View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by International Institute for Science, Technology and Education (IISTE): E-Journals

# A Review of Analog Audio Scrambling Methods for Residual Intelligibility

A.Srinivasan[1] P.Arul Selvan[2*]

1. Dept of Information Technology, MNM Jain Engineering College, Chennai – 600097, Tamil Nadu, India. Email: asrini30@gmail.com
2. Dept of Electronics and Communication Engineering, Sathyabama University, Chennai, Tamil Nadu, India

* E-mail of the corresponding author: arulp6874@gmail.com

## Abstract

In this paper, a review of the techniques available in different categories of audio scrambling schemes is done with respect to Residual Intelligibility. According to Shannon's secure communication theory, for the residual intelligibility to be zero the scrambled signal must represent a white signal. Thus the scrambling scheme that has zero residual intelligibility is said to be highly secure. Many analog audio scrambling algorithms that aim to achieve lower levels of residual intelligibility are available. In this paper a review of all the existing analog audio scrambling algorithms proposed so far and their properties and limitations has been presented. The aim of this paper is to provide an insight for evaluating various analog audio scrambling schemes available up-to-date. The review shows that the algorithms have their strengths and weaknesses and there is no algorithm that satisfies all the factors to the maximum extent.

**Keywords:** residual Intelligibility, audio scrambling, speech scrambling

## 1. Introduction

In communication systems, audio which includes speech and music signifies either analog or digital audio. The transmission of digital audio of good quality requires a channel bandwidth (up to 32kbps) that is greater than the channel bandwidth needed for analog audio (up to 4 KHz). Scrambling of digital audio results in a signal whose characteristics is similar to white noise. Hence it has zero residual intelligibility with high cryptanalytic strength, but this scrambled digital audio signal needs a higher channel bandwidth for transmission. Another class of analog scrambling operates on the digital codes of pulse code modulation (PCM), adaptive differential pulse code modulation (ADPCM) and delta modulation (DM). In this case, the scrambled bits are converted into analog form for transmission over analog channels. This is a kind of nonlinear transformation which results in poor recovered speech quality; hence it has lesser practical usage [S.C.Kak et al 1983]. Scrambling of the analog audio reduces the residual intelligibility, but the signal has lesser cryptanalytic strength. Moreover, the signal bandwidth is kept at a comparatively low level, so that transmissions through analog channels are feasible.

The key factors that characterize the scrambling algorithm are Residual Intelligibility, Encoding Delay and Key-Space. This paper reviews the available analog audio scrambling algorithms for the above mentioned factors. The auxiliary factors Bandwidth Expansion and Cryptanalytic Strength are also considered in the review.

The paper is organized as follows. In section 2, the main factors pertaining to analog audio scrambling algorithms are summarized. In section 3, the algorithms are categorized based on the methodology used. Next in section 4, the algorithms have been discussed for the three key factors with the tabulation of results, merits-demerits and future work. The paper concludes with final remarks.

## 2. Factors of Analog Audio Scrambling algorithms

In an analog scrambler, the analog signal is first converted into a discrete signal and then processed for scrambling using digital processing techniques; finally the scrambled signal is again converted back to analog signal. Since, the scrambler output is an analog signal; the scrambling scheme is termed as analog scrambling. Analog scrambling is the preferred method for secure speech communications over the telephone channel. Moreover, analog scrambling can only provide good privacy in the context of casual eavesdropping. For high security applications digital encryption has to be used.

### *2.1 Residual Intelligibility*

The amount of redundant information in the scrambled signal is termed as residual intelligibility, which helps in easier recovery of the original information. Scrambling effectiveness is determined by the amount of residual intelligibility, key-space, rate of change of the key and the distortion produced by the key. The above factors are linked to the complexity of the system and the resultant encoding delay. Thus for low intelligibility levels and high key-space the scrambling effectiveness is higher, but the system complexity and the encoding delay increases [S.C.Kak et al 1977].

Intelligibility is a subjective quantity that is evaluated by using trained and untrained human listeners to listen to the scrambled audio. Intelligibility is commonly expressed in terms of word, sentence and digit intelligibility levels. Test materials used for word intelligibility testing include a list of monosyllabic words, for sentence intelligibility its a combination of monosyllabic words and for digit intelligibility recordings of N-Digit numbers are used. In the redundancy scale, sentences have the highest redundancy followed by words and digits have the lowest amount of redundancy. In most cases the analog scrambler performance is benchmarked with respect to the digit intelligibility because of limited vocabulary and lesser redundancy of the digits. Intelligibility scores are given in the range of 100-0 percent, with zero percent being the ideal value of zero residual intelligibility which will resemble white noise. A level of 10% is termed as the lower threshold, 30% is the medium level and 50% is the higher level [N.S.Jayant, R.V.Cox, B.J.McDermott, A.M.Quinn 1983].

### *2.2 Encoding Delay*

The amount of time taken per unit by the scrambling algorithm to complete the scrambling operation is termed as encoding delay; in general the unit is taken as block or segment. The encoding delay is directly proportional to the number (N) of units, length (L) of each unit and the number of samples (S) present in one unit. When N,Land S increases the recovered speech quality increases because of the availability of more number of permutable samples, but the encoding delay also increases. Considering the two complementary factors of encoding delay and recovered speech quality an appropriate segment length chosen is between 16 to 32 ms or 256 samples per frame [N.S.Jayant 1982].

### *2.3 Key-space*

The procedure used for transforming the signal is commonly called as Key. The level of security offered by an analog scrambling algorithm is a complex function of the number of usable keys called as key-space, length, rate of change of the key, properly selected limited key dictionary, proper time variation and distribution of the keys.[N.S.Jayant 1982]. For casual privacy the key is independent of time and for high security the key is time dependent. Similarly high security needs a larger key-space, but, when the key-space is larger or the key is time dependent the system complexity increases, thereby increasing the encoding delay. Within a given key-space, the keys selected have to be statistically independent for increased security. Moreover, the keys selected have to distort the scrambled signal to a larger extent.

### *2.4 Bandwidth Expansion*

When the speech signal is scrambled, discontinuities are introduced in the scrambled signal, which results in an increase in the scrambled signal bandwidth. For higher scrambling effectiveness, larger amount of discontinuities are introduced, which in turn increases the bandwidth. Bandwidth expansion limits the capability of the scrambled signal to be transmitted through narrow-band channels. In general the time-frequency permutation introduces bandwidth expansion. To keep the bandwidth expansion minimal, linear orthogonal invertible transformations can be used. In this review paper, bandwidth denotes analog bandwidth.

## 3. Analog Audio Scrambling Classification

### *3.1 Analog Audio Scrambling Classification-First level*

Scrambling of the audio signal can be done in analog and digital domain. This section captures the classification of the analog audio scrambling techniques.

