Circuits Syst Signal Process (2008) 27: 893–913
DOI 10.1007/s00034-008-9066-3

NONLINEAR CIRCUITS AND SYSTEMS

A Bandwidth Extension Technique for Signal Transmission Using Chaotic Data Hiding

Siyue Chen · Henry Leung

Received: 1 November 2007 / Revised: 10 March 2008 / Published online: 29 October 2008

Abstract Bandwidth extension can be defined as a deliberate process of expanding the frequency range (bandwidth) for signal transmission. Its significant advancement in recent years has led to the technology being adopted commercially in several areas, including psychoacoustic bass enhancement of small loudspeakers and the high frequency enhancement of perceptually coded audio. In this paper, a novel bandwidth extension method based on chaotic data hiding is proposed. More specifically, a hidden channel is created by removing the imperceptible components from the transmitted audio or video. The out-of-band information is then encoded and embedded into the hidden channel without degrading the quality of the bandlimited signal. At the receiver, when the out-of-band information is extracted from the hidden channel, it can be used to combine with the bandlimited signal, providing a signal with a wider bandwidth. To balance the tradeoff between robustness and payload of transmitting the out-of-band information through the hidden channel, the technique of chaotic code division multiple access (CDMA) is employed. Since minimizing the multiple access interference (MAI) is crucial to the detection performance of a CDMA system, we employ a hybrid algorithm based on genetic programming (GP) and DNA computing to generate the desired spreading sequences with low cross-correlation. The effectiveness of the proposed method is validated by applying it to enhance telephony speech quality.

The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Wing-Kuen Ling.

S. Chen (✉) · H. Leung
Department of Electrical and Computer Engineering, University of Calgary, 2500 University Drive NW, Calgary, Alberta, T2N 1N4, Canada
e-mail: chens@ucalgary.ca

H. Leung
e-mail: leungh@ucalgary.ca

Keywords Chaos · Communication · Data hiding · Linear prediction · Spread spectrum

1 Introduction

Bandwidth scarcity is a common problem of contemporary telecommunications. For example, speech in communication networks is mostly transmitted in narrowband (NB) by using an audio bandwidth (0.3–3.4 kHz) and a sampling frequency (8 kHz) originating from conventional pulse coding modulation (PCM). Both the pleasantness and intelligibility of NB speech suffer from the limited bandwidth. A similar situation also exists for video communication. Although wideband (WB) transmission will become more prevalent in the future, it will take time before all the terminals and networks support WB transmission. During the long transition period, alternative solutions have to be sought to extend the bandwidth virtually, instead of physically.

Many efforts have been devoted to artificially generate WB audio/video from NB audio/video [3, 6, 15]. The rationale is that while WB and NB audio/video are generated from the same source, they are highly correlated to each other [8]. By exploiting the mutual dependence between these NB and WB signals, the missing WB components can be estimated from the NB signal. However, "dependence" is not "deterministic". This artificial bandwidth extension (ABE) approach has the limitation of lacking sufficient information to estimate the out-of-band (the frequency range beyond NB but within WB) information accurately [9].

In this paper, we propose to embed the out-of-band information into the NB signal by data hiding, which makes use of a hidden channel to transmit out-of-band information [1, 19, 21]. More precisely, the NB components below the perceptual threshold are removed. While this removal does not degrade the perceptual quality of the NB audio/video, it creates a hidden channel, in which the out-of-band information can be transmitted. At the receiver, when the hidden information is retrieved, a WB signal can be constructed by combining the out-of-band signal that was transmitted through the hidden channel and the NB signal. Since the proposed method uses real out-of-band information for bandwidth extension, it does not have the problems of conventional ABE methods, which use estimated values to construct the WB signal. Furthermore, the proposed method is compatible with conventional NB terminal equipments, e.g., a plain ordinary telephone set (POTS). In other words, conventional NB receivers can still access the NB audio/video properly without additional hardware, while a customized receiver is able to extract the hidden signal and provide WB services.

There are two challenges for this data hiding-based bandwidth extension framework. On one hand, we want as much out-of-band information as possible to be transmitted through the hidden channel. On the other hand, the hidden signal should be robust to noise corruption or the quantization process. Because the direct sequence code division multiple access (DS-CDMA) technique is well recognized by its high capacity and robustness to interference, it is employed in this study for data transmission through the hidden channel. In particular, the data bits carrying the out-of-band information are spread by multiplying with spreading sequences before embedding.

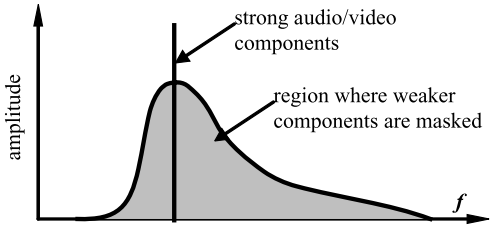


Fig. 1 Illustration of perceptual masking

The resulted spread signals are then multiplexed to form the hidden signal for transmission. At the receiver, the corrupted hidden signal is extracted, and a multiuser detector is used to retrieve the out-of-band information.

To minimize the multiple access interference (MAI) caused by the other simultaneously transmitted data information, spreading sequences with low cross-correlations are preferred. Conventional spreading sequences, e.g., Walsh-Hadamard codes, are able to achieve an optimal cross-correlation performance, i.e., orthogonal to each other, only when the sequence length is a power of 2, i.e., 2^k, $k = 1, 2, 3, \ldots$. However, in our application of bandwidth extension, the sequence length can be any positive integer. In order to meet the design requirement that the spreading sequences should have optimal cross-correlation performance for any sequence length, a chaotic system with a hybrid algorithm based on genetic programming (GP) and DNA computing is employed in this study. The proposed hybrid algorithm uses GP to design a chaotic map with a maximum number of desired spreading sequences, and it uses DNA computing to determine the initial conditions of the chaotic map.

The paper is organized as follows. Section 2 describes the proposed chaotic data hiding method for bandwidth extension. Section 3 presents the design of chaotic spreading sequences by GP-DNA. The performances of the proposed scheme under additive white Gaussian noise (AWGN) and quantization error are analyzed in Sect. 4. Section 5 presents a system design of telephony speech bandwidth extension using chaotic data hiding. Experimental results using real speech signals are reported in Sect. 6, and the concluding remarks are given in Sect. 7.

2 Chaotic Data Hiding for Bandwidth Extension

Data hiding is based on the knowledge that the presence of strong audio/video components makes a spectral neighborhood of weaker components imperceptible, as shown in Fig. 1. Therefore, the spectrum components of an NB audio/video signal below the perceptual thresholds can be removed without degrading the perceptual quality. Assume that the NB audio/video signal is denoted by $\{x_{\mathrm{NB}}(k)\}$, $k = 0, 1, \ldots, N-1$. Its corresponding spectrum components can be represented by $\{X(k)\}$,[1] $k = 0, 1, \ldots, N-1$, where N is the total number of samples. This below-

[1] A two-dimensional signal can be transformed to a one-dimensional signal by raster scan, zig-zag scan, etc.

