Dimensionality Reduction of Speech Signals using Singular Value Decomposition and Karhunen-Loeve

Domy Kristomo[1] and Yudhi Kusnanto[2]
[1]*Department of Informatics Engineering, STMIK AKAKOM, Yogyakarta, Indonesia*
[2]*Department of Computer Engineering, STMIK AKAKOM, Yogyakarta, Indonesia*

Keywords: Speech, singular value decomposition, wavelet, karhunen-loeve.

Abstract: The design of speech recognition system requires the reliable feature in order to improve the performance of speech recognition system. Thus it requires the efficient feature in order to minimizing computational time and to obtaining the optimal classification result. This paper proposes the combined method of various time-frequency feature extraction techniques with singular value decomposition (SVD) for extracting, selecting, and classifying the Indonesian stop consonants in initial position of Consonant-Vowel (CV) syllables as well as the word of stop consonant. The results of the study are divided into two parts, first: the implementation of the extraction method and selection of features based on Singular Value Decomposition (SVD) on stop consonant data, second: the implementation of the extraction method and selection of features based on Singular Value Decomposition (SVD) on word sound data formed by stop consonants. The experimental result shows that SVD gives improved the classification scores. The classification of stop consonants is more difficult than classifying of word of stop consonants.

1 INTRODUCTION

Speech recognition technology is currently growing rapidly. Speech recognition technology enables a computer to recognize and understand language spoken by humans (speakers). The technology is currently widely applied in various applications, such as security systems, smart devices, smartphones, and so on. Some researches related to speech recognition have been carried out by previous researchers, however, research to recognize the sounds of stop consonant words in Indonesian as well as to apply the method of dimensionality reduction to the voice data is still very limited and received less attention from local researchers.

In research related to speech recognition systems, the main stages commonly used by researchers are to be able to classify or recognize sound cues, including: preprocessing, segmentation, feature extraction, feature selection, and classification or recognition. Feature selection becomes an important stage in speech signal recognition system, this is intended to determine the featuress that are efficient, relevant and appropriate so that the optimal speech recognition or classification results are obtained. Feature selection is a process of selecting a subset of original features so that the dimension / size of features is optimally reduced according to evaluation criteria. Dimension features that are too large will affect the performance of classification and computational load, because the number of features that many will make the number of parameters in the classifier (for example the number of synaptic weights in the Neural Network). Therefore, the urgency of this research is to choose the right traits through the Singular Value Decomposition (SVD) and Karhunen-Loeve (KL) -based dimensionality reduction methods which have not been done by previous (local) researchers. So that this research is expected to be able to provide new references in research in the field of Indonesian speech recognition and also improve the performance of the Indonesian speech recognition system through the dimensionality reduction method.

Singular Value Decomposition (SVD) based, Karhunen-Loeve (KL) or Principal Component Analysis (PCA) based methods, Correlation based Feature Selection (CFS), and other feature selection methods have been used by previous researchers to reduce feature dimensions in data 1 dimension (1-D) and 2-dimensional (2-D) data. In research (Hariharan et al., 2009), SVD is used to reduce the features of Mel Frequency Band Energy Coefficients (MFBECs).

Kristomo, D. and Kusnanto, Y.
Dimensionality Reduction of Speech Signals using Singular Value Decomposition and Karhunen-Loeve.
DOI: 10.5220/0009432200780084
In *Proceedings of the International Conferences on Information System and Technology (CONRIST 2019)*, pages 78-84
ISBN: 978-989-758-453-4