The taxonomy of analog audio scrambling algorithms is depicted in Figure 1.

In the Sample Amplitude based technique, the amplitude of the analog audio samples is altered with a simple reordering in the time domain, resulting in change of the magnitude spectrum of the scrambled signal. In the Time Domain based technique, the samples are grouped together into segments and these segments are then reordered. In the Frequency Domain based technique, the frequency contents of the segments are extracted as sub-bands and these sub-bands are permuted thereby altering the frequency spectrum. When scrambling is done in both the time and frequency domains, it is called as two-dimensional technique. The audio signal is transformed using an appropriate transformation technique and the transform coefficients are permuted to produce the Transformation based audio scrambling.

### *3.2 Analog Audio Scrambling Classification-Second level*

This section sub-classifies the above techniques based on the analog scrambling techniques available in the literature up-to-date.

#### 3.2.1 Sample Amplitude based techniques

In the sample amplitude based technique, the amplitude samples of the original signals are taken up for scrambling. Typical operations include interchange or permutation of speech samples [J.Phillips, M.H.Lee, J.E.Thomas 1971], linear addition of pseudorandom noise amplitudes and non-linear modulo-arithmetic additions [S.C.Kak et al 1977]. Two basic types of permutations available are Uniform (U) permutations and Shift-Register generated Pseudo-Random (PR) permutations. Some types of scramblers involve addition of masking signals to the amplitude samples, these masking signals can be a PR binary or modulo-arithmetic sequence.

#### 3.2.2 Time Domain based techniques

In the time domain based technique, the audio signal is divided into segments and the segments are then permuted. Main time domain techniques are Time-Inversion, Time Segment Permutation (TSP), Hopping-Window and Sliding Window TSP, Time Shifting of Speech Sub-bands, Reverberation [N.S.Jayant 1982] and time-domain based scrambler which does not need synchronization [F. Huang, E. V. Stansfield 1983].

#### 3.2.3 Frequency Domain based techniques

In this class of scramblers the speech signal spectrum is divided into many sub-bands and the position of these sub-bands are then permuted. Main frequency domain techniques are Frequency Inversion, Band-splitting, Band-splitting with Frequency Inversion and Frequency Inversion followed by Cyclic Band-shift [N.S.Jayant 1982].

#### 3.2.4 Two-Dimension based techniques

Two-Dimensional Scramblers perform manipulations in both the time and frequency domains simultaneously. Important types of scramblers are Frequency Inversion combined with Block TSP, Frequency Inversion and Cyclic Band-shift combined with time manipulations and Time-Frequency Segment Permutation (TFSP).

#### 3.2.5 Transform based techniques

This class of analog scramblers is based on operations performed on the linear transform coefficients of the audio samples. Types of transforms used are Discrete Prolate Spheroidal Transform (DPST), Fast Fourier Transform (FFT), Discrete Cosine Transform (DCT), Modified discrete cosine transform (MDCT), Hadamard Transform (HT), Circulant transformation, Wavelet Transform, parallel structure of two different types of wavelets with the same decomposition levels, combination of QAM mapping method and an orthogonal frequency division multiplexing (OFDM).

## 4. Review of the techniques

### *4.1 Review on Sample-Amplitude based techniques*

Sample interchange method is the simplest technique where the individual samples are reordered. The reordering can be achieved by using delay networks as shown in the Figure 2. The figure given shows the scrambling order for a block-4 sequence. This reordering produces sideband components that mask or alters the amplitude of the adjacent audio samples. But, this method still retains a substantial amount of residual intelligibility; the word intelligibility is about 22% and the digit intelligibility is about 24% for a block size of 128 samples. When the sample displacement is larger the sideband components are stronger, this increases the effect of masking. A variation of this method is to have a sample-sequence reordering which is closer to the completely reverse sequence, in this case a word intelligibility level of 2% is obtained. A second variation is to have a more complex sample interchange that takes place between samples of different segments, this is termed as 'running exchange' and it gives much lower value of residual intelligibility. The two variations given above leads to a scrambled signal bandwidth that exceeds the analog channel bandwidth of 4KHz[J.Phillips, M.H.Lee, J.E.Thomas 1971].

Majority of the scrambling schemes involves permutation of the samples or transform coefficients. The permutation must result in a marked difference between the original and scrambled blocks both perceptually and spectrally. The effectiveness of permutation is measured in terms of the rank correlation coefficient which ranges from 0-1 with zero denoting a highly effective permutation. Spearman's coefficient and Kendall's coefficient are the two most frequently used methods for determining the rank correlation.[S.C.Kak et al 1983]

The class of scramblers based on temporal permutation of the speech samples has efficiency which is dependent on the order of the permutation matrix and the randomness of the matrix coefficients. Two types of permutation possible are U-permutations and PR-Permutations. U-permutations results in the frequency spectrum of the scrambled signal that is flat in an average sense. Whereas, for PR-Permutation the frequency spectrum of the scrambled signal is flat in the average sense and also the transitions of the adjacent samples produces a smoother spectrum. This flatness ensures a decrease of residual intelligibility. Since the speech-silence pattern is recognizable, the residual intelligibility of both the technique is considered to be essentially higher. [S.C.Kak et al 1977]

Contiguous time-sample permutation has an important limitation of bandwidth expansion of the scrambled signal. To overcome this, individual samples are grouped together into time-segments on which the U or PR permutations can be applied. Typical segment duration chosen is 10-30ms. [S.C.Kak et al 1977]. A common issue that needs to be taken care of for a sample/segment based scrambler is synchronization of the frames between the scrambler and descrambler.

Masking based scramblers are an alternative to permutation based scramblers. Main types of masking techniques include linear addition of PR noise or modulo-m addition to the samples/segments; these techniques provide a low level of residual intelligibility. The masking signal has to be slow changing with respect to the audio waveform, this ensures that the entire audio sample is impacted by the masking signal and the spectrum is closer to the white signal spectrum. The frequency spectrum after scrambling is shown in Figure 3. [S.C.Kak et al 1977]. Moreover the speech-to-channel noise ratio is lower in the case of linear masking thereby resulting in higher receiver complexity. Non-linear masking techniques are more robust to real-channel imperfections, but it leads to bandwidth expansion [N.S.Jayant 1982]. A significant advantage of the masking techniques is the removal of speech-silence patterns.

Technique based on chaotic encryption in conjunction with lookup tables is discussed in [K.Ganesan, R.Muthukumar, K.Murali 2006]. The lookup tables are constructed by using an appropriate chaotic system (like Arnold map), the entries in this table include index number and iterated decimal value. The amplitude values of the quantized audio samples are converted based on the lookup table entry. The input quantized audio data that varies between 0 and 19512 is converted to the amplitude values that vary between 0 and 65284. Thus the randomized amplitude value is generated with a higher dynamic range; this ensures a lower level of residual intelligibility.

