Evaluation and design of wavelet packet cepstral coefficient (WPCC) for a noisy Indonesian vowels signal

By Muhammad Tajuddin

WORD COUNT	4068	TIME SUBMITTED	06-JUN-2023 02:35PM
		PAPER ID	100275369

Evaluation and design of wavelet packet cepstral coefficient (WPCC) for a noisy Indonesian vowels signal

IOP Conf. Series: Journal of Physics: Conf. Series **1211** (2019) 012023 doi:10.1088/1742-6596/1211/1/012023

Evaluation and design of wavelet packet cepstral coefficient (WPCC) for a noisy Indonesian vowels signal

S Hidayat[1], Abdurahim[1], M Tajuddin[1]

[1]Informatic Engineering, STMIK Bumigora, Mataram, Indonesia

Email: syahroni.hidayat@stmikbumigoa.ac.id

Abstract. Wavelet feature,Wavelet Packet Cepstral Coefficient (WPCC),is widely used both in speaker and speech recognition systems. WPCC is designed to replace the role of MFCC. Unfortunately, WPCC still needs to be developed primarily in the process of extracting a noisy signal.Since the possibility of noise mapped into nodes as a result of wavelet decomposition and being a feature coefficient is very high. Therefore, this study aims to analyze the raw design of WPCC filters which extracted from a noisy signal.Mean Best Basis (MBB)algorithm with wavelet function db44 and db45 is applied to obtain it. The results show, based on the average SNR value, all data used in this study categorized as noisy signals. The implementation of MBB with db44 and db45 wavelet function able to generate the raw design of WPCC filter. Only two types of raw design created. Both have a different number of nodes, each with 59 nodes and 37 nodes. The noise does not influence this difference,but the effect of the entropy usage, the spectral properties of Indonesian vowel signals, and the wavelet function applied.

1. Introduction

In an Automatic Speech Recognition (ASR) system, feature extraction acts as one of the essentialparts. Where the performance of an ASR relies on it. There are several feature extraction methods have been developed by the researcher. Among them, a feature of Mel Frequency Coefficient[1][2][3] and Wavelet coefficient[4][5][6]most applied in the ASR system. These methods give the best performance of recognition.

MFC feature is gained from the transformation of a raw speech signal into the frequency domain using a Short Time Fourier Transform (STFT) algorithm[7]. At the first step, the raw speech signalsegmented into smaller frames. Each frametransformed with overlapped frames between 20 % - 50 % of its width. This procedure will be appliedeffectively to the raw signal which meets the requirements of length 2^n.However, the actual length of the signal does not always meet these requirements. So, the possibility of losing information from the Fourier transform process is enormous[7]. Therefore, wavelet can be the solution since its capability could transform the entire raw signal into the domain of time-frequency simultaneously, whatever the length[8].

Based on the previous description, several studies [9], [10] have designed alternative filters which, respectively, are called Wavelet Packet Cepstral Coefficient (WPCC) and Wavelet Cepstral Coefficient (WCC). These filters are trying to combine the advantages possessed by MFCC and Wavelet, as an energy-based feature extractor for automatic speech recognition systems. Pavez [9] focused on designing the WPCC filter using Wavelet Packet decomposition. Wavelet Daubechies

applied in this study for reasons of its ability to cover signal frequencies. As for Adam [10] applied wavelet dyadic to gain the WCC feature. Precisely by replacing the DFT block with Discrete Wavelet Transform (DWT) in the MFCC feature extraction process flow.

The increase of the recognition results canbe obtained from the implementation of the two types of features above.Unfortunately, it happens when the wavelet features compared to the MFCC features under the clean speech signal condition. Therefore, this study aims to analyze the WPCC which extracted from a noisy signal. The Mean Best Basis (MBB) algorithm is applied in this research to obtain the WPCC raw design. The analysis is carried out at several levels of noise, so there is a possibility of changes and differs forms of WPCC filter design compares to what hasobtained in previous studies conducted by Pavez[8], and Adam [9].

