

Practical HF Digital Voice

High-quality voice communication is possible without exceeding SSB bandwidth or expensive broadcast studio equipment.

By Charles Brain, G4GUO, and Andy Talbot, G4JNT

[Editor's note: We goofed! The transmission of telephony in digital format (emission designator J1E or J2E) is perfectly legal in the phone bands. The restriction placed on transmitted baud rate by §97.307(f)(3) of the FCC rules does not apply. In fact, there is no upper limit on the bit rate for this mode. See the sidebar by ARRL Technical Relations Manager Paul Rinaldo, W4RI. Some of this material is from Charles' paper in the Proceedings of the 18th ARRL/TAPR Digital Communications Conference, some from Andy's paper in RadCom, March 1999.]

This whole project began with a conversation over the telephone: Andy said that it would be fun to transmit "real-time" digital speech on the amateur bands. Now there was a challenge! As he is located some 70 km

away over a fairly obstructed path, it would need to be on HF—even more of a challenge!

For several years, digitized voice has been transmitted in a bandwidth comparable with normal analog voice communications using existing transmitters and receivers. After our phone call finished, Charles then went away and had a long think.

Digital Communication Techniques

To fully appreciate why one data communication technique is employed over another, we need to cover digital communication techniques and the problems of the HF environment. When properly implemented, digital communications can show considerable advantages over their analog counterparts

with regard to quality and robustness through noisy transmission media: Consider the quality of CD music recordings over the old vinyl or tape systems and the new digital telephone networks versus the old systems. There are several major issues to be resolved before the conversion is made.

Sampling Rate Selection

To digitize an analog signal such as voice, it first must be sampled; that is, turned into a series of numerical values. Sampling theory dictates that the sampling rate must be at least twice the highest-frequency component present (the Nyquist criterion). Any components at more than half the sampling rate will appear as spurious components at other frequencies, causing distortion. This is called *aliasing*. The high-frequency components need to be removed by conventional filtering before digitization. For a voice signal as transmitted using telephone or SSB, the frequency range of 300-3300 Hz is usually considered important and therefore requires a

7 Elverlands Close
Ferring, West Sussex
BN12 5PL
United Kingdom
chbrain@dircon.co.uk

15 Noble Rd
Hedge End, Southampton
SO30 0PH
United Kingdom
g4jnt@arrrl.net

sampling rate of at least 6.6 kHz. In practice, to ease the anti-aliasing filter's design, a sampling rate of 8.0 kHz is often adopted.

Bit-Resolution and Quantization Noise

Since an analog signal has an infinite number of instantaneous amplitude levels, these cannot be represented exactly; it is necessary to choose a suitable number of levels to represent the signal. Instead of levels, it is more convenient to think of the number of bits (N) needed to give the corresponding quantization: 8 bits give 2^8 (256) absolute levels. Sixteen bits per sample give 2^{16} (65,536) levels. The effect of the random instantaneous error at each sampling point is to add a noise component to the signal, referred to as quantization noise.

A simple rule of thumb can be applied here: The best SNR that can be achieved is given by:

$$\text{SNR} \approx (6N - 1.75) \text{ dB} \quad (\text{Eq 1})$$

The 1.75 dB is a "fiddle factor" that sometimes has slightly different values in various textbooks, but SNR is approximately $6N$. If a figure of 40 dB is taken as good communications quality, then 8-bit quantization—allowing about 48-dB of SNR—would be adequate. This is the system we adopted—although in slightly modified form—on the public telephone network.

Choosing a Data Rate

We can see that for 8000 samples per second, sampling at 8 bits per sample, a total of $(8)(8000) = 64,000$ bits per second (b/s) are generated.¹ The digital telephone network has enough bandwidth with optical fiber and microwave links to pass 64 kb/s directly, but a radio communications link does not have this luxury! At HF, we want to pass digital voice over a bandwidth comparable with SSB (3 kHz). At VHF—if NBFM is taken for the standard channel width—we can increase this figure to 12 kHz, but to preserve the enhanced voice quality that good-SNR FM can give, more quantization levels should be used.

