International Journal of Engineering & Technology, 7 (4.40) (2018) 61-65

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Comparison of Different Wavelet Sub-Band Features in the Classification of Indonesian Stop Consonants in CV Syllable Context

Domy Kristomo[1,2]*, Risanuri Hidayat[1], Indah Soesanti[1]

[1]*Department of Electrical Engineering and Information Technology, Universitas Gadjah Mada, Indonesia*
[2]*Study Program of Computer Engineering, STMIK AKAKOM Yogyakarta, Indonesia*
**Corresponding author E-mail: domy.kristomo@mail.ugm.ac.id*

Abstract

In the research field of signal processing by using wavelet method there are some factors affecting the accuracy of recognition such as the selection of the sub-band parameter, the selection of suitable mother wavelet or coefficient, and the determination of decomposition level. This paper presents a comparative study of three wavelet-based sub-bands (WBSB) combined with the moving average energy (MAE) features for classification of Indonesian stop consonants in consonant-vowel (CV) context. Three different feature sets used in this study are the MAE of each different wavelet sub-band using mother wavelet of daubechies2. The first feature set is the MAE taken from standard wavelet packet (WP) sub-band at the 4th level of decomposition denoted as WBSB. Whereas the second and third feature sets are the MAE taken from WP which the sub-band is selected based on the previous research denoted as WBSB1 and WBSB2. For the classification of the stops sound signal after feature extraction process, two different classifiers were used, based on multi-layer perceptron and random forest. The experimental result showed that the performance rank of feature extraction method were WBSB, WBSB1, and WBSB2, respectively.

Keywords: *Feature extraction; frequency sub-band; stop consonants; wavelet.*

1. Introduction

The Wavelet Transform (WT) can be analogized as a mathematical microscope [1], [2], through which a speech signal in the time domain can be decomposed in a representation into both time and frequency. One of the challenging tasks in speech recognition is to classify the stop consonants. Several factors which make it difficult to classify are the speaker-dependent factor, short event in signal (i.e. burst in a slow time varying speech signal), variable context, and dynamic nature of stops [3]. The stop consonants are consonants that occur with the nasal cavity close, a rapid closure or opening affected, at some points, the oral cavity. Behind the point of closure, a pressure is built, which is suddenly released with the release of closure in the vocal tract. In the overall speech recognition or classification scheme, feature extraction stage plays a very important role because a better feature is good for improving the classification rate.

Based on the standard Indonesian grammar literature, Indonesian language has six stop consonants (*/p/*, */b/*, */t/*, */d/*, */k/*, and */g/*) [4], [5]; similar to English, but different from the other language i.e. Hindi which has sixteen stop consonants [3]. The features adopted for English and Hindi language may not be suitable for Indonesian due to the difference pronunciation and characteristic of each language. Thus, the study of Indonesian stop consonant features is important in order to discover their time and frequency domain characteristics. There are some differences in Indonesian and English stop articulation. In Indonesian, */k/*, */p/*, and */t/* are articulated without an aspiration or a release of air. It is completely different from "standard" American English which articulates those consonants with a release of air. There are also different place of articulation for stops between the two languages. In Indonesia, the consonants stops of */d/* and */t/* occur when the tip of tongue compresses the upper teeth making it called as dental sounds, while in English those stops are considered to be alveolar since it is produced when the blade of the tongue is on the alveolar ridge. The previous study which aims to examine the phonetic characteristics of the Indonesian stop consonants has been done by previous researcher [4]. Nonetheless, research on Indonesian stop consonants features and its classification has not been done frequently by the local researchers. Several studies related to Indonesian speech were focused on vowel [6], phoneme [7], syllable [8][9], word [10], and digit [11] sound signal.

