A universal data transfer technique over voice channels of cellular mobile communication networks

Tikui Zhang | Sensen Li | Bin Yu

Zhengzhou Information Science and Technology Institute, No. 62 Science Avenue, High-Tech Zone, Zhengzhou, China

Correspondence
Sensen Li, Zhengzhou Information Science and Technology Institute, No. 62 Science Avenue, High-Tech Zone, Zhengzhou, Henan, China. Email: vestincloud@163.com

Abstract
A universal technique for transmitting data over voice (DoV) links that can be applied to various cellular networks is proposed. In the scheme, a generic modulation model is established based on waveform symbol mapping, where the data to be transmitted are mapped into finite waveform symbols of a codebook generated offline and demodulated with soft decision results at the receiving end. To make the proposed scheme applicable to various kinds of cellular networks, first sinusoidal signals are selected to synthesise waveform symbols due to their stable transmission characteristics over various voice channels. Then, an analytical method based on Surface Packing is proposed to optimise modulation codebooks. Finally, the target demodulation codebook is obtained through learning from the modulation codebook online. Simulation results show that the proposed scheme performs well over voice channels with different vocoders and has low symbol error rate on various vocoders. Compared with previous schemes designed for some specific vocoders in global system for mobile communication (GSM), the proposed scheme can extend the application scope of DoV to cross-network scenarios consisting of GSM, universal mobile telecommunications system (UMTS) and long-term evolution (LTE).

1 | INTRODUCTION

Cellular mobile networks, especially the fifth generation (5G) networks, have been developing and evolving rapidly over the past several decades. However, it is noticeable that the main types of networks in service in some countries and regions are still the second generation (2G) and the third generation (3G) networks, the data transmission efficiency of which is much more inferior than the more advanced ones. Unlike data transmission, which is provided mainly in modern networks, real-time voice communication is a relatively more basic service in all kinds of cellular mobile networks and therefore occupies a wider range of service areas. Besides, compared to the best-effort features of data transmission services, the voice services possess much higher priority and timeliness. Therefore, when a secret call is required, the encryption scheme based on voice service can be applied to a wider range of regions and networks and is more reliable than that based on data service.

In many cellular mobile communication networks, voice encryption is only supported during the over-the-air transmission stage and is controlled by network operators, thus the voice is transmitted in plain mode in the core networks. As a result, secret calls need to be implemented by end-to-end voice encryption schemes. But during the over-the-air transmission stage, only the signals with speech characteristics can pass through the vocoders located at the mobile terminals and the base stations. Therefore, the encrypted voice data that lose speech characteristics needs to be transmitted over the voice channels of cellular mobile networks by the data over voice (DoV) technique. In a regular implementation of DoV, data are modulated into speech-like signals and transmitted over voice channels of cellular mobile networks transparently in camouflage, which would be demodulated at the receiving end. For convenience of description, the voice channels of cellular mobile networks will be referred to as voice channels in the following.

Apart from voice encryption, the DoV technique can be further applied to the field of instant messaging, which can be regarded as an emergency supplement to cellular network data links. The pan-European eCall vehicle emergency call...
system [7] could extend the E112 system based on the voice channels of GSM networks and automatically transmit data to the public safety answering node in an emergency. The WikiWalk project [8] could provide wizards for pedestrians by sending data commands over the voice channels. The Federal Communications Commission (FCC), a subsidiary of the US government, requires that the emergency transmission of Baudot codes be provided on all 911 emergency calls. [9] The applicable fields of DoV technique also include ATM and POS terminals with high real-time communication requirements in the financial field. However, it should be clear that the research on DoV is not intended to compete with the data services of cellular mobile communication networks, but only acts as an emergency supplement.

To allow greater channel capacity, vocoders are used in the voice channels to perform compression coding on the speech signals transmitted through them. The source signals are encoded into speech parameters at the sending end and reconstructed by the vocoder at the receiving end according to those parameters. [7] These reconstructed speech signals just sound like the original ones, but have inevitable waveform distortions. [7] At the same time, in order to improve the voice channel efficiency, techniques such as voice activation detection (VAD), discontinuous transmission (DTX), noise cancellation and echo cancellation are widely employed in vocoders, [1, 6] which will further aggravate the signals’ waveform distortions or even suppress their transmission. Thus, there are two important issues to address in order to implement DoV. The first is to design speech-like signals that could be transmitted over voice channels with slight distortions, and the second is to effectively deal with the speech-like signals’ waveform distortions and demodulate data accurately.

According to the status quo that multi-generation and multi-type cellular mobile networks coexist in many countries and regions, we focus on designing a universal DoV technique, aiming to provide technical support for transmitting encrypted voice and emergency DoV channels. We make the following contributions.

• We design a modulation model with soft decision in demodulation. Based on the maximum likelihood rule, the received signals are demodulated into the most likely multiple symbols with their corresponding probability values, which provides an excellent support for reducing the error rate by channel coding.

• We propose a method to optimise modulation codebooks offline based on Surface Packing. First, the codebook’s minimum distance and the Surface Packing are defined, respectively. Then, the analytical relationship between the optimal minimum distance value of a codebook and the parameters used to generate the codebook is presented. Based on these analytic relationships, the Surface Packing makes the waveform symbols in a codebook distributed evenly on the surface of a unit sphere in a multi-dimensional Euclidean space, which makes it possible to keep the received symbols as separable as possible. And compared with the previous heuristic-based optimisation algorithms, the computational cost of this scheme is much smaller.

• We design a learning algorithm to obtain the demodulation codebook online. In the algorithm, the waveform symbols of the demodulation codebook are made from the stable outputs of their corresponding symbols in the modulation codebook used in operation. This algorithm makes the same modulation codebook could be used over different vocoders, which is a prerequisite for the proposed DoV scheme to be applied to various types of cellular mobile networks.