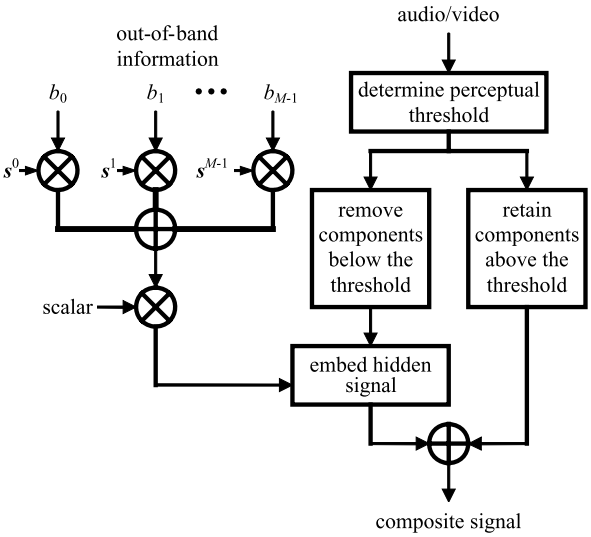


Fig. 2 Block diagram of embedding the encoded out-of-band information into NB signal by chaotic CDMA

perceptual-threshold removal process can be expressed by

$$X(k) = \begin{cases} X(k), & \text{if } |X(k)| \geq T(k), \\ 0, & \text{if } |X(k)| < T(k), \end{cases} \quad k = 0, 1, \ldots, N-1, \tag{1}$$

where $T(k)$ is the perceptual threshold obtained by passing the audio/video signal through a psychoacoustic model [16] or a psychovisual model [22].

Assume that L components are set to zero during the removal process and that the indices of these components are $k_0, k_1, \ldots, k_{L-1}$ respectively. We also assume that the out-of-band information is encoded into a sequence of binary digits, i.e., $\{b_m\}$, $b_m \in \{-1, 1\}$, $m = 0, 1, M-1$, where M denotes the total number of digits. Figure 2 plots the embedding procedure of the proposed chaotic hiding approach. As shown, each data bit is spread out by multiplying with a spreading sequence, i.e., $b_m \mathbf{s}^m$. The length of the spreading sequence $\mathbf{s}^m$ is equal to L. All these spreading signals are then added together to generate the hidden signal. That is,

$$\mathbf{h} = \sum_{m=0}^{M-1} b_m \mathbf{s}^m, \tag{2}$$

where $\mathbf{h}$ denotes the vector of the hidden signal. $\mathbf{h}$ is embedded into the NB signal by

$$X(k) = \alpha h_l, \quad k = k_0, k_1, \ldots, k_{L-1}, \tag{3}$$

where h_l is the lth element of $\mathbf{h}$, and α is a scalar that controls the hidden signal to stay below the perceptual threshold, i.e., $\alpha^2 h_l^2 \leq T(k)$. Hence, a suitable value of α can be determined by $\alpha = \sqrt{T(k)/h_l^2}$. Considering that $|b_m s_l^m|$ is always equal to 1, we have

$$\alpha = \sqrt{\frac{T(k)}{M}}. \tag{4}$$

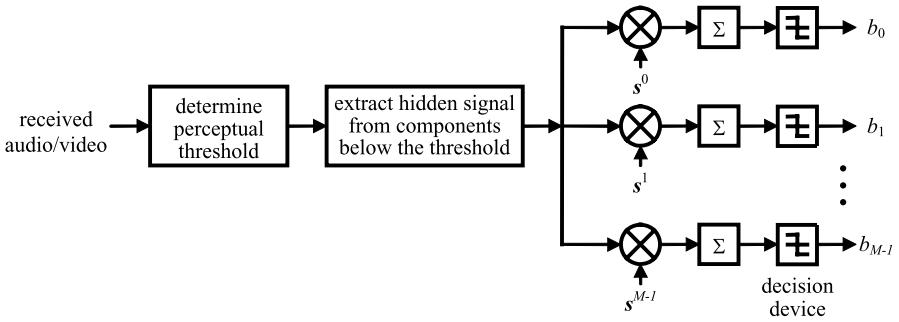


Fig. 3 Block diagram of decoding the out-of-band information by chaotic CDMA

After embedding, $\{X(k)\}$, $k = 0, 1, \ldots, N-1$ is transformed back to the time domain, providing the composite signal $x'_{\mathrm{NB}}(k)$ to be sent through the communication channel.

At the receiver side, conventional terminals treat the received signal, i.e., $X(k)$, as an ordinary audio/video signal. Since the hidden signal is under the perceptual threshold, it will not be perceptible to humans. Meanwhile, a pre-designed decoding mechanism is able to retrieve the digital data $\{b_m\}$ from the hidden channel. As shown in Fig. 3, the perceptual threshold is first evaluated again at the receiver. Since the hidden signal is embedded under the perceptual threshold while the other spectrum components are unaltered, the perceptual threshold estimated at the receiver should be the same as the one obtained at the transmitter. The hidden signal **h** can thus be properly extracted by

$$\hat{h}_l = \hat{X}(k), \quad \text{if } \hat{X}_p(k) < T(k). \tag{5}$$

We use $\hat{h}_l$ instead of αh_l to denote the extracted hidden signal because the extracted hidden signal is subjected to noise corruption. A multiuser detector [17] is used here to decode the digital data. That is,

$$\hat{b}_m = \operatorname{sign}\left(\sum_{l=0}^{L-1} \hat{h}_l s_l^m\right). \tag{6}$$

In a noise-free environment, $\hat{h}_l = \alpha h_l$. Substituting it into (6), the decoding process can be further interpreted as

$$\begin{aligned}
\hat{b}_m &= \operatorname{sign}\left(\sum_{l=0}^{L-1} \alpha h_l s_l^m\right) \\
&= \operatorname{sign}\left(\alpha \sum_{l=0}^{L-1} \left(b_m s_l^m s_l^m + \sum_{n=0, n\neq m}^{M-1} b_n s_l^n s_l^m\right)\right) \\
&= \operatorname{sign}\left(\alpha L b_m + \alpha \sum_{n=0, n\neq m}^{M-1} b_n \sum_{l=0}^{L-1} s_l^n s_l^m\right).
\end{aligned} \tag{7}$$

It can be seen that if $\sum_{l=0}^{L-1} s_l^n s_l^m = 0$, the MAI coming from other simultaneously transmitted data can be completely removed. In other words, in order to decode b_m successfully, the spreading sequences are required to be orthogonal to each other. Therefore, the design of the spreading sequences is critical in order to reduce MAI.

3 Design of Chaotic Spreading Sequences with GP-DNA Algorithm

The design of chaotic discrete sequences for DS-CDMA using ergodic properties of chaotic systems has been applied to one-dimensional (1-D) chaotic systems [2, 14]. In these studies, a particular chaotic map $\mathcal{F}(\cdot)$ is usually taken and iterated using a randomly selected initial condition $\mathbf{c}_0^m$. That is,

$$\mathbf{c}_l^m = \mathcal{F}^l\left(\mathbf{c}_0^m\right), \quad l = 1, \ldots, L-1,\ m = 0, \ldots, M-1. \tag{8}$$

Then the time series are converted to binary spreading sequences using an appropriate thresholding rule as follows:

$$\mathbf{s}_l^m = \mathcal{T}\left(\mathbf{c}_l^m - \frac{1}{L}\sum_{l=0}^{L-1} \mathbf{c}_l^m\right), \tag{9}$$

where

$$\mathcal{T}(x) = \begin{cases} 1, & x > 0, \\ -1, & x \le 0. \end{cases} \tag{10}$$

The ergodic properties of chaotic maps are then used to estimate the correlation performance. These are basically asymptotic performance measures, and the theoretical bounds are approached only when the sequence length is very long. However, in this study, the sequence length is much shorter than what is required to approach the theoretical limit. To overcome the problem, an efficient GP-DNA algorithm [18, 20] is employed for a quick design of the desired chaotic maps and identification of the initial conditions.

A DNA computation model for the selection of the initial conditions is depicted in Fig. 4. As shown, "Lor" represents the Lorenz system that produces the perturbations and controls the well-designed chaotic map $\mathcal{F}(\cdot)$. Initially, $\mathbf{c}^1$ is obtained as the base sequence by optimizing its autocorrelation function. The control feedback CF_2 then evaluates the correlation properties of $\mathbf{c}^2$ with respect to $\mathbf{c}^1$. It outputs a logical 0 when the new spreading sequence is optimized (i.e., orthogonal). Similarly, the mth control feedback CF_m evaluates the correlation properties of $\mathbf{c}^m$ with all the previously optimized spreading sequences. An "OR" operation is then performed for all the CF outputs. The resulting signal is then multiplied by δ, which is generated by the Lorenz system, to produce a perturbation value ξ. A new initial condition can be obtained by adding ξ to the current one.