The WT is often used as the technique to analyze the non-stationary and short duration signal i.e. the stop consonant signal, because it has a good resolution in time-frequency, hence it has good performance in representing stationary as well as non-stationary parts of the speech signal [3], [12]. In addition, wavelet based techniques have multi-resolution properties that enable them to capture stop consonants because they have sudden bursts of high frequency. One of the factors that can influence the discriminative information in the feature is the selection of the right sub-band parameter [3], [13]–[17]. In [3], a new feature extraction technique adopting wavelet-based sub-band parameters (WBSP) was proposed to classify un-aspirated Hindi stop consonants. The WT was applied to the splitting of a sound signal into 8 sub-bands of different sub-band with different variation of energy. In [13], a new WP based features have been proposed which were based on sixteen wavelet sub-band and harmonic energy features combined with WP Cepstral (WPCC) for recognizing Hindi phoneme. The result showed that the proposed technique gives significant improvement compared to the conventional MFCC. In [14], 24-band admissible wavelet packet-based feature was used to recognize Hindi phoneme. A research was performed in order to compare two wavelet-based sub-band structures in the form of two-stage two-band and the two-stage full binary decomposition structures [15]. From the comparison, it is obtained that the features produced from the signal of two-stage two-band structure is better than the signal of the two-stage full binary structure. Another comparison was also performed in [16] between the feature extraction methods; discrete wavelet transform (DWT) and wavelet packet decomposition (WPD). The comparison was evaluated using Artificial Neural Networks (ANN) which aims to classify the Isolated Spoken Words. The result shows that when DWT and ANN methods are combines, it gives a higher result than the combination of WPD and ANN methods. In [17], a comparison of the proposed feature extraction technique based on WPT by using daubechies1 at the 1st until the 4th level of decomposition with MFCC. Based on the experiment, the features extracted using wavelets show better results than MFCC for both classifiers (SVM and MLP).

However, the performance of some sub-band-based feature has not been compared to the other different wavelet-based sub-band structure. This research compared three different wavelet-based sub-band (WBSB) which is derived with moving average energy (MAE). This study used WT to extract and classify the Indonesian Stop consonants in CV syllable context. Three feature sets were performed in this study. The 1st until the 3rd feature sets were the moving average of the WT (WBSB, WBSB1, and WBSB2) with the 4th until the 5th level of decomposition. The wavelet type used was daubechies2, and the classifiers type used were multi-layer perceptron back propagation (MLP-BP) and random forest (RF).

2. Methodology

2.1. Speech Data

This research uses speech data files taken from six Indonesian native speakers with different dialects. Four male speakers are from Java, and two male other speakers are from Sumatera. Based on the literature on language-specific phonetics, Indonesian consonant can be grouped based on three factors: articulation area, articulation manner, and condition vocal record. The parts of articulation area are bilabial, labiodental, palatal, alveolar, velar, or glottal [18]. The voiced stop consonants */b, d, g/* and unvoiced stop consonants */p, t, k/* followed by the three vowels */a, i, u/* were used as a database. The consonants */g, k/* represent the place of articulation (POA) of the velar, */t, d/* represent POA of dental, and */p, b/* represent POA of the labial [3], [4]. There were in total 18 classes of syllable recorded by 5 times repetitions; it means that 90-syllable data were accumulated from each speaker. So the total numbers of the data were 6 x 90 = 540 utterances with 8 kHz sampling frequency and stored in the 16 bits mono per sample.

The next step was the signal segmentation process. To obtain the relevant and precision segmented sound signal, each CVC unit was manually segmented to form the CV syllables. Based on the acoustic study done by the previous researchers, the duration for all of the relevant acoustic parameters to form a CV unit was about 60 ms. Hence, the signal length taken for each CV unit in this study was about 60 ms. The release burst of the associate consonant was used as an initial position for the segmentation until the following vowel steady state [3].

The signal normalization process was conducted after the segmentation process. It was carried out to adjust the volume to ensure the optimal use of media distributed in the phase of recording. There are some type of audio normalization, such as loudness normalization, and peak normalization. In this study, we used peak normalization which is a process aims to produce the highest value or peak of Pulse-code modulation (PCM) samples of the analogue signal by changing the gain up to the desired level. On the other hand, loudness normalization is a process adjusting a signal's gain in order that it equals to the desired level.