It should be noted that the standard modem scheme over Public Telephone Networks (PSTNs) could achieve error-free transmission rates up to 56,000 bits per second (bps), [10] while the rate of the proposed scheme is only about 500–2,000 bps with higher bit error rates (BERs). However, Katugampala et al. [5] pointed out that the standard data modem over PSTNs cannot be employed for transmitting data over lossy voice compression channels of cellular mobile networks.

The remaining is organised as follows. In Section 2, the relevant works on DoV are introduced. In Section 3, a universal modulation model with soft decision in demodulation is presented. The codebook generation and optimisation are described in Section 4. The simulation setup and results are provided in Section 5. We conclude the paper in Section 6.

2 RELATED WORKS

Generally, the methods for implementing DoV could be classified into three categories, which are parameter mapping, parameter modulation and waveform symbol mapping.

2.1 Parameter mapping

The parameter mapping method generate speech-like signals dynamically by mapping the data to be transmitted into parameters which are extracted at the receiving end. Katugampala et al. [5, 11, 12] mapped data to pitch, energy and linear spectral frequencies (LSFs) to implement DoV, while Rashidi et al. [13–15] used parameters such as speech formants, pitch and phases. The speech-like signals synthesised by such methods were prone to have phase and amplitude mutations at the splicing, which included high frequency harmonics and further led to high symbol error rate (SER) in demodulation. Kotnik et al. [16] selected two sets of LSFs to synthesise new LSFs by mapping the data to be transmitted bit-by-bit, and then used a predefined signal to excite the filters composed of the synthesised LSFs to generate speech-like signals. However, the signals generated by this scheme do not necessarily possess speech features. Kondoz et al. [17] mapped data to pulse positions based on linear algebraic code-excited speech coding and data interleaving techniques, and Werner et al. [18] mapped data to bipolar pulses. In these two schemes high data transfer rates could be achieved, but the polar pulses contain many harmonics, which will lead to a high BER. Boloursaz et al. [19] reduced the BER by repeating coding and clustering, but resulted in a lower data transmission rate.
Similar to the above-mentioned methods, some speech processing technologies in the field of speech recognition provide some possible alternative methods for DoV’s implementation. Lojka et al. [20] developed an acoustic events detection system with modified viterbi decoder operating over hidden Markov models (HMM). The feature mechanism of the system can be employed to train the acoustic models and extract feature vectors from the audio signals during demodulating for DoV. Vavrek et al. [21] proposed a weighted fast sequential dynamic time warping (WFSDTW) algorithm. The acoustic modelling method of the scheme can be employed to build phonetic units, which could be transmitted over cellular mobile networks and the synthetic phonetic units could be identified by WFSDTW at the receiving end.

2.2 Parameter modulation

The parameter modulation method is similar to the conventional data modulation, in which the data to be transmitted are modulated into amplitudes, frequencies and phases of the carrier signals, and performs data demodulation at the receiving end. Dhananjay et al. [22] modulated data to a combination of two frequency sinusoidal signals, which can achieve self-synchronisation at the receiving end. However, when the data transfer rate is higher than 1000 bps, high frequencies of the carriers are needed, which will cause severe waveform distortions and high BER. Chmayssani et al. [23] used frequency shift keying (FSK) and Quadrature Amplitude Modulation (QAM) to implement DoV, respectively, and concluded that the VAD and DTX had great impacts on the performance of QAM. Ali et al. [8] described a modulation scheme based on M-ary frequency shift keying (M-PSK), but the data transfer rate is no more than 800 bps. Sheikh et al. [24] implemented a modulator based on quadrature phase shift keying (QPSK). Xu [25] proposed two schemes based on binary phase shift keying (BPSK) and QPSK, respectively. As the channel utilisation is low in these schemes, the BER is very high when the data are transferred at a high rate.

2.3 Waveform symbol mapping

In the waveform symbol mapping method, the data to be transmitted are mapped into speech-like waveform symbols of a codebook generated offline. LaDue et al. [26] first proposed the waveform symbol mapping method for GSM and used the cooperative genetic algorithm (GA) to perform codebook optimisation. However, the optimisation algorithm converges slowly and requires large amount of calculations. Sapozhnykov et al. [27] introduced a pattern search (PS) algorithm to improve the speed of codebook optimisation. But specific codebook optimisations for different vocoders are still needed in their schemes. Shahbazi et al. [28, 29] concluded that the waveform symbols generated by LaDue’s scheme did not necessarily possess speech features, so they generated codebooks based on TIMIT, which is a real speech library. In Shahbazi’s schemes, pitch modifications were required to make the codebook symbols have the same pitch. Mashhadi et al. [30] pointed out that the pitch modification might adversely affect the transmission of the speech-like signals and removed them from codebook generation. In the schemes using real speech libraries, the signals synthesised from the codebook symbols do not necessarily have speech characteristics as a whole.

Compared to the DoVs with the waveform symbol mapping method, those based on the parameter mapping or parameter modulation method have higher BERs and require more complex calculations in the frequency domain. Therefore, the waveform symbol mapping method is chosen to design a DoV scheme for cellular mobile communication networks.

In addition, most of the schemes are optimised for some specific vocoders in GSM in previous research on DoV. When these schemes are applied to various kinds of cellular networks, the speech-like signals will suffer from severe waveform distortions, which will lead to high SERs. To improve this situation, we propose a scheme suitable to various types of cellular networks.