The experimental results show that SVD provides improved performance in the pathology sound signal classification. In research (Lukasik, 2000), SVD is used to reduce the entropy matrix of the Wavelet Packet (WP) method for each class of plosive consonant sounds /k/, /t/, /p/. In research (Chakroborty and Saha, 2010), an alternative to feature selection using QR decomposition with column pivoting (SVD-QRcp) is proposed. The experimental results show that SVD-QRcp increases the ratio compared to MFCC and LFCC. The KL or PCA method has also been applied to medical cues such as heart sounds (SaraçOğLu, 2012)(Yazdanpanah et al., 1999). The study (SaraçOğLu, 2012) used the PCA method to select the features of heart sounds for heart valve abnormality, then classified using the Hidden Markov Model (HMM). The results showed that by selecting / reducing the dimensions of features using PCA can improve the performance of the classification of heart sound signals. Research (Yazdanpanah et al., 1999) analyzes the performance of four different approaches to feature selection using the Karhunen-Loeve Expansion (KLE) method to select the most discriminant feature set to classify the status of bioprostetic heart valves. Previous researchers analyzed the characteristic of SVD in a number of word similarity extraction tasks (Gamallo and Bordag, 2011). The results lead them to conclude that SVD makes the extraction less computationally efficient and much less precise than other more basic models for extracting word data.

On the side of speech recognition in Indonesian has also been done by previous researchers (Yessivirna and Marji, 2013) (Nafisah et al., 2016)(Fachrie and Harjoko, 2015). In research (Yessivirna and Marji, 2013), spectral domain-based spectral (Spectral Centroid and Spectral Flux) feature extraction methods using the K-Nearest Neighbor (KNN) classification system to classify the sounds of the word "I" that are spoken normally by adults. The results of this study indicate the accuracy of the sound classification based on gender with the KNN method is quite good. The lowest accuracy in the trial with a frame width value of 1024, frame shift 31.25%, and an alpha value of 0.97 is 71.2% and the highest accuracy is 77.1%. In research (Nafisah et al., 2016), the MFCC-based feature extraction method with a variety of window functions is used to classify word sound data from an isolated word database in Indonesian (Database for Isolated Word) using the Back Propagation Neural Network (BPNN) classification system. The results showed that the MFCC method and the rectangle window (rectwin) function in the frame blocking process can improve the performance of Automatic System Recognition (ASR). In research (Fachrie and Harjoko, 2015), the MFCC method combined with natural logarithm of Frame Energy (lnFE) was used to extract features digit word sounds in Indonesian by using Elman Recurrent Neural Network (ERNN) as the classification. In research (Ferdiansyah and Purwarianti, 2011), the ASR system was developed from an existing system model to recognize words in Indonesian. However, there are no studies that apply the SVD and Karhunen Loeve feature selection method for Indonesian word sound data.

This research is a development from previous research that applies the SVD method to reduce the dimension of Indonesian speech features (Kristomo et al., 2018), which is still limited to the second reduction level and only to the stop consonant data.This study aims to obtain efficient and effective traits in Indonesian word sound cues by applying the Singular Value Decomposition (SVD) and Karhunen-Loeve (KL) -based trait selection method. So that the trait selection is expected to be able to improve the performance of speech recognition systems. This research will be divided into 4 main stages namely preprocessing, feature extraction, feature dimensionality reduction, and classification, which this research will be more focused on the dimensionality reduction stage.

2 MATERIAL AND METHODS

2.1 Database

The research was conducted by firstly collected the speech database from several speakers that were used for training and testing to the system. Stop consonant sound data used in this study were 560 utterances and for word data were 300 utterances. The stop consonant data was segmented for 60 ms while for word data was segmented for 480 ms. The set of word are listed in Table 1.

Table 1: The List of Stop Consonants Words Data.

Words in Indonesian	English Translation
Kakak	Older sibling/cousin
Tutup	Closed
Bibit	Seed
Papan	Board
Duduk	Sit
Gigit	Bite

2.2 Feature Extraction

2.2.1 Time-frequency Features

We used three main method namely Wavelet, Autoregressive Power Spectral Density (AR-PSD), and Renyi Entropy. The wavelet transform (WT) has a strength to localize the transient events emergence (Boccaletti et al., 1997). It is considered to be the best in describing the signal anomaly, pulses, and other events which appear in the brief duration of the signal (Fugal, 2009), e.g. speech signal of the stop consonants. In conducting this feature extraction process, DWT was used at the decomposition level-7. In addition, a lower frequency band or also referred to as approximation was used in the process of DWT decomposition. The decomposition which was conducted as the 7th level gave the lowest frequency band of 0–31.25 Hz. Therefore, since it results in a very low frequency, there is no more decomposition was conducted as such frequency would be insignificant and would not have any discriminatory information. After the decomposition of the the speech signal frequency sampling of 8 kHz, the frequency bands obtained were 2000-4000, 1000-2000, 500-1000, 250-500, 125-250, 62.5-125, 31.25-62.5, and 0-31.25 Hz.

