Pengenalan Tutur Vokal Bahasa Indonesia Menggunakan Metode DWT dan DTW

A.Asni B.[1], Risanuri Hidayat [2], Noor Akhmad Setiawan [3]

Mahasiswa S2 Teknik Elektro dan Teknik Informasi, Universitas Gadjah Mada[1]
a_asni_b@yahoo.com
Teknik Elektro dan Teknik Informasi,Fakultas Teknik , Universitas Gadjah Mada[2]
Teknik Elektro dan Teknik Informasi,Fakultas Teknik , Universitas Gadjah Mada[3]

Abstract

Bunyi tutur vokal bahasa indonesia masih sulit dibedakan oleh sistem pengenalan tutur. Sifat non-stasioner, perbedaan kecepatan, dan noise merupakan faktor yang mempengaruhi hasil pengenalan tutur. Penelitian ini bertujuan mengukur kesamaan dan perbedaan antar isyarat-isyarat tutur vokal Bahasa Indonesia dengan melakukan ekstraksi ciri berbasis DWT. Dekomposisi WPT full binary level 3 dan 5 diterapkan untuk ekstraksi ciri. Algoritma DTW diterapkan untuk validasi dengan cara mengukur kesamaan dua isyarat tutur . Hasil yang dicapai menunjukkan tingkat akurasi pengenalan yang tinggi hingga 100 % . Selisih pengukuran terbaik dari dekomposisi WPT full binary level 3 sebesar 72 % sedangkan dekomposisi level 5 hanya 12 % .

Kata Kunci: *Dynamic Time Warping* ,DTW, *Discrete Wavelet Transform*, DWT

I. PENDAHULUAN

Identifikasi satu kata atau satu huruf vokal yang dituturkan menjadi masalah tersendiri bagi sistem pengenalan tutur. Contoh isyarat tutur yang sama dari satu sumber penutur dan diulang diwaktu berbeda sehingga memiliki kecepatan dan waktu pencuplikan yang berbeda akan menjadi masalah pada suatu sistem identifikasi tutur, berbeda dengan otak manusia yang dengan cerdas mampu mengidentifikasi hal tersebut dengan mudah. Metode *Dynamic Time Warping* (DTW) merupakan salah satu metode untuk mengatasi perbedaan kecepatan yang pertama kali diusulkan oleh Saoke dan Chiba [1]. Faktor lain yang mempengaruhi sistem pengenalan isyarat tutur diantaranya, sifat isyarat tutur yang tidak stasioner dan noise yang tidak bisa lepas dari lingkungan isyarat tutur.Berbagai Algoritme ekstraksi ciri dan pengenalan pola telah dikembangkan untuk memperoleh hasil yang optimal yang diukur berdasarkan tingkat akurasi pengenalan hingga efisiensi dari segi komputasi[1-9].

Sistem pengenalan tutur yang handal adalah sistem yang mampu mengatasi sifat non-stasioner dari isyarat tutur dan bisa menyaring kebisingan yang ikut dalam isyarat tutur serta mampu mengatasi perbedaan kecepatan isyarat tutur. Metode DTW sudah banyak diteliti dan diterapkan dalam pengenalan isyarat tutur diantaranya, untuk pengenalan kata terisolasi angka digit menggunakan bahasa Inggris, dengan menerapkan ekstraksi ciri *Mel Frequency Cepstral Coefficient* (MFCC)[2-4].

Penelitian tentang perbandingkan metode DTW dan *Hidden Markov Models* (HMM) dengan ekstraksi *Mel Frequency Cepstrum Coefficient* (MFCC) menyimpulkan bahwa metode HMM lebih unggul dalam penerapan isyarat tidak starsioner dibandingkan metode DTW[5].Untuk menyamai tingkat akurasi pengenalan pola HMM , filter median ditambahkan pada metode DTW[6]. Selanjutnya pengembangan metode DTW untuk peningkatan akurasi pengenalan, dengan penerapan algoritme Shape Averaging (SA) pada DTW dilakukan oleh peneliti[7]. Kemudian berdasarkan review peneliti lainya disimpulkan bahwa metode DTW memiliki keunggulan dalam mengatasi distorsi akibat pergeseran waktu dan tidak memerlukan komputasi yang kompleks[8].

