2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)

Feature Selection using Singular Value Decomposition for Stop Consonant Classification

Domy Kristomo
Dept. of Computer Engineering
STMIK AKAKOM Yogyakarta
Yogyakarta, Indonesia
domy@akakom.ac.id

Risanuri Hidayat
Dept. of Electrical Engineering and Information Technology
Universitas Gadjah Mada
Yogyakarta, Indonesia
risanuri@ugm.ac.id

Indah Soesanti
Dept. of Electrical Engineering and Information Technology
Universitas Gadjah Mada
Yogyakarta, Indonesia
indahsoesanti@ugm.ac.id

Abstract—**In the research field of pattern recognition, especially in the signal classification, the process of determining the suitable and the relevant feature is important to obtain the better classification result. This paper presents the feature selection of stop consonant by using singular value decomposition (SVD) and the classification by applying multi-layer perceptron (MLP). The feature sets were derived by using the wavelet packet transform (WPT) at 4th decomposition level with daubechies2 wavelet family, and the WPT-based feature after being dimensionality reduced by using SVD with varying reduction index which denotes as SVD1 and SVD2. Each CVC stop consonant is windowed at a certain length of duration to obtain a relevant CV unit. The experimental result shows that SVD gives improved classification scores.**

Keywords-Feature selection; singular value decomposition; stop consonants; wavelet.

I. Introduction

Research in speech recognition which focuses on stop consonants had been done for several decades [1]–[5]. In 1955 [1], the research for analyzing the voiced stops /g, d, and b/ has been done. In 1957 [2], examination of stop consonants time variation and at short-time spectra at burst of stop release was conducted. In 1979 [3], The research was done by measuring the spectrum sampled of a large number of consonant-vowel (CV) as well as vowel-consonant (VC) syllable containing both voiced and voiceless stop consonants uttered by some speakers [3]. There are six stop consonants in the Indonesian language (*/g/, /d/, /b/, /k/, /t/, and /p/*) [6], [7] which is same with English but different in the pronunciation. Recent speech classification system has achieved high accuracy, especially for classifying vowel, digit, word, etc. However, classifying the stop consonant is more difficult than classifying others.

One of challenge in the speech classification system is to determine the efficient feature with low dimension in order to minimizing computational time [8], [9]. Thus, it requires the appropriate feature selection techniques which can reduce the dimensionality of feature in order to obtain the optimal result of classification. Various feature selection techniques such as the Principle Component Analysis (PCA) [10], Singular Value Decomposition (SVD) [11]–[16], Correlation-based Feature Selection (CFS), and the other feature selection methods have been adopted for selecting feature of the speech signal by previous researcher. In [10], the PCA was used to reduce dimensionality of the Hindi Stop consonant features. The 39 generated features of proposed feature extraction techniques namely WBSP and the conventional MFCC reduce to 24 principal features by using PCA. The result showed that the feature selection give the improvement to the classification accuracy. In [11], the SVD was used to reduce the Mel Frequency Band Energy Coefficients (MFBECs) feature. The experimental results showed that the SVD give improvement to classification accuracy of the vocal fold pathology signal. In [12], the SVD was used reduce the matrices of entropy calculated from the wavelet packets (WP) method for each class of voiceless plosive consonants /k/, /t/, /p/. In [13], an alternative method for selecting feature using SVD followed by QR Decomposition with Pivoting of Column (SVD-QRcp) was proposed. The result showed that SVD-QRcp outperforms F-Ratio compared to the MFCC and LFCC. In [14], the SVD was combined with mean-normalized Stochastic Gradient Descent (MN-SGD) for classifying speech from TIMIT database. In [15], SVD was utilized on the weight matrices in trained deep neural network (DNNs). In [16], SVD was combined with spectra-perceptual linear prediction (RASTA-PLP) for analyzing Malay language speech signals.

