Dimension Reduction Analysis of Signal Manifold for Vowels in Time and Frequency Domain

JINQING SHEN[1], ZHONGXIAO LI[2], QIANQIAN CHEN[2], XIAODONG ZHUANG[2]
[1]Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, Shenzhen, CHINA
[2]Electronics & Information College, Qingdao University, Qingdao, CHINA

Abstract: In this paper, the LLE and ISOMAP algorithms in manifold learning are applied them to the analysis of vowel signals in time and frequency domain. Time domain simulation results show that the two dimensionality reduction methods can implement two-dimensional visualization of signals while preserving the high-dimensional manifold structure of original signals as much as possible. The time-frequency domain dimension reduction analysis of vowel signal manifold effectively solves the problem that high-dimensional speech signals can't be intuitively felt, and provides a new potential way for signal classification. The frequency domain analysis is further optimized on the basis of time domain simulation. Because half of the amplitude values in DFT is used in the simulation, the two-dimensional manifold of the signal is roughly linearly distributed, which can effectively reduce redundancy and make the signal more compactly expressed in the frequency domain.

Key-Words: dimension reduction, manifold learning, LLE, ISOMAP, vowel signal

Received: March 20, 2021. Revised: July 11, 2021. Accepted: July 29, 2021. Published: August 14, 2021.

1 Introduction

Currently, the amount of available information increases exponentially. It is necessary to effectively reduce the dimension of high-dimensional data in order to solve the "dimension disaster" [1,2], and obtain as much valuable information as possible. Data dimension reduction studies low-dimensional manifolds embedded in high-dimensional space and their different types, so as to find out the essential properties of data set. The purpose of dimension reduction technology based on manifold learning is to find out the inherent geometric structure of high-dimensional data in order to solve practical problems. Tenenbaum, etc. replaced Euclidean distance in MDS algorithm with geodesic distance, and proposed an ISOMAP algorithm based on global attributes [3]. Roweis, etc. considered the local information of samples in the process of data dimension reduction, and proposed a nonlinear LLE algorithm [4]. ISOMAP and LLE have greatly promoted the development of nonlinear dimension reduction methods based on manifold learning.

At present, manifold learning method has been applied to feature extraction of sound signals, and the results show that the essential features of stationary sound signals can be discovered by manifold learning. Speech signal is a non-stationary random signal with high redundancy[5]. In speech signal processing, appropriate dimension reduction methods are also needed for efficient feature extraction and analysis [6]. In order to analyze vowel signals in more details, it is necessary to analyze the time and frequency domain features of speech. Therefore, this paper mainly studies the data dimension reduction method based on manifold learning and applies it to the time and frequency domain analysis of vowel signals.

2 Introduction to Manifold Learning Methods

2.1 Locally linear embedding algorithm (LLE)

The core of LLE algorithm is to assume that the points adjacent to each other in the high-dimensional sample space are usually located on the same manifold, and LLE algorithm can make the nonlinear high-dimensional data keep the original local linear property as much as possible after dimension reduction. The LLE algorithm can be summarized in three steps [8,9]:

Step1. For each data point X_i, compute its k-nearest neighbors(based on Euclidean distance or some other appropriate definition of 'nearness').

Step2. Compute weights W_{ij} that best reconstruct each data point X_i, from its neighbors, minimizing the reconstruction error J:

$$J(W) = \sum_{i=1}^{m} ||X_i - \sum_{j=1}^{k} W_{ij} X_j||^2 \qquad (1)$$

Step3. Compute the low dimensional embeddings Y_i , best reconstructed by the weights W_{ij} , minimizing the cost function φ:

$$\varphi(Y) = \sum_{i=1}^{m} ||Y_i - \sum_{j=1}^{k} W_{ij} Y_j||^2 \qquad (2)$$

In **Step 2**, the reconstruction error is minimized subject to two constraints: first, that each input is reconstructed only from its nearest neighbors, or W_{ij} = 0 if X_i is not a neighbor of X_j; second, that the reconstruction weights for each data point sum to one, or $\sum_j W_{ij}$= 1 $\forall i$. The optimum weights for each input can be computed efficiently by solving a constrained least squares problem.

