

# Pengenalan Isyarat Tutur Vokal Bahasa Indonesia Menggunakan Metode *Dynamic Time Wrapping* (Dtw) Berbasis Fungsi Jarak

<sup>1</sup>Dwi Novianto, <sup>2</sup>Risky Via Yuliantari

<sup>1</sup>Jurusan Teknik Elektro, Universitas Tidar

Jln. Kaptem Suparman 39 Magelang 56116

<sup>1</sup>nopi\_jte@untidar.ac.id, <sup>2</sup>rviayulantari@gmail.com

**Abstrak** - Developing of vowel recognition system is explained. DWT (Discrete Wavelet Transform) level 3 and DTW (Dynamic Time Wrapping) with euclidean and manhattan distance function were used. DTW was used for pattern recognition and DWT was for feature extraction. First, Vowels of seven speakers were recorded, second they were extracted by DWT and the last, the waves were recognized by DTW with euclidean and manhattan distance. The best result for those processes were 80,82% for euclidean distance function and 83,7% for manhattan distance function

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

## I. PENDAHULUAN

Sistem identifikasi satu vokal pengenalan isyarat tutur dengan menggunakan algoritme pembelajaran pada mesin pengenalan, belum mampu memberikan pengenalan sejauhnya otak manusia. Hal tersebut berbeda dengan otak manusia dalam proses identifikasi yang mudah dalam waktu yang singkat, sehingga perlu dilakukan eksplorasi pada algoritme yang sudah ada.

Sifat yang terdapat pada isyarat tutur antara lain: sinyal yang tidak stasioner, adanya perubahan kecepatan suara dan derau. Hal tersebut dipengaruhi oleh lingkungan sekitar merupakan masalah dalam sistem pengenalan isyarat tutur. *Dynamic Time Wrapping* (DTW) sebagai metode pengenalan isyarat tutur digunakan untuk mengoptimalkan hasil pengenalan suara tanpa harus mengurangi komputasi [1]. Sedangkan *Discrete Wavelet Transform* (DWT) sebagai metode ekstraksi ciri digunakan untuk mengatasi isyarat tutur yang mengandung derau dan sinyal yang tidak stasioner, serta mengurangi panjang sinyal input [2].

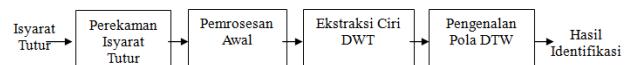
Banyak penelitian yang dilakukan untuk meningkatkan kemampuan pengenalan tutur. Sehingga sistem pengenalan tutur menggunakan aplikasi komputer bukan merupakan hal yang baru [3]. Beberapa contoh penerapan pengenalan isyarat tutur diantaranya tentang penambahan filter median pada metode DTW untuk menyamai tingkat akurasi pengenalan pola *Hidden Markov Model* (HMM) [4]. Penerapan algoritma *Shape Averaging* (SA) pada DTW untuk peningkatan akurasi pengenalan [5]. Penerapan ekstraksi ciri *Mel Frequency Cepstral Coefficient* (MFCC) untuk pengenalan kata terisolasi angka menggunakan bahasa Inggris [6]. Pengenalan isyarat tutur bahasa Indonesia secara konvensional untuk vokal [7] dan beberapa kata terisolasi menggunakan *Mel Frequency Cepstral Coefficient* (MFCC) secara otomatis [8].

Metode klasifikasi pengenalan pola pengukuran fungsi jarak yang dilakukan Yaniar dan Nugraheny dengan cara membandingkan metode pengukuran fungsi jarak *euclidean*, *manhattan*, *mahalanobis*, *correlation*, *angle based*, *mean squared error* (MSE), dan *sum squared error* (SSE) untuk menghasilkan akurasi terbaik [9][10]. Pengukuran fungsi jarak dilakukan dengan pengambilan data berdasarkan kemiripan atau perbedaan antara satu pola dengan pola yang lain [11]. Fungsi jarak yang paling umum digunakan adalah fungsi jarak *euclidean* dan *manhattan* dikarenakan tidak banyak membutuhkan komputasi dan memiliki tingkat akurasi yang cukup tinggi [9][12][13].

Dalam penelitian ini pengukuran fungsi jarak digunakan sebagai metode klasifikasi pengenalan pola isyarat tutur vokal bahasa Indonesia untuk dapat meningkatkan proses pengenalan pada fonem sengau bahasa Indonesia. Ekstrasi ciri dilakukan dengan menyeleksi keseluruhan ciri yang ada untuk mendapatkan ciri yang unik dari suatu sinyal suara agar dapat dibedakan antara sinyal yang satu dengan sinyal suara lain sehingga dapat digunakan sebagai data acuan untuk mengenali setiap masukan sinyal suara.