3 Proposed Modulation Model

To accurately describe the general voice channel’s transmission characteristics, we abstract it into a channel composed of a discrete memoryless component and a memory component. For a speech symbol \( y_j \) with \( n \) samples in length, its output \( y'_j \) over such a voice channel is represented by

\[
y'_j = \Psi(y_j) + \xi(y'_{j-1}, y'_{j-2}, \ldots, y'_{j-p}),
\]

where \( \Psi(y_j) \) denotes the stable output of \( y_j \) over the memoryless component, which is only related to \( y_j \), while \( \xi(\cdot) \) denotes the dynamic output of \( y_j \) over the long-time memory component, which is related to the precious received \( p \) symbols prior to \( y'_j \).

Using the waveform symbol mapping method, a universal DoV modulation model combined with soft decision in demodulation as shown in Figure 1 is established based on (1). Where \( i \) denotes the data to be transmitted, \( C^M_m \) denotes the modulation codebook generated offline, which contains \( m \) speech-like waveform symbols with \( n \) samples in length, \( C^D_m \) denotes the demodulation codebook, and its acquisition will be described in detail in Section 4. The recorder records the latest \( p \) received symbols, while the compensator compensates for the distortion of the received signals. \( \tilde{r}_k : p_k \) denotes the demodulation output set with soft decision, and \( \tilde{r}_k : p_k \) represents the probability of demodulating the received signal \( y''_j \) to \( r_k \).

In operation, the modulator maps \( y_j \) into one symbol \( s_j \) in the codebook \( C^M_m \) indexed by the data \( i \). The compensator compensates the output \( y'_j \) of \( y_j \) based on the previous \( p \) symbols recorded by the recorder, and its output is denoted by \( y''_j \). The demodulator gives the probabilities of demodulating \( y''_j \) into the most probable \( v \) waveform symbols in \( C^D_m \) according to the maximum likelihood rule.
For the symbol \( y_j \) in Figure 1, its output over a voice channel \( y''_j \) is demodulated by

\[
\{ \hat{i}_k : p_k, k = 1, 2, ..., v \} = \text{arg top}_v \left[ P(sD^i | y''_j), v \right],
\]

where \( sD^i \in C_{\text{doa} \times n}^D \), \( \text{arg top}_v \) denotes getting the indexes of the top \( v \) elements of vector \( x \) in descending order of their values.

The \( sD^i \) and \( y''_j \) can be regarded as two \( N \)-dimensional vectors, whose similarities are denoted by cosine similarity. Therefore, \( P(sD^i | y''_j) \) can be equivalently calculated by

\[
P(sD^i | y''_j) \approx \cos(sD^i, y''_j) = \frac{sD^i \cdot y''_j}{\|sD^i\| \|y''_j\|},
\]

where \( \| \cdot \| \) denotes the Euclidean norm calculation.

In the study of phonetics, the p-order all-pole model is generally used as the speech generation model. Therefore, the speech compensation could be performed by

\[
s(n) = s(n) - \sum_{i=1}^{p} \alpha_i \cdot s(n-i),
\]

where \( \alpha_i \) denotes the influence coefficient of the previous \( i \)th samples on the current sample. The values of \( \alpha_i \) are determined by autoregressive model, and Wan et al. [31] described it in detail.

It needs to be clear that the compensation and demodulation operations in the modulation model need to be performed on the basis of accurate signal synchronisations. Due to the distortion of the speech-like signals, it is difficult, especially in the scenarios of different cellular networks, to achieve precise synchronisations by the traditional synchronisation methods. Here the focus is on the general techniques of modulation and demodulation, and the signal synchronisation is studied as a separate work. [32] Wherein, we reviewed that Kotnik et al. [16] employed chirp signals as the synchronisation signals for their DoV scheme. However, such a scheme was prone to missing synchronisation due to the weakening of the correlation of the synchronisation signals caused by the vocoders in cross-network scenarios. Therefore, we proposed a synchronous scheme combining time domain analysis with fast correlation method in [32].

### 4 CODEBOOK GENERATION AND OPTIMISATION

**Definition 1.** The minimum value of the Euclidean distances between any two symbols in codebook \( C \), is defined as the minimum distance of \( C \), denoted as \( d(C) \).

**Definition 2.** For a codebook \( C \) and a symbol \( s(\notin C) \), the minimum distance between \( s \) and \( C \), denoted as \( d(s, C) \), refers to the minimum value of the Euclidean distances between \( s \) and all of the symbols in \( C \).

**Definition 3.** The maximum cosine similarities of a codebook \( C \), denoted as \( \cos(C) \), represents the maximum value of the cosine similarities between any two symbols in \( C \).

**Definition 4.** For a codebook \( C \), its optimisation has the following propositions

\[
\min d(C)
\]

s.t. \( C = \{ s_i \mid i = 1, 2, ..., m \} \),

\[
s_i = (s_{i0}, s_{i1}, ..., s_{in}),
\]

\[
\| s_i \| = 1.
\]

**Definition 5.** For \( m \) points and a unit sphere in a multi-dimensional space, the process of distributing these points uniformly on the surface of this sphere is defined as Surface Packing.

**Theorem 1.** If the Euclidean norm of any symbol in codebook \( C \) is 1, then the maximisation of \( d(C) \) is equivalent to the minimisation of \( \cos(C) \).
Theorem 1 holds.

4.1 Generation of waveform symbols

Since sinusoidal signals can be stably transmitted over various types of voice channels, we choose a series of sinusoidal signals with equal frequency intervals in the voice frequencies to synthesise the waveform symbols. The analog form $e_i(t)$ of the waveform symbols in the codebook $C_{MxN}$ is represented by

$$e_i(t) = \sum_{j=1}^{q} a_{ij} \cdot \sin(2\pi f_j t),$$

where $q$ stands for the number of subcarriers, $f_j(= f_0 + (j - 1) \cdot \Delta f)$ represents the frequency of the $j$th subcarrier, and $a_{ij} \in [-1, 1]$ denotes the coefficient of the $j$th subcarrier in the $i$th waveform symbol. We use $f_0$ and $\Delta f$ to indicate the starting frequency of the subcarrier and the interval of the subcarrier frequency, respectively.

