Zeng et al. EURASIP Journal on Advances in Signal Processing 2012, 2012:257
http://asp.eurasipjournals.com/content/2012/1/257

EURASIP Journal on Advances in Signal Processing
a SpringerOpen Journal

RESEARCH Open Access

Scrambling-based speech encryption via compressed sensing

Li Zeng[1*], Xiongwei Zhang[1], Liang Chen[2], Zhangjun Fan[2] and Yonggang Wang[2]

Abstract

Conventional speech scramblers have three disadvantages, including heavy communication overhead, signal features underexploitation, and low attack resistance. In this study, we propose a scrambling-based speech encryption scheme via compressed sensing (CS). Distinguished from conventional scramblers, the above problems are solved in a unified framework by utilizing the advantages of CS. The presented encryption idea is general and easily applies to speech communication systems. Compared with the state-of-the-art methods, the proposed scheme provides lower residual intelligibility and greater cryptanalytic efforts. Meanwhile, it ensures desirable channel usage and notable resistibility to hostile attack. Extensive experimental results also confirm the effectiveness of the proposed scheme.

Keywords: Speech encryption, Scrambling, Compressed sensing, Residual intelligibility, Keyspace

1. Introduction

Encryption, dating back to BC, is essential for information security in modern society [1]. Information espionages, including illegal surveillance and wiretapping, have emerged due to the wide applications of speech communication in national defense, economy and trade, scientific research, and social affairs. With security an ever more vital requisite of communications systems, speech encryption has attracted substantial acceptance as an effective means of enhancing protection in both military and civilian applications.

Two main categories of technologies have been developed for this purpose. The first one is content protection through encryption, e.g., speech scrambler [2-9]. Proper decryption of the data requires a key or the so-called scrambling matrix. The second one is digital watermarking, which aims at embedding messages into the multimedia data [10]. Intuitively, the time domain sample scrambling method is thus far the most attractive and desired, because it simply takes a segment of time domain sample values and directly scrambles them into a different segment of samples. In this article, we focus on the scrambling-based encryption.

Earlier speech scramblers disorder the original signal using specific sequence or matrix, such as pseudorandom sequence, Fibonacci transform [2], Hadamard matrix [3,4], and so on. The main disadvantage shared by these methods is that the decryption key is invariable. Since the performance of computer hardware has incredibly been improved, these methods could easily be deciphered. To alleviate this problem, researchers proposed to employ new key schedules, such as stochastic matrix [5] and Latin square [6], to improve the strength of security [7,8]. However, the improved algorithms also result in heavy transmission load due to their disability of compressing the original signal. Consequently, the speech compression methods are integrated into the process of encryption, e.g., G.729 mixed excitation linear prediction (MELP) and AMR [9]. Indeed, the combination of compression and scrambling leads to less costly encrypted data. But the parametric coding algorithms are of low robustness in the presence of noise or other hostile attacks. Besides, the performances of such algorithms depend heavily upon the encryption operator, and the character of speech itself is not well utilized. More recently, researches in nonlinear process have shown that the chaotic signal is very suitable for secure communications. However, chaotic system is sensitive to disturbance and requires strict self-synchronization, which limits its practical applications [11,12].

* Correspondence: mofistova@qq.com
[1]College of Command Information Systems, PLA University of Science and Technology, Nanjing, China
Full list of author information is available at the end of the article

According to Del Re et al. [13], the degree of security (deciphering difficulty) provided by a speech encryption system is related to (1) residual intelligibility (the amount of intelligibility left over in the encrypted signal) and (2) keyspace (the number of keys available for encryption). In general, the lower a scrambling system's residual intelligibility and the bigger its keyspace, the higher its degree of security. After the propose of the stochastic approach [14], the keyspace of a scrambler is commonly measured by the number of encryption operators, namely the scrambling matrix [3,4,6,13].

To summarize, a channel-saving and anti-attack speech scrambler is a major issue to be addressed. In the meantime, it should attain residual intelligibility as low as possible and provide keyspace as large as possible to increase the cryptogram immunity to cryptanalysis. Despite the improvements achieved by the aforementioned works, few investigations manage to simultaneously address these problems. In light of this consideration, we apply compressed sensing (CS) [15-17] to speech encryption, due to its promising capability in signal compression and its notable robustness to hostile attacks.

In this article, we tackle the issue on scrambling-based speech encryption via CS by exploiting the sparsity of speech over the Karhunen–Loeve (K–L) incoherent dictionary [18]. Distinguished from existing schemes, we scramble the dimensional-reduced measurements instead of the original speech. The algorithm proposed in this article is motivated by the following idea: if two independent signals $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are aliased and scrambled in the same space by a stochastic matrix, the intelligibility of the original signal can be eliminated, and it is hard for the eavesdroppers to get any information barely from the mixture [19]. Observations show that the measurement vector of speech exhibits noise-like nature. However, the envelope of compressed data still retains considerable information of the original speech [20,21], we therefore alias and scramble the two measurement vectors of separated speech instead of using the envelope as ciphered data directly.

To be specific, the presented scheme contains two stages: encryption and decryption. At the encryption stage, we compress and encrypt the speech. The original signal is separated into two independent parts and sparsely represented by the corresponding K–L incoherent dictionaries. Next, the sparse vectors of the two parts are measured by stochastic matrices. Afterwards, these low-dimensional measurements are mixed and scrambled by the scrambling matrix, which is constructed from the null space basis (NSB) of the aligned sensing matrix using singular value decomposition (SVD). At the decryption stage, the inverse operator is constructed to eliminate each aliased measurements part individually. At last, the separated speech parts are reconstructed using the orthogonal matching pursuit (OMP) [22], and assembled to recover the speech. Experimental results demonstrate that the encrypted signal of proposed scheme has low transmission cost and low residual intelligibility. It provides immense cryptogram immunity and exhibits notable attack resistance.

The rest of this article is outlined as follows. In Section 2, we introduce the K–L incoherent dictionary to sparsely represent speech signals. Section 3 explains the encryption idea that we seek to address, and expatiates on the detailed procedure of encryption and decryption. The residual intelligibility, encryption strength, and robustness performance of the proposed scheme, together with experimental results, are presented and discussed in Section 4. Conclusions are drawn in Section 5.

2. The sparse representation of speech

Sparse representation is a critical step of CS [16], since one can obtain the dimension-reduced signal on the basis of sparse vectors. In this section, we employ a practical sparsifying dictionary to sparsely represent the speech signal.

The K–L expansion [18] describes a stochastic process in the form of incoherent random principle components and the corresponding deterministic orthogonal basis. Thus, the main structure of the process can be captured by a few expansion terms. Assume a real second-order moment stochastic process $\{x(t), t \in [0, 1]\}$, its K–L expansion is

$$x(t) = \sum_{k=1}^{\infty} a_k \phi_k(t) \tag{1}$$

where $a_k = \int_0^1 x(t)\phi_k(t)dt$. The orthogonal bases $\{\phi_k(t)\}$ are eigenfunctions of the signal autocorrelation $R_x(t,u)$. They can be used to design the atoms of signal dictionary. $\{\phi_k(t)\}$ and the corresponding eigenvalues y_k satisfy the Fredholm integral equation [18,23]

$$y_k \phi_k(t) = \int_0^1 R_x(t,u)\phi_k(u)du \tag{2}$$

However, for a practical signal, Equation (2) is hard to solve due to the complexity of the signal autocorrelation. Since the autocorrelation of speech signal decreases rapidly within a low delay, we approximate it by the exponential function: $R_x(t, u) = R_x(0)e^{-\mu|t-u|}$, where μ is the attenuation coefficient. Substituting $R_x(t,u)$ into (2) yields

$$\begin{aligned} y_k \phi(t) &= \int_0^1 R_x(t,u)\phi(u)du \\ &= R_x(0)\left(e^{-\mu t}\int_0^t e^{\mu\mu}\phi(u)du + e^{\mu t}\int_t^1 e^{-\mu\mu}\phi(u)du\right) \end{aligned} \tag{3}$$

