Research of Speech Amplitude Distribution Based on Hadamard Transformation

Jingxue Tu[1], Jingyun Xu[1](✉), and Xiaoqun Zhao[2]

[1] School of Engineering, Huzhou University, Huzhou 313000, China
happy_865709718@qq.com, xujingyunsh@gmail.com
[2] School of Electronic and Information Engineering, Tongji University, Shanghai 201804, China
Zhao_Xiaoqun@tongji.edu.cn

Abstract. In view of PCM (Pulse Code Modulation) of speech signal, this paper puts forward a method of speech processing based on hadamard matrix transformation to change the amplitude distribution of speech signal, which can reduce the standard deviation of speech signal. Experiments show that the hadamard matrix transformation algorithm can obviously reduce the amplitude range of speech signal. Speech signal standard deviation is reduced by 20% after the transformation. At the same time, speech quality after decoding is not decreased according to listening experimenter. The algorithm reduces amplitude range and standard deviation of the speech signal, which can code the speech signal with less bits, and compression efficiency can be further improved.

Keywords: Speech · Amplitude distribution · Hadamard matrix · Standard deviation

1 Introduction

Speech are transmitted and processed in analogy manner in early stage of research. Since the PCM method has been proposed, Speech can be stored and transmitted as a digital data. However, A huge amount of data is a big issue, therefore it is necessary to compress speech signal. From the original 64 KB/s standard PCM waveform encoder to the at or below 4 KB/s parameter coding vocoder now, speech compression coding is gaining steam rapidly for decades. Speech coding and compression are extensively used in many applicants such as GSM system and IP telephone system [1–4]. GSM mobile communication use wireless channel transmission, since the frequency of the wireless channel resources are limited, the utilization rate of channel using speech compression technology is improved. Telephone system usually adopts linear PCM, which is sample rate of 8 kHz and quantization number is 11 bit, Coding rate reaches up to 88 kbit/s if using 8 bit non-uniform quantization, the coding rate reaches up to 64 kbit/s [5].

The most important business is speech business in mobile communication, precious wireless spectrum resources requires each user to take up the narrow spectrum as soon

XU. Jingyun—Project supported by the National Nature Science Foundation of China (No. 61271248), Natural Science Foundation of Huzhou City (No. 2015YZ04).

X.-L. Huang (Ed.): MLICOM 2016, LNICST 183, pp. 267–273, 2017.
DOI: 10.1007/978-3-319-52730-7_27

as possible. The size of the spectrum is directly related in the compression ratio of speech. We need to compress the speech signal in order to save storage media. The purpose of the speech compression coding is to meet the demand of the narrowband channel low bit rate transmission and realize speech storage efficiently on the premise of guarantee the quality of speech [5].

At the same time, due to the special nature and construction method of hadamard matrix, which has a wide range of USES in communication [6, 7]. Thus, the paper puts forward a method of speech processing based on hadamard matrix transformation to change the amplitude distribution of speech signal, which can reduce the standard deviation of speech signal. Experiments show that speech compression efficiency can be further improved by the hadamard matrix transformation algorithm.

2 Speech Signal Amplitude Distribution

Suppose that the sample size of speech signals is $K = M \times N$ point, denoted as matrix $X = (X_1, X_2, \ldots, X_K)$, among K, M, N is a positive integer. The matrix form of the data points can be expressed as

$$X_{N\times \mathrm{M}} = \begin{bmatrix} X_1 & \cdots & X_M \\ \vdots & \ddots & \vdots \\ X_{\mathrm{N}-1\times \mathrm{M}+1} & \cdots & X_{N\times \mathrm{M}} \end{bmatrix} \tag{1}$$

The duration of speech signal is 10 min, the sampling rate of speech signal is 8 kHz, the quantitation precision and range is 8 bit and between 128 ~ 128 respectively in the experiment.

The most part of speech signal sample data points fall within the scope of the (0.2 ~ 0.2) from Fig. 1, so we make the scope of the (0.2 ~ 0.2) as the benchmark for comparisons in this paper. The experimental results show that the probability of data point fall within the scope is 91.78%.

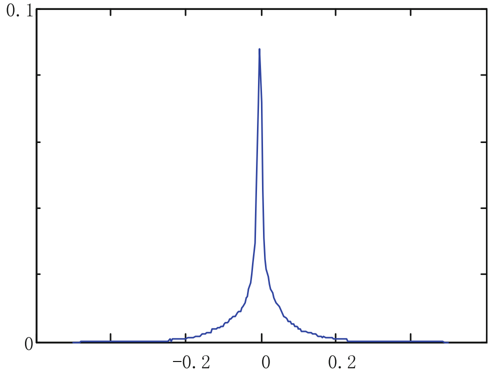

Fig. 1. Amplitude distribution of speech

Speech signal is non-uniform and has significant correlation from the time domain. Speech signal has strong unevenness and certain redundancy from the power spectral density. Due to the special nature and construction of hadamard matrix, We can take advantage of the hadamard matrix of speech signal to eliminate the correlation operation, which reduces the dynamic range of speech signal and improve the coding efficiency.