2. Literature Review

The design of the wavelet filter as the MFCC-like feature extractor for an Automatic Speech Recognition system developed lately. Pavez[9]developed wavelet filter using wavelet packet called wavelet packet cepstral coefficient (WPCC). It applied to several fidelity measurements for validation. Design of filter obtained from this method is two WPCC filter with 24 bands and 26 bands of the filter. It is developed using wavelet Daubechies db44. This design closed to wavelet filters developed by Farooq and Datta[11] and Choueiter and Glass[12]. The result shows that the recognition rate using these two filters still couldn't surpass the MFCC recognition rates.

Adam [10]designed WCC filters based on wavelet dyadic method, Discrete Wavelet Transform (DWT). To gain WCC Adam [5] replace the Discrete Cosine Transform (DCT) block process in MFCC by DWT block. DWT than applied to the raw speech signal and decomposed until level 8, level 5, and level 3 of decomposition. The filter gained from the decomposition process then arranged like MFCC filter. A number of a coefficient obtained from this method is 90, 60, and 40 coefficients respectively. Adam [10]applied this filter especially for speaker recognition using Neural Network. The result shows that the accuracy of recognition using WCC is comparable to MFCC result.

There are several methods developed to make the best tree of wavelet packet decomposition. Like Coifman [13] did to obtain the best tree by calculating the entropy value of each leaf in the wavelet packet tree. Galka [14]accomplished theworks did by Coifman by developing a method called Mean Best Basis (MBB) Algorithm. In MBB method applied the DCT and Shannon entropy to gain the Wavelet Packet Cosine Transform (WPCT) coefficient.

3. Theoretical Background

3.1. Noise to Signal Ratio (SNR)

Noise to Signal Ratio (SNR) is the comparison between the log power of a signal without noise and a noisy signal[15], [16]. SNR formula denoted as Equation (1). SNR value is always calculated to identify the signal is categorized as information or not.

$$SNR_{dB} = 10\ x\ log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) \quad (1)$$

3.2. Wavelet Transform

Wavelet is a limited duration of the wave which has zero average value. It is not like a sinusoidal wave which theoretically has a length from -∞ to +∞. Wavelet has the beginning and the ending. It called as a shortwave. Wavelet concentrates its energy into frequency and time simultaneously, so it is suitable to analyze signals that are temporary, like an earthquake signal. That process is the implication of the translation and scaling of limited energy called mother wavelet($\psi(t)$) and scaling function ($\varphi(t)$). Mother wavelet acts as the High Pass Filter (HPF) while scaling function acts as the Low Pass Filter (LPF).

Let s[n] bea discrete signal with N number of period, the discrete wavelet transformation (DWT) of s[n]expressed as [17]:

$$DWT[n, 2^j] = \sum_{m=0}^{N-1} s[n]\psi_{2^j}^*[m-n] \quad (2)$$

Where j is the level of decomposition, and

$$\psi_{2^j} = \frac{1}{\sqrt{2}}\psi\left(\frac{n}{2^j}\right) \quad (3)$$

DWT decomposed input signal to be a set of approximation coefficient and details coefficient. There are two methods ofwavelet decomposition, dyadic wavelet and wavelet packet. Dyadic decomposition is recursively decomposed signal only on its approximation side. While WPT is one of DWT variation which decomposes the two sides of sub-band frequency, decomposition of approximation signal uses LPF and for detail signals using HPF. In Figure 1 every level-j for both side decomposition (approximation and details) gives 2^j equally spaced sub-band. This equally spaced sub-band is called a decomposition tree. The energy in every sub-band can be calculated to form a feature vector.

