ANALOG SPEECH SCRAMBLING VIA THE GABOR REPRESENTATION

S. Farkash, S. Raz and D. Malah

Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel

ABSTRACT

This paper presents a Gabor-domain scrambling scheme which enables a variety of scrambling options - in time, in frequency, and in the combined time-frequency spaces. The scrambling operation is performed by Modifying on the Gabor representation of the input speech signal. The proposed scheme can perform any invertible modification, and not necessarily the commonly used permutation operation, rendering the scheme more secure. It is shown that when using the oversampled Gabor expansion there are modifications which are not legal, and the class of all legal modifications is determined. The synchronization problems, usually accompanying the scrambling process, is studied. It is shown that certain channel delays are overcome when using the proposed scheme. The effect of the Gabor expansion parameters on the synchronization problem and on the security level of the scheme is examined. The performance of the proposed scheme is demonstrated by simulations.

1. INTRODUCTION.

In spite of significant progress in digital speech technology, analog speech scramblers continue to be important for achieving privacy in many types of voice communication [1-3], due to the desire for secure communication over existing channels. Most scrambling algorithms are based on rearranging or reordering speech samples, rendering the speech unintelligible. Three approaches for speech scrambling are usually considered. These are, time scrambling, frequency scrambling, and a combined time-frequency scrambling. Gabor representation of the speech signal is ideally suited for scrambling speech samples in time, in frequency or in time-frequency (Gabor) domains.

Gabor representation of a single variable function (e.g., a speech signal) constitutes a mapping onto a two-dimensional discrete space, describable by a two-dimensional array. The columns of the array correspond to M-size transforms (usually DFT) of the signal multiplied by a set of windows, shifted R samples apart.

In this paper we propose a Gabor-domain based analog speech scrambling technique, which uses the approach presented in [4] for performing linear time-varying operation. The proposed scrambling scheme creates a variety of scrambling options, it can perform any invertible modification, and not necessarily the commonly used permutation operation, rendering the proposed scrambling scheme more secure. The speech signal is transformed into Gabor space where the **Gabor coefficients** are modified by a linear system and inverse-transformed into the time domain, yielding the scrambled speech. At the receiver end the scrambled speech is transformed back into Gabor space, inverse modified, and inverse Gabor transformed, resulting in the original speech signal in time (assuming a distortion-less channel). When applying the modification in either time or frequency, the proposed scrambling scheme coincides with the time and frequency scrambling schemes suggested in [2,3]. In these schemes it was argued that the synchronization problem (channel delays) accompanying the scrambling process is resolved, as long as the shift parameter R is less then or equal to the transform size M. In this paper it is shown that only channel delays of $r \cdot R$ samples (where r is an integer) are resolved by the schemes described in [2,3]. Thus, a small R should be chosen to properly reduce the synchronization problem. This contradicts the claim raised in [2,3] that any $R \le M$ will do. Furthermore, it is shown that in the oversampled case ($R < M$), not all modifications are legal, in accordance with the results stated in [5]. (A modified Gabor representation is considered legal if it is identical to the Gabor representation of the scrambled signal). Hence, the number of potential modifications decreases, resulting in a less secure scrambling scheme. The class of all legal modifications for a given set of parameters R , M, and the analysis window associated with the Gabor representation is determined, generalizing the special cases for time and frequency scrambling considered in [5]. The tradeoff in choosing R which controls the synchronization problem on one hand and the security level on the other hand is considered.

2. THE GABOR REPRESENTATION

The Gabor representation comprises a superposition of sliding window functions multiplied by a Fourier kernel. Throughout we shall consider the discrete-time case which was addressed in [6] and is briefly presented here for the sake of completeness. Unless otherwise indicated, summation indices range from $-\infty$ to $+\infty$.

The Gabor expansion of a discrete-time function $x(k)$ is given by

$$x(k)=\sum_{p}\sum_{q=0}^{N-1} x_{p,q}\,\psi_{p,q}(k)=\sum_{p}\sum_{q=0}^{N-1} x_{p,q}\,\psi(k-pR)\exp\left(j\frac{2\pi}{M}qk\right) \quad (1)$$

where $\psi(k)$ is the synthesis window, R describes the discrete-time shift and M is the transform size. The Gabor coefficients $x_{p,q}$ are evaluated using a biorthogonal function (analysis window) $\hat{\psi}(k)$ via

$$x_{p,q}=\sum_{k} x(k)\hat{\psi}^{*}_{p,q}(k)=\sum_{k} x(k)\hat{\psi}^{*}(k-pR)\exp\left(-j\frac{2\pi}{M}qk\right) \quad (2)$$

where * denoted the complex conjugate. Similar expansions are applicable in the frequency domain.

