



VoIP to 20 kHz: Codec Choices for High Definition Voice Telephony

Jeffrey Rodman
CTO/Co-Founder
Polycom

July 2008

VoIP to 20 kHz: Codec Choices for High Definition Voice Telephony

July 2008

White Paper

Table of Contents

Introduction	3
How the Codecs Compare.....	5
Keeping Up With the VoIP Industry	9

VoIP to 20 kHz: Codec Choices for High Definition Voice Telephony

July 2008

White Paper

Introduction

As voice over IP (VoIP) becomes more and more prevalent throughout business and personal communications, it's important to understand the full potential of this technology and how it can dramatically improve our day-to-day communications experiences. One of the most important new capabilities that VoIP provides is the transformation of voice quality from conventional narrowband, "cellphone-sound" audio, to wideband, or HD Voice, in which the full range of human speech can be carried among all the participants. By delivering voice over an IP network, we have the opportunity to bring the highest quality, most lifelike conversations to our daily interactions both in the workplace and at home.

With the rapid transformation of business telephony from circuit-switched to VoIP (at the time of this writing, more than 70 percent of new lines are VoIP), there is an accompanying move from low-fidelity 3 kHz audio-only telephones to devices that take advantage of the IP network to deliver integrated higher-fidelity (also called "wideband," "High Definition Voice," or "HD Voice") sound, features, applications, and the whole range of integrated communication capabilities sometimes known as "Unified Communications." All of this is happening because with the move from plain old telephone service (POTS) to VoIP, the network is suddenly transparent and can deliver an enormous new range of capabilities.

The transition from narrowband audio (3 kHz) to VoIP wideband telephony is accelerating. One essential element of this transition is the addition of wideband-capable codecs, from 7 kHz to 20 kHz, to the codec repertoire. This paper explains the codec alternatives available and the "yardsticks" used to compare and select among them.

WHAT IS A VOIP WIDEBAND CODEC?

A VoIP codec ("coder - decoder") is an algorithm that squeezes (the "coder") digitized audio so it fits more easily into a VoIP data channel, and then re-expands it (the "decoder") so you can hear the audio once again. "Wideband" means that the audio has much higher fidelity than old-fashioned analog phones (see "The Effect of Bandwidth on Speech Intelligibility" and other papers by this author), and 20 kHz is wideband audio, extended to the very top edge of human hearing.

VoIP to 20 kHz: Codec Choices for High Definition Voice Telephony

July 2008

White Paper

The Codec's Role in a Phone Call

VoIP codecs operate by taking uncompressed digital audio, and applying an agreed, standardized algorithm to reduce the number of bits it takes to represent that audio. It's important for both ends of a phone call to agree on what that algorithm should be, of course, and in a VoIP phone call (using standard SIP signaling), this agreement is achieved when the call is first placed, through a process called "capabilities exchange."

Every VoIP phone contains one or more codecs, and during call establishment, they share their lists of supported codecs. One phone, for example, may say "Hey stranger, I can support codecs A, B, or C", and the other one will respond "Nice to meet you, I can support codecs B, C, or D." At this point, both phones recognize that they could converse in either B or C (this process is can easily be compared to two multilingual strangers meeting on the street, figuring out what languages they share, then deciding which of the shared languages to proceed in). Depending on how they have been set to prioritize various parameters, one phone may then say "well, since C gives better audio bandwidth than B, let's proceed with that," or "B uses a lower bit rate and my company thinks that's more important, so let's proceed with that."

Most VoIP phones contain a number of different codecs covering a range of performance levels and, often, bandwidths. Having wideband capability doesn't mean that a phone is unable to connect to a narrowband phone, it just means that it has a wider repertoire and can do both, like a musician who can play both clarinet and saxophone (perhaps not at the same time). So now let's consider the following question: On what basis does one evaluate and choose those wideband codecs?

Comparing Codecs

Here's a first glance at the most important codec characteristics; we'll cover these in more detail later. It is quickly evident from the large number of parameters, however, that no codec is likely to be "best" in all categories at any given time.