Berdasarkan beberapa hasil penelitian di atas akurasi pengenalan di fokuskan pada pengembangan algoritme DTW, sebagian besar menerapkan metode ekstraksi ciri MFCC, peneliti yang lain menerapkan algoritme tambahan untuk menyaring isyarat. Penelitian metode ekstraksi ciri menggunakan DWT dengan menghitung nilai entropy minimum dari hasil lokalisasi adaptif-frekuensi untuk mencari basic terbaik hasil dekomposisi DWT telah dilakukan peneliti [9]. Peneliti yang lain menggunakan metode DWT untuk mengatasi isyarat yang mengandung derau dengan melakukan dekomposisi hingga level 5. Prosedur pada ekstraksi ciri yang sebelumnya menggunakan *Mel Scale filter-bank* digantikan hasil dari paket wavelet [10]. Peneliti yang lain menerapkan ekstraksi ciri yang menggunakan energi dari frekuensi *sub-band* hasil dekomposisi Wavelet Transform (WT) dan diterapkan bersama metode pengenalan pola GMM[11].

Hasil peneliti terdahulu dari uraian di atas belum ada yang penerapan metode DWT bersama DTW tradisional pada vokal Bahasa Indonesia. Penelitian ini telah mengupayakan sebuah ekstraksi ciri menggunakan metode DWT yang dapat mengoptimalkan hasil pengenalan DTW tradisional. Tiga metode ekstraksi ciri dibandingkan yaitu; metode pertama menggunakan metode *dyadic DWT* level 8 yang terdiri dari 9 ciri, metode kedua menggunakan *full binary DWT* level 3 yang terdiri dari 8 ciri dan metode ketiga menggunakan *full binary DWT* level 5 yang terdiri dari 32 ciri. Pengukuran DTW dilakukan untuk menentukan metode DWT yang optimal. Metode kedua dan ketiga adalah metode yang diusulkan untuk dibandingkan dengan metode ekstraksi ciri pertama dari peneliti [9].

II. *DISCRETE WAVELET TRANSFORM* (DWT)

Wavelet adalah gelombang dengan durasi terbatas yang memiliki nilai rata-rata nol. Tidak seperti isyarat sinusoida yang secara teoritis memiliki panjang dari minus ke plus tak terhingga, wavelet memiliki awal dan akhir.

Era tahun 80-an wavelet muncul sebagai revolusi frekuensi-waktu dalam pemrosesan sinyal. Pada tahun 1989 Mallat mengusulkan algoritme Fast *Discrete Wavelet Transform* (DWT) untuk menguraikan isyarat menggunakan satu set dekomposisi *Quadrature Mirror Filter* (QMF), yang memiliki sifat khusus wavelet untuk setiap *band-pass* dan *low-pass*. Sejak periode ini wavelet telah diterapkan dalam berbagai bidang termasuk dinamika fluida, teknik, geofisika keuangan, studi nada musik, audio, pemampatan gambar dan *de-noising* .Dalam analisis wavelet diskrit ,informasi yang tersimpan dalam koefisien wavelet tidak diulang, memungkinkan regenerasi lengkap dari sinyal asli tanpa redundansi atau pengulangan informasi yang sama[10-11].

DWT diaplikasikan dalam data diskrit untuk menghasilkan keluaran diskrit yang mentransformasikan isyarat dari domain waktu (domain asli dari isyarat tutur) ke domain wavelet. Proses dekomposisi dan rekonstruksi menggunakan *Fast* DWT merupakan proses konvolusi antara isyarat dan koefisien filter, hasil konvolusi kemudian diseleksi menggunakan faktor 2 untuk proses *downsamping* dan *upsampling.*

Persamaan proses dekomposisi :

$$a_k^{(j+1)} = \sum_{n=-\infty}^{\infty} h_{n-2k} a_n^{(j)} = (a^{(j)} * h^{(0)})(2k) \quad \textbf{(1)}$$

$$d_k^{(j+1)} = \sum_{n=-\infty}^{\infty} g_{n-2k} a_n^{(j)} = (a^{(j)} * g^{(1)})(2k) \quad \textbf{(2)}$$