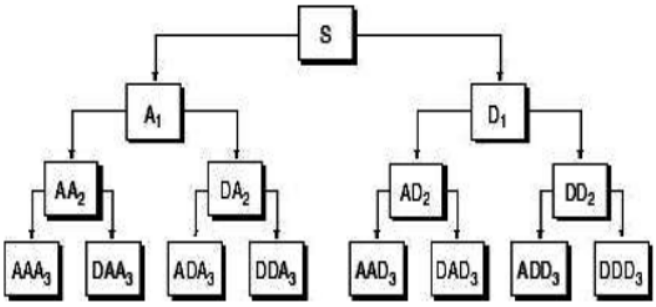

Figure 1. Decomposition of full WPT Level-3

Total energy for all subband is calculated using equation (4).E_i is energy at frequency sub-band, andX_i (k) is the value of frequency sub-band at series-k.

$$E_i = \sqrt{\sum_{k=1}^{N} |X_i(k)|^2} \quad (4)$$

3.3. Wavelet Daubechies

One of the wavelet functions which usedwidely in the speech signal processing, especially phoneme and vowels of Indonesian language is wavelet Daubechies family[5], [18]. Since the filter coefficient, which forms the wavelet Daubechies function, is very similar to the Indonesian vocal sound signal [18]. It is proofed by [18]and obtained that the best wavelet Daubechies for Indonesian vowels speech signal is wavelet db44 and db45. Wavelet Daubechies categorized by its filters coefficient, dbN. N referrs to the order of the filters. The higher the coefficient is, the closer the shape of the filter coefficients to the original signal.

3.4. Mean Best Basis Algorithm

Mean Best Basis (MBB) algorithm is the alternative to determine the basis of wavelet decomposition. This method is the improvement of Best Basis algorithm which developed by Coifman[14]. The determination of the best basis from wavelet packet decomposition is one of the ways in selecting the best feature of the speech signal,i.e., by evaluating the entropy value between signal and noise. Steps of MBB algorithm described as below:

- For each element of signal s, calculate WPCT. WPCT is the process of wavelet decomposition.

$$\boldsymbol{W^{WPCT}} = \{\boldsymbol{W_{m,j}^{WPCT}}\}: \boldsymbol{W_{m,j}^{WPCT}} \leftrightarrow \boldsymbol{\hat{d}_{m,j}} \tag{5}$$

$\hat{d}_{m,j}(k)$ is WPCT for each sub-band. It is defined as:

$$\hat{d}_{m,j}(k) = \sum_{n=1}^{N_m} d_{m,j}(n).cos\left(2\pi\frac{nk}{N_m}\right) \tag{6}$$

- Calculate the entropy $\boldsymbol{\varkappa_{m,j}^{i}}$of each node in the decomposition tree.It is defined as:

$$\boldsymbol{\varkappa_{m,j}^{i}} = \boldsymbol{\varkappa}\left(\boldsymbol{\hat{d}_{m,j}^{i}}\right) \tag{7}$$

$\varkappa\left(\hat{d}_{m,j}^{i}\right)$ is the Shannon entropy of the WPCT coefficient. It defined as:

$$\varkappa\left(\hat{d}_{m,j}^{i}\right) = -\sum_{n=1}^{N_d}\left(\frac{\hat{d}_{m,j}^{2}(n)}{\left\|\hat{d}_{m,j}\right\|^2} log\left(\frac{\hat{d}_{m,j}^{2}(n)}{\left\|\hat{d}_{m,j}\right\|^2}\right)\right) \tag{8}$$

- For each entropy value obtained from the node of the decomposition tree, normalized the entropy value based on its root entropy $\forall_i \forall_{m,j}$ using (9). This makes entropy independent of different signal energy values.

$$\forall_{\boldsymbol{i}} \forall_{\boldsymbol{m,j}} = \frac{\boldsymbol{\varkappa_{m,j}^{i}}}{\boldsymbol{\varkappa_{0.1}^{i}}} \tag{9}$$

- Determine the general tree of mean entropy values$\bar{W}^{\varkappa}$ over all signals with all entropy values normalized.

$$\bar{W}^{\varkappa} = \left\{\bar{W}_{m,j}^{\varkappa} \leftrightarrow \bar{\varkappa}_{m,j}\right\}: \bar{\varkappa}_{m,j} = \frac{1}{|\{s\}|}\sum_{\varkappa_{m,j}^{i} \leftrightarrow W_i^{\varkappa}} \varkappa_{m,j}^{i} \tag{10}$$