Let the sampling frequency be $f_s$, then the sampling values in $X_i(= (X_{i0}, X_{i1}, ..., X_{in-1}))$ of $e_i(t)$ starting from $t = 0$ is denoted by

$$X_{ik} = \sum_{j=1}^{q} a_{ij} \cdot \sin(2\pi k f_j f_s), \quad k = 0, 1, 2, ..., n - 1. \quad (9)$$

The waveform symbol $s_i$ is obtained by normalising $X_i$ as

$$s_i = \frac{X_i}{\|X_i\|}. \quad (10)$$

4.2 Optimisation for modulation codebook

The performance of a DoV scheme is mainly evaluated by the data transfer rate and the BER. For a codebook $C_{MxN}$, the data transfer rate $r$ is calculated by

$$r = \frac{f_s}{n} \cdot \log_2 m. \quad (11)$$

Since $f_s$ is only related to the codebook sizes of $m$ and $n$, therefore, we take minimising the BER at various data transfer rate as the goal of codebook optimisation.

When $\hat{\ell}_t$ is used as the hard decision result in demodulation as shown in (2), the BER in probability is calculated by

$$P_t = \frac{1}{m} \sum_{j=1}^{m} P_{ij}, \quad (12)$$

where $P_{ij}$ denotes the probability of demodulation misjudgment of symbol $s_i$ after being transmitted.

LaDue et al. [26] considered that the output signal $s_i$ of $s_j$ over a voice channel conforms to the Gaussian distribution, while Kazemi et al. [33] assumed that $s_j$ conforms to the Weibull distribution and the chi-square distribution. Regardless of which distribution is most realistic, $s_j$ is scattered around a statistical mean $\hat{s}_i$. If the distribution is in a limited range, $P_t$ takes its minimum value when the proposition shown in Definition 4 has its optimal value.

According to Shannon, [34] the waveform symbols in a codebook $C_{MxN}$ can be regarded as $m$ points on the surface of a unit sphere in the $n$-dimensional Euclidean space

$$\mathbb{S}^{n-1} = \{ \mathbf{s} \in \mathbb{R}^n \ | \ \| \mathbf{s} \| = 1 \}. \quad (13)$$

Wherein, the center of $\mathbb{S}^{n-1}$ is located at the origin of the space $\mathbb{R}^n$.

If there is no constraint on $s_i$, the essence of the proposition shown in Definition 4 is a Surface Packing problem as shown in Definition 5. According to Leopardi, [35] the solution for it is to divide the spherical surface $\mathbb{S}^{n-1}$ into $m$ parts evenly, and locate $s_i$ at the center of each area. Leopardi [35] studied this issue through his whole doctoral thesis. For an illustrative example shown in Figure 2, the surface of an unit sphere in the three-dimensional space is divided into 32 equal-area regions according to his scheme.

Due to the constraints of the symbols shown in (9), the proposition of Definition 4 cannot be solved directly using Leopardi’s scheme. Suppose that the subcarrier coefficient $a_{ij}$ of $s_j$ as a point $\mathbf{a}(= a_{i1}, a_{i2}, ..., a_{iq})$ in the space $\mathbb{R}^q$ and normalise it by (10), it can be divided into two steps of determining the optimal frequency interval and searching for the optimal number of subcarriers to optimise the modulation codebook based on Surface Packing.
4.2.1 Determining the optimal frequency interval

Given \( m \) and \( n \), and ensuring that \( f_j \) is within the speech frequency range of [300, 3400]Hz by initialising \( f_0 \) and \( q \), the optimal deployment point \( a^*_i \) can be obtained by the \( q \)-dimensional Surface Packing, and then the codebook \( C \) and \( \cos(C) \) can be calculated respectively. Taking minimising \( \cos(C) \) as the goal, a one-dimensional searching for the frequency interval \( \Delta f \) can be carried out by Algorithm 1.

Based on the calculations of Algorithm 1, the relationship between \( \Delta f \) and \( \cos(C) \) is shown in Figure 3. It can be seen from that, when \( q \geq m/2, \cos(C) \) takes its minimum values, and when \( f_0 = f_j/n \), the minimum value of \( \cos(C) \) is 0. It should be clear that, when \( q < m/2 \), the value of \( \cos(C) \) in Figure 4 is not necessarily the optimal value under its condition.

According to Algorithm 2, the relationship between \( q \) and \( \cos(C) \) is shown in Figure 4. It can be seen from that, when \( q \geq m/2, \cos(C) \) takes its minimum values, and when \( f_0 = f_j/n \), the minimum value of \( \cos(C) \) is 0. It should be clear that, when \( q < m/2 \), the value of \( \cos(C) \) in Figure 4 is not necessarily the optimal value under its condition.

At various data transfer rates, the optimal parameters used to generate codebooks are listed in Table 1. Taking the parameters that \( m = 16 \) and \( n = 32 \) listed in Table 1 as an illustrative example, some waveform symbols of codebook generated by these parameters are shown in Figure 5.

