A New Speech Signal Scrambling Method for Mobile Radio Applications

Enrico Del Re, Romano Fantacci
Dipartimento di Ingegneria Elettronica - Università di Firenze
Via di Santa Marta, 3 - 50139 Firenze - Italy

Graziano Bresci, Damiano Maffucci
OTE ISC - Via Barsanti, 8 - 50127 Firenze - Italy

Abstract. Secure communications systems have acquired a particular importance today, especially as concerns mobile radiocommunications. At present, however, highly secure signal scrambling systems are limited in that they require frame synchronization; this not only limits the system's reliability but complicates its use. This paper presents a two-dimensional signal scrambling method based on digital signal processing techniques which eliminate the need for frame synchronization, but maintain an extremely high level of communications security. Such techniques include short-time Fourier analysis and the filter bank concept. The paper also discusses the use of special digital FIR filters which make it possible to completely implement the system algorithm via commercial processor software. As a result, the system can be configured with very little hardware. Finally, the paper presents an investigation of valid keys and available key space.

1. INTRODUCTION

The need for speech communications protection for a wide variety of applications is constantly growing. As a result, speech scramblers are finding more and more applications in today's communications systems. In the field of mobile radiocommunication where preventing unauthorized users from eavesdropping is almost impossible, the use of speech scramblers is virtually indispensable.

2. SPEECH SIGNAL CIPHERING

There are two speech ciphering approaches commonly used today:
- digital encryption
- analog scrambling [1].

In the first approach (digital encryption) the signal is digitized and, when possible, compressed to reduce bit rate. The cipher operates over the bit sequence to produce a different bit series using an encryption method based either on block ciphering (such as the DES), or on stream ciphering (such as a multiregister non-linear combination). The modified sequence is then transmitted via appropriate digital modulation.

Although this method is highly secure, it requires a transmission technique different from that generally used in present radio and telephone systems. Moreover, it generates a signal with a bandwidth much greater than the original, unless a speech compression technique is used. This, however, requires extensive processing and alters voice quality [2].

In the second approach (analog scrambling), traditional methods operated directly on the analog signal with no digitazion, and processing was carried out through analog circuits. But recent advances in microprocessors, LSI and VLSI technologies as well as in digital signal processing methods have given way to a new era in analog scrambling. Modern scramblers use analog transmission (thus are still referred to as "analog" scramblers), but the signal itself is entirely digital-processed. Digital techniques allow the implementation of very complex and secure algorithms not possible with conventional analog methods. Furthermore, digital techniques are unaffected by environmental factors, can be repeated perfectly and require no trimming.

With modern analog scrambling methods the voice signal is first digitized, then digitally processed with a certain algorithm and, finally, digital-to-analog converted and transmitted. The receiver redigitizes the signal, inversely processes it, converts it to analog from and, thus, reconstructs the original voice signal.

3. PARAMETERS OF SPEECH ENCRYPTION SYSTEMS

The main parameters of a speech encryption system are:

- residual intelligibility
- key space
- degree of security

Residual intelligibility is the percentage of intelligibility left in the encrypted speech signal; key space is the number of keys that can be used in the encryption; degree of security is the degree of difficulty with which an eavesdropper can decipher the signal. Low residual intelligibility and a large key space are necessary - yet not the only - conditions for a high degree of security. For instance, it is possible that some of the keys in the large key space produce a non-intelligible signal, which, however, may still be readily deciphered via certain cryptoanalytic attacks. Thus, bandwidth expansion, delay time, resistance to channel impairments (noise, distortion, group delay), and recovered speech quality are also important parameters.

4. A NEW SPEECH SCRAMBLING METHOD

Cipher and decipher synchronization poses a particular problem in speech scrambling systems. Until now, all methods permitting a high degree of communications security required frame synchronization. This not only complicates implementation of the scrambling method, but means that speech quality is dependent upon channel condition, and may even impair system performance. This is especially true in the area of mobile radiocommunications.

The most significant feature of the speech signal scrambling method presented in this paper is the fact that no synchronization is required - without jeopardizing security. Moreover, signal bandwidth is not increased and only a slight delay time (of approx. 100 ms) is introduced.