2.2 Isometric mapping algorithm (ISOMAP)

The main idea of ISOMAP is to replace Euclidean distance in MDS algorithm with geodesic distance between all data points, and try to ensure the global relationship between data and explore the internal geometric structure on the low-dimensional manifold of samples in the process of dimension reduction, so as to better mine the essential characteristics of samples and reveal their internal manifold information more reasonably [9].

While ISOMAP and LLE have similar aims, ISOMAP is based on a different principle than LLE. In particular, ISOMAP attempts to preserve the global geometric properties of the manifold while LLE attempts to preserve the local geometric properties of the manifold.

As with LLE, the ISOMAP algorithm consist of three steps[10]:

Step1. Construct a neighborhood graph - Determine which points are neighbors on the manifold based on distances d(i,j) between pairs of points i, j in the input space (as in step1 of LLE).These neighborhood relations are then represented as a weighted graph over the data points with edges of weight d(i, j) between neighboring points.

Step2. Compute the shortest path between all pairs of points among only those paths that connect nearest neighbors using a technique such as Floyd's algorithm.

Step3. Apply classical MDS to embed the data in a d-dimensional Euclidean space so as to preserve these geodesic distances.

3 Nonlinear Dimension Reduction of Vowel Signal Data

The dimension reduction method based on manifold learning focuses on preserving some geometric attributes in manifold after dimension reduction. Some scholars have studied the geometric structure of natural speech data by using the acoustic pipe model generated by speech, and verified the correctness of manifold hypothesis in speech data[9].

Analysis of speech signal features is the premise and foundation of speech signal processing. According to the different features of speech signal analysis, speech signal feature analysis can be divided into time domain, frequency domain and cepstrum domain analysis. In this study, the vowel signal is analyzed in time and frequency domain based on data dimension reduction, and the signal processing tools in MATLAB software and Voice box toolbox are used [11]. The experimental data are monophones [e] and [u] with a sampling frequency of 16 kHz.

Firstly, the signal is preprocessed by framing and windowing. Several speech sampling points are divided into a frame (denoted by N), in which the signal characteristics are considered to remain unchanged [14]. In order to make the transition between frames smooth, there will be partial overlap between frames (as shown in Fig. 1), and the overlapping area is the frame shift (denoted by M), and the ratio of frame shift to frame length is generally 0 ~ 1/2 [7]. In order to make both ends of the signal edge continuous after framing, in practical problems, each frame signal should be multiplied by a window function to prevent truncation effect and spectrum leakage. Therefore, in this paper, firstly, the Hamming window with frame length of 25 *ms* and frame shift of 16ms is used to preprocess the vowel signal, and the simulated non-stationary speech signal is changed into a short-time stationary signal [1].

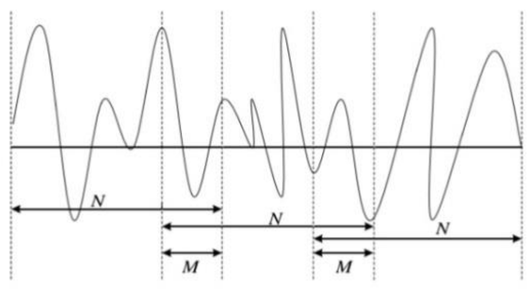

Fig. 1. Schematic diagram of frame length and frame movement

3.1 Analysis of Vowel Signal in Time Domain

Time domain analysis is the most intuitive method for signal analysis. On the basis of the above pre-processing operations, two nonlinear dimension reduction algorithms, ISOMAP and LLE, are directly used to reduce the dimension of the high-dimensional vowel signal, and the two-dimensional geometric structure of the vowel signal is obtained for subsequent analysis while keeping the manifold structure of the original high-dimensional signal unchanged as much as possible[13].