4.3 Demodulation codebook acquisition via online learning

For a waveform symbol \( y \), its stable output over a voice channel can be calculated by

\[
E(Y) = E(y^\prime) - E(\xi) \tag{14}
\]

according to (1).
FIGURE 4 The relationship between $q$ and $\cos(C)$

TABLE 1 Optimal parameters used to generate codebooks

| $n$ | $m$ | $r$ | $f_0$ | $\Delta f$ | $q$ | $\cos(C)$ |
|-----|-----|-----|-------|-----------|-----|-----------|
| 16  | 8   | 1500| 500   | 500       | 4   | 0         |
| 16  | 16  | 2000| 500   | 250       | 8   | 0         |
| 20  | 8   | 1200| 400   | 400       | 4   | 0         |
| 20  | 16  | 1600| 400   | 200       | 8   | 0         |
| 20  | 32  | 2000| 400   | 180       | 16  | 0.17      |
| 32  | 16  | 1000| 500   | 250       | 8   | 0         |
| 32  | 32  | 1250| 500   | 125       | 16  | 0         |
| 40  | 16  | 800 | 400   | 200       | 8   | 0         |
| 40  | 32  | 1000| 400   | 100       | 16  | 0         |
| 40  | 32  | 1000| 400   | 200       | 16  | 0         |
| 40  | 64  | 1200| 400   | 90        | 32  | 0.14      |
| 80  | 32  | 500 | 400   | 100       | 16  | 0         |

Where $f_s$ is 8 kHz.

Suppose that in practice the previous $p$ waveform symbols of $y$ conform to a uniform random distribution, if the amount of transmission of $y$ is very large, then $E(\xi) \to 0$. Therefore, (14) can be simplified as

$$E(Y) = E(y') = \frac{\sum_{i=1}^{N_y} y'_i}{N_y}, \quad N_y \to \infty,$$

(15)

where $N_y$ denotes the amount of transmission of $y$, and the addition operation on $y'$ is processed as vectors.

According to (15), the demodulation codebook could be obtained by online learning, and a demodulation codebook learning model is established as shown in Figure 6. Where $D[d_j \in [1, m]] j = 1, 2, \ldots$ denotes a predefined random number set, and $y_j$ denotes a copy of the $k^{th} = (d_j)$ symbol in the codebook $C_{mol}^M$.

In operation, the modulator takes $d_j$ sequentially from $D$ and maps it into $y_j$ at the sending end. The codebook renovator learns and updates the demodulation codebook $C_{mol}^D$ based on the received signals $y_j'$ and $D$. Each symbol is sent $N_y$ times, and the total amount of data to be sent during the learning process is $m \cdot N_y$. Based on (15), a demodulation codebook learning algorithm is given in Algorithm 3.

5.1 Simulation settings

According to the proposed DoV modulation model, a simulated platform is built on Matlab with two pairs of vocoders employed, as shown in Figure 7. Where $D$ is generated randomly in $[1, m]$, and its front part is used for demodulation.
codebook learning and the latter part is for SER test, $C_m^{M \times n}$ is generated offline and $C_m^{D \times n}$ is obtained online via Algorithm 3. Both pairs of the vocoders contain an encoder and a decoder. The channel encoder and decoder are used to make use of the soft decision results in demodulation to improve the demodulation accuracy.

The vocoders currently employed by various types of cellular mobile networks are shown in Table 2.

Since HR is rarely used currently and the ANSI source codes of the vocoders except EVRC and EVRC-B are available from ETSI, therefore, we choose FR, EFR, AMR and EVS to construct the simulation scenarios. Wherein, the sampling frequency of EVS is set to 8 kHz.

Given the parameters shown in Table 1, when $\Delta f = k \cdot f_s/n (k = 1, 2, \ldots)$, the generated waveform symbols are phase-continuous, which can reduce the waveform distortion of the speech-like signals. Therefore, the optimal parameters employed to generate codebooks are set at each data transfer rate as shown in Table 3.

In the channel coding scheme, $b$ symbols are grouped into a group, which could represent $(b \cdot R - c)$ bits. The first $(b \cdot R - c)$ bits are set to the data to be transmitted, and the last $c$ bits are set to the checksum of the first $(b \cdot R - c)$ bits. In the simulation, $b$ is set to 5 and $c$ is set to 2. It needs to be clear that this channel coding scheme is just a simple demonstration of using the soft decision. Some error correction mechanisms can be used in a dedicated channel coding scheme.

Among other parameters, $f_s$ is set to 8 kHz, $v$ is set to 3, $N_t$ is set to 5, and 1,000,000 symbols will be sent in each test.

### Table 3: The optimal parameters used to generate codebooks at various data transfer rate

| r(bps) | $n$ | $m$ | $f_s$(Hz) | $\Delta f$(Hz) | $q$ | $f_{q-1}$(Hz) | cos(C) |
|-------|-----|-----|-----------|---------------|----|---------------|--------|
| 500   | 80  | 32  | 400       | 100           | 16 | 1900          | 0      |
| 800   | 40  | 16  | 400       | 200           | 8  | 1800          | 0      |
| 1000  | 32  | 16  | 500       | 250           | 8  | 2250          | 0      |
| 1200  | 20  | 8   | 400       | 400           | 4  | 1600          | 0      |
| 1500  | 16  | 8   | 500       | 500           | 4  | 2000          | 0      |
| 2000  | 16  | 16  | 500       | 250           | 8  | 2250          | 0      |

Where $f_{q-1}$ denotes the highest subcarrier frequency.

### 5.2 Result analysis

When the vocoder pair A and B are set to the same vocoder with the same channel coding rate, the test results of SER at each data transfer rate are shown in Table 4 and Figure 8.

Table 4 and Figure 8 show that the SER are correlated to the data transfer rate and the channel coding rate in the following ways.