In this study, we used SVD to reduce the feature dimension of WPT-based feature for classifying the Indonesian Stop consonants in CV context. Three feature sets are performed in this study, namely the original WPT (Feature Set 1), WPT after selecting feature using SVD with 1 index of reduction denotes as WPT+SVD1 (Feature Set 2), and WPT after the selection using SVD with 2 index of reduction denotes as WPT+SVD2 (Feature Set 3). The coefficient wavelet family used is daubechies2 (db2) at 4th decomposition level, and the type of classifier used is multi-layer perceptron (MLP).

II. Methodology

A. Speech Data

The speech data used in this study were obtained from normally utterances by 6 male Indonesian native speakers with aged between twenty-five and thirty-five years. They were asked to utter a set of CVC syllables of stop consonants. There were in total 18 different syllables recorded with 5 times repetitions of each type of syllable; it means that each speaker

978-1-5386-5163-6/18/$31.00 ©2018 IEEE

utter 90-syllable data. Finally, the stop consonant data of 540 utterances with 8 kHz of frequency sampling were collected.

The next step is the signal segmentation process. In order to obtain the precision and relevant segmented signal, each CVC unit was manually segmented to form the CV syllables sound signal with the length 60 ms [10]. The release burst of the associate consonant was used as a starting point for the segmentation until the following vowel steady state [10]. After the segmentation process, the next step was the peak signal normalization.

B. Wavelet

The difficult part to of stop consonants lies in the non-stationary structure of the signal in the burst region and the transition (i.e. the transition of V-C or C-V) [17]. The Wavelet Transform (WT) is a powerful tool which is suitable for representing both the short duration and transient event in the stops signal compared to Fourier Transform based.

In this work, WPT at 4th level of decomposition was conducted. WPT allows a balanced binary tree structure by performing decomposition to both the higher frequency sub-bands or so called ‘detail’ and the lower frequency sub-bands or so called ‘approximation’ to provide a better frequency resolution of speech signal to be analyzed. In the WPT, both approximation and detail were used to generate the feature.

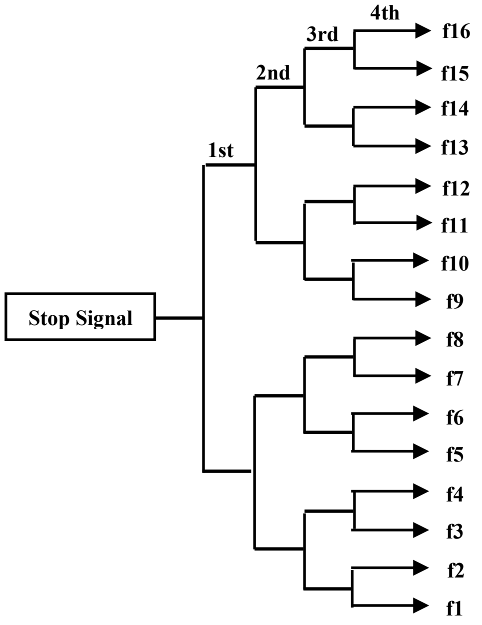


Figure 1. Sixteen sub-band of WPT feature extraction.

The sub-band tree structure of WPT feature extraction method used in this work is shown in Fig. 1. The stop consonants in CV syllables context in this study have 8 kHz of sampling frequency, giving 4 kHz bandwidth signal. A frame size of 60 ms has been used to derive the WPT. All these frequency band was decomposed using full 4-level WP to obtain sixteen sub-band each of 250 Hz. So all the frequency bands obtained after decomposition from the higher to the lower frequency band were 3.75-4 kHz (f16), 3.5-3.75 kHz (f15), 3.25-3.5 kHz (f14), 3-3.25 kHz (f13), 2.75-3 kHz (f12), 2.5-2.75 kHz (f11), 2.25-2.5 kHz (f10), 2-2.25 kHz (f9), 1.75-2 kHz (f8), 1.5-1.75 kHz (f7), 1.25-1.5 kHz (f6), , 1-1.25 kHz (f5), 0.75-1 kHz (f4), 0.5-0.75 kHz (f3), 0.25-0.5 kHz (f2), and 0-0.25 kHz (f1), respectively [18].