### 4.1.1 Experimental Overview

The Table 1 given below lists the comparative values of the factors for the various algorithms.

The sample interchange method leaves a considerable amount of residual intelligibility in the scrambled speech, because the interchange happens within a finite distance. Since finite numbers of samples are taken up for interchange, key-space and encoding delay are both lower. Improvement in residual intelligibility and increase in key-space value is obtained when the number of samples taken up for applying this method is larger, but this will increase the encoding delay. In comparing the PR and U permutation scramblers for a given value of block length N, its found that both the techniques have a relatively higher level of residual intelligibility. The presence of speech-silence pattern increases the residual intelligibility. As the value of N is increased, the key-space increases. For a value of N=256, PR permutation has a key-space of 4080 and U permutation has a key-space of 63232. In permutation based algorithms the segment duration has to be kept between 10-30 ms for limiting the bandwidth expansion [S.C.Kak et al 1977].

Masking and permutation techniques can be applied concurrently on the speech samples to improve the residual intelligibility levels. The coefficients of the permutation matrices can be time-varying to increase the crypt-analytic strength. It is also important to note that the Hamming Distance can be used as a measure of scrambling for permutation based scramblers. When more elements are moved from their original place because of permutation the Hamming Distance is larger and the residual intelligibility will be lesser.

4.1.2 Merits and Demerits

The encoding delay is low because the scrambling is done on a finite set of samples at a time. The bandwidth expansion is minimal when the segment duration is kept between 10-30ms.

The various algorithms in this category retain a significant amount of residual intelligibility. The presence of speech-silence pattern in the techniques other than masking decreases the cryptanalytic strength. The key-space available is low because these algorithms work on a subset of speech samples.

4.1.3 Future Work

Sample interchange in a random manner is theoretically possible, but it has practical difficulties that needs to be explored.[J.Phillips, M.H.Lee, J.E.Thomas 1971] Permutation matrix coefficients can be generated by following a look-up-table approach.

The possibility of using higher dimensional chaotic system for better scrambling results is yet to be explored.[K.Ganesan, R.Muthukumar, K.Murali 2006]

*4.2. Review on time domain based techniques*

In this class of scramblers the speech segments of length 10-30ms is taken up for permutation, because this will result in a bandwidth-preserving operation [S.C.Kak et al 1977]. The basic unit taken up for scrambling is a block of samples or segments; variation in the scrambling technique depends on the operation that is performed on the blocks. The block-wise operation introduces a time-delay which is directly proportional to the block size. Permutation based scramblers do not change the characteristics like frequency, phase and amplitude of the speech components, but the time or frequency order of the components are only changed. The coordinates of 1's in the permutation matrix defines the scrambling key, thus for an NxN matrix the key-space is of N! keys. In this key-space, only 10-20% of the keys provide low residual intelligibility, hence key selection is an important factor in this type of scrambler. The main advantage of permutation based scramblers is that it does not increase the signal bandwidth [D. B.Sadkhan, D. Abdulmuhsen, N. F.Al-Tahan 2007].

The scrambling efficiency (S) is given by the function

$S = F(B,N,D)$

Where B = block size, N = number of samples or Segments in a block and D = the temporal distance of segment separation.

The average segment separation and residual intelligibility are monotonically related. For permutation scramblers given a constant B and N, the scrambling efficiency is directly proportional to the temporal distance D. Temporal distance is the time-distance between a pair of segments in the original speech that appears as adjacent segments in the scrambled speech. Its important to note that B and D are related as given below

$\mathrm{Max}(D) = 2B - 1(\mathrm{InSegments})$

For a block size of 8, the maximum temporal distance achievable is 15 segments[N.S.Jayant, R.V.Cox, B.J.McDermott, A.M.Quinn 1983].

In the Time-Inversion technique, block length of the order of 128ms or 256 ms is chosen. The order of the speech samples are inverted within the block. This resulted in reduction of residual intelligibility, but the level is still comparatively higher. The total encoding delay is of the order of 256ms or 512ms. Time-inversion is a deterministic operation; hence the cryptanalytic strength is lesser.

In the TSP technique, segments of speech are permuted and transmitted in a pseudo-random fashion by following a segment-mapping algorithm. Two types of TSP techniques available are block and sequential technique. In the Block

TSP, all the Segments in a given block are scrambled and transmitted before the segments of the next block are brought into the scrambler memory. But in a sequential TSP, individual segments are transmitted instead of waiting for the block. Optimal segment duration in both the cases is about 16-32ms [N.S.Jayant, R.V.Cox, B.J.McDermott, A.M.Quinn 1983]. In a block TSP Scrambler known as Hopping-Window TSP Scrambler, segments of 16ms duration in blocks of b segments are used. Thus the number of permutations available is b!, but only 0.1% of the permutations are good from a scrambling point of view. The sequential TSP scrambler known as Sliding-window TSP scrambler has a memory to store b segments; the segment that is outputted is determined by a pseudo-random selector. The maximum staying time in the memory permissible for a segment is t=2b which ensures an optimal residual intelligibility. This staying time is termed as communication delay. TSP based scramblers gives a higher residual intelligibility level of the order of 80-100% for an communication delay of 256ms, for larger communication delay (512ms) the intelligibility level of block TSP improves to 60%. Using mu-law compression of speech, the residual intelligibility level can be improved up to 45%. In the TSP technique the need to synchronize scrambler-descrambler is the main disadvantage [N.S.Jayant 1982].

In the Time Shifting of Speech Sub-bands technique, different time segments of speech are differentially delayed. Normally the time segment corresponding to the lower frequency signal is delayed by time interval τ and added to the time segment corresponding to the higher frequency signal for transmission. The reverse happens at the descrambler and the total encoding delay introduced is τ. Scramblers in this category provide a better residual intelligibility level, but with higher encoding delay.[N.S.Jayant 1982]

In the Reverberation technique, multiple number of time-discrete echoes with fixed interval are mixed with the current speech amplitude to generate the scrambled output. In the Forward type technique the echoes decreases exponentially and in the Reverse type the echoes increases exponentially. These schemes have higher value of encoding delay and a lower residual intelligibility [N.S.Jayant 1982].

Time-domain based scrambler which does not need synchronization uses a time varying transversal filter, where the incoming time-samples are selected randomly and multiplied by constant values. Conversely in frequency domain this is equivalent to having narrow band filters that have different center frequencies, each of the filter passes a given input frequency sub-band whose center frequency is then shifted. The amount of frequency shift is controlled by either constant key or variable key, which results in scrambling of the input frequency sub-bands. For most of the keys the speech signal is not intelligible, but for a subset of the keys the scrambled spectrum shows perfect symmetry for certain sub-bands, which results in the presence of sufficient intelligibility [F. Huang, E. V. Stansfield 1983].