Although it is theoretically possible to transmit 64 kb/s in a 3000-Hz bandwidth, the SNR that is required for a sufficiently low error rate is very high—around 64 dB according to Shannon's information theorem. Therefore, other techniques must be adopted to transmit digitized voice signals. A data rate comparable with

the RF bandwidth is wanted for optimum transmission at SNR ratios that would be just acceptable for poor speech quality: around 3000 b/s for 10-15 dB SNR in a 3-kHz bandwidth.

Choosing a Voice Coder (Vocoder)

A number of candidate systems were studied. The vocoder must operate at a low data rate, be inexpensive, stand alone and be reasonably available. The systems considered were: LPC-10e (linear predictive coding), MELP (multiband excited linear-predictive coding), AMBE (advanced multiband excited coding) and various CELP (codebook-excited linear-predictive coding) systems.²

We experimented with LPC-10e and even managed to implement a version of it on a Motorola 56002EVM. The speech was understandable but we never did get it to track the pitch correctly. Having listened to a commercial implementation of LPC-10e, we decided that it did not have acceptable speech quality anyway.

We then went on to find an implementation of MELP (the 2.4-kbps DOD standard) on the Internet. We got the code to compile and added some Win95 sound-handling routines. The

speech quality was much better, but it consumed about 90% of the CPU resources on Charles' P133 machine. In addition, after contacting the patent holders, we found that they were not at all happy with what we were doing.

We then looked at CELP-based systems. These require large codebooks and clever search algorithms, some things we thought were beyond the ROM capability of the Motorola evaluation module and available programming skills. Finally, we settled on the AMBE vocoder chip manufactured by DVS, Inc.³ This chip is relatively cheap, has very good sound quality, may use data rates between 2400 and 9600 bps, and the manufacturer would sell us some!

The technique adopted codes the voice to reduce the number of bits/s needed for transmission. There has been a considerable amount of research done on various techniques for doing this over the last decade or so, and some very effective compression schemes are now available. The techniques are too complex to cover in any detail here; they usually involve modeling the human voice tract and coding the various elements, such as voiced and unvoiced sounds, in efficient ways.

As an example, GSM mobile phones

Is Digital Voice Permissible under Part 97?

There has been some discussion about Part 97 of the FCC Rules and whether digital voice is "legal." A careful reading of the Rules will show that digital voice is indeed provided for. Read on.

Q. Is HF digital voice classified as "Data," thus subject to the provision in §97.307(f)(3), namely "The symbol rate must not exceed 300 bauds..."

A. No. It is "Phone," also called "Telephony." The *Data* symbol-rate limitations do not apply to this mode.

Q. What is the emission designator for HF digital voice?

A. Digital voice is *Phone*, defined in §97.3(c)(5) as: "Speech and other sound emissions having designators with A, C, D, F, G, H, J or R as the first symbol; 1, 2 or 3 as the second symbol; E as the third symbol." (It rambles on...)

The first symbol of the emission symbol depends upon the modulation of the main carrier. Typically, the output of the digital-voice modem would be fed into a single-sideband, suppressed-carrier (SSB-SC) transmitter, in which case the first symbol would be "J." (If the main carrier of the transmitter is modulated in some other way than SSB-SC, then choose from the permissible ones: A, C, D, F, H or R, which are explained in §2.201 in Part 2 of the Rules, readily available in *The ARRL's FCC Rule Book*.)

The second symbol in this case is "2," meaning: "A single channel containing quantized or digital information with the use of a modulating subcarrier, excluding time-division multiplex."

The third symbol is "E" for "Telephony."

So, the most likely HF digital voice emission symbol will be "J2E."