- The SER increases as the data transfer rate increases. On the one hand, when the data transfer rate is increased by increasing the number of symbols of the codebooks, the mutual Euclidean distances between the symbols become smaller, which will reduce the symbols’ separability. On the other hand, when the data transfer rate is increased by reducing the symbol length, more harmonics outside the speech spectrum will be generated, which will exacerbate the waveform distortions of the speech-like signals.

- The SER increases as the channel coding rate decreases. At a lower channel coding rate, a higher compression rate will be selected in the vocoders in operation, which leads to a higher information loss rate and more severe distortions of
## TABLE 4
The SER under the same vocoder with the same channel coding rate condition (%)

| Vocoder | 500 bps | 800 bps | 1000 bps | 1200 bps | 1500 bps | 2000 bps |
|---------|---------|---------|----------|----------|----------|----------|
| FR      | 0.0002  | 0.0002  | 0.0236   | 0.0078   | 0.1461   | 0.2119   |
| EFR     | 0.0000  | 0.0000  | 0.0009   | 0.0000   | 0.0012   | 0.0032   |
| AMR 4.75| 2.9331  | 3.4819  | 12.5033  | 7.9386   | 13.3680  | 24.4932  |
| AMR 5.15| 2.3143  | 2.4947  | 6.4243   | 3.2318   | 8.5083   | 18.8318  |
| AMR 5.9 | 1.7423  | 1.5582  | 5.7812   | 2.6423   | 7.0921   | 12.7233  |
| AMR 6.7 | 0.5132  | 0.6389  | 3.7369   | 0.9793   | 2.3217   | 8.0191   |
| AMR 7.4 | 0.1510  | 0.3141  | 0.7513   | 0.3432   | 0.8926   | 4.9123   |
| AMR 7.95| 0.0531  | 0.1525  | 0.4814   | 0.2326   | 0.6013   | 2.0275   |
| AMR 10.2| 0.0008  | 0.1307  | 0.1912   | 0.0618   | 0.1210   | 0.8075   |
| AMR 12.2| 0.0000  | 0.0000  | 0.0000   | 0.0000   | 0.0000   | 0.0034   |
| EVS 7.2 | 0.1416  | 1.8013  | 5.5132   | 1.9124   | 6.1166   | 19.5233  |
| EVS 8   | 0.0013  | 0.4617  | 1.6031   | 1.1152   | 2.9672   | 10.6889  |
| EVS 9.6 | 0.0000  | 0.0512  | 0.0533   | 0.1021   | 0.5041   | 2.9623   |
| EVS 13.2| 0.0000  | 0.0009  | 0.0000   | 0.0263   | 0.0314   | 0.2981   |
| EVS 16.4| 0.0000  | 0.0000  | 0.0000   | 0.0131   | 0.0009   | 0.0019   |
| EVS 24.4| 0.0000  | 0.0000  | 0.0000   | 0.0004   | 0.0007   | 0.0007   |

Where 0.0000 indicates that the SER is less than $1.0 \times 10^{-6}$.

## FIGURE 8
The relationship between SER and data transfer rate under the same vocoder with the same channel coding rate condition.

The speech-like signals. In practical scenarios, the AMR and EVS vocoders will negotiate to use a lower rate of coding rate in environments with poor signals.

It can also be seen from Figure 8 that the SER at the data transfer rate of 1200 bps is not higher than that at 1000 bps when other conditions are the same. Therefore, in the range of 800–1200 bps, the data transfer rate of 1200 bps is preferred.

When the vocoder pair A and B are set to different types of vocoders, the test results of SER at each data transfer rate are shown in Table 5 and Figure 9, which show that in a simulated environment composed of different vocoders, the SER is higher than that of the same vocoder, and there is a high correlation between the vocoders and their channel coding rates used in both parties. If the channel coding rates of both parties are high, the SER could be low. Any party using a lower channel coding rate will result in a higher SER.

### 5.3 Comparison with previous schemes

In terms of codebook optimisation, different algorithms are proposed in [26, 27] and [30]. Among them, the GA algorithm proposed in [26] takes 25,000 iterations to generate the codebook symbols, the pattern search algorithm introduced in [27] needs 8000 iterations, while the algorithm used in [30] still needed a large amount of calculations with a roughly exponential convergence rate. The optimisation algorithm proposed is an analytical method, which has a better performance in the computational cost than the other three.

In terms of performance in SER, as shown in Table 6, the results show that the proposed scheme is superior to [27] over the FR vocoders, and is equivalent to [27] and [28] over the EFR vocoders. Under the AMR 12.2 vocoders conditions, the proposed scheme is slightly inferior to [30], equivalent to [27] and better than [25].

In terms of performance in adaptability in cross-network scenarios, the schemes in previous studies are designed for some specific types of vocoders, which makes it difficult to adapt to different networks. The proposed scheme takes adaptability as an important feature, and the simulation results show that it achieves excellent performance in scenarios composed of various types of vocoders.
TABLE 5  The SER under different types of vocoders condition (%)

