THE ANALOG VOICE PRIVACY SYSTEM

Richard V. Cox, Donald E. Bock, Keith B. Bauer, James D. Johnston, and James H. Snyder

***Keith B. Bauer** is a senior technical associate and **Donald E. Bock** is a member of technical staff in the Speech Research Department at AT&T Bell Laboratories in Murray Hill, New Jersey. **Richard V. Cox, James D. Johnston**, and **James H. Snyder** are members of technical staff in the Signal Processing Research Department at AT&T Bell Laboratories in Murray Hill, New Jersey. Mr. Bauer works on real-time hardware for speech processing applications. He joined the company in 1979 and has a B.S. in electrical engineering from Pennsylvania State University. Mr. Bock, who joined the company in 1963, works on real-time hardware for speech processing applications. Mr. Cox does research in speech coding and voice privacy. His work tends to emphasize real-time implementations of algorithms.*

(continued on page 131)

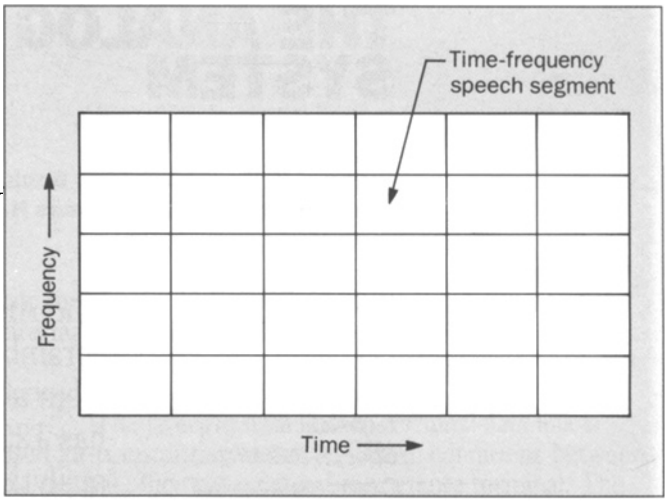
The Analog Voice Privacy System (AVPS) is a voice scrambler that permutes individual output samples from a subband coder analysis filterbank. The system has 125! possible permutation keys, giving it the cryptanalytical strength of a digital encryption system. However, it retains the good voice-quality characteristics of analog scramblers. The AVPS has been implemented in a real-time hardware prototype designed for evaluation in telephone environments and works with any modular telephone and standard 120V ac electrical power. The unit contains two circuit boards—one for analog and one for digital processing—that each use four digital signal processors. To date, we have successfully tested it over long-distance telephone connections, several analog and digital PBXs and telephone switches, and a channel simulator. The quality of the decrypted speech is considered very natural, and speaker recognition is retained—a significant advantage over digital vocoders. This paper describes the underlying principles of the AVPS algorithm, its implementation, and laboratory test results.

Perspective

A telephone bandwidth, voice privacy system has simple goals:

1. The system must transmit the encrypted speech over ordinary analog telephone bandwidths, 200 to 3200 Hz.
2. The encrypted speech must be unintelligible.
3. It must be difficult to decrypt the speech if the decryption key is not available.
4. The decrypted speech must be of good quality. It preserves both the intelligibility of the speech and the characteristics of the talker's voice.

Figure 1. Example of the time and frequency permutation concept. The speech has been divided into five frequency bands and six time segments.



5. Also desirable, any system that satisfactorily achieves the first four goals should also be manufacturable at a low cost.

Today, voice privacy systems in the U.S. are used almost solely by the military. Their widespread use in the civilian and commercial sectors has been inhibited, at least in part, by their cost.

Traditionally, voice privacy systems have been grouped into two categories: *digital encryption systems* and *analog scramblers.*

Digital encryption systems typically include three elements:

- *Digital speech coder*—transforms the speech into a digital bit stream.
- *Digital encryption and decryption unit*—encodes and decodes the bit stream.
- *Modem* (modulator/demodulator)—transmits and receives the encrypted digital information over an analog telephone channel.

In such systems, the limiting factor has been that the modem could reliably transmit and receive only at low bit rates, but the speech quality from speech coders at such low rates was poor. While some low-bit-rate speech coders can maintain intelligibility, the naturalness of the speech and retention of the individual's voice characteristics have suffered.

Recently, promising breakthroughs have occurred in both modem[1] and speech coder[2] design. These breakthroughs could eventually lead to high-quality, low-bit-rate, voice privacy systems. However, such coders are still experimental.

Analog scramblers, the second category of privacy systems, scramble the speech in several ways but do not transmit the speech as a digital bit stream. Instead, the scrambling operation produces a waveform from which the original speech can be derived. The problem with this technique has been that simple scrambling operations, such as frequency inversion or time-segment permutation, are easy to break.

Frequency inversion is like a one-key system. Once an eavesdropper recognizes that the speech is frequency inverted, he or she can reinvert the speech to decrypt it. Time-segment permutation provides many more keys, but requires a substantial amount of processing delay to render the encrypted speech unintelligible. If the segments are made too short, the decrypted speech becomes more noisy. As a result, analog scramblers have been considered less secure than digital systems.

The System Concept. We wanted to design a system that had the good features of both the digital and analog systems. It would have the potential for many digital keys and a digital system's high degree of security, but retain the voice-quality characteristics of a good analog system. The system is based on a technique called time-frequency segment permutation.[3,4]

In the current implementation, we divide the speech into eight equal frequency bands and discard three of the bands. The signal from each remaining band is divided into time segments that are then permuted and transmitted through the channel. In Figure 1, a simple diagram of the concept, we represent the speech as having been divided into five frequency bands and six time segments. There are 30! possible permutations, or over 10^{32} possible keys. We have been able to divide the speech into 25 different time segments for each of the five bands, which corresponds to 125! (or more than 10^{203}) possible keys. In the current implementation, the length of each time segment is a single time-frequency sample.