3 Hadamard Matrix and Its Properties

Hadamard matrix [4] is made up of element +1 and 1 and nonsingular N order phalanx, 2 order hadamard matrix can be defined as

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} N \times N \tag{2}$$

4 order hadamard matrix can be defined as:

$$H_4 = \begin{bmatrix} H_2 & H_2 \\ H_2 & -H_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \tag{3}$$

The general equation for hadamard matrix:

$$H_{2N} = \begin{bmatrix} H_N & H_N \\ H_N & -H_N \end{bmatrix} \tag{4}$$

Where H_N is a hadamard matrix of size $N \times N$, Hadamard matrix (H matrix) and its main properties:

(A) any two rows (columns) of the matrix are orthogonal.
(B) the square sum of all the elements in any row (column) is equal to the squares of order number.
(C) the hadamard matrix order number is 2 or 4 multiples.

4 Speech Signal Distribution Transformed

4.1 Mixed Speech Signal Distribution

We can get the matrix of speech X, N × M order. The matrix X is transformed by hadamard method to the mixed matrix Y:

$$Y = \mathrm{H} \times X' = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & -1 \end{bmatrix} \times \begin{bmatrix} X_1 & \cdots & X_M \\ \vdots & \ddots & \vdots \\ X_{\mathrm{N}-1\times\mathrm{M}+1} & \cdots & X_{N\times\mathrm{M}} \end{bmatrix} \tag{5}$$

The experimental results show that the probability of the mixed matrix Y data point fall within the scope is 94.11% (Fig. 2).

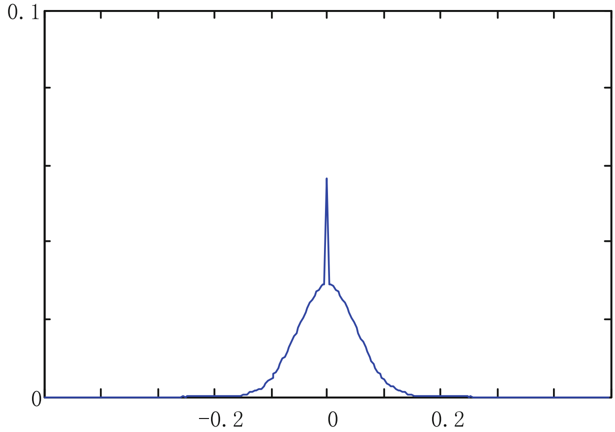

Fig. 2. Mixed speech signal distribution

4.2 Speech Signal Distribution Transposed

In order to further eliminate the correlation of speech signal among the samples, the speech signals are divided into a set of N samples, and carries on the hadamard transformation, the resulting speech signal is equivalent to transpose for mixed speech signals, so called transposed speech signal. The matrix form of the speech signal F is expressed as:

$$F_{\mathrm{M}\times N} = \begin{bmatrix} X_1 & \cdots & X_N \\ \vdots & \ddots & \vdots \\ X_{\mathrm{M}-1\times\mathrm{N}+1} & \cdots & X_{\mathrm{M}\times N} \end{bmatrix} \tag{6}$$

The correlation of speech signal mainly embodied in the (X_1, X_2, … X_N), …, (X_{M-N+1}, X_{M-N+2}, … X_M), in order to take advantage of the nature of the hadamard matrix to remove the correlation of speech signal, to transpose for F, we can get F′ are expressed as

$$F'_{N\times\mathrm{M}} = \begin{bmatrix} X_1 & \cdots & X_{\mathrm{M}-1\times\mathrm{N}+1} \\ \vdots & \ddots & \vdots \\ X_{\mathrm{N}} & \cdots & X_{\mathrm{M}\times N} \end{bmatrix} \tag{7}$$

N order hadamard matrix H multiply matrix X′, We can get the transposed matrix Z:

$$Z = HX' = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & -1 \end{bmatrix} \begin{bmatrix} X_1 & \cdots & X_M \\ \vdots & \ddots & \vdots \\ X_{(N-1)M+1} & \cdots & X_{NM} \end{bmatrix} \tag{8}$$

The dynamic range of speech signal are compressed and the coding bits is reduced by the method, The experimental results show that the probability of the transposed matrix Z data point fall within the scope is 95.78% (Fig. 3).

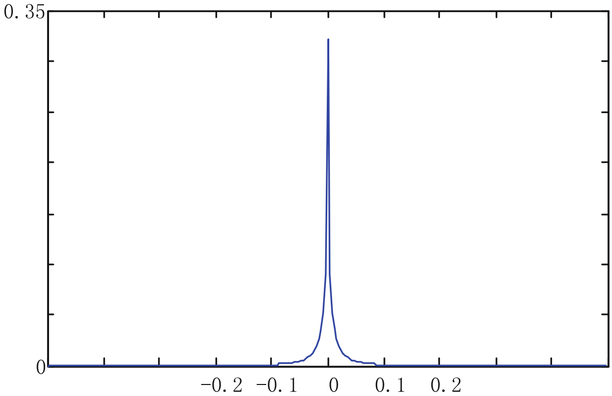

Fig. 3. Speech signal amplitude distribution transposed

5 Order Selection of Hadamard Matrix

The speech signal is transformed by using different order hadamard matrix and is to do its statistics of amplitude distribution with the increase of hadamard matrix of order, the speech signal compression ratio is higher from the simulation results. However, when the hadamard matrix has achieved a certain order, signal compression rate no longer increase, if we further increase the order number of hadamard matrix, compression effect decreases instead.