- Find the best subtree using the Wickerhouser's Best Basis algorithm with a mean-entropy tree

$$\boldsymbol{\bar{W}^{opt}} = \boldsymbol{\underset{\bar{W}^{*}}{\operatorname{argmin}}} \sum_{\boldsymbol{\bar{\varkappa}_{m,j}^{\sigma} \leftrightarrow W_i^{*}}} \boldsymbol{\bar{\varkappa}_{m,j}^{\sigma}} \tag{11}$$

4. Methods and Materials

This research carried out following the steps shown in Figure 2. The process starts with audio recording. The recorded audio is the Indonesian vowel utterance, *a*, *i*,*u*, *e*, *é*, *o*, and *ó*. The audio obtained from 50 speakers, each of 25 male and female speakers. The recording process is done once for each vocal sound. Its duration is 1-second. The sampling frequency is 16000 Hz. The standard of pronunciation follows the International Phonetic Association (IPA) standard. The recording processis done in an open area. The aim is to ensure that the sound recorded contains noise. The recording tool is a headphone while the recording software is Audacity. The recorded sound saved as '*001a.wav*'. Initial '*001*' refers to sample numbers while '*a*' refers to recorded vowel sounds. The storage file format is * .wav. The recording properties are shown briefly in Table 1.

The SNR value of all recorded sounds is calculated. After that, the average SNR value calculated for each vowel. Furthermore, the sound transformed with db44 and db45 wavelet functions. The transformation type used is the wavelet packet transform (WPT). Its maximum level of decomposition

is six levels. The MBB algorithm is then applied to all channels of decomposed signals to obtain the best tree. This best tree considered as the raw design of WPCC filter.

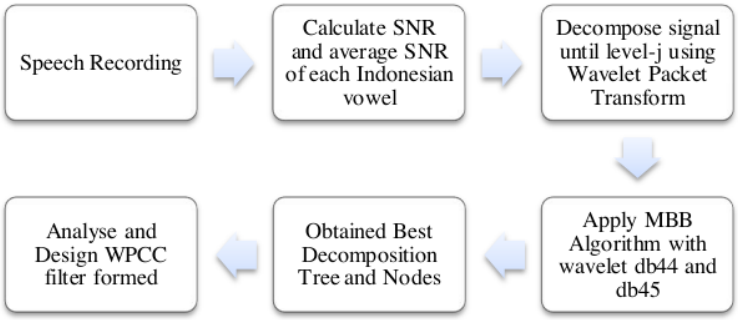

Figure 2. Research steps

Table 1. Properties of recording

No	Properties	Variable
1	Data Recorded	Vowels
2	Number of Samples	50 persons
3	Recording Repetition	1
4	Recording Tempo	IPA standard
5	Sampling frequency (*fs*)	16 kHz
6	Duration	1 second
7	Recording Environment	Open area
8	Data format	**.wav*
9	Recording tools	*Microphone*
10	*Software*	Audacity

5. Result and Discussion

5.1. Speech Recording

Recorded Indonesian vowel speech signal has various lengths. There is head in the beginning and tail at the end of the signal. This head and tail considered as the noise. Details of the speech signal sample recorded is shown in Figure 3. Each signal has a period of utterance under 1 second. It is about 0.4 – 0.5 *millisecond*on average. The magnitude of the signal varied between ± 0.2 - ± 1. The signal with magnitude is under ± 0.03 usually categorized as the noise.

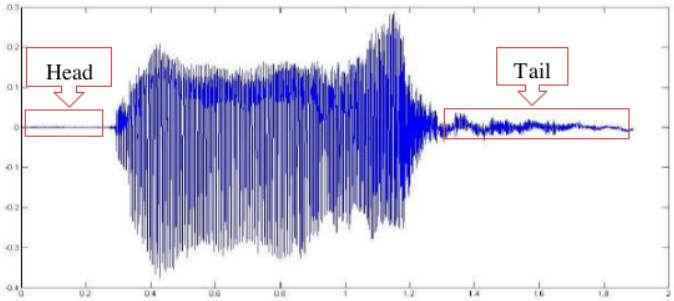

Figure 3. Sample signal of vowel *i* with head in the beginning and tail at the end of the signal.