2.2. Wavelet

The stop consonant has central difficulties that lay in the non-stationary and nonlinear structure of the signal in the transition and burst regions [19]. The WT is a powerful tool which is suitable for representing both the stationary and non-stationary signal compared to Fourier Transform based.

In this part, feature extraction using WPT at 4th and 5th level of decomposition was conducted. WPT produces various signal analysis since it is the generalization of DWT decomposition. It also provides more and better frequency resolution features of the speech signal by decomposing both the lower (approximation) and higher frequency bands (detail) which produces a balanced binary tree structure. In the decomposition phase of WPT, a low frequency and high frequency band were used [8].

A proper mother wavelet is important for optimum classification in feature extraction using WPT. As stated in the previous research in [3], [20], [7], [16]. Daubechies was the one of the effective mother wavelets. The Daubechies wavelet for D-2N is as follow:

$$\psi(x) := \sqrt{2}\sum_{k=0}^{2N-1}(-1)^k h_{2N-1-k}\varphi(2x-k) \qquad (1)$$

$h0,\ldots, h2N\text{-}1 \in \mathbb{R}$ are the constant filter coefficients satisfying the condition and φ is the transformation scalling function (Daubechies). The WT transformation process is resulting in a frequency domain signal. The next step was calculating the MAE from twenty samples of signal magnitude until the maximum sample of the signal magnitude.

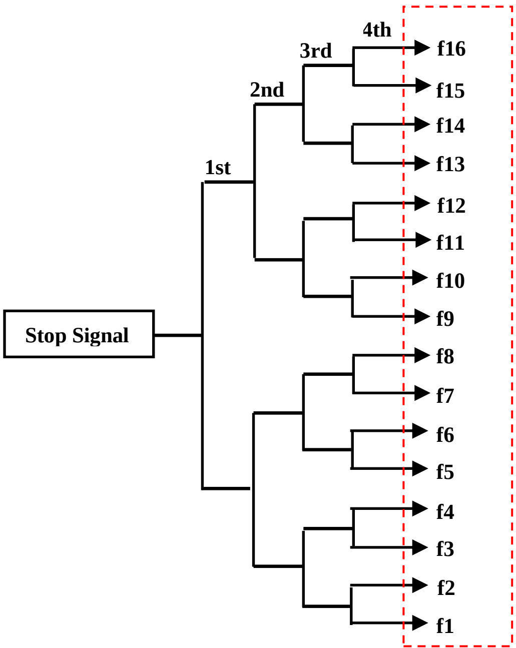

Fig. 1: Sixteen sub-bands of WBSB feature extraction.

The first wavelet structure used in this study namely WBSB was also equal to WPT. Fig. 1 shows the sub-band structure of WBSB feature extraction method. The stop consonant in CV syllables context was sampled at 8 kHz, giving 4 kHz bandwidth signal. A frame size of 60 ms was used to derive the WBSB. The whole frequency band was decomposed using full 4-level WP to obtain sixteen sub-band each of 500 Hz. So all the frequency bands obtained after decomposition 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) [7].

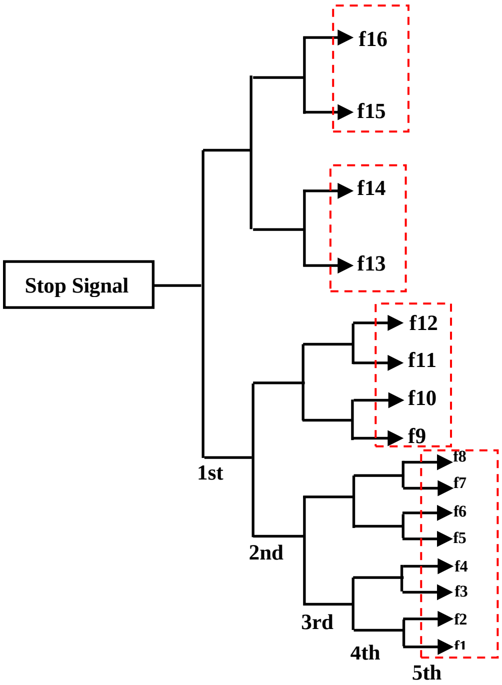

Fig. 2: Sixteen sub-bands of WBSB1 feature extraction.