Persamaan proses rekonstruksi:

$$a_k^{(j)} = \sum_{n=-\infty}^{\infty} h_{k-2n} a_n^{(j+1)} + \sum_{k=-\infty}^{\infty} g_{k-2n} d_n^{(j+1)} \quad \textbf{(3)}$$

$$a_k^{(j)} = (\breve{a}^{(j+1)} * h)(k) + (\breve{d}^{(j+1)} * g)(k)$$

Dengan:

$$\breve{a}_k^{(j+1)} = \begin{cases} a_p^{(j+1)} & if \quad k = 2p \\ 0 & if \quad k = 2p+1 \end{cases} \quad \textbf{(4)}$$

dan

$$\breve{d}_k^{(j+1)} = \begin{cases} d_p^{(j+1)} & if \quad k = 2p \\ 0 & if \quad k = 2p+1 \end{cases}$$

$\breve{a}^{(j+1)}$ dan $\breve{d}^{(j+1)}$ adalah koefisien aproksimasi dan detail pada level j+1 yang nilainya berasal dari $a_k^{(j+1)}$ dan $d_k^{(j+)}$ yang melalui operasi *dyadupsampling* seperti pada persamaan 2-6 ,yaitu menambahkan nilai nol diatara 2 titik interval, jika interval ganjil akan diisi dengan nol, kemudian hasilnya akan dikonvolusikan dengan koefisien *filter* h (LPF) dan g (HPF)

Isyarat sebelumnya dinormalisasi menggunakan *dc removal*, dan isyarat diam dibuang sebelum proses dekomposisi. Aplikasi fungsi "wpdec" yang ada pada Matlab *wavelet toolbox* digunakan untuk dekomposisi isyarat tutur. Tiga cara berbeda untuk memperoleh vektor ciri diterapkan untuk mencari karakteristik isyarat tutur vokal. Jenis wavelet Daubechies (db-N ,orde (N=2 dan N=10) akan diterapkan dalam memperoleh vektor ciri.

Metode pertama menggunakan metode *dyadic DWT* level 8 yang terdiri dari 9 ciri, metode kedua menggunakan *full binary DWT* level 3 yang terdiri dari 8 ciri dan metode ketiga menggunakan *full binary DWT* level 5 yang terdiri dari 32 ciri. Proses pembentukan vektor ciri dengan menghitung energi masing-masing frekuensi *sub-band* hasil rekonstruksi [9]:

$$E_{i=} = \sqrt{\sum_{k=1}^{N} \left| X_i(k) \right|^2} \quad \textbf{(5)}$$

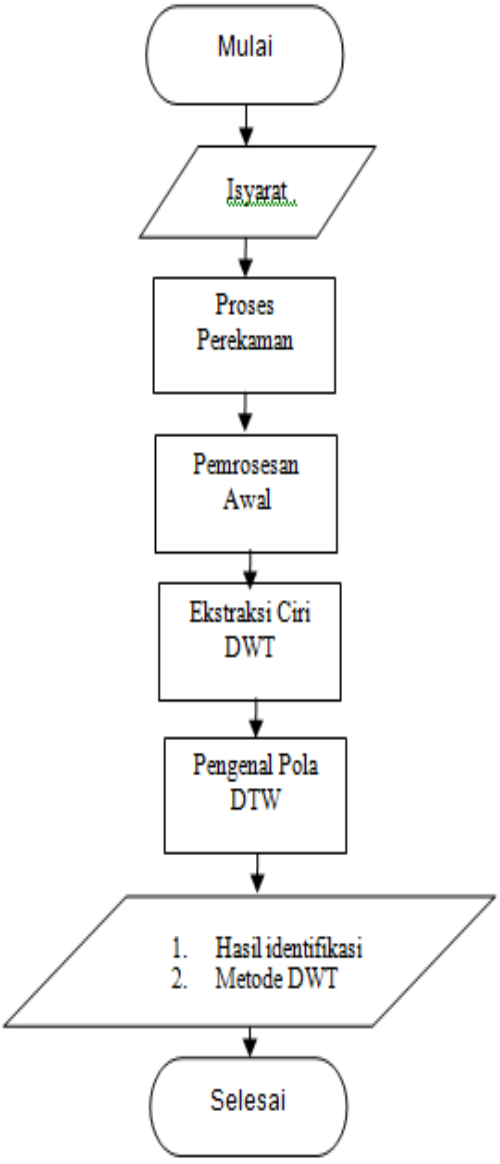
Secara umum proses pengenalan isyarat tutur dilakukan seperti pada Gambar 1.