By solving (3) with the boundary conditions and eliminating the zero particular solution $\phi_0(t)=0$ (it cannot be

used as an eigenfunction), we obtain the orthogonal basis

$$\phi_k(t) = \left(\frac{k\pi}{\mu}\right)\cos(k\pi t) + \sin(k\pi t),\ \ k \in \boldsymbol{Z}\setminus\{0\} \quad (4)$$

After adding $\phi_0(t)=1$, the complete orthogonal K–L dictionary $\boldsymbol{E}$ is represented by

$$\boldsymbol{E} = \{\phi_0(t)\} \cup \{\phi_k(t), k \in Z\setminus\{0\}\} \quad (5)$$

where k stands for the number of atoms in the dictionary. For digital signal processing, the bases of $\boldsymbol{E}$ are sampled in the range of $0 \le t \le 1$ by uniform sampling. Let μ^* denote the optimal value of parameter μ, it is estimated by solving the following optimization problem:

$$\mu* = \arg\min_{\mu>0} \left\| R_x(\tau) - \hat{R}_x(\mu) \right\|_2^2 \quad (6)$$

where $R_x(\tau) = \frac{1}{(n-\tau)}\sum_{i=1}^{n-\tau} x(i)x(i+\tau)$ is the unbiased estimation of the autocorrelation of speech frame $\boldsymbol{x} \in \boldsymbol{R}^n$ with delay τ and $\hat{R}_x(\tau) = R_x(0)e^{-\mu|\tau|}, \tau = 0, 1, \cdots, n-1$.

Thus, the optimal discrete atoms are $\boldsymbol{e}_k=[e_k(1), \ldots, e_k(i), \ldots, e_k(n)]^T$, where

$$e_k(i) = \frac{k\pi}{\mu*}\cos\left(\frac{k\pi(i-1)}{n-1}\right) + \sin\left(\frac{k\pi(i=1)}{n-1}\right)(i = 1, \cdots, n) \quad (7)$$

Then add with $\boldsymbol{e}_0=[1, \ldots, 1]^T$, we construct the complete discrete speech dictionary as

$$\boldsymbol{D} = \{e_0\} \cup \{e_k, k \in \boldsymbol{Z}\setminus\{0\}\} \quad (8)$$

In this case, discrete uniform sampling changes the orthogonality of bases $\{\phi_k(t)\}$. Though they are not mathematically orthogonal, atoms in $\boldsymbol{D}$ are of low coherence and subsequently, $\boldsymbol{D}$ turns out to be an overcomplete incoherent dictionary.

Concisely, the presented dictionary is codetermined by the sinusoidal atoms (given in (7)) and the parameter μ^*. The structure of the sinusoidal atoms is based on the K–L expansion. It is the general paradigm of the dictionary, thus can be shared by both the compression and the recovering part. On the other hand, μ^* is affected by the character of each frame. It determines the detailed structure of the current dictionary. Hence, with μ^*, the corresponding dictionary can be rebuilt to recover the original speech in the recovering part.

Three types of speech frame (unvoiced, voiced, and transition sound) and their corresponding sparse vectors over the K–L incoherent dictionaries are shown in Figure 1. Here, k is set equal to the length of a speech frame.

3. Proposed encryption scheme

This section details the specification of the proposed scheme. Section 3.1 illustrates the derivation of proposed encryption scheme. The course of scrambling matrix designing is addressed in Section 3.2. Section 3.3 describes the decryption and recovering process.

3.1. CS-based scrambling

According to Candès and Wakin [17], the implication of sparsity is now clear: when a signal has a sparse expansion, one can discard the small coefficients without much perceptual loss. Hence, some minor but non-zero entries of the sparse vectors can be discarded before the measuring to further reduce the compression rate. In addition, they prove that, for a K-sparse signal $\boldsymbol{s} \in \boldsymbol{R}^n$ and a fixed basis $\boldsymbol{\Phi} \in \boldsymbol{R}^{m\times n}$ with atoms selected uniformly at random, the exact reconstruction of $\boldsymbol{s}$ from the measurements $\boldsymbol{y}=\boldsymbol{\Phi s} \in \boldsymbol{R}^m$ ($m \ll n$) is of overwhelming probability, as long as the number of observations obeys

$$m \ge C \cdot K \cdot \log n \quad (9)$$

for some real positive constant C. In this case, the original speech is compressed and the compression ratio is m:n. Here, we concentrate on the issue of speech encryption, and the quality of reconstructed speech will be given in Section 4.

The speech has proved to be a robust signal that can be perturbed in many different ways while remaining intelligible [24]. As depicted in Figure 2 (the compression ratio is set as 1:20), though the measurement vector exhibits some noise-like nature (Figure 2b), it is observed that the envelopes (red, dashed line) of the original speech and the compressed signal are of high similarity. This means the CS retains considerable information within the low-dimensional measurements.

Actually, neurons in the auditory brainstem sensitive to speech envelope have been found in mammalian physiological studies [25]. The envelope extracted using the Hilbert transform reveal that the envelope is most important for speech reception, namely the words are identified according to the envelope [20]. Research on the relationship between speech envelope and audio perceptual comprehending is still intensely ongoing [26]. More recently, Mehmet Cenk et. al. [21] have investigated the perceptual features for automatic emotion recognition with temporal envelopes.

Since the strong connection between residual intelligibility and speech envelope, our goal aims to come up with a new algorithm which is able to utilize the sparsity of speech signal as well as to decrease the residual intelligibility of the compressed data. In view of this consideration, the CS-based scrambling approach is employed for its straightforwardness.

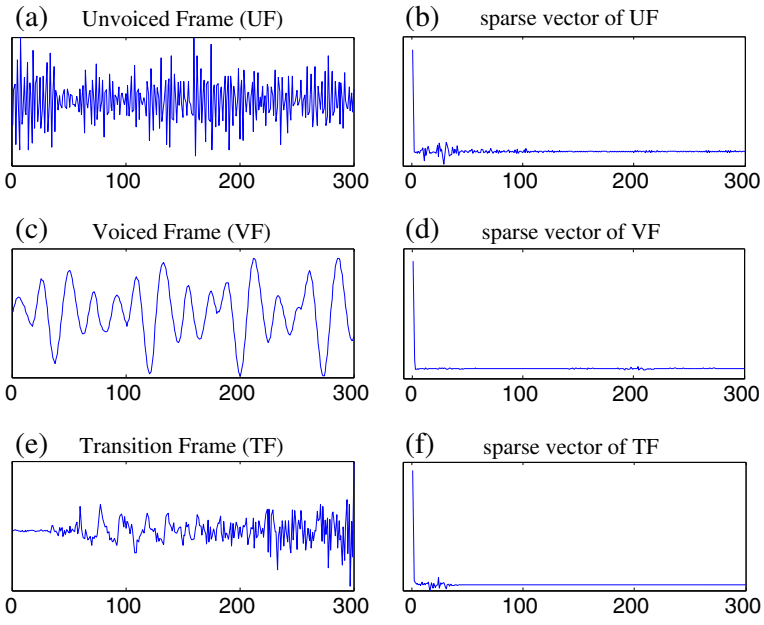

Figure 1 Sparsity of three types of speech over K–L incoherent dictionary. The horizontal axes stand for the amplitudes of signals, the vertical axes stand for the indices of vectors.