| Vocoder A | Vocoder B | 500 bps | 800 bps | 1000 bps | 1200 bps | 1500 bps | 2000 bps |
|-----------|-----------|---------|---------|----------|----------|----------|----------|
| EFR AMR   | AMR 5.15  | 1.5453  | 0.3627  | 13.1728  | 11.8199  | 18.5817  | 25.1446  |
| EFR AMR   | AMR 7.4   | 0.3601  | 0.0018  | 0.0047   | 0.0064   | 0.0136   | 0.0112   |
| EFR AMR   | AMR 12.2  | 0.0812  | 0.0512  | 0.0071   | 0.0064   | 0.0136   | 0.0112   |
| EFR AMR   | EVS 7.2   | 0.8215  | 0.2658  | 2.3731   | 0.4633   | 2.0067   | 10.9289  |
| EFR AMR   | EVS 9.6   | 0.0000  | 0.0000  | 0.0000   | 0.0000   | 0.0000   | 0.0000   |
| EFR AMR   | EVS 13.2  | 0.0000  | 0.0000  | 0.0000   | 0.0000   | 0.0000   | 0.0000   |
| EFR AMR   | EVS 24.4  | 0.0000  | 0.0514  | 1.1128   | 0.2701   | 0.9568   | 4.8972   |
| AMR EVS   | EVS 7.2   | 10.5835 | 16.7787 | 18.4099  | 23.9317  | 25.3812  | 36.8876  |
| AMR EVS   | EVS 24.4  | 0.1025  | 0.3987  | 6.3683   | 0.5723   | 2.9273   | 4.7791   |
| AMR EVS   | EVS 9.6   | 2.1610  | 7.0028  | 4.6227   | 2.0414   | 2.3933   | 3.1482   |
| AMR EVS   | EVS 13.2  | 0.0006  | 0.1801  | 0.0000   | 0.0000   | 0.0000   | 0.0000   |
| AMR EVS   | EVS 24.4  | 0.0000  | 0.0000  | 0.0000   | 0.0000   | 0.0000   | 0.0000   |

Where 0.0000 indicates that the SER is less than 1 × 10⁻⁶.

FIGURE 9  The relationship between SER and data transfer rate under different types of vocoders condition

TABLE 6  The comparison of SERs in various schemes

| Scheme        | Rate (bps) | SER over FR | SER over EFR | SER over AMR12.2 |
|---------------|------------|-------------|--------------|------------------|
| [25]          | 1330       | –           | –            | 1.5 × 10⁻⁴       |
| [27]          | 1000       | 1.0 × 10⁻³  | 2.0 × 10⁻⁶   | 3.0 × 10⁻⁶       |
|               | 1330       | 3.0 × 10⁻³  | 3.0 × 10⁻⁶   | 5.0 × 10⁻⁶       |
|               | 2000       | 1.2 × 10⁻²  | 4.0 × 10⁻⁶   | 6.0 × 10⁻⁶       |
| [28]          | 1200       | –           | 5.0 × 10⁻⁷   | –                |
|               | 2000       | –           | 1.5 × 10⁻⁵   | –                |
| [30]          | 1000       | –           | –            | 6.3 × 10⁻⁸       |
|               | 1200       | –           | –            | 1.4 × 10⁻⁷       |
| Proposed      | 1000       | 2.4 × 10⁻⁴  | 9.0 × 10⁻⁶   | < 1.0 × 10⁻⁵     |
|               | 1200       | 7.8 × 10⁻⁵  | < 1.0 × 10⁻⁶ | < 1.0 × 10⁻⁶     |
|               | 2000       | 2.1 × 10⁻³  | 3.2 × 10⁻⁵   | 3.4 × 10⁻⁵       |

CONCLUSION

This paper studied transmitting digital data through cellular mobile network voice channels. A universal implementation scheme for DoV was constructed based on the waveform symbol mapping method.

In order to make the proposed scheme applicable to various cellular mobile networks, we selected sinusoidal signals to synthesise waveform symbols due to their stable transmission characteristics over various cellular networks. On the other hand, we designed a learning algorithm to obtain demodulation codebooks online, which avoided optimising codebooks for different networks or vocoders.

To achieve high performance in SER, we proposed an analytical algorithm based on Surface Packing to optimise modulation codebooks offline firstly. The proposed algorithm makes the waveform symbols distributed evenly on the surface of a multi-dimensional unit sphere, which ensures that each symbol is as far away as possible from other symbols in the codebook. Then, we used soft decision in demodulation, which could provide an excellent support to reduce SER by channel coding.

Simulation results show that the SER of the proposed scheme is not higher than 6 × 10⁻⁵ at the data transfer rate of 1200 bps between any two vocoders of EFR, AMR1220 and EVS2440. Therefore, the proposed scheme performs well in cross-network scenarios.

ORCID

Tikui Zhang  https://orcid.org/0000-0003-0582-8205

REFERENCES

1. 3GPP: Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; Services and service capabilities. 3rd Generation Partnership Project (3GPP), 2018.
2. 3GPP: Digital cellular telecommunications system (Phase 24) (GSM); Universal Mobile Telecommunications System (UMTS); 3G security; Security architecture. 3rd Generation Partnership Project (3GPP), 2018. 33.102. version 15.1.0

3. 3GPP: Digital cellular telecommunications system (Phase 24) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; 3G security; Access security for IP-based services. 3rd Generation Partnership Project (3GPP), 2018. 33.203. version 15.1.0

4. 3GPP: 5G; Security architecture and procedures for 5G System. 3rd Generation Partnership Project (3GPP), 2018. 33.501

5. Katugampala, N., et al.: Secure voice over GSM and other low bit rate systems. In: Proceedings of the International Conference on IEE Seminar Secure GSM and Beyond: End to End Security for Mobile Communications (IET), pp. 3/1–3/4. IET, London (2003)

6. Nantogjan, C., et al.: A survey of voice and communication protection solutions against wiretapping. Comput. Electr. Eng. 77, 163–178 (2019)

7. 3GPP: Digital cellular telecommunications system (Phase 24) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; Transferring of emergency call data. 3rd Generation Partnership Project (3GPP), 2018. 22.967

8. Ali, R.T., et al.: Data transmission over mobile voice channel based on M-FSK modulation. In: Proceedings of the International Conference on IEEE Wireless Communications and Networking Conference (WCNC), pp. 4416–4421. IEEE, Shanghai, China (2013)