. Total energi dihitung dengan persamaan

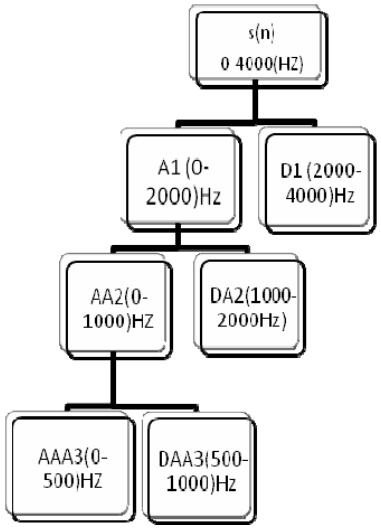
$$E_{tot} = \sqrt{\sum_{i=1}^{I} E_i^{\,2}} \quad \textbf{(6)}$$

N adalah panjang isyarat, l adalah jumlah sub-band frekuensi ., karakteristik vektor ciri diperoleh dengan membagi setiap total energi *sub-band* dengan total energi yang ada pada level j dengan persamaan (7)

$$V_{energi} = \frac{E_i}{E_{tot}} \quad \textbf{(7)}$$



Gambar 1 . Proses Pengenalan Isyarat



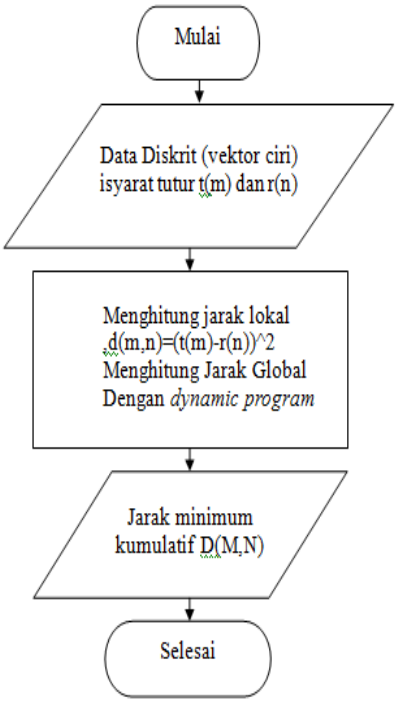
Gambar 2. Gambar Dekomposisi *Dyadic* DWT lev-3

III. *DYNAMIC TIME WARPING* (DTW)

Dynamic Time Warping adalah algoritme berbasis kesamaan ukuran yang memberikan hasil pengukuran jarak antara dua isyarat. Asumsikan dua isyarat tutur, didefinisikan mengatakan x (t_i) dan x (t_j), masing-masing dengan basis waktu sendiri, t_i dan t_j. Juga menganggap bahwa awal dan akhir dari isyarat suara yang dikenal, masing-masing dinotasikan sebagai (t_{is},t_{if}) dan (t_{js}, t_{jf}). Jika kedua isyarat adalah sampel pada tingkat yang sama, maka sample t kedua isyarat mulai i = j = 1. Pemetaan fungsi, i = j. (i / j), adalah menuju linear. Isyarat tutur bersifat tidak linear, sehingga fungsi *non-linear time warping* harus dihitung, dengan beberapa asumsi. Misal fungsi, w (k), didefinisikan sebagai urutan titik: c (1), c (2), ..., c (k), dimana c (k) = (i (k), j (k)) adalah pencocokan dari titik i (k) pada basis waktu pertama dan titik j (k) pada basis waktu kedua.