5.2. SNR Value

The SNR of the signal is calculated using the equation (1) above. The result of this process shown in **Table 2**.The average SNR for each Indonesian vowel *a, i, u, e, é, o,* and *ó* respectively is -6.3 dB, 5.3 dB, 2.6 dB, 0.14 dB, 0.6 dB, -0.35 dB, and -6.2 dB. The speech signal is classified as a noisy signal if its SNR is below 30 dB[19]. It can be concluded from it $\overline{SNR}$ that all the vowels signal used in this research are a noisy signal.

Table 2. The average SNR value of Indonesian Vowel

Vowel	$\overline{SNR}$ (db)
a	-6.3
i	5.3
u	2.6
e	0.14
é	0.6
o	-0.35
ó	-6.2

5.3. The Best Mean Best Basis Tree

The raw design of the WPCC filter has been obtained from the application of the MBB algorithm [14]. In its application, the MBB algorithm forms the best basis signal from the wavelet packet decomposition tree. The entire basis tree obtained from this algorithm used as the feature of the sound signal. In this study, MBB was applied only to determine the best tree for Indonesian vowel sound signals. Daubechies db44 and db45 wavelet functions implemented in the decomposition process. The application of these two functions based on their resemblance to all Indonesian vocals [18]. The results show that only two types of MBB best tree formed. Each represents the wavelet Daubechies function used. These two best MBB trees designated as WPCC filter raw designs.

The result,in figure 4 and figure 5, shows that there is the difference between the best tree obtained from db44 and db45 implementation.Figure 4 showsthe difference where the number of the node obtained from the implementation of db44 wavelet function is 59 nodes in total. While from using db45 wavelet function obtained 37 nodes only. The difference is starting from the 5^{th}level until the 6^{th}level of decomposition. Where in figure 5 shows that, for db45 function,it is almost all of the leaf onits last two levels of decomposition not considered as the best tree.It is the effect of the implementation of the entropy and the spectral properties[14]of Indonesian vowels signal. And it can be assumed that not only entropy but this result also influenced by the wavelet function used.

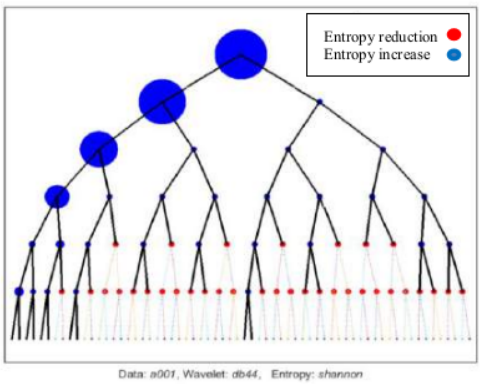

Figure 4. The WPCC filter raw designfor wavelet db44 implementation

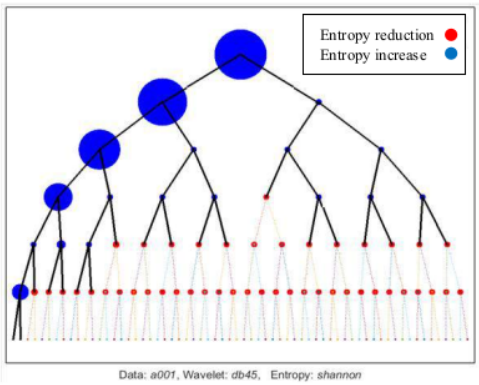

Figure 5. The WPCC filter raw design for wavelet db45 implementation

Table 3. The Best Node and *Leaf* obtained from the implementation of wavelet function db44