In this study, PSD using Yule-Walker AR algorithm was performed. This algorithm is used for transformation of the speech signal from time domain to frequency domain. Whereas Renyi Entropy is used to obtain the speech signal features in time domain (Kristomo et al., 2016).

2.2.2 Wavelet

The sub-band tree structure of WPT feature extraction method adopted in this work refers to the previous research. The words data in this research have 8 kHz of sampling frequency, giving 4 kHz bandwidth signal. A frame size of 480 ms has been used to derive the WPT. All these frequency bands were decomposed using full 4-level WP to obtain sixteen sub-bands each of 0.25 kHz. So the sixteen frequency bands obtained after decomposition from the lower to the higher frequency band were 0-0.25 kHz (f1), 0.25-0.5 kHz (f2), 0.5-0.75 kHz (f3), 0.75- 1 kHz (f4), 1-1.25 kHz (f5), 1.25-1.5 kHz (f6), 1.5- 1.75 kHz (f7), 1.75-2 kHz (f8), 2-2.25 kHz (f9), 2.25-2.5 kHz (f10), 2.5-2.75 kHz (f11), 2.75-3 kHz (f12), 3-3.25 kHz (f13), 3.25-3.5 kHz (f14), 3.5-3.75 kHz (f15), and 3.75-4 kHz (f16) respectively.

2.3 Dimensionality Reduction

This study uses the Singular Value Decomposition (SVD) and Karhunen-Loeve (KL) method to select the feature of Indonesian word sound signal.

2.3.1 Singular Value Decomposition

Singular Value Decomposition (SVD) is a matrix factorization that can be used to for a real matrix and a complex matrix. SVD is a classic and reliable method in linear algebra that is used for dimension reduction and ranking in pattern recognition. SVD allows factorization of feature matrices into three matrices denoted as USVT. Where U represents N x N orthogonal matrix (N = amount of data), S represents N x n diagonal matrix with singular values of the original feature value matrix on the diagonal, and V shows the orthogonal matrix n x n (n = number of features). VT is the Hermitian transpose of V. Figure 1 shows the interpretation of the matrix product related to SVD (Theodoridis and Koutroumbas, 2009).

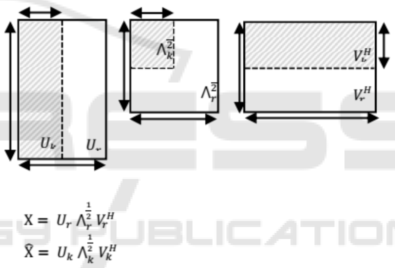

Figure 1: The factorization and reduction ilustration using SVD in the estimation of X by X.

2.3.2 Karhunen Loeve

Karhunen-Loeve (KL) or Principal Component Analysis (PCA) is one of the most popular methods for feature generation and dimension reduction in pattern recognition.

The step in the selection of features using the KL transformation is shown in equations 1 to 8. First of all a data matrix in the form:

$$x_i = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, x = [x_1, x_2, x_3, \ldots, x_n] \quad (1)$$

