

FULL-HD VOICE IS NEARLY HERE

Combining codecs for voice and music will yield perfectly clear calls



> HD Voice, the first major upgrade to telephone sound quality since the vacuum-tube era, has finally become widely available—just in time for a new generation of phone service called Full-HD Voice to take its place.

At the Mobile World Congress in Barcelona earlier this year, Fraunhofer IIS (Institute for Integrated Circuits) demonstrated a system based on a combination of powerful standard algorithms that can encode and decode in real time the full audio spectrum to 20 kilohertz in stereo. Switching to Full HD, which could be done in many devices as early as next year, would also mark the complete merging of voice into the mobile data stream, a goal long in the making.

Full-HD Voice converts speech into packets that can flow through the Internet along with data traffic, incorporating algorithms that can recover from packet loss, which turns today's Voice over Internet Protocol (VoIP) calls into

choppy, unintelligible hash. The technology includes algorithms that encode music and other nonspeech audio, sounds that are typically mangled by codes optimized to squeeze many voice calls into narrow slices of the spectrum. Because Full-HD Voice carries the whole audio spectrum, calls sound as if everybody's in the same room; you can even hear soft background sounds, like the faint clatter of fingers on a keyboard. And the powerful coding-decoding (codec) software can run as a smartphone app.

"We want to bring telephony into the 21st century," just as HD television has done for video, says H.P. Baumeister, director of Fraunhofer IIS's U.S. branch, in San Jose, Calif.

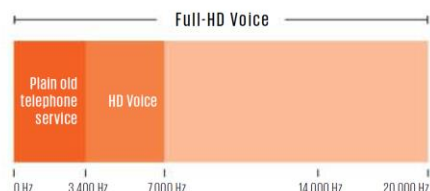
There's no doubt that voice telephony still has a foot in the 20th century. Modern landline phones have a frequency range of 300 to 3,400 hertz, a standard based on Bell Labs stud-

ies of the requirements for intelligible speech dating back to the 1920s. That range cuts off high frequencies needed to discriminate between consonants such as *f* and *s*, but it fit the limited bandwidth of old analog copper phone lines.

In 1988, the International Telecommunication Union approved the G.722 standard for HD Voice, which allows digital phone lines to carry 50 to 7,000 Hz. But it was little used because it would have required upgrading the landline phone network. The first three generations of cellular phones instead retained the 3,400-Hz narrowband landline audio, but they often sounded worse because of the way they compressed speech to squeeze more calls into the limited mobile spectrum. [See "Why Mobile Voice Quality Still Stinks—and How to Fix It," *IEEE Spectrum*, October 2014.]

The broader bandwidth of the Internet allowed Skype and some other VoIP services to carry 7,000-Hz HD Voice, but VoIP calls into the phone network have been limited to 3,400 Hz. Most 4G smartphones include dedicated circuits running algorithms to code and decode 7,000-Hz HD Voice, but they can connect at that rate only if both phones and every link between them can handle the signals. In practice, that means it works only between 4G phones on the same carrier.

Full HD will be able to bridge the audio gap regardless of the network or the device connected to it. The technological heart of Full-HD Voice is a standard called the Enhanced Voice Services (EVS) codec. Its speech compression algorithms are more complex and powerful »



EASY LISTENING: Unlike earlier systems, Full-HD Voice will cover the whole range of human hearing.

GETTY IMAGES

than those used for the decade-old HD Voice system, and it can squeeze stereo speech spanning the whole audible range into data rates as low as 9.6 kilobits per second. The codec also includes other algorithms developed to compress music.

The separate algorithms are vital because speech and music are compressed in different ways. Voice compression typically relies on algorithms called code-excited linear prediction (CELP), which is built on the physics underlying the human vocal system. CELP can reduce the data rate of voice signals by about a factor of 10. "That coding did a good job on speech but was terrible on everything else," says Richard Stern, an electrical and computer engineering professor at Carnegie Mellon University, in Pittsburgh.

