An Efficient Transmission Coding Technique of Digitized Speech Data

Tetsuya Shimamura and Katsuaki Yaguchi
Department of Information and Computer Sciences
Saitama University, 255 Shimo-Okubo, Saitama 338-8570, Japan
Tel: +81-48-858-3496, Fax: +81-48-858-3716
E-mail: shima@sie.ics.saitama-u.ac.jp

Abstract: Speech transmission is common in many communications systems. In this paper, a technique to reduce the total bits required for expressing the speech data is proposed for the purpose of a packet transmission. A novel coding method is derived based on the concept of finding common information in sequential speech samples. Computer simulations demonstrate that the proposed scheme reduces the total bits required in PCM approximately by half.

1 Introduction

Pulse code modulation (PCM) [1] is a source coding technique for analog speech data and has been widely used in digital transmission systems. However, when we send the speech data in a packet form, PCM is not always fruitful. This is because continuous speech sometimes gives similar values for sequential time samples and are sometimes periodic, which are obviously redundant.

From this point of view, we consider how to express the continuous speech data efficiently in this paper. The use of logarithmic PCM, differential PCM (DPCM) and adaptive DPCM (ADPCM) may invoke a more efficient transmission than that of PCM [2][3]. They, however, also belong to a source coding technique for analog speech data. In this paper, we derive a coding method which can be commonly applied to all the above source codes. The derived code is a "code for transmission". If the source codes are efficiently transmitted by means of the transmission code, then it is expected that a faster communication is accomplished.

For simplicity, we deal with a PCM source code hereafter as the basis of the transmission code. In this case, the transmission encoder and transmission decoder are used after the PCM transmitter and before the PCM receiver, respectively, as shown in Figure 1. Our purpose in this paper is to find a code which is attractive at the transmission encoder and decoder.

Let us assume that a speech data sample is expressed by the PCM transmitter as

$$Sample = [a_1\ a_2\ ...a_M] \tag{1}$$

where $a_i, i = 1, 2, ..., M$ consist of "1" or "0" and M corresponds to the number of bits. In such a case, if there are N speech data samples to be transmitted, we can express them in a matrix form as

$$Data = \begin{bmatrix} a_1 & a_2 & \cdots & a_M \\ b_1 & b_2 & \cdots & b_M \\ c_1 & c_2 & \cdots & c_M \\ \vdots & \vdots & \vdots & \vdots \\ *_1 & *_2 & \cdots & *_M \end{bmatrix} \tag{2}$$

where the length of $a_i b_i ... *_i$ is N. In a packet transmission, the above matrix should be sent as efficiently as possible. To achieve this, we would consider an another expression of (2).

2 Code for Sample

Any speech data samples can be expressed by using only an integer. Thus, a fixed point expression is enough for arranging the elements $a_i, i = 1, 2, ..., M$ in (1), where the first bit, a_1, is allocated for the sign of the sample. If the sample is positive, then the sign bit is "0" and the absolute value is expressed by the remaining bits. However, if the sign is negative, then the sign bit is "1" and usually the remaining bits are complementally used to express the sample. This is a so-called complement. Because we can simplify logical circuits for processing the sample by using such a complement, the complemental expression is usually used. However, simplifying circuits is not our purpose. We thus employ the following coding scheme for each sample:

- the first bit is allocated for sign expression.
- regardless of the sign of the sample, the remaining bits are used to express the absolute value of the sample.

This expression of the sample is more effective to compress the total data bits.

3 Code for Data

When we transmit the total data expressed by (2), if we can find common bits, for example, such as

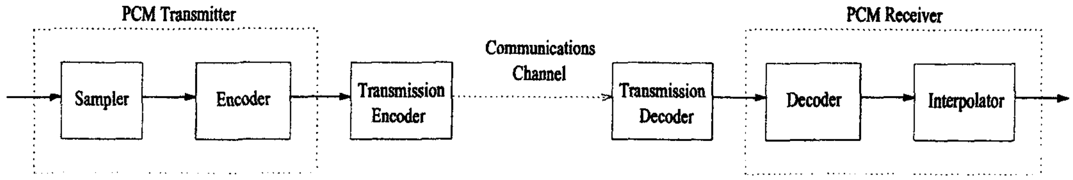

Figure 1: Transmission scheme

$a_1 = b_1 = ... = *_1$ and $a_2 = b_2 = ... = *_2$, then it is straightforward to use the following expression:

$$Data' = a_1 a_2 \begin{bmatrix} a_3 & a_4 & \cdots & a_M \\ b_3 & b_4 & \cdots & b_M \\ c_3 & c_4 & \cdots & c_M \\ \vdots & \vdots & \vdots & \vdots \\ *_3 & *_4 & \cdots & *_M \end{bmatrix}. \quad (3)$$

In this case, we can dramatically reduce the total bits to be transmitted. The compression rate (CR) is

$$CR = \frac{N(M-2)+2}{NM}. \quad (4)$$

However, this may not be realistic, because speech is nonstationary and its waveform is of various shapes.