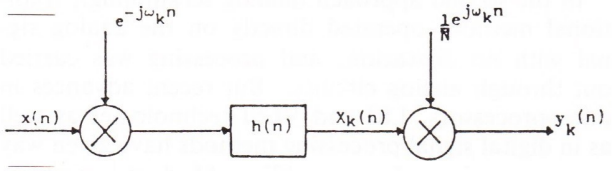

Fig. 1 - The k-th filter-bank channel.

4.1 *The Scrambling Algorithm*

Recent advances in digital signal processors have permitted cost-effective implementation of scrambling algorithms based on the one presented in [3]. This method, which operates in the frequency domain, consists essentially of dividing the spectrum into several bands and subsequently permuting their positions. Although the cipher method itself is not new, the number of bands made available is significantly greater than was possible with conventional methods (with 5 bands, for example). To increase the level of communications protection, a time domain ciphering operation is performed, making the complete system a two-dimensional scrambler. For computational efficiency, spectral analysis using the FFT algorithm must be performed. But the direct use of FFT requires frame synchronization [4]. This obstacle can, nevertheless, be overcome by using the FFT to compute a bank of N filters as illustrated in [5] for the implementation of a digital vocoder. Fig. 1 shows the bank's k-th channel: the block $h(n)$ is a prototype low-pass filter constituting, along with the amplifier, the analysis section, while the other multiplier constitutes the synthesis section.

To perfectly reconstruct signal $x(n)$, the following relationship between input and output signals must exist:

$$x(n) = y(n) = \sum_{k=0}^{N-1} y_k(n) = \frac{1}{N}\sum_{k=0}^{N-1} X_k(n)e^{jw_k n} =$$

$$= \frac{1}{N}\sum_{k=0}^{N-1}\sum_{m=-\infty}^{+\infty} x(m)h(n-m)e^{-jw_k m}e^{jw_k n} \qquad (1)$$

where $w_k = 2\pi/N \cdot k$.

Although an ideal low-pass filter with a cutoff frequency of $w_c = \pi/N$ is not necessary, it is necessary and sufficient that the first Nyquist criterion be satisfied:

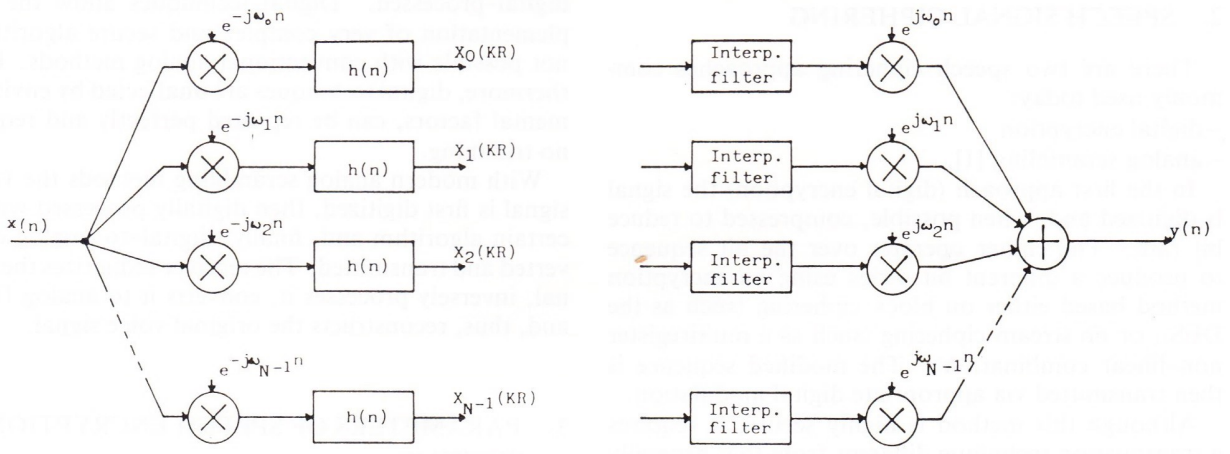

Fig. 2 - The filter-bank for analysis and synthesis procedures.

$$h(0) = 1$$
$$h(n) = 0 \text{ for } n = \pm N, \pm 2N, \pm 3N, \ldots \quad (2)$$

for non-causal filters.

Signals $X_k(n)$, $k = 1, 2, \ldots, N$, are approximately band-limited for the frequency range $-\pi/N < w < \pi/N$. It thus follows from the sampling theorem that $X_k(n)$ can be computed only once every R, where $R \leq N$. Signal $y(n)$ must then be constructed by interpolation.