In a speech scrambling algorithm based on blind source separation, unknown and mutually independent source signals which are in the form of mixtures are used. The algorithm proposed combines the time element scrambling and masking methods, wherein segments of speech signal are mixed with equal number pseudorandom key signals. The process of mixing reduces the number of the segments; hence decryption without knowing the key signals will not be possible. A significant aspect of this algorithm is that the speech segments are taken up together for mixing with the key signals thereby rendering more complete scrambling, hence keeping the residual intelligibility lower[Q.H. Lin, F.L. Yin, T.M. Mei, H.L. Liang 2004].

### 4.2.1 Experimental Overview

The Table 2 given below lists the comparative values of the factors for the various algorithms. The Time-Inversion method is applied on a segment of speech samples and hence the scrambled speech retains significant amount of residual intelligibility. To have a lower the residual intelligibility level the segment size is made larger. In the TSP scrambling method, scrambling takes place at the segment level hence the residual intelligibility is reduced, but not by a significant amount. By performing scrambling of the speech samples within the segment together with the segment level scrambling, the residual intelligibility can be lowered. In the reverberation method, the residual intelligibility is controlled by the number of the past speech samples that impacts the present speech sample; when this number is higher the residual intelligibility becomes lower.

### 4.2.2 Merits and Demerits

The bandwidth expansion is low. Time-Domain operations remove the speech-silence rhythm that is present in the scrambled signal. Time-Domain based scramblers are robust to real channel imperfections, wherein the overall nature of the speech signal is intact for segment length within 16ms. Hence synchronization is essential for segment lengths more than 16ms.

The various algorithms in this category have a comparatively higher value of residual intelligibility and the digit intelligibility is of the order of 60%. For lower values of residual intelligibility the encoding delay will be above 512ms. The key-space is low because these algorithms work on a subset of speech samples.

### 4.2.3 Future Work

The effect of loss of synchronization between the transmitter and the receiver on the intelligibility of the unscrambled speech needs to be examined.

## *4.3. Review of frequency domain based techniques*

In frequency domain based scramblers, the frequency sub-bands of the audio signal are divided into segments and scrambling of these segments is performed. The Frequency Inversion as shown in Figure 4 is a technique based on one reference frequency which is termed as the key. Though this technique provides a residual intelligibility level of 30%, the characteristics of the scrambled speech are identifiable; hence it has the least crypt-analytic strength. A marginal variation of this technique is the frequency hopping inversion which involves a varying reference frequency; here the residual intelligibility is only slightly better. This technique offers a digit intelligibility score of 30% when untrained listeners are used.[S.C.Kak et al 1977].

Band-splitting technique which involves permuting the frequency sub-bands offers a better residual intelligibility. With f sub-bands, the total number of permutations available is f!. When the correct position of one or two main sub-bands are found out, then information of the phonemes can be recognized, hence the crypt-analytic strength is lower.[N.S.Jayant 1982]

With Band-splitting and Frequency Inversion technique, specific frequency sub-bands are subjected to frequency inversion. With f sub-bands the total number of scrambler mappings possible is of the order of f!2f of which only 5% of the mappings are effective. The word intelligibility level using this technique is of the order of 45 to 70% with trained listeners.[N.S.Jayant 1982]

In the Frequency Inversion followed by Cyclic Band-shift technique each sub-band is shifted by the factor n (modulo k). For the case of 16 sub-bands and the shift variation rate of 50 per second, the residual intelligibility level is of the order of 55% for digits and 30% for words. [N.S.Jayant 1982]

### 4.3.1 Experimental Overview

The Table 3 given below lists the comparative values of the factors for the various algorithms.

### 4.3.2 Merits and Demerits

The encoding delay and bandwidth expansion are low. These classes of scramblers do not need to have synchronization between the transmitter and receiver for segment length upto 200Hz. The spectral characteristics of the individual phonemes are altered which increases the security. The various algorithms discussed in this category have higher levels of residual intelligibility and the presence of speech-silence rhythm. The key-space is low because these algorithms work on a subset of speech samples. Crypt-analytic strength is lesser because identification of certain frequency components gives information that leads to deciphering the information of the remaining content. A common disadvantage of all frequency domain scramblers is the effect of group delay distortion in the transmission channel.

### 4.3.3 Future Work

Specification of the spectral distortions that provide optimal speech scrambling needs to be done. Cascading of multiple stages of the techniques to realize better residual intelligibility levels can be done.

## *4.4. Review on two dimensional scrambling techniques*

Two-Dimensional Scramblers operates on time-segments of 16ms duration which are subsequently partitioned into frequency sub-bands, manipulations of both the time and frequency domain components are done simultaneously. The Time-domain manipulations destroy the speech-silence rhythm and the frequency-domain manipulations alter the spectral characteristics of some of the audio components, thereby cumulatively reducing the residual intelligibility up to 15-25%. These types of scramblers come with increased complexity, encoding delay and sensitivity to channel imperfections.[N.S.Jayant, R.V.Cox, B.J.McDermott, A.M.Quinn 1983]

Frequency Inversion combined with Block TSP technique produces digit intelligibility of the order of 20% for a block size of 256 samples. The scrambler is operated with a delay of 1024ms. Dynamic cyclic band-shift schemes that are used for analog scramblers have two variations. First type of scrambler system uses a type of time-manipulation called dynamic time reverberation. This system produces digit intelligibility of the order of 18-28% and word intelligibility close to zero. In this system the preferred order is frequency scrambling followed by time scrambling. Second type of system uses time-shifting between two frequency sub-bands, this system produces digit intelligibility of the order of 25-38% and word intelligibility of 2-3%. In this system the preferred order is time scrambling followed by frequency scrambling. The above two types of systems operate well for channels that introduce heavy signal distortion and fading. [N.S.Jayant 1982]

In the TFSP based scrambling system shown in the Figure 5, f frequency sub-bands in each of the b time-segments are collected to form fb time-frequency segments. These time-frequency segments are outputted randomly either sequentially or in blocks from the scrambling system memory. The maximum memory retention time of one segment being t = 2fb segment-durations, this retention time denotes the encoding delay. The average digit intelligibility of a TFSP scrambler for 256ms encoding delay is about 25 percent. The word intelligibility score is close to zero [N.S.Jayant 1982]. Two problems that need to be addressed in the TFSP scrambler are synchronization and recovered speech quality. Synchronization is established by sending signaling chirps from the transmitter. Channel equalization is done to increase the recovered speech quality. [R.V.Cox, T.M.Tribolet 1983].

### 4.4.1 Experimental Overview

The Table 4 given below lists the comparative values of the factors for the various algorithms. In most algorithms the residual intelligibility is lower with typical level of 30%. In all the algorithms the usable key-space is very low, encoding delay is moderate with typical value of 256 ms, bandwidth expansion is low.