Previous studies [5,7,8] have demonstrated the security of using stochastic matrix as the key. Coincidentally, the sensing matrix with respect to CS is also a stochastic matrix. It therefore can be used as the scrambling matrix to decrease the residual intelligibility of the dimensional-reduced signal. Nevertheless, we notice that the dimension of the compressed signal is not in accordance with that of the sensing matrix. In other words, $\boldsymbol{y}\in\boldsymbol{R}^{m}$ cannot be scrambled directly by $\boldsymbol{\Phi}\in\boldsymbol{R}^{m\times n}$ without dimensional variation. To solve this problem, one feasible way is to select a group of random atoms from $\boldsymbol{\Phi}$ to form an $m\times m$ scrambling matrix. But the selection schedule will bring about additional communication load. As a consequence, based on the compressed speech sensing, we

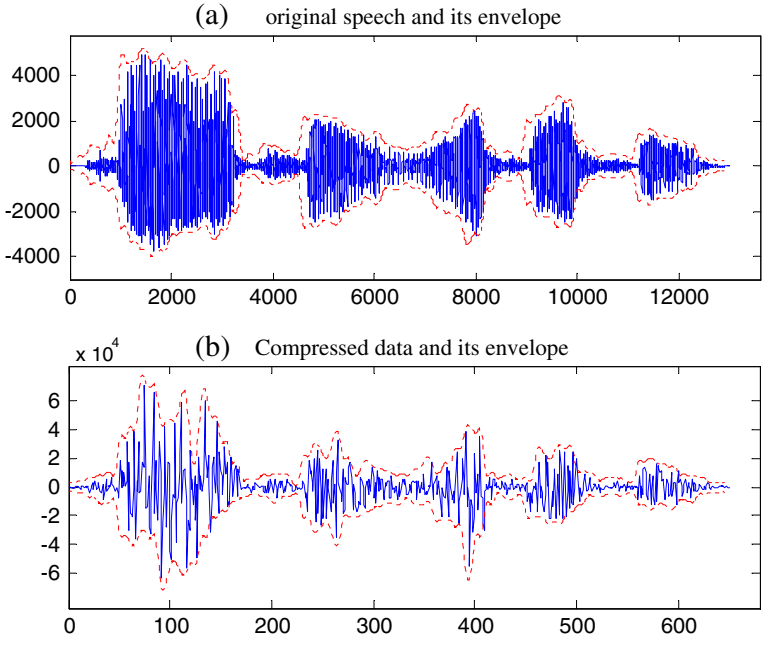

Figure 2 Comparison of the envelopes (red, dashed line) of original and compressed signal.

have designed a new paradigm of scrambling matrix that alias and scramble two volumes of compressed data $\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix}$ together.

3.2. Design of the scrambler

As mentioned above, one can hardly get any information barely from the result if two independent signals $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are aliased and scrambled in the same space. As shown in Figure 3, the original speech is thereby separated into two parts. For the sake of the independency of new speech segments, we set every four consecutive frames (800 samples at sampling rate 8 kHz) as a segmentation piece by considering two facts: (1) the quasi-periodic property of speech endures about 50 ms; (2) the auditory tolerance to delay is about 200 ms. Each cube in Figure 3 represents such a piece.

Next, the new signals are sparsely represented over the corresponding K–L incoherent dictionaries, and then measured by different stochastic matrices to get the compressed signals individually. Since using only one fixed matrix does not always hold the restricted isometry property [17] and will result in undesirable reconstructions, we randomly choose the matrices from the stochastic matrix dictionary $\boldsymbol{B}=\{\boldsymbol{\Phi}_j: j=1, 2, \ldots, L\}$.

Such matrix dictionary is constructed in advance. During the construction process, each randomly generated stochastic matrix is tested with a group of different speech frames. If the correct reconstruction rate of this matrix is acceptable, it is chosen as a dictionary atom. In this study, we set the accepting threshold of correct reconstruction rate as 80%. As a matter of fact, almost all random matrices are CS matrices [15], thus the number of atoms L in the dictionary can be set according to practical requirements.

The compressed data $\boldsymbol{y}_i$ are pre-reconstructed until the final selected sensing matrices $\boldsymbol{\Phi}_i$, i=1, 2 ensure precise reconstruction. Similarly, let α_i, i=1,2 denote the normalized indices of these matrices in the dictionary and let $\boldsymbol{D}_i$, i=1,2 stand for the sparsifying matrices of the two speech parts. Then, the encrypted signal $\boldsymbol{y}^D$ is obtained by aliasing and scrambling the two low-dimensional measurement vectors with the selected matrices $\boldsymbol{\Phi}_i$.

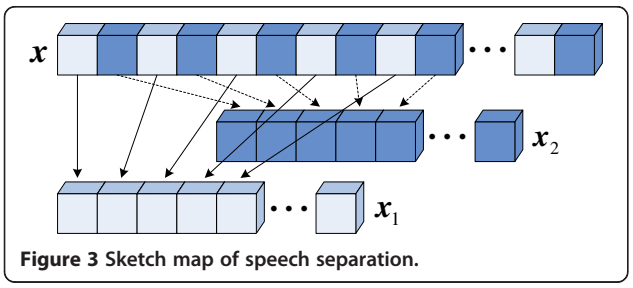

Figure 3 Sketch map of speech separation.

$$\boldsymbol{y}^D = f(\boldsymbol{y}_1, \boldsymbol{y}_2; \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2) \tag{10}$$

Subsequently, one heuristic approach is to design a scrambling matrix schedule that is of high security together with its inverse operator for decryption. Due to the independency of the two speech parts, their corresponding measurements $\boldsymbol{y}_i=\boldsymbol{\Phi}_i\boldsymbol{D}_i^T\boldsymbol{x}_i$, i=1, 2 are also incoherent. We can remove any one of them by its orthocomplement without damaging the other one. Without loss of generality, we take $\boldsymbol{y}_1$ for the following illustrations.

Assume there exists a vector $\boldsymbol{z}$ from $(\boldsymbol{\Phi}_1^T)^{\perp}$ that is orthogonal to $\boldsymbol{\Phi}_1^T$, i.e., $\boldsymbol{z}^T\boldsymbol{\Phi}_1=0$. Multiplying $\boldsymbol{z}$ with the encrypted data $\boldsymbol{y}^D$ yields

$$\boldsymbol{z}^T\boldsymbol{y}_1 = \left(\boldsymbol{z}^T\boldsymbol{\Phi}_1\right)\boldsymbol{D}_1^T\boldsymbol{x}_1 = 0 \tag{11}$$

thus the $\boldsymbol{y}_1$ part comprised in $\boldsymbol{y}^D$ is eliminated. Then, we can reconstruct $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ by reconstruction algorithms and assemble them to further obtain the recovered speech $\hat{\boldsymbol{x}}$.

Since the operation objects are matrices and vectors in practice, the orthocomplement designing problem turns out to be orthogonal vector designing. Following related linear space theories [27], $\boldsymbol{z}$ can be presented by a linear combination of the vectors in the non-trivial NSB of $\boldsymbol{\Phi}_1$, denoted as Null($\boldsymbol{\Phi}_1$). The rank of the stochastic matrix $\boldsymbol{\Phi}_1\in\boldsymbol{R}^{m\times n}(m<n)$ is m. However, the dimensions of $\boldsymbol{\Phi}_1$ and the null space of $\boldsymbol{\Phi}_1^H$ have following relationship

$$\dim\left[Null\left(\boldsymbol{\Phi}_1^T\right)\right] = m - \dim\left(\boldsymbol{\Phi}_1^T\right) = 0 \tag{12}$$

According to (12), if we choose $\boldsymbol{\Phi}_i$ to scramble $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ directly, the two measurement parts cannot be separated for decryption since Null($\boldsymbol{\Phi}_1$) does not exist. To ensure the inverse operation, we construct a non-full row rank matrix $\boldsymbol{\Phi}_1^D = \begin{bmatrix} \boldsymbol{\Phi}_1 \\ \boldsymbol{\Phi}_1 \end{bmatrix} \in \boldsymbol{R}^{2m\times n}$. The dimension of $\boldsymbol{\Phi}_1^D$ is m. By employing the conclusion drawn from [27], we have

$$\dim\left(\boldsymbol{\Phi}_1{}^D\right) + \dim\left[Null\left(\left(\boldsymbol{\Phi}_1{}^D\right)^H\right)\right] = 2m \tag{13}$$

Then,

$$\begin{aligned}\dim\left[Null\left(\left(\boldsymbol{\Phi}_1{}^D\right)^H\right)\right] &= 2m - \dim\left(\boldsymbol{\Phi}_1{}^D\right) \\ &= m\end{aligned} \tag{14}$$

This ensures the existence of null space of $\boldsymbol{\Phi}_1^D$ and it can be constructed by SVD: for a matrix $\boldsymbol{\Phi}_1^D \in \boldsymbol{R}^{2m\times n}$ with rank($\boldsymbol{\Phi}_1^D$)=m, it can be decomposed as

$$\boldsymbol{\Phi}_1{}^D = \boldsymbol{U}\sum\boldsymbol{V}^H \tag{15}$$

Then the m left singular vectors $\{\boldsymbol{u}_{m+1}, \boldsymbol{u}_{m+1}, \ldots, \boldsymbol{u}_{2m}\}$ that correspond to the non-zero singular values are

orthonormal basis of the conjugate transpose matrix $(\boldsymbol{\Phi}_1^D)^H$, that is

$$Null\left(\left(\boldsymbol{\Phi}_1{}^D\right)^H\right) = Span\{\boldsymbol{u}_{m+1}, \boldsymbol{u}_{m+1}, \ldots, \boldsymbol{u}_{2m}\} \quad (16)$$

where $\boldsymbol{U}\in\boldsymbol{R}^{2m\times 2m}$ and $\boldsymbol{V}\in\boldsymbol{R}^{m\times m}$ are unitary matrices. Here $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 0 \\ 0 \quad 0 \end{bmatrix}$, where $\boldsymbol{\Sigma}_1$=diag(σ_1, σ_2, …, σ_m), and σ_i are the eigenvalues of $(\boldsymbol{\Phi}_1^D)^H\boldsymbol{\Phi}_1^D$.