C. Singular Value Decomposition (SVD)

SVD is a factorization of a real matrix as well as a complex matrix. It is a classic and powerful algorithms in linear algebra which was developed by differential geometers for dimensionality reduction and rank in pattern recognition [8]. Dimensionality reduction can be performed by using following equation.

$$\mathrm{X} \simeq \widehat{\mathrm{X}} = [\boldsymbol{u}_0, \boldsymbol{u}_1, \dots, \boldsymbol{u}_{k-1}] \begin{bmatrix} \sqrt{\lambda_0} v_0^H \\ \vdots \\ \sqrt{\lambda_{k-1}} v_{k-1}^H \end{bmatrix} \qquad (1)$$

$$= U_k[\boldsymbol{a}_0, \boldsymbol{a}_1, \dots, \boldsymbol{a}_{n-1}]$$

Where U_k contains the first k columns of U and the k-dimensional vectors $\boldsymbol{a}_{i,}$. From Eq. 1, each column vector, $\boldsymbol{x}_i$ of X, can be approached as

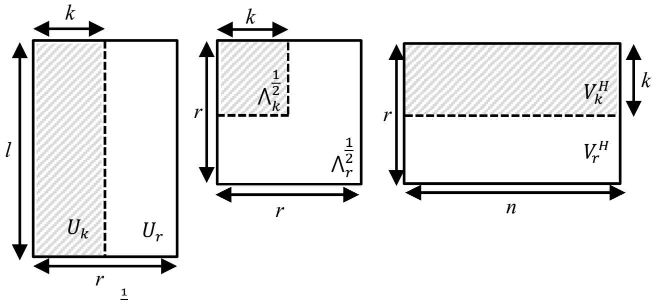
$$X_i \simeq U_k a_i = \sum_{m=0}^{k-1} u_m a_i(m), \qquad i = 0, 2, \dots, n-1 \qquad (2)$$

Where $a_i(m)$, $m = 0, 1, \dots, k-1$, represent the elements of the respective vector $\boldsymbol{a}_i$. Because of the columns $\boldsymbol{u}_i$ orthonormality, $i = 0, 1, \dots, k-1$, of $U_{k,}$ it can be directly to examine that

$$||X_i - X_j|| = ||U_k\left(a_i - a_j\right)|| = ||\sum_{m=0}^{k-1} U_m\left(a_i(m) - a_j(m)\right)|| \qquad (3)$$

$$= ||a_i - a_j||, \quad i,j = 0, 1, \dots, n-1$$

Figure 2 shows an interpretation of the matrix products associated to SVD. Where and $\Lambda_k^{\frac{1}{2}}$ is the diagonal matrix which has square roots elements of the respective k singular values [8].



$$\mathrm{X} = U_r \Lambda_r^{\frac{1}{2}} V_r^H$$

$$\widehat{\mathrm{X}} = U_k \Lambda_k^{\frac{1}{2}} V_k^H$$

Figure 2. Illustration of the matrix results using SVD. In the estimation of X by $\widehat{\mathrm{X}}$.

D. Multi-layer Perceptron (MLP)

The MLP which is also called ANN is a common used classifier in pattern recognition system. It has a training algorithm namely back propagation algorithm which is a supervised learning network based on gradient descent learning rule for correcting error in a backward direction. For validating our experimental result in classification process, we used k-fold cross validation [19].

III. Result and Discussion

A. Feature Selection using SVD

The SVD performs a factorization of a stop consonant feature matrix (N x 24) into three matrices which is expressed as USV^T, where U represents an N x N orthogonal matrix, S represents N x 24 diagonal matrix with singular values of original feature matrix on its diagonal, and V represents a 24 x 24 orthogonal matrix. V^T is the Hermitian transpose of V. The column field of the matrix U is termed as the left singular vectors, the diagonal elements of S is termed as the singular values, and the column field of the matrix V is termed as the right singular vectors. The stop consonants features matrix is factorized using SVD into three matrices (USV^T) with variation of index reduction in each singular value then each reduced matrix is multiplied to form a new stop consonants feature matrix. The decomposition of a stop consonant feature matrix with 1-index SVD reduction or SVD1 can be described as follows:

$$U^1 = \begin{bmatrix} U_{1,1} & \cdots & U_{1,539} & 0 \\ \vdots & \ddots & \vdots & 0 \\ \vdots & \cdots & \vdots & 0 \\ U_{540,1} & \cdots & U_{540,539} & 0 \end{bmatrix} \qquad S^1 = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & 0 & \lambda_{23} & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

$$V^{T1} = \begin{bmatrix} V_{1,1} & \cdots & \cdots & V_{1,24} \\ \vdots & \ddots & \cdots & \vdots \\ V_{23,1} & \vdots & \ddots & V_{23,24} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

B. Classification

Table I shows the classification result of three feature extraction methods by using MLP with 10-fold cross validation method (10 FCV). SVD1 represents the 1-level of singularity index reduction whereas SVD2 represents the 2-level of singularity index reduction. In vowel */a/* case, the score for the consonant of velar */k/* is 73.3%, 80%, and 90% by using WPT, WPT+SVD1, WPT+SVD2 based featured, respectively. For consonant */g/*, the highest score was 73.3% by using WPT+SVD2. For consonant */b/*, the highest score is only 56.7% by using WPT and WPT+SVD. The performance degraded on the consonant /b/ due to the sound signal similarity with /p/ which has the same place of articulation (labial POA). The previous study done by Sharma et al. has shown that the classification of stop consonant /b/ was more difficult than other as shown by classification score of 50%, 34.61%, 0% (with following vowel */a, i, u/*) by using MFCC and 57.69%, 19.12%, 88.46% (with following vowel */a, i, u/*) by using WBSP [10]. For consonant */d/*, the highest score is 80 by using WPT+SVD. For consonant */t/* the highest score is 90% by using WPT and also the highest score in the experiment.

In the vowel */i/* case, the average classification score for WPT, WPT+SVD1, and WPT+SVD2 were 62.23%, 61.65%, and 61.68%, respectively. In the vowel */u/* case, the average classification score for WPT, WPT+SVD1, and WPT+SVD2 were 61.67%, 62.22%, and 66.65%, respectively. From the result in the Table I, it can be observed that WPT+SVD2 has a better performance than WPT and WPT+SVD. It indicates that SVD at the level 2 can perform more discriminate features and improve the classification performance.

TABLE I. Classification Result of Each Stop Consonant for Each Feature Extraction Techniques

Methods	**Vowels**	**Consonants**						**Average % classification**
		/k/	***/g/***	***/b/***	***/d/***	***/p/***	***/t/***	
WPT	*/a/*	73.3	60	50	76.7	76.7	90	71.11
	/i/	70	70	56.7	60	50	66.7	**62.23**
	/u/	63.3	70	56.7	66.7	70	43.3	61.67
WPT+SVD1	*/a/*	80	53.3	56.7	80	73.3	90	72.22
	/i/	63.3	73.3	53.3	63.3	50	66.7	61.65
	/u/	70	70	53.3	60	70	50	62.22
WPT+SVD2	*/a/*	**90**	**63.3**	50	76.7	73.3	83.3	**72.77**
	/i/	70	70	50	**66.7**	46.7	66.7	61.68
	/u/	**80**	**73.3**	53.3	53.3	**76.7**	**63.3**	**66.65**

IV. Conclussion

In this paper, a feature selection technique based on Singular Value Decomposition (SVD) combined with Wavelet Packet Transform (WPT) was performed for classifying the Indonesian stop consonants in the context of CV syllable. Based on the classification result, it indicated that SVD gives improved classification score. The result showed that the WPT+SVD2 based feature outperforms the WPT and WPT+SVD1 as shown by average classification percentage of 67.03% versus 65% and 65.3%, respectively. In the future work, it is recommended to use the bigger data of the stop consonants, to try the higher reduction index in the SVD, and to compare with different feature selection technique.

Acknowledgment

This work was supported by Ministry of Research, Technology and Higher Education (RISTEKDIKTI) of Indonesia.