### 4.4.2 Merits and Demerits

The various algorithms in this category have low levels of residual intelligibility with digit intelligibility of the order of 20%. The presence of speech-silence rhythm is removed. The bandwidth expansion is also low. These classes of scramblers do not need to have synchronization between the transmitter and receiver. These scramblers are robust to transmission channel characteristics with problems only at the spectral and temporal segment boundaries, thus the loss of speech quality is lesser [N.S.Jayant 1982].

The key-space is low because these algorithms work on a subset of time and frequency segments. Crypt-analytic strength is lesser, because identification of certain frequency components gives information that leads to deciphering the information of the remaining content. The encoding delay is moderate because these types of scramblers involve both time and frequency domain manipulations.

### 4.4.3 Future Work

There is scope for devising appropriate techniques to bring the digit intelligibility value closer to the lower bound of 10% and word intelligibility value of zero percent. Problems due to the channel characteristics on the spectral and temporal boundaries need to be addressed.

## *4.5.Review on Transform domain based techniques*

The class of analog audio scramblers based on operations performed on the linear transform coefficients of the speech samples is known as transform based scrambler. The transform based scramblers have larger number of usable permutations which increases the cryptanalytic strength and offers very low levels of residual intelligibility.

A speech sample block is first converted into transform coefficient blocks (F). These transform coefficient blocks are scrambled based on operations like permutation or non-linear modulo-arithmetic masking (P). The scrambled speech blocks are generated by performing inverse transformation operation (I). The reverse of this is done at the receiver. The process for transform domain scrambling is shown in Figure 6.

The transformations to be used in these class of algorithms has to be linear orthogonal type, the reason being that it will not increase the level of the noise component in the scrambled sequence. For example, consider F as the transformation and x as the input sequence, the transformed sequence is given by Fx, when the noise gets added the transformed sequence becomes Y = Fx+n. When inverse transformation is applied on Y then $F^{-1}Y = F^{-1}(Fx+n)= x+ F^{-1}n$, hence when $F^{-1}$ is orthogonal $F^{-1}n = n$ and the noise component can be easily filtered out, thereby preserving the sequence energy.

A simple scrambling scheme is the permutation of the coefficients of the transform sequence, where a band-limited input sequence results in a band-limited scrambled sequence at the output. The Discrete Prolate Spheroidal Transform (DPST) is a type of linear orthogonal transform that is used for this purpose. The crypt-analytic strength of this scheme is much higher compared to traditional analog scramblers. The residual intelligibility level is lower, but the limitation is high complexity and usage in narrowband channels only [A. D.Wyner 1979].

In the FFT based scrambling, the FFT coefficients selected are scrambled using a permutation matrix which is either stored in the ROM memory or generated instantaneously from a key value. FFT based scrambling expands the bandwidth of the scrambled signal, hence when the transmission is done on a band-limited channel the recovered speech quality is reduced. The recovered speech quality can be increased by having a large number of samples per FFT frame; typical frame lengths for this purpose are 128,256 and 512 samples. As the number of permutable FFT coefficients is higher the crypt-analytic strength is increased, but the encoding delay also increases considerably. A reasonable frame size with tradeoff between recovered speech quality and encoding delay is 256 samples per frame. An alternative to limit the bandwidth is to take up a subset of the FFT coefficients for scrambling, commonly 85 FFT coefficients corresponding to frequencies from 288 to 2976 Hz is taken for scrambling. To further increase the crypt-analytic strength a multi-frame structure where different permutation is used for each frame can be used. FFT based scheme is impacted by group delay distortion which is equalized using a digital transversal filter. Preservation of signal energy, talk spurts and original intonation decreases the security of the FFT based scheme. [K. Sakurai, K. Koga, T. Muratan 1984].

A scrambling scheme based on FFT coefficient permutation and adaptive dummy spectrum insertion is used to prevent the detection of the talk spurts. Dummy spectrum insertion introduces noise at the receiver; syllabic companding operation is used to reduce this noise. To enhance security, FFT coefficients of lesser energy are adaptively selected and replaced with dummy coefficients prior to permutation. The values of these dummy coefficients are selected so that the scrambled speech signal is of constant energy. This scheme is sensitive to channel impairments whereby the scrambled speech undergoes parabolic group delay distortion as shown in Figure 7, which induces high amounts of delay at the spectral boundaries. An equalizer is used to suppress this distortion [K. Sakurai, K. Koga, T. Muratan 1984].

DCT has good energy compaction property, hence DCT based scrambling systems are superior when compared to DFT and DPST. When the bandwidth limitation is taken into account the DCT based scrambler has 197 coefficients available for permutation in the band 300 to 3300 Hz. This results in a total of 197! possible permutations which increases the crypt-analytic strength and provides lower levels of residual intelligibility. These systems have lower encoding delay and better recovered speech quality. To prevent detection of talk spurts dummy transform coefficients are substituted for a predefined block of components in the original speech spectrum [S.Sridharan, E.Dawson, B.Goldburg 1993].

In the technique based on Modified discrete cosine transform (MDCT), the audio samples transformed by MDCT are sorted and packetized according to its importance by an index. A subset of the important packets is selectively scrambled and the rest of the packets are either discarded or left in its original form. The primary focus is ensuring high energy efficiency and it is seen that as the number of packets scrambled are increased, then the dissimilarity between the original and scrambled audio increases. [H. Wang, M. Hempel, D. Peng, W. Wang, H. Sharif, H.H. Chen 2010]

Speech scramblers based on Hadamard (H) matrices which are a linear transformation of speech components is an effective alternative to permutation based scrambler. Main advantage of this method is that the signal energy is distributed more uniformly over the scrambling frame hence making pattern matching impossible. Other advantages include no bandwidth expansion, lower residual intelligibility, larger key-space, lower encoding delay and simpler system implementation. The results for the listening test for sentence intelligibility for a frame length of 64ms indicate that for permutation scrambler 20% correct guess was obtained and for H-based scrambler correct guess was approximately zero percent [D. B.Sadkhan, D. Abdulmuhsen, N. F.Al-Tahan 2007]. A significant advantage is that the speech segments are both scrambled and altered in terms of amplitude, frequency and phase (This is because the entries of the H-Matrix is 1,-1) thereby giving lower values of residual intelligibility [V. Milosevic, V. Delic, V. Senk 1997].

When the speech signal is subjected to circulant transformation, phase distortion is introduced. This phase distortion redistributes the signal energy to the entire frame. When the frame length is higher the order of circulant matrix increases, thus the redistribution of energy covers more area thereby reducing the residual intelligibility to a lower value. Another property of this scheme is the distortion of the formant frequencies and introduction of new formants which significantly contributes in reducing the residual intelligibility. As the row values of the circulant matrix functions as the key, theoretically an infinite number of keys are possible. [G.Manjunath, G.V.Anand, 2002] .