Q. Will other amateur stations think that digital voice stations are unauthorized or even intruders?

A. It's likely that some will, until digital voice is more familiar and accepted. Old timers will recall that, in the days of yesteryear when wall-to-wall full-carrier DSB-AM reigned supreme, the introduction of SSB wasn't without angst. The best approach is to follow **The Amateur's Code** and inform other stations on conventional SSB what you're doing.—*Paul Rinaldo, W4RI, ARRL Technical Relations Manager*

¹Notes appear on [page 8](#).

use a technique that allows transmission at 13,000 bits/s. Whatever technique is used for voice encoding, there is usually a trade-off between data rate and the quality of the resulting speech. Some of the early systems had a very synthetic-sounding, "Dalek-like" result. Modern variants provide very much better toll-quality speech.

AMBE appears to offer major improvements over earlier systems. It moves away from the concept of modeling the voice tract and instead models the spectrum of the signal every 20 ms. Not many technical details appear to be available to date, as it is still a commercial system. Nonetheless, the results of test programs show the technique to be better than any of the 'ELPs. It has thus been adopted for at least one of the new satellite-based mobile-phone systems. Much more importantly for us, a single-chip solution is available for converting from microphone input to encoded digits. So rather than try to write vocoding DSP software based on published algorithms, we decided to just buy a chip to do the job.

The AMBE1000 chip by DVS implements the whole process and provides the user with extensive tradeoffs between data rate and link quality, as well as forward error correction (FEC). Eventually, the data rate we adopted was 2400 bits/s of voice data plus 1200 b/s of FEC, giving a total of 3600 b/s to be transmitted over the RF link. The IC produces samples every 20 ms and can be regarded as a real-time system in this sense. Any 20-ms samples that get lost just create glitches in the speech that cause minimal disturbance and often go unnoticed.

Programming the Vocoder Module

In use, the AMBE chip must be programmed at turn-on to set the operating conditions, and the easiest way to do this is to include an on-board PIC microprocessor. The digitized output samples at a rate of 3600 bits/s are sent via an EIA-232 interface to the modem—in packets of nine bytes for each 20-ms frame. The data rate for this part of the link is 19.2 kbaud. If you do the math on this, you will find there is a lot of spare capacity for programmers who want to use the development board for their own purposes. An example of this would be for inclusion of data and control signals.

Choosing a Modem

After a literature search, we came to the conclusion that the HF modem must use parallel-tone, PSK technol-

ogy.⁴ (See Fig 1.) It is relatively easy to implement and well proven; it runs on Charles' DSP evaluation module and is more suitable for digital voice transmission than serial-tone modems. Serial-tone modems tend to produce long bursts of errors when the equalizer fails, as opposed to the more random errors produced by a parallel-tone modem. Speech is unlike computer data in that occasional errors do not significantly affect its intelligibility.

PSK

In bipolar phase-shift keying (BPSK), instead of changing the transmission frequency for binary 1s or 0s, the phase is reversed—or effectively, the signal is inverted—between 0 and 1 states. It is possible to show that there is at least a 3-dB improvement in SNR-versus-error-rate performance over frequency-shift keying (FSK) given an "ideal" demodulator for each mode, and very much better than this is possible in practice. PSK has begun to replace keyboard-to-keyboard RTTY on the amateur bands recently in the form of PSK31 (see the articles in December 1998 and January 1999 *RadCom* or July/August 1999 *QEX* by Peter Martinez, G3PLX). For very nearly the same data rate as RTTY, the bandwidth needed has shrunk from around 300 Hz to 30 Hz with a corresponding increase in reliability and error rate. By using four phase states instead of two (90° apart, quaternary phase-shift keying or QPSK), it is possible to encode two bits at once without increasing the bandwidth. This does incur a 3-dB penalty because the transmission power is shared between twice as many bits in a given time.

This technique is available in PSK31, where it is included as an option for adding the extra data needed

for FEC in noisy environments. A properly filtered PSK signal has a bandwidth that can approach the baud rate (in fact, PSK31 is optimized to do just this). If it is not implemented correctly—with waveform control and filtering—the bandwidth of the signal can easily spread alarmingly in a manner analogous to CW key clicks.