Both LLE and ISOMAP algorithms need to manually adjust the nearest neighbour parameter *K* in order to obtain the best experimental results. When *K* is 9, 12, 15, 18, 21, 23, 25, 28, 35 and 48, the pre-processed experimental data are subjected to ISOMAP and LLE dimension reduction processing, and the LLE and ISOMAP two-dimensional geometric structures of vowel signals [e] and [u] are obtained respectively, as shown in the following Fig.2, Fig.3 and Fig.4. Experiments on vowel signals show that there is a low-dimensional manifold structure in vowel signal data, that is, the high-dimensional vowel signals are nonlinear manifolds. Comparing the LLE and ISOMAP two-dimensional embedding results of vowel signals given below, it can be seen that the low-dimensional embedding results of ISOMAP change little and have certain regularity when the *K* value changes little; On the contrary, the low-dimensional geometric structure obtained by LLE is distorted seriously when the value of k changes little, so it can be considered that the two-dimensional embedded graph obtained by ISOMAP dimension reduction can express the essential characteristics of vowel signals more stably. Experiments on more vowel signals show that the two-dimensional ISOMAP embedding results obtained from different vowel signals are different and have their own characteristics. Compared with LLE low-dimensional embedding results, the low-dimensional ISOMAP embedding results are more stable, so it can be considered to be applied to subsequent classification analysis.

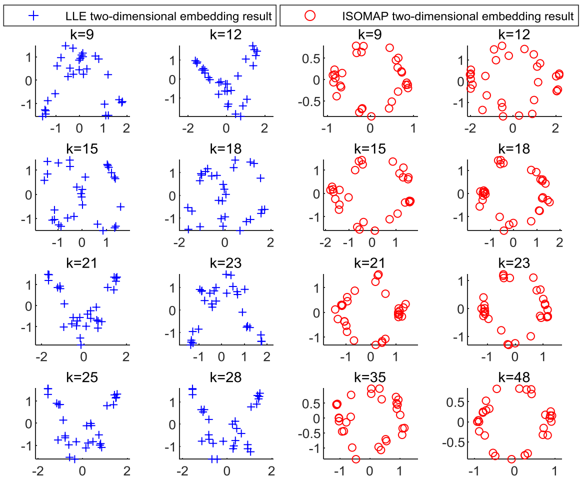

Fig. 2. Two-dimensional embedding results of LLE and ISOMAP under different *K* values of vowel [e]

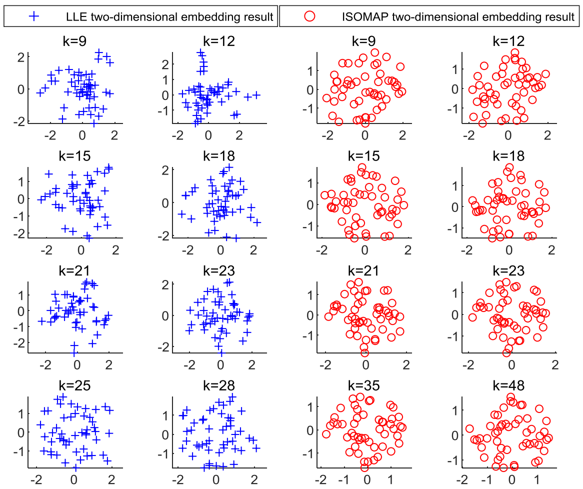

Fig. 3. Two-dimensional embedding results of LLE and ISOMAP under different *K* values of vowel [o]