Fig. 2 shows the sub-band structure of WBSB1 feature extraction method. The stop consonant in CV syllables context was sampled at 8 kHz, giving 4 kHz bandwidth signal. A frame size of 60 ms was used to derive the WBSB2. At the beginning, the whole frequency band was fully decomposed at 2nd level of WP decomposition to obtain eight sub-bands, each of Next 0–0.5 and 0.5–1 kHz frequency band was further fully decomposed into eight sub-bands, each of 125 Hz at 3rd level of WP decomposition. The resulting sub-band division finely emphasizes frequencies between 0 and 1 kHz, which normally contains a large portion of voiced signal energy. Then 1–1.5 and 1.5–2 kHz frequency bands were fully decomposed at 4th level of WP decomposition to produce four sub-bands each of 250 Hz. Four frequency bands 2–2.5, 2.5–3, 3–3.5 and 3.5–4 kHz were kept unchanged. Lastly, 16 frequency WP sub-bands were achieved. The first 12 sub-bands obtained a particular concern since its voice signals range up to 4000 Hz, while most of the speech energy usually lies below 2000 Hz. Most discriminative information of speech signal lie below 4 kHz. Therefore, WP filter is expected to be able to extract certain information from the speech sound signal by using WP decomposition [13].
Fig. 3 shows the sub-band structure of WBSB2 feature extraction method. The choosing of the sub-band referred to the previous research [17], which consist of eight selected sub-bands. This wavelet tree structure used 4th level decomposition. By using sampling frequency of 8 kHz, the selected frequency bands used after decomposition were 0–0.25 kHz (f1), 0.25–0.5 kHz (f2), 0.5–0.75 kHz (f3), 0–2 kHz (f4), 1–1.5 kHz (f5), 2–2.25 kHz (f6), 2–3 kHz (f7), and 3–3.25 kHz (f8).

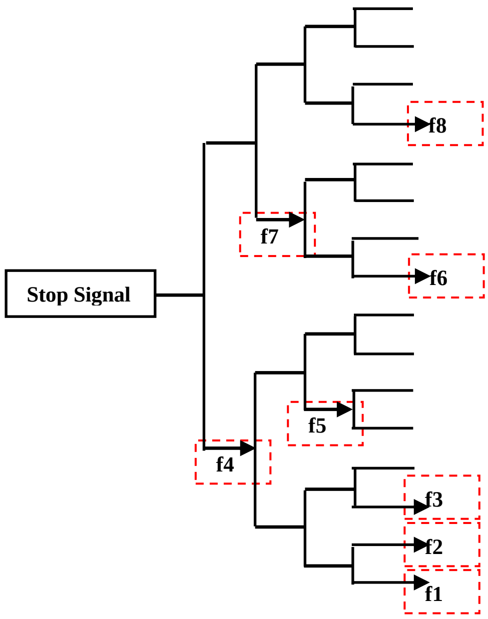

Fig. 3: Eight sub-bands of WBSB2 feature extraction.

2.3. Multi-layer Perceptron

The classifier used in this study was multi-layer perceptron (or Artificial Neural Network – ANN). MLP or ANN is a computational tool for pattern classification that has an information-processing paradigm inspired by the biological nervous system. The MLP is trained by using a learning method namely back propagation algorithm in order to solve particular problems such as classification.
The MLP-BP used in this study consists of three layers; they are input layer, hidden layer, and output layer. The input layer represents the features of each feature extraction technique (WBSB, WBSB1, or WBSB2). The number of hidden neurons was set in 55 nodes with one hidden layer. The parameters of the learning rate and the momentum were used at 0.3 and 0.2, respectively. The output layer consists of eighteen neurons, which represents the classification result of the stop consonant syllables.

