

PENENTUAN *FILTERBANK* WAVELET MENGGUNAKAN ALGORITMA *MEAN BEST BASIS* UNTUK EKSTRAKSI CIRI SINYAL SUARA BER-*NOISE*

Abdurahim^{*1}, Syahroni Hidayat²

^{1,2}STMIK Bumigora Mataram

Email: ¹aim.math13@gmail.com, ²syahroni.hidayat@stmikbumigora.ac.id

*Penulis Korespondensi

(Naskah masuk: 22 November 2018, diterima untuk diterbitkan: 13 Januari 2020)

Abstrak

Belakangan ini *filterbank* berbasis wavelet sebagai ekstraktor ciri mulai banyak dikembangkan untuk dapat menggantikan peran ciri Mel Frequency Cepstral Coefficient (MFCC) dalam sistem pengenalan suara otomatis. Salah satu *filterbank* ciri wavelet yang dikembangkan adalah Wavelet-Packet Cepstral Coefficient (WPCC). Namun sejauh ini pengembangannya hanya difokuskan untuk suara tanpa *noise*. Sehingga penelitian ini bertujuan untuk mendesain WPCC untuk suara yang mengandung *noise*. Algoritma Mean Best Basis (MBB) dan fungsi wavelet db44 dan db45 digunakan untuk memperoleh desain *filterbank* WPCC. Suara yang digunakan adalah rekaman suara vokal bahasa Indonesia *a*, *i*, *u*, *e*, *é*, *o*, dan *ó* yang mengandung *noise*. Hasil menunjukkan telah terbentuk dua buah desain *filterbank* WPCC. Masing-masing merupakan hasil penerapan fungsi *daubechies* db44 dan db45. *Noise* tidak memberikan pengaruh terhadap pembentukan kedua *filterbank* WPCC tersebut. Kedua bentuk *filterbank* telah memenuhi standar bentuk filter MFCC terutama untuk variabel range dan skala frekuensinya. Range frekuensinya berkisar antara 125 Hz - 1000 Hz dengan bentuk skala yang linier untuk frekuensi di bawah 1000 Hz. Sehingga dapat disimpulkan kedua bentuk *filterbank* WPCC ini dapat dipertimbangkan untuk digunakan sebagai ekstraktor ciri suara ber-*noise*.

Kata kunci: sistem pengenalan suara otomatis, WPCC, MFCC, MBB.

DETERMINATION OF WAVELET FILTERBANKS USING THE MEAN BEST BASIS ALGORITHM FOR FEATURE EXTRACTION OF NOISY SPEECH SIGNAL

Abstract

Recently wavelet-based filterbanks as feature start extractors have been widely developed to replace the role of the Mel Frequency Cepstral Coefficient (MFCC) feature in automatic speech recognition systems. One of the wavelet feature filterbanks developed is Wavelet-Packet Cepstral Coefficient (WPCC). But so far the development has only been focused on clean speech signal. So, the aim of this study is designing WPCC for a noisy speech signal. The Mean Best Basis (MBB) algorithm and db44 and db45 wavelet functions are applied to obtain the WPCC filterbank design. The noisy speech signal used is the recorded utterance Indonesian vowels *a*, *i*, *u*, *e*, *é*, *o*, and *ó*. The results show that two WPCC filterbank designs have been formed. Each of them is the result of applying the *daubechies* db44 and db45 functions. Noise has no effect on the establishment of both the WPCC filterbanks. Both filterbank designs have met MFCC filter form standards, especially for its range of frequency and frequency scale. Its range of frequency is between 125 Hz - 1000 Hz with a linear scale for frequencies below 1000 Hz. Therefore it can be concluded that the two forms of WPCC filterbank can be considered to be used as a feature extractor for a noisy speech signal.

Keywords: automatic speech recognition system, WPCC, MFCC, MBB.

1. PENDAHULUAN

Sistem pengenalan suara otomatis adalah sebuah sistem yang dibangun agar komputer dapat mengenali suara makhluk hidup. Sistem ini terbagi menjadi dua yaitu untuk mengenali pembicara dan mengenali suara yang diucapkan (Anusuya dan Katti,

2009; Rashmi, 2014). Secara umum cara kerja sistem ini terbagi menjadi dua fase, fase pelatihan dan fase pengenalan/pengujian. Di dalam dua fase ini terdapat beberapa tahapan yang sama yaitu tahap perekaman, tahap pengolahan awal, dan tahap ekstraksi ciri. Adapun pembeda di antara keduanya adalah tahap

pelatihan untuk fase pelatihan dan tahap pengenalan/pengujian untuk fase pengujian. Kedua fase ini sangat bergantung kepada tahapan proses sebelumnya, terutama tahap ekstraksi ciri.

Ekstraksi ciri adalah proses pengubahan sinyal suara asli menjadi vektor akustik yang mencirikan suara tersebut (Turner dan Joseph, 2015). Ciri yang baik harus memiliki kekhasan tersendiri dan ukuran vektornya lebih kecil dari sinyal suara asli. Banyak metode ekstraksi ciri yang telah dikembangkan dan digunakan oleh para peneliti. Di antara metode-metode tersebut ciri koefisien frekuensi mel (*Mel Frequency Coefficient*) (Gaikwad, Gawali dan Yannawar, 2010; Desai, Dhameliya dan Desai, 2013) dan ciri wavelet (Farooq dan Datta, 2003; Pavez and Silva, 2012; Adam, Salam dan Gunawan, 2013; Abo-Zahhad, Ahmed dan Abbas, 2016) yang telah banyak diaplikasikan baik pada pengenalan suara maupun pembicara terbukti memberikan hasil pengenalan terbaik. Selain itu, MFCC akhir-akhir ini juga telah banyak dikembangkan dan ditingkatkan dari bentuk standarnya.