It can be observed that with minor modifications, the discrete-time Gabor representation coincides with the thoroughly investigated Short Time Fourier Transform (STFT) [7]. In this paper, the algebraic approach for the description of the STFT, introduced in [8], is adopted to

describe Gabor representations and operations in Gabor space. With this approach the Gabor representation of an input signal is obtained by multiplying each segment of the signal by a matrix A, which represents the analysis window $\hat{\psi}(k)$, and then multiplying the result by an appropriate DFT matrix W. The matrix representation of the Gabor analysis is given therefore by

$$g_x = W \cdot A \cdot x \triangleq \tilde{A} \cdot x \tag{3}$$

where x is a vector representing a segment of the input signal $x(k)$, and g_x is a vector representing its Gabor expansion, which is part of the Gabor coefficients array $x_{p,q}$. The synthesis is performed by multiplying the Gabor representation of the signal g_x with the inverse (or generalized inverse when needed) of the above matrices in reverse order. The matrix form of the Gabor synthesis is given therefore by

$$x = S \cdot W^{-1} \cdot g_x \triangleq \tilde{S} \cdot g_x \tag{4}$$

Because of lack of space, only the case where the transform size M is equal to the analysis window length N is considered. A generalization to the case where $N > M$ can be conducted via the weighted overlap add (WOLA) method [7], and is reported in [9]. In addition, it is assumed, without loss of generality, that the shift parameter R satisfies $mod(M, R) = 0$. For given values of the parameters M and R, the size of the analysis matrix A is taken to be $[M^2/R] \times [2 \cdot M - R]$, and the corresponding DFT matrix W is a square block-diagonal matrix with $[M/R]$ blocks of dimension $M \times M$ each. In each operation represented by (3), a segment of $[2 \cdot M - R]$ samples of the input speech signal is multiplied by the matrices A and W yielding $[M/R]$ vectors, of length M each, representing the Gabor expansion. The operation in (3) is repeated with a shift of M samples at a time.

3. THE PROPOSED SCRAMBLING SCHEME

The proposed scrambling scheme is based on the approach presented in [4] for the description and operation of linear time-varying systems in Gabor time-frequency space. A block diagram describing the proposed scrambling scheme is given in Fig. 1. The speech signal $x(n)$ is transformed into Gabor space, in which the **Gabor coefficients** are modified by a linear system (the scrambler) and inverse-transformed into the time domain, yielding the scrambled speech $y(n)$. At the receiver end the scrambled signal is transformed back into Gabor space, descrambled, and inverse Gabor transformed, resulting in, the original speech signal in time (assuming a distortion-less channel).

The general form of the scrambling operation is described by the superposition sum

$$y(k) = \sum_l h(k, l)\, x(k-l) = \sum_l h(k, k-l)\, x(l) \tag{5}$$

substitution of (5) into the Gabor representation of the scrambled speech signal $y(n)$ yields

$$y_{m,n} = \sum_p \sum_{q=0}^{N-1} h_{m,n,p,q} x_{p,q} \tag{6}$$

where $y_{m,n}$ are the Gabor coefficients of the scrambled speech with another analysis window $\phi(k)$, and $x_{p,q}$ are the Gabor coefficients of the original speech with analysis window $\psi(l)$. The four-dimensional array $h_{m,n,p,q}$ in (6) represents the Gabor expansion coefficients of the scrambling operation $h(k, l)$ according to

$$h_{m,n,p,q} = \sum_{k,l} h(k, l)\, \psi_{p,q}(k-l) \hat{\phi}^*_{m,n}(k) \tag{7}$$

where $\psi^*(k-l)\hat{\phi}(k)$ is the resulting analysis window.

To facilitate the manipulation of (6), the four dimensional array $h_{m,n,p,q}$ representing the scrambling operation is converted into a two-dimensional matrix H such that the scrambling operation remains the same. Thus, the matrix form of the scrambling operation in (6) is given by

$$g_y = H \cdot g_x \tag{8}$$

The proposed scrambling scheme creates a variety of scrambling options by using any invertible four dimensional array $h_{m,n,p,q}$ (or any invertible matrix H) rather then a permutation array usually used in scrambling schemes [1-3].

