Vocoding: Creating Digital Voice

BY DON ROTOLO,* N2IRZ

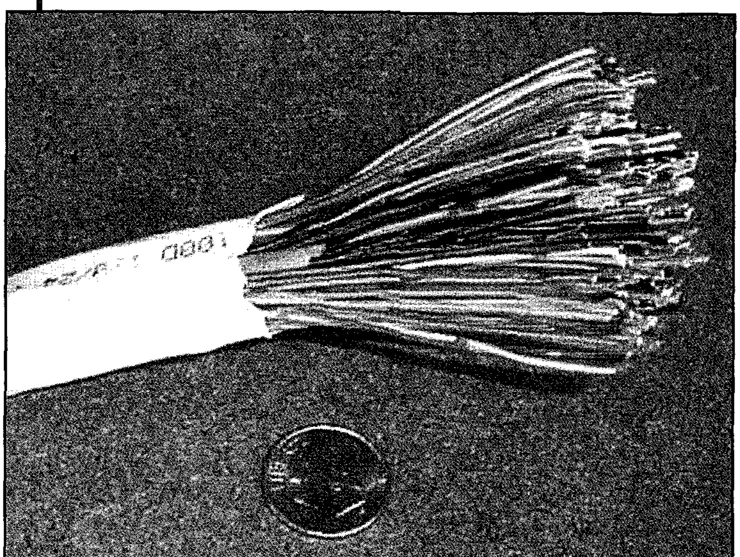
How do we put the digital into digital voice? As digital voice continues to become more popular, I thought we should take a closer look at how it works. Thus, this time we'll swing way over to the technical side and learn quite a bit (pun intended) about encoding a human voice into a digital data stream, a process known as voice encoding, or vocoding.

In the beginning, there was a voice. We used the electronic waveforms that represented that voice to first change the amplitude, and then the frequency, phase, and other characteristics of a radio signal as a means of transmitting that voice over great distances without the burden of running wires. The advent of voice communications over radio was a major driving force in scientific awakening in our culture, the icing on the technological revolution that began in the mid 19th century.

However, radio couldn't replace (or even compete economically with) the telephone, despite the tremendous expense of building and maintaining a wired network and its associated equipment. Ma Bell could add more twisted pairs, or multiplex thousands of voice signals onto a single cable, but radio spectrum was essentially a finite resource.

What does this have to do with digital voice? The short answer is spectrum—or using it more efficiently, to be more precise. The telephone company still has to deliver about 3 kHz of amplitude- and phase-controlled passband through its system and is not concerned as much about spectrum, since it is not limited to using it only once. The phone company can just add another wire, and all of its spectrum is empty and available again. Radio, in order to avoid interference (when *nobody* communicates) can only use a given slice of spectrum once.

*P.O. Box 114, Park Ridge, NJ 07656
e-mail: <n2irz@cq-amateur-radio.com>



A somewhat small twisted-pair telephone trunk cable with only 50 pairs. The phone company can increase the available bandwidth by adding more wires, while radio users have to work a lot harder to increase spectrum usage efficiency. The dime is to show the scale.

However, if we implement ways to fit a pound of flour into a half-pound bag, more communications can happen. Digital voice is that magical "flour compressor" that can make it fit. Now let's take a look at how we manage to smoosh all that voice into a smaller space.

Analog Signal to Data Stream

First, let's step sideways a moment and review how an analog signal is converted into a data stream. First, we take a sample of the voltage of the analog signal and convert that voltage into a number. When it's time to take the next sample, we measure the voltage again, and continue this until we want to stop. How often we take a voltage sample—known as the *sampling rate,* measured in samples per second—depends on the highest frequency we want to capture. A basic principle of analog-to-digital conversion is that you generally need to have a sample rate greater than twice the highest frequency that exists in the signal you are sampling. Google "Nyquist" if you want to learn more about that, including some exceptions. That means a toll-quality telephone signal, with a bandwidth of about 3 kHz, needs about 6 kiloSamples per second (kS/s) as a sampling rate.