With (16), we obtain the NSB of $(\boldsymbol{\Phi}_1^D)^H$: $\boldsymbol{N}\in\boldsymbol{R}^{2m\times m}$, denoted as $\boldsymbol{N} = \begin{bmatrix} \boldsymbol{N}_1 \\ \boldsymbol{N}_2 \end{bmatrix}$ with property $\boldsymbol{N}_1$=−$\boldsymbol{N}_2$. $\boldsymbol{N}_1$, $\boldsymbol{N}_2$ are full rank matrices and therefore have inverse matrices, as proved in Appendix. In this case, the matrix $\boldsymbol{\Phi}_1^D$ provides available NSB. In other words, if we use the NSB matrix to construct a scrambling matrix for the aligned measurements $\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix}$, the inverse operator for decryption is available.

For this consideration, the scrambling matrix is designed as $\boldsymbol{S} = \begin{bmatrix} \alpha_1\boldsymbol{I} & \boldsymbol{N}_1 \\ \boldsymbol{N}_2^{-1} & \alpha_2\boldsymbol{I} \end{bmatrix}\in R^{2m\times 2m}$ to alias and disorder the measurements $\boldsymbol{y}_1$, $\boldsymbol{y}_2$. Here, α_i are the corresponding normalized indices of $\boldsymbol{\Phi}_i$ in the dictionary $\boldsymbol{B}$, so that every frame of encrypted signal is aliased in different proportion to enhance the encryption complexity. The identity matrix $\boldsymbol{I}\in\boldsymbol{R}^{m\times m}$ is used to adapt the dimensions of $\boldsymbol{S}$ and $\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix}$. The final encrypted signal $\boldsymbol{y}^D$ is given in (17). Every compressed part $\boldsymbol{y}_i$ is scrambled and the scrambled data are aliased with each other. The whole process of encryption is shown in Figure 4.

$$\boldsymbol{y}^D = \begin{bmatrix} \alpha_1\boldsymbol{I} & \boldsymbol{N}_1 \\ \boldsymbol{N}_2^{-1} & \alpha_2\boldsymbol{I} \end{bmatrix}\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} \alpha_1\boldsymbol{y}_1 + \boldsymbol{N}_1\boldsymbol{y}_2 \\ \boldsymbol{N}_2^{-1}\boldsymbol{y}_1 + \alpha_2\boldsymbol{y}_2 \end{bmatrix} \quad (17)$$

For communication, the encrypted data $\boldsymbol{y}^D$, the dictionary parameters μ_i^*, and the indices α_i of sensing matrix are transmitted to the decryption end. Then the decryption operator can be constructed by $\boldsymbol{\Phi}_i$. As such, the stochastic matrix dictionary $\boldsymbol{B}$ is also the key book shared by both the encryption and the decryption parts. This key book will never be exposed in the channel. In practice, it is irregularly updated and the number of its atoms L can adaptively be set to meet the requirements of the system.

3.3. Decryption

Figure 5 depicts the decryption procedure. We get $\boldsymbol{\Phi}_1$ with the index α_1 from the transmitted data, and construct the matrix $\boldsymbol{S}'$=$[\boldsymbol{N}_2^{-1}\ -\alpha_2\boldsymbol{I}_1]$ with $\boldsymbol{\Phi}_1^D$ by SVD. The $\boldsymbol{y}_1$ part is removed as follows.

$$\begin{aligned} \boldsymbol{S}'\cdot\boldsymbol{y}^D &= \begin{bmatrix} \boldsymbol{N}_2^{-1} & -\alpha_2\boldsymbol{I}_1 \end{bmatrix}\begin{bmatrix} \boldsymbol{y}_1 + \boldsymbol{N}_1\boldsymbol{y}_2 \\ \alpha_1\boldsymbol{y}_1 + \boldsymbol{N}_1^{-1}\boldsymbol{y}_2 \end{bmatrix} \\ &= \left(\boldsymbol{N}_2^{-1}\boldsymbol{N}_1 - \alpha_2^2\boldsymbol{I}\right)\boldsymbol{y}_2 \\ &= \boldsymbol{t}\in\boldsymbol{R}^{m\times 1} \end{aligned} \quad (18)$$

As mentioned above, $\boldsymbol{N}_1$ and $\boldsymbol{N}_2$ are full rank matrices. With the inverse matrix $\boldsymbol{N}_2^{-1}$, we can get $\boldsymbol{y}_2$ by multiplying the $\boldsymbol{y}_1$—removed data $\boldsymbol{t}$ with the matrix $(\boldsymbol{N}_2^{-1}\boldsymbol{N}_1\ -\alpha_2^2\boldsymbol{I})^{-1}$. Analogously, $\boldsymbol{y}_1$ is decrypted as

$$\boldsymbol{y}_1 = \left(\alpha_1\alpha_2\boldsymbol{I} - \boldsymbol{N}_1\boldsymbol{N}_2^{-1}\right)^{-1}[\alpha_2\boldsymbol{I} \quad -\boldsymbol{N}_1]\boldsymbol{y}^D \quad (19)$$

When the measurement vectors $\boldsymbol{y}_i$ and the dictionary parameters μ_i^* are derived, the two speech parts can be recovered using OMP algorithm [22]. Finally, by assembling

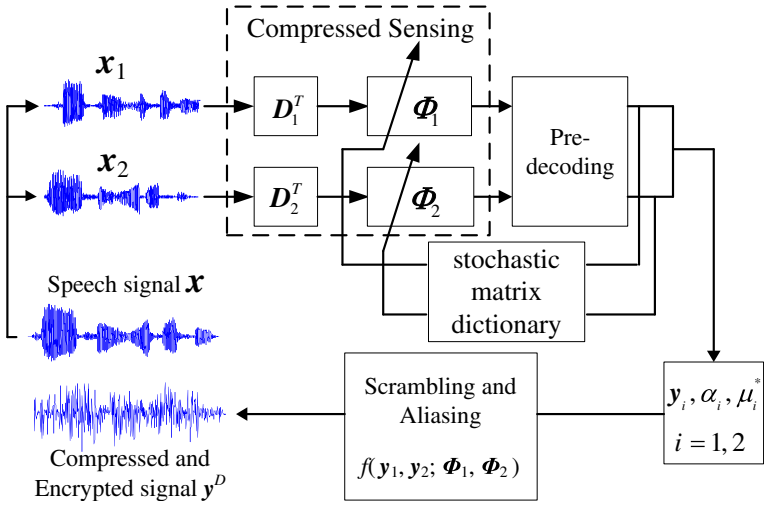

Figure 4 The overall framework of encryption.

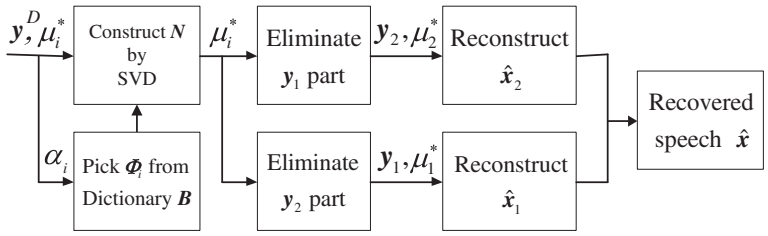

Figure 5 The overall framework of decryption.

the two reconstructed parts, we obtain the recovered speech $\hat{x}$.