Ciri MFC diperoleh melalui transformasi isyarat tutur ke dalam domain frekuensi menggunakan transformasi fourier, khususnya *Short Time Fourier Transform* (STFT). Transformasi dilakukan dengan terlebih dahulu membagi isyarat tutur menjadi isyarat dengan ukuran yang lebih kecil (*frame*). Selanjutnya setiap *frame* ditransformasi dengan diberikan perlakuan overlap sebesar 20 – 30 ms. Perlakuan ini akan sangat efektif diterapkan jika panjang total isyarat tutur memenuhi syarat 2^n . Akan tetapi, pada kenyataannya panjang isyarat sebenarnya tidaklah selalu memenuhi persyaratan tersebut sehingga kemungkinan terjadinya kehilangan informasi dari penerapan proses transformasi fourier sangatlah besar (Anusuya dan Katti, 2011). Oleh karena itu, wavelet yang pada proses transformasinya dapat bekerja pada domain frekuensi-waktu secara bersamaan dan langsung merubah seluruh isyarat, berapapun panjangnya, dapat dijadikan sebagai solusi. Wavelet juga memiliki sifat-sifat yang serupa dengan transformasi fourier. Sehingga penerapannya untuk mengekstraksi ciri seperti MFCC juga pastinya akan memberikan akurasi hasil yang tidak jauh berbeda.

Berangkat dari uraian di atas, Pavez dan Silva (2012) serta Adam, et. al. (2013) telah mendesain *filter* alternatif yang serupa dengan *Mel Frequency Cepstral Coefficient* (MFCC), secara berurutan, disebut sebagai *Wavelet Packet Cepstral Coefficient* (WPCC) dan *Wavelet Cepstral Coefficient* (WCC). Keduanya berusaha menggabungkan kelebihan yang dimiliki oleh MFCC dan wavelet, sebagai pengekstrak fitur berbasis energy untuk sistem pengenalan tutur otomatis. Pavez dan Silva (2012) memfokuskan desain *filter* WPCC menggunakan dekomposisi wavelet paket. Keluarga wavelet yang digunakan adalah wavelet daubechies dengan alasan kemampuannya dalam meng-cover frekuensi sinyal. Adapun Adam, et. al. (2013) memanfaatkan wavelet

dyadic untuk memperoleh ciri WCC. Tepatnya menggantikan blok DFT dengan DWT pada alur ekstraksi ciri MFCC.

Penentuan desain wavelet yang telah dilakukan Pavez dan Silva (2012) merujuk pada penelitian yang telah dilakukan oleh Farooq dan Datta (2001) serta Choueiter dan Glass (2007). Nilai yang digunakan dalam nilai *fidelity* dari sinyal terdekompresi. Hasilnya adalah sebuah desain *filter Wavelet Packet Cepstral Coefficient* yang serupa dengan hasil (Farooq dan Datta, 2001; Choueiter dan Glass, 2007) namun dengan jumlah *filter* yang lebih sedikit.

Filter WPCC yang menjadi hasil dari beberapa penelitian di atas dibentuk dari sinyal tanpa *noise*. Sayangnya penelitian tentang desain filter WPCC untuk sinyal ber-*noise* belum dilakukan. Oleh karena itu, tujuan dari penelitian ini adalah menganalisis dan mendesain filter WPCC dari sinyal suara ber-*noise*. Analisis dilakukan pada beberapa tingkatan *noise* sehingga ada kemungkinan terjadinya perubahan dan perbedaan bentuk desain *filter* WPCC yang terbentuk. Penelitian ini menerapkan algoritma *Mean Best Basis* (MBB) dan fungsi wavelet daubechies untuk mendapatkan desain *filter* WPCC.

2. TINJAUAN PUSTAKA

2.1 Sinyal Suara

Suara adalah sebuah gelombang longitudinal yang bergerak melalui udara karena vibrasi/getaran molekul. Di udara suara ditransmisikan sebagai perubahan tekanan, antara tekanan tinggi dan rendah, dengan laju perubahan tekanan dari rendah ke tinggi kemudian ke rendah lagi, menunjukkan frekuensi (Huang, Acero dan Hon, 2001).

Sinyal suara merupakan sinyal yang bervariasi lambat sebagai fungsi waktu, dalam hal ini ketika diamati pada durasi yang sangat pendek (5 sampai 100 mili detik) karakteristiknya masih stasioner. Tetapi bilamana diamati dalam durasi yang lebih panjang ($> 1/5$ detik) karakteristik sinyalnya berubah untuk merefleksikan suara ucapan yang keluar dari pembicara.

Untuk akuisisi suara frekuensi sampling (fs) yang digunakan dapat bervariasi mulai dari 8 kHz, untuk suara kualitas telepon, sampai 48 kHz, untuk suara kualitas tinggi. Rata-rata frekuensi sampling suara manusia untuk aplikasi pengenalan suara otomatis yang baik adalah 16 kHz atau di atasnya. Berdasarkan kriteria Nyquist, frekuensi tertinggi yang dapat direpresentasikan adalah setengah dari fs (McLoughlin, 2009).

2.2 Signal to Noise Ratio (SNR)

SNR adalah log perbandingan (rasio) antara daya sinyal (*signal strength*) tanpa *noise* terhadap daya sinyal ber-*noise* (*noise level*) (Huang, Acero dan Hon, 2001), dinyatakan sebagai :

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

Nilai SNR dipakai untuk menunjukkan kualitas sinyal tersebut. Oleh karena itu, semakin besar nilai SNR-nya, makin tinggi kualitas sinyal tersebut.

2.3 Transformasi Wavelet

Wavelet adalah salah satu alat matematika yang digunakan untuk mendekomposisi (memotong-motong) sebuah fungsi atau sinyal. Dalam hal ini, wavelet membagi fungsi tersebut kedalam komponen atau kumpulan frekuensi yang berbeda. Lebih jauh, wavelet adalah gelombang dengan durasi terbatas yang memiliki nilai rata-rata nol, tidak seperti sinusoid yang secara teoritis memiliki panjang dari minus ke plus tak hingga, sementara wavelet memiliki awal dan akhir. Atau dapat dikatakan sebagai gelombang pendek atau singkat.