2.4. Random Forest

Random forest (RF) is an algorithm used in the classification of large amounts of data. Random forest classification is carried out through a tree combination by conducting training on the sample data. The use of more trees will affect to the better accuracy obtained. Determination of classification by random forest is taken based on the results of the voting of the tree formed. Winners of the tree formed are determined by the most votes. Build a tree on a random forest until it reaches the maximum size of the data tree. However, the construction of a random forest tree is not pruned which is a method to reduce space complexity. Development is carried out by applying the random feature selection method to minimize errors.
The verification technique used on the classification was k-fold cross validation or the hold out method, i.e. to designate 9/10 of the data as the training set and the remaining 1/10 as the test set, by rotation, so all data have chance to become testing data and training data. The hold out estimated accuracy is shown in Eq. 2

$$acc_h = \frac{1}{h} \sum_{(v_i,y_i)\in \mathcal{D}_h} \delta(\mathcal{I}(\mathcal{D}_t, v_i), y_i \qquad (2)$$

Where (i, j) = 1 if i = j and 0 otherwise [21].

3. Result and Discussion

The comparison of three feature extraction methods was evaluated in this section. The purpose of this study is to determine the best wavelet-based sub-band structure selection as well as the wavelet decomposition level for extracting and classifying Indonesian stop consonants. Two classifiers namely MLP and RF were used at the classification process.

3.1. Feature Extraction

To capture the signal characteristic, feature extraction was used. The aim of this process is to obtain the signal characteristic which discriminates between certain types of stop sound and the other stop sound signals uttered by different speakers. The methods used were WBSB, WBSB1, and WBSB2 with each method have twenty-four numbers of features. After the feature extraction, the next process was the classification that uses MLP-BP and RF.

3.2. Classification

The classification of six Indonesian consonants with three following vowels using three different feature extraction techniques were compared. It is shown in table I which illustrates the percentage of classification scores of Indonesian consonants based on its POA in respect to the following vowel.

Table I shows the percentage classification scores using 10-fold cross validation (10 FCV). For the MLP classifier, in case of */a/*, the score for the velar consonant */k/* is 73.3% and 70% by using WBSB, and WBSB2 based featured, and increases to 76.7% by using WBSB2-based features. For consonant */g/*, the highest score is 80% by using WBSB1. For consonant */b/*, the highest score is 63.3% by using WBSB1. For consonant */p/*, the highest score is 76.7 by using WBSB. For consonant */t/* the highest score is 90% by using WBSB and also the highest score in the experiment. The average score for */i/* for WBSB, WBSB1, and WBSB2 are 71.11%, 63.88%, and 58.35%, respectively.

In case of */i/*, the average recognition score for WBSB, WBSB1, and WBSB2 are 62.23%, 60%, and 52.23%, respectively. In case of /u/, the average recognition score for WBSB, WBSB1, and WBSB2 are 61.67%, 53.33%, and 45.56%, respectively. From the result in Table I, it is clear that WBSB has a better performance than WBSB1 and WBSB2. It indicates that WBSB-based feature that uses 16 sub-bands provide more and better frequency resolution features for extracting Indonesian stop consonant signal compared to WBSB1, and WBSB2.

In order to observe of performance comparison, classification using an RF is also performed. For RF classifier, the average classification rate for WBSB, WBSB1, and WBSB2 were 67.77%, 67.76%, and 60.54% which is also indicated that WBSB based feature shows better performance than other feature sets for both RF and MLP classifier. The result shows that RF outperform MLP by 2.76% for WBSB features, however, the optimal number of hidden neurons in the MLP hidden layer need further investigation. In the Table I, we also present a comparative classification performance of similar study by previous researchers [3] [22]. In [3], WBSP, PCA LDA methods were used for classification of Hindi stop consonants. In [22], multilayer feed forward neural networks (MLFFNNs) was used for classifying 80 utterances of Hindi stop consonant-vowel (SCV).