With DNA computation, we can search initial conditions for any given chaotic map to generate desired spreading sequences. Meanwhile, the chaotic map itself also

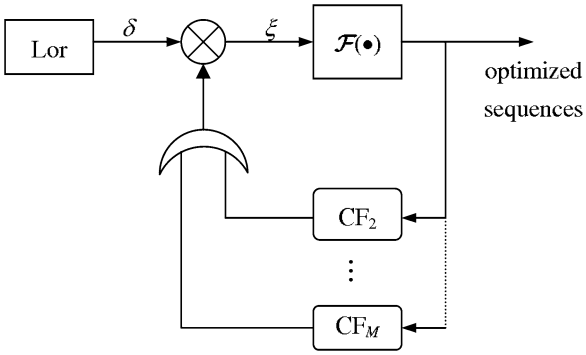


Fig. 4 Schematic diagram of DNA computation based on chaos

has to be optimized. GP is employed for this purpose. To illustrate the optimization process, the tree representation of the dynamical system

$$\mathbf{c}_l^m = \left(2.876 * \mathbf{c}_{l-1}^m\right) \bmod 1 + 1.4327 * \mathbf{c}_{l-1}^m \tag{11}$$

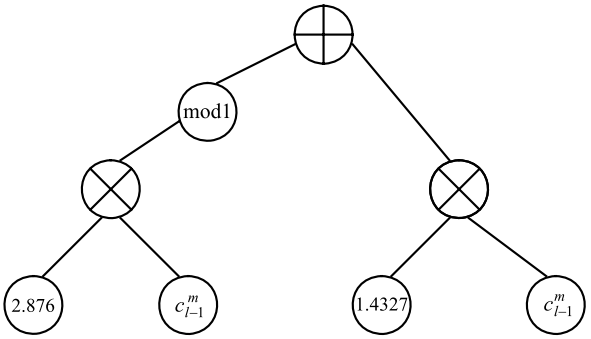
is plotted in Fig. 5. It is seen that the leaf nodes are variables from the initial terminal set, i.e., $\mathbb{C} = \{c_{l-1}^m,$ random ephemeral constant$\}$, and the internal nodes are functional operators from the functional set, i.e., $\mathbb{F} = \{+, -, *, \bmod 1\}$. Since the objective is to obtain the chaotic map that can provide the maximum number of low cross-correlation spreading sequences for the length L, the fitness of each individual in any population of GP is assigned a rank according to the number of optimal sequences that it produces. We can summarize the GP-based searching algorithm as follows:

Step 1: Set up an initial population of random functions (trees) with the "ramped half-and-half" method. An equal amount of trees are assigned to depths $d = \{2, 3, \ldots, D_{\max}\}$. At each depth, half the trees are generated with the "full" method, and the other half with the "grow" method. In the "full" method, all the leaf nodes have the same depth d_i. Internal nodes are randomly selected from the functional set $\mathbb{F}$, and leaf nodes are selected from the terminal set $\mathbb{C}$. The probability density among the choices from $\mathbb{F}$ and $\mathbb{C}$ is uniformly distributed. The "grow" method starts by randomly choosing a function in $\mathbb{F}$ or an input in $\mathbb{C}$ for the root node (50% probability for each). Then a corresponding symbol from $\mathbb{C}$ or $\mathbb{F}$ is selected to achieve the depth d_i by recursively applying the "grow" method to that node.

Step 2: Rank each individual according to the number of optimal spreading sequences that can be obtained. Sort the population according to the rank and assign the fitness to the individual by linearly interpolating from the best to the worst rank. Average the fitness of individuals with the same rank so that all of them are sampled at the same rate.

Step 3: Create a new population by reproducing or combining selected functions in the current population. Reproduction of a function is simply a copying of it to the new population. The genetic operator of crossover is used to combine functions. Crossover involves crossing over two random subtrees of two respective parents. The probability of an individual being selected for reproduction or crossover is

Fig. 5 A GP functional tree



proportional to that individual's fitness rank. The probability of performing reproduction vs. crossover is set to 0.2 : 0.8. Also, the population is forced to be 100% diverse; in other words, no duplicates.

Step 4: Repeat steps 2 and 3 until a convergence to an optimal solution occurs or the maximum number of generations is exceeded.

It is found that with the GP-DNA algorithm we can generate orthogonal sequences for any length that satisfies $L = 4k$, $k = 1, 2, \ldots$. When $L = 4k - 1$, we can generate the sequences which have the optimal cross-correlation as Gold sequences. Therefore, the length of spreading sequences which have the optimal cross-correlation performance is significantly extended compared to that of conventional sequences. In addition, since the proposed algorithm uses only one chaotic map for generation by simply adjusting the initial conditions, the system complexity is greatly reduced.

4 Performance Analysis of Chaotic Data Hiding

Although using orthogonal chaotic signature sequences can guarantee a correct decoding of b_m in a noise-free environment, in practice, there is always a noise corruption during transmission. The most common one is additive white Gaussian noise (AWGN), which comes from electronic devices. Mathematically, we can express the received NB signal as

$$\hat{x}_{\mathrm{NB}}(k) = x'_{\mathrm{NB}}(k) + d(k), \tag{12}$$

where $d(k)$ denotes AWGN with zero mean and a variance of σ_d^2. Its Fourier transform $X(k)$ is given by

$$\hat{X}(k) = \sum_{n=0}^{N-1} \hat{x}_{\mathrm{NB}}(n)\mathrm{e}^{-j\frac{2\pi}{N}kn} = \sum_{n=0}^{N-1} \big(x'_{\mathrm{NB}}(n) + d(n)\big)\mathrm{e}^{-j\frac{2\pi}{N}kn}. \tag{13}$$

In other words, the extracted hidden signal can be expressed as

$$\hat{h}_l = \alpha h_l + \sum_{n=0}^{N-1} d(n)\mathrm{e}^{-j\frac{2\pi}{N}kn} \tag{14}$$

based on (5) and (3). Defining $\epsilon_l = \sum_{n=0}^{N-1} d(n)\mathrm{e}^{-j\frac{2\pi}{N}ln}$, (14) becomes

$$\hat{h}_l = \alpha h_l + \epsilon_l. \tag{15}$$

When a multiuser detector as given in (6) is applied, we have

$$\begin{aligned}\hat{b}_m &= \operatorname{sign}\left(\sum_{l=0}^{L-1}(\alpha h_l + \epsilon_l)s_l^m\right)\\ &= \operatorname{sign}\left(\alpha L b_m + \sum_{n\neq m}\alpha b_n \sum_{l=0}^{L-1} s_l^n s_l^m + \sum_{l=0}^{L-1}\epsilon_l s_l^m\right).\end{aligned} \tag{16}$$

Since MAI is completely removed by the use of orthogonal chaotic sequences, (16) can be further written as

$$\hat{b}_m = \operatorname{sign}\left(\alpha L b_m + \sum_{l=0}^{L-1}\epsilon_l s_l^m\right). \tag{17}$$