In theory, without the key book B, it is very hard for the eavesdroppers to decipher the encrypted signal even though they have cryptanalyzed the encryption mechanism. The strength of security is discussed in the following section.

4. Experimental results and discussions

In this section, the performances of the proposed encryption scheme are evaluated from three perspectives: (1) the residual intelligibility of the encrypted signal; (2) the strength of security; (3) resistance to hostile attacks. We test over 20,000 frames of speech coming from several speakers with unlike characteristics (gender, age, pitch, regional accent). These test signals are taken randomly from TIMIT database and are sampled at 8kHz with length 25ms, that is, 200 samples per frame. In Section 4.1, residual intelligibility test results and discussions are presented. Section 4.2 analyzes the strength of security, and a possible deciphering technique is considered. Section 4.3 verifies the robustness of the proposed scheme in two conditions: in the presence of noise and low-pass filtering (LPF).

4.1. Residual intelligibility

The amount of intelligibility left over in the encrypted signal is measured by the envelope relevance between the original speech and the processed signal, given as

$$\rho = \frac{< \boldsymbol{E}_0, \boldsymbol{E}_p >}{\left(\|\boldsymbol{E}_0\|_2 \cdot \|\boldsymbol{E}_p\|_2 \right)^{\frac{1}{2}}} \tag{20}$$

where $\boldsymbol{E}_o$ and $\boldsymbol{E}_p$ denote the envelopes of original speech and the processed signal, respectively. Naturally, we interpolate the vector of $\boldsymbol{E}_p$ to reach the same dimension as $\boldsymbol{E}_o$ due to the operation of inner production in (20). We test two kinds of processed signal: the compressed signal (CoS) and the encrypted signal (ES).

According to the experimental statistics, when the compression rate (m/n) is above 5%, the salient information of speech can be captured, and acceptable reconstruction quality is derived with the K–L incoherent dictionary. On the other hand, though reconstruction quality improves with the increasing of compression rate, it is of no significance for signal compression with a high compression rate. Therefore, the average residual intelligibilities are performed at compression rates ranged from 5 to 10%.

As seen in Table 1, despite of the noise-like nature, the low-dimensional measurements still retain considerable information of the original speech. The envelope relevance between the aliased, scrambled signal, and the original speech exhibits a dramatic decrease in terms of residual intelligibility. In the meantime, it is noticed that the compression rates and the residual intelligibility are not remarkably related, which means one can choose the compression ratio adaptively without increasing the residual intelligibility. In addition, as a subjective method for predicting the quality of narrow-band speech, the mean opinion score (MOS) recommended by ITU-T P.862 [28] is adopted to evaluate the perceptual quality of the recovered speech, and the results are also presented in Table 1 to illustrate the relationship between speech quality and compression rate.

4.2. Strength of security

Following Shannon's landmark article [14], the majority of literatures on key generation may roughly be categorized into four basic approaches: information theory approach, system theory approach, complexity theory approach, and stochastic approach. Considering the key schedule, our encryption scheme belongs to the stochastic approach, and its security is generally measured by the scale of the keyspace. Thus, the keyspace of proposed scheme is analyzed in two cases to evaluate its strength of security.

Table 1 Residual intelligibilities of processed signals at different compression rates and the corresponding MOS scores of the recovered speech

Compression rate (%)		5	6	7	8	9	10
ρ	CoS	0.964	0.967	0.957	0.959	0.975	0.967
	ES	0.326	0.335	0.342	0.372	0.347	0.364
MOS		2.42	2.49	2.54	2.63	2.73	2.96

First, we consider the optimal case, in which the scrambling mechanism is thoroughly unknown to the unauthorized listener, including the key schedule, the structures of the scrambling matrix, and the sparsifying dictionary. Consequently, the essential approaches for decryption is given by

$$\begin{matrix} \boldsymbol{y}^D \\ \boldsymbol{S}(unknown) \end{matrix} \} \rightarrow \begin{matrix} \boldsymbol{y}_i \\ \boldsymbol{\Phi}_i(unknown) \end{matrix} \} \rightarrow \begin{matrix} \boldsymbol{s}_i \\ \boldsymbol{D}_i(unknown) \end{matrix} \} \rightarrow \hat{\boldsymbol{x}}_i \rightarrow \hat{\boldsymbol{x}} \tag{21}$$

In this case, all of the above three unknown matrices can be regarded as the key since one cannot obtain the original information without any one of them. Meanwhile, the extreme low residual intelligibility and the noise-like nature of the encrypted signal have hampered the statistical analysis methods [29]. Therefore, the most feasible way is to search for the key, and its complexity is directly decided by the scale of keyspace. Given the dimensions of $\boldsymbol{S}$, $\boldsymbol{\Phi}_i$, and $\boldsymbol{D}_i$, their keyspace sizes are $O\left(10^{(2m)^2}\right)$, $O(10^{mn})$, and $O(10^{mn})$, respectively. These are also the computational complexity for the searching process.

In the second case, we assume that an eavesdropper has a complete knowledge of the system, and has the necessary hardware to synchronize and isolate the frames. In other words, he knows that the scrambling matrix $\boldsymbol{S}$ can be constructed with $\boldsymbol{\Phi}_i$ and the sparsifying matrix $\boldsymbol{D}_i$ can simply be rebuilt for its characteristic structure. Hence, the security of the system is assumed to reside entirely with the selection of a key $\boldsymbol{\Phi}_i$. For the eavesdropper, the only task is to find the key. Since the randomness of $\boldsymbol{\Phi}_i$, the keyspace size is $O(10^{mn})$. In practical situations, the speech frames is of length n=180–220 and if we choose the compression rate as 5%, the length of compressed signal is m=9–11. Therefore, the order of magnitude of the keyspace is about 10^{2000}.

Table 2 compares the keyspace sizes and the compression ratios of the proposed scheme and some prior scramblers, which employ representative key schedules, including Hadamard matrix [4], Latin Square [6], dimensional-variable matrix [7], and chaos system [12]. As seen in Table 2, the proposed scheme provides larger keyspace and requires lower communication overhead.

Now let us consider a possible deciphering technique by dictionary learning regardless of deciphering delay. A cryptanalyst trying to break the system may be in possession of large amounts of encrypted signal except the key book, since it is held by both the encryption and decryption parts and never transmitted through the channel. He knows the complete specification of the system (scrambling mechanism, structure of sparsifying dictionary); he would like to deduce the key without considering the real-time requirement.

In mathematical language, the jth encryption operation is represented by

$$\boldsymbol{y}_{(j)}^D = \boldsymbol{S}^{(j)} \begin{bmatrix} \boldsymbol{y}_1^{(j)} \\ \boldsymbol{y}_2^{(j)} \end{bmatrix} \tag{22}$$

By wiretapping, the cryptanalyst has obtained enough encrypted signal $\boldsymbol{y}_{(j)}^D$. He would like to learn $\hat{\boldsymbol{S}}^{(j)}$ from $\boldsymbol{Y}^D$, where $\boldsymbol{Y}^D=[\boldsymbol{y}_{(1)}^D\ \boldsymbol{y}_{(2)}^D\ \cdots\ \boldsymbol{y}_{(j)}^D]$. Each $\begin{bmatrix} \boldsymbol{y}_1^{(j)} \\ \boldsymbol{y}_2^{(j)} \end{bmatrix}$ is unknown. Unfortunately, this is an optimization problem with no constraint conditions and thus cannot be solved, let alone the scrambling matrix $\boldsymbol{S}$ is variable but not fixed. To say the least, even though he is able to find some fixed $\hat{\boldsymbol{S}}$, he still cannot rebuild the sensing matrix $\boldsymbol{\Phi}$. This can be verified through Equations (15) and (16).

In fact, there would not be enough data and delay tolerance for cryptanalysis. For instance, in secure communications of military information or intelligence of espionage activities, the key information is expected to be as briefly as possible to ensure short durations. Therefore, the dictionary learning may not be a feasible approach in real cases.