No	Node	Leaf	Node	Leaf	Node	Leaf
1	0	[0,0]	20	[4,5]	44	[5,13]
2	1	[1,0]	21	[4,6]	47	[5,16]
3	2	[1,1]	22	[4,7]	48	[5,17]
4	3	[2,0]	23	[4,8]	51	[5,20]
5	4	[2,1]	24	[4,9]	52	[5,21]
6	5	[2,2]	25	[4,10]	59	[5,28]
7	6	[2,3]	26	[4,11]	60	[5,29]
8	7	[3,0]	27	[4,12]	61	[5,30]
9	8	[3,1]	28	[4,13]	62	[5,31]
10	9	[3,2]	29	[4,14]	63	[6,0]
11	10	[3,3]	30	[4,15]	64	[6,1]
12	11	[3.4]	31	[5,0]	65	[6,2]
13	12	[3,5]	32	[5,1]	66	[6,3]
14	13	[3,6]	33	[5,2]	67	[6,4]
15	14	[3,7]	34	[5,3]	68	[6,5]
16	15	[4,0]	35	[5,4]	71	[6,8]
17	16	[4,1]	36	[5,5]	72	[6,9]
18	17	[4,2]	39	[5,8]	95	[6,32]
19	18	[4,3]	40	[5,9]	96	[6,33]
20	19	[4,4]	43	[5,12]		

Table 4. The Best Node and *Leaf* obtained from the implementation of wavelet function db45

No	Node	Leaf	Node	Leaf
1	0	[0,0]	19	[4,4]
2	1	[1,0]	20	[4,5]
3	2	[1,1]	21	[4,6]
4	3	[2,0]	22	[4,7]
5	4	[2,1]	25	[4,10]
6	5	[2,2]	26	[4,11]
7	6	[2,3]	27	[4,12]
8	7	[3,0]	28	[4,13]
9	8	[3,1]	29	[4,14]
10	9	[3,2]	30	[4,15]
11	10	[3,3]	31	[5,0]
12	11	[3.4]	32	[5,1]
13	12	[3,5]	33	[5,2]
14	13	[3,6]	34	[5,3]
15	14	[3,7]	35	[5,4]
16	15	[4,0]	36	[5,5]
17	16	[4,1]	63	[6,0]
18	17	[4,2]	64	[6,1]
19	18	[4,3]		

The entropy applied in this study is wavelet entropy. Wavelet entropy cananalyze transient features of non-stationary signals. It is the combination of wavelet decomposition and entropy calculation. The calculation implemented on the high-frequency sub-bands of wavelet decomposition to quantify its energy distribution. This energy distribution is the representation of minimum information contained on it. Also, it acts as an indicator of the possibility of reconstructing a decomposed signal into the original signal.Wavelet entropy has influenced the tree decomposition, especially on the last two levels of decomposition used in this study (5^{th} level and 6^{th} level).

The effect of wavelet function usage also considered. The wavelet function db44 and db45 applied in this research are based on the similarity obtained using cross-correlation parameter [18]. Even tough

wavelet function db44 and db45 coefficientsare most similar to the Indonesian vowels signal under different levels of decomposition, the energy function of the noise contained in it also decreases exponentially. It causes the function of the reference energy formed becomes stable at each level of decomposition [20].Consequently, only two forms of the WPCC raw design obtained in this study.

From the previous explanation, the influence of noise on the raw design ofWPCC can be ignored since it does not affect the best tree generated. It is certainly due to the nature of the wavelet transform process. Where in this process the mother wavelet function which implemented to the signal acts as a high pass filter and low pass filter. And by implementing the mean best basis algorithm and the entropy calculation, the node which containedthe noise signalis automatically excluded as the best tree criterion. The information of the node and leaf which considered as the WPCC filter raw designshown in Table 3 and Table 4.

6. Conclusion

The raw design of wavelet packet cepstral coefficient (WPCC) can be obtained by implementing the mean best basis algorithm. After implementing it using wavelet db44 and db45 function concluded that only two of the best tree generated as the raw design of WPCC. There aredifferences between these twodesigns.This differencewas the effect of the entropy usage, the spectral properties of the Indonesian vowels signal, and the wavelet function applied. As for the noise, it does not affect the raw design of WPCCobtained.

In the future work, this research will determine the best frequency and range frequency for the wavelet packet cepstral coefficient. These two variables will be considered in designing the most similar wavelet filter to the MFCC filter based on its properties.After that, this design will be implemented as the feature extractor in speech recognition system to obtain its best performance.