9. 3GPP: Digital cellular telecommunications system (Phase 24) (GSM); Universal Mobile Telecommunications System (UMTS); Cellular text telephone modem; General description. 3rd Generation Partnership Project (3GPP), 2018. 26.226

10. ITU: A digital modem and analogue modem pair for use on the Public Switched Telephone Network (PSTN) at data signalling rates of up to 56 000 bit/s downstream and up to 33 600 bit/s upstream. International Telecommunication Union (ITU), 1998. 90. version 90

11. Katugampala, N.N., et al.: Real-time data transmission over GSM voice channel for secure voice and data applications. In: Proceedings of the International Conference on 2nd IEE Secure Mobile Communications Forum: Exploring the Technical Challenges in Secure GSM and WLAN, London, pp. 7/1–7/4 (2004)

12. Katugampala, N., et al.: Real-time end-to-end secure voice communications over GSM voice channel. In: Proceedings of the International Conference on 13th European Signal Processing Conference, pp. 1–4. IEEE, Antalya, Turkey (2005)

13. Rashidi, M., Sayadian, A.: A new approach for digital data transmission over GSM voice channel. In: Proceedings of the International Conference on 2nd WSEAS International Conference on Circuits, Systems, Signal and Telecommunications, Acapulco, Mexico, pp. 193–196 (2008)

14. Rashidi, M., et al.: Data mapping onto speech-like signal to transmission over the GSM voice channel. In: Proceedings of the International Conference on 40th Southeastern Symposium on System Theory (SSST), pp. 54–58. IEEE, New Orleans, Louisiana (2008)

15. Rashidi, M., et al.: A harmonic approach to data transmission over GSM voice channel. In: Proceedings of the International Conference on 3rd International Conference on Information and Communication Technologies: From Theory to Applications, pp. 1–4. IEEE, United States (2008)

16. Kotnik, B.: Data transmission over GSM voice channel using digital modulation technique based on autoregressive modeling of speech production. Digit. Signal Process. 19(4), 612–627 (2009)

17. Ahmet, K., et al.: Data transmission. US Patent PCT/GB2005/001729, 2007

18. Werner, M.E., et al.: Cellular in-band modem solution for eCall emergency data transmission. In: Proceedings of the International Conference on VTC Spring 2009-IEEE 69th Vehicular Technology Conference, pp. 1–6. IEEE, Barcelona (2009)

19. Boloursaz, M., et al.: Secure data over GSM based on algebraic codebooks. In: Proceedings of the International Conference on East-West Design & Test Symposium (EWDTS 2013), pp. 1–4. IEEE, Rostov-on-Don, Russia (2013)

20. Lojka, M., et al.: Modification of widely used feature vectors for real-time acoustic events detection. In: Proceedings on the International Conference on ELMAR, 2013 55th International Symposium, Zadar, Croatia (2013)

21. Vavrek, J., et al.: Weighted fast sequential DTW for multilingual audio query-by-example retrieval. J. Intell. Inf. Syst. 51(2), 439–455 (2018)

22. Dhananjay, A., et al.: Hermes: data transmission over unknown voice channels. In: Proceedings of the International Conference on Proceedings of the sixteenth annual international conference on Mobile computing and networking, pp. 113–124. ACM, Chicago, Illinois, USA (2010)

23. Chmaysani, T., Baudoin, G.: Data transmission over voice dedicated channels using digital modulations. In: Proceedings of the International Conference on 18th International Conference Radioelektronika, Rzeszów, pp. 1–4. IEEE (2008)

24. Sheikh, J.A., Akhtar, S., Parah, S.A., Bhat, G.: ‘A new method of Haar and Db10 based secured compressed data transmission over GSM voice channel. Proc. Int. Conf. Intelligent Techniques in Signal Processing for Multimedia Security, pp. 401–426. Springer (2017)

25. Xu, Z.: Data transmission method based on single carrier over GSM voice channel. Revista de la Facultad de Ingeniera 32(9), 23–29 (2017)

26. LaDue, C.K., et al.: A data modem for GSM voice channel. IEEE Trans. Veh. Technol. 57(4), 2205–2218 (2008)

27. Sapozhnykov, V.V., Fienberg, K.S.: A low-rate data transfer technique for compressed voice channels. J. Signal Process. Syst. 68(2), 151–170 (2012)

28. Shahbazi, A., et al.: A novel speech-like symbol design for data transmission through GSM voice channel. In: Proceedings of the International Conference on IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), pp. 478–483. IEEE, Ajman, United Arab Emirates (2009)

29. Shahbazi, A., et al.: Data transmission over GSM adaptive multi rate voice channel using speech-like symbols. In: Proceedings of the International Conference on 2010 International Conference on Signal Acquisition and Processing, Chennai, India, pp. 63–67. IEEE (2010)

30. Mashhadial, M.B., Behnia, F.: Efficient codebook design for digital communication through compressed voice channels. IET Commun. 10(18), 2613–2620 (2016)

31. Wang, H., Xiao, L.: Variational Bayesian learning for robust AR modeling with the presence of sparse impulse noise. Digit. Signal Process. 59, 1–8 (2016)

32. Zhang, T., et al.: Research on speech-like signal synchronization scheme of encrypted communication in cellular mobile networks. J. Electron. Meas. Instrum. 34(1), 163–170 (2020)

33. Kazemi, R., et al.: Capacity bounds and detection schemes for data over voice. IEEE Trans. Veh. Technol. 65(11), 8964–8977 (2016)

34. Shannon, C.E.: Probability of error for optimal codes in a Gaussian channel. Bell Syst. Tech. J. 38(3), 611–656 (1959)

35. Leopold, P.: Distributing points on the sphere: partitions, separation, quadrature and energy. PhD thesis, University of New South Wales (2007)