Proses *warping*, w (k), hanya boleh dengan batasan yang diberikan, dengan pengaturan yang disebut:

1. *Monotonic*; $i(k-1) \leq i(k)$ dan $j(k-1) \leq j(k)$, yaitu langkah jalur tidak akan kembali ke waktu (indeks) ,sehingga tidak ada pengulangan jalur pada ciri isyarat yang sama.
2. *Continuity*; $i(k)-i(k-1) \leq 1$ dan $j(k)-j(k-1) \leq 1$, yaitu fungsi warping tidak akan melompoti waktu (indeks), hal ini menjamin jalur tidak akan mengabaikan ciri isyarat yang penting.



Gambar 2**.** Proses *Dynamic Time Warping*

3. Kedua batasan pertama dan kedua dituliskan pada persamaan (8)

$$c(k-1)=\begin{cases}(i(k),\ j(k-1),\\ (i(k)-1,\ j(k)-1),\\ or(i(k)-1,\ j(k).\end{cases} \quad \textbf{(8)}$$

4. *Boundary*; i(1)=1, j(1)=1, dan i(K)= I , j(K)=J, yaitu langkah penjajaran (*warping)*, dimulai dari titik (1,1) dan berakhir pada titik (I,J), jika dalam matriks maka berawal dari posisis kiri atas dan berakhir pada posisi kanan bawah..

Metode DTW digunakan untuk menentukan kesamaan atau perbedaan antara dua isyarat tutur yang dibandingkan tanpa proses pelatihan terlebih dahulu dengan menggunakan diskriminasi jarak. Keluaran algoritme DTW ada dua yaitu, nilai jarak DTW dan isyarat yang dinormalisasi dengan DTW. Dalam penelitian ini yang dingunakan adalah nilai jarak DTW saja. Data diperoleh dari pengukuran DTW berdasarkan hasil pengukuran jarak terkecil yang digunakan dalam pengenalan pola menggunakan persamaan logika (10) untuk mengambil keputusan :

$$dwt_{(x,x)}=\begin{cases}1 & jika\ \ dwt(x,x)<d(x,y)\\ 0 & jika\ \ dwt(x,x)\geq d(x,y)\end{cases} \quad \textbf{(9)}$$

IV. IMPLEMENTASI METODE YANG DIUSULKAN

Bahan penelitian berupa rekaman tutur yang diperoleh dari satu sumber penutur yang menuturkan huruf vokal; “a” ,”e”,”i”,”o”,”u”, masing-masing diulang sebanyak 3 kali, yang di simpan dalam format “wav”. Sehingga diperoleh 15 isyarat tutur.

Proses ekstraksi ciri sesuai langkah pada Bagian II yang menguraikan metode ekstraksi ciri sehingga diperoleh 15 vektor ciri pada masing-masing metode. Dari 15 vektor ciri dibuat berpasangan hingga diperoleh 225 kemungkinan pasang data yang diukur dengan menerapkan metode DTW yang dijelaskan pada Bagian III sehingga menghasilkan 225 hasil pengukuran DTW, diataranya ada 9 yang merupakan pasangan target untuk masing-masing isyarat vokal . Jarak DTW hasil pengukuran dirangkum dalam lembar kerja menggunakan microsoft excel.

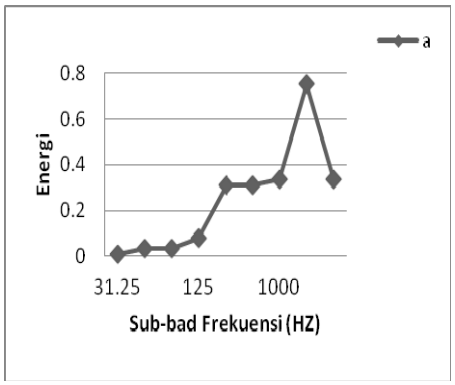
Langkah selanjutnya adalah membandingkan hasil pengukuran dari masing-masing metode ekstraksi ciri untuk kemudian dianalisis guna memperoleh metode yang lebih baik untuk di terapkan lebih lanjut.