Wavelet mengkonsentrasikan energinya dalam frekuensi dan waktu sehingga cocok untuk menganalisis isyarat yang sifatnya sementara saja. Proses tersebut merupakan hasil dari proses pergeseran dan penskalaan suatu energi terbatas yang disebut *mother wavelet* ($\psi(t)$) yang bertindak sebagai *high pass filter* (HPF) sementara fungsi skala ($\varphi(t)$) sebagai *low pass filter* (LPF) (Rioul dan Vetterli, 1991).

Diberikan sinyal diskrit $s(n)$ dengan periode N , maka *Discrete Wavelet Transform* (DWT) dinyatakan sebagai (Adam, Salam dan Gunawan, 2013):

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

Di mana j merupakan level dekomposisi, dan

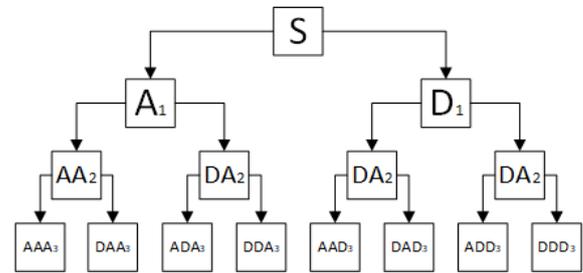
$$\psi_{2^j} = \frac{1}{\sqrt{2}} \psi\left(\frac{n}{2^j}\right) \tag{3}$$

DWT dapat diaplikasikan pada data diskrit untuk menghasilkan *output* diskrit dengan cara mentransformasikan isyarat dari domain waktu ke domain frekuensi. Selanjutnya, DWT mendekomposisi sinyal masukan menjadi sekelompok koefisien aproksimasi dan detail. Ada dua metode dekomposisi pada wavelet, yaitu *dyadic wavelet packet*. Dekomposisi *dyadic* secara rekursif hanya melakukan dekomposisi pada koefisien aproksimasi. Adapun *wavelet packet* (WP) melakukan dekomposisi baik terhadap koefisien aproksimasi maupun koefisien detail sinyal. Sehingga hasil dari dekomposisi berbentuk seperti struktur pohon biner, kita sebut pohon dekomposisi.

2.4 Wavelet Packet Transform (WPT)

WPT adalah variasi dari DWT yang melakukan dekomposisi pada dua sisi *sub-band* yaitu dekomposisi isyarat aproksimasi A pada sisi frekuensi rendah menggunakan *Low Pass Filter* (LPF) dan dekomposisi isyarat detail D pada sisi frekuensi tinggi menggunakan *High Pass Filter* (HPF). Pada Gambar 1 tiap level ke- j hasil dekomposisi WPT menghasilkan 2^j *sub-band* dengan

lebar yang sama. Energi hasil rekonstruksi pada masing-masing *sub-band* dapat dihitung untuk membentuk karakteristik vektor ciri (Hidayat, Hidayat dan Adji, 2015).



Gambar 1. Dekomposisi Full WPT Level-3

Total energi pada seluruh frekuensi *sub-band* dihitung menggunakan persamaan (4).

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

dengan E_i adalah energi pada frekuensi *sub-band* dan $X_i(k)$ menyatakan nilai pada runtun ke- k pada frekuensi *sub-band*.

2.5 Wavelet Daubechies

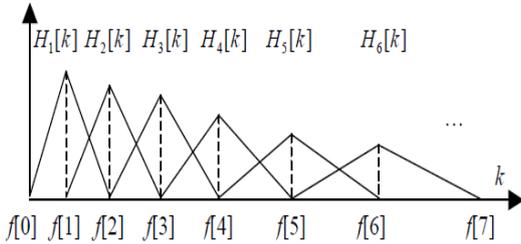
Wavelet daubechies merupakan salah satu jenis wavelet yang aplikasinya banyak diterapkan pada pengolahan sinyal suara, terutama suara fonem dan vokal bahasa Indonesia (Hidayat, Hidayat dan Adji, 2015; Hidayat, Negara dan Kumoro, 2017). Hal tersebut dikarenakan koefisien *filter* yang membentuk fungsi wavelet daubechies sangat mirip dengan sinyal suara vokal bahasa Indonesia. Hidayat, et. al (2017) telah membuktikannya dan diperoleh hasil bahwa wavelet daubechies terbaik untuk suara vokal bahasa Indonesia adalah wavelet db44 dan db45.

Filter wavelet Daubechies memiliki koefisien respon yang dinyatakan sebagai DbN , notasi N adalah orde *filter* wavelet yang digunakan. Besaran nilai orde N memiliki pengaruh terhadap hasil dekomposisi wavelet. Di mana kelebihan menggunakan wavelet DbN dengan nilai N lebih besar adalah adanya peningkatan pada hasil analisis multi resolusinya, yaitu isyarat yang lebih halus. Karena peningkatan nilai N menyebabkan fluktuasi nilai magnitud isyarat menjadi lebih kecil. Sifat ini sangat baik diterapkan untuk kompresi dan penghapusan derau terutama pada isyarat rata-rata dengan resolusi yang rendah

2.6 Mel Frequency Cepstral Coefficient (MFCC)

MFCC merupakan representasi dari cara kerja sistem pendengaran manusia yang bersifat linier. Sebagai acuan, penskalaan antara frekuensi dalam Hz dan skala mel bersifat linier pada frekuensi di bawah 1000 Hz dan bersifat logaritmik pada frekuensi diatasnya (Anusuya dan Katti, 2011, 2009). Untuk menerapkan skala mel ke cepstrum, sebuah *filter*

bank dengan M *bandpass filter* segitiga diterapkan pada isyarat hasil DFT. *Filter bandpass* segitiga ini memiliki sejumlah M jarak frekuensi titik tengah nilai mel yang sama. Bentuk umum *filter* segitiga MFCC ditunjukkan pada Gambar 2.



Gambar 2. Bentuk *Filter* Segitiga MFCC

Nilai yang umum untuk mendefinisikan konstanta bank *filter* ditunjukkan pada Tabel 1.