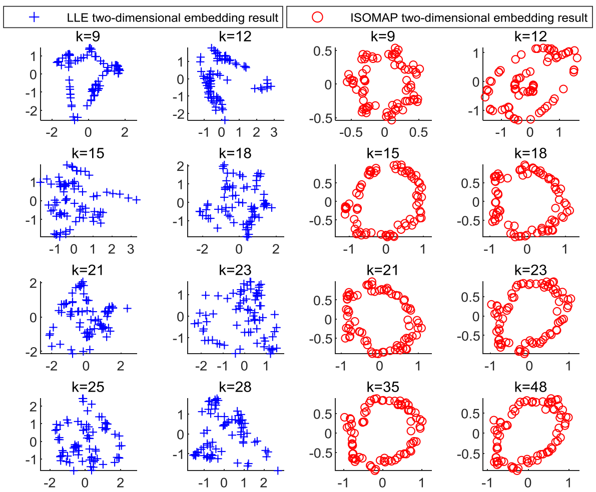

Fig. 4. Two-dimensional embedding results of LLE and ISOMAP under different *K* values of vowel [u]

3.2 Analysis of Vowel Signal in Frequency Domain

Due to the redundancy of time domain features such as short-term energy, it is vulnerable to noise interference, the features are unstable and the signal energy that can be directly obtained is less. However, the speech features are stable in frequency domain and can be used as the coefficients of back-end recognition[14]. On the basis of LLE and ISOMAP dimension reduction analysis in time domain, this paper further analyzes the frequency domain characteristics of vowel signals. The discrete Fourier transform (DFT) is applied to each frame of time domain signal to obtain its frequency domain signal. Because of the high computational complexity of DFT, this paper uses Fast Fourier Transform (FFT) to realize it [15]. The energy spectrum of speech features is obtained after FFT, and the phase spectrum in the energy spectrum actually contains very little information, so it is generally chosen to keep the amplitude spectrum and discard the phase spectrum. Due to the symmetry of the amplitude-frequency curve, the vowel signals [e], [o] and [u] are subjected to 512-point FFT and modulo, and then half of the amplitude spectrum is taken for subsequent LLE and ISOMAP dimension reduction analysis. When k is 15, 17, 19, 21, 23, 25, 27 and 29, the preprocessed experimental data are subjected to ISOMAP and LLE dimension reduction processing in frequency domain, and the LLE and ISOMAP two-dimensional geometric structures of vowel signals [e] and [u] are obtained respectively, as shown in the following figures 5, 6 and 7.

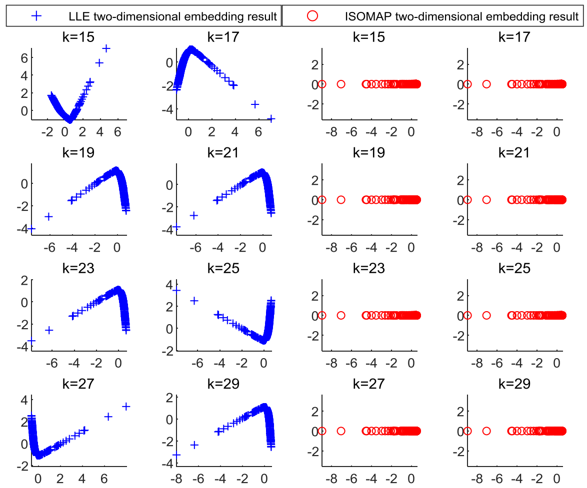

Fig. 5. Results of two-dimensional embedding of vowel signal [e] with different *K* (values are obtained by taking FFT modulus first and then taking half of amplitude spectrum for LLE and ISOMAP dimension reduction)

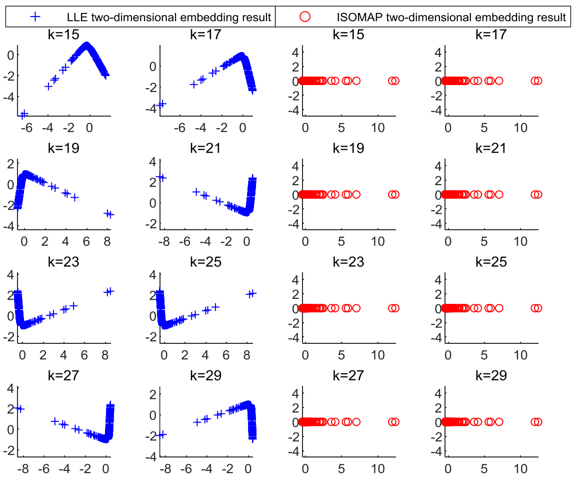

Fig. 6. Results of two-dimensional embedding of vowel signal [o] with different *K* (values are obtained by taking FFT modulus first and then taking half of amplitude spectrum for LLE and ISOMAP dimension reduction)