Our goal here was to produce prototype units for use with any modular telephone on any dialed-up telephone line. We wanted to show that the basic concept would work under a variety of conditions. Therefore, we limited some

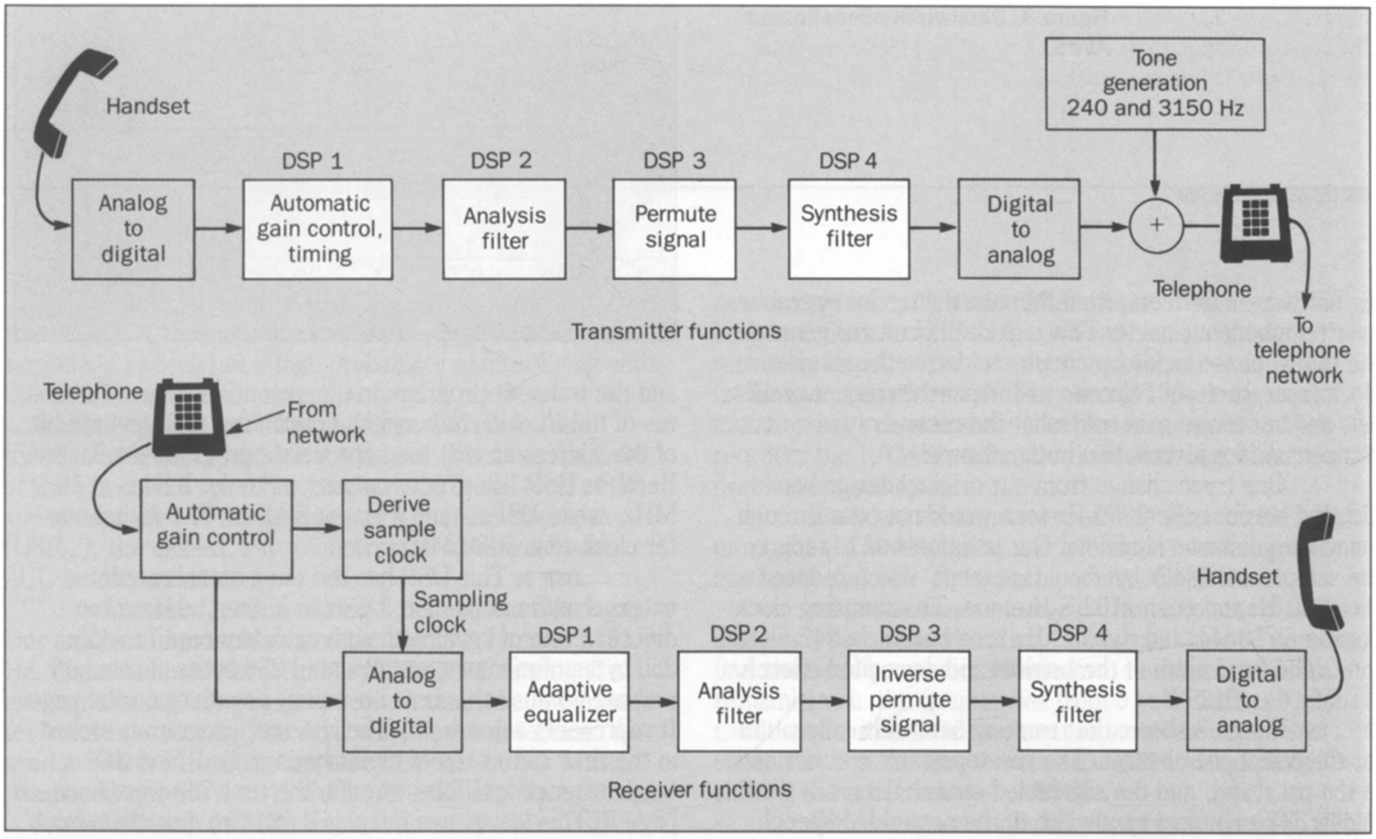


Figure 2. The Analog Voice Privacy System (AVPS). Its receiver and transmitter each contain four digital signal processor (DSP) chips.

features that were not crucial to demonstrating feasibility. For example, although the number of possible keys is very large, we included ROM (read-only memory) space for only ten possible keys.

The implementation described is a half-duplex system; it cannot transmit and receive simultaneously. There were two main reasons for a half-duplex implementation:

- It helped us avoid redesign of the telephone hybrid and design of an echo canceler that a full-duplex implementation would require.
- It only required four DSP (digital signal processor) chips, the most expensive part of the system. A full-duplex implementation would have needed twice as many DSP chips and, therefore, a larger package.

The Original System. Figure 2 is a block diagram of our original idea for the system. The transmitter and receiver would each use four DSP chips.

In the transmitter, the four DSPs are:

- *Silence detector*—determines when to reequalize the channel and resynchronize the receiver.
- *Analysis filterbank*—divides the speech into five 500-Hz bands, from 0 to 2500 Hz.
- *Permuter*—divides the speech into time segments and then uses one of the possible permutations to scramble the speech.
- *Synthesis filterbank*—combines the scrambled speech segments to form a signal in the 500- to 3000-Hz band.

Analog tone generators, driven by the 8-kHz sampling clock, would produce signals of 250 and 3250 Hz to be combined with the scrambled speech. At the receiver, these tones would be used to rederive the sampling clock.

The receiver's first DSP would do phase equalization. Its second, third, and fourth DSPs would do the analysis filterbank, inverse permutation, and synthesis filterbank tasks.

While the basic digital functions have not changed in the implementation we will describe, considerable ana-

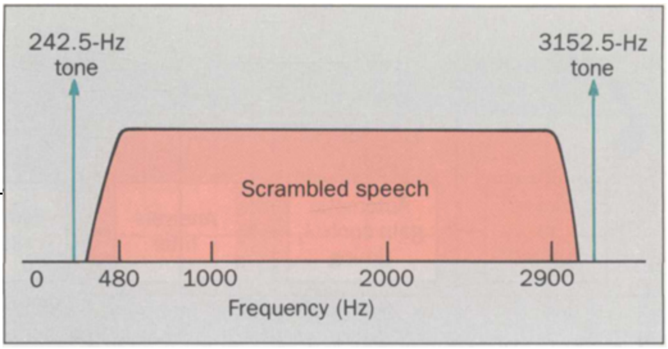


Figure 3. Bandwidth allocation for AVPS.

log hardware has been added to make the device operable over telephone channels. This includes circuits to generate the tones, phase lock loop circuits to derive the sample clock from the tones, circuits to correct the frequency offset, and automatic gain control at the receiver to compensate for attenuation in the channel.