From (17), a detection error occurs if $|\sum_{l=0}^{L-1}\epsilon_l s_l^m| \geq \alpha L$ and $\sum_{l=0}^{L-1}\epsilon_l s_l^m$ has an opposite sign with b_m. According to the central limit theorem [4], the conditional probability density function (PDF) of $\hat{b}_m$ can be expressed by

$$p(\hat{b}_m|-1) = \frac{1}{\sqrt{2\pi\sigma_{\epsilon s}^2}}\mathrm{e}^{-\frac{(\hat{b}_m+\alpha L)^2}{2\sigma_{\epsilon s}^2}}, \tag{18}$$

provided $b_m = -1$. Here, $\sigma_{\epsilon s}^2$ is the variance of the random variable $\sum_{l=0}^{L-1}\epsilon_l s_l^m$, which can be expressed by

$$\sigma_{\epsilon s}^2 = E\left[\left(\sum_{l=0}^{L-1}\epsilon_l s_l^m\right)^2\right] = E\left[\sum_{l=0}^{L-1}\sum_{k=0}^{L-1}\epsilon_l\epsilon_k s_l^m s_k^m\right]. \tag{19}$$

Considering that ϵ_l is uncorrelated with s_l^m and s_l^m is a random variable with zero mean and a variance of 1, (19) can be further written as

$$\sigma_{\epsilon s}^2 = E\left[\sum_{l=0}^{L-1}\epsilon_l^2\left(s_l^m\right)^2\right] = \sum_{l=0}^{L-1}E\left[\epsilon_l^2\left(s_l^m\right)^2\right] = L\sigma_\epsilon^2, \tag{20}$$

where σ_ϵ^2 denotes the variance of ϵ_l. Similar to (18), the conditional PDF for $b_m = 1$ can be formulated as

$$p(\hat{b}_m|1) = \frac{1}{\sqrt{2\pi\sigma_{\epsilon s}^2}}\mathrm{e}^{-\frac{(\hat{b}_m-\alpha L)^2}{2\sigma_{\epsilon s}^2}}. \tag{21}$$

Therefore, the conditional PDF of deciding in favor of "1" given that "−1" was transmitted can be written as

$$p(1|-1)=\int_0^{\infty} p(\hat{b}_m|-1)\,\mathrm{d}\hat{b}_m=\frac{1}{\sqrt{2\pi\sigma_{\epsilon s}^2}}\int_0^{\infty}\mathrm{e}^{-\frac{(\hat{b}_m+\alpha L)^2}{2\sigma_{\epsilon s}^2}}\,\mathrm{d}\hat{b}_m. \tag{22}$$

The above equation can be expressed in terms of the complementary error function as

$$p(1|-1)=\frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{\alpha^2L^2}{2\sigma_{\epsilon s}^2}}\right), \tag{23}$$

where $\operatorname{erfc}(x)=\frac{2}{\sqrt{\pi}}\int_x^{\infty}\mathrm{e}^{-t^2}\,\mathrm{d}t$. Similarly, we can derive the error function for deciding in favor of $\hat{b}_m=-1$ when $b_m=1$. That is,

$$p(-1|1)=\frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{\alpha^2L^2}{2\sigma_{\epsilon s}^2}}\right). \tag{24}$$

Based on $p(1|-1)$ and $p(-1|1)$, the average probability of detection error can be obtained as

$$P_{\mathrm{err}}=p(-1)p(1|-1)+p(1)p(-1|1)=\frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{\alpha^2L^2}{2\sigma_{\epsilon s}^2}}\right), \tag{25}$$

assuming that $b_m=-1$ and $b_m=1$ are equiprobable. Substituting (20) and (4) into (25), we have

$$P_{\mathrm{err}}=\frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{LT(k)}{2M\sigma_{\epsilon}^2}}\right). \tag{26}$$

Recalling that $\epsilon_l=\sum_{n=0}^{N-1}d(n)\mathrm{e}^{-j\frac{2\pi}{N}ln}$, σ_{ϵ}^2 can be obtained as

$$\sigma_{\epsilon}^2=E\left[\left(\sum_{n=0}^{N-1}d(n)\mathrm{e}^{-j\frac{2\pi}{N}ln}\right)^2\right]=E\left[\sum_{n=0}^{N-1}d^2(n)\mathrm{e}^{-j\frac{4\pi}{N}ln}\right]=\sigma_d^2\sum_{n=0}^{N-1}\mathrm{e}^{-j\frac{4\pi}{N}ln}. \tag{27}$$

Since $\sum_{n=0}^{N-1}\mathrm{e}^{-j\frac{4\pi}{N}ln}$ can be treated as a scalar of σ_d^2 when N is set as a constant, we can reduce (27) to

$$\sigma_{\epsilon}^2=C\sigma_d^2, \tag{28}$$

where $C=\sum_{n=0}^{N-1}\mathrm{e}^{-j\frac{4\pi}{N}ln}$. Substituting (28) into (26), the error probability under AWGN can be obtained as

$$P_{\mathrm{err}}^{\mathrm{AWGN}}=\frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{LT(k)}{2CM\sigma_d^2}}\right). \tag{29}$$

Note that $T(k)/C\sigma_d^2$ is actually the signal-to-noise ratio (SNR) of the hidden channel, i.e., $\mathrm{SNR}_{\mathrm{hidden}} = \frac{T(k)}{C\sigma_d^2}$. That is,

$$P_{\mathrm{err}}^{\mathrm{AWGN}} = \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{L\mathrm{SNR}_{\mathrm{hidden}}}{2M}}\right). \tag{30}$$

The performance of the proposed method under AWGN depends on the number of inaudible components that are removed from the NB speech, the number of simultaneously transmitted data bits, and the SNR of the hidden channel. For video signals, although a discrete cosine transform (DCT), instead of a Fourier transform, is used to obtain $X(k)$, an analysis result similar to (30) can be obtained.

Besides AWGN, the hidden signal is also required to be robust to the quantization process, which converts analog waveforms to digital bits. The quantized NB signal can be modeled as

$$\hat{x}_{\mathrm{NB}}(k) = x'_{\mathrm{NB}}(k) + \underline{d}(k), \tag{31}$$

where $\underline{d}(k)$ is the quantization error. Hence, $X(k)$ can be expressed by

$$\hat{X}(k) = \sum_{n=0}^{N-1} \hat{x}_{\mathrm{NB}}(n)\mathrm{e}^{-j\frac{2\pi}{N}kn} = \sum_{n=0}^{N-1} \big(x'_{\mathrm{NB}}(n) + \underline{d}(n)\big)\mathrm{e}^{-j\frac{2\pi}{N}kn}. \tag{32}$$

Correspondingly, the extracted hidden signal is given by

$$\hat{h}_l = \alpha h_l + \sum_{n=0}^{N-1} \underline{d}(n)\mathrm{e}^{-j\frac{2\pi}{N}kn}. \tag{33}$$

Comparing (33) with (14), except that $\underline{d}(n)$ is used instead of $d(n)$, the expressions of the extracted hidden signal under AWGN and under quantization have the same formula. Hence, if we define $\epsilon_l = \sum_{n=0}^{N-1} \underline{d}(n)\mathrm{e}^{-j\frac{2\pi}{N}ln}$, the derivation of the error probability under quantization should be the same as that under AWGN. In other words, the error probability under quantization can be expressed by

$$P_{\mathrm{err}} = \frac{1}{2}\operatorname{erfc}\left(\sqrt{\frac{LT(k)}{2M\sigma_\epsilon^2}}\right). \tag{34}$$

The problem then becomes finding out what σ_ϵ^2 is for the quantization error.

