PERIODICA POLYTECHNICA SER. EL. ENG. 33/3, pp. 135-140 (1989)

A COMPACT VOICE PRIVACY SYSTEM

F. MARX

Institute of Communication Electronics*
Technical University of Budapest

Received June 6, 1988

Abstract

An analogous voice privacy system was developed using a single TMS32010 DSP processor. The system is used for encrypting of human voice and transmitting it via a standard analogous telephone channel without frame synchronization. The full duplex system was implemented on a single EU printed circuit board.

Keywords: voice privacy system, speech transmission, voice encryption.

Introduction

Voice encryption has already a four-decade history (see DIFFIE and HELLMANN (1979)). The first approaches were based on analogous methods, of course. A typical system splits the input speech into about 5 bands and then heterodyned them into other bands according to the permutation key. The key space was very small: $5! = 120$. Rolling code scramblers changed the key several times a second and inverted some of the bands according to a pseudorandom sequence generated by using the key as a seed. These systems were still breakable, because the speech spectrum does not change abruptly and a band could be tracked during changing the permutation.

Time division scramblers make the scrambling procedure in time domain and their key space can be very large but have an impractically large transmission delay at the same time.

The methods used nowadays through analogous telephone lines fall into two categories. The first uses a low bit rate speech quantizer followed by a sophisticated data encrypter and a modem. But the speech quality achieved at these rates (1200–2400 bps) is not good enough. It still has some losses in intelligibility and naturality.

The second category we used is called sample based scrambling. The

*On leave from Research Institute for Telecommunication.

digitized speech vector is transformed into a certain transform domain, permuted, inversely transformed and transmitted in analogous form.

Sample Based Scrambling

The transformation used for speech encryption is expected

- to be orthogonal (i.e. not to enhance the noise on decrypting)
- to leave bandlimited speech bandlimited
- to be computationally efficient
- to eliminate the need of any synchronization between the transmitter and the receiver.

WYNER (1979) proposed a new concept of using discrete prolate spheroidal sequences as a basis of the transformation. Using them the degree of security improved tremendously due to the much lower residual intelligibility and much larger key space size and the fact that the transform domain is not in connection with any physical feature of the speech. Unfortunately, the problem of frame synchronization remains untouched. Another price paid for these good properties is that the factorization of the transformation matrix has not been solved yet, so this system has not been implemented in real-time due to its relatively high computational complexity.

To overcome the above limitations a new frequency domain scrambling system was investigated by LEE and CHOU (1986) in which the short-time Fourier analysis-synthesis is used.

Short-Time Fourier Analysis-Synthesis

The analysis by definition in discrete time and discrete frequency is

$$X_n(k) = \sum_{m=-\infty}^{\infty} h(n-m)\cdot x(m)\cdot \exp(-j\cdot 2\pi\cdot k\cdot m/N) \quad 0 \le k \le N-1 \quad (1)$$

where $x(n)$ is the input sequence, $h(n)$ is the impulse response of a low-pass filter, n is the time index and k is the frequency index. It can be interpreted (see PORTNOFF (1980)) at fixed k as a down-converted output sequence from a band-pass filter with centre frequency $2\pi k/N$.

Supposing that $h(n)$ is finite in time and observing the fact that $\exp(-j2\pi km/N)$ is periodic in m with period N, one can simplify the expression Eq. (1) by defining $u_n(q)$:

$$u_n(q) = \sum_{r=-L}^{L-1} x(n+N\cdot r+q)\cdot h(-N\cdot r-q) \quad 0\le q\le N-1 \quad (2)$$

$$X_n(k) = \sum_{q=0}^{N-1} u_n(q)\cdot \exp(-j\cdot 2\pi\cdot k\cdot q/N) \qquad 0\le k\le N-1 \quad (3)$$

where Eq. (3) is exactly of the form of an N-point DFT.

The two expressions describe an N band filter bank which can efficiently be calculated by means of an FFT algorithm. Since all the $X_n(k)$ sequences have the bandwidth π/N, they can adequately be described by sample values at the period N. The synthesis part is

$$x(n) = \frac{1}{N}\sum_{k=0}^{N-1}\sum_{s=-\infty}^{\infty} f(n-s\cdot N)\cdot X_k(s\cdot N)\cdot \exp(j\cdot 2\pi\cdot n\cdot k/N) \quad (4)$$

where $f(n)$ is the impulse response of an interpolating filter.

Using the same idea (see PORTNOFF (1980)) as in Eq. (1), Eq. (4) can be replaced by

$$y_{r\cdot N}(i) = \frac{1}{N}\sum_{k=0}^{N-1} X_{s\cdot N}(k)\cdot \exp(j\cdot 2\pi\cdot k\cdot i/N) \quad (5)$$

$$v_n(i) = \begin{cases} y_{s\cdot N}(i) & , n = s\cdot N \quad , s = 1, 2, \ldots \\ 0 & , n \neq s\cdot N \end{cases} \quad (6)$$

$$x(n) = \sum_{m=n-L\cdot N}^{n+L\cdot N-1} v_m(n)\cdot f(n-m) \quad (7)$$

In Eq. (5), the form of an inverse DFT appears.

Now *Fig. 1* shows the new scrambling algorithm. Block P_1 performs expression Eq. (2), P_2 performs Eq. (6) and Eq. (7). M is a scrambling matrix which permutes $X_n(k)$ into an encrypted version. The descrambling process in the receiver is exactly the same except M is replaced by its inverse M^{-1}.

The method uses Fourier basis which is complex. Therefore, the scrambled speech will be complex in general. To keep it real, matrix M must save the complex conjugate symmetry in $X_n(k)$ LEE and CHOU (1986).