V. HASIL SIMULASI DAN DISKUSI

Contoh hasil vektor ciri dari metode pertama (dekomposisi *dyadic* DWT level 8) pada Gambar 4. Metode kedua (dekomposisi *full binary* DWT level 3) dan Metode ketiga (dekomposisi *full binary* DWT level 3), pada Gambar 5 dan 6.

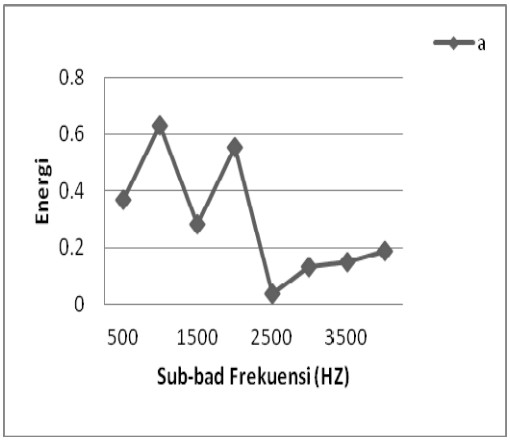
Hasil pengujian metode yang diusulkan dari 225 pasangan pengukuran menggunakan 15 isyarat tutur vokal dari sumber penutur yang sama, sebagai berikut:

1. Nilai persentase pengenalan 100 %

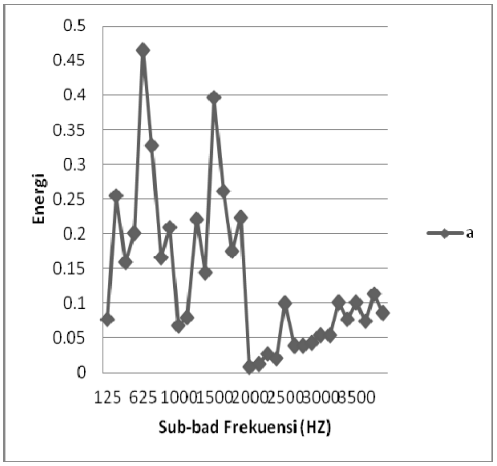


Gambar 3. Hasil Vektor Isyarat Vokal “a” dengan Metode I

menggunakan vektor ciri orde 2 dan orde 10 *mother* wavelet Daubechies



Gambar 4. Hasil Vektor Isyarat Vokal "a" dengan Metode II



Gambar 5. Hasil Vektor Ciri Isyarat vokal "a" dengan Metode III

2. Nilai persentase pengenalan 100 % diperoleh dari metode satu (*dyadic* DWT level 8), metode dua (*full binary* DWT level 3), dan metode tiga (*full binary* DWT level 5)

Vektor ciri dari metode I, II dan III diuji dengan menggunakan pengukuran DTW. Hasil yang diperoleh dari pengujian mencapai tingkat akurasi 100 % untuk masing-masing metode.

Analisis lebih lanjut dilakukan untuk membandingkan ketiga metode dengan menganalisis jarak yang terbaik dalam pengukuran menggunakan vektor ciri masing-masing. dibandingkan jarak hasil pengukukuran DTW untuk melihat . Karakteristik Vektor ciri masing-masing metode dapat dilihat pada Gambar 3, gambar 4, Gambar 5.

Vektor ciri dikatakan baik jika dibandingkan dengan vektor ciri dari kelas yang sama maka akan menghasilkan pengukuran yang paling kecil sebaliknya jika dibandingkan dengan vektor ciri yang berasal dari kelas yang berbeda maka jarak pengukuran menjadi lebih besar.

Perbandingan hasil pengukuran dari masing-masing metode disajikan pada Tabe 1. Nilai yang di cetak tebal menandakan hasil pengukuran yang terbaik diantara ketiga metode yang diujikan.Tiap metode terdapat 25 hasil pengukuran DTW. Metode kesatu memberikan 4 dari 25 hasil terbaik (16%), **metode kedua memberikan 18 dari 25 hasil terbaik (72%)**, sementara metode ketiga hanya memberikan 3 dari 25 hasil pengukuran terbaik (12%).