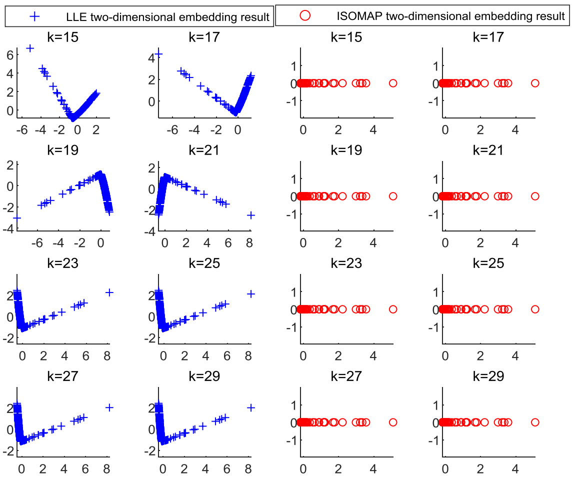

Fig. 7. Results of two-dimensional embedding of vowel signal [u] with different *K* (values are obtained by taking FFT modulus first and then taking half of amplitude spectrum for LLE and ISOMAP dimension reduction)

Because of the symmetry of amplitude-frequency curve, taking half of the amplitude value for LLE and ISOMAP dimension reduction analysis, while realizing visualization, its greatest advantage is to prevent data duplication and reduce redundancy. From the experimental results, it can be seen that the distribution of data points obtained by dimension reduction processing of a given signal in frequency domain is more concentrated than that obtained by time domain processing, and the vowel signal is expressed more effectively and compactly in frequency domain, and the two-dimensional embedding result graph of ISOMAP is roughly linear.

The generation of speech is a very complicated process. How to extract and select the characteristic parameters that can represent the essential characteristics of speech signals is an important link in speech recognition and other technologies. The data dimension reduction technology based on manifold learning can not only get the low-dimensional embedding results of signals [9,17], but also get LLE and ISOMAP feature parameters representing vowel signal features, which is expected to be applied to speech recognition and other technologies in order to discover more signal features.

4 Conclusion

In order to solve the problem that the traditional feature analysis method based on linear system theory can't find the essential features of speech signals well, we studied data dimension reduction method based on manifold learning and applies it to the analysis of vowel signals in both time and frequency domain. The speech signal is preprocessed, then the preprocessed signal is reduced in dimension by using ISOMAP and LLE directly. On the other hand, the preprocessed signal is transformed into the frequency domain by FFT, and then the signal is reduced in dimension by ISOMAP and LLE. By manually adjusting the neighbor parameter *K* in both time and frequency domain, the 2D embedding result of high-dimensional speech signal is obtained. Time domain analysis experiments show that the dimension reduction method based on manifold learning can eliminate redundant information and realize two-dimensional visualization of signals while preserving the high-dimensional manifold structure of original signals as much as possible. Moreover, the LLE and ISOMAP dimension reduction analysis are carried out by taking half of the amplitude values in frequency domain, which can reduce redundancy and realize visualization, and at the same time, the vowel signal can be expressed more effectively and compactly in frequency domain than in time domain.

However, there are some subjective factors in choosing the nearest neighbor parameter K, which has a great influence on the analysis results. The way to choose the appropriate parameter *K* is the key content future research. In addition, in order to make the frequency domain analysis more comprehensive, the combination of the signal phase spectrum analysis is also worth considering.