Then the covariance (C_x) of the data matrix is calculated:

$$C_x = \frac{1}{n} \Sigma_{k=1}^{n} X_k X_k^t - (m_x m_x^t) \quad (2)$$

Where m_x is the average of the data matrix

$$m_x = \frac{1}{n}\Sigma_{k=1}^{n} X_k \tag{3}$$

After (C_x) is obtained then the eigen value (λ) is calculated

$$det|c_x - \lambda i| \tag{4}$$

And also the eigen vector (e)

$$(c_x - \lambda i)v = 0 \tag{5}$$

Then the eigen value (λ) is sorted from the largest to the smallest

$$\lambda_1 > \lambda_2 \tag{6}$$

Based on the order of the eigen value (λ) arrange e_i^t into a transformation matrix as follows

$$A = \begin{bmatrix} e_1^t \\ e_2^t \end{bmatrix} \tag{7}$$

Then transform the data matrix in a way

$$y1 = A(x_i - m_x) \tag{8}$$

3 RESULTS AND DISCUSSION

The results of the study are divided into two parts The first result is applying the SVD method to speech signal data with a reduction index variation from 1 to 30 for three types of feature sets namely WS, WPSDS, and WRPSDS. In this experiment, we compare the performance of the time-frequency features without feature reduction using SVD and the time-frequency features with feature reduction using SVD. The feature set without dimensionality reduction is denoted as WS, WPSDS, and WRPSDS, whereas the feature set with dimensionality reduction is denoted as WS+SVD2, WPSDS+SVD1, and WRPSDS+SVD10 as shown in Figure 2.

Feature extraction	Following vowels	/k/	/g/	/b/	/d/	/p/	/t/	Average % classification
WS	/a/	63.3	50	63.3	56.7	70	80	63.88
	/i/	**76.7**	66.7	40	40	63.3	73.3	60
	/u/	80	70	36.7	53.3	56.7	23.3	59.07
WPSDS	/a/	**93.3**	46.7	60	**70**	73.3	83.3	71.1
	/i/	66.7	63.3	50	40	63.3	80	60.55
	/u/	**86.7**	70	50	53.3	60	**66.7**	64.45
WRPSDS	/a/	**93.3**	46.7	70	53.3	**80**	76.7	68.33
	/i/	66.7	63.3	46.7	50	60	80	60.56
	/u/	**86.7**	70	36.7	63.3	**70**	50	62.78
WS+SVD2	/a/	70	**56.7**	63.3	56.7	76.7	76.7	66.68
	/i/	66.7	60	43.3	36.7	66.7	70	57.23
	/u/	83.3	73.3	43.3	60	60	40	59.98
WPSDS+SVD1	/a/	76.7	**56.7**	**73.3**	**70**	70	**86.7**	**72.23**
	/i/	73.3	**73.3**	56.7	60	53.3	80	**66.1**
	/u/	83.3	76.7	**53.3**	56.7	**70**	**66.7**	67.78
WRPSDS+SVD10	/a/	83.3	50	66.7	60	73.3	80	68.88
	/i/	70	70	53.3	56.7	60	**83.3**	65.55
	/u/	80	**80**	**53.3**	**70**	66.7	60	**68.33**

Figure 2: Classification Result of stop consonants using 10-Fold Cross Validation.

From the result shown in Figure 2, it can be seen that SVD gives improved the classification scores as shown by accuracy of 72.23%, 66.1%, and 68.33% for WPSDS+SVD1 /a, i/ and WRPSDS+SVD10 /u/, respectively. However, some parts of stop consonant syllables shows better result without feature selection using SVD, such as /ki/ in WS, /ka, ku/ in WPSDS and WRPSDS; /da/ in WPSDS; /pa, pu/ in WRPSDS; and /tu/ in WPSDS.

Based on Figure 1 it can be seen that the optimal classification results are achieved in the reduction indices 2, 1, and 10 for the WS, WPSDS, and WRPSDS feature sets respectively. The WS feature set starts to decrease continuously at the 10th reduction index and reaches the minimum classification results on the 27th reduction index and so on. The results of WRPSDS classification are better than WPSDS and WS but in certain reduction indices the results of WS classification are better than WRPSDS.

