

Wavelet Based Multi-Scale Principal Component Analysis for Speech Enhancement

Mr. Vijaykumar D. Shinde^{*1}, Mr. C. G. Patil^{#2}, Mr. Sachin D. Ruikar^{#3}

^{*}ME (VLSI & Embedded Systems)Department of E & TC
Sinhgad Academy of Engineering Kondhwa(Bk) pune-48(India)

[#]Assistant Professor, Department of E & TC
Sinhgad Academy of Engineering Kondhwa(Bk) pune-48(India)

Abstract- The goal of speech enhancement varies according to specific applications, such as to reduce listener fatigue, to boost the overall speech quality, to increase intelligibility, and to improve the performance of the voice communication device. This paper presents Multiscale principal component analysis (MSPCA) for denoising of single channel speech signal. Principle Component Analysis (PCA) is a standard tool in modern data analysis because it is simple method for extracting relevant information from complex data matrix using eigenvalues and eigenvectors. The multiscale principal component generalizes the usual PCA of a multivariate signal seen as a matrix by performing simultaneously a PCA on the matrices of details of different levels. In multi scale Principal Component Analysis (MSPCA) decorrelate the variables by extracting a linear relationship and wavelet analysis. In addition, a PCA is performed also on the coarser approximation coefficients matrix in the wavelet domain as well as on the final reconstructed matrix. By selecting conveniently the numbers of retained principal components, interesting simplified signals can be reconstructed. Wavelet analysis of speech signal segments the voice information content at different Wavelet scales. At subband levels or scales multivariate data matrix are formed using Wavelet coefficients extracted from the same scales of voice signals. At each subband matrix or scales, PCA is used for noise reduction. Qualitative performance is evaluated and quantitative performance of denoising effect is measured by input/output signal-to-noise ratio (SNR), segmental SNR and IS measure.

Keywords- Multiscale Principal Component Analysis (MSPCA), Principal component Analysis (PCA), Speech Enhancement, Signal Denoising, Wavelet Analysis.

I. INTRODUCTION

Communication via speech is one of the essential functions of human beings. Humans possess varied ways to retrieve information from the outside world or to communicate with each other, and the three most important sources of information are speech, images and written text. For many purposes, speech stands out as the most efficient and convenient one. Speech not only conveys linguistic contents, but also communicates other useful information like the mood of the speaker. Language communication through speech is closely intertwined with the evolution of human civilization.

Speech processing is an interdisciplinary field that studies acoustic signals using both signal processing techniques and knowledge from hearing science, phonetics, linguistics, and psychology. Thanks to the explosive advances in digital signal processing, the ease and speed of representing, storing, retrieving and processing speech data has boosted the development of speech processing techniques to address different application areas: speech enhancement, speech synthesis, speech coding and speech recognition.

One of the most important branches of speech processing, speech enhancement focuses on finding an optimal (i.e., preferred by a human listener) estimate $\hat{s}(t)$ given a noisy measurement. Application areas include the reduction of noise for hearing purposes, the preprocessing of speech coding or recognition systems and hearing aid research. Noise reduction or speech enhancement has always been a non-trivial problem for engineers. The total removal of background noise is practically impossible and the distortion of the speech content is inevitable. The ease of implementation is $y(t)=s(t) + n(t)$ another consideration when real-time on-site performance is expected, for example in hearing aid devices.

Traditional speech enhancement algorithms include spectral subtraction, Wiener filtering, time varying speech model-based or state-based methods, and microphone array based techniques. The wavelet transform distinguishes itself in the analysis of non-stationary signals such as speech. So far all the speech enhancement techniques discussed are based on the spectral information obtained through the short time Fourier transform analysis of the target signal. These are all frequency-based methods intending to preserve the slow-varying short time spectral characteristics of the speech such as the low-frequency harmonics of vowels, which is still not enough to maintain speech quality after the processing. In 1995, Donoho introduced wavelet shrinkage as a powerful tool in denoising signals corrupted by additive white noise. Wavelet shrinkage employs non-linear thresholding in the wavelet domain and has proved broad asymptotically near-optimal properties for a wide class of signals corrupted by additive noise [13]. Recent efforts made for speech enhancement using wavelet shrinkage include those of [11],

[5], [6], [12].