- Audio bandwidth (higher is better)
- Data rate or bit rate (how many bits per second, fewer is better)
- Audio quality loss (how much does it degrade the audio, lower is better)
- Kind of audio (does it only work with speech, or with anything?)
- Processing power required (less is better)
- Processor memory required (less is better)
- Openly available to vendors? ("yes" is essential)
- Inserted delay (audio latency caused by the algorithm, less is better)
- Resilience (how insensitive to lost or corrupted packets, more is better)
- ITU standards-based (standardized by the International Telecommunications Union - "yes" is better)

As you read through this, it's possible you already have experience in evaluating these parameters among narrowband codecs for existing VoIP systems (if you've ever compared G.711 against G.729, for example). Except for boosting the audio bandwidth to wideband, the other tradeoffs are much the same. Let's look at some of the key parameters, and compare them among the most popular wideband VoIP codecs.

VoIP to 20 kHz: Codec Choices for High Definition Voice Telephony

July 2008

White Paper

Principal Wideband VoIP Codecs Today

L256. The simplest of all wideband codecs, the 7 kHz L256 directly sends all the bits of digital audio sampled into 16-bit words at 16 kilosamples per second (ksps), using no compression whatever, hence the name ("Linear 256" kbps). L256 is a basic requirement in all VoIP phones, but is seldom used because of its high bit rate.

G.719. Perhaps the best match among requirements for communication systems at 20 kHz, G.719 is a recent ITU-approved arrival that combines excellent quality for music and voice with low latency, modest processor load, and network-friendly bit rates.

G.722. This is the grandfather of 7 kHz wideband VoIP codecs, and the most widely deployed so far. G.722 applies adaptive differential pulse code modulation (ADPCM) to high and low frequencies separately, yielding an algorithm that works equally well with music or voice.

G.722.1. Also known as "Siren 7," this modern 7 kHz audio codec is in almost every videoconferencing system today and is gaining traction in VoIP because of its higher efficiency and lower bit rate. G.722.1 is a "transform" (as in "Fourier transform") codec and works by removing frequency redundancies in any kind of audio.

G.722.2. This codec, "AMR-WB," is a 7 kHz wideband extension of the popular adaptive multi-rate (AMR) cellphone algorithm, and excels in delivering wideband high-quality voice at the lowest bit rates. G.722.2's algebraic code excited linear prediction (ACELP) algorithm is optimized for speech, and works by sending constant descriptions of how to shape and stimulate a human speech tract to reproduce the sound you feed into it.

G.722.1 Annex C. Also known as "Siren14," this is a 14 kHz extension of G.722.1 and is popular because of its wider bandwidth, its efficiency, and its availability (under license) for zero royalty.

Speex. Speex is an open-source CELP codec.

MPEG. There are more than 25 versions of the moving pictures expert group (MPEG) transform codecs, each delivering a set of performance levels optimized for various parameters. The variant best suited to telecommunications is MPEG4 AAC-LD, a lower-delay version of the intended MP3 successor, MPEG4 AAC.

MP3. The popular MP3 format uses a form of transform coding, and is optimized for media distribution.

FLAC. The Free Lossless Audio Codec (FLAC) produces much higher bit rates than most other codecs, but compensates by preserving complete audio quality.

How the Codecs Compare

Each conferencing environment has its own acoustical challenges that require an appropriately designed conferencing solution. Let's examine some of these differences.

Audio Bandwidth

Audio bandwidth corresponds to audio fidelity, that is, the ability to carry sounds ranging from very low pitches, like a kettledrum or a sonic boom, to very high pitches, like a cymbal or a plucked guitar. Therefore, more bandwidth is better. The human voice has important content beyond 14 kHz (this is why wideband telephony, even at 7 kHz, delivers such a telling improvement over older 3 kHz analog phones). The human ear can be sensitive to 20 kHz, and virtual every medium we experience today carries sound over this full range.

VoIP to 20 kHz: Codec Choices for High Definition Voice Telephony

July 2008

White Paper

SIDENOTE

This same word, "bandwidth", is sometimes used to refer to two very different things. Audio bandwidth is the amount of fidelity, or how high are the audio frequencies, that are carried by a connection. Data bandwidth is how many bits per second, the bit rate, are used to send the digital signal. These two qualities are also described as wideband (meaning higher audio bandwidth) and broadband (a high data-rate digital connection). This can be confusing at first. But in any event, one goal of a codec is to decouple the two, and deliver a lot of audio bandwidth for very little data bandwidth.

VoIP is also working its way toward supporting up to 20 kHz, but today, there are more codecs available to support 7 kHz audio than these even higher bandwidths. This is because 7 kHz in and of itself provides an easily achievable and dramatic improvement in voice-only communications. Desktop phones supporting 7 kHz are available from many vendors today, but to-date, only Polycom has introduced conference phones at 14 kHz and 20 kHz.