When a uniform quantization is applied, $x_{\mathrm{NB}}(k)$ can be modeled as being uniformly distributed within each cell, i.e., $(K\Delta - \frac{\Delta}{2}, K\Delta + \frac{\Delta}{2}]$, where K is an integer and Δ is the quantization step size. Thus, we can obtain the expected squared error by quantization as

$$\sigma_d^2 = E\big[\underline{d}^2(k)\big] = \int_{-\Delta/2}^{\Delta/2} \underline{d}^2(k)\, p\big(\underline{d}(k)\big)\, \mathrm{d}\underline{d}(k) = \int_{-\Delta/2}^{\Delta/2} \frac{1}{\Delta}\underline{d}^2(k)\, \mathrm{d}\underline{d}(k) = \frac{\Delta^2}{12}. \tag{35}$$

Substituting (35) into (28), we have

$$\sigma_\epsilon^2 = \frac{C\Delta^2}{12}. \tag{36}$$

Given (36) and (34), the error probability under uniform quantization can be derived as

$$P_{\text{err}}^{\text{UQ}} = \frac{1}{2}\,\text{erfc}\left(\sqrt{\frac{6LT(k)}{CM\Delta^2}}\right). \tag{37}$$

Besides uniform quantization, non-uniform quantization is also popular in analog-to-digital conversion of audio/video signals. For example, in the 8-bit pulse coding modulation (PCM), quantization errors are proportional to signal amplitudes. In other words, a large quantization error is generated when the quantization is applied to low-probability high-magnitude signal samples, while a relatively small quantization error occurs for the high-probability low-magnitude samples. In this case, it is difficult to derive a general formula of σ_d^2. However, there is usually a constraint on the signal-to-quantization error ratio (SQER). For example, in PCM, the SQER is required to lie within 38 dB [5], so that the quality of digitized speech can be retained. That is, assuming the power of NB speech as σ_x^2, we have $\frac{\sigma_x^2}{\sigma_d^2} = 38\ \text{dB} = 6309.6$. Hence, the error probability under 8-bit PCM can be written as

$$P_{\text{err}}^{\text{PCM}} = \frac{1}{2}\,\text{erfc}\left(\sqrt{\frac{LT(k)}{2\times 1.585\times 10^{-4} CM\sigma_x^2}}\right) = \frac{1}{2}\,\text{erfc}\left(\sqrt{\frac{3154.8LT(k)}{CM\sigma_x^2}}\right). \tag{38}$$

The payload of the hidden channel is defined as the maximum number of data bits that can be transmitted simultaneously while a certain level of the error probability is retained. If the error probability requirement is denoted by P_0, in order to meet P_0 under AWGN, we have

$$\frac{L\text{SNR}_{\text{hidden}}}{2M} \geq \eta_0 \tag{39}$$

according to (30), and η_0 is the value that satisfies $\frac{1}{2}\,\text{erfc}(\sqrt{\eta_0}) = P_0$. With simple manipulations, we obtain

$$M \leq \frac{L\text{SNR}_{\text{hidden}}}{2\eta_0}. \tag{40}$$

Therefore, the payload of the hidden channel under AWGN can be expressed by

$$M_{\text{AWGN}} = \left\lfloor \frac{L\text{SNR}_{\text{hidden}}}{2\eta_0} \right\rfloor, \tag{41}$$

where $\lfloor\cdot\rfloor$ denotes the operator rounding the positive number to the nearest integer towards zero. Similarly, the payload of the hidden channel under uniform quantization and PCM can be derived as

$$M_{\text{UQ}} = \left\lfloor \frac{6LT(k)}{\eta_0 C\Delta^2} \right\rfloor \tag{42}$$

and

$$M_{\mathrm{PCM}} = \left\lfloor \frac{3154.8LT(k)}{\eta_0 C\sigma_x^2} \right\rfloor \tag{43}$$

respectively.

5 Application of Chaotic Data Hiding to Telephony Speech Bandwidth Extension

The current public switched telephone network (PSTN), which has been part of our daily life for more than a century, is designed to transmit toll-quality voice only. This design target has been inherited in most modern and fully digitized phone systems, such as digital private branch exchange (PBX) and voice over IP (VoIP) phones. As a result, these systems are only able to deliver analog signals in a relatively narrow frequency band, about 300–3400 Hz. This bandwidth limitation causes the characteristic sound of "telephone speech". Listening experiments have shown that an increased frequency bandwidth can significantly improve the speech quality and intelligibility. However, it is difficult to change the current PSTN infrastructure to provide a wider bandwidth. In this section, we propose to extend the bandwidth of telephony speech virtually, not physically, by the proposed chaotic data hiding method.

Figure 6 plots the system design. As shown, the original WB (0–7000 Hz) speech, with a sampling rate of 16 kHz, is split into NB (0–3500 Hz) and highband (HB, 3500–7000 Hz) signals by a low-pass and a high-pass filter respectively. The HB signal then undergoes a frequency down-shift operation. As a result, a frequency-shifted version of HB, i.e., HB_S, fits into the NB frequency range. Both NB and HB_S should be band-limited to below 4 kHz, so they are down-sampled, i.e., decimated, so that their sampling rate is halved to 8 kHz, which is the sampling rate of the NB channel.

Let us denote the decimated HB_S signal as $x_{\mathrm{NH}}(k)$, and the decimated NB signal as $x_{\mathrm{NB}}(k)$. $x_{\mathrm{NH}}(k)$ is encoded into binary digits by using linear predictive coding (LPC)

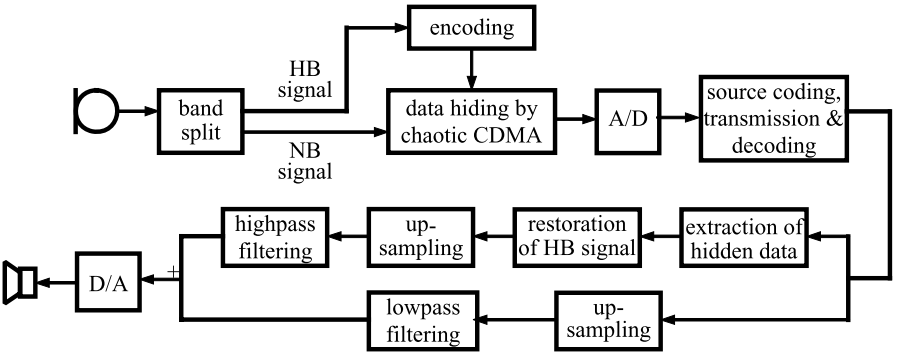


Fig. 6 Block diagram of a telephony speech bandwidth extension system using the chaotic data hiding method

analysis and a vector quantization (VQ) codebook. The LPC analysis is performed by employing a part of ITU-T G.729 [7]. More specifically, $x_{\mathrm{NH}}(k)$ is segmented into frames, and for each frame, with N speech samples, the autocorrelation coefficients are calculated by

$$u(n) = \sum_{k=0}^{80} x_{\mathrm{NH}}(k) x_{\mathrm{NH}}(n-k), \quad n = 0, \ldots, N_a, \tag{44}$$

where N_a denotes the order of an autoregressive (AR) filter. The value of $u(0)$ has a lower boundary of $u(0) = 1.0$. Based on $u(n)$, the AR filter coefficients $a_{\mathrm{NH}}(i)$, $i = 1, \ldots, N_a$, are obtained by applying the Levinson–Durbin algorithm [13] to solve the set of equations

$$\sum_{i=1}^{N_a} a_{\mathrm{NH}}(i) u'\big(|i-n|\big) = -u'(n), \quad n = 1, \ldots, N_a, \tag{45}$$

where $u'(n)$ represents the modified autocorrelation coefficients, i.e.,

$$u'(n) = \exp\left[-\frac{1}{2}\left(\frac{3\pi}{200}n\right)^2\right] u(n) \tag{46}$$

and $u'(0) = 1.0001u(0)$. The relative gain of $x_{\mathrm{NH}}(k)$ against $x_{\mathrm{NB}}(k)$ is also calculated as $G_{\mathrm{rel}} = \frac{G_{\mathrm{NH}}}{G_{\mathrm{NB}}}$. $a_{\mathrm{NH}}(i)$, $i = 1, \ldots, N_a$, is then combined with G_{rel} to form a representation vector of the decimated HB_S signal. That is, $\mathbf{a} = [G_{\mathrm{rel}}, a_{\mathrm{NH}}(1), \ldots, a_{\mathrm{NH}}(N_a)]$.