One basic change from our original design was dictated because the 3250-Hz tone would not pass through many long-distance channels. Our solution was to reduce the sample rate clock by about 3 percent, which reduced the 3250-Hz tone to a 3152.5-Hz tone. The sampling clock became 7760 Hz, and the 250-Hz tone became a 242.5-Hz tone. The bandwidth of the original and decrypted speech is really 0 to 2425 Hz.

Figure 3 shows the resulting bandwidth allocation for the telephone channel. The two tones are at either end of the passband, and the scrambled speech data is in the middle. The nominal bandwidth of the scrambled speech signal is 485 to 2910 Hz. However, because the digital filterbanks are not ideal, there are transition bands to either side. The digital filters were designed so that the two tones would be in the filterbank's cutoff regions to prevent the tones from interfering with the speech data. In our later tests, we show that the system's success or failure depends primarily on the successful transmission of the two tones.

Digital Circuits

As Figure 2 shows, the scrambler's principal digital components are the μ-law codec and the four DSP chips. Each DSP is supported by two 2732a EPROMs (erasable programmable ROMs) that contain 4096 16-bit words for program storage. Because a DSP generates a 10-bit address, it can only address up to 1024 words in external memory. We used external logic to generate a 12-bit address, which allows us to use as much of the 4k words of ROM as we need.

DSPs 1, 2 and 4. Each of these DSPs requires two programs, one for transmit and one for receive. The receive program is stored on the first 1k page of ROM, and the transmit program on the second 1k page. The status of the *Push-to-Talk* switch determines the eleventh bit of the address so that the appropriate program is selected. Because DSP 1 has more operations to do, it runs at 10 MHz, while DSPs 2 and 4 run at 5 MHz. The digital master clock runs at 20 MHz.

DSP 3. This DSP has the most complicated external addressing logic, because it must address ten different sets of keys. Although each key requires only 250 bytes of memory, we allocated 256 bytes of memory per key because this is a convenient boundary on the page. It represents a *quadrant.* The receiver program is stored in the first 768 words of memory on page 1, and the transmitter program is stored in the first 768 words on page 3. This leaves two full pages and two quadrants for key storage.

Digital Operations in the Transmitter

The digital functions in the transmitter are still the same as in Figure 2.

DSP 1—Silence Detection and Reequalization Timing. To equalize the phase, the phase characteristics of the channel must be known and, to be known, must be measured. Most conventional telephone channels are nearly stationary, although some radio links may not be. The equalizer must be updated before channel drift renders the previous measurement useless. We update the equalizer about every 2 seconds.

The silence detector tries to find a long enough silent interval to insert the reequalization sequence unobtrusively. Because we don't want to send reequalization sequences continuously during silence intervals, we further required at least a 1-second spacing between successive sequences. This gives the silence detector about a 1.25-second window to find a silence interval. If it can't find one during this time, it will interrupt the speech anyway.

A second function of DSP 1 is automatic gain control (AGC). If the input speech has too high a level, the scrambled speech has a high probability of interfering with the warning signal. On the other hand, if the level of the scrambled speech is too low, the signal to noise ratio (SNR) of the decrypted speech is poor. Because the level of the scrambled speech is proportional to the level of the speech supplied to the transmitter's analysis filterbank (DSP 2), the easiest way to control the level is with an AGC before scrambling.

Another function of DSP 1 is low-pass filtering. Our analysis filterbank retains the signal only up to 2425 Hz. Signals above that frequency can cause aliasing. To prevent aliasing, we put all speech through the low-pass filter before performing the AGC and silence detection. We used a 17-tap FIR (finite-impulse response) filter whose passband region is 0 to 1940 Hz. At 2425 Hz, the speech is down 4 dB and, at 3100 Hz, it is down by over 40 dB. We found that this filter prevented a significant amount of aliasing in the decrypted speech and improved the overall quality.

DSP 2—Analysis Filterbank. This DSP functions as a 0 to 2425-Hz analysis filterbank. It divides the output signal from DSP 1 into five equally spaced bands. Each has a nominal bandwidth of 485 Hz and is sampled at 970 Hz.

The filter-bank structure, which uses 48-tap FIR filters, is based on Rothweiler's generalized quadrature-mirror-filter (QMF) structure.[5,6] The filters were designed to have the following property: Their outputs could be decimated at 8:1, and then interpolated and recombined to reproduce the original signal with a very high degree of fidelity, if sample-to-sample integrity is maintained between the transmitter and receiver.

DSP 3—Permuter. The heart of the scrambling algorithm is the 125! permuter. At any given time, the data in the DSP consists of 125 frequency-band samples. Initially, one can picture the data as being all in order. The permutation key determines the order in which the 125 samples will be passed to DSP 4.

As each sample is sent according to this key, a new sample takes its place. After all the original samples have been sent to DSP 4, the 125 new samples that replaced them are now in a permuted sequence in RAM (random access memory). The second 125-element permutation key releases these samples in a sequence that reorders the RAM according to time of arrival when the first key is used again.

Each key contains two permutations of size 125, or a total of 250 elements, and is divided into five parts, one for each band. Thus, each part has 50 elements.

DSP 4—Synthesis Filterbank. The telephone channel includes a high-pass filter that removes 60 Hz and its second and third harmonics. Therefore, no data can be transmitted in the frequencies from 0 to 200 Hz. As a result, the synthesis filter must transmit the scrambled speech data in the frequencies 485 to 2910 Hz, in effect shifting the spectrum up by 485 Hz.

We reassigned the band numbers so that 485 to 970 Hz is known as band 1; 970 to 1455 Hz is band 2, etc. Each synthesis filter interpolates its band by 8:1, which means that the output rate of DSP 4 is equal to the original sampling rate, 7760 Hz.