Acknowledgment

We would say thanks to DRPM of Ministry of Research, Technology and Higher Education (RISTEKDIKTI) of the Republic of Indonesia for funding this research with Research Grant number SP DIPA-042.06.1.401516/2018.

References

[1] Gaikwad S K, Bharti W G and Pravin Y 2010A Review on Speech Recognition Technique*International Journal of Computer Applications***10** 16–24

[2] Desai N, Dhameliya K and Desai V 2013Feature Extraction and Classification Techniques for Speech Recognition : A Review*International Journal of Emerging Technology and Advanced Engineering***3** 367–71.

[3] Anusuya M A and Katti S K 2009Speech Recognition by Machine: A Review*International Journal of Computer Science and Information Security***6** 181–205

[4] Farooq O and Datta S 2003Phoneme Recognition Using Wavelet Based Features*Information Sciences***150** 5–15.

[5] Hidayat S, Hidayat R and Adji T B 2015Speech Recognition of KV-Patterned Indonesian Syllable using MFCC, Wavelet,and HMM.*Jurnal Ilmiah Kursor***8** 67–78.

[6] Turner C and Anthony J 2015A Wavelet Packet and Mel-Frequency Cepstral Coefficients-Based Feature Extraction Method for Speaker Identification*Procedia Computer Science***61** 416–21.

[7] Anusuya M A and Katti S K 2011Front End Analysis of Speech Recognition: A Review*International Journal of Speech Technology***14** 99–145.

[8] Shah A F and B Anto P 2017Wavelet Packets for Speech Emotion Recognition*3rd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB17)* 3-5.

[9] Pavez E and Silva J F 2012Analysis and Design of Wavelet-Packet Cepstral Coefficients for Automatic Speech Recognition*Speech Commun.***54** 814–35.

[10] Adam T B, Salam M S and Gunawan T S 2013Wavelet Cesptral Coefficients for Isolated Speech Recognition*Telkomnika***11** 2731–38.
[11] Farooq O and Datta S 2001Mel Filter-Like Admissible Wavelet Packet Structure for Speech Recognition*IEEE Signal Process. Lett.***8** 196–98.
[12] Choueiter G F and Glass J R 2007An Implementation of Rational Wavelets and Filter Design for Phonetic Classification*IEEE Trans. Audio, Speech Lang. Process.***15** 939–48.
[13] Coifman R R and Wickerhauser M V 1992Entropy-Based Algorithms for Best Basis Selection*IEEE Trans. Inf. Theory***38** 713–18.
[14] Galka J and Ziolko M 2009Mean Best Basis Algorithm for Wavelet Speech Parameterization*Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing* 1110–13.
[15] Vondrasek M and Pollak PMethods for Speech SNR Estimation: Evaluation Tool and Analysis of VAD Dependency 2005 *RADIOENGINEERING***14** 6–11.
[16] Huang X, Acero A and Hon H W 2001 *Spoken Language Processing: A Guide to Theory, Algorithm and System Development* (New York: Prentice Hall).
[17] Rioul O and Vetterli M 1991Wavelets and Signal Processing*IEEE SP Magazine***8** 14 - 38.
[18] Hidayat S, Negara H R P and Kumoro D T 2017Determination of the Optimum Wavelet Basis Function for Indonesian Vowel Voice Recognition*J. Elektron. dan Telekomun.***17** 42–47
[19] Ellis D P W 2000 ICSI Speech FAQ: 4.1 How is the SNR of a speech example defined? Online: http://www1.icsi.berkeley.edu/Speech/faq/speechSNR.html.
[20] Sang Y F 2012A Practical Guide to Discrete Wavelet Decomposition of Hydrologic Time Series*Water Resour. Manag.***26** 3345–65.

Evaluation and design of wavelet packet cepstral coefficient (WPCC) for a noisy Indonesian vowels signal

ORIGINALITY REPORT

12%
SIMILARITY INDEX

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

★karyailmiah.uho.ac.id 2%
Internet

EXCLUDE QUOTES	OFF	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	OFF	EXCLUDE MATCHES	OFF