To reduced the amount of information that is to be embedded, the representation vector $\mathbf{a}$ is further quantized to a closest entry in a VQ codebook. We formulate this process as

$$\bar{\mathbf{a}}_{\mathrm{id}} = Q(\mathbf{a}), \quad \bar{\mathbf{a}}_{\mathrm{id}} \in \mathbf{A}, \tag{47}$$

where $Q(\cdot)$ denotes the quantization operation and $\mathbf{A}$ is a 2^M level, $N_a + 1$ dimensional VQ codebook, i.e., $\mathbf{A} = \{\bar{\mathbf{a}}_j; j = 0, \ldots, 2^M - 1\}$. The generation of the VQ codebook is performed by the popular binary-split Linde–Buzo–Gray training (LBG) algorithm [12]. The binary representation of the index, i.e., $(b_0 b_1 \cdots b_m \cdots b_{M-1})$ is embedded into $x_{\mathrm{NB}}(k)$ by the proposed chaotic data hiding method to produce a composite signal $x'_{\mathrm{NB}}(k)$ that can be transmitted through an NB channel.

At the receiver end, the observed signal, i.e., $\hat{x}'_{\mathrm{NB}}(k)$, is partitioned into frames in the same way as at the transmitter. When the hidden information is extracted, the entry index, i.e., id, can be obtained and the corresponding $\bar{\mathbf{a}}_{\mathrm{id}}$ is retrieved from the codebook. Meanwhile, the excitation signal is obtained as the residual of an LPC analysis on the received NB signal, i.e.,

$$r(k) = \hat{x}'_{\mathrm{NB}}(k) - \sum_{i=1}^{N_a} a_{\mathrm{NB}}(i) \cdot \hat{x}'_{\mathrm{NB}}(k-i), \tag{48}$$

where $r(k)$ is the residual and $a_{\mathrm{NB}}(i)$ denotes the AR coefficients for the received NB signal. For unvoiced sounds, $r(k)$ has a flat spectrum like white noise. For voice

sounds, such as vowels and semi-vowel consonants, $r(k)$ appears as harmonic peaks plus the flat noise-like spectrum. These peaks occur in multiples of the pitch—the fundamental voice frequency of a speaker. Hence, there is no need for the proposed system to have any explicit voiced/unvoiced decision and control.

The all-pole AR filter described by $\bar{\mathbf{a}}_{\mathrm{id}}$ is excited by $r(k)$ to produce $\hat{x}_{\mathrm{NH}}(k)$. The energy of $\hat{x}_{\mathrm{NH}}(k)$ should ideally be equal to the energy of the original HB signal. This is achieved by multiplying $\hat{x}_{\mathrm{NH}}(k)$ by the estimated HB gain, i.e., $\hat{G}_{\mathrm{NH}}$. Since the relative gain G_{rel} can be extracted from $\bar{\mathbf{a}}_{\mathrm{id}}$, the estimated HB gain can be obtained by $\hat{G}_{\mathrm{NH}} = \hat{G}_{\mathrm{NB}} \cdot G_{\mathrm{rel}}$, where $\hat{G}_{\mathrm{NB}}$ is the NB gain estimated from the received NB signal. At this point, $\hat{x}_{\mathrm{NB}}(k)$ and $\hat{x}_{\mathrm{NH}}(k)$ are still the signals sampled at 8 kHz. They should be up-sampled to 16 kHz, the sampling rate of the WB output. In addition, $\hat{x}_{\mathrm{NH}}(k)$ should be shifted to its destination frequency band by a high-pass filter, providing the restored HB signal $\hat{x}_{\mathrm{HB}}(k)$. Finally, the WB speech is generated by adding $\hat{x}_{\mathrm{HB}}(k)$ to the up-sampled NB speech.

6 Experimental Results

Experiments have been carried out to evaluate the proposed method for telephony speech bandwidth extension. The NB speech is segmented into non-overlapped frames. Each frame contains 256 samples so that the frame size is 32 milliseconds (ms).

6.1 Payload of Hidden Channel

As shown in (41), the payload of the hidden channel under AWGN is determined by the number of inaudible components in NB speech, the SNR value of the hidden channel, and the error probability requirement. Therefore, we have to find out the values of these parameters in real practice.

First, the number of inaudible components in NB speech is investigated. Twenty NB speech signals are used in this experiment. The perceptual threshold is estimated by the method in [16]. It is found that the number of inaudible components within one speech frame, i.e., L, varies significantly from 8 to 48. This occurs because each frame has its own perceptual characteristics. A frame carrying voiced fractive phonemes must have a different L than a frame carrying vowels. In average, we have $L = 26$.

Second, the SNR of the hidden channel is evaluated. In particular, Gaussian noise that is distributed is added to NB speech in the time domain with a certain SNR value, i.e., $\mathrm{SNR}_{\mathrm{NB}} = \sigma_x^2/\sigma_d^2$, where σ_d^2 is the noise variance. Then, the SNR of the hidden channel, i.e., $\mathrm{SNR}_{\mathrm{hidden}}$, is obtained by averaging $\frac{T(k)}{(X(k)-\hat{X}(k))^2}$ over all the inaudible components. Figure 7 plots $\mathrm{SNR}_{\mathrm{hidden}}$ versus $\mathrm{SNR}_{\mathrm{NB}}$. It is seen that $\mathrm{SNR}_{\mathrm{hidden}}$ increases approximately linearly with $\mathrm{SNR}_{\mathrm{NB}}$. However, there is a degradation around 12.5 dB in $\mathrm{SNR}_{\mathrm{hidden}}$ compared to $\mathrm{SNR}_{\mathrm{NB}}$. Considering that $\mathrm{SNR}_{\mathrm{NB}}$ varies from 25 to 35 dB in real life [11], $\mathrm{SNR}_{\mathrm{hidden}}$ would be within 12.5 and 22.5 dB.

Given the above parameters, experiments are performed to find out the payload of the hidden channel. We plot the experimental results versus P_0 and L in Fig. 8

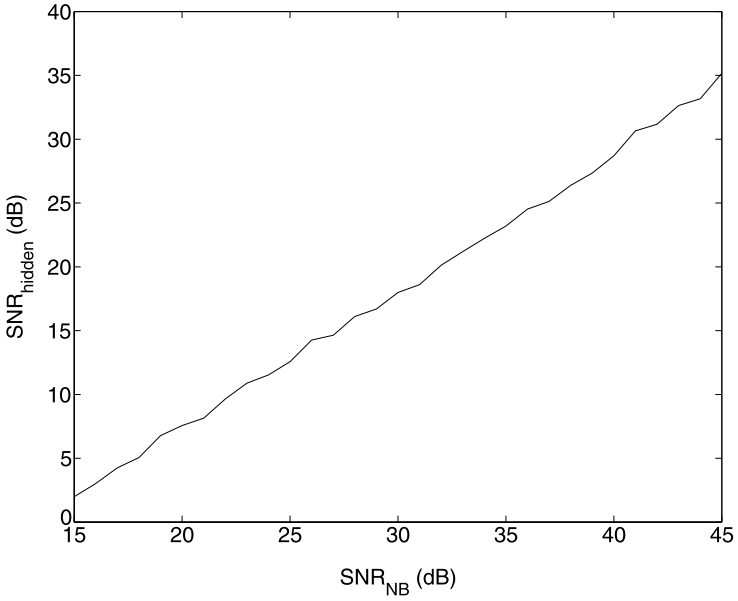


Fig. 7 SNR of the hidden channel versus SNR of NB channel

when the SNR of NB speech is 25 dB and correspondingly $\mathrm{SNR}_{\mathrm{hidden}} = 12.57$ dB. It is seen that when the error probability requirement is $P_0 = 10^{-4}$, we always have $M \geq L$. As P_0 decreases to 10^{-5}, M takes almost the same value as L. If P_0 becomes even smaller, e.g., 10^{-6} or 10^{-7}, the number of data bits that can be transmitted simultaneously is always less than L. Therefore, we improve the robustness of the hidden signal at the cost of payload. We calculate the theoretical payloads based on (41) and find that the simulation results are consistent with the theoretical ones.