4. VECTOR-SPACE APPROACH TO GABOR EXPANSION AND SCRAMBLING.

The Gabor Expansion and the scrambling operation described in the previous two sections, has a very clear description when a vector-space approach is used. The general analysis matrix $\tilde{A}$ in (3) which represents the Gabor expansion is of size $[M^2/R] \times [2 \cdot M - R]$. Thus, it can be considered as a transformation $\boldsymbol{A}$ from a vector space $\mathcal{V}$ of dimension $[2 \cdot M - R]$ to a vector space $\mathcal{W}$ of dimension $[M \times M/R]$. When R is equal to M, the *critically-sampled* case, the dimensions of both vector spaces are identical and equal to M, in which case the generalized synthesis matrix $\tilde{S}$ in (4) is the simple inverse of $\tilde{A}$. In the oversampled case, $R < M$, the dimension of the destination space $\mathcal{W}$ is larger then the dimension of $\mathcal{V}$, and $Range\{\boldsymbol{A}\}$ (the range of the transformation $\boldsymbol{A}$) forms a sub-space of $\mathcal{W}$ which is spanned by the columns of the matrix $\tilde{A}$. The scrambling process, described by the four-dimensional array $h_{m,n,p,q}$, is considered as a transformation $\boldsymbol{H}: \mathcal{W} \rightarrow \mathcal{W}$. The synthesis of the scrambled speech is described as a transformation $\boldsymbol{S}: \mathcal{W} \rightarrow \mathcal{V}$. This vector-space approach is visualized in Fig. 2. It can be observed that in the oversampled case the transformation $\boldsymbol{A}: \mathcal{V} \rightarrow \mathcal{W}$ is one-to-one but not onto $\mathcal{W}$, i.e., there are vectors in the space $\mathcal{W}$ which are not in the range of $\boldsymbol{A}$. For such vectors $w' \in \mathcal{W}' = \mathcal{W} \setminus Range\{\boldsymbol{A}\}$ there is no vector $v \in \mathcal{V}$ such that $\boldsymbol{A}v = w'$, and the scrambling transformations (modifications) which yields these vectors are considered non-legal. This problem was first addressed in [5] for time and frequency scrambling, and the results reported there constitute special cases of the following results. A modification $\boldsymbol{H}$ is considered legal if the modified Gabor representation is identical to the Gabor representation of the scrambled speech, i.e.,

$$Hg_x = Ay \tag{9}$$

Since (9) is valid for every vector $u \in Range\{\boldsymbol{A}\}$ it is valid for a basis of $Range\{\boldsymbol{A}\}$ (the columns of A, for example). Thus, the class of all legal modifications (LM) $\boldsymbol{H}_L: \mathcal{W} \rightarrow \mathcal{W}$ for a given analysis transformation $\boldsymbol{A}$ is composed of all invertible modifications obeying the following matrix

equation

$$H_L \cdot \tilde{A} = \tilde{A} \cdot B \tag{10}$$

where H_L is the legal modification matrix of size $[M^2/R] \times [M^2/R]$, $\tilde{A}$ is the general analysis matrix of size $[M^2/R] \times [2 \cdot M - R]$ representing the transformation $\boldsymbol{A}$, and B is an arbitrary nonsingular matrix of size $[2 \cdot M - R] \times [2 \cdot M - R]$ which controls the modification characteristics. In this case there exists a scrambling system as in Fig. 1 that reconstructs the original input signal without error. The result expressed in eqn. (10) is important, as it enables checking whether a given modification matrix is legal. In addition, a legal modification with special characteristics can be synthesized using (10) through the selection of B, as it is demonstrated for the time and frequency scrambling in the next section. When the scrambling transformation is not legal, the issue of optimal synthesis arises. This problem is dealt with in detail in [9]. In the critically sampled case, the analysis transformation $\boldsymbol{A}$ is one-to-one and onto $\mathcal{W}$. Thus, there is always a unique synthesis transformation $\boldsymbol{S} = \boldsymbol{A}^{-1}$, and all scrambling transformations $\boldsymbol{H}$ are legal.

5. TIME AND FREQUENCY SCRAMBLING SCHEMES

Because of the close resemblance between the Gabor expansion and the STFT, the time and frequency scrambling techniques first introduced in [2,3] constitute special cases of the proposed scheme. In these schemes a permutation is performed on the STFT vectors of the signal, or on their IDFT, yielding frequency-domain scrambling or time domain scrambling correspondingly. These scrambling methods are claimed to overcome the effect of synchronization error usually encountered in scrambling schemes. In this section these time and frequency schemes are analyzed. It is shown that only channel delays of $r \cdot R$ samples are resolved, as opposed to the claim in [2,3] which says that any shift parameter R less then or equal to the transform size M will resolve the synchronization problem, In addition, in accordance with the previous section results, the class of legal modifications for the time and frequency scrambling schemes is determined. This class of legal modification constitute a special case of the class of all legal modifications defined in (10), and coincides with the legal modifications found in [5].