Analog speech scrambler based on Wavelet Transform scrambles the speech signal in both time and frequency domains. As this resembles a two-dimensional scrambler very low levels of residual intelligibility is obtained. In this method the speech signal is converted into wavelet-analyzed signal by means of the filter bank which is based on wavelet basis. These wavelet signals are then multiplexed and collected as frames of constant length, scrambling involves permutation of these frames. The spectrum of the scrambled signal is highly irregular and the formant frequencies of the speech signal are hidden completely. Thus the scheme provides very low values of residual intelligibility. [F. Ma, J. Cheng, Y. Wang, 1996].

A technique based on parallel structure of two different types of wavelets with the same decomposition levels has been discussed in [D. B.Sadkhan, D. Abdulmuhsen, N. F.Al-Tahan 2007]. The combinations of wavelets used are Db1 wavelet along with Haar wavelet, Db2 wavelet along with Sym2 wavelet and Db4 wavelet along with Sym4 wavelet for the same level. The speech signal is divided into two sub-frames of equal size and the two sub-frames are applied to the parallel wavelet structure and the wavelet coefficients are generated. These coefficients are then suitably permuted. For a level 3 type of Haar wavelet the lowest value of Segmental SNR (SEGSNR) distance measure that is achieved for a SNR of 15db is -4.7093. These results show that using wavelet transforms give lower values of residual intelligibility. Since high computation time is involved, the wavelet structure level is restricted to three.

In the technique based on the combination of QAM mapping method and an orthogonal frequency division multiplexing (OFDM), the speech signal in PCM format is converted to complex valued frequency components by QAM mapping. These components are permuted and then inverse transformed to get the time-domain signal. To control bandwidth expansion the number of components is restricted to 93 corresponding to frequencies from 375-3250 Hz. The length of the scrambling key is equal to 93 and hence the key-space is 93!. The formant and pitch information are totally removed in the scrambled speech thereby lowering the residual intelligibility. [D.C.Tseng, J.H.Chiu, 2007]

### 4.5.1 Experimental Overview

The Table 5 given below lists the comparative values of the factors for the various algorithms. In most of the algorithms the residual intelligibility is lower. In all the algorithms the usable key-space is high but limited to the number of transform coefficients selected. Encoding delay is higher because of the larger number of samples available in each frame and bandwidth expansion is comparatively high.

For a given frame length the residual intelligibility and key-space of DPST and FFT based algorithms are comparable, but the encoding delay of the DPST algorithm is higher as it involves more number of calculations. In a DCT based system when the samples/frame is greater than 256, the residual intelligibility decreases, but the encoding delay increases. Circulant transformation based system is capable of distorting the silent portions of the speech thereby reducing the intelligibility levels to very low values. In this scheme, since the phase vector is the key, theoretically infinite choices of keys are possible for the phase range 0 to π.

### 4.5.2 Merits and Demerits

The algorithms in this category have very low levels of residual intelligibility with sentence intelligibility closer to zero percent. The presence of speech-silence rhythm is removed. The key-space is high because these algorithms work on a considerably larger number of permutable transform coefficients, crypt-analytic strength is also higher. The noise components in the original signal are not enhanced and will be kept at the same level. Moreover, the energy of the scrambled signal is held constant.

The encoding delay is high because these types of scramblers involve time and frequency domain manipulations, the bandwidth expansion is also higher.

### 4.5.3 Future Work

1) The effect of distortions introduced by the channel, on the performance of the scrambling algorithms needs to be studied.

2) An efficient key distribution scheme is required for the uniform filter bank and wavelet packet based scrambling algorithms. The problem of finding an efficient key distribution schemes for the time-frequency scrambling algorithms is still open.

3) DCT based system on Multi-frame structure is to be explored.

## 5. Conclusion

Audio Scrambling has been an active research topic for the past decades due to its wide areas of application. For satisfying specific factors of audio scrambling, a variety of algorithms have been proposed. For better comparison of the various features, algorithms discussed in this paper have been classified into five categories. The basic principles of the algorithms have been presented and their strengths and weakness have been analyzed. The review shows that there is no algorithm that is suitable for all applications. The design of fast audio scrambling algorithms with zero residual intelligibility, large key-space, low bandwidth expansion and good recovered speech quality remains therefore a goal for future research.

## References

R.V.Cox, T.M.Tribolet (1983), Analog voice privacy systems using tfsp scrambling: Full duplex and half duplex., The Bell System Technical Journal 62 (Jan 1983) 47-61.

K.Ganesan, R.Muthukumar, K.Murali, (2006) Look-up table based chaotic encryption of audio files, in: Circuits and Systems, 2006. APCCAS 2006. IEEE Asia Pacific Conference on, volume 5, pp. 1951-1954. Doi: 10.1109/APCCAS.2006.34224 http://dx.doi.org/10.1109/APCCAS.2006.34224

F. Huang, E. V. Stansfield, (1993) Time sample speech scrambler which does not require synchronization, IEEE Transactions on communications 41 (November 1993) pp.1715-1722. Doi: 10.1109/26.241752 http://dx.doi.org/10.1109/26.241752

N.S.Jayant, (1982) Analog scramblers for speech privacy, Computers and Security, North-Holland Publishing Company pp.275-289.Doi:10.1016/0167-4048(82)90047-5 http://dx.doi.org/10.1016/0167-4048(82)90047-5

N.S.Jayant, R.V.Cox, B.J.McDermott, A.M.Quinn, (1983) Analog scramblers for speech based on sequential permutations in time and frequency, The Bell System Technical Journal 62 (Jan 1983) pp.25-46.

S.C.Kak, et al, (1977) On speech encryption using waveform scrambling, The Bell System Technical Journal 56 (May-June 1977) pp.781-808.

S.C.Kak,et al, (1983) Overview of analog signal encryption, IEE Proceedings 130 (August 1983) pp.399-404. Doi: 10.1049/ip-f-1:19830066 http://dx.doi.org/10.1049/ip-f-1:19830066

Q.-H. Lin, F.-L. Yin, T.-M. Mei, H.-L. Liang, (2004) A speech encryption algorithm based on blind source separation, in: ICCCAS 2004.International Conference on Communications, Circuits and Systems., pp. 1013-1017.