## II. METODE PENELITIAN

Proses pengenalan isyarat tutur digambarkan pada Gambar 1, yang terdiri dari; tahap perekaman isyarat tutur secara langsung, pemrosesan awal, ekstraksi ciri, dan pengenalan pola. Tiap tahapan sangat penting dalam mengoptimalkan hasil pengenalan pola isyarat tutur.



Gambar 1. Proses pengenalan isyarat tutur

### A. Perekaman Isyarat Tutur

Pengumpulan data isyarat tutur dilakukan melalui proses perekaman secara langsung yang dapat tersimpan dengan ekstensi .wav. Keuntungan dari format.wav yaitu dapat dikenali pada software aplikasi Matlab dengan frekuensi *sampling* yang bervariasi antara 8000 Hz sampai 48000Hz tergantung pada spesifikasi *soundcard* computer yang mempengaruhi kecepatan *sampling* [14].

Frekuensi *sampling* (fs) merupakan nilai yang memenuhi kriteria Nyquist, pada persamaan (1).

$$fs \geq 2f, \quad (1)$$

$f_s$  = frekuensi sampling (diskrit) (Hz)  
 $f$  = frekuensi isyarat (analog) (Hz)

## B. Pemrosesan Awal

Pada pemrosesan awal dilakukan untuk normalisasi isyarat tutur dengan tiga tahap yang meliputi DC removal, normalisasi amplitudo dan menghilangkan isyarat diam.

- DC removal

DC Removal dilakukan dengan menghilangkan komponen DC dari isyarat tutur menjadi 0 (nol). Pada persamaan (2) cara menghilangkan komponen DC dilakukan dengan mengurangkan tiap nilai pada isyarat  $S(n)$  dengan hasil rerata isyarat itu sendiri (rerata  $S(n)$ ).

$$DC_{offset}(n) = S(n) - \frac{\sum_{n=1}^N S(n)}{N} \quad (2)$$

$D_{offset}(n)$  = runtun isyarat keluaran  
 $S(n)$  = runtun isyarat masukan  
 $n$  = urutan runtun  
 $N$  = merupakan panjang runtun isyarat

- Normalisasi amplitudo

Normalisasi amplitudo dilakukan untuk mengatasi tingkat energy yang tidak konsisten antara tiap isyarat. Sehingga kualitas ciri dapat ditingkatkan dan semua data memiliki standar pengukuran yang sama. Proses normalisasi amplitudo diperoleh dengan membagi setiap nilai  $S(n)$  pada runtun ke-n dengan nilai absolut amplitudo tertinggi yang terdapat pada isyarat dengan nilai batasan maksimal antara -1 dan 1, dirumuskan pada persamaan (3):

$$S_{nor}(n) = \frac{DC_{offset}(n)}{\max(\text{abs}(DC_{offset}(n)))} \quad (3)$$

$S_{nor}(n)$  = runtun isyarat keluaran  
 $D_{offset}(n)$  = runtun isyarat masukan  
 $n$  = urutan runtun

- Menghilangkan isyarat diam

Proses menghilangkan isyarat diam bertujuan untuk mengefektifkan komputasi pada segmentasi karena derau dan isyarat diam bukan merupakan informasi yang dibutuhkan dalam pengolahan isyarat tutur.

Proses segmentasi dilakukan dengan membagi-bagi isyarat dalam *frame* dengan durasi tertentu. *Frame* yang tidak mengandung isyarat tutur akan diobservasi dan dieleminasi untuk menentukan nilai ambang. Jika isyarat  $S(n)$  pada runtun ke-n dan lebar *frame* adalah  $N$  runtun, maka *frame* ke-i dapat dituliskan dalam persamaan (4).

$$F_i = (S(n))_{n=(i-1)*N+1}^{i*N} \quad (4)$$

$F_i$  = *frame* pada indeks ke-i  
 $N$  = lebar *frame* (160 runtun)  
 $i$  = nomor/ indeks *frame*  
 $S(n)$  = nilai isyarat pada runtun ke-n

## C. Ekstrasi ciri menggunakan metode DWT

DWT digunakan untuk mentransformasikan isyarat dari domain waktu ke domain frekuensi yang dapat diaplikasikan pada data diskrit untuk menghasilkan keluaran diskrit [7]. DWT dikatakan sebagai *Low Pass Filter* (LPF) dan *High Pass Filter* (HPF). Frekuensi rendah dan frekuensi tinggi dipisahkan dari sinyal asli dengan menggunakan transformasi dekomposisi. Semakin rendah pendekatan sinyal frekuensi maka semakin tinggi sinyal frekuensi yang dihasilkan [15].