4 Proposed Code for Data

In deriving the proposed code, we note the fact that for a high probability, common bits of speech samples appear during a short time length. Based on this, we consider L data samples as one data set. Then, we divide the data bits matrix (2) into N/L matrices as

$$Data = \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \\ \mathbf{D}_3 \\ \vdots \\ \mathbf{D}_{N/L} \end{bmatrix} \quad (5)$$

where each element $\mathbf{D}_i, i = 1, 2, ..., N/L$ consists of an L by M matrix (for simplicity, it is assumed that N can be divided by L without remainders).

We apply the following coding rule to each submatrix $\mathbf{D}_i$ and make a new code shown in Figure 2. The new code consists of 4 parts:

- Decision bit for omission (D)
- Omission number bits (O)
- Sign bits (S)
- Remaining data bits (R)

For (D) part, 1 bit is required. For (O) and (S) parts, 3 bits are commonly required. A variable number of bits are allocated to (R) part.

[Coding Rule]

Let us assume that the i-th submatrix $\mathbf{D}_i$ is expressed by

$$\mathbf{D}_i = \begin{bmatrix} s^i_{11} & s^i_{12} & \cdots & s^i_{1M} \\ s^i_{21} & s^i_{22} & \cdots & s^i_{2M} \\ s^i_{31} & s^i_{32} & \cdots & s^i_{3M} \\ \vdots & \vdots & \vdots & \vdots \\ s^i_{L1} & s^i_{L2} & \cdots & s^i_{LM} \end{bmatrix}. \quad (6)$$

For this matrix, we do as follows.

1. find the number of the columns having common "0"s from lower order columns except for the first column $s_{i1}, i = 1, 2, ..., L$.
2. allocate the number found in 1 to (O) part in binary bits.
3. When the number found in 1 is equivalent to the corresponding number in the previous submatrix $\mathbf{D}_{i-1}$, allocate "0" to (D) part and omit (O) part. On the other hand, when the number found in 1 is not equivalent to the corresponding number in the previous submatrix $\mathbf{D}_{i-1}$, allocate "1" to (D) part.
4. allocate each sign bit to (S) part in order, as $s_{11}, s_{12}, ..., s_{1L}$.
5. allocate the remaining bits of each data to (R) part in order.

As an example, consider the case where $M = 8$ and $L = 3$. The data samples coded by PCM are ... 65, 14, 8, -5, ... We recognize {14, 8, -5 } as one data set. In this case,

$$65 = [01000001] \quad (7)$$

$$14 = [00001110] \quad (8)$$

$$8 = [00001000] \quad (9)$$

$$-5 = [10000101] \quad (10)$$

and the submatrix becomes

$$\mathbf{D}_i = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}. \quad (11)$$

In (11), (S) part is 001. Because we can find common three "0"s in the absolute value expression, the number

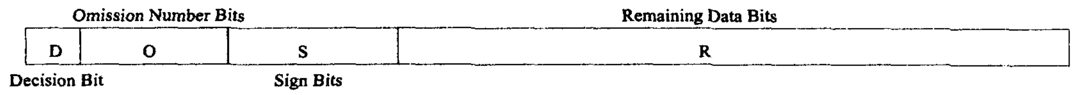

Figure 2: Proposed code

Table 1: Compression Rate

Speech	Total Data Bits	Coded Data Bits	CR
Male 1	880000	387607	44.0
Female 1	880000	402124	45.7
Male 2	880000	385714	43.8
Female 2	880000	400285	45.5
Male 3	880000	433681	49.3
Female 3	880000	456472	51.9
Male 4	880000	453077	47.2
Female 4	880000	417253	47.4

of common "0"s is 3 and the resulting binary expression 001 is allocated to (O) part. Such three "0"s are not included in the expression of 65, that is in (7). Thus, (D) part is 1. The remained bits are 1110 in the expression of 14 in (8), 1000 in the expression of 8 in (9), and 0101 in the expression of -5 in (10). These are allocated to (R) part. As a result, the following code is obtained.

$$1011001111010000101 \qquad (12)$$

The proposed code mentioned above can be uniquely decoded by implementing a reversed procedure of that of coding.

5 Simulations

To verify the performance of the proposed coding scheme, we carried out computer simulations. Speech data in the experiments was taken from "20 Countries Language Database (NTT Advanced Technology)". The speech signals used were uttered by 4 male and 4 female speakers. Each of the speech signals consisted of about 10s Japanese sentences, which were sampled by a rate of 10 kHz and quantized by a PCM scheme with 8 bits. In the proposed code, the setting of $L = 3$ was used and the compression rate (CR) was evaluated in percentage. The results have been summarized in Table 1. From this table, we see that a compression with 50 percent or more is achieved by the proposed code.

6 Concluding Remarks

Packet transmission of speech data has been considered. It has been shown that by using a code which eliminates redundant information for continuous speech data, we can transmit the source-coded speech data more efficiently. Future work will aim at investigating the dependency of the proposed code on the source coding schemes such as log-PCM, DPCM, ADPCM as well as PCM.

References

[1] C.C.Bissell and D.A.Chapman, "Digital Signal Transmission", Cambridge University Press, 1992.

[2] N.S.Jayant and P.Noll, "Digital Coding of Waveforms: Principles and Applications to Speech and Video", Prentice-Hall Inc., 1984.

[3] J.G.Proakis, "Digital Communications", Third Edition, McGraw-Hill, Inc., 1995.