The experiments shows that N = 256 order of hadamard matrix has the optimal compression effect. So, this paper uses the hadamard matrix of 256 order.

6 Result and Analysis

The matrix of original speech signal X, the hadamard matrix H, the matrix Y is the results of transformation, Y is expressed as

$$Y = H \times X \tag{9}$$

In order to restore the original speech signal X, do the following operation for Y:

$$\frac{1}{N} \times H^{'} \times \mathrm{Y} = \frac{1}{N} \times H^{'} \times \mathrm{H} \times \mathrm{X} = \frac{1}{N} \times \mathrm{N} \times \mathrm{X} = \mathrm{X} \tag{10}$$

The original signal can be restored after the above operation, Fig. 4 shows the original speech and speech after two kinds of inverse transformation of time domain waveform, it can be seen that speech waveform recovered is no distortion.

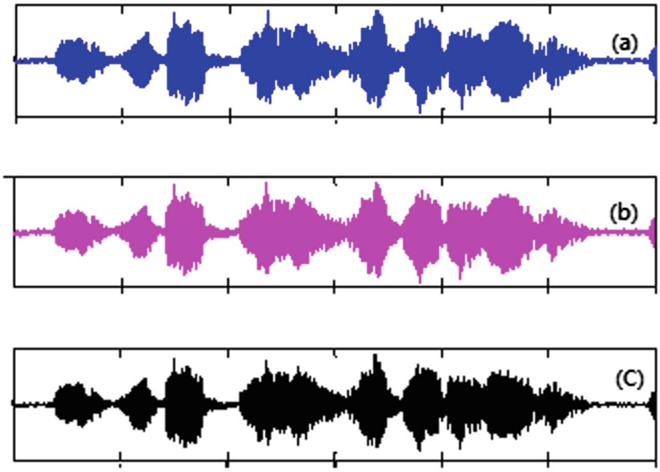

Fig. 4. Time domain waveform of speech signal: (a) original speech; (b) mixed speech; (c) transposed speech

The simulation data as shown in Table 1. The standard deviation is defined as the overall standard units with the square root of the arithmetic average of the mean square deviation. It reflects the degree of discrete between individuals in the group. In the experiment, the standard deviation is normalized, the results are shown in Table 2.

Table 1. Original signal standard deviation

	Original	Mixed	Transposed
Standard deviation	0.9178	0.9411	0.9578

Table 2. Signal standard deviation before DC eliminated

	Original	Mixed	Transposed
Standard deviation	1.00	0.974	0.952

Table 3. Signal standard deviation after DC eliminated

	Original	Mixed	Transposed
Standard deviation	1.00	0.862	0.805

After the original speech is eliminated the dc to, the standard deviation of the original signal and signal transformed are calculated, the result as shown in Table 3. From Table 3, the standard deviation of mixed signal is reduced by 14%, and standard deviation of the transposed signal is reduced by 20% relative to the original signal.

7 Conclusion

This paper first analyzes the correlation of speech signals between adjacent samples using hadamard transformation, the experimental results show that this method can significantly reduce the speech signal dynamic range, which can improve the compression ratio, at the same time greatly reduce the standard deviation of speech signals. The method in speech coding and wireless communication has certain reference value.

References

1. Ahmed, N., Rao, K.R.: Orthogonal Transforms for Digital Signal Processing. Springer Science & Business Media, Heidelberg (2012)
2. Kaur, H., Kaur, R.: Speech compression and decompression using DWT and DCT. Int. J. Comput. Technol. Appl. **3**(4), 1501–1503 (2012)
3. Hillebrand, F.: The creation of standards for global mobile communication: GSM and UMTS standardization from 1982 to 2000. IEEE Wirel. Commun. **20**(5), 24–33 (2013)
4. Amira, A., Bouridane, A., Milligan, P., et al.: Novel FPGA implementations of Walsh-Hadamard transforms for signal processing. IEE Proc. Vis. Image Sign. Process. **148**(6), 377–383 (2001). IET
5. Wang, G., Huang, H., Liu, Y., Zhang, X., Wang, Z.: Uncertainty estimation of reliability redundancy in complex systems based on the Cross-Entropy method. J. Mech. Sci. Technol. **23**(10), 2612–2623 (2009)
6. Fu, Z., Zhao, J.: Information Theory and Coding. Electronic Industry Press, Beijing (2008)
7. Kumar, A., Singh, G.K., Rajesh, G., et al.: The optimized wavelet filters for speech compression. Int. J. Speech Technol. **16**(2), 171–179 (2013)