*Low Pass Filter* (LPF) maupun *High Pass Filter* (HPF) merupakan salah satu fungsi yang paling banyak digunakan pada pemrosesan sinyal. Perwujudan *wavelet* dapat berupa penskalaan ulang dengan iterasi. Resolusi sinyal diukur dari jumlah informasi sinyal ditentukan oleh operasi *filtering* dan menggunakan skala operasi *upsampling* dan *downsampling* [16][17]. Perhitungan DWT dapat dilakukan dengan menkonvolusi koefisien LPF ( $h$ ) dan HPF ( $g$ ) [7] yang ditunjukkan pada persamaan (5) dan (6).

$$a_k^{(j+1)} = \sum_{n=-\infty}^{\infty} h_{n-2k} a_n^{(j)} = (a^{(j)} * h^{(0)})(2k) \quad (5)$$

$$d_k^{(j+1)} = \sum_{n=-\infty}^{\infty} g_{n-2k} a_n^{(j)} = (a^{(j)} * g^{(1)})(2k) \quad (6)$$

## D. Pengenalan pola menggunakan metode DTW

*Dynamic Time Warping* (DTW) merupakan algoritme berfungsi untuk mencari jarak antara dua isyarat dengan meminimalkan fungsi biaya. Namun, dengan elemen tertentu ada beberapa hambatan pada urutan poin  $W = \{ w_1, w_2, \dots, w_k \}$  sebagai berikut :

- *Boundary*;  $w_1 = (1,1)$  dan  $w_k = (N, M)$ . Dimulai dari titik (1,1) dan berakhir pada titik (N, M), jika dalam matriks maka berawal dari posisi kiri atas dan berakhir pada posisi kanan bawah.
- *Monotonicity*;  $i_{p-1} \leq i_p \leq i_{p+1}$  dan  $j_{p-1} \leq j_p \leq j_{p+1}$ ,  $\forall p \in [1, k]$ . Kondisi dimana tidak ada pengulangan jalur pada ciri isyarat yang sama untuk mempertahankan waktu pemesanan,, konsekuensi, dan sebab akibatnya
- *Step size*;  $i_p - i_{p-1} \leq 1$  dan  $j_p - j_{p-1} \leq 1$ ,  $\forall p \in [1, k]$ . untuk membatasi warping yang satu dengan yang lain.

Dari ketiga hambatan tersebut didapatkan persamaan (7) yang berasal dari kondisi *monotonicity* dan *step size* sebagai berikut :

$$c(k-1) = \begin{cases} (i(k), j(k-1)), \\ (i(k)-1, j(k)-1), \\ \text{or } (i(k)-1, j(k)). \end{cases} \quad (7)$$

Untuk menentukan kesamaan atau perbedaan antara dua isyarat tutur yang dibandingkan tanpa proses pelatihan terlebih dahulu dengan menggunakan diskriminasi jarak pada metode DTW. Nilai jarak dan isyarat yang dinormalisasi merupakan keluaran algoritme DWT. Dalam penelitian ini yang digunakan adalah nilai jarak DTW saja dan dikembangkan secara *realtime*. Hasil pengukuran diperoleh dari jarak minimum yang digunakan dalam pengenalan pola untuk mengambil keputusan seperti pada persamaan (8).

$$dwt_{(x,x)} = \begin{cases} 1 & \text{jika } dwt(x,x) < d(x,y) \\ 0 & \text{jika } dwt(x,x) \geq d(x,y) \end{cases} \quad (8)$$

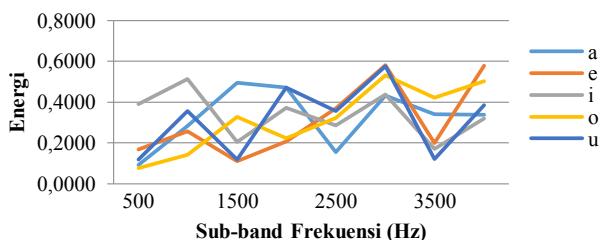
Adapun rumus perhitungan akurasi pengenalan ditunjukkan pada persamaan (9) berupa persentase pengenalan terbaik [18].

$$\% \text{ Pengenalan} = \frac{\Sigma \text{ keluaran yang dikenali}}{\Sigma \text{ total keluaran yang diujji}} \times 100 \% \quad (9)$$

Pengenalan isyarat tutur dikembangkan dengan berbagai macam metode secara konvensional maupun *realtime*. Akurasi pengenalan dengan menggunakan metode secara *realtime* yang didapatkan cukup tinggi dan bervariasi. Pencapaian persentase rata-rata pengenalan terbaik menurut [19] sebesar 88%, [20] sebesar 77,2 %, dan [18] sebesar 86.2%.