The second research result is applying the SVD method to the word voice signal data with a reduction index variation from 1 to 25 with the Wavelet Packet Transform (WPT) feature extraction method in decomposition 4. The singular values and the matrix reduction process are listed in descending order as follow (Equation 8 to 10):

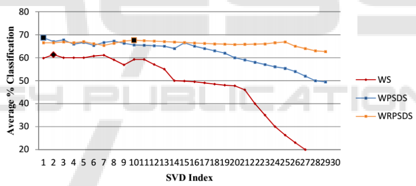

Figure 3: Classification results of stop consonants by using three sets of features with SVD reduction index variations.

$$U^1 = \begin{bmatrix} U_{1,1} & \ldots & U_{1,299} & 0 \\ \vdots & \ddots & \vdots & 0 \\ \vdots & \ldots & \vdots & 0 \\ U_{30Q1} & \ldots & U_{30Q299} & 0 \end{bmatrix}$$

$$U^1 = \begin{bmatrix} -0.03834 & \ldots & -0.0406 & 0 \\ \vdots & \ddots & \vdots & 0 \\ \vdots & \ldots & \vdots & 0 \\ 0.003032 & \ldots & -0.03778 & 0 \end{bmatrix}$$

$$U^2 = \begin{bmatrix} -0.03834 & \ldots & 0 & 0 \\ \vdots & \ddots & 0 & 0 \\ \vdots & \ldots & 0 & 0 \\ 0.003032 & \ldots & 0 & 0 \end{bmatrix} \tag{9}$$

$$S^1 = \begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & 0 & \lambda_{23} & \vdots \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}$$

$$S^1 = \begin{bmatrix} 2.296941 & 0 & 0 & \dots & 0 \\ 0 & 1.657405 & 0 & \dots & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & 0 & 0.001605 & \vdots \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}$$

$$S^2 = \begin{bmatrix} 2.296941 & 0 & 0 & \dots & 0 \\ 0 & 1.657405 & 0 & \dots & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & 0 & 0 & \vdots \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix} \quad (10)$$

$$V^{T1} = \begin{bmatrix} V_{11} & \dots & \dots & V_{125} \\ \vdots & \ddots & \dots & \vdots \\ V_{241} & \vdots & \ddots & V_{2425} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$V^1 = \begin{bmatrix} -0.72144 & \dots & \dots & -0.00165 \\ \vdots & \ddots & \dots & \vdots \\ 0.000035 & \vdots & \ddots & -0.03348 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$V^2 = \begin{bmatrix} -0.72144 & \dots & \dots & -0.00165 \\ \vdots & \ddots & \dots & \vdots \\ 0 & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (11)$$

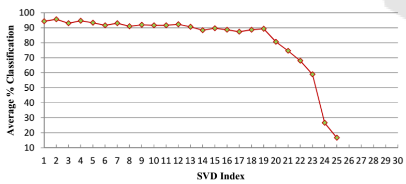

Figure 4: Classification results of word by using WPT feature set with SVD reduction index variations

Table 2: Confusion Matrices of Each Reduction Index.