In this paper, our investigation is focused on the integration of the Multiscale principal component analysis into wavelet based denoising for the single-channel noise reduction of speech corrupted by broadband white noise.

II. PRINCIPAL COMPONENT ANALYSIS

Principal components analysis (PCA) is a technique used to reduce a multidimensional data to lower dimensions for analysis. PCA consists of computation of the eigenvalue decomposition or singular value decomposition of a data set, usually after mean centering the data for each attribute. Usually, the results of a PCA are discussed in terms of scores and loadings [10],[4].

PCA is quite much the same as SVD. The only difference is that PCA is calculated based on the covariance (or correlation) matrixes but SVD could be applied on any type of matrices. As a result PCA can be applied on square matrices only, where this is not the case for SVD. In general both methods are based on the Karhunen-Loève Transform (KLT) and the difference is trivial.

"Like SVD, PCA involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called *principal components*. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. In other words, suppose you have samples located in environmental space or in species space. If you could simultaneously envision all environmental variables or all species, then there would be little need for ordination. However, with more than three dimensions, we usually need a little help. What PCA does is that it takes your cloud of data points, and rotates it such that the maximum variability is visible. Another way of saying this is that it identifies your most important gradients [1].

Conventionally, we perform PCA on the symmetric Covariance matrix or on the symmetric Correlation matrix. These matrices can be estimated from the data matrix. In the covariance matrix we have scaled sums of squares and cross products. We obtain the correlation matrix in the same way as we estimate the covariance matrix, but first we should standardize the variables, i.e., the columns [8]. This standardization is necessary if the variances of variables differ much, or if the units of measurement of the variables differ. PCA is based on eigen analysis; we solve for the eigenvalues and eigenvectors of a square symmetric matrix with sums of squares and cross products. The eigenvector associated with the largest eigenvalue has the same orientation as the first principal component. The eigenvector associated with the second largest eigenvalue determines the orientation of the second principal component. The sum of the eigenvalues

equals the trace of the square matrix and the number of eigenvectors equals the number of rows (or columns) of this matrix [4], [2].

III. MULTISCALE PCA

The aim of multiscale PCA is to reconstruct, starting from a multivariate signal and using a simple representation at each resolution level, a simplified multivariate signal. The multiscale principal component generalizes the normal PCA of a multivariate signal represented as a matrix by performing a PCA on the matrices of details of different levels simultaneously. A PCA is also performed on the coarser approximation coefficients matrix in the wavelet domain as well as on the final reconstructed matrix [9], [3]. By selecting the numbers of retained principal components, interesting simplified signals can be reconstructed [10].

IV. ALGORITHM

The multiscale principal component generalizes the usual PCA of a multivariate signal seen as a matrix by performing simultaneously a PCA on the matrices of details of different levels. In addition, a PCA is performed also on the coarser approximation coefficients matrix in the wavelet domain as well as on the final reconstructed matrix. By selecting conveniently the numbers of retained principal components, interesting simplified signals can be reconstructed [10].

- 1) Read a multivariate signal.
- 2) Set the wavelet parameters such as level of decomposition and mother wavelet.
- 3) Select the number of retained principal components using Kaiser's rule.
- 4) Perform a simple multiscale PCA.
- 5) Improve the obtained result by retaining less principal components.
- 6) Suppress the details at levels 1 to n, update the numbers of retained principal components selected by Kaiser's rule.
- 7) Perform multiscale PCA again.
- 8) Estimate Performance parameters
 - a) Calculate snr
 - b) Calculate ssnr
 - c) Calculate is (Itakura-Saito) measure.
- 9) Display the original Signal, Noisy Signal and Denoised signals.