The other critical sampling parameter is the number of sample bits, or *bit depth.* For example, we can represent 1024 different voltages with 10 bits, which may or may not yield the desired level of fidelity when converted back to analog. If we generate 10 bits for each sample, at 6000 samples per second, we end up with a data stream of 60,000 bits per second. This amount of data requires far more bandwidth to transmit than the original analog signal, even allowing for data compression and other techniques.

The conclusion is that just digitizing the analog signal waveform actually increases the necessary bandwidth, contributing to spectrum inefficiency, which is exactly the wrong way to go.

Project 25

Then how can we even think of using digital voice on the radio? Like the shady businessman who keeps two sets of books, we use some sneaky tricks. Instead of just digitizing a waveform, we can recognize that the human voice has some very predictable characteristics, and we can exploit those characteristics to dramatically reduce the digitized bandwidth while maintaining that "human voice" sound. One commercially popular digital voice system, called Project 25 (P25), uses a vocoder that implements one such exploitative trick, and that is what I will explain in the rest of the column.

A brief explanation of P25: Radio users from various emergency services and commercial and manufacturing sectors recognized a need for a digital voice communications standard and created Project 25 (http://www.project25.org) to develop and define these standards. It is an open standard (like AX.25 or D-Star), meaning anyone can use it to build a compliant radio or system. It has become arguably the most popular standard for digital voice in the land mobile radio sector, although several other available systems are highly competitive. Amateur radio can learn a lot from the work put into the standards, since most of the lessons are equally applicable to HF channels as they are to VHF, UHF, and above.

It should go without saying that interoperability is one good reason for amateur radio to get involved with standards such as P25. Another really cool thing is that our software-defined radios can—if someone clever programs the mode—also operate with P25 and other digital voice signals. More on that later.

Vocoders

A few years ago, I wrote about Digital Radio Mondial (DRM) and how it was able to fit a near-FM-quality music signal into a 4.5-kHz shortwave channel, using a nifty trick that fools the ear into hearing more than is really there. The energy in a music signal is concentrated below 3 kHz, with only a very small portion of the overall energy content appearing above that frequency. What DRM does is digitize the lower frequencies with good fidelity, and digitize the high frequencies only in terms of the amount of energy in a certain frequency band. These energies are then recreated synthetically at the receiving end. For example, a cymbal crash is characterized as a noise burst in one or more frequency bands, requiring only a few bytes to fully communicate. The receiver synthesizes and recreates approximations of those noise bursts, and the human ear can hardly tell the difference, with a significant savings in required bandwidth. A slightly different encoding scheme is used for voice-only broadcasts with similar results.

Well, voice encoders (vocoders) that claim good fidelity at low bandwidth are as plentiful as used antenna cable, but a certain class of vocoders, known as Multi-Band Excitation (MBE), seems to be head and shoulders above the rest when it comes to delivering on its claims. It should be no surprise that the P25 system has chosen one of these types of vocoders as the standard for all digital voice.

A company called Digital Voice Systems, Inc. (DVSI) has built upon research on voice encoding and MBE that was originally conducted several years ago at the Massachusetts Institute of Technology (MIT), coming up with what is now a family of MBE vocoders. P25 has chosen the Improved Multi-Band Excitation (IMBE) vocoder as its standard, since in testing it significantly outperformed all other vocoder technologies available at the time, even those using a data rate several times higher. Since then, Advanced MBE (AMBE) and AMBE2+ chips have been developed by DVSI, and they are even more efficient than their predecessors.

IMBE is available only as software, while AMBE is available only as integrated circuits. DVSI also sells AMBE chips assembled into evaluation boards or assembled OEM systems. DVSI's technology is a trade secret, but anyone can buy the hardware or license the software.