Frekuensi Sampling (Hz)	16000	11025	8000
Jumlah <i>Filter</i> (M)	40	36	31
Frekuensi Minimum (Hz)	130	130	200
Frekuensi Maksimum (Hz)	6800	5400	3500

2.7 Wavelet Packet Cepstral Coefficient (WPCC)

WPCC merupakan koefisien ciri suara yang diperoleh dari ekstraksi ciri menggunakan wavelet. Munculnya WPCC dikarenakan penggunaan wavelet sebagai ekstraktor ciri memberikan hasil akurasi yang seimbang dengan menggunakan ciri *Mel Frequency Cepstral Coefficient* (MFCC) yang diperoleh dari penerapan *Short Time Fourier Transform* (STFT) terhadap sinyal suara. Namun pada beberapa kasus tertentu, khususnya pengenalan suara fonem (Farooq and Datta, 2003; Hidayat, Hidayat dan Adji, 2015) MFCC masih lebih baik dibandingkan WPCC.

2.8 Algoritma Mean Best Basis (MBB)

Algoritma *Mean Best Basis* (MBB) merupakan algoritma alternatif pemilihan basis dekomposisi wavelet. Algoritma ini merupakan pengembangan dari algoritma *Best Basis* yang telah dikembangkan oleh (Coifman dan Wickerhauser, 1992). Pemilihan basis terbaik dari hasil dekomposisi wavelet merupakan salah satu cara untuk memilih fitur terbaik dari sinyal suara, yaitu dengan menghitung entropy antara sinyal dan *noise*. Berikut adalah langkah-langkah algoritma MBB (Galka dan Ziolkko, 2009):

Untuk setiap elemen sinyal s , hitung *Wavelet Packet Cosine Transform* (WPCT). WPCT merupakan proses dekomposisi *wavelet packet*.

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

di mana

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

Hitung entropy dari setiap simpul pohon dekomposisi :

$$\kappa_{m,j}^i = \kappa(\hat{d}_{m,j}^i) \tag{7}$$

di mana

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

Untuk setiap entropy yang diperoleh dari simpul pohon dekomposisi normalisasi nilai entropy sesuai nilai entropy akarnya menggunakan :

$$\forall_i \forall_{m,j} = \frac{\kappa_{m,j}^i}{\kappa_{0,1}^i} \tag{9}$$

Menghitung pohon nilai entropi rata-rata seluruh sinyal dengan nilai entropi ternormalisasi menggunakan:

$$\bar{W}^\kappa = \{\bar{W}_{m,j}^\kappa \leftrightarrow \bar{\kappa}_{m,j}\} : \bar{\kappa}_{m,j} = \frac{1}{|S|} \sum_{\kappa_{m,j}^i \leftrightarrow W_i^\kappa} \kappa_{m,j}^i \tag{10}$$

Menentukan sub pohon terbaik menggunakan algoritma Basis Terbaik Wickerhauser melalui pohon entropy rata-rata \bar{W}^{opt} .

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

3. METODOLOGI

Secara umum tahapan penelitian yang dilakukan merujuk pada Gambar 3. Penelitian dimulai dengan melakukan pengambilan data penelitian berupa suara sesuai dengan properti yang ditunjukkan pada Tabel 2.



Gambar 3. Tahapan Penelitian

Jenis suara vokal yang direkam adalah suara vokal bahasa Indonesia $a, i, u, e, \acute{e}, o,$ dan \acute{o} . Berikutnya setiap suara rekaman sejenis yang memiliki *noise* akan dihitung SNR-nya. Kemudian

akan dikelompokkan lagi dan dihitung nilai rata-rata SNR-nya. Setelah itu dilakukan dekomposisi terhadap sinyal suara tersebut sampai level ke-*j* dengan metode *wavelet packet*. Kemudian pada sinyal suara terdekomposisi akan diterapkan algoritma MBB kepadanya untuk mendapatkan simpul dan bentuk *tree* terbaik.

Kumpulan simpul terbaik ini kemudian dianalisis sehingga diperoleh bentuk *filter* WPCC terbaik untuk masing-masing level SNR. Pada penelitian ini digunakan wavelet Daubechies db44 dan db45 dengan dekomposisi hingga level-6. Penerapan db44 dan db45 berdasarkan hasil penelitian (Hidayat, Negara dan Kumoro, 2017). Entropy yang digunakan adalah entropy Shannon.

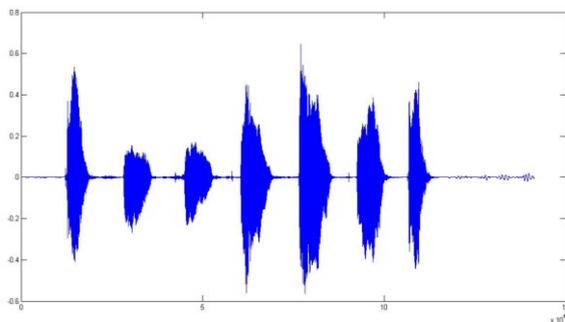
Tabel 2. Properti Pengumpulan Data

No	Properti	Ket
1	Jenis data yang direkam	Vokal
2	Jumlah sampel	50 orang
3	Pengulangan Perekaman	1 kali
4	Tempo Perekaman	Sesuai IPA
5	Frekuensi sampling (<i>f_s</i>)	16 KHz
6	Durasi	1 detik
7	Lingkungan Perekaman	Area terbuka
8	Format Penyimpanan Data	*.wav
9	Alat rekam	Microphone
10	Software perekaman	Audacity

4. HASIL DAN PEMBAHASAN

4.1. Hasil Perekaman

Perekaman telah dilakukan terhadap 50 orang dewasa, masing-masing 25 orang pria dan wanita. Perekaman menggunakan *software* Audacity. Contoh hasil perekaman diberikan pada Gambar 4. Terlihat bahwa amplitudo maksimum setiap suara vokal yang terekam tidaklah sama. Selain itu, data rekaman suara tersebut masih mengandung *noise*. Hasil perekaman kemudian disegmentasi menggunakan perangkat lunak yang sama dan dikelompokkan berdasarkan jenisnya dengan contoh format nama 'a001.wav' merupakan data suara vokal *a* untuk orang pertama.



Gambar 4. Sampel hasil perekaman suara vokal bahasa Indonesia *a, i, u, e, é, o, dan ó*.