For the 3600 bits/s needed for the digitized voice experiments, either binary PSK (BPSK) at 3600 baud or QPSK at 1800 baud would be adequate. The QPSK signal at potentially 1800-Hz bandwidth could even be transmitted unmodified over SSB radios. However, while this technique is ideal for UHF or "clean" VHF links, there are particular characteristics on a typical HF transmission path that make simple high-baud-rate signals very prone to errors and frequently unusable.

Designing the Modem

Amateur Radio equipment has very poor filtering compared to military equipment. The filters tend to be quite narrow and have poor group-delay characteristics. This means the modem must use a narrower bandwidth than it would with the equivalent military equipment. This ruled out the MIL-STD-188-110A 39-tone modem.

In the end, we decided on a 36-tone modem, with a baud rate to match the 20-ms frame length of the AMBE vocoder chip. This provides a raw data rate of $(36)(2)/(20 \text{ ms}) = 3600 \text{ bits/s}$ and enough time for a 4-ms guard period. The guard period is required to give the modem some multipath tolerance.

Each tone carries two bits of data in each baud interval. Unlike military modems, our modem has no Doppler-correction tone and no slow "sync-on-data" facility. So far, both of these facilities

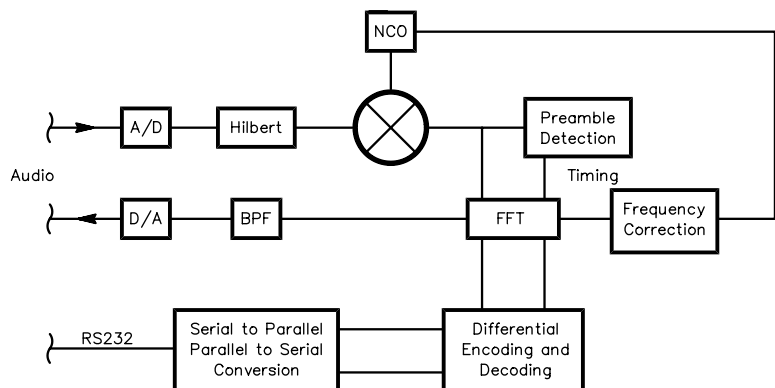


Fig 1—A block diagram of the modem.

have been unnecessary. The modem remains in lock for long periods, well beyond our ability to carry on a dialogue.

We then did some *MATLAB* computer simulations that showed that the modems must be within 5 Hz of the correct frequency to work properly. To achieve initial timing and frequency-offset correction, the modem used three BPSK-modulated preamble tones. It differentially decodes them using a delay of one baud interval. It then integrates the received symbol over that time; from this, it deduces the timing. Then, by looking at the energy in the FFT bins on either side of the preamble tones, it calculates the frequency error and makes a correction by translating the received signal in frequency using a complex mixer. The reason for three tones is to provide some frequency diversity, as on-air testing showed a single tone could get lost during deep fades.

Each symbol consists of 160 samples; the sample rate is 8 ksamples/s. The 36 tones were created by using a 128-point complex FFT. The guard period is added by taking the last 32 samples from the output of the FFT and adding them to the beginning of the FFT samples to form a total of 160 samples. These 32 samples form the 4-ms guard period. The data are differentially coded and mapped to the output phases using Gray coding before transmission.

After the preamble has been sent, the modem sends a reference vector by transmitting a known phase on each of the 36 tones. A "synch" sequence follows this. When the receiving modem detects the synch sequence, it ceases hunting for the preamble and starts passing (we hope!) valid data to the vocoder board.

When the operator releases the PTT, the modem detects the loss of voice data and transmits an EOM (end-of-message) sequence embedded in the data stream. This message is, in fact, the SOM (start-of-message) sequence, inverted. Transmit/receive control of the modem is triggered by the presence/absence of data from the vocoder; there is, at present, no formal protocol between them.