The set of M-dimensional short-time-vectors (STV) is produced by multiplying the input signal $x(n)$ by a sliding window sequence $\hat{\psi}(n)$. Using the assumption of a finite extent window of length M, the $p-th$ vector in the STV is given by

$$X_p(n) = \hat{\psi}^*(n - pR) \cdot x(n); \quad 0 \le n \le M-1, \quad -\infty < p < \infty \tag{11}$$

The case where the window length is greater than M is carried out utilizing the WOLA method [7]. The STFT is obtained by applying an M-size DFT on each of the STV, i.e.,

$$X_{p,q} = DFT\{X_p(n)\} \tag{12}$$

As proven in [5], the existence and properties of scrambling systems such as introduced here, are independent of the specific transform used. Thus, without loss of generality, by using the *identity transformation* instead of the DFT, the STV are considered instead of the STFT. As a result, the legality of the modifications as well as their properties concerning synchronization problems are identical in the time and in the frequency scrambling schemes.

The special case of time-invariant scrambling in time is represented by a $M \times M$ nonsingular matrix Q operating on each of the STV, i.e.,

$$Y_p(n) = Q\, X_p(n) \tag{13}$$

It is a well known fact that decimation systems are inherently time-varying, thus the Gabor expansion scheme which comprises of a decimation system becomes a time-varying system. Indeed, the Gabor expansion of a delayed signal generally does not coincide with the delayed Gabor expansion of the signal. The Gabor expansion constitute a *time-invariant* system only for delays of $r \cdot R$ samples (where R is the shift parameter, and r is any integer). Thus, the encoding, i.e., inverse modification of the Gabor expansion of the delayed version of the scrambled speech, and transformation back to the time-domain, would not produce a delayed version of the original speech, unless the delay equals $r \cdot R$. This contradicts the claim raised in [2,3], that any delay would be resolved as far as $R \le M$.

The above argumentation indicates that small values of R should be chosen to properly overcome the synchronization problem. However, as it is shown below, small values of R reduces the set of legal modifications, rendering the scrambling scheme less secure.

The class of legal modifications for the time-invariant special case is derived from (10) by imposing additional constraints, such as in (13). For this class of legal modifications, H_L is composed of M/R block-diagonal square matrices Q of size M having the following form

$$Q = diag\{P_i\} \quad , \quad 1 \le i \le M/R \tag{14}$$

where P_i is a set of square matrices of size R given by

$$P_i = \hat{A}_i \cdot \hat{B} \cdot \hat{A}_i^{-1} \tag{15}$$

$\hat{A}_i$ is the $R \times R$ sub matrix beginning at coordinate $((i-1) \cdot R + 1, (i-1) \cdot R + 1)$ of the analysis matrix A, and $\hat{B}$ is an arbitrary nonsingular matrix of size $R \times R$ which controls the scrambling characteristics. When $\hat{B}$ is chosen to be a permutation matrix, the resulting legal modification is in agreement with the modification obtained in [5]

Obviously, in order to overcome the synchronization problem, the time-invariant case should be used. In this case the class of legal modifications is reduced as R becomes smaller, resulting in a less secure scrambling scheme. Furthermore, since the modification matrix Q is a block-diagonal matrix with blocks of size R, the scrambling is performed on small segments of the speech signal, further reducing the security level of the scheme. Thus, as a conclusion from the above discussion, there is trade-off in choosing R which controls the synchronization error on one hand, and the level of security achieved by the scrambling scheme on the other.