Table 1. Tabel Perbandingan Metode Ekstraksi Ciri Isyarat Tutur Vokal bedasarkan hasil pengukuran DTW

Vokal	Metode	a	e	i	o	u
a	I	0.0215	0.1484	0.2965	0.1635	0.2812
	II	**0.0180**	**0.2423**	**0.5814**	**0.2077**	**0.4748**
	III	0.0274	0.1547	0.2472	0.1973	0.1797
e	I	0.1484	0.0276	**0.3733**	0.1769	**0.5210**
	II	**0.2423**	**0.0264**	0.2807	**0.2226**	0.3661
	III	0.1547	0.0573	0.2968	0.2060	0.3333
i	I	0.2965	**0.3733**	0.0141	0.2210	0.1300
	II	**0.5814**	0.2807	**0.0028**	0.2543	0.1238
	III	0.2472	0.2968	0.0341	**0.2630**	**0.1907**
o	I	0.1635	0.1769	0.2210	0.0098	0.2314
	II	**0.2077**	**0.2226**	**0.2543**	**0.0036**	**0.2663**
	III	0.1973	0.1573	0.2630	0.0583	0.1823
u	I	0.2812	**0.4542**	0.1300	0.1732	0.0261
	II	**0.4748**	0.3661	0.1238	**0.2663**	**0.0196**
	III	0.1797	0.3247	**0.1926**	0.1823	0.0571

VI. KESIMPULAN

Hasil penelitian menunjukkan metode DWT dan DTW dapat diterapkan dalam pengenalan isyarat tutur vokal Bahasa Indonesia, sebuah metode ekstraksi ciri yang lebih efektif dengan 8 vektor ciri dan pengenalan pola DTW tradisional dapat digunakan sehingga waktu komputasi dapat dihemat. Analisis pengenalan masih bersifat hitungan manual, dan dapat dikembangkan untuk dibuat otomatis oleh peneliti berikutnya sehingga dapat diujikan untuk jumlah data yang lebih besar.

REFERENSI

[1] H. Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition," *IEEE Trans. Acoust.*, vol. ASSP-26, no. 1, pp. 43–49, 1978.

[2] S. D. Dhingra, G. Nijhawan, and P. Pandit, "Isolated Speech Recognition using MFCC and DTW," *IJAREEIE*, pp. 4085–4092, 2013.

[3] A. Bala, "Voice Command Recognition System Based on MFCC AND DTW," vol. 2, no. 3491, pp. 7335–7342, 2010.

[4] L. Muda, M. Begam, and I. Elamvazuthi, "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques," *Jouranal Comput.*, vol. 2, no. 3, pp. 138–143, 2010.

[5] S. C. Sajjan and C. Vijaya, "Comparison of DTW and HMM for Isolated Word Recognition," *IEEE*, no. 1, pp. 466–470, 2012.

[6] Z. Yuxin, Y. Miyanaga, and C. Siriteanu, "New Robust Speech Recognition Using DTW in Noise," *IEEE Isc. 2010*, no. 1, pp. 34–38, 2010.

[7] D. Srisai and C. A. Ratanamahatana, "Efficient Time Series Classification under Template Matching using Time Warping Alignment," *IEEE Int. Conf. Comput. Sci. Converg. Inf. Technol.*, pp. 685–690, 2009.

[8] P. Senin, "Dynamic Time Warping Algorithm Review," Hawaii,USA, 2008.

[9] C. J. Long and S.Datta, "Wavelet Based Feature Extraction for Phonem Recognition," *IEEE Spok. Lang. 1996. ICSLP 96. Proceedings., Fourth Int. Conf.*, vol. 1, pp. 264–267, 1996.

[10] X. Wu, F. Tian, and J. Liu, "An Improved Speech Feature Extraction Algorithm Using DWT," pp. 1086–1090, 2008.

[11] X. Zhao, Z. Wu, J. Xu, K. Wang, and J. Niu, "Speech Signal Feature Extraction Based on Wavelet Transform," *IEEE Int. Conf. Intell. Comput. Bio- Med. Instrum.*, no. 1, pp. 1–4, 2011.