According to the DVSI website, the IMBE vocoder works by first splitting the voice signal into several frequency bands. It then looks at each band to characterize the audio energy it sees there. Human speech has two major sound components, *voiced* and *unvoiced.* Voiced energy is periodic in nature, containing tones or frequencies, while unvoiced energy is like noise. To better understand this concept, say the word "wash" out loud. The first part of the sound is voiced at a relatively constant frequency, changing in its harmonic content, while the "sh" ending is unvoiced and essentially a burst of noise. The word "hot" has different kinds of unvoiced sounds at the beginning and end (mixed in with some voiced sounds), but they are still noise-like in nature.

Okay, so we take these narrow bands of frequency, and classify the amount of energy from voiced and unvoiced sounds, along with some information about the tone and harmonics of the voiced energy and the dynamics of the unvoiced energy. Rather than digitize the actual analog voice signal, we assign a value to the parameters of each frequency band and send that instead at a raw data rate of 3.6 kb/s. We then use several compression and error-correction techniques (such as Reed-Solomon, Golay, and Hamming codes) to help handle any radio channel fade, noise, or multipath, with an end result of a 7.2-kb/s data stream. (The P25 standard adds data on top of that for control and other purposes, for a 9600-baud on-air data rate).

At the receiver end, we recreate the 3.6-kb/s data stream as best we can using the error-correction information, and then use a bank of harmonic oscillators and noise generators to reproduce the voice signal. You really need to hear it to believe just how good it sounds, and for that DVSI has several speech samples you can hear on its website (http://www.dvsinc.com/).

One downside to the IMBE and AMBE vocoders is that since they are highly optimized for voice, they are poor at reproducing sounds such as DTMF tones. They also are not very good with music, but again, that's not their purpose. If you want music, the Digital Radio Mondial standard (anywhere but the ham bands!) may be a better choice. (For an adaptation of the DRM standard for amateur HF SSB use, visit <http://n1su.com/windrm>.)

The ARD9000 MK2 Digital Voice Modem from AOR uses the AMBE vocoder and FEC, allowing any SSB radio to operate robust digital voice, while occupying no more bandwidth than an analog SSB signal.

What about our software defined radios (SDRs)? Can we use them to operate with P25 radios, for example? The short answer is no—not easily and not yet. The "not easily" part has more to do with the non-voice information that the P25 system uses to direct calls and manage the overall communications *system* that is P25. Since the standard is open and widely available, it's not a problem for someone to build a controller, or write computer software, to allow a software defined radio to communicate in "P25-speak." We just need to find someone who has the skill and interest in doing it. That's the "not yet" part: I don't think it is a matter of *if* someone will do it, just *when.*

Note that to *build* a P25-compliant radio using an SDR is not a small task. Instead, it is a major project requiring thousands of man-hours, along with considerable financial investment (for the IMBE vocoder, for example)—definitely not for the average "Joe Ham." On the other hand, developing such a system has real commercial possibilities. I have not been able to find anything like it in the commercial sector, so here's a market ripe for the picking (just remember me when you make your first million dollars!).

Conclusion

This time we learned something about how digital voice works, and why it's not enough to just buy an analog-to-digital converter and connect it to a radio. There are some very clever techniques we can (and do) use to optimize the digitization of a human voice for radio transmission, and in so doing we can use other digital techniques such as forward error correction (FEC) to greatly increase the reliability and range of our signals – all while occupying less bandwidth than ever before.

Our friends at AOR (http://www.aorusa.com) use the AMBE vocoder in their ARD9800 and ARD9000 MK2 Digital Voice Modems, which I have been writing about for years. I also saw thecompany's ARD25 Multimode Data Receiver at Dayton a few months ago, which allows one to decode Project 25 signals. In other words, amateurs also have equipment available that takes advantage of these technologies.

Again, amateur radio is right around the cutting edge in communications technology. We have SDRs with outstanding capabilities that can be bought for a tiny fraction of what the commercial world has available. We have the capability of developing better and more efficient technologies, if we choose. While some areas of study are not as advanced as in the commercial world, we're not doing too badly, either.

Today's amateur radio experimenter is as likely to use a keyboard as a soldering iron for experiments, and as a digital enthusiast, I can only cheer and encourage you to get involved and have some fun. Until next time . . .

73, Don, N2IRZ