4.2. Hasil Perhitungan SNR

Dengan menerapkan rumus perhitungan SNR di atas maka diperoleh nilai rata-rata SNR untuk suara

vokal bahasa Indonesia seperti yang disajikan pada Tabel 3.

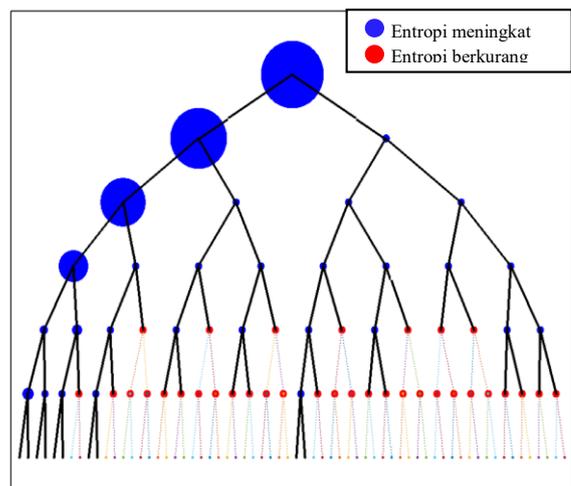
Tabel 3. Nilai Rata-Rata SNR Vokal Bahasa Indonesia

Vokal	SNR (db)
<i>a</i>	-6.3
<i>i</i>	5.3
<i>u</i>	2.6
<i>e</i>	0.14
<i>é</i>	0.6
<i>o</i>	-0.35
<i>ó</i>	-6.2

Dari hasil perhitungan rata-rata SNR yang ditampilkan pada Tabel 3 diperoleh nilai rata-rata SNR untuk masing-masing suara vokal bahasa Indonesia *a, i, u, e, é, o, dan ó* secara terurut adalah -6.27 db, 5.29 db, 2.64 db, 0.14 db, 0.60 db, -0.34 db, dan -6.16 db. Nilai rata-rata SNR ini menunjukkan bahwa sinyal suara yang digunakan adalah suara *ber-noise*. Karena suara dengan SNR > 30 db maka suara tersebut dikategorikan sebagai sinyal yang bersih dari *noise* (Ellis, 2000).

4.3. Bentuk *Tree* dan Simpul Terbaik

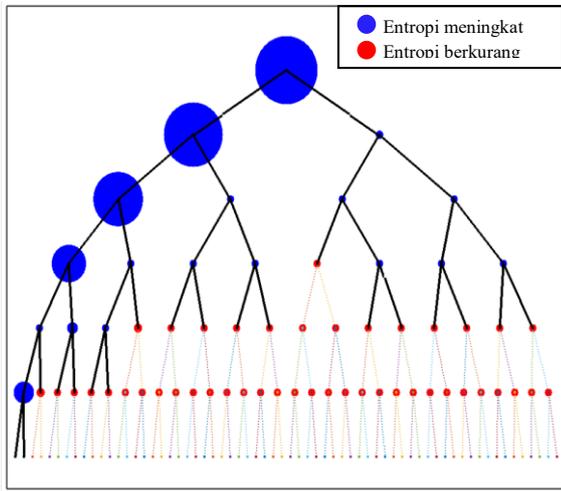
Dari proses penentuan simpul terbaik dengan menggunakan algoritma *Mean Best Basis* diperoleh hasil bahwa dengan wavelet db44 total terdapat 59 simpul yang masuk dalam kategori simpul terbaik. Adapun dengan menggunakan wavelet db45 jumlahnya lebih sedikit, yaitu total hanya 37 simpul yang masuk dalam kategori simpul terbaik. Hasil ini dapat dilihat lebih jelas dengan merujuk pada Gambar 5 dan Gambar 6 tentang bentuk pohon dekomposisi hasil MBB.



Gambar 5. MBB *Tree* untuk wavelet Db44

Pada Gambar 5 diketahui bahwa pada level dekomposisi ke 6, yang memiliki range frekuensi terkecil, terdapat 10 simpul yang masuk kategori *filter* terbaik. Di mana 6 simpul pertama merupakan simpul-simpul awal yang saling berurutan, kemudian 4 simpul selanjutnya mulai memiliki jarak. Jarak 6 simpul pertama hanya terpaut 2 satuan dari 2 simpul berikutnya, sedangkan dengan 2 simpul sisanya

terpisah dengan jarak yang cukup jauh yaitu hingga 26 satuan. Sedangkan pada Gambar 6 hanya 2 simpul pertama saja yang masuk kategori *filter* terbaik, sedangkan sisanya dinihalkan oleh algoritma MBB.



Gambar 6. MBB Tree untuk wavelet Db45

Tabel 4. Simpul dan Leaf Terbaik untuk Db44

No	Simpul	Leaf	Simpul	Leaf	Simpul	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]		

Pada dekomposisi level ke 5 dengan panjang range frekuensi 250, kondisi antara hasil MBB pada

db44 dan db45 masih serupa, di mana jumlah *filter* db45 yang terbentuk masih lebih sedikit daripada *filter* db44. Namun pada level ke 4 dengan panjang frekuensi 500, jumlah *filter* yang terbentuk relatif sama. Perbedaannya hanya terpaut 2 *filter* saja. Adapun pada level 1-3 penerapan MBB masih memberikan hasil yang sama. Ini menunjukkan bahwa, algoritma MBB mempengaruhi koefisien *filter* hasil dekomposisi wavelet mulai dari dekomposisi ke 4 dan seterusnya. Dan dapat disimpulkan bahwa level dekomposisi minimal agar dapat memperoleh koefisien ciri suara menggunakan wavelet adalah dengan menerapkan level dekomposisi ke 4. Namun penggunaan level dekomposisi yang sangat tinggi juga sangat tidak dianjurkan, dikarenakan koefisien *filter* yang terbentuk nantinya tidak dapat merepresentasikan ciri suara.