Fig. 2 shows the entire system of speech signal spectral analysis and synthesis.

The two sampling frequencies used are:

- frequency f_s for input and output signals
- frequency $f_d = f_s/R$ for spectral components.

If the number of frequency bands N is a power of 2, it can be shown that both the analysis and the synthesis parts can be efficiently computed by using the FFT algorithm [5]. The procedure is as follows:

let the low-pass filter be FIR with a length of $M = 2LN + 1$ (L being an integer). The most recent M samples of signal $x(n)$ are multiplied by the window $h(-n)$. The resulting weighted sequence is partitioned in sections of length N each. These sections are then added together - each sample with its counterpart in the other $(2L - 1)$ sections. An N point sequence function of the time index n, denoted as $\bar{x}_m(n)$, $m = 0, 1, \ldots, N - 1$, is thus obtained. The sequence $\bar{x}_m(n)$ is circularly shifted (in m) modulo N by n samples to obtain the new sequence:

$$x_m(n) = \bar{x}_{[m-n]}(n) \quad (3)$$

where $[p]$ is p modulo N.

By computing the N-point FFT of the sequence $x_m(n)$, $m = 0, 1, \ldots, N - 1$, we obtain the N complex values X_p, $p = 0, 1, \ldots, N - 1$ (Fig. 2) which represent the frequency components of the speech signal. As cited above, this procedure need be repeated only once every R, where $R \leq N$.

The synthesis procedure which reconstructs the speech signal from its frequency components $X_p(kR)$ can be denoted as follows:

a computation is made of the inverse N-point FFT of the complex vector $X_p(kR)$, $p = 0, 1, \ldots, N - 1$, sampled at the rate of f_s/R. From this the vector $s_m(kR)$, $m = 0, 1, \ldots, N - 1$ is obtained. Let the interpolating filter be FIR of length $2QR + 1$ with integer Q, and let $f(n)$, $n = 0, 1, \ldots, 2QR$ be its impulse response. R output samples are then obtained from the $2Q$ most recent vectors $s_m(kR)$ using the formula:

$$y(n) = \sum_{k=L_1(n)}^{L_2(n)} s_{[n]}(kR) f(n - kR) \quad (4)$$

where

$$L_1(n) = \left\lceil \frac{n}{R} \right\rceil - Q + 1$$
$$L_2(n) = \left\lceil \frac{n}{R} \right\rceil + Q$$

and where $\lceil k \rceil$ is the largest integer contained in k.

The scrambling operation consists of permuting the frequency components X_p, $p = 0, 1, \ldots, N - 1$, before performing the synthesis procedure. The cipher thus multiplies the vector X_p by the scrambing matrix M, while the decipher performs the same operations using the descrambling matrix M^{-1}. The entire system requires no synchronization. Due to band permutations, a synchronization error (a difference in timing in the decimation process) introduces a phase error in the descrambled signal only. This, however, is of little significance as the human ear is unable to perceive such errors. A rigorous proof of this statement is given in [3].

The scrambling system also performs a time domain ciphering operation to implement a two-dimensional scrambler by adding a masking signal to the output signal. To avoid the need for synchronization, the masking signal is a variably delayed and weighted version of the transmitted signal itself. To save memory this operation should be performed over the components X_p because these signals are sampled at a lower rate. In turn, the descrambler must subtract the delayed and weighted received component from the component actually received. Figure 3 shows the structures used in the scrambler (a) and in the descrambler (b) for the processing of each component X_p.

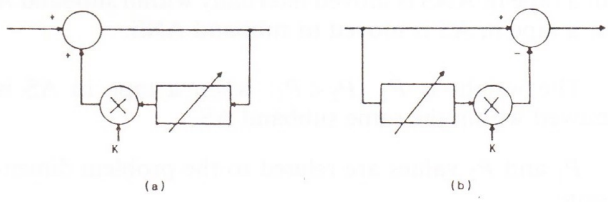

Fig. 3 - The structures used to perform the time domain ciphering: (a) cipher and (b) decipher.

The box represents a variable delay, while the multiplication by factor K represents the weighting factor. This additional cipher procedure further reduces the residual intelligibility of the scrambled signal.