Data Rate

The data rate required by a compressed audio channel becomes important when network bandwidth is limited, especially when supporting multiple phone connections. This is a common issue in narrowband VoIP telephony (comparing the bit rates of G.711 vs. G.729 is a common discussion), and its importance in wideband-capable systems is no different. Table 1 shows some typical numbers.

BW (kHz)	Typical bit rate (kbps)
3.3	8 (G.729), 56 (G.711)
7	10 (G.722.2), 24 (G.722.1), 64 (G.722)
14	32 (G.722.1C)
20	32 (G.719), 64 (AAC-LD)
22	32 (Siren22)

Table 1: Audio bandwidth versus bit rate for some popular codecs

One thing we see here is that the typical bit rates don't necessarily rise with rising audio bandwidth; the bit rate has as much to do with the codec chosen as with the bandwidth. The reasons for this are twofold: audio contains most of its information in the lower frequencies, so there's less information to be coded and sent in the higher frequencies, and the human ear is less sensitive to inaccuracies at the higher frequencies, so a compression algorithm can be a little less precise without being noticed.

Another point to note is the span of bit rates among wideband codecs. For example, 7 kHz audio requires 64 kbps from G.722, but only 10 kbps from G.722.2. Here, the difference is due to the assumptions made by the codecs. G.722.2, an ACELP codec, assumes that it's working on human speech. It knows that it's not going to be fed the sounds of a violin or a speeding locomotive, so it

VoIP to 20 kHz: Codec Choices for High Definition Voice Telephony

July 2008

White Paper

takes a whole different approach to compression and consequently can be extremely efficient about it. This is why G.722.2 is preferred for cellphone use, where the cost of the bit rate is high, but another codec such as G.722.1 would be preferred if the application were broader and included multiple talkers, or music.

Figure 1 shows how these different codecs stack up when comparing bandwidth to bit rate.

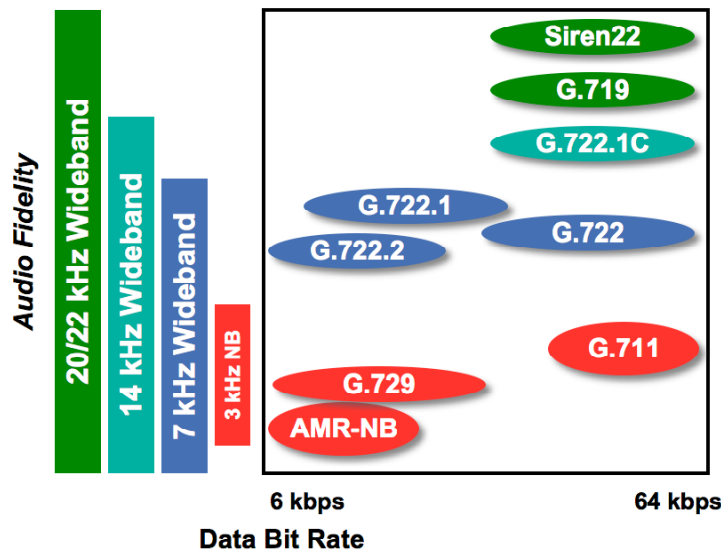


Figure 1: Audio bandwidth versus bit rate for different codecs

Processor loading

High-complexity codecs drive up the cost of a VoIP phone or endpoint because they require faster, more expensive processors and more memory. The issue multiplies with VoIP phones that perform multi-party bridged calls internally, which is a common VoIP-enabled feature today. Table 2 gives a couple of good example of how codecs can differ in their appetites for processor power.

BW (kHz)	Codec	MIPS
7	G.722.2	38
7	G.722	14
7	G.722.1	5.5
20	G.719	18
20	MPEG-4 AAC-LD	36

Table 2: MIPS versus audio bandwidth for some popular codecs

VoIP to 20 kHz: Codec Choices for High Definition Voice Telephony

July 2008

White Paper

At 7 kHz, G.722.2 shows the highest demand for processor power, but we remember that its operation also results in the lowest bit rate. G.722.1, at one-seventh the processor power of G.722.2 and 40 percent the bit rate of G.722, is a good compromise.

Comparing the two 20 kHz codecs, the difference in processor loading is striking, but surprisingly, there's no compensating advantage in bit rate or quality. This may be because the MPEG codec adapts technology originally intended for media streaming and recording where G.719 was always targeted for VoIP and telecommunications, but it does form a good demonstration of how important differences can pop up.