Tabel 5. Simpul dan Leaf Terbaik untuk Db45

No	Simpul	Leaf	Simpul	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]		

4.4. Desain Filterbank Wavelet Packet Cepstral Coefficient (WPCC) Terbaik

Dengan melakukan analisis dari Tabel 4 dan 5 serta Gambar 5 dan 6 dapat dibentuk koefisien *filter wavelet packet cepstral coefficient* (WPCC) untuk suara vokal bahasa Indonesia menggunakan db44 dan db45. Penentuannya dimulai dari hasil dekomposisi yang berubah setelah diterapkan algoritma MBB kepadanya, yaitu dekomposisi level 4 hingga level 6. *Filter* dipilih mulai dari yang memiliki jarak frekuensi terendah, yaitu 125 Hz pada hasil dekomposisi ke 6, kemudian beranjak ke frekuensi di atasnya, 500 Hz pada level dekomposisi ke 5, demikian seterusnya hingga tidak terdapat perubahan simpul pada level dekomposisi yang dipilih.

Pemilihan *filter* berpindah pada level dekomposisi di atasnya dilakukan jika dan hanya jika pada level dekomposisi tersebut terdapat *filter* yang dinihalkan oleh algoritma MBB (simpul berwarna merah), meskipun pada level dekomposisi tersebut masih ada simpul yang termasuk dalam kategori simpul terbaik. Asumsinya adalah simpul *filter* yang diabaikan tersebut nantinya dapat diwakili oleh

simpul *filter* di atasnya yang berasosiasi dengannya (simpul berwarna biru).

Hasil pemilihan koefisien *filter* WPCC suara vokal bahasa Indonesia untuk wavelet db44 dan db45 ditunjukkan berturut-turut pada Tabel 6 dan Tabel 7. Variabel yang terdapat pada kedua tabel tersebut adalah Frekuensi (F), Range Frekuensi (RF) dan *Leaf* (L). RF terendah adalah 125 Hz dan tertinggi 4000 Hz. Frekuensi tertingginya adalah 8000 Hz yang diperoleh dari persamaan frekuensi Nyquist. Karena frekuensi sampling yang digunakan saat proses perekaman suara vokal bahasa Indonesia adalah 16000 Hz.

Dari hasil penentuan koefisien *filter* diperoleh untuk db44 total terdapat 22 *Leaf* yang menjadi koefisien *filter* dengan range frekuensi terendah 125 Hz dan tertingginya 500 Hz. Sedangkan untuk db45 total terdapat 16 *Leaf* yang menjadi koefisien *filter* dengan range frekuensi terendah 125 Hz dan tertingginya 1000 Hz.

Untuk mempermudah analisis hasil *filter* yang dibentuk dari penerapan algoritma MBB, digunakan nilai magnitude maksimalnya = 1 untuk setiap nilai tengah. Nilai ini adalah nilai dummy untuk dapat membentuk grafik segitga. Dapat dilihat dari Gambar 7 dan Gambar 8 bahwa bentuk desain *filterbank* WPCC mendekati bentuk *filterbank* MFCC.

4.5. Pembahasan

MFCC merupakan *filterbank* yang merepresentasikan cara kerja sistem pendengaran manusia. Skala frekuensi sistem pendengaran manusia merupakan gabungan antara skala frekuensi linier dan skala frekuensi non linier. Skala frekuensi linier berkisar antara 0-1000 Hz, sedangkan jika > 1000 Hz maka skala frekuensinya non linier. Adapun pada penelitian ini, bentuk *filterbank* WPCC yang diperoleh dari penerapan algoritma MBB juga memiliki sifat-sifat tersebut.

Tabel 6. Solusi *Filterbank* untuk Gambar 6

F		RF	Leaf
LF _{cut of}	HF _{cut of}		
1	125	125	[6,0]
126	250	125	[6,1]
251	375	125	[6,2]
376	500	125	[6,3]
501	625	125	[6,4]
626	750	125	[6,5]
751	1000	250	[5,3]
1001	1250	250	[5,4]
1251	1500	250	[5,5]
1501	2000	500	[4,3]
2001	2500	500	[4,4]
2501	3000	500	[4,5]
3001	3500	500	[4,6]
3501	4000	500	[4,7]
4001	4500	500	[4,8]
4501	5000	500	[4,9]
5001	5500	500	[4,10]
5501	6000	500	[4,11]
6001	6500	500	[4,12]
6501	7000	500	[4,13]
7001	7500	500	[4,14]
7501	8000	500	[4,15]

Seperti yang ditunjukkan pada Gambar 7 dan Gambar 8, untuk db44 dan db45 memiliki sifat skala *filter* yang linier pada range frekuensi < 1000 Hz. Hasil db44 dimulai dari RF 0 – 750 Hz, sedangkan untuk db45 dimulai dari RF 0 – 250 Hz. Kemudian setelah itu mulai terjadi perubahan skala frekuensi di atas masing-masing nilai ambang batas tersebut. Sehingga dari hasil desain *filter* tersebut dapat disimpulkan bahwa *filterbank* WPCC dari DB44 lebih mendekati bentuk *filterbank* MFCC konvensional.

Jika dibandingkan dengan hasil penentuan *filterbank* wavelet yang telah dilakukan oleh Farooq dan Datta (2001) serta Choueiter dan Glass (2007) maka bentuk *filter* wavelet setelah menerapkan MBB sangatlah mirip. Perbedaannya hanya terletak pada jumlah *band* frekuensi. Hasil dari MBB jumlah *band* frekuensinya lebih sedikit dari apa yang telah dilakukan oleh Farooq dan Datta serta Choueiter dan Glass, yaitu secara berurutan 22 *band* untuk db44 dan 16 *band* untuk db45. Akan tetapi untuk kisaran RF-nya relatif sama yaitu mulai dari 125 Hz hingga 1000 Hz.