There is another property M must dispose of. Existing analogous telephone channels have the bandwidth 0.3 – 3.4 kHz, while $X_n(k)$ has

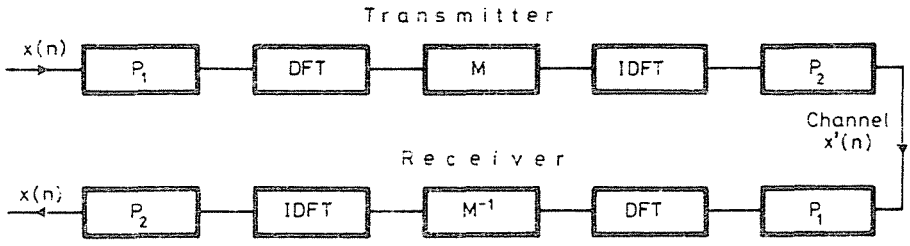

Fig. 1. Block diagram of the scrambling system

information about the band 0−4 kHz at the usual sampling frequency 8 kHz. So M must not change the samples of $X_n(k)$ corresponding to the bands $0 - 0.3$ kHz and $3.4 - 4$ kHz. It means that the corresponding submatrices must be identity matrices.

Why synchronization is not required? Assume the receiver is out of frame synchronization by r samples. It means that while the transmitter calculates the $X_k(s \cdot N)$ samples, the receiver has the $X_k(s \cdot N + r)$ ones. However, the sequence can be described by its samples at period N, no matter whether these samples are given at the moments $s \cdot N$ or $s \cdot N + r$. The synchronization error introduces only a linear phase error which becomes a random phase error during the operation with M^{-1}. Fortunately, human ears are not sensitive to phase distortion at all.

Hardware and Software Considerations

A single TMS32010 digital signal processor is used to implement the encryption system. The hardware consists of the processor, of its 4 kword program ROM, of an additional 2 kword data RAM, of the A/D D/A converters and auxiliary elements. The frame length is chosen to be 128 and the filter length is 6×128. The 6×128 samples of the input sequence are stored in the outer RAM and so are the output samples. An interrupt routine is used to handle the input-output operations.

The most important part of the program is how to calculate the DFT. A time efficient radix-4 FFT algorithm is used for this purpose (Kocsis and Marx (1987)) as the most suitable one on this processor. Because of straight-line coding the same routine must perform the inverse transformation. Fortunately it can be fulfilled by taking into account the reversed indexing of output points on inverse operation.

To increase the degree of security the permutation changes in every 256 msec. The permutation matrices are calculated from the output of a pseudorandom generator stepped in every 256 msec. In the first version of the program a simple shift register with feed-back was used as a random generator. But it is known that it can be broken by knowing its twice as long output sequence as the register length itself. The DES (DATA ENCRYPTION STANDARD) algorithm is implemented in feed-back mode instead. There is no enough memory for the DES in the basic ROM so it must be extended using paging technique.

Because of using variable permutation, the synchronization of the pseudorandom generators in the transmitter and the receiver is needed. For this, the transmitter begins its work with transmitting a synchronization tone. According to the experiences this single operation is enough to keep the receiver synchronous during a telephone call. Storage requirement and execution time can be seen on *Table 1.*

Table 1 Time and storage requirements

Routines	Prog. mem. [word]	Exec. time [ms]
I/O op.	23	0.58
P_1 and P_2	110	2×2.8
DFT, IDFT	2730	2×0.83
Permut.	340	0.21
DES	1460	1.96
Total	5083, inc. tables	10.01

Experiences, Conclusions

The residual intelligibility test (LEE and CHOU (1986)) has resulted in 13% average digit intelligibility using Hungarian language (the theoretically best one is 10%). The standard deviation of the test was about 4%. The speech quality is quite good but there is a little crack at permutation change.

Some authors consider DES breakable with an exhaustive search in an acceptable time. But in data encryption the situation is quite different. The plain text and encrypted text are well defined so they can be a source for a search if they are both known or supposed. Here, in speech encryption there is only a little chance to know both the source speech and the corresponding encrypted speech exactly (!) at the same time. And how can a computer de-

cide whether it has found exactly the right source speech sequence, because it is not the problem of speech recognition but of a much more difficult one. The key is only given in the case of exact decryption, so exhaustive search by a computer is more difficult then in data encryption. Because of these reasons we consider DES a good choice in speech encryption.

Further experiments are needed to check the system through real telephone lines.

References

DIFFIE, W. – HELLMAN, M.E. (1979): Privacy and Authentication: An Introduction to Cryptography. *Proc. IEEE.* Vol. *67.*, No. 3., pp. 397–427, Mar. 1979.

KOCSIS, F. – MARX, F. (1987): Periodogram-type Spectrum Analysis Program for Advanced Digital Signal Processing μP. *Conf. on Microcomputers, Microarchitectures and Developments.* Brno, Sept. 1987.

LEE, L.S. – CHOU G.C. (1986): A General Theory for Asynchronous Speech Encryption Techniques. *IEEE Journal on SAC.* Vol. *4.*, No. 2., pp. 280–286, Mar. 1986.

PORTNOFF, M.R. (1980): Time-Frequency representation of Digital Signals and Systems Based on Short-time Fourier Analysis. *IEEE Trans. on ASSP.* Vol. *28.*, No. 1., pp. 55–69, Feb. 1980.

TMS 32010 User's Guide. Texas Instruments Incorporated. 1983.

WYNER, A.D. (1979): An Analog Scrambling Scheme which Does Not Expand Bandwidth. *IEEE Trans. on Inf. Theory.* Vol. *25.*, No. 3–4., pp. 261–275, 415–425, May-July 1979.

Address:

Ferenc MARX
Budapest, Mező u. 63.
Hungary H-1089