The bandpass nature of the telephone channel (200 to 3200 Hz) is our real motivation for reducing the bandwidth of the encrypted speech to 2425 Hz. This is all we can squeeze through. Ordinarily, eliminating the upper three bands of the analysis filter would cause aliasing in the fifth band, but the low-pass filter we used in DSP 1 prevents it. Eliminating bands 6, 7, and 8 in the receiver does not cause additional aliasing, because zeros are transmitted in those bands in the transmitter's synthesis filterbank.

The synthesis filterbank is based on the prototype low-pass filter used to design the analysis filterbank. However, the phase of each synthesis bandpass filter is the opposite of its analysis filter counterpart. Their impulse responses are reversed in time, and their amplitudes have been increased by a factor of 8 to account for the interpolation. The combination of two filters with opposite phase gives a total filterbank with linear phase.

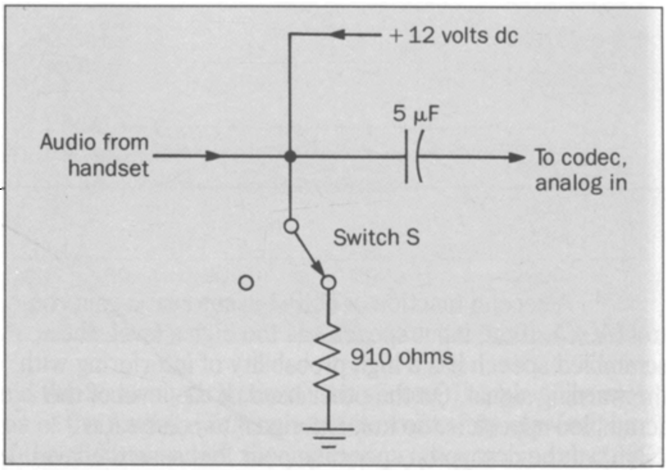


Figure 4. Analog interface circuit connects the carbon-button microphone to the transmitter's analog to digital converter. For handsets with an electret microphone, the switch S must be open.

Analog Operations in the Transmitter

Originally, the analog circuitry in the transmitter was simple. It generated the 242.5- and 3152.5-Hz tones and set the transmission levels of the tones and the scrambled speech. But as the project evolved, we realized that the characteristics of the telephone sets and switching systems through which the scrambler had to perform varied more than we had expected.

We needed to provide different input circuits for the two most common types of microphones: carbon button or electret. We also needed to accommodate the 18-dB variation in phone-to-phone gain of the telephone switches. It permits the user to set the scrambler's output level according to the type of switch.

Tone Generation. The system's master clock is an 11.919-MHz crystal (4.096 MHz × 3 × 0.97) on the digital board. Two digital counters divide this signal into a frequency of 93.12 kHz, which is further divided by 12 (yielding 7760 Hz) and 32 (yielding 2910 Hz).

The 7760-Hz signal is divided by 32 to give a 242.5-Hz signal that is passed through a high-Q analog filter to provide a sine wave with a SNR of better than 50 dB. This filtered signal then is modulated by the 2910-Hz signal. The upper product (3152.5 Hz) is selected by two filters—a notch filter at 2667.5 Hz, and a high-Q bandpass filter at 3152.5 Hz—to create the other sidetone.

Finally, the 242.5- and 3152.5-Hz tones are combined with the scrambled signal from the digital to analog (D/A) converter and transmitted over the channel.

In addition, the output signal is fed back into the receiver's analog AGC circuit. The two sidetones are then used to generate the sampling clock, which drives the transmitter's A/D (analog to digital) and D/A clocks. This also ensures that the receiver and transmitter have identical sampling clocks. Deriving the sampling signal this way expedited switching between transmit and receive modes.

Microphone Switch. Because we designed this scrambler to be placed in line between the handset and the telephone set, it must interface directly to the handset's microphone. Figure 4 shows the analog circuit that connects a carbon-button microphone to the A/D converter. For telephone handsets with the newer electret microphone, this interface is slightly different.

We added the switch S, shown closed in Figure 4. It must be open for the electret microphone, which has a much higher output impedance and less output drive capacity. As most AVPS terminals would be attached indefinitely to a modular phone, this switch—which is mounted on the back of the scrambler chassis—will only need to be set when the AVPS unit is installed.

Analog and Digital Telephone Switches. The first time we tried the system over a digital switch, we found that digital PBXs (private branch exchanges) have virtually no loss, unlike analog switches. (By *loss* in the following text, we mean that we measured the loss from the output of the transmitter AVPS to the input of the receiver AVPS.) An analog switch or PBX typically has 18 dB of loss. In addition, standard desk telephones have a 6-dB loss over most of the bandwidth and an additional 6-dB loss at 240 Hz in the hybrid. But the new telephones with an electronic hybrid have no loss over most of the band, only the 6-dB loss at 240 Hz.

If we set the signal level for an analog (passive) hybrid telephone and analog switch but used the AVPS unit on an electronic hybrid and digital switch, the transmitter's output level would be above the standard limits for line signals. Therefore, we opened the scramblers and internally adjusted them to the correct transmission gain. However, the user would need outside adjustment to set the gain.

For a conventional telephone connected to an analog switch, a user would probably set the gain at –6 or –12 dB, depending on which sounded better. To talk over a digital PBX with a new electronic telephone, the user would probably choose either –24 or –30 dB, again based on personal preference. The gain control knob, which we placed on the back of the unit—next to the microphone switch—can be adjusted at any time, even during the conversation. If it is set incorrectly, the unit will still function but the scrambling-descrambling process will inject significantly greater noise.

Analog Operations in the Receiver

In the original proposal for scrambling through time and frequency sequential permutation (TFSP),[4] the entire scrambler was simulated digitally. We even suggested how the different parts of the scrambler could be implemented on DSP chips. However, analog processing in the receiver is far more efficient and economical than a digital implementation.

The receiver has the most extensive analog circuitry. It's analog *front end* does three basic functions:

- Automatic gain control to set the correct level for the signal
- Derivation of the sampling clock for the digital processing
- Correction of any frequency offset that may have occurred in the channel.

The receiver also has an analog *back end.*

The bandwidth of the decrypted speech is only 2425 Hz, but an anti-aliasing filter in the μ-law codec is much wider. We discovered that the quality of the decrypted speech could be substantially improved if an analog bandpass filter was used after the codec. In addition, a final stage, adjustable amplifier allows the listener to set the output level to his or her preference. Also, both front end and back end have telephone-line interfaces that are transformer coupled to the telephone lines. The back-end interface is isolated from ground to prevent grounding the telephone line.