Audio Quality

One reason that comparing codecs can get tricky is that audio quality, a somewhat subjective measure, is also tightly related to bit rate within a particular codec. One codec may tout extremely low bit rates, but a quick listening will reveal that the audio quality at those lowest bit rates is almost unusable. In this paper, I have tried to relate bit rates at comparable audio qualities, so the "typical" figures here will often be higher than the provider's "minimum" figures. But they are realistic, and appropriate for VoIP usage.

There are standard objective measures of audio quality (MOS, PESQ, etc.), but if you are making a serious comparison of codecs, it's best to do a real "apples-to-apples" test and apply the same standard test track to all candidate algorithms, with each candidate running at its planned bit rate. Private and open-source codec suppliers will be glad to provide an algorithm simulator that runs on a standard PC, making it possible for you to do this test yourself.

Give some thought to this test track. Even in VoIP applications, we're not always dealing with just the human voice. You often will find two or more voices speaking at the same time, or someone talking in a room with lots of reverberation: two situations that can really throw off a human voice tract codec (some CELP and ACELP implementations, for example, can be particularly sensitive to these things). Even a door closing or pencil dropping while someone is talking can come across very strangely, so it's good to have a full test in order to build a good comparison.

Latency

Latency is the time delay from when you say a word until the other person hears it, also referred to as the "mouth-to-ear" delay. When it gets too long, conversations become difficult and stilted, with participants frequently but inadvertently interrupting each other and not understanding why. Twenty years ago, we often heard very long latencies on long-distance calls because of the widespread use of satellite links (even at the speed of light, a couple of hops off of satellites perched 22,500 miles above the earth add up to the better part of a second), but today long latencies are mostly the consequence of older videoconferencing systems or carelessly planned VoIP phone systems.

A common recommendation is that one-way latency, which includes the codec, should not exceed 150 milliseconds. This is not a problem in a well-designed system using telecom codecs such as G.722, G.722.1, etc. But occasionally a media or streaming codec will find its way into a VoIP system with disruptive results. Media codecs, such as those used to transmit streaming audio over the internet, are often not optimized for latency because one-way streaming connections are not sensitive to latency. Because they can insert an appreciable fraction of a second delay, they should be avoided in VoIP and telecommunication systems.

VoIP to 20 kHz: Codec Choices for High Definition Voice Telephony

July 2008

White Paper

Another contributor to latency in a VoIP system is a "jitter buffer." This is a kind of shock absorber for data flow that soaks up the momentary variations that occur in any IP network. These are sometimes embedded within a codec, which makes it important to be sure that multiple, redundant jitter buffers are not inadvertently built into a system (a jitter buffer can be 20 to 80 milliseconds or more).

Cost

As users, we usually do not see the cost of a codec, but cost can influence its selection into a phone system or a phone. There are license fees, or royalties, associated with some codecs; often, a per-year minimum fee, with a per-port or per-phone fee, and perhaps an initial fee as well.

Some of the codecs we use today, such as G.722, are royalty-free because the underlying patents have expired. Some codecs, such as G.722.1 Annex C, are royalty-free because their vendor has decided that the industry is better served if high performance codecs are widely deployed. Other codecs, such as MPEG4 AAC-LD, still bear royalties.

Royalties are not necessarily a bad thing, because codecs are often the result of long and expensive research resulting in valuable characteristics (such as low bit rate in G.722.2, which saves money in use). They are simply something to be aware of when considering your VoIP network deployment plans.

Standardization and Availability

The ITU is the de facto worldwide agency for standardization of telecommunications codecs. This is the industry organization that assigns the numbers to our familiar codecs; G.722.1's full name, in fact, is ITU-T G.722.1, because it is a product of the ITU-T Telecommunications Standards Sector, and like all ITU standards has been subjected to open, rigorous multi-vendor evaluation before being accepted. While proprietary codecs may be incorporated in limited-use systems, it's of paramount importance that business VoIP telephony systems, which require worldwide interoperability and high reliability, be configured with ITU-approved codecs. The ITU sanction also ensures that codecs are available to all vendors on fair and reasonable terms.

Keeping Up with the VoIP Industry

There are two trends to keep in mind in VoIP audio bandwidth today; one is strategic, and one is technical.