1- REDUCTION INDEX

	A	B	C	D	E	F
A	48	0	0	2	0	0
B	0	50	0	0	0	0
C	0	0	43	0	0	7
D	2	0	0	48	0	0
E	0	1	0	0	49	0
F	0	0	5	0	0	45

2- REDUCTION INDEX

	A	B	C	D	E	F
A	48	0	0	2	0	0
B	0	50	0	0	0	0
C	0	0	46	0	0	4
D	1	0	0	49	0	0
E	0	1	0	0	49	0
F	0	0	5	0	0	45

3- REDUCTION INDEX

	A	B	C	D	E	F
A	**46**	0	0	**4**	0	0
B	0	50	0	0	0	0
C	0	0	43	0	1	6
D	3	0	0	47	0	0
E	0	0	0	1	49	0
F	0	0	6	0	0	44

4- REDUCTION INDEX

	A	B	C	D	E	F
A	**48**	0	0	2	0	0
B	0	50	0	0	0	0
C	0	0	46	0	0	4
D	3	0	0	47	0	0
E	0	1	0	0	49	0
F	0	0	6	0	0	44

5- REDUCTION INDEX

	A	B	C	D	E	F
A	48	0	0	2	0	0
B	0	50	0	0	0	0
C	0	0	45	0	0	5
D	4	0	0	46	0	0
E	0	0	0	1	49	0
F	0	0	8	0	0	42

6- REDUCTION INDEX

	A	B	C	D	E	F
A	47	0	0	3	0	0
B	0	50	0	0	0	0
C	0	0	39	0	0	11
D	3	0	0	47	0	0
E	0	1	0	0	49	0
F	0	0	7	0	0	43

7- REDUCTION INDEX

	A	B	C	D	E	F
A	48	0	0	2	0	0
B	0	50	0	0	0	0
C	0	0	41	0	0	9
D	2	0	0	48	0	0
E	0	1	0	0	49	0
F	0	0	7	0	0	43

8- REDUCTION INDEX

	A	B	C	D	E	F
A	**46**	0	0	**4**	0	0
B	0	50	0	0	0	0
C	0	0	**40**	0	0	**10**
D	**3**	0	0	**47**	0	0
E	0	**2**	0	0	**48**	0
F	0	0	**8**	0	0	**42**

9- REDUCTION INDEX

	A	B	C	D	E	F
A	48	0	0	2	0	0
B	0	50	0	0	0	0
C	0	0	41	0	0	9
D	2	0	0	48	0	0
E	0	2	0	0	48	0
F	0	0	8	0	0	42

10- REDUCTION INDEX

	A	B	C	D	E	F
A	47	0	0	3	0	0
B	0	50	0	0	0	0
C	0	0	40	0	1	9
D	1	1	0	48	0	0
E	0	**2**	0	0	**48**	0
F	0	0	8	0	0	42

11- REDUCTION INDEX

	A	B	C	D	E	F
A	47	0	0	3	0	0
B	0	50	0	0	0	0
C	0	0	40	0	0	10
D	1	1	0	48	0	0
E	0	2	0	0	48	0
F	0	0	8	0	0	42

...25- REDUCTION INDEX

	A	B	C	D	E	F
A	**10**	**5**	**10**	**10**	**15**	**0**
B	**10**	**5**	**10**	**10**	**15**	**0**
C	**10**	**5**	**10**	**10**	**15**	**0**
D	**10**	**5**	**10**	**10**	**15**	**0**
E	**10**	**5**	**10**	**10**	**15**	**0**
F	**10**	**5**	**10**	**10**	**15**	**0**

Based on Figure 4, it appears that the best classification results are on the second SVD index that is equal to 95.67%. The classification results begin to decrease continuously at the 20th index, and achieve the lowest classification results at the 25th index. This indicates that the greater the reduction index at a certain threshold will reduce the accuracy of classification, while for a reduction that is not too large can allow an increase in the classification results, because the new matrix with a reduction value that is not too large can be more discriminatory. For the classification results with the variation of the reduction index in the form of a confusion matrix are shown in Table 2, where the class of data is as follows a = kakak, b = tutup, c = bibit, d = papan, e = duduk, f = gigit. Based on Table 2 it appears that the data class f or sound ”gigit” is always the lowest accuracy in each reduction index, this is likely due to the similarity of the sound signal / characteristic between the sound ”gigit” with data c or ”bibit”. While the highest classification results are in the data class b or ”tutup” because it has the most discriminant characteristic of other data classes. In the 2nd reduction index there was an increase in data classes c ($43->46$) and d ($48->49$), but again decreased in the 3rd reduction index for data classes c ($46->43$) and d ($49->47$), and is still changing (up and down) fluctuatively in the next reduction index.

Table 3 shows the results of word sound classification using WPT features before and after being reduced by SVD and KL feature reduction methods. Based on Table 3 it can be seen that the results of word sound classification using WPT features without reduction reached 95% whereas after reduction with the SVD method the 2nd reduction index increased to 95.67%. The use of the KL method in the reduction of feature dimensions reduces the classification accuracy level by 89%. However, the KL method is able to reduce the features to 18 which indicates that the KL method is more efficient.

Table 3: Word of stop consonants classification result.