V. PERFORMANCE MEASURES

In this paper, we will develop new algorithms to enhance speech signals in noise. To evaluate their performance, we need performance measures which are consistent with human perception. Many different objective measures have been developed, especially in the context of speech coding. Herein we use the following objective performance measures:

A. Global signal to noise ratio

$$SNR = 10 \cdot \log \frac{\sum_n s^2(n)}{\sum_n [s(n) - \hat{s}(n)]^2} \quad [\text{dB}]$$

where s is the clean and \hat{s} the enhanced signal. The summation is performed over the signal length and \log is the logarithm base 10. This is the most frequently used measure but it has the inconvenience of not being well correlated to subjective results.

B. Segmental signal to noise ratio

This measure makes a conversion of the SNR into dB prior to averaging it, in order to give equal weights to loud and soft parts of speech.

$$SNR_{seg} = \frac{1}{M} \sum_{m=0}^{M-1} 10 \cdot \log \frac{\frac{1}{N} \sum_{n=0}^{N-1} s^2(n + Nm)}{\frac{1}{N} \sum_{n=0}^{N-1} [s(n + Nm) - \hat{s}(n + Nm)]^2} \quad [\text{dB}]$$

where M represents the number of frames in the signal and N the number of samples per frame. In order to improve the measure, it is possible to suppress the periods of silence before calculating the SNR. Furthermore, there exists a frequency weighted segmental SNR.

C. Log Likelihood Ratio (Itakura-Saito distortion)

Distortion between all-pole parameters. This measure performs a comparison between the spectral envelopes and is more influenced by a mismatch in formant location than in spectral valleys. It is computed with the following expression:

$$IS(m)(\vec{a}_s, \vec{a}_{\hat{s}}) = 10 \log \left[\frac{\vec{a}_s \mathbf{R}_{\hat{s}} \vec{a}_s^T}{\vec{a}_{\hat{s}} \mathbf{R}_{\hat{s}} \vec{a}_{\hat{s}}^T} \right]_m$$

where \vec{a}_s and $\vec{a}_{\hat{s}}$ are vectors containing the coefficients of the linear prediction from the m^{th} frame of the clean and enhanced speech. $\mathbf{R}_{\hat{s}}$ is the autocorrelation matrix of the enhanced speech.

VI. RESULTS AND DISCUSSION

The proposed method is evaluated using speech file with different SNR (SNR= -10,-5, 0, 5, 10). All the speech files are sampled at default sampling rate chosen by Matlab and are corrupted by white noise and other noises. To de-noise noisy signal using soft heuristic SURE thresholding and scaled noise option, on detail coefficients obtained from the decomposition of signal, at level 5 by sym8 wavelet. By selecting the number of retained principal components using Kaiser's rule the speech signal is enhanced. Figure 1 shows original and enhanced signal. The following bar graph (figure 2) shows the magnitude plot of SNR, segmental SNR and IS measure for SNR= -10 dB to 10 dB input additive white gaussian noise. Signal plot (figure 1) shows it also helps to remove the initial hissing sound and provides extra clarity to the voice signal.