Table 1: Performance comparison of each wavelet based sub-band using 10-fold cross validation (10 fcv) & similar study

Feature extraction and classification methods	Following vowels	/k/	/g/	/b/	/d/	/p/	/t/	Average % classification
WBSB+MLP	/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
WBSB1+MLP	/a/	70	80	63.3	63.3	56.7	50	63.88
	/i/	63.3	73.3	76.7	40	40	66.7	60
	/u/	56.7	23.3	80	36.7	53.3	70	53.33
WBSB2+MLP	/a/	76.7	56.7	46.7	53.3	60	56.7	58.35
	/i/	56.7	70	26.7	50	50	60	52.23
	/u/	60	46.7	40	36.7	56.7	33.3	45.56
WBSB+RF	/a/	83.3	83.3	56.7	73.3	63.3	60	69.98
	/i/	70	76.7	43.3	56.7	53.3	80	63.33
	/u/	63.3	90	60	80	66.7	60	70
WBSB1+RF	/a/	83.3	73.3	60	63.3	80	73.3	72.2
	/i/	70	70	50	56.7	40	83.3	61.67
	/u/	66.7	80	63.3	73.3	73.3	60	69.43
WBSB2+RF	/a/	63.3	73.3	46.7	63.3	56.7	80	63.88
	/i/	63.3	80	20	73.3	46.7	66.7	58.33
	/u/	76.7	70	43.3	50	83.3	33.3	59.43
WBSP+LDA [3]	/a/	50	26.9	92.3	80.8	88.5	46.2	64.12
	/i/	11.5	34.6	42.3	88.5	30.8	30.8	39.75
	/u/	65.4	73.1	69.2	84.6	50	57.7	66.67
WBSP+PCA+LDA [3]	/a/	65.4	42.3	57.7	69.2	88.5	50	62.18
	/i/	80.8	42.3	19.1	11.5	53.8	34	40.25
	/u/	73.1	80.8	88.5	61.5	53.8	50	67.95
MLFFNN [22]	/a/	42	83	75	42	75	100	69.5
	/i/	48	67	75	58	67	67	63.67
	/u/	83	25	75	75	83	67	68

4. Conclusion

In this paper, a comparison of three different feature extraction methods based on Wavelet based sub-band combined with moving average energy was performed to classify the Indonesian stop consonants in the context of CV syllable. Based on the experiments result presented in this paper, it can be concluded that the WBSB based feature has a better performance than WBSB1 and WBSB2. The classification result by using RF classifier is better than MLP, however, the optimal number of hidden neurons need further investigation. In the future work we recommend to use bigger stop consonant dataset, to separate Indonesian speech data according the speaker dialect, and to use different feature extraction technique.

Acknowledgement

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

References

[1] Boccaletti S, Giaquinta A, and Arecchi FT (1997), “Adaptive recognition and filtering of noise using wavelets,” *Physical Review Journal. Rev. E*, vol. 55, no. 5, pp. 5393–5397. DOI: 10.1103/PhysRevE.55.5393.

[2] Holschneider M (1988), “On the wavelet transforms of fractal objects,”*Journal of Statistical Physics*, vol. 50, pp. 953–993. DOI: 10.1007/BF01019149.

[3] Sharma RP, Farooq O, and Khan I (2013), “Wavelet based sub-band parameters for classification of unaspirated Hindi stop consonants in initial position of CV syllables,” *International Journal of Speech Technology*, vol. 16, no. 3, pp. 323–332. DOI: 10.1007/s10772-012-9185-x.

[4] Hardjono FL and Fox RA (2011), “Stop Consonant Characteristics: VOT and Voicing in American-Born-Indonesian Children’s Stop Consonants,” The Ohio State University.