Tabel 7. Solusi *Filterbank* untuk Gambar 7

F		RF	Leaf
LF _{cut of}	HF _{cut of}		
1	125	125	[6,0]
126	250	125	[6,1]
251	500	250	[5,1]
501	750	250	[5,2]
751	1000	250	[5,3]
1001	1250	250	[5,4]
1251	1500	250	[5,5]
1501	2000	500	[4,3]
2001	2500	500	[4,4]
2501	3000	500	[4,5]
3001	3500	500	[4,6]
3501	4000	500	[4,7]
4001	5000	1000	[3,4]
5001	6000	1000	[3,5]
6001	7000	1000	[3,6]
7001	8000	1000	[3,7]

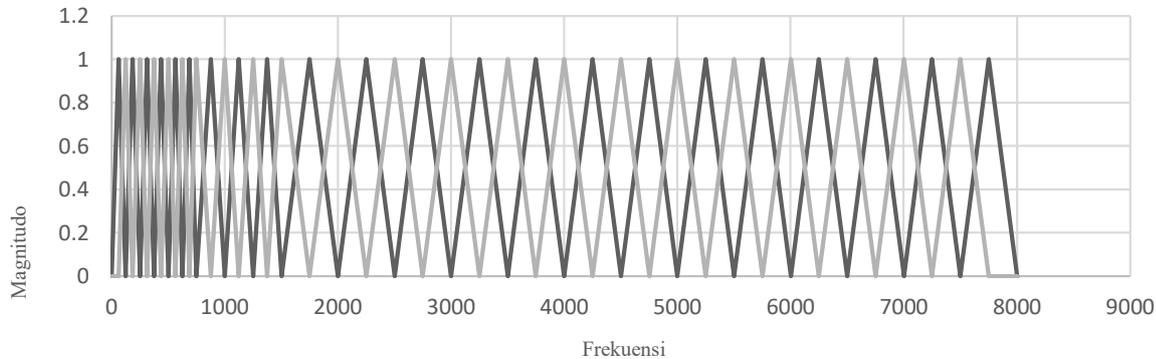
Perbedaan jumlah ini bisa dipengaruhi oleh algoritma yang diterapkan. Namun secara tidak langsung, Farooq dan Datta (2001) serta Choueiter dan Glass (2007) berusaha merekayasa agar jumlah *filter* yang terbentuk sesuai dengan desain mereka. Sedangkan pada hasil MBB desain terbentuk secara alami setelah menerapkan algoritma tersebut. Meskipun demikian, jumlah *band* yang terbentuk masih dalam batas jumlah yang wajar, lihat Tabel 1, dikarenakan pada aplikasi MFCC sebenarnya, hanya sekitar 13 koefisien *filter* saja yang dijadikan sebagai koefisien ciri suara, adapun sisanya diabaikan.

Di dalam algoritma yang diterapkan oleh (Farooq and Datta, 2001) terdapat proses *Discrete Cosine Transform* (DCT) untuk menentukan *band* terbaik. Proses ini juga diterapkan pada algoritma MBB. Bedanya hanya pada langkah penerapannya, di mana DCT oleh (Farooq dan Datta, 2001) diimplementasikan terakhir setelah *tree* terbentuk sedangkan (Galka dan Ziolko, 2009) menerapkannya

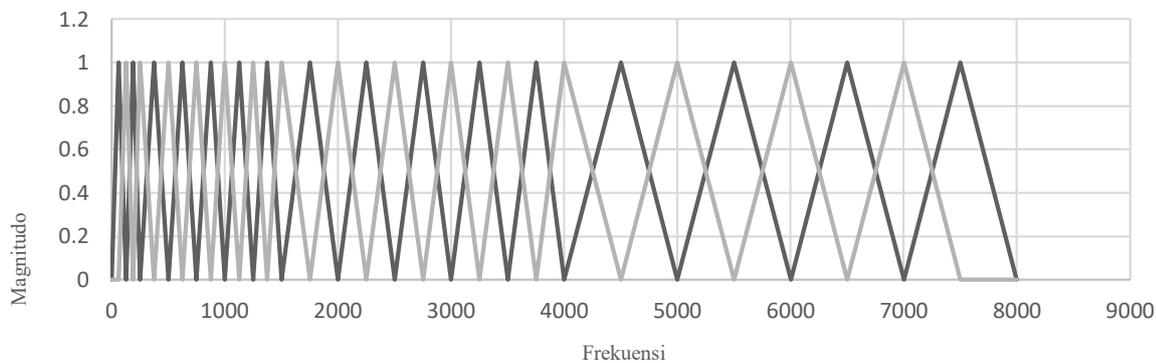
ditengah – tengah proses sebelum melakukan desain *tree* terbaik dari hasil dekomposisi wavelet paket.

Jika memperhatikan variabel *noise*, dapat dikatakan bahwa *noise* tidak berpengaruh terhadap pembentukan filter WPCC. Filter yang terbentuk justru lebih dipengaruhi oleh jenis fungsi wavelet

yang digunakan. Terlihat hanya dua jenis filter wavelet saja yang terbentuk untuk mewakili seluruh suara vokal bahasa Indonesia. Selain itu, keduanya masing-masing merupakan representasi *filterbank* wavelet dari fungsi wavelet db44 dan db45.



Gambar 7. Bentuk *filterbank* WPCC dummy untuk Tabel 5



Gambar 8. Bentuk *filterbank* WPCC dummy untuk Tabel 6

5. KESIMPULAN DAN SARAN

Telah dilakukan penerapan algoritma MBB untuk menentukan *filterbank* WPCC bagi sinyal suara ber-*noise*. Dari hasil yang diperoleh dapat disimpulkan bahwa *noise* tidak berpengaruh dalam pembentukan *filterbank* WPCC. Akan tetapi bentuk filter lebih dipengaruhi oleh jenis fungsi wavelet yang digunakan. Sehingga hanya ada 2 jenis filter WPCC yang terbentuk dan masing-masing merupakan *filterbank* untuk fungsi wavelet db44 dan db45.

Adapun jumlah total *filterbank* untuk kedua *filterbank* WPCC tersebut masih lebih sedikit dari syarat jumlah *filterbank* MFCC. Meskipun demikian keduanya telah memenuhi syarat bentuk *filterbank* MFCC terutama dari bentuk skala dan range frekuensinya. Untuk skala frekuensi, bentuknya telah menggabungkan skala frekuensi linier (0-1000 Hz) dan skala frekuensi non linier (>1000 Hz). Adapun range frekuensi yang terbentuk minimumnya 125 Hz dan maksimumnya 1000 Hz. Sehingga, dapat disimpulkan kedua bentuk *filter* ini dapat dipertimbangkan untuk digunakan sebagai *filter* WPCC.