Using PCM, the ratio between $T(k)$ and $C\sigma_x^2$ is determined to be around 12.5 dB experimentally. Substituting this value into (38), we plot the theoretical payloads in Fig. 9 versus P_0 and L. The theoretical payloads under PCM are much larger than those under AWGN; i.e., compared to PCM, the hidden signal is more vulnerable to AWGN. To further clarify this, we compare the error probabilities under AWGN and PCM when $M = L$. The SNR of the NB speech is set at 25 dB, which is the worst transmission situation in practice. It is found that $P_{\mathrm{err}}^{\mathrm{AWGN}} = 1.064 \times 10^{-5}$ and $P_{\mathrm{err}}^{\mathrm{PCM}} = 1.053 \times 10^{-8}$. Therefore, the decoding of hidden information under PCM is much more reliable than that under AWGN. In other words, the number of data bits that are to be hidden depends on the payload under the worst AWGN situation, not the one under PCM.

6.2 Perceptual Quality of Composite Signal

The perceptual quality of the composite signal $x'_{\mathrm{NB}}(k)$ after embedding is subjectively assessed by the mean opinion score (MOS). The subjects are asked to compare $x_{\mathrm{NB}}(k)$ with $x'_{\mathrm{NB}}(k)$ in a quiet environment and then provide their results. The MOS scaling is divided into 4 grades as follows:

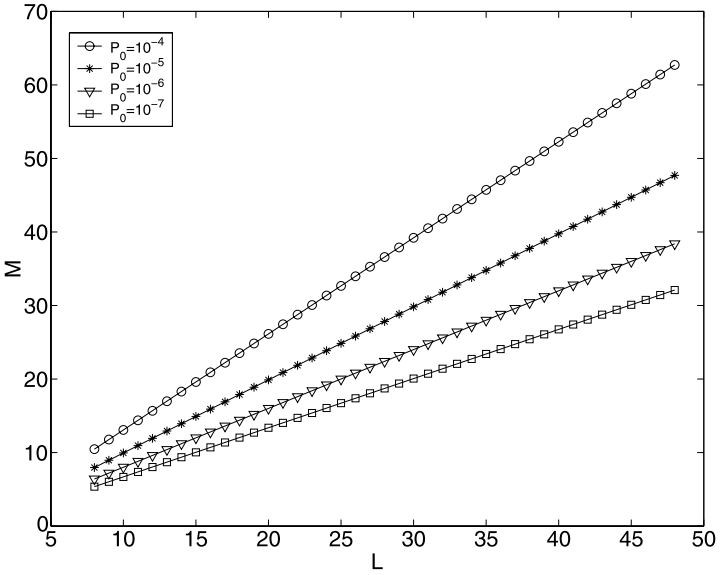


Fig. 8 Payload of the hidden channel versus P_0 and L under AWGN

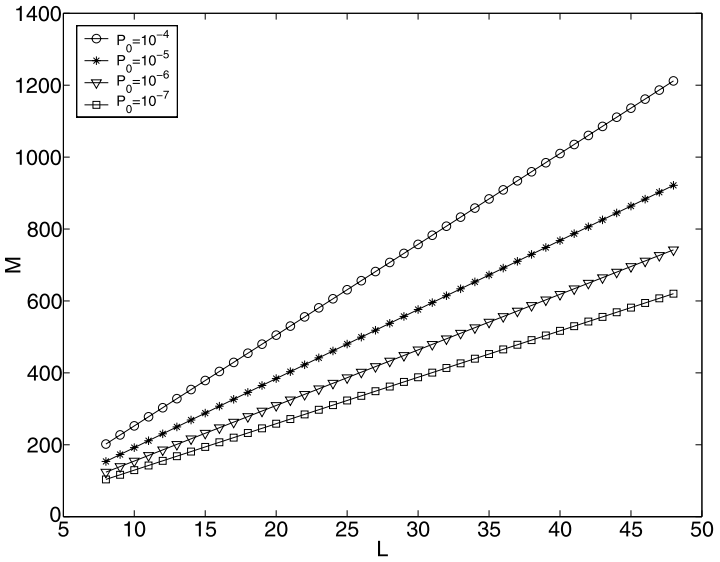


Fig. 9 Payload of the hidden channel versus P_0 and L under PCM

1. The two signals are quite different.
2. The two signals are similar, but the difference is easy to detect.
3. The two signals sound very similar; only a little difference exists.
4. The two signals sound identical.

The two signals are played in random order. The sound pressure level is set at 63 dB. Ten healthy adults, comprising 5 females and 5 males aged 25 to 38, are involved in the test. The subjects are asked to compare $x_{\mathrm{NB}}(k)$ with $x'_{\mathrm{NB}}(k)$ in a quiet environment and then provide their results. The resultant average MOS over all subjects and all test signals is 3.58. Therefore, although the embedding process has some negative impact on the perceptual quality, the degradation is within a tolerable level.

The Diagnostic Rhyme Test (DRT) is also performed to evaluate the intelligibility of the composite signal. The same 10 persons are still employed. Out of 96 word pairs in the DRT list, only those that are able to hide extra information are used for testing. One word is randomly taken from each word pair to generate $x'_{\mathrm{NB}}(k)$, and presented to the subjects. The subjects are then asked to mark on answer sheets which one they think was played. The error rate, defined as the fraction of wrong answers, is computed to evaluate the performance. An average error rate of 12.7% over all the 10 subjects has been found from experiments. This result is close to that of the original NB speech, which is 10.2%.

6.3 Perceptual Quality of the Reconstructed WB Speech

It has been shown in Fig. 8 that even under the worst telephony transmission situation, i.e., $\mathrm{SNR}_{\mathrm{hidden}} = 12.57$ dB, at least 6 data bits can be sent through the hidden channel with an error probability as low as 10^{-7}. Based on this result, we can use 6 digits to encode the entry index of the VQ codebook; i.e., the VQ codebook has a size of $2^6 = 64$. When these 6 digits can be decoded properly at the receiver, an HB signal can be reconstructed and combined with the NB signal to provide a WB signal.

The objective measure used to evaluate the perceptual quality of the reconstructed WB signal is spectral distortion (SD), defined as

$$\mathrm{SD} = \sqrt{\frac{1}{7000}\frac{G_{\mathrm{WB}}^2}{\hat{G}_{\mathrm{WB}}^2}\int_0^{7000}\left|10\log_{10}\frac{|\hat{X}_p(f)|^2}{|X_p(f)|^2}\right|^2 \mathrm{d}f}, \tag{49}$$

where $\hat{X}_p(f)$ and $X_p(f)$ are the reconstructed and the original WB LPC spectra, and $\hat{G}_{\mathrm{WB}}$ and G_{WB} are the gains of the reconstructed and the original WB speech respectively. Figure 10 plots the SD values of using the chaotic data hiding method. For comparison, the SD performance of a conventional ABE method based on codebook mapping [10] is also plotted. In [10], the HB signal is estimated from the NB signal by exploiting the training results from a hidden Markov model (HMM). The training can be either speaker dependent or speaker independent. From Fig. 10, we observe that when $\mathrm{SNR}_{\mathrm{NB}} \geq 25$ dB, the chaotic data hiding method consistently outperforms the conventional ABE method. Only when NB speech is transmitted at 20 dB does the chaotic data hiding method not perform as well as the conventional ABE method with speaker-dependent training. This happens because the number of error bits is increased due to the low SNR of the NB channel. However, considering that the telephony speech is usually transmitted at 25 dB or even higher in real life, chaotic data hiding is a better choice than the conventional ABE method to generate WB speech artificially.