References:

[1] Claudio Turchetti, Laura Falaschetti, A Manifold Learning Approach to Dimensionality Reduction for Modeling Data, *Information Sciences*, 2019, 491, 16-29.

[2] Panagis Magdalinos,Christos Doulkeridis, Michalis Vazirgiannis, Enhancing Clustering Quality Through Landmark Based Dimensionality Reduction[J], *ACM Transactions on Knowledge Discovery fr om Data (TKDD)*,2011,5(2),1-44.

[3] Tenenbaum J B, Silva V D, Langford J C, A Global Geome-tricframework for Nonlinear Dimensionality Reduction, *Science,* 2000, 290(5500): 2319-2323. a

[4] Roweis S T, Saul L K. Nonlinear Dimensionality Reduction by Locally Linear Embedding, *Science*, 2000, 290 (5500): 2323-2326.

[5] Zhong Zhao, Guocan Feng, Jiehua Zhu, Qi Shen, Manifold Learning: Dimensionality

Reduction and High Dimensional Data Reconstruction via Dictionary Learning, *Neurocomputing*, 2016, 216, 268-285.

[6] M. M. Hasan, H. Ali, M. F. Hossain and S. Abujar, Preprocessing of Continuous Bengali Speech for Feature Extraction, *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2020, pp. 1-4.

[7] Hui Liu, Jun 'an Yang, Yi Wang, Research on Feature Extraction Method of Acoustic Target Based on Manifold Learning, *Journal of Information Physics*, 2011,60 (7): 1729-1736.

[8] Liu Fangyuan; Xia Kewen; Niu Wenjia, An Improved Locally Linear Embedding Algorithm for Weight Reconstruction, *Chinese Journal of Image and graphics*, 2018,23(01) : 52-60.

[9] Andrew Errity, John McKenna, An Investigation of Manifold Learning for Chinese Handwriting Analysis, *International Journal of Systems and Control*, 2007,(03):213-220.

[10] Zhao Yuanyuan; Wang Li, Research on Speech Feature Extraction and Application Based on Manifold Learning ISOP Algorithm [J], *Microcomputer and application,* 2014,33 (01): 43-45 + 49.

[11] Xi Hua, Speech Signal Processing Based on MATLAB [J]. *Modern information technology*, 2020, (17): 76-78.

[12] Paul S Bibin Sam; Glittas Antony Xavier; Gopalakrishnan Lakshminarayanan, A Low latency Modular Level Deeply Integrated MFCC Feature Extraction Architecture for Speech Recognition, *Integration*, 2021,76,69-75.

[13] Shaeela Ayesha, Muhammad Kashif Hanif,Ramzan Talib, Overview and Comparative Study of Dimensionality Reduction Techniques for High Dimensional Data, *Information Fusion*,2020,59,44-58.

[14] Garima Sharma, Kartikeyan Umapathy, Sridhar Krishnan. Trends in Audio Signal Feature Extraction Methods, *Applied Acoustic*s, 2020, 158.

[15] Zhao Xiaolei, Research on Speech Signal Processing Based on Short-time Fourier Transform [J], *Ship Electronic Engineering*, 2018,38 (04): 19-22

[16] YuJie, Analysis of Speech Signals Based on Short time Fourier Method, *Proceedings of 7th International Conference on Advanced Materials and Computer Science* (ICAMCS 2018),2018.

[17] Bi Datian, Qiu Changbo, Zhang Han, Research Status and Development of Data Dimension Reduction Technology, *Intelligence theory and practice*, 2013,36(02) : 125-128.

Contribution of individual authors to the creation of a scientific article (ghostwriting policy)

Jinqing Shen carried out the simulation, and prepared the draft of the manuscript.
Xiaodong Zhuang directed the algorithm design and simulation experiments.
Zhongxiao Li and Qianqian Chen gave suggestions to improve the manuscript.

Sources of funding for research presented in a scientific article or scientific article itself

This work is supported by the National Science Foundation of China (41804110).

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)