Receiver Analog Front End. The diagram in Figure 5 describes the receiver analog front end. The signal from the telephone-line interface is buffered and then sent to the AGC circuit. The level of AGC is inversely proportional to the level of the 3152.5-Hz signal that it received from the telephone channel.

Notch filters extract the two tones that are modulated via a phase lock loop to derive the sampling clock. The loop's frequency is 93.12 kHz, or 32 times the difference frequency of the two tones. The frequency-offset-compensation circuit first modulates the entire signal by 3637.5 Hz minus the frequency offset and then demodulates it by a pure 3637.5 Hz. This circuit's output is fed to the A/D converter.

The AGC circuit has a range of 40 dB. A nominal gain is set for the expected loss in the telephone line as follows. The output of the 3152.5-Hz notch filter is rectified, and the peaks of the rectified signal are compared to a reference voltage. If the peaks are smaller, the gain is increased; if larger, the gain is decreased.

Figure 5 shows two such notch filters:

- The first filter, which consists of low-pass and bandstop sections, isolates the 242.5-Hz tone. Its bandstop output contains the input signal minus the 242.5-Hz tone. Its low-pass output contains the 242.5-Hz tone and is fed to the modulator for deriving the sampling clock.
- The second filter, which consists of high-pass and bandstop sections, isolates the 3152.5-Hz tone. Its bandstop output contains the input signal minus both the 242.5- and 3152.5-Hz tones. The band edge of the high-pass section is 3152.5 Hz. This output is fed to the modulators and the AGC circuit.

The two tones are modulated together to produce a 2910-Hz signal. If there is some frequency offset in the channel, then both tones will be offset by the same amount, but their difference will still be 2910 Hz. Thus, we can use this 2910-Hz tone to drive the sampling clock.

For modulation, we use a switching modulator. The 3152.5-Hz tone is converted into a digital signal that drives the switching modulator and switches the 242.5-Hz tone on and off. The resulting signal has many harmonic components; the principal ones are at 242.5, 2910, 3152.5, and 3395 Hz. All but the 2910-Hz tone are filtered out before the signal enters the phase lock loop.

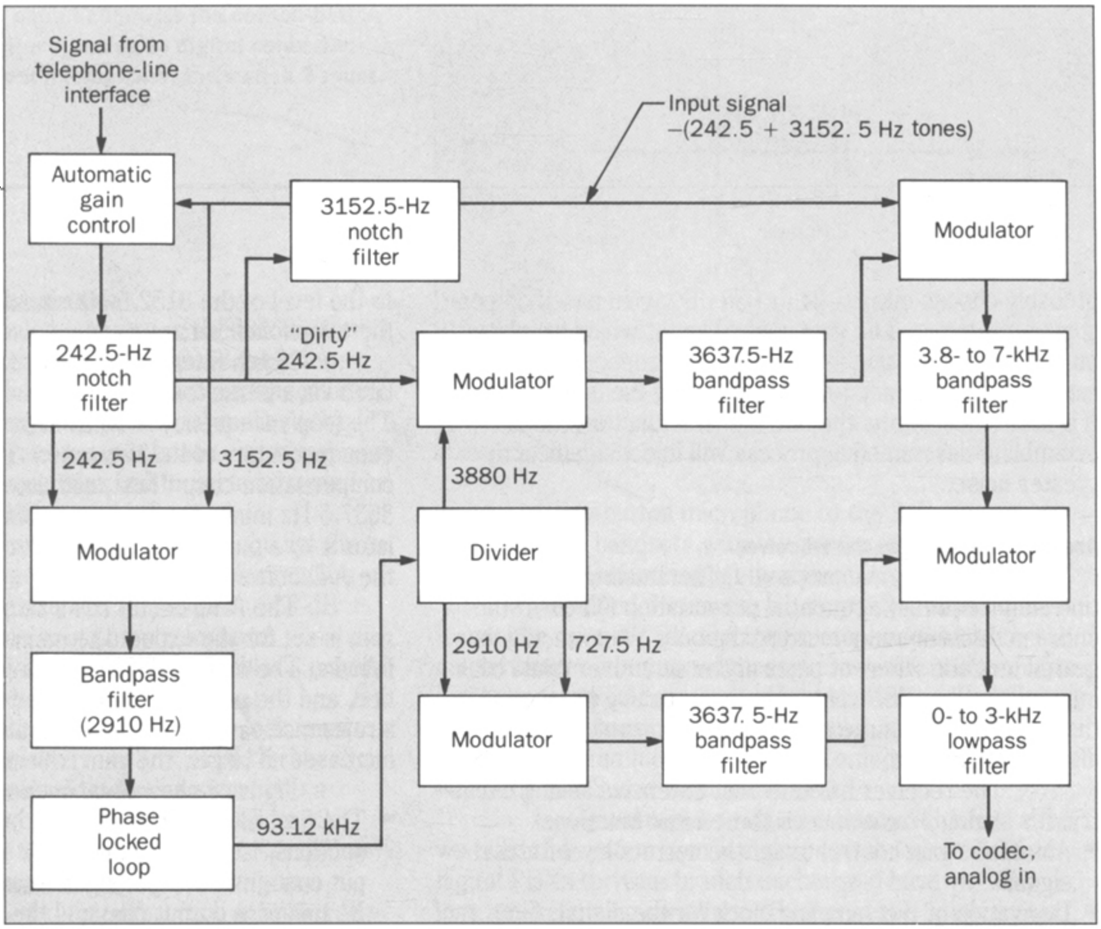


Figure 5. Analog circuits for the receiver's front end.

The 2910-Hz component falls inside the band of the analysis filterbank and is then turned into white noise by the inverse permuter. One way to reduce this signal's level is to reduce the level of the 242.5-Hz tone at the transmitter. We found that the 242.5-Hz tone could be received at a level 27 dB below that of the 3152.5-Hz tone without disturbing operation of the phase lock loop. The clock for the phase lock loop, at 93.12 kHz, can be divided to produce the 7760-Hz sampling clock and the other frequencies—such as 727.5, 2910, and 3880 Hz—required for the frequency-offset compensation.