4.2 *System Implementation*

Due to the decimation process, signal reconstruction in the synthesis section is necessarily imperfect; in fact, neither the low-pass filter $h(n)$ nor the interpolating filter $f(n)$ are perfect. Furthermore, even if a Nyquist low-pass filter were used and there were no decimation process, signal reconstruction at the receiving end could not be perfect due to permutation. Thus, sharper filters such as FIR type (designed using the equiripple approximation method [6]), may be employed instead of theoretical filters (i.e., the windowed sinc

function); the maximum decimation factor $R = N$ can then be chosen without sacrificing signal quality significantly. Moreover, if $R = N$, another ciphering operation can be more easily introduced: A frequency component $X_p(kN)$ can be alternatively multiplied by ± 1 to cause the inversion of that band. Finally, in order to avoid band expansion, permutation has been restricted to a number P of central bands, while the extreme bands have been set at zero.

This system was implemented using software with $N = 64$ and $P = 24$ (48 center frequencies in the interval $0 - f_s$). The sampling frequency used was $f_s = 8$ kHz and the 24 bands adequately covered the mobile radio baseband range.

5. USEFUL PERMUTATION KEY SPACE

This study was conducted using a speech bandwidth of 250 - 2600 Hz. The bandwidth is divided into 24 subbands of 95 Hz each called "tape" which, in turn, can be regrouped into two smaller subbands: subband A covering the first 12 tapes, subband B the remaining 12. The low-frequency subband A is composed of 6 significant tapes (240 Hz - 850 Hz) referred to as AS, and of 6 insignificant tapes called ANS.

Frequency scrambler efficiency for listening intelligibility was measured by a "weighted displacement" (wdp) parameter. This parameter measures the shift in each weighted tape. The weight associated to each tape depends on its relative position and can assume two values: P_1 and P_2.

The weight is P_1 when:

a) a tape in B is moved to B
b) a tape in ANS is moved internally within subband A
c) a tape in AS is moved to subband ANS.

The weight is P_2 ($P_2 < P_1$) when a tape in AS is moved within the same subband AS.

P_1 and P_2 values are related to the problem dimension:

$$P_1 = (1/a) \cdot NNS$$
$$P_2 = P_1/4 \tag{5}$$

where NNS is the number of insignificant tapes (18 in this case) and "a" is the basic weight (150, for example). As a result:

$$P_1 = 0.12$$
$$P_2 = 0.03 \tag{6}$$

The weighted displacement is thus the sum of the shift of each tape (divided by the total number of tapes (24)), multiplied by its relative weight. The greatest value is obtained in this case of spectral inversion.

A high weighted displacement value means that a large number of tapes have moved from one subband to another. Although this improves signal intelligibility, it lessens permutation key robustness. Permutation key robustness is indicated by the "weighted mobile average" (wma) denoted as follows:

Let the sequence of tapes $AA, BB, CC, DD, EE, \ldots$ be scrambled. After selecting the element CC we can evaluate the average distance of the original adjacent elements BB and DD:

$$(|CC - BB| + |CC - DD|)/2 \tag{7}$$

The "weighed mobile average" is obtained by the sum of these values (called "local means") for all the tapes and then dividing this value by the total number of tapes (24). In this case, each addendum must also be weighed in order to consider the most significant tape in the speech spectrum, as follows:

a) if in a local mean only one significant tape exists, the weight is unitary;
b) if in a local mean more than one significant tape exists, the weight decreases as the number of significant tapes increases;
c) if in a local mean more than one significant tape exists, the weight is even less if significant tapes are adjacent.

Since:

W = weight
KW = constant
N = the number of significant tapes present in a generic local mean
NA = the number of adjacent significant tapes in a generic local mean,

then weight W is:

$$\begin{aligned} &W = 1 \\ &\quad \text{if } N \leq 1 \\ &W = PW = KW/2^{N-1} \\ &\quad \text{if } N \geq 2 \text{ and } NA = 0 \\ &W = (PW + PW/(1 + NA))/(1 + NA) \\ &\quad \text{if } N \geq 2 \text{ and } NA \geq 1 \end{aligned} \tag{8}$$

A good permutation key must have high "weighed displacement" and "weighed mobile average" values. However these parameters increase in contrasting mode; an applicative example is shown in Fig. 4. An excessive increase in weighed displacement produces a heavy concentration of those tapes which were originally close together. This results in a decreased weighed mobile average and, consequently, in reduced key robustness.