One problem with parallel-tone modems is that they tend to produce signals with very high peak-to-mean ratios. To combat this, our modem uses different initial phases on each of the tones and applies clipping and filtering to the output signal. This allows the transmitter to be driven quite heavily before errors begin to appear in the received signal. The simplest way to set the audio level is to increase

the drive level until ALC action occurs, then back it off a bit.

The modem is capable of full-duplex operation. It does not require a feedback channel and so can be used in broadcast operation; that is, with one sender and many listeners. The modem also incorporates a CW-ID feature to comply with UK regulations—the old meets the new. The CW call sign is hand coded into the DSP software, but it can be switched off. It is not sent at the end of each transmission, but after a programmable period.

Fig 2 shows a compressed spectrogram of an off-air transmission. The three preamble tones can be clearly seen along with the selective fading (diagonal stripes), as can the carrier of an AM broadcast station in the background and a burst of interference at the end. The distinct vertical stripes were, in fact, pauses in the speech.

The greatest problem during operation is multipath. Sky-wave signals frequently arrive after several ionospheric hops, with the same instantaneous element of signal arriving at

different times after having traveled different distances. For a signal such as SSB voice, these two or more signals will cause alternate cancellation and reinforcement, giving the characteristic multipath fading as a notch passes through the audio passband. Differences in arrival time of typically up to 5 ms are often observed, and in poor propagation conditions, can reach a lot more than this.

The effect on digital signals can be much more catastrophic than it is for analog speech. A particular bit of information arrives at several different points in time, so it can easily land on top of another bit arriving via an alternative path. This mixing up of received information causes *intersymbol interference* and is the major cause of bit errors on what might otherwise appear to be a good link with a strong signal.

There is a way around this. If we can send one symbol in such a way that it remains uncorrupted when mixed with a delayed (say, 0 to 5 ms) version of itself, the error rate due to intersymbol interference can be reduced or even elimi-

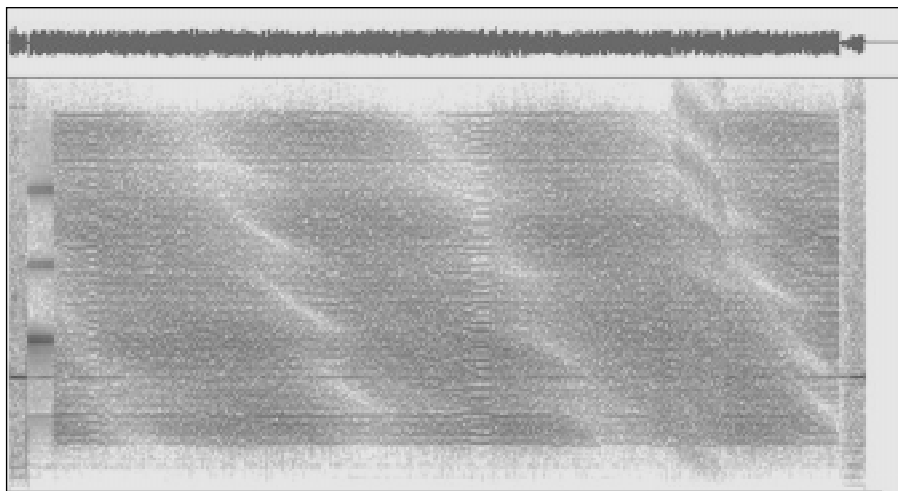


Fig 2—Off-air spectrogram of M36 modem waveform.