F. Ma, J. Cheng, Y. Wang, (1996) Wavelet transform-based analogue speech scrambling scheme, Electronics Letters 32 (11th April 1996) pp.719-720. Doi: 10.1049/el:19960471 http://dx.doi.org/10.1049/el:19960471

V. Milosevic, V. Delic, V. Senk, (1997) Hadamard transform application in speech scrambling, in: Digital Signal Processing Proceedings,1997. DSP 97., 1997 13th International Conference, volume 1, pp. 361-364. Doi: 10.1109/ICDSP.1997.628102 http://dx.doi.org/10.1109/ICDSP.1997.628102

G.Manjunath, G.V.Anand, (2002) Speech encryption using circulant transformations, in: 2002 IEEE International Conference on Multimedia and Expo, pp. 553-556. Doi: 10.1109/ICME.2002.1035841 http://dx.doi.org/10.1109/ICME.2002.1035841

J.Phillips, M.H.Lee, J.E.Thomas, (1971) Speech scrambling by the re-ordering of amplitude samples, Radio and Electronic Engineer 41 (March 1971) pp.99-112. Doi: 10.1049/ree.1971.0038 http://dx.doi.org/ 10.1049/ree.1971.0038

K. Sakurai, K. Koga, T. Muratan, (1984) A speech scrambler using the fast fourier transform technique, IEEE Journal on selected areas in communications 2 (May 1984) pp.434-442. Doi: 10.1109/JSAC.1984.1146074 http://dx.doi.org/10.1109/JSAC.1984.1146074

S.Sridharan, E.Dawson, B.Goldburg,(1993) Design and cryptanalysis of transform-based analog speech scramblers, IEEE Journal on selected areas in communications 11(June 1993) pp.735-744. Doi: 10.1109/49.223875 http://dx.doi.org/10.1109/49.223875

D. B.Sadkhan, D. Abdulmuhsen, N. F.Al-Tahan, (2007) A proposed analog speech scrambler based on parallel structure of wavelet transforms, in: 24th National Radio Science Conference (NRSC 2007), pp. C13/1-C13/12.

D.C.Tseng, J.H.Chiu, (2007) An ofdm speech scrambler without residual intelligibility, in:TENCON 2007 - 2007 IEEE Region 10 Conference, pp. 1-4 Doi: 10.1109/TENCON.2007.4428903 http://dx.doi.org/ 10.1109/TENCON.2007.4428903

A. D.Wyner, (1979) An analog scrambling scheme which does not expand bandwidth, part 1: Discrete time, IEEE Transactions on information theory 25 (May 1979) pp.261-274. Doi: 10.1109/TIT.1979.1056050, http://dx.doi.org/10.1109/TIT.1979.1056050,

H. Wang, M. Hempel, D. Peng, W. Wang, H. Sharif, H.-H. Chen, (2010) Index-based selective audio encryption for wireless multimedia sensor networks, IEEE Transactions on multimedia 12 (April 2010) pp.215-223. Doi: 10.1109/TMM.2010.2041102 , http://dx.doi.org/10.1109/TMM.2010.2041102

**Srinivasan Arulanandam** completed his ME, PhD in computer Science and Engineering at Madras Institute of Technology, Anna University, Chennai. He has finished his Post Doctorate at Nan yang Technological University, Singapore. He has 20 years of Teaching and Research Experience in Computer Science and Engineering field and one year of Industrial Experience. He has successfully guided 32 M.E projects and currently guiding 8 PhD students. He has published more than 52 Research publications in National and International journals and conferences. He is on the editorial board in Journal of Computer Science and Information Technology [JCSIT] and Review Board Member to ten reputed International Journals in Computer Science and Engineering field. Currently he is working as Senior Professor and Head, Department of Information Technology, Misrimal Navajee Munoth Jain Engineering College, Anna University, Chennai, India. His fields of interests are Digital Image processing, Face Recognition and Distributed Systems.

**Arul Selvan Palanisamy** received his Bachelor's degree from College of Engineering, Guindy, Anna University Chennai and M.E (Electronics) from Madras Institute of Technology, Anna University, Chennai. Currently he is pursuing his PhD in Sathyabama University Chennai in the field of Audio and Video scrambling techniques. He has 11 years of industry experience and 3 years of teaching experience. His research interests are Audio and Video signal Processing.

Table 1. Feature comparison of sample-amplitude based techniques

| Algorithm | Residual Intelligibility | Key-Space | Encoding Delay |
|---|---|---|---|
| Sample Interchange | Medium,24% for a Block Size of 128 | Very Low | Low |
| Linear addition of PR Noise | Medium | Low | Low |
| Uniform Permutation | Medium, Recognition of speech-silence pattern | Medium, 63232 for a block size of 256 | Low |
| PR Permutation | Medium, Recognition of speech-silence pattern | Low, 4080 for a block size of 256 | Low |
| Chaotic Encryption | Low | High, 65536 | High |

Table 2. Feature comparison of time domain based techniques

| Algorithm | Residual Intelligibility | Key-Space | Encoding Delay |
|---|---|---|---|
| Time Inversion | Medium | Low | Medium,256/512ms |
| Time Segment Permutation(TSP) | Medium, Digit intelligibility 30% | Very Low,0.1% | High,512ms |
| Reverberation | Medium | Low | Medium, depends on segment length |
| Scrambler without synchronization | Medium | Low,group of few constant and variable keys | Low |
| Blind Source Separation | Medium | Medium,PR Keys used | Medium |

Table 3. Feature comparison of frequency domain based techniques

| Algorithm | Residual Intelligibility | Key-Space | Encoding Delay |
|---|---|---|---|
| Frequency Inversion | Medium, Digit intelligibility 30% | Very Low, One Key | Zero |
| Band-splitting | Medium | Low, F! | Low |
| Band-splitting with Frequency Inversion | High, 50-70% | Low, F!*2F | Low |
| Frequency Inversion followed by cyclic band-shift | Medium, 55% | Low | Low |

Table 4. Feature comparison of two-dimensional techniques

| Algorithm | Residual Intelligibility | Key-Space | Encoding Delay |
|---|---|---|---|
| Frequency Inversion | Low, Digit intelligibility 20% | Very Low | Medium |

| | | | |
|---|---|---|---|
| with Block TSP | | | |
| Dynamic-time reverberation | Low, Digit intelligibility 20% | Very Low | Medium |
| Time-shifting of the frequency sub-bands | Medium,Digit intelligibility 38% | Very Low | Medium |
| Time-Frequency segment permutation(TFSP) | Low, Digit intelligibility 25%, Word intelligibility close to 0% | Very Low, fb time-frequency segments | Medium, 256ms |