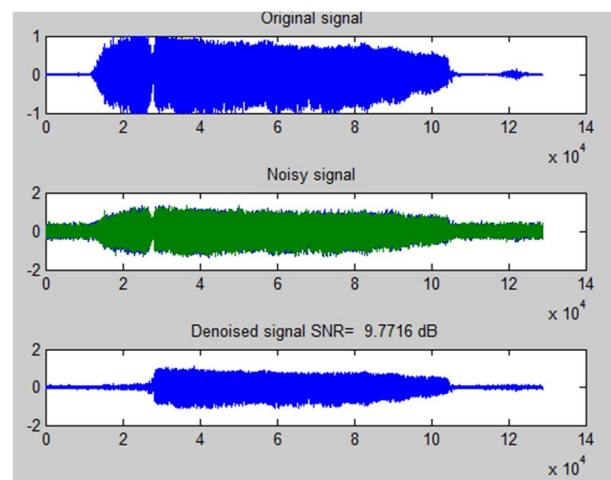


Figure 1. Time magnitude plot of speech signal

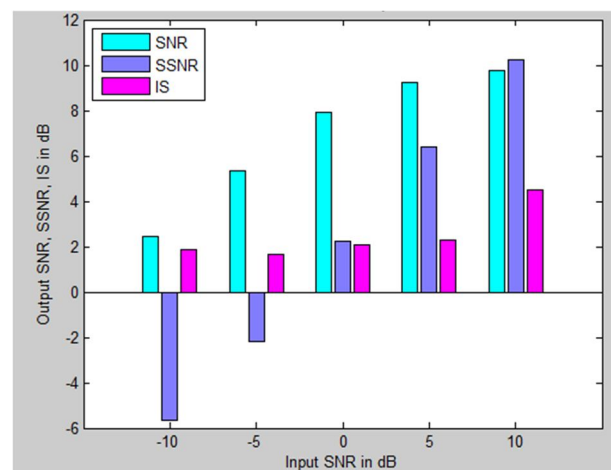


Figure 2. Enhancement result of speech corrupted by white noise.

VII. CONCLUSION

The proposed speech enhancement system has efficiently eliminated the background noise, and it can be almost free

from the musical residual noise, results show clearly that our algorithm provides a lower signal distortion and musical noise and a higher noise reduction than existing enhancement methods. This algorithm suppresses initial hissing noise from the speech as shown in the fig 1. A prominent point in our method is that it does not require any empirical parameter.

References

- [1] Abolhassani, Amin Haji / Selouani, Sid-Ahmed / O'Shaughnessy, Douglas / Harkat, Mohamed-Faouzi (2007), "Speech enhancement using PCA and variance of the reconstruction error model identification", In *INTERSPEECH-2007*, 974-977.
- [2] Tetsuya Takiguchi, Yasuo Arika. (2007), "PCA-Based Speech Enhancement for Distorted Speech Recognition" *Journal of Multimedia*, Vol 2, No 5, 13-18, Sep 2007
- [3] Aminghafari, M.; Cheze, N.; Poggi, J-M. (2006), "Multivariate de-noising using wavelets and principal component analysis," *Computational Statistics & Data Analysis*, 50, pp. 2381-2398.
- [4] Jonathon Shlens (2005), "A Tutorial on Principal Component Analysis", Systems Neurobiology Laboratory, Salk Institute for Biological Studies La Jolla, CA92037 and Institute for Nonlinear Science, University of California, San Diego La Jolla, CA 92093-0402.
- [5] Chang, S., Y.Kwon (2002). Speech enhancement for non-stationary noise environment by adaptive wavelet packet. International conference on audio,speech and signal processing of IEEE.
- [6] Bahoura, M. a. J. R. (2001). "Wavelet speech enhancement based on the Teager energy operator." *IEEE Signal Processing Letters* 8(1): 10-12.
- [7] S. Valle, W. Li and S.J. Qin (1999), "Selection of the number of principal components: the variance of the reconstruction error criterion with a comparison to other methods", *Industrial and Engineering Chemistry Research* 38, pp. 4389-4401.
- [8] Rousseeuw, P.; Van Driessen, K. (1999), "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, 41, pp. 212-223.
- [9] Bakshi, B. (1998), "Multiscale PCA with application to MSPC monitoring," *AIChE J.*, 44, pp. 1596-1610.
- [10] Qin, S. J.; Dunia R.(1998), Determining the number of principal components for best reconstruction. In *IFAC DYCOPS'98*, Greece, June 1998.
- [11] Seok, J. W. (1997). Speech enhancement with reduction of noise components in the wavelet domain. International conference on audio,speech and signal processing of IEEE.
- [12] Sheikhzadeh, H. (2001). An improved wavelet-based speech enhancement system. *Eurospeech*.
- [13] Donoho, D. L. (1995). "Denosing by soft thresholding." *IEEE Trans. on Information Theory* 41(3): 613-627.