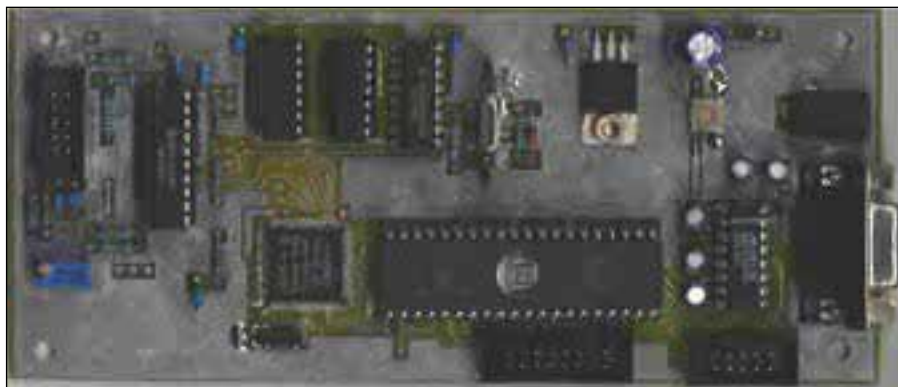


Fig 3—Prototype vocoder board.

nated. One method is to reduce the baud rate so much that the 5-ms multipath period becomes insignificant. A figure of 20 ms is often used in practice, resulting in 50-baud signals. It is no coincidence that the data rate adopted for RTTY signals for many years has been in the 45 to 75-baud region!

To reduce our 3600 bits/s to 50-baud signaling means trying to compress 72 bits into one symbol. While there are some direct techniques of doing this, such as quadrature amplitude modulation, these are prone to other types of errors and inefficiencies. Another system is needed that is more resistant to in-band interference. The technique is as follows.

Instead of using a single carrier modulated with a complex multilevel waveform, we use a large number of multiple carriers, each one modulated with a simple waveform. If there are N carriers, each one independently modulated with 50-baud QPSK, then it is possible to transmit data at $(2)(50N)$ bits/s.

The spacing between each carrier pair must be consistent with the baud rate, and carrier spacing equal to at least the symbol rate is required. If we do a few calculations, it soon becomes evident that many solutions are possible for 3600 bits/s in a voice bandwidth.

FEC

The modem has no inherent FEC capability; instead, it uses the FEC in the AMBE vocoder chip itself. The vocoder tailors the FEC to match the significance of bits in the data stream, so it can probably do a much better job than we can. It is a shame, however, to waste the soft-decision information generated by the modem.

The AMBE chip uses both Golay and Hamming codes for error detection and correction. It follows the normal convention during periods of errors, trying to guess what was sent by looking at previous frames, then ultimately giving up. The format used is 2400 bits/s speech and 1200 bits/s FEC.

The first tests were done without the FEC enabled—whoops! The system worked quite well; but occasionally gave off very loud screeches. After the FEC was enabled, however, fewer strange noises came from the system. When the modem was initially tested without FEC, one third of the tones were in fact transmitting no data whatever and were just wasting energy.

Some experiments were done using



Fig 4—Current digital voice station at G4GUO.

a data interleaver, but they were abandoned because the interleaver adds a large delay to the voice. During deep fades or periods of interference this spreads errors over multiple vocoder frames and so prolongs the dropout.

Development of the Vocoder PC Board

The vocoder board consists of a Motorola MC14LC5480P codec using μ -law coding, an AMBE chip, a PIC17C44JW microcontroller, some HC-series glue logic and an EIA-232 interface. The AMBE is a 100-pin surface-mount chip that Charles soldered onto the board by hand. After five boards, he began to get very tired of doing it!

For the PC board, he used the services of ExpressPCB in the US. In hindsight, this was a mistake; their free PC board software is not compatible with anyone else's, so it pretty much locked us into using them once we had started. Their service is very good, however: Charles e-mailed the files on Monday and had the boards back in the UK by Thursday. He also found most of the components from DigiKey⁵ in the US as well; it worked out cheaper than buying them in the UK, especially for the microcontrollers.

We used the 17C44 PIC microcontroller for a number of reasons: first, so we could use one crystal to drive both the AMBE and the PIC (the AMBE requires a 27-30 MHz clock); second, the 17C44 PIC has enough

ROM available to allow addition of quite complex code at a later date (in fact, Charles has since done a version of his software that can encrypt the speech using triple DES encryption in real-time); finally, because we already had the development tools available. Each board costs us about \$150 to make, and we have so far made five.