4.3. Robustness performance

Since readily decipherable unintelligibility signals may also be generated in large keyspace, other factors, including bandwidth expansion, delay times, channel resistance (to noise, distortion, etc.), cannot be ignored in assessing security. Two types of attack are performed with the encrypted signal: (1) in the presence of additive white Gaussian noise (AWGN); (2) LPF.

Representative speech scrambling schemes are chosen to compare with the CS scheme. To be specific, the time-domain scrambling (TDS) [5] is adopted to stand for non-compressional scramblers. In parallel to it, the approximate 13 line μ-law pause code modulation (PCM) and the MELP [9] are, respectively, chosen to represent waveform coding and parametric coding, with respect to compressional scramblers.

Table 2 Comparison of keyspace size and compression ratio

	Proposed	Hadamard [4]	Latin Square [6]	ASVDS [7]	Chaos [12]
Keyspace	$O(10^{n^2})$	$O(10^n(n!)^2$	$O(10^{120})$	$O(n^2)$	$O(10^{64})$
Compression ratio	$\ll$1:1	1:1	1:1	1:1	1:1

The MOS is chosen as the subjective criterion. In addition, average-subsection signal-to-noise ratio (SNRseg) [30] is adopted as the objective criterion to evaluate the quality of recovered speech, given by (23).

$$SNRseg = 101g\left(\frac{1}{N_{frame}}\sum_{j=1}^{N_{frame}} SNRseg_j\right) \tag{23}$$

where $SNRseg_j = \sum_{i=1}^{n} \frac{x_j^2(i)}{\left[x_j(i) - \hat{x}_j(i)\right]^2}$ and N_{frame} denotes the total number of frames. The results are calculated and averaged for a test set of approximately 100 sentences randomly selected from the TIMIT database.

4.3.1. Noise resistance

AWGN is added to the encrypted signal of each scheme. The performances of the proposed and comparative schemes are compared. The compression ratio of CS is set as 1:10.

As shown in Figure 6a, it is observed that CS scheme always outperforms the comparative schemes for all degrees of contamination. As the signal-to-noise ratio (SNR) becomes higher, the superiority of CS scheme becomes more obvious and leads to a more favorable comparison; the compressional schemes, including PCM and MELP, perform worse, and these are verified by the SNRseg decrements as well (Figure 6b).

The results are mainly due to the use of stochastic matrix: it has extreme low column coherence. The studies [31,32] have shown that for a noiseless signal $\boldsymbol{y}=\boldsymbol{\Phi s}$, if the K-sparse vector $\boldsymbol{s}$ satisfies $\|\boldsymbol{s}\|_0 < \frac{1}{2}spark(\boldsymbol{\Phi})$, then the reconstruction $\hat{\boldsymbol{s}}$ from the contaminated measurements ($\boldsymbol{y}+\boldsymbol{e}$) satisfies

$$\|\boldsymbol{s} - \hat{\boldsymbol{s}}\|_2^2 \leq O\left(\frac{\|\boldsymbol{e}\|_2^2}{1 - M(2K-1)}\right) \tag{24}$$

where spark($\boldsymbol{\Phi}$) stands for the minimum number of columns of $\boldsymbol{\Phi}$ that are linearly dependent, and M is the "mutual coherence" of $\boldsymbol{\Phi}$ [31], defined as

$$M = \max_{1\leq i,j\leq n, i\neq j} \left|\frac{\boldsymbol{\Phi}_i^T \boldsymbol{\Phi}_j}{\|\boldsymbol{\Phi}_i\|_2 \|\boldsymbol{\Phi}_j\|_2}\right| \tag{25}$$

In a word, one can stably reconstruct the sparse vector $\boldsymbol{s}$ with error proportional to the noise level, provided that (1) the columns of the sensing matrix $\boldsymbol{\Phi}$ are weakly mutual correlated, and (2) the vector $\boldsymbol{s}$ is to some extent sparse. The reason why the reconstruction error is restricted is geometric in nature: summarily, the reconstruction $\hat{\boldsymbol{s}}$ is restrained within a tiny tubular wedge that surrounds the original vector $\boldsymbol{s}$, which ensures the stability of recovering (for further details please refer to [31], subsection 5.3). In particular, the method of convex

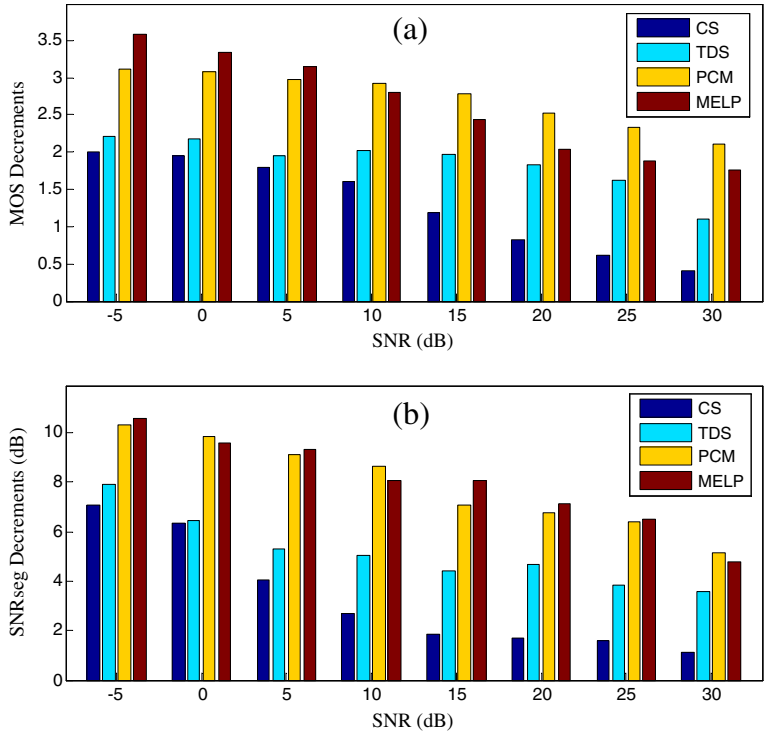

Figure 6 Comparison of scramblers in the presence of AWGN. (**a**) MOS decrements. (**b**) SNRseg decrements.

relaxation can identify a sparse signal in AWGN [33], and more sophisticated stable recovery schemes and boundaries have been investigated [34].

Technically speaking, the waveform coding scheme has no noise-resistant precaution and thus is vulnerable to noise. In terms of the parametric coding scheme, the noise will bring in errors to the feature parameters such as pitch period and voiced/unvoiced judgment. Once such parameters are contaminated, undesirable reconstruct distortion happens. For reasons given above, the noise resistance of the proposed scheme is better than the counterparts.

4.3.2. LPF

In terms of LPF, the decrements of MOS and SNRseg between speeches reconstructed from the filtered and the original encrypted data are compared. Also, the compression ratio of CS is set as 1:10.

As seen in Figure 7, CS scheme slightly outperforms the comparative schemes when the cutoff frequency ranges from 2400 to 2600 Hz. It gradually performs better along with the increasing of cutoff frequency and exhibits obvious competitive advantages. The TDS scheme ranks in the second place and the PCM scheme shows the worst performance.

The encrypted signal obtained from CS has dramatically removed the speech characters by aliasing and scrambling, thus it is of the best resistance to LPF. On the contrary, the TDS scheme still retains some speech structures, which makes its performances inferior to the CS scheme. The TDS scheme outperforms the other two comparative ones due to its robustness to time domain perturbations [24]. As for PCM and MELP schemes, their encoding signals have no spectral structures. All parts of the signal share the similar importance and therefore are vulnerable to this type of attack. Any damages to the encrypted signal would bring about serious reconstruction errors and deteriorate the auditory quality. As a consequence, these two schemes have the worst robustness to LPF.