Frequency-offset compensation works on a simple principle. If the entire signal has been shifted by a fixed amount, *s* Hz, then we can shift it down by first modulating up by 3637.5 – *s* and then down by 3637.5 Hz.

In the first stage of the frequency-offset-compensation circuit, we generate a *dirty* 3637.5-Hz tone by modulating the 242.5-Hz notch filter output with the 3880-Hz signal derived from the 93.12-kHz clock. This produces a 3637.5 – *s* Hz signal and other harmonics, which the 3637.5-Hz bandpass filter must remove. The signal is then used to modulate the output of the second notch filter (Figure 5) that contains the scrambled speech data. This modulator's output is bandpass filtered. By modulating the pure 2910- and 727.5-Hz tones, we derive a pure 3637.5-Hz tone that can then be used to demodulate the scrambled signal. The resulting signal is then restored to its original bandwidth.

There is also an upper-product image—beginning at 7275 Hz—that is equal in magnitude to the baseband image. We found that the anti-aliasing filter in the μ-law codec did not remove enough of the upper image. To correct this, we added another low-pass filter with a cutoff point at 3000 Hz and a stopband that begins at 7000 Hz. It reduces signals above 7000 Hz by at least 36 dB. When used with the μ-law codec anti-aliasing filter, this filter eliminated the noise that the upper image caused.

Receiver Analog Back End. As mentioned above, we found it necessary to filter the output from the μ-law codec. Perceptually, we determined the best bandwidth for the back end's final stage filter. If the filter was too sharp, the resulting speech reverberated.

Also, both a high-pass and a low-pass section were desirable. The high-pass section, which *crispens* the speech, was designed for a stopband frequency of 100 Hz and a transition band of about 150 Hz. The low-pass section eliminates the high-frequency aliasing that the too-wide anti-aliasing filter could not remove. This section has a cutoff of 2400 Hz and a stopband of 3000 Hz. While their effect is not dramatic, these sections decidedly improved the quality of the decrypted speech.

Because the codec output is too high for the speaker in a telephone handset, we also included a buffer stage in the back end. This buffer has an externally adjustable gain control that the listener can set to his or her preference.

Digital Operations in the Receiver

As Figure 2 shows, the digital operations in the receiver consist of phase equalization, analysis and synthesis filterbanks, and an inverse permuter to unscramble the speech. Because the analysis and synthesis filter programs are similar to those of the transmitter, we do not need to discuss them further. Instead, we will concentrate on the equalizer, synchronization, and individual-sample permutation.

DSP 1—Equalizer. In the receiver, DSP 1 is used for the equalizer. System performance depends greatly on how reliably this DSP recognizes the warning signal. Our warning signal consists of two wideband 1-kHz tone bursts that are separated by silence.

The detection algorithm requires that three different events occur—a tone burst, followed by silence, followed by tone burst—all with precise timing. Both tone bursts must be in the correct relative phase, and the silence interval must be between them. Scrambled speech may give the same behavior, but is unlikely to duplicate it. The thresholds for the matched filters are adaptive. Each time a warning pulse is detected, the thresholds are reset, taking into account the most recent detection.

Once DSP 1 detects a warning pulse, it goes into measurement mode. During this time, there is no output from the equalizer. We know the approximate time from detection of a warning pulse to arrival of the channel impulse response. Slightly before the time of the expected arrival, DSP 1 begins to save the input samples.

The next step is to search through the samples to find the maximum value (the maximum value of the impulse response is close to the beginning). We store 64 samples as the impulse response, beginning five samples before the maximum value occurs. We have found empirically that five samples is a good value.

At this point, the equalizer goes back to normal mode. It resumes scanning for a warning pulse and resumes equalization, using the new equalizer coefficients.

DSP 2—Synchronization Detection. The synchronization signal is really the impulse response of the 1940- to 2425-Hz synthesis filter. After a reequalization, when DSP 1 sends its *s*1 status bit to signal DSP 2, the analysis filterbank uses only this filter.

There are eight possible decimation phases. For the correct decimation phase, this filter's output will look like an impulse with four zeros on either side. The output of the other seven phases will have additional samples with significant values. We use a matched filter approach to determine the correct phase.

DSP 3—Inverse Permuter. After DSP 2 detects the synchronization signal, it signals DSP 3 that the new key is already underway and that the inverse permuter has to begin the key in a different place. Otherwise, the permuter's operations are still normal. Thereafter, each time the key is repeated, DSP 3 reads it from the beginning.

Performance Results

The usefulness of an analog encryption system depends on its behavior in four areas:

- Resistance of the system to casual eavesdropping.
- Resistance of the system to determined penetration.
- Fidelity of the decrypted system to the original waveform.
- Robustness of the system in the presence of channel impairments. (Loosely speaking, *impairments* also include the variation in specified performance of telephone sets and switching systems.)

To characterize our system's robustness, we observed the changes in the decrypted voice quality when channel impairments occurred. In addition, subjective testing determined the residual intelligibility of the transmitted signal. The comments of those who have tried our system support our claim that its speaker recognizability is far superior to that of vocoders.

Channel Simulator Tests. One method used for testing modems is to check their performance over several different possible telephone channels. Commercially available telephone channel simulators, such as the AEA S-3, provide this capability. The AEA S-3 could vary the following parameters:

- Channel SNR
- Color or spectral shaping of the noise
- Phase jitter at 120 Hz
- Frequency offset
- Second- and third-harmonic distortion
- Envelope delay divided into three bandwidths
- Spectral tilt
- Channel roll off at both the low and high frequency ends.

With this simulator, we tested a pair of AVPS devices by varying each parameter individually. For each, we tried to detect when the device began to fail, i.e., when the quality of the decrypted speech degraded.

The system was most sensitive to the bandwidth of the channel. The AGC at the receiver is based on the level of the 3152.5-Hz tone, so the system is more sensitive to additional attenuation of this tone. If this tone is attenuated more than the scrambled signal, too much gain will be applied to the input signal. To compensate for this, we built some tolerance into the digital algorithms.

