HD Voice: Why Phones Are Going HD Fast

The HD Phenomenon

Almost everything that makes sound today creates not just "sound," but good, high-fidelity sound. Movies, televisions, iPods, radios, just about anything you can name covers the whole audio spectrum that we hear. Anything, that is, except the telephone. While normal sound sources today commonly extend their audio performance to 20 kHz, the traditional telephone network transmits only one-third to one-sixth of this. That's why phones have that congested "telephone" sound: they're cutting out significant amounts of the sound your ear is accustomed to hearing.

And yet, this "missing" sound is especially important in phones because so much of the useful information in the human voice happens in those higher frequencies, the ones that are cut off on a traditional telephone call. Consonants often distinguish one word from another, and over half the energy in consonants occurs above the 3 kHz range. Think of "sail" and "fail," for example: that hissing "s" sound is strongest in the high frequencies; this effect is true for many other consonants, and in most languages. This means that when normal phone lines muffle these out, we are making more mistakes and getting fatigued sooner: not a good situation when business is at stake.

The loss is even greater as global companies turn to telephony in order to connect countries, continents, and cultures. The challenges are multi-dimensional — we all know that when someone talks with an accent different than our own, they are much more difficult to understand over the phone than in person. It's much harder to tell people apart by the sound of their voice when they're on the phone. And when groups get together and talk over a speakerphone, the addition of room noise, reverberation and echoes within the room, soft talkers, and multiple talkers often leads to a very stressful time.

The History of Wideband and Narrowband Audio

Until now, there have been few alternatives to sub-par voice quality on telephone calls. If someone is talking too softly, we may ask them to speak up — this works for a few minutes, until they fall back into their normal pattern. We ask them to turn off their projector fan, or move the projector away, or turn down the air conditioner. We try to get them to move closer to the phone, if there's an extra seat. And we ask them to repeat what they just said, sometimes more than once, sometimes to spell it, sometimes to send an email repeating what they said. After a while, we just give up, make out what we can, maybe tell them we'll just "catch up next time we're out there," and surreptitiously start browsing our in-baskets. Sound familiar?

These are all problems caused by the need of conventional telephones to discard most of our voices to fit over a narrowband phone line, and this is why standards-compatible wideband audio (also known as HD audio or HD voice) is rapidly taking hold of the market. By restoring the higher frequencies, we are restoring the missing half of human speech: We are boosting our ability to distinguish one word from another, and the fidelity that lets us easily distinguish who is speaking on the other end of a conference call. This means less fatigue, less "what did you say," fewer errors and misunderstandings, and sharply increased attention and engagement.

POLYCOM®
TOGETHER, GREAT THINGS HAPPEN.

If it's so important to have HD voice, why hasn't it happened before now? Well, back in the 1920's and 1930's, 3 kHz was about the limit of the technology. That's all that could be done with carbon microphones and the networks and phones of the time. As time passed, we learned to live with this and to laugh at its shortcomings (movies of the 1930's carried a lot of gags about poor telephone sound although phones have really not improved much since 1940). And once the network began its transition to digital Pulse-code modulation (PCM), there was no going back for traditional POTS (plain old telephone service) systems; the selection of an 8 kHz sample rate set a theoretical limit below 4 kHz and a practical limit well below that.

What has happened in the last few years is that the network has suddenly become transparent. As we shift from POTS to voice over IP (VoIP), we are shifting from a network that is designed to do only one thing — that is, to carry only narrowband voice — to a network that is designed to carry high-bandwidth applications ranging from data to voice to multimedia transmissions. By converting voice into data, it can be carried over the IP network in the same way as computer files, video, streaming audio, and all other forms of data and digital media. So by moving from POTS telephony to VoIP telephony, we not only get the advantages of a unified high-speed network, we can also positively impact the quality of the communications.

The advantage of this network transparency is that if a network is VoIP-capable, it is already capable of carrying HD voice. In the same way that the ITU's narrowband codecs like G.711 and G.729 convert narrowband audio to data streams, there are also ITU-approved standard wideband codecs, the "seven twenty-twos" (G.722, G.722.1, G.722.2, G.722.1 Annex C), which convert wideband audio to data rates comparable to their narrowband counterparts. The old perception that wideband audio requires more data bandwidth is actually inaccurate due to the accessibility of more advanced compression techniques.

A Look Today and Tomorrow

Many vendors today are making wideband audio VoIP devices that fully interoperate with one another. TIA-920 specifies the basic characteristics for a wideband telephone, and many phones today comply with this standard. The most common wideband codec at the moment is G.722, because it has been around the longest (since 1988), and operates at data rates comparable to G.711 (G.722 is specified at 48 kbps, 56 kbps, and 64 kbps). Most wideband phones include G.722 for these reasons. G.722.1 (Siren7) is also coming along, because it provides excellent quality for both speech and music at data rates of 24 kbps and 32 kbps, and G.722.2 (AMR-WB) can provide wideband speech below 14 kbps, which makes it very attractive for wireless applications.

While the endpoints are essential, there are other elements to truly achieve HD voice. The telephone switch — either the IP PBX, soft switch, or service provider — needs to recognize that a wideband codec is being used so it can properly manage calls. Interestingly however, because standard SIP telephony passes the audio stream directly between the endpoints, the PBX only needs to understand that these other codecs are being used; it doesn't actually "see" the audio, so it does not need to actively encode or decode. Many vendors have added G.722 to their PBX capabilities, so standard SIP calls can be made with wideband audio today. The PBX may also want to process wideband audio itself to provide features such as voicemail and speech recognition in wideband as well as narrowband.

Beyond the PBX, the data then needs to pass over a wide-area network (WAN). Major networks are now planning for wideband deployment although it is not yet in place, so wideband audio is currently being carried among facilities over corporate networks and via bridging service providers such as ZipDX. A number of service providers are also establishing a consortium to provide enhanced audio and features among all their users, which will greatly enhance the penetration of interoperable HD voice.

Regardless of whether a direct wideband connection is available between two wideband phones, those phones can always connect because wideband audio is an additional capability in a device, not a replacement. Wideband devices still support narrowband communication (the SIP protocol incorporates a capabilities negotiation in which the endpoints agree which codec they will use), so if one end doesn't understand wideband, they'll just connect in narrowband.

Finding a Solution

A well-designed wideband phone, in fact, often will provide better audio, even in narrowband, than a phone designed solely for narrowband. This is because with a good wideband phone, the whole phone is designed to extend far beyond normal narrowband limits, while narrowband-only phones begin cutting off well before the absolute limits of G.711. Hence, a narrowband phone will often provide a narrowband connection just through the conventional narrowband limits of 300–3,300 Hz, where a wideband phone can provide full performance from the same narrowband connection, running 150 Hz–3,700 Hz or better. That's a full octave extension at the low end where speech "presence" is felt, and more than 10 percent at the high end, all over a G.711 channel. And over a G.722 or other wideband channel, of course, it is much better than this.

Summary

The move to HD voice is happening today; and it is likely that wideband telephony will be in common use within the next two to three years. Because wideband audio endpoints are compatible with both narrowband and wideband calls, the best way to safeguard against early obsolescence is to be sure that any new phones are compliant with TIA920 and incorporate at least the G.722 wideband standard. Doing this will help ensure that you are protecting your customers' investments while providing an existing technology with tremendous growth potential.