On-Air Testing

The system has been tested over a 70-km path using frequencies in the 40-meter band. We made our first successful contact at the first attempt on the 27th of March, 1999. This is not a weak-signal mode; it requires about a 25 dB of SNR to function. When working, however, it makes HF sound like a telephone conversation. There is no background noise—total silence—except for the "comfort noise" inserted during gaps in the speech by the vocoder itself. The system can tolerate strong CW interference and the multipath-induced selective fading found on HF. SSB interference is more troublesome—it affects more than one of the tones. If RTTY/CW interference gets too bad, it is possible to switch a DSP notch filter into the circuit: There is enough power in the FEC to cope with the missing tones. The notch filter must be switched out during the preamble phase.

The most effective and impressive demonstration was one evening in April when a QSO lasted for an hour and a half as the sun set. Copy started out as

perfect, with no lost preambles or garbled messages. The multipath became worse as dusk arrived so copy worsened slightly, but it wasn't until nearly dark when the link had almost faded out completely that it became unusable.

The weakest part of the modem is the preamble phase. To help remedy this, we added the ability to save the frequency offset correction and timing epoch after each successful preamble synchronization. If, for some reason, the receiving modem misses the start of the transmission, it is then possible to press a button on the front panel and revert to the last set of synch information. In a one-to-one QSO, this works most times.

Another change allows the different tones to be given different amplitudes to compensate for the amplitude response of the transceiver. The group delay in the transceiver does reduce the modem's tolerance to multipath. With the new generation of IF-DSP radios, this will not be a problem as their filter characteristics are much more suited to this kind of operation.

Along with the HF testing, Charles has also used the system on 2 m, both on SSB and FM. There is no reason it would

not work via a repeater since there is no ARQ (but we have not tried it).

Conclusion

It is now possible for the home constructor to build—for about \$300—a portable, working digital voice system for HF with near-toll-quality audio. This system can be used equally well to experiment with digital speech using different DSP modems on different frequencies. For further information and a full technical description, plus some sound files, surf along to Charles' Web site.⁶ Once some boards have been made up in the US, we hope to be able to try some transatlantic tests.

Notes

¹Consider CD music recording. A sampling rate of 44.1 kHz is chosen to allow a 20-kHz maximum audio frequency; 16-bit quantization is used to give a dynamic range of greater than 90 dB. With two independent channels for stereo, this results in a data rate in excess of 1.4 MB/s.

²More information on these technologies may be found in L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, (Upper Saddle River, NJ: Prentice-Hall, ISBN 0-13-213603-1); and V. K. Madisetti and D. B. Williams, eds,

The Digital Signal Processing Handbook, (Boca Raton, FL: CRC Press, ISBN 0-8493-8572-5).

³Digital Voice Systems (DVS), Inc may be found on the Web at <http://www.dvsinc.com/>.

⁴It is no coincidence that this low-baud-rate, parallel-tone approach has been adopted for digital TV transmission where 2048 parallel tones are employed in an 8-MHz bandwidth. Multipath on the UHF TV frequencies is typically a few microseconds in duration, and the individual baud rate for each tone is consistent with this. The technique is further refined to minimize bandwidth by using the minimum carrier spacing and ensuring that side lobes from one modulated carrier do not interfere with those adjacent to it. The system is referred to as Coded Orthogonal Frequency Division Multiplexing (COFDM). A similar coding method with 1536 tones of 1-kHz spacing is used for the terrestrial Digital Audio Broadcasting network. Parallel-tone modems are one of the candidate technologies for HF digital broadcasting and there is a lot of professional interest in parallel-tone technology.

⁵Digi-Key Corporation, 701 Brooks Ave S, PO Box 677, Thief River Falls, MN 56701-0677; tel 800-344-4539 (800-DIGI-KEY), fax 218-681-3380; <http://www.digikey.com/>.

⁶G4GUO's Web page is found at www.chbrain.dircon.co.uk/dvhf.html. The page includes sample audio and project updates. □□