6. SIMULATION RESULTS

The performance of the proposed scrambling scheme was examined on an actual speech signal. The performance, i.e., the residual intelligibility and the degradation introduced when using the scheme, was tested using a distortion-less channel, as well as in the presence of channel delays. Throughout, an Hamming window of length $N = 64$ was used as the analysis window, and the transform size M was taken to be $M = 64$. First, an arbitrary (but legal) modification matrix was tested under the conditions of a distortion-less channel. The results showed that there wasn't any residual intelligibility in the scrambled speech and that the reconstructed signal at the receiver was perfect. Next, the time-invariant scrambling scheme was considered for two values of the shift parameter $R = 2$ and $R = M$. For $R = 2$ the scrambled speech contained some residual intelligibility, and channel delays of $r \cdot 2$ were reconstructed perfectly. Channel delays different then $r \cdot 2$ destroyed the reconstructed signal. For $R = M$ the scrambled speech was unintelligible, and again channel delays different then $r \cdot R$ totally degraded the reconstructed speech. Finally, we tried to find, on a trial and error basis, a modification matrix which reasonably withstands channel delays other then $r \cdot R$. We succeeded in finding such a modification, however we can't propose a rigorous way of finding such matrices, and more research is needed concerning this problem.

7. SUMMARY AND CONCLUSIONS

This paper proposes a Gabor-domain analog speech scrambling technique which creates a variety of scrambling options, and generalizes current scrambling techniques in time and in frequency. The proposed scheme can perform any invertible modification, and not necessarily the commonly used permutation operation, rendering the scheme more secure. In addition, the synchronization problem which usually accompanies scrambling schemes was studied. It is shown that the claim raised by other authors [2,3] that the problem is resolved by such a scheme fails to be correct. The proposed scheme overcomes only channel delays which are a multiple of the shift parameter R, thus, small values of R should be chosen to properly reduce the synchronization problem. The selection of small values of R leads to the problem of illegal modifications which appears in the oversampled Gabor case, and decreases the security level of the scrambling scheme. The class of all legal modifications in the general case, and in the time and frequency scrambling special cases is determined. The trade-off in choosing R which controls the synchronization problem and the security level is considered. The performance of the proposed scheme is demonstrated by simulations.

REFERENCES

[1] N. S. Jayant, "Analog Scramblers for Speech Privacy" Computers and Security J., New York: North-Holland, 1982, pp. 275-189.

[2] L. S. Lee, "A Speech Security System not Requiring Synchronization", IEEE Communications Magazine, Vol. 23, No. 7 July 1985 pp 42-55.

[3] L. S. Lee and G. C. Chou, "A General Theory for Asynchronous Speech Encryption Techniques", IEEE Journal on Selected Areas in Communications, Vol. SAC-4, No. 2, March 1986, pp. 280-287.

[4] S. Farkash and S. Raz, "Linear Systems in Gabor Time-Frequency Space", submitted to IEEE Trans. on ASSP.

[5] A. Dembo and D. Malah, "Signal Synthesis from Modified Discrete Short Time Transform", IEEE Trans. on Acoust., Speech, Signal Processing, Vol. ASSP-36 no. 2 pp 168-181, Feb. 1988.

[6] J. Wexler and S. Raz, "Discrete Gabor Expansion", submitted.

[7] R. E. Crochiere and L. R. Rabiner, Multirate Digital Signal Processing. Englewood Cliffs, N.J.: Prentice-Hall 1983.

[8] Z. Shapiro and D. Malah, " An Algebraic Approach to Discrete Short-Time Fourier Transform Analysis and Synthesis", in Proc. ICASSP 1984, pp. 2.3.1-2.3.4.

[9] S. Farkash, "Gabor Representation and its Application to Signal Processing", D.Sc dissertation (in Hebrew), Technion-Israel Inst. Technology, 1991, in preparation.

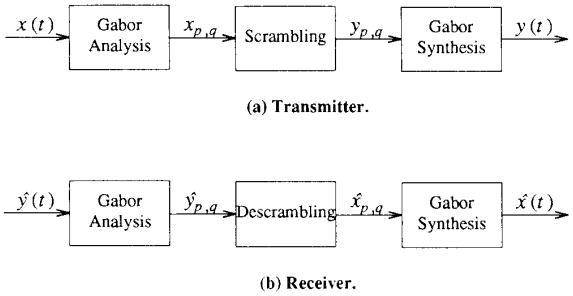

Fig. 1 : Block diagram of the Gabor scrambling scheme.

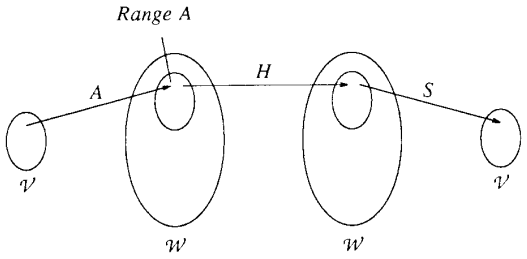

Fig. 2 : Visualization of the vector space approach to Gabor representation and the scrambling process.