Music-compression algorithms, such as the MP3 and AAC codecs used for streaming audio, are optimized for human auditory perception. For example, the algorithms don't bother to accurately reproduce the soft components of sounds likely to be masked by louder sounds at other frequencies and times. That method can represent a wider range of sound, but it requires more bits per second than a speech-based codec, Stern says.

The new EVS codec is a hybrid, containing algorithms for both voice and music, and it switches between them as needed. The new voice algorithms are substantially more complex than those of the decade-old 7,000-Hz codec. Rather than being developed around characteristics of specific languages, as earlier codecs were, these are nearly language independent. The music part is the latest low-latency version of the AAC algorithm, developed for real-time streamed communications. Called AAC-ELDv2, it delivers CD-quality stereo sound in a stream of only 32 kb/s by transmitting one stereo

channel plus a lower-data-rate signal that represents the difference between that channel and the other stereo channel.

An important feature of the combined package, says Baumeister, is that EVS is the first codec designed to compensate for packet loss. Such losses degrade voice quality and are inevitable on IP networks such as 4G LTE.

To verify performance of the codec and its loss tolerance, Fraunhofer IIS and II partners—including Ericsson, Huawei, Qualcomm, and Samsung—spent millions of euros on human listening tests. Full-HD Voice quality was possible even at data rates as low as 9.6 kb/s.

The processing power of modern smartphone chips is a key enabler for the new codecs. They can be implemented in digital signal processing chips as the 7,000-Hz codecs in 4G smartphones are, or as apps running on a smartphone's applications processor. The EVS codec "is not complex compared to the apps in a smartphone," says Baumeister.

Because Full-HD Voice can tolerate packet losses, it could feed compressed data directly into the Internet data stream for routing directly to other equipped devices, like a Skype-to-Skype call between computers or smartphones. Fraunhofer's Mobile World demonstration did that using apps on Google Nexus 5 phones. With no need for network upgrades, Baumeister says, "you could conceptually roll out service this year, but next year is more realistic."

You can hear samples at <http://www.full-hd-voice.com>, but be sure to use good headphones in a quiet environment. Stern compares the difference to the shift from standard resolution to HD television. "It's going to be subtle, not a huge difference in intelligibility, but it will sound better and more natural, like a high-quality speaker system," he says.

—JEFF HECHT

SMART SEWERS FOR PUBLIC HEALTH

With robotic samplers, public-health officials will study cities' microbiomes



Newsha Ghaeli reached down an open manhole

one morning in April and grabbed a hose dangling in the rushing streams of the sewer below. As she pulled up the hose, she spotted a wad of toilet paper clinging to its end and recoiled in disgust, accidentally splashing a few drops of dirty water on strands of exposed hair sticking out from her face mask. A Cambridge, Mass., sanitation worker offered comfort: "A little caca won't hurt you," he said.

Ghaeli is an architect and research fellow at MIT who recently helped coordinate a 24-hour effort to collect water samples from the sewer beneath an East Cambridge neighborhood. Every hour, the team pulled up half a liter of precious sludge. Ghaeli and her colleagues were doing the groundwork for an ambitious project

that aims to understand the well-being of a city by tracking its residents' biological and chemical waste. Eventually, with robotic samplers placed below the streets, Cambridge may have a "smart sewer" that will let public-health officials study the city's collective microbiome—the communities of microorganisms that live in humans' guts.

"The idea is to look at patterns of sewage relative to how we live our daily lives," says Yaniv Jacob Turgeman, who until recently was the research director of the project. "We want this to be something that has actionable insights that enable public health in a meaningful way."

The project—dubbed Underworlds and supported by a US \$4 million grant through the Kuwait-MIT Center for Natural Resources and

the Environment—is one of a few with that goal, and it has been embraced by the city that the research team calls home. "If you get information closer to real time, you can rally resources to try to address a problem before it becomes a much bigger problem," says Sam Lipson, director of environmental health for Cambridge. "Those kinds of metrics are generally not available in public health."

Until recently, this type of "sewage epidemiology" was used mostly to monitor population-level trends in illicit drug use. In Europe, for example, a recent study of sewage from 42 cities throughout the continent revealed that cocaine and ecstasy use was greatest in large metropolises on weekends, whereas cannabis and methamphetamine use was