5. Conclusions

This article presented a scrambling-based speech encryption algorithm via CS. A high degree of security can be achieved due to low residual intelligibility and large keyspace size. The immense complexity associated with the task of finding the scrambling matrix ensures the effectiveness of encryption. It affords notable robustness to common hostile attacks, while requires lower communicational costs and introduces only a slight (about 200ms) processing delay. Experimental results are included which demonstrate the improved performance of the scheme compared with state-of-the-art speech scramblers. As a future work, it is planned to investigate

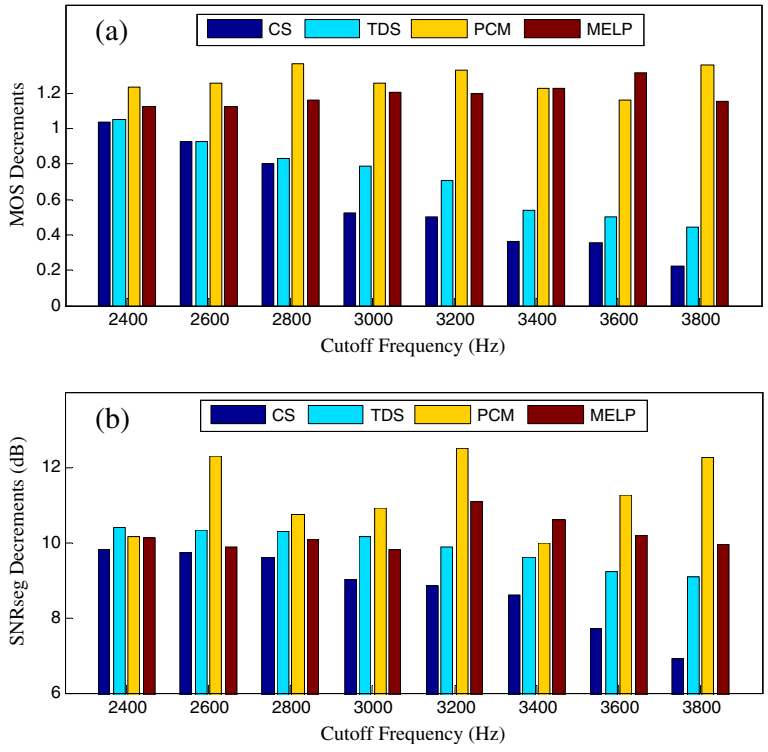

Figure 7 Comparison of scramblers for LPF performances. (**a**) MOS decrements. (**b**) SNRseg decrements.

more sophisticated speech sparse representation and reconstruction algorithms to further reduce the compression ratio and improve the auditory quality of the recovered speech.

Appendix

Proof of full rank property mentioned in Section 3.2 is given as follows.

As for full row rank matrix $\boldsymbol{P}\in\boldsymbol{R}^{m\times n}(m<n)$, rank($\boldsymbol{P}$)=$m$, denote its SVD as $\boldsymbol{P}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H$, where $\boldsymbol{\Sigma}=[\boldsymbol{\Sigma}_m\ \boldsymbol{O}_{m\times(n-m)}]\in\boldsymbol{R}^{m\times n}$, $\boldsymbol{\Sigma}_m=\mathrm{diag}(\sigma_1,\ \sigma_2,\ldots,\sigma_m)$. Here $\boldsymbol{U}$ and $\boldsymbol{V}$ are $m\times m$, $n\times n$ unitary matrices, respectively. $\sigma_i>0$ denote the square roots of non-zero eigenvalues of $\boldsymbol{P}^H\boldsymbol{P}$. With the SVD of $\boldsymbol{P}$, we have

$$\begin{aligned}\boldsymbol{P}^H\boldsymbol{P}&=\left(\boldsymbol{V}\boldsymbol{\Sigma}^H\boldsymbol{U}^H\right)\left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H\right)\\&=\boldsymbol{V}\begin{bmatrix}\boldsymbol{\Sigma}_m^2 & \boldsymbol{O}_{(n-m)\times(n-m)}\\ \boldsymbol{O}_{(n-m)\times(n-m)} & \boldsymbol{O}_{(n-m)\times(n-m)}\end{bmatrix}\boldsymbol{V}^H\end{aligned}\tag{26}$$

Then for $\boldsymbol{Q}=\begin{bmatrix}\boldsymbol{P}\\ \boldsymbol{P}\end{bmatrix}=\boldsymbol{U}'\boldsymbol{\Sigma}'\boldsymbol{V}'^H\in\boldsymbol{R}^{2m\times n}$ and rank($\boldsymbol{Q}$)=m, $\boldsymbol{Q}^H\boldsymbol{Q}$ is represented as

$$\begin{aligned}\boldsymbol{Q}^H\boldsymbol{Q}&=\begin{bmatrix}\boldsymbol{P}^H\boldsymbol{P}^H\end{bmatrix}\begin{bmatrix}\boldsymbol{P}\\ \boldsymbol{P}\end{bmatrix}=2\boldsymbol{P}^H\boldsymbol{P}\\&=2\boldsymbol{V}'\begin{bmatrix}\boldsymbol{\Sigma}_m^2 & \boldsymbol{O}_{(n-m)\times(n-m)}\\ \boldsymbol{O}_{(n-m)\times(n-m)}\boldsymbol{O}_{(n-m)\times(n-m)}\end{bmatrix}\boldsymbol{V}'^H\end{aligned}\tag{27}$$

where $\boldsymbol{U}'\in\boldsymbol{R}^{2m\times 2m}$, $\boldsymbol{\Sigma}'\in\boldsymbol{R}^{2m\times n}$, $\boldsymbol{V}'\in\boldsymbol{R}^{n\times n}$.

Comparing (26) and (27), it is noticed that $\boldsymbol{V}'=\boldsymbol{V}$, $\boldsymbol{\Sigma}'=\begin{bmatrix}\boldsymbol{\Sigma}_m & \boldsymbol{O}_{m\times(n-m)}\\ \boldsymbol{O}_{m\times n} & \boldsymbol{O}_{m\times(n-m)}\end{bmatrix}$ and $\boldsymbol{U}'\boldsymbol{U}'^H=2\boldsymbol{I}_{2m}$, where $\boldsymbol{I}_{2m}$ denotes the $2m\times 2m$ identity matrix. In this case, if there is a proper $\boldsymbol{U}\prime$, the SVD of $\boldsymbol{Q}$ can be obtained. It is verified that $\boldsymbol{U}'=\begin{bmatrix}\boldsymbol{U} & -\boldsymbol{U}\\ \boldsymbol{U} & \boldsymbol{U}\end{bmatrix}$ satisfies $\boldsymbol{U}'\boldsymbol{U}'^H=2\boldsymbol{I}_{2m}$, then (27) is rewritten as

$$\begin{aligned}\boldsymbol{Q}^H\boldsymbol{Q}&=2\boldsymbol{V}'\begin{bmatrix}\boldsymbol{\Sigma}_m^2 & \boldsymbol{O}_{(n-m)\times(n-m)}\\ \boldsymbol{O}_{(n-m)\times(n-m)} & \boldsymbol{O}_{(n-m)\times(n-m)}\end{bmatrix}\boldsymbol{V}'^H\\&=2\boldsymbol{V}'\boldsymbol{\Sigma}'^H\boldsymbol{\Sigma}'\boldsymbol{V}'^H=\boldsymbol{V}'\boldsymbol{\Sigma}'^H(2\boldsymbol{I}_{2m})\boldsymbol{\Sigma}'\boldsymbol{V}'^H\\&=\left(\boldsymbol{U}'\boldsymbol{\Sigma}'\boldsymbol{V}'^H\right)^H\left(\boldsymbol{U}'\boldsymbol{\Sigma}'\boldsymbol{V}'^H\right)\end{aligned}\tag{28}$$

Thus, with the SVD of $\boldsymbol{Q}=\boldsymbol{U}'\boldsymbol{\Sigma}\boldsymbol{V}'^H$, the null space basis of $\boldsymbol{Q}^H$ can be constructed.

$$\begin{aligned}Null\left(\boldsymbol{Q}^H\right)&=Span\left\{\boldsymbol{u}'_{m+1},\boldsymbol{u}'_{m+2},\ldots,\boldsymbol{u}'_{2m}\right\}\\&=\begin{bmatrix}-\boldsymbol{U}\\ \boldsymbol{U}\end{bmatrix}=\begin{bmatrix}\boldsymbol{N}_1\\ \boldsymbol{N}_2\end{bmatrix}\in\boldsymbol{R}^{m\times 2m}\end{aligned}\tag{29}$$

where $\boldsymbol{u}'_{m+i}$ denotes the column vector of $\boldsymbol{U}'$.