The robustness parameter must be high enough to avoid that simple manipulations of the crypted audio signal permit deciphering the signal. Numerous tests were conducted to prevent this from occuring. A few seconds of a message were recorded on the mass storage memory of a computer and then scrambled. Participants not knowing the original voice and asked to listen to the scrambled voice were unable to identify the voice.

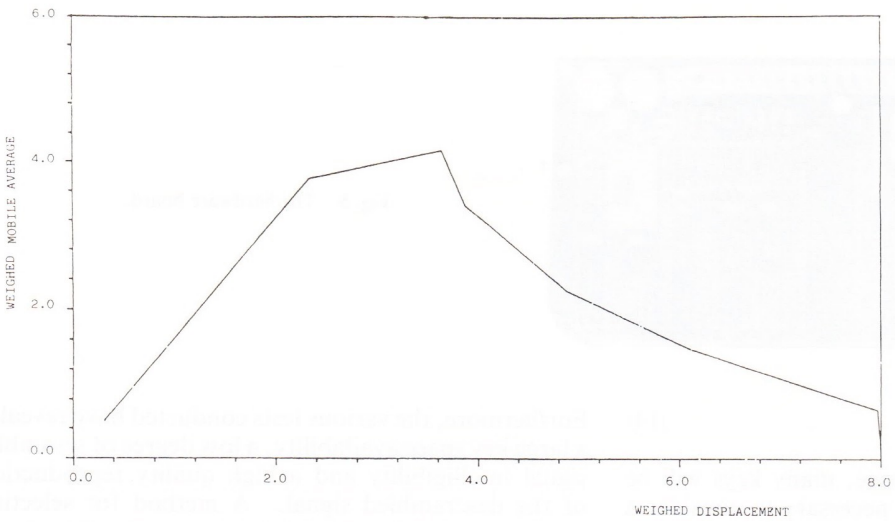

Fig. 4 Behaviour of weighed mobile average vs. weighed displacement.

The test revealed that if a scrambling method can effectively produce a non-intelligible voice, it can do the same, if not better, with numbers due to their intercorrelation.

In conducting these tests threshold values were assigned to the two parameters to provide good permutation keys; however, this puts some constraints on the scrambling type:

1) no significant tape must remain in its original position;
2) there can be no more than two significant tapes in the AS subband, though these need not be in adjacent positions originally;
3) at least 8 tapes (4 from subband A, and 4 from subband B) must change subband;
4) at least 2 of the 4 tapes in the low frequency subband A must be significant;
5) in the entire group of 4 tapes there must not be any more than 3 significant elements;
6) the robustness parameter requires that no more than 8 tapes change from subband A to B and vice versa.

6. APPROXIMATE KEY SPACE EVALUATION

As cited above, of the 8 tapes at least which must change subband, two of these must be significant elements, moreover no more than 16 elements must change subband. The number of all possible permutations on 24 elements is 24!. By dividing these into two subbands of 12 elements each and making permutations on each subband the number is:

$$12! \cdot 12!$$

Since some elements can change subband the number of possible permutations (K) is:

$$12! \cdot 12! < K < 24! \tag{9}$$

And since from 4 to 8 elements per subband must change, the result is:

$$K = (12! \cdot 12!) \cdot \left(\sum_{z=4}^{8} \binom{12}{z} \cdot \binom{12}{z} \right) \tag{10}$$

This equation can be expressed in general terms as:

$$K = NA! \cdot NB! \cdot \left(\sum_{z=vi}^{vs} \binom{NA}{z} \cdot \binom{NB}{z} \right) \tag{11}$$

where NA and NB are the number of elements in subband A and B respectively and "vi" and "vs" are the respective lower and upper bonds of the number of elements (per subband) that change subband. By introducing all of the limitations noted in the paragraph above, it can be shown that a minimum value exists for the number of possible keys:

$$K_{\min} = m \cdot N_c \cdot 6! \cdot 6! \cdot \cdot \left(\sum_{z=4}^{8} \binom{12}{z} \cdot \left(\sum_{j=0}^{z-2+h} \binom{6}{z-j-h} \cdot \binom{6}{j+h} \right) \right) \tag{12}$$

where:

$h = 0$ if $z \leq 6$

$h = z - 6$ if $z > 6$

and:

- m is approx. $6.8 \cdot 10^7$
- $N_c \geq 38$

This relation was obtained by making pessimistic assumptions about the tape distribution, thus resulting in a lower than real limitation for available key space. The value for $K_{\min}$ is:

$$K_{\min} \approx 1.3 \cdot 10^{21} \tag{13}$$

A further constraint can be made with reference to point (2) of the previous paragraph which states that no more than two significant tapes can remain in the AS subband. Considering that the two tapes remaining in the subband need not be adjacent, $K_{\min}$ becomes:

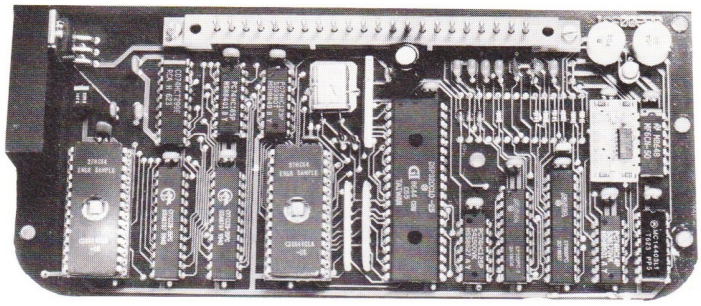

Fig. 5 - The hardware board.

$$K_{min} \approx 2.3 \cdot 10^{20} \qquad (14)$$

Within the available key space, many keys will be close together. It is therefore necessary to consider a narrower space only between those useful keys which are different enough from one another.

7. SOME REMARKS ABOUT HARDWARE

The foregoing tests were conducted using a commercial signal processor. A board (Fig. 5) was constructed based on a pin-to-pin compatible DSP and clocked with a 24 MHz xtal, with which the frequency scrambler was then implemented. This board performs both scrambling and descrambling functions on a simplex basis; the operating mode is activated by the PTT (push to talk) command.

In this study, extensive use was made of programmable logic components (i.e., the Eprom Programmable Logic Device). The A/D and D/A converters were obtained with a CODEC, which also includes anti-aliasing filters.

Since sampling is carried out at a standard frequency of 8 kHz, the audio signal bandwidth was limited to between 250 - 3400 Hz (approximate) using high-slope switched capacitor filters to avoid aliasing.

8. CONCLUSIONS

The use of advanced digital processing techniques and a commercial signal processor have made it possible to implement a highly reliable secure communications system based on a two-dimensional scrambling algorithm requiring no frame synchronization and no expansion of the signal bandwidth.

Furthermore, the various tests conducted have revealed a large key space availability, a low degree of scrambled signal intelligibility and a high quality reproduction of the descrambled signal. A method for selecting significant keys has also been presented. Finally, a hardware board suitable for assembly in OTE-produced mobile radio sets has been constructed.

This new speech signal scrambling method has been used in one of OTE's recent productions - a mobile radio set operating in the 400 MHz band - effectively providing some of the highest levels of communications security available in mobile radiocommunications today.

Manuscript received on February 11, 1988.

REFERENCES

[1] H.J. Beker, F.C. Piper: *Secure Speech Communications.* Academic Press, London, 1985.

[2] V. Cappellini (editor): *Data Compression and Error Control Techniques with Applications.* Academic Press, London, 1985.

[3] L.S. Lee, G.C. Chou, C.S. Chang: *A New Frequency Domain Speech Scrambling System Which Does Not Require Frame Synchronization.* "IEEE Trans. Communications", April 1984, vol. COM-32, p. 444-456.

[4] K. Sakurai, K. Koga, T. Muratani: *A Speech Scrambler Using the Fast Fourier Transform Technique.* "IEEE Journal on Selected Areas in Communications", May 1984, vol. SAC-2, n. 3.

[5] M.R. Portnoff: *Implementation of the Digital Phase Vocoder Using the Fast Fourier Transform.* "IEEE Trans. Acoustic, Speech, and Signal Processing", June 1976, vol. ASSP-24, p. 243-248.

[6] A.V. Oppenheim, R.W. Schafer: *Digital Signal Processing.* Prentice-Hall, New Jersey, 1975.