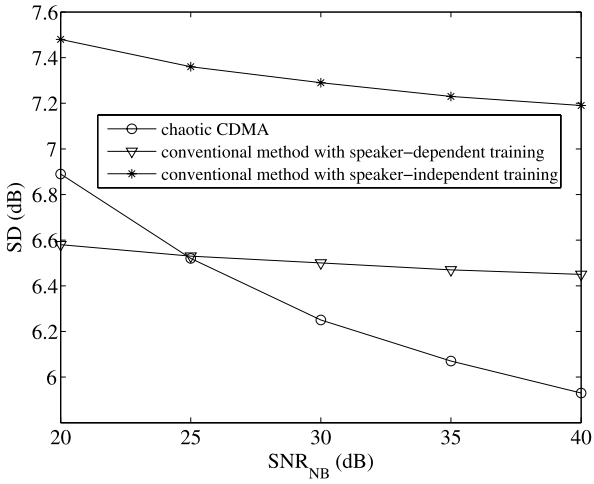


Fig. 10 SD performances of chaotic data hiding and the conventional ABE method

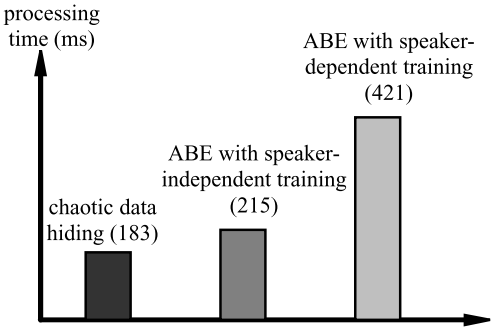


Fig. 11 Illustration of the processing time for chaotic data hiding and the conventional ABE method with both speaker-dependent and speaker-independent training

The subjective evaluation of the perceptual quality of the reconstructed WB speech is also carried out by MOS. The final average MOS is 2.97 when comparing the reconstructed WB by chaotic data hiding with the original WB speech, while the MOS is 1 when comparing the NB and the WB signals. Hence, there is a significant improvement of perceptual quality of the speech signal through the use of the proposed data hiding method.

6.4 Computational Complexity

Computational complexity is measured by the processing time. A PC with Intel Pentium IV 1.5 GHz CPU has been used to run MATLAB programs. Although using C or C++ may be more appropriate and faster than using MATLAB, the relative time difference between the tested schemes will not change much. Figure 11 illustrates the results. Compared to the conventional ABE method [10], the proposed chaotic data hiding method is found to have absolute superiority in processing speed, especially when compared to the conventional ABE method with speaker-dependent training.

7 Conclusion

In this paper, a novel bandwidth extension scheme for communication is proposed based on chaotic data hiding. A hidden channel is created by removing the inaudible components from an NB audio/video signal to transmit the out-of-band information. Since the hidden channel is imperceptible to humans, the embedded signal will not degrade the perceptual quality of the NB signal. At the receiver, when the signal extracted from the hidden channel is decoded, the out-of-band information can be restored and used to reconstruct a WB signal.

The hidden channel is in the least significant part of NB speech, and is thus vulnerable to noise corruption. The proposed chaotic data hiding is found to be robust to channel noise and quantization error while achieving a high payload. We then apply the proposed method to a real application—telephony speech bandwidth extension. Both theoretical and experimental results show that the proposed method is effective and the resulting payload is high enough to transmit all the out-of-band information required. Compared to the conventional ABE methods, the proposed chaotic CDMA method is able to artificially generate WB speech with a better quality and a reduced computational cost.

References

1. W. Bender, D. Gruhl, N. Morimoto, A. Lu, Techniques for data hiding. IBM Syst. J. **35**(3–4), 313–336 (1996)
2. C.-C. Chen, K. Yao, K. Umeno, E. Biglieri, Design of spread spectrum sequences using chaotic dynamical systems and ergodic theory. IEEE Trans. Circuits Syst.-I: Fundam. Theory Appl. **48**(9), 1110–1114 (2001)
3. Y.M. Cheng, D. O'Shaughnessy, P. Mermelstein, Statistical recovery of wideband speech from narrowband speech. IEEE Trans. Speech Audio Process. **2**(4), 544–548 (1994)
4. W. Feller, *An Introduction to Probability Theory and Its Applications*, 3rd edn. (Wiley, New York, 1970)
5. L. Hanzo, F.C.A. Somerville, J.P. Woodard, *Voice Compression and Communications: Principles and Applications for Fixed and Wireless Channels* (IEEE Press, Hoboken, 2001)
6. M. Hosoki, T. Nagai, A. Kurematsu, Speech signal bandwidth extension and noise removal using subband HMM, in *Proc. of IEEE ICASSP*, Orlando, FL, USA, May 2002
7. International Telecommunications Union, Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP), ITU-T Recommendation G. 729, March 1996
8. B. Iser, W. Minker, G. Schmidt, *Bandwidth Extension of Speech Signals* (Springer, New York, 2007)
9. P. Jax, P. Vary, An upper bound on the quality of artificial bandwidth extension of narrowband speech signals, in *Proc. of IEEE ICASSP*, Orlando, FL, USA, May 2002
10. P. Jax, P. Vary, On artificial bandwidth extension of telephone speech. Signal Process. **83**, 1707–1710 (2003)
11. B.E. Keiser, E. Strange, *Digital Telephony and Network Integration* (Van Nostrand Reinhold, New York, 1995)
12. Y. Linde, A. Buzo, R.M. Gray, An algorithm for vector quantizer design. IEEE Trans. Commun. **28**, 84–95 (1980)
13. J.D. Markel, A.H. Gray, *Linear Prediction of Speech* (Springer, Berlin, 1976)
14. G. Mazzini, G. Setti, R. Rovatti, Chaotic complex spreading sequences for asynchronous DS-CDMA—Part I: System modeling and results. IEEE Trans. Circuits Syst.-I: Fundam. Theory Appl. **44**(10) (1997)
15. M. Nilsson, W.B. Kleijn, Avoiding over-estimation in bandwidth extension of telephony speech, in *Proc. of IEEE ICASSP*, Istanbul, June 2000
16. T. Painter, A. Spanias, Perceptual coding of digital audio. Proc. IEEE **88**(4), 451–513 (2000)

17. J.G. Proakis, *Digital Communications*, 2nd edn. (McGraw-Hill, New York, 1989)
18. S.K. Shanmugam, V. Varadan, H. Leung, Design of 1-D piece-wise maps for multi-user CDMA communications using a novel GP-DNA algorithm, in *Recent Advances in Computers, Computing and Communications* (WSEAS, Athens, 2002), pp. 490–495
19. M.D. Swanson, B. Zhu, A.H. Tewfik, Data hiding for video-in-video, in *Proc. Int. Conf. Image Processing (ICIP)*, vol. 2, Santa Barbara, CA, USA, October 1997, pp. 676–679
20. V. Varadan, H. Leung, Design of piecewise maps for chaotic spread spectrum communications using genetic programming. IEEE Trans. Circuits Syst.-I: Fundam. Theory Appl. **49**(11), 1543–1553 (2002)
21. M. Wu, B. Liu, *Multimedia Data Hiding* (Springer, New York, 2003)
22. A.B. Watson, DCT quantization matrices visually optimized for individual images, in *Human Vision, Visual Processing, and Digital Display IV*. Proc. SPIE, vol. 1913 (1993), pp. 202–216,