Feature	Acc. classification (%)
WPT (25 features)	95
WPT + SVD2	95,67
WPT + KL (18 features)	89

4 CONCLUSIONS

In this paper, a dimensionality reduction method based on SVD combined with time-frequency features was performed for classifying the Indonesian stop consonants in the context of CV syllable as well as the word of stop consonants. Based on the experimental result presented in this paper, it can be concluded that the SVD gives improved the classification scores as shown by average classification rate of 68.7%, and 67.58 for WPSDS+SVD1 and WRPSDS+SVD10, respectively which are 3.34% and 3.69% increase than WPSDS and WRPSDS without dimensionality reduction by using SVD. The application of the SVD method in the dimension of word sound features, at a certain level of the reduction index (index-2) can increase the classification results, however an increase in the reduction index that is too high tends to reduce the results of the classification. Classification of stop consonants is more difficult when compared to words of stop consonant. The highest classification result for the words is 95.67%.

ACKNOWLEDGEMENTS

This work has been supported by Directorate General of Research, Technology, and Higher Education (RISTEKDIKTI) of Indonesia under "*Penelitian Dosen Pemula*" scheme with contract number B/1435.28/L5/RA.00/2019.

REFERENCES

Boccaletti, S., Giaquinta, A., and Arecchi, F. (1997). Adaptive recognition and filtering of noise using wavelets. *Physical review E*, 55(5):5393.

Chakroborty, S. and Saha, G. (2010). Feature selection using singular value decomposition and qr factorization with column pivoting for text-independent speaker identification. *Speech Communication*, 52(9):693–709.

Fachrie, M. and Harjoko, A. (2015). Robust indonesian digit speech recognition using elman recurrent neural network. *Konferensi Nasional Informatika (KNIF)*, 2015:49–54.

Ferdiansyah, V. and Purwarianti, A. (2011). Indonesian automatic speech recognition system using english-based acoustic model. In *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics*, pages 1–4. IEEE.

Fugal, D. L. (2009). *Conceptual wavelets in digital signal processing: an in-depth, practical approach for the non-mathematician*. Space & Signals Technical Pub.

Gamallo, P. and Bordag, S. (2011). Is singular value decomposition useful for word similarity extraction? *Language resources and evaluation*, 45(2):95–119.

Hariharan, M., Paulraj, M. P., and Yaacob, S. (2009). Identification of vocal fold pathology based on mel frequency band energy coefficients and singular value decomposition. In *2009 IEEE International Conference on Signal and Image Processing Applications*, pages 514–517. IEEE.

Kristomo, D., Hidayat, R., and Soesanti, I. (2018). Feature selection using singular value decomposition for stop consonant classification. In *2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, pages 432–435. IEEE.

Kristomo, D., Hidayat, R., Soesanti, I., and Kusjani, A. (2016). Heart sound feature extraction and classification using autoregressive power spectral density (arpsd) and statistics features. In *AIP Conference Proceedings*, volume 1755, page 090007. AIP Publishing LLC.

Lukasik, E. (2000). Wavelet packets based features selection for voiceless plosives classification. In *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 00CH37100)*, volume 2, pages II689–II692. IEEE.

Nafisah, S., Wahyunggoro, O., and Nugroho, L. E. (2016). An optimum database for isolated word in speech recognition system. *Telkomnika*, 14(2).

SaraçOğLu, R. (2012). Hidden markov model-based classification of heart valve disease with pca for dimension reduction. *Engineering Applications of Artificial Intelligence*, 25(7):1523–1528.

Theodoridis, S. and Koutroumbas, K. (2009). Pattern recognition, edition.

Yazdanpanah, M., Allard, L., Durand, L., and Guardo, R. (1999). Evaluation of karhunen-loeve expansion for feature selection in computer-assisted classification of bioprosthetic heart-valve status. *Medical & biological engineering & computing*, 37(4):504–510.

Yessivirna, R. and Marji, D. E. R. (2013). Klasifikasi suara berdasarkan gender (jenis kelamin) dengan metode k-nearest neighbor (knn). *Jurnal Doro*, 2(3).