The strategic trend is this: VoIP telephony is moving toward full bandwidth 20 kHz sound, because the VoIP endpoint is undergoing transformation to a multi-purpose, multimedia device that integrates communications, applications, and even entertainment. As you have seen, there's little cost or bit rate penalty in going to wideband telephony using modern codecs (even the fullband G.719 codec has a lower bit rate than G.711), and competitive pressure will drive VoIP vendors to achieve full human compatibility in a very few years. Some applications will remain at the voice-friendly 7 kHz point due to tight cost or size constraints, but we'll see an increasing tide of fully capable 20 kHz VoIP systems with unified capabilities.

The technical trend follows from the strategic: which are the codecs will bring us to this 20 kHz world?

VoIP to 20 kHz: Codec Choices for High Definition Voice Telephony

July 2008

White Paper

At 7 kHz, G.722 is mature, free, and already widely deployed in endpoints and in PBXs and softswitches. G.722.2 will be deployed in applications where its higher cost is offset by its very low bit rate and high quality, much of this driven by cellphones. Its adoption there will push the network, and consequently wired endpoints, to follow. And finally, G.722.1 adds multimedia capability at less than half the bit rate of G.722, and one-seventh the processing cost of G.722.2. These three codecs form a functionally complete set for 7 kHz performance.

The choice at 14 kHz is G.722.1 Annex C because of its maturity, modest bit rate and processing needs, and zero-cost license.

And finally, 20 kHz performance in the VoIP world will come from the ITU's new G.719, as the likely successor to Siren22 (Siren22 is a principal predecessor of G.719, however).

Want to Know More?

If you're interested in more information on wideband audio and HD Voice, here are some places to look.

hdvoice.tmcnet.com

www.siren7.com

www.wikipedia.com, www.voip-info.org, etc: There's a lot of good information on the internet, but take note of when the material was posted because this wideband VoIP is moving very quickly.

Contact us at:
www.polycom.com
sales@polycom.com

VoIP to 20 kHz: Codec Choices for High Definition Voice Telephony

July 2008

White Paper

APPENDIX

Characteristics of Wideband Audio Codecs

Codec	Audio BW (kHz)	Industry Use	Data Rate	Latency (ms)	Source	MIPS	Data rate at good audio
G.729A/B	3.3	Conventional telephone	8 kbps	15	Voice	13	64 kbps
G.711	3.3	Conventional telephone	64 kbps	0.125	Voice	0	64 kbps
L256	7	Wideband VoIP standard	256 kbps		All	0	256 kbps
G.722/1988	7	Wideband VoIP common	48, 56, and 64 kbps	0.125	All	14	56 kbps
G.722.1/1999	7	Videoconferencing, Microsoft, growing VoIP	24 and 32 kbps	40	All	5.5	24 kbps
G.729J	7	Not known	14 to 32kbps at 7kHz	49	Speech	36	20 kbps
AMR-WB/G.722.2/2001	7	Cellular, 3G	6.6-23.85Kbps	114	Speech	38	12 kbps
GIPS- iPCM WB	7	Skype	80 kbps	0.125	All	8.6	
Speex	7	Not known	10 - 40 kbps	34	Speech	41	20 kbps
Vorbis	7	Not known	28 kbps at 7kHz	82	All	Hi	64 kbps
GSM-FR	7	Mobile	13 kbps	Long	Speech	>100	13 kbps
G.722.1 Annex C	14	Videoconferencing, Microsoft, growing VoIP	24, 32, and 48 kbps	40	All	10.9	24 kbps
MP3	20	Consumer music	(mono) 64-512 kbps	>54	All	38	96 kbps
MPEG-4 AAC-LD	20	Consumer Multimedia	(mono) 24-192 kbps	30	All	36	64 kbps
Siren 22	22	Polycom Video	32, 48, and 64 kbps	40	All	18.4	32 kbps
G.722FB	20	VoIP, Video	32, 48 and 64 kbps	40	All	16	32 kbps
Microsoft WMA9 Voice	22	Microsoft UC, Windows Media, etc	64-192 kbps		Speech	Hi	64 kbps

Polycom Headquarters: 4750 Willow Road, Pleasanton, CA 94588 (T) 1.800.POLYCOM (765.9266) for North America only. For North America, Latin America and Caribbean (T) +1.925.924.6000, (F) +1.925.924.6100

Polycom EMEA: 270 Bath Road, Slough, Berkshire SL1 4DX, (T) +44(0)1753 723000, (F) +44(0)1753 723010

Polycom Asia Pacific: 8 Shenton Way, #11-01 Temasek Tower, Singapore 068811 (T) +65.6389.9200, (F) +65.6323.3022