References

[1] P. C. Delattre, A. M. Liberman, and F. S. Cooper, "Acoustic Loci and Transitional Cues for Consonants," *J. Acoust. Soc. Am.*, vol. 27, no. 4, pp. 769–773, 1955.

[2] M. Halle, G. W. Hughes, and J.-P. Radley, "Acoustic properties of stop consonants," *J. Acoust. Soc. Am.*, vol. 29, no. 1, pp. 107–116, 1957.

[3] S. E. Blumstein and K. N. Stevens, "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristic of stop consonants," *J. Acoust. Soc. Am.*, vol. 66, no. 4, pp. 1001–1017, 1979.

[4] K. N. Stevens, S. Y. Manuel, and M. Matthies, "Revisiting place of articulation measures for stop consonants : implications for models of consonant production," in *Proceeeding International Congress of Phonetic Sciences*, 1999, pp. 1117–1120.

[5] A. Suchato, "Classification of Stop Consonant Place of Articulation," *Ph.D. Diss. Submitt. to Massachusetts Inst. Technol.*, 2004.

[6] F. L. Hardjono and R. A. Fox, "Stop Consonant Characteristics: VOT and Voicing in American-Born-Indonesian Children's Stop Consonants," The Ohio State University, 2011.

[7] A. Hasan and S. Dardjowidjojo, *Tata Bahasa Baku Bahasa Indonesia (Indonesian Grammar)*, Vol.3. Jakarta: Balai Pustaka, 2003.

[8] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. United States ofAmerica, 2009.

[9] M. A. Anusuya and S. K. Katti, "Front end analysis of speech recognition: a review," *Int. J. Speech Technol.*, vol. 14, no. 2, pp. 99–145, Jun. 2011.

[10] R. P. Sharma, O. Farooq, and I. Khan, "Wavelet based sub-band parameters for classification of unaspirated Hindi stop consonants in initial position of CV syllables," *Int. J. Speech Technol.*, vol. 16, no. 3, pp. 323–332, 2013.

[11] M. Hariharan, M. P. Paulraj, and S. Yaacob, "Identification of vocal fold pathology based on Mel frequency band energy coefficients and singular value decomposition," *ICSIPA09 - 2009 IEEE Int. Conf. Signal Image Process. Appl. Conf. Proc.*, pp. 514–517, 2009.

[12] E. Lukasik, "Wavelet packets based features selection for voiceless plosives classification," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2000, pp. 689–692.

[13] S. Chakroborty and G. Saha, "Feature selection using singular value decomposition and QR factorization with column pivoting for text-independent speaker identification," *Speech Commun.*, vol. 52, no. 9, pp. 693–709, 2010.

[14] C. Cai and K. Su, "A Combination of Multi-state Activation Functions , Mean-normalisation and Singular Value Decomposition for Learning Deep Neural Networks," pp. 0–7, 2015.

[15] S. Xue, H. Jiang, L. Dai, and Q. Liu, "Speaker Adaptation of Hybrid NN/HMM Model for Speech Recognition Based on Singular Value Decomposition," *J. Signal Process. Syst.*, vol. 82, no. 2, pp. 175–185, 2016.

[16] M. Amirul, A. Zulkifly, and N. Yahya, "Relative Spectral-Perceptual Linear Prediction (RASTA-PLP) Speech Signals Analysis Using Singular Value Decomposition (SVD)," pp. 3–7, 2017.

[17] B. Gidas and A. Murua, "Classification and clustering of stop consonants via nonparametric transformations and wavelets," in *1995 International Conference on Acoustics, Speech, and Signal Processing*, 1995, pp. 872–875.

[18] D. Kristomo, R. Hidayat, and I. Soesanti, "Wavelet Based Feature Extraction for the Indonesian CV Syllables Sound," *TELKOMNIKA Indones. J.*, vol. 16, no. 3, pp. 925–933, 2018.

[19] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," *Int. Jt. Conf. Artif. Intell.*, vol. 14, no. 12, pp. 1137–1143, 1995.