Table 5. Feature comparison of transform domain based techniques

| **Algorithm** | **Residual Intelligibility** | **Key-Space** | **Encoding Delay** |
|---|---|---|---|
| Discrete Prolate Spheroidal Transform(DPST) | Low | Limited to the order of the permutation matrix M | High, depends on the order of the permutation matrix M |
| Fast Fourier Transform (FFT) | Low | Limited to the number of FFT coefficients permuted (84!) | Medium, for 256 samples/frame |
| Discrete Cosine transform (DCT) | Low | Limited to the number of DCT coefficients permuted (197!) | Medium |
| Selective Encryption using modified discrete cosine transform (MDCT) | Medium, Since a Portion of the transform coefficients are encrypted | Medium, Limited to the index value K. | High, For Frame Lengths greater than 256 |
| Hadamard Transform | Approx 0% Sentence Intelligibility for frame length of 64ms. | High, depends on the order of H matrix ( N!*2*N)*2 | Medium |
| Circulant transformation | Low, Dependent on the order of the circulant matrix and frame length. | Very High, Theoretically Infinite. | High, For Frame Lengths greater than 256 |
| Wavelet Transform | Very Low | Depends on the Frame Size chosen | Medium |
| Parallel structure of two different types of wavelets | Low | Medium, Limited to 30% of total frame length. | High, when more than 3 Wavelets are used |
| QAM plus orthogonal frequency division multiplexing (OFDM) | Low, removes talk spurts and original intonation | Medium, Limited to (Number of Frequency Components)!,which is 93!. | Medium, when Number of Frequency Components is limited to 93. |

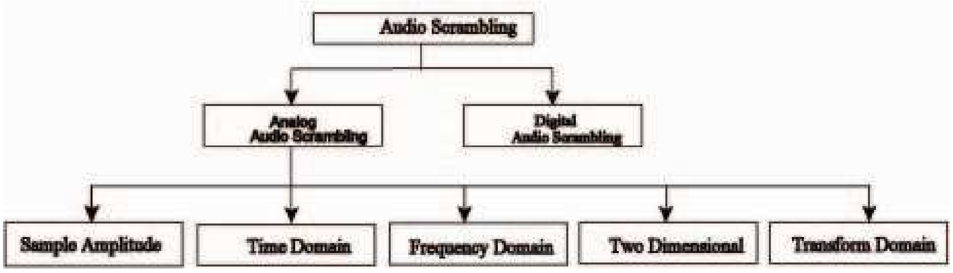
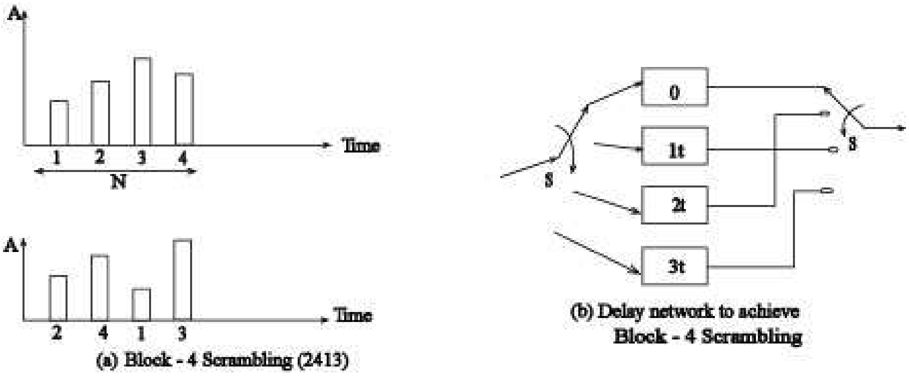


**Figure 1: Taxonomy of analog audio scrambling algorithms**



(a) Block - 4 Scrambling (2413)

(b) Delay network to achieve Block - 4 Scrambling

**Figure 2: Block-4 Scrambling**

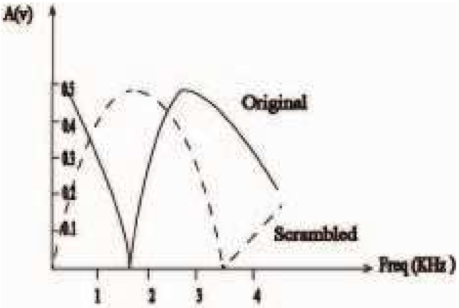


**Figure 3: Spectral Amplitudes after scrambling**

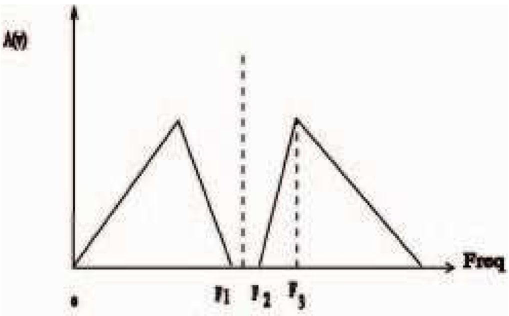


**Figure 4: Frequency Inversion Scheme**

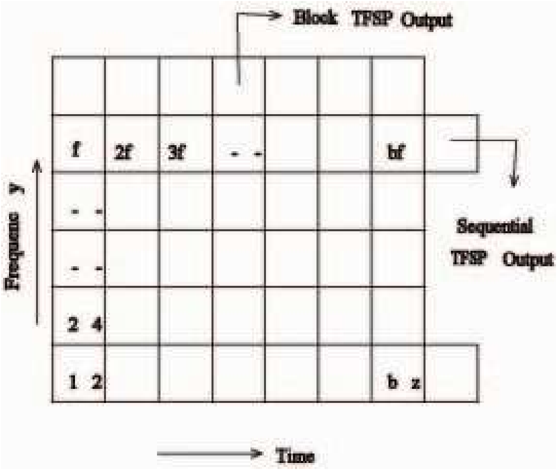


**Figure 5: TFSP Scrambling Scheme**

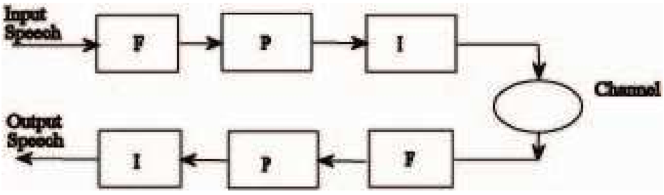


**Figure 6: Transform Domain Scrambling**

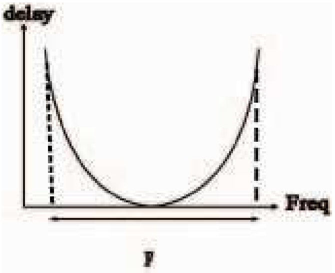


**Figure 7: Parabolic Group Delay Distortion**

This academic article was published by The International Institute for Science, Technology and Education (IISTE). The IISTE is a pioneer in the Open Access Publishing service based in the U.S. and Europe. The aim of the institute is Accelerating Global Knowledge Sharing.

More information about the publisher can be found in the IISTE's homepage: http://www.iiste.org

The IISTE is currently hosting more than 30 peer-reviewed academic journals and collaborating with academic institutions around the world. **Prospective authors of IISTE journals can find the submission instruction on the following page:** http://www.iiste.org/Journals/

The IISTE editorial team promises to the review and publish all the qualified submissions in a fast manner. All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Printed version of the journals is also available upon request of readers and authors.

**IISTE Knowledge Sharing Partners**

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library , NewJour, Google Scholar