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

[6] Amalia N, Fahrudi AE, and Nasrulloh AV (2013), “Indonesian Vowel Recognition using Artificial Neural Network based on the Wavelet Features,” *International Journal Electronic and Computer Engineering*, vol. 3, no. 2, pp. 260–269.

[7] Hidayat R, Kristomo D, and Togarma I, “Feature extraction of the Indonesian phonemes using discrete wavelet and wavelet packet transform,” in *2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE)*, 2016, pp. 478–483. DOI: 10.1109/ICITEED.2016.7863310.

[8] S. Hidayat, R. Hidayat, and T. B. Adji, “Speech recognition of CV-patterned indonesian syllable using MFCC, wavelet and HMM,” *J. Ilm. Kursor*, vol. 8, no. 2, pp. 67–78, 2015.

[9] Abriyono and A. Harjoko (2012), “Pengenalan Ucapan Suku Kata Bahasa Lisan Menggunakan Ciri LPC, MFCC, dan JST,” *Indonesian Journal Computing and Cybernetics Systems*, vol. 6, no. 2, pp. 23–34. DOI: 10.22146/ijccs.2149.

[10] Nafisah S, Wahyunggoro O, and Nugroho LE (2016), “An Optimum Database for Isolated Word in Speech Recognition System,” *Telkomnika*, vol. 14, no. 2, pp. 588–597, 2016.

[11] Fachrie M and Harjoko A, “using Elman Recurrent Neural Network Robust Indonesian Digit Speech Recognition using Elman Recurrent Neural Network,” in *Prosiding Konferensi Nasional Informatika (KNIF)*, 2015, no. March, pp. 49–54.

[12] Fugal DL (2009), *Conceptual Wavelets in Digital Signal Processing*. San Diego, California: Space & Signals Technical Publishing.

[13] Biswas A, Sahu PK, Bhowmick A, and Chandra M (2015), “Admissible wavelet packet sub-band-based harmonic energy features for Hindi phoneme recognition,” *IET Signal Processing Journal*, vol. 9, no. 6, pp. 511–519.

[14] Biswas A, Sahu PK, and Chandra M (2014), “Admissible wavelet packet features based on human inner ear frequency response for Hindi consonant recognition,” *Computer Electronic Engineering*, vol. 40, no. 4, pp. 1111–1122.

[15] Chen YH and Yu SN (2006), “Comparison of DifferentWavelet Subband Features in the Classification of ECG Beats Using Probabilistic Neural Network,” in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, 2006, pp. 1398–1401.

[16] Sunny S, Peter D, and Jacob KP (2013), “A Comparative Study of Wavelet Based Feature Extraction Techniques in Recognizing Isolated Spoken Words,” *International Journal Signal Processing System*, vol. 1, no. 1, pp. 49–53.

[17] Kulkarni P, Kulkarni S, Mulange S, Dand A, and Cheeran AN (2014), “Support Vector Machines for Isolated Word Recognition using Wavelet Packet Features,” *International Journal Engineering Technoogy. Res.*, no. 2, pp. 31–37.

[18] Suyanto and Hartati S (2013), “Design of Indonesian LVCSR using Combined Phoneme The Approaches of LVCSR,” *ICTS*, pp. 191–196.

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

[20] Hidayat R, Priyatmadi, and Ikawijaya W (2015), “Wavelet based feature extraction for the vowel sound,” in *2015 International Conference on Information Technology Systems and Innovation (ICITSI)*, pp. 1–4. DOI: 10.1109/ICITSI.2015.7437702.

[21] Kohavi R (1995), “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection,”*International Joint Conference on Artificial Intelligent*, vol. 14, no. 12, pp. 1137–1143.

[22] Chandra C and Yegnanarayana B (2002), “A constraint satisfaction model for recognition of stop consonant-vowel (SCV) utterances,” *IEEE Transactions on Speech and Audio Processing*, no. 7, pp. 472–480. DOI: 10.1109/TSA.2002.804298.