting wideband audio across speaker, handset, and headset.

Additionally, other newer Cisco Unified IP Phones 7900 Series phones (Cisco Unified IP Phone 7906G, 7911G, 7921, 7931G, 7941G-GE, 7961G, 7961G-GE, 7970G, and 7971G-GE models) support G.722 with an optional wideband handset or headset.

Wideband audio is most useful for on-system calls, Session Initiation Protocol (SIP) calls, or H.323 trunk calls, and all Cisco products are moving toward G.722 support. However, because the PSTN and Primary Rate Interface (PRI) connections are still G.711 a-law or G.711 μ -law based, when the audio stream leaves the IP network, the voice is returned to a PCM encoding, thereby losing the full benefit of wideband audio. Even though calls through a PRI or PSTN do not use wideband, Cisco Unified IP phone users should still have an improved experience with the enhanced acoustics and hardware on their phone.

Conclusion

With the increase in awareness of audio quality in today's society, the "toll-quality" standard is being reevaluated and improved. The frequency limitations on PCM-based encoding make conversations a challenge because listeners have to concentrate on context rather than content to decipher the speaker's words. The expansion of the frequency range coverage in a wideband audio codec reduces the potential for word confusion and allows you to enjoy a more natural sounding conversation.



Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 527-0883

Asia Pacific Headquarters
Cisco Systems, Inc.
168 Robinson Road
#28-01 Capital Tower
Singapore 068912
www.cisco.com
Tel: +65 6317 7777
Fax: +65 6317 7799

Europe Headquarters
Cisco Systems International BV
Haarlerbergpark
Haarlerbergweg 13-19
1101 CH Amsterdam
The Netherlands
www-europe.cisco.com
Tel: +31 0 800 020 0791
Fax: +31 0 20 357 1100

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

©2007 Cisco Systems, Inc. All rights reserved. CCVP, the Cisco logo, and the Cisco Square Bridge logo are trademarks of Cisco Systems, Inc.; Changing the Way We Work, Live, Play, and Learn is a service mark of Cisco Systems, Inc.; and Access Registrar, Aironet, BPX, Catalyst, CCDA, CCDP, CCIE, CCIP, CCNA, CCNP, CCSP, Cisco, the Cisco Certified Internetwork Expert logo, Cisco IOS, Cisco Press, Cisco Systems, Cisco Systems Capital, the Cisco Systems logo, Cisco Unity, Enterprise/Solver, EtherChannel, EtherFast, EtherSwitch, Fast Step, Follow Me Browsing, FormShare, GigaDrive, HomeLink, Internet Quotient, IOS, iPhone, IP/TV, iQ Expertise, the iQ logo, iQ Net Readiness Scorecard, iQuick Study, LightStream, Linksys, MeetingPlace, MGX, Networking Academy, Network Registrar, PIX, ProConnect, ScriptShare, SMARTnet, StackWise, The Fastest Way to Increase Your Internet Quotient, and TransPath are registered trademarks of Cisco Systems, Inc. and/or its affiliates in the United States and certain other countries.

All other trademarks mentioned in this document or Website are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (0708R)