Bandwidth tests on the simulator showed that the system cannot tolerate totally losing either the 3152.5- or 242.5-Hz tones. A loss at either frequency relative to the scrambled signal will not disable the system but will introduce additional noise.

As a whole, the tests showed that the AVPS was robust. It can withstand some severe conditions, such as severe phase jitter and severe phase distortion; most modems would not do as well. In other areas, such as harmonic distortion and channel SNR, the AVPS performs robustly before breaking down, and its behavior is probably similar to most modems. Its one great sensitivity is to attenuation at the ends of the bandwidth. We will propose a solution for this problem later.

Phone Circuit and Switch Tests. Before the channel simulator tests, we did some field testing of early models of the system. From them we learned about the characteristics of the different PBXs, electret telephones, etc., that led to our final design. In all these tests, our measure of performance was the quality of the decrypted speech. If the background noise was low and performance was reliable, we considered the test a success.

Our initial development of the AVPS was on the local loop at AT&T Bell Laboratories in Murray Hill, New Jersey, where the telephones are connected to an ESSX telephone switch. It has all the standard characteristics of an analog switch, most noticeably the amount of attenuation between the transmitter and receiver. Once we had learned about the proper transmission levels and the amount of loss to expect, the AVPS performed well.

Next, we tested the AVPS over three different types of PBXs: the Dimension® 2000 analog PBX, and the System 75 and 85 digital PBX. Our experiences with digital PBXs led to further modifications, such as the transmission level adjustment for digital switches and the interface for the electret microphone handsets used in newer PBXs instead of the carbon-button microphone. The system worked well with these changes.

Our long-distance testing has been more limited.

We believe that loops between Murray Hill and Holmdel, New Jersey, were made on T1 lines. In addition, we tested the AVPS on long-distance DDD (direct distance dialing) calls and calls on the AT&T Cornet system.

On long-distance calls, the AVPS's greatest sensitivity is to narrow bandwidth. If either the 242.5- or 3152.5-Hz tone is lost, the system fails. In addition, if the received level is too low, then the levels of these two tones will be lost in the channel noise, and the system will fail.

Residual Intelligibility Tests. The digit intelligibility test has been used frequently[3] as a benchmark for residual intelligibility of the channel signal. In this test, listeners are told that they will hear four-digit strings of numbers and, for each string, are to write down the four digits they heard. An average score of 10 percent would correspond to pure guessing, implying perfect scrambling.

Such a test has been applied to time-segment permutation, frequency inversion, those two combined, and time-and-frequency segment permutation scramblers. However, no previous systems had time segments as short as 1 ms, as in the current AVPS.

In a test based on a computer simulation with 16-ms segments and a communication delay of 256 ms, the overall recognition rate was 24.5 percent.[3] The longer communication delay allows more mixing of segments. Apparently, the effect of the shorter segment lengths compensated for the effect of less mixing of segments. The current system has a communication delay (from scrambling) of 50 ms. We hoped that going to 1-ms segments would further decrease the residual intelligibility.

We performed an informal digit-intelligibility test where new speakers recorded the digit strings. Therefore, we cannot compare these results directly with those for male and female voices listed in the previous tests. Nevertheless, the total results should be comparable.

Ten subjects listened to 30 four-digit strings for each speaker, a total of 2400 digits. The strings were balanced so that each digit occurred 240 times in the test. To reduce listener fatigue and masking, we filtered out the two tones so that only the scrambled signal in the 485- to 2910-Hz band was heard. Any clever eavesdropper could do this bandpass filtering, and we don't consider the tones to be part of the system's security.

Table I compares the results of this test (1-ms segments) and the previous test (16-ms segments). The overall digit intelligibility, 22.7 percent, was reduced only slightly from the other test. If we examine the recognition rates for the individual digits, we see that the new system's lowest recognition rates were for the digits 0, 1, 4, 5, 8 and 9. For the digits 2 and 3, its recognition rates are still low, 15 and 18 percent, respectively. The results for 6 and 7 are most perplexing, with recognition rates of 79 and 52 percent, respectively. We suspect that only a larger communication delay can make the scrambled versions of these digits sound like the other digits. Because the rates are lower for the other eight digits, we believe that the current AVPS is a stronger scrambler than all the previously cited systems.

There is an alternate way to examine the results. Table I shows the fraction of correct guesses for digit x, given that digit x was actually spoken. The alternate possibility is the fraction of correct guesses for digit x, given that digit x was guessed. This method of analysis was not discussed in the analysis of our previous tests. Table II shows a comparison for the same scramblers as Table I.

Here, the individual variation between the digits 6 and 7 and the other digits is less than in Table I. If a person guesses that the digit 6 was spoken, he or she is probably not correct. Thus, 60 percent of the time that the digit 6 was guessed, the guess was incorrect. In other words, the scrambler tends to make other digits sound like 6. It also tends to make them sound like 7, 2, or 3 since those four digits got a disproportionate share of the guesses, as Table III shows.

It is also worth noting that, of 600 strings, no one correctly identified an entire string. The digit intelligibility test is a worst-case bound on residual intelligibility. Listeners knew that only ten possible words were being spoken, and each string had four of them. They could use any clues available to help guess the digits. Even so, the

Table I. Digit Recognition Rate (Percent) by Digit Spoken

Scrambler Characteristics	Digits 0	1	2	3	4	5	6	7	8	9
16-ms segment; 256-ms delay	27	17	16	17	21	48	44	22	18	20
AVPS 1-ms segment; 50-ms delay	12	8	15	18	5	17	79	52	10	11

Table II. Digit Recognition Rate (Percent) by Digit Guessed

Scrambler Characteristics	Digits 0	1	2	3	4	5	6	7	8	9
16-ms segment; 256-ms delay	23	16	22	22	23	34	33	22	24	24
AVPS 1-ms segment; 50-ms delay	17	17	11	14	9	19	40	32	15	23

recognition rate for complete strings was zero. This suggests that, unless the strings consisted only of sixes and sevens, it would be safe to transmit digit strings over the scrambler. Using the complete vocabulary of English words rather than just ten digits would be even safer.