Untuk penelitian selanjutnya, hasil dari penelitian ini akan diterapkan pada proses pengenalan suara vokal bahasa Indonesia. Performa pengenalan akan digunakan untuk menentukan *filter* WPCC terbaik, di antara kedua bentuk *filter* WPCC db44 dan db45, dalam mengekstraksi ciri suara vokal bahasa Indonesia. Performa WPCC sebagai koefisien ciri suara untuk jenis suara dengan dan tanpa *noise* juga akan dievaluasi.

Ucapan Terima Kasih

Terima kasih kepada Direktorat Riset dan Pengabdian Masyarakat, Direktorat Jendral Penguatan Riset dan Pengembangan, KEMENRISTEDIKTI, yang telah mendanai penelitian ini melalui skema Penelitian Dosen Pemula (PDP) dengan nomor hibah SP DIPA-042.06.1.401516/2018.

DAFTAR PUSTAKA

ABO-ZAHHAD, M., AHMED, S.M. dan ABBAS, S.N., 2016. Biometrics from heart sounds: Evaluation of a new approach based on wavelet packet cepstral features using HSCT-11 database. *Computers and*

- Electrical Engineering*, [online] 53, pp.346–358. Tersedia di: <<http://dx.doi.org/10.1016/j.compeleceng.2016.05.004>>.
- ADAM, T.B., SALAM, M.S. dan GUNAWAN, T.S., 2013. Wavelet Cepstral Coefficients for Isolated Speech Recognition. *Telkomnika*, 11(5), pp.2731–2738.
- ANUSUYA, M.A. dan KATTI, S.K., 2011. Front end analysis of speech recognition: a review. *International Journal of Speech Technology*, [online] 14(2), pp.99–145. Tersedia di: <<http://link.springer.com/10.1007/s10772-010-9088-7>> [Diakses 11 May 2015].
- ANUSUYA, M. dan KATTI, S., 2009. Speech recognition by machine: A review. *International Journal of Computer Science and Information Security*, [online] 6(3), pp.181–205. Tersedia di: <http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1454427>.
- CHOUeiter, G.F. dan GLASS, J.R., 2007. An implementation of rational wavelets and filter design for phonetic classification. *IEEE Transactions on Audio, Speech and Language Processing*, 15(3), pp.939–948.
- COIFMAN, R.R. dan WICKERHAUSER, M.V., 1992. Entropy-Based Algorithms for Best Basis Selection. *IEEE Transaction on Information Theory*, 38(2), pp.713–718.
- DESAI, N., DHAMELIYA, P.K. dan DESAI, P.V., 2013. Feature Extraction and Classification Techniques for Speech Recognition: A Review. *International Journal of Emerging Technology and Advanced Engineering*, 3(12), pp.367–371.
- ELLIS, D.P.W., 2000. *ICSI Speech FAQ: 4.1 How is the SNR of a speech example defined?* [online] ICSI Berkeley. Tersedia di: <<http://www1.icsi.berkeley.edu/Speech/faq/index.html>> [diakses 10 Oct. 2018].
- FAROOQ, O. dan DATTA, S., 2001. Mel Filter-Like Admissible Wavelet Packet Structure for Speech Recognition. 8(7), pp.196–198.
- FAROOQ, O. dan DATTA, S., 2003. Phoneme recognition using wavelet based features. In: *Information Sciences*. pp.5–15.
- GAIKWAD, S.K., GAWALI, B.W. dan YANNAWAR, P., 2010. A Review on Speech Recognition Technique. *International Journal of Computer Applications*, 10(3), pp.16–24.
- GALKA, J. dan ZIOLKO, M., 2009. Mean Best Basis Algorithm for Wavelet Speech Parameterization. In: *Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. pp.1110–1113.
- HIDAYAT, S., HIDAYAT, R. dan ADJI, T.B., 2015. Speech Recognition of KV-Patterned Indonesian Syllable using MFCC, Wavelet and HMM. *Jurnal Ilmiah Kursor*, 8(2), pp.67–78.
- HIDAYAT, S., NEGARA, H.R.P. dan KUMORO, D.T., 2017. Determination of the Optimum Wavelet Basis Function for Indonesian Vowel Voice Recognition. *Jurnal Elektronika dan Telekomunikasi (JET)*, [online] 17(2), pp.42–47. Tersedia di: <<http://www.jurnalet.com/jet/article/view/194>> [Diakses 16 Mar. 2018].
- HUANG, X., ACERO, A. dan HON, H.-W., 2001. *Spoken Language Processing: A Guide to Theory, Algorithm and System Development*. [online] Processing. Prentice Hall. Tersedia di: <<http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0130226165>>.
- MCLOUGHLIN, I., 2009. *Applied Speech And Audio Processing : With Matlab Examples*. 1st ed. [online] United Kingdom: Cambridge University Press. Tersedia di: <www.cambridge.org/9780521519540>.
- PAVEZ, E. dan SILVA, J.F., 2012. Analysis and design of Wavelet-Packet Cepstral coefficients for automatic speech recognition. *Speech Communication*, [online] 54(6), pp.814–835. Tersedia di: <<http://dx.doi.org/10.1016/j.specom.2012.02.002>>.
- RASHMI, C.R., 2014. Review of Algorithms and Applications in Speech Recognition System. *International Journal of Computer Science and Information Technologies (IJCSIT)*, 5(4), pp.5258–5262.
- RIOUL, O. dan VETTERLI, M., 1991. Wavelets and Signal Processing. *IEEE SP Magazine*. Oct.
- TURNER, C. dan JOSEPH, A., 2015. A Wavelet Packet and Mel-Frequency Cepstral Coefficients-Based Feature Extraction Method for Speaker Identification. *Procedia Computer Science*, [online] 61, pp.416–421. Tersedia di: <<http://dx.doi.org/10.1016/j.procs.2015.09.177>>.

Halaman ini sengaja dikosongkan