As the randomness of $\boldsymbol{P}$, dim [Null($\boldsymbol{Q}^H$)]=m, namely the NSB matrix $\boldsymbol{N}=\begin{bmatrix}\boldsymbol{N}_1\\ \boldsymbol{N}_2\end{bmatrix}$ is a full column rank matrix. By elementary row operations, $\boldsymbol{N}$ can be converted to the form $\begin{bmatrix}\boldsymbol{N}_1\\ \boldsymbol{O}\end{bmatrix}$ without rank changing. Therefore, rank ($\boldsymbol{N}_1$)=m, which means $\boldsymbol{N}_i$, i=1, 2 are full rank matrices and possess inverse matrices. This completes the proof.

Abbreviations

AMR: Adaptive multi-rate; AWGN: Additive white Gaussian noise; BC: Before Christ; CoS: Compressed signal; CS: Compressed sensing; ES: Encrypted signal; K–L: Karhunen–Loeve; LPF: Low-pass filtering; MELP: Mixed excitation linear prediction; MOS: Mean opinion score; NSB: Null space basis; OMP: Orthogonal matching pursuit; PCM: Pause code modulation; SNR: Signal-to-noise ratio; SNRseg: Average-subsection signal-to-noise ratio; SVD: Singular value decomposition; TDS: Time-domain scrambling.

Competing interests

The authors declare that they have no competing interests.

Acknowledgments

This study was supported by the National Natural Science Foundation, China (61072042), the Natural Science Foundation of Jiangsu Province, China (BK2012510), and the Pre-research Foundations of PLA University of Science and Technology (20110211). The authors appreciate Professor Shou-sheng Liu and Professor Zu-ping Qian for their useful discussions and valuable suggestions from the bottom of our hearts. The authors would like to thank the anonymous reviewers for their constructive comments and questions which greatly improved the article.

Author details

[1]College of Command Information Systems, PLA University of Science and Technology, Nanjing, China. [2]College of Communications Engineering, PLA University of Science and Technology, Nanjing, China.

Received: 15 December 2011 Accepted: 4 December 2012
Published: 28 December 2012

References

1. JA Clark, Nature-inspired cryptography: Past, Present and Future, in *Congress on Evolutionary Computation*, ed. by 3rd edn. (Newport Beach, USA, 2003), pp. 1647–1654
2. L Nan, S Yanhong, Z Jiancheng, An audio scrambling method based on Fibonacci transformation. J. North China Univ. Technol. **16**(3), 8–11 (2004)
3. V Senk, VD Delic, VS Milosevic, A new speech scrambling concept based on Hadamard matrices. IEEE Signal Process. Lett. **4**(6), 161–163 (1997)
4. SK Pal, Fast, reliable & secure digital communication using Hardmard matrices, in *Proceedings of the International Conference on Computing: Theory and Applications*, ed. by 1st edn. (Kolkata, India, 2007), pp. 526–532
5. H Li, Z Qin, XP Zhang, LP Shao, An *n*-dimensional space audio scrambling algorithm based on random matrix. J. Xi'an Jiaotong Univ. **44**(4), 13–17 (2010)
6. M Satti, S Kak, Multilevel indexed quasi-group encryption for data and speech. IEEE Trans. Broadcast. **55**(2), 270–281 (2009)
7. H Li, Z Qin, Audio scrambling algorithm based on variable dimension spaces, in *International Conference on Industrial and Information Systems*, ed. by 1st edn. (West Bengal, India, 2009), pp. 316–319
8. W Honggang, H Michael, S Hamid, DM Peng, W Wang, C Hsiao-Hwa, Index-based selective audio encryption for wireless multimedia sensor networks. IEEE Trans Multimed **12**(3), 215–223 (2010)

9. S Antonio, M Juan Carlos, Perception-based partial encryption of compressed speech. IEEE Trans. Speech Audio Process **10**(8), 637–643 (2002)
10. M Pierre, JA O'Sullivan, Information-theoretic analysis of information hiding. IEEE Trans. Inf. Theory. **49**(3), 563–593 (2003)
11. H Li-Lian, Y Qi-tian, A chaos synchronization secure communication system based on output control. J. Electron. Inf. Technol. **31**(10), 2402–2405 (2009)
12. T Liangrui, Z Lin, Y Xue, Chaos synchronization based on observer and its application in speech secure communication, in *Proceedings of IC-NIDC*, ed. by 2nd edn. (Beijing, China, 2010), pp. 773–777
13. E Del Re, R Fantacci, D Maffucci, A new speech signal scrambling method for secure communications: theory, implementation, and security evaluation. IEEE J. Sel. Areas Commun. **7**(4), 474–480 (1989)
14. CE Shannon, Communication theory of secret systems. Bell Syst. Tech. J. **28**(4), 656–715 (1949)
15. DL Donoho, Compressed sensing. IEEE Trans. Inf. Theory. **52**(4), 1289–1306 (2006)
16. RG Baraniuk, Lecture notes: compressive sensing. IEEE Signal Process. Mag. **24**(4), 118–121 (2007)
17. EJ Candès, MB Wakin, An introduction to compressive sampling. IEEE Signal Process. Mag. **25**(2), 21–30 (2008)
18. W Tian-jing, Z Bao-yu, Y Zhen, A speech signal sparse representation algorithm based on adaptive overcomplete dictionary. J. Electron. Inf. Technol. **33**(10), 2372–2377 (2011)
19. C Zhang-hua, T Yuan-sheng, Secure communication based on network coding. J. Commun. **31**(8A), 188–194 (2010)
20. ZM Smith, D Bertrand, AJ Oxenham, Chimaeric sounds reveal dichotomies in auditory perception. Nature. **416**, 87–90 (2002)
21. S Mehmet Cenk, G Bilge, K Gunes Karabulut, Perceptual audio features for emotional detection. EURASIP J. Audio Speech Music Process **16**, 1–21 (2012)
22. JA Tropp, AC Gilbert, Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans. Inf. Theory. **53**(12), 4655–4666 (2007)
23. J Navarro-Moreno, JC Ruiz-Molina, Nonlinear estimation using correlation information. IEEE Trans. Signal Process. **54**(7), 2822–2827 (2006)
24. K Saberi, DR Perrot, Cognitive restoration of reversed speech. Nature **398**, 760 (1999)
25. PX Joris, TC Yin, Envelope coding in the lateral superior olive. I. Sensitivity to interaural time differences. J. Neurophysiol. **73**, 1043–1062 (1995)
26. D Bendor, X Wang, The neuronal representation of pitch in primate auditory cortex. Nature. **436**, 1161–1165 (2005)
27. Z Xian-da, *Matrix Analysis and Applications* (Tsinghua University Press, Beijing, 2004)
28. P Itu-T, *862, Perceptual Evaluation of Speech Quality (PESQ), and Objective Method for End-toEnd Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs* (ITU-T Recommendation, Geneva, 2001)
29. P Georgiev, F Theis, A Cichocki, Sparse component analysis and blind source separation of underdetermined mixtures. IEEE Trans. Neural Netw. **16**(4), 992–996 (2005)
30. X Tingting, Y Zhen, S Xi, Novel speech secure communication system based on information hiding and compressed sensing, in *The 4th International Conference on System and Networks Communications*, ed. by, 4th edn. 2009, pp. 201–206
31. DL Donoho, E Michael, T Vladimir, Stable recovery of sparse overcomplete representations in the presence of noise. IEEE Trans. Inf. Theory. **52**(1), 6–18 (2006)
32. M Babaie-Zadeh, C Jutten, On the stable recovery of the sparsest overcomplete representations in presence of noise. IEEE Trans. Signal Process. **58**(10), 5396–5400 (2010)
33. JA Tropp, Just relax: convex programming methods for identifying sparse signal in noise. IEEE Trans. Inf. Theory. **52**(3), 1030–1051 (2006)
34. Q Sun, Sparse approximation property and stable recovery of sparse signal from noisy measurements. IEEE Trans. Signal Process. **59**(10), 5086–5090 (2011)

doi:10.1186/1687-6180-2012-257
Cite this article as: Zeng *et al.*: **Scrambling-based speech encryption via compressed sensing.** *EURASIP Journal on Advances in Signal Processing* 2012 **2012**:257.