In this test, we added one more question. After a subject listened to 30 strings by one speaker, we asked if that speaker was male or female. The question was repeated after the second set of 30 strings. Because one person already knew which was which, his responses were thrown out. Of the other 18 responses, nine subjects were right and nine were wrong. This result is an indicator of how well the scrambler hides the speaker's identity.

User and Listener Reactions. Since completion of the early prototypes, we have been giving briefings and demonstrations of AVPS to AT&T personnel and potential users. Their reactions to AVPS as users describe its perceived strengths and weaknesses.

Most listeners agreed that the naturalness of the voice quality was the system's strongest point. They found it more natural than vocoder speech and felt that the talker was more identifiable with AVPS than with a vocoder. Negative reactions centered around two points:

- The security of any analog scrambling system versus digital systems
- The half-duplex or Push-to-Talk nature of the system.

There were also questions about possible key-distribution systems.

The system's half-duplex nature is not as big a problem as it might be. Many military communication systems are four wire and would not require a special equalizer designed for AVPS. Thus, full-duplex capability could be offered immediately. In addition, some civilian applications, such as cellular mobile radio, can be classified as four wire in nature.

For the future, we can study how to achieve full-duplex operation over two-wire circuits. The problem is not necessarily insoluble; echo cancelers placed at the correct points may solve it.

In general, analog systems are perceived as having less strength than digital systems. We believe that an analog system using a secure digital key can be made just as secure as a digital system using the same key. We have no exact quantification of the strength of AVPS. (A companion paper[7] attempts to establish bounds on its cryptanalytic strength.)

The important point is to roll the keys so that every block must be broken independently of previous and future blocks. Using individual sample permutation in the time-frequency matrix should render each block unbreakable by itself. With rolling keys, no long-term statistics could be used to break a key. A simple microprocessor could be used to change the key and distribute the keys. We believe that the customer, in part, should determine the method of key distribution. The current system is flexible enough that a key distribution system could easily be built into AVPS if one more microprocessor were included. The microprocessor used for key distribution could also be used to change the keys.

To summarize, it would be fair to say that listeners were favorably impressed with the voice quality but expressed concerns about cryptographic aspects of the system.

Conclusions

The Analog Voice Privacy System is based on our previous work in time-frequency segment permutation. In this new implementation, we have continued to refine and improve AVPS. For example, reducing the segment size from 8 ms to 1 ms for a single sample should improve the algorithm's cryptanalytic strength. It certainly increases the number of keys, from about 10^{32} to 10^{203}. However, the

Table III. Guessing Rate (Percent) for Each Digit

Scrambler Characteristics	Digits									
	0	1	2	3	4	5	6	7	8	9
16-ms segment; 256-ms delay	12	11	7	8	9	14	13	10	8	8
AVPS 1-ms segment; 50-ms delay	7	5	13	13	6	9	20	16	7	5

real novelty of this implementation is that AVPS is ready to be interfaced directly to the telephone and the telephone network. It can be used with any modular telephone, with both conventional carbon-button and electret handsets. Its 30-dB range of transmission levels compensates for the network switch to which the telephone is connected. We have tested AVPS extensively, both in the laboratory and in the field. Now, it is ready for others to use and decide how they like it.

Some may feel that AVPS is not secure enough, but our tests for residual intelligibility showed that the new system is quite secure. Currently, the system lacks a key-distribution mechanism. The digital hardware needed to compute any possible key would make it possible to change keys on the fly. Thus, the additional hardware needed for a key-distribution and key-generation system could also be used to enhance the cryptological strength of AVPS.

The system's greatest sensitivity is to narrow channels that would eliminate one or both of the tones. An alternative design could move both tones closer to the center of the channel. If this digital filterbank were based on a prototype with a narrower transition band, the tones could be moved inward.

In summary, we can say that the current half-duplex prototypes demonstrate that the basic system works robustly in the real world. If AVPS is to be developed further, the two issues that require the most study are:

- Make the system full duplex.
- Design a key distribution and generation system that will make AVPS stronger cryptologically.

References

1. J. J. Werner, "An Echo-Cancellation-Based 4800 Bit/s Full Duplex DDD Modem," *IEEE Journal of Selected Areas in Communications*, Vol. SAC-2, No. 5, September 1984, pp. 722-730.
2. M. R. Schroeder and B. S. Atal, "Code-Excited Linear Prediction (CELP): High Quality Speech at Very Low Bit Rates," *Proceedings of ICASSP '83*, IEEE International Conference on Acoustics, Speech, and Signal Processing, March 1985, pp. 937-940.
3. N. S. Jayant et al, "Analog Scramblers for Speech Based on Sequential Permutations in Time and Frequency," *The Bell System Technical Journal*, Vol. 62, No. 1, January 1983, pp. 25-46.
4. R. V. Cox and J. M. Tribolet, "Analog Voice Privacy Systems Using TFSP Scrambling: Full Duplex and Half Duplex," *The Bell System Technical Journal*, Vol. 62, No. 1, January 1983, pp. 47-61.
5. J. H. Rothweiler, "Polyphase quadrature filters—a new sub-band coding technique," *Proceedings of ICASSP '83*, IEEE International Conference on Acoustics, Speech, and Signal Processing, April 1983, pp. 1280-1283.
6. R. V. Cox, "Quadrature Mirror Filters for Speech and Audio Processing," *Proceedings of ISCAS 86*, International Symposium on Circuits and Systems, May 1986, pp. 285-288.
7. N. S. Jayant, "Effective Number of Keys in a Voice Privacy System Based on Permutation Scrambling," *AT&T Technical Journal*, Vol. 66, No. 1, January/February 1987, pp. 132-136.

Biographies (continued)

He joined the company in 1979 and has a B.S. from Rutgers University and an M.A. and Ph.D. from Princeton University, all in electrical engineering. Mr. Johnston works on high-quality audio, analog hardware, and audio transmission. He joined the company in 1976 and has a B.S. and an M.S. in electrical engineering from Carnegie Mellon University. Mr. Snyder does research in speech applications of digital signal processing hardware. He joined the company in 1979, and has a B.S. in mathematics and physics from Rice University and a Ph.D. in high-energy physics from Yale University.

(Manuscript received May 2, 1986)