### III. HASIL DAN PEMBAHASAN

Penentuan metode pengenalan terbaik berdasarkan metode ekstrasi ciri *Discrete Wavelet Transform* (DWT) level 3 menggunakan fungsi jarak *Euclidean* dan *Manhattan* menghasilkan pengenalan terbaik dengan persentase 80,82% dan 83,27%.

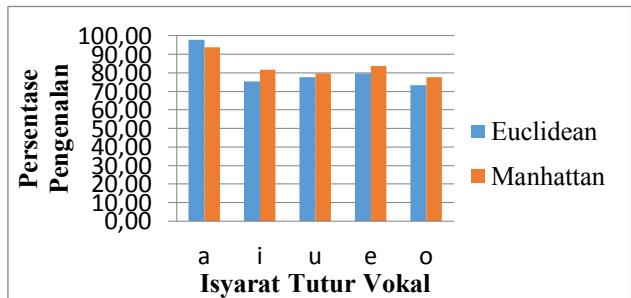


Gambar 2. Hasil vektor ciri isyarat vokal menggunakan metode ekstrasi ciri *Discrete Wavelet Transform* (DWT) level 3.

Pengujian dilakukan pada 7 penutur yang berbeda tanpa dilakukan pelatihan terlebih dahulu secara bergantian sebagai data referensi. Vektor ciri isyarat vokal yang dihasilkan oleh penutur pertama menggunakan metode ekstrasi ciri *Discrete Wavelet Transform* (DWT) level 3 ditunjukkan pada gambar 2.

Berdasarkan pengujian yang dilakukan dihasilkan persentase rata-rata tiap vokal bahasa Indonesia yang

ditunjukkan gambar 3.



Gambar 3. Persentase rata-rata tiap vokal bahasa Indonesia

### IV. KESIMPULAN

Ekstrasi ciri isyarat tutur berupa suara vokal dengan menggunakan metode *Discrete Wavelet Transform* (DWT) level 3 terdapat 8 ciri suara vokal sebanyak 400 pasang data. Kemudian menentukan frekuensi sampling (fs) sebesar 8000Hz sehingga frekuensi yang digunakan adalah 500 Hz – 4000 Hz dengan lebar frekuensi yang sama.

Setelah dilakukan ekstrasi ciri isyarat tutur berupa suara vokal dengan menggunakan metode *Discrete Wavelet Transform* (DWT) level 3 maka diperoleh rata-rata pengenalan terbaik sebesar 80,82% menggunakan fungsi jarak *Euclidean* dan 83,27% menggunakan fungsi jarak *manhattan*.

### DAFTAR PUSTAKA

- [1] H. Sakoe and S. Chiba, “Dynamic Programming Algorithm Optimization for Spoken Word Recognition,” in *IEEE Transactions on Acoustic Speech and Signal Processing*, 1978, vol. 26, pp. 43–49.
- [2] L. O. B. R.C. Guido, J.F.W. Slaets, R. Koberle and J. C. P. Almeida, “New Technique to construct a wavelet transform matching a specified signal with applications to digital, real-time, spike and overlap pattern recognition,” in *Digital Signal Processing*, Elsevier, v. 16, n. 1, 2006, p. 24–44.,
- [3] A. N. R. Adipranata, “Implementasi Sistem Pengenalan Suara Menggunakan SAPI 5.1 dan Delphi 5,” vol. 4, pp. 107–114, 2003.
- [4] Z. Yuxin and Y. Miyanaga, “An improved dynamic time warping algorithm employing nonlinear median filtering,” *2011 11th Int. Symp. Commun. Inf. Technol. Isc.*, pp. 439–442, 2011.
- [5] D. S. and C. A. Ratanamahatana, “Efficient Time Series Classification under Template Matching using Time Warping Alignment,” *IEEE Int. Conf. Comput. Sci. Converg. Inf. Technol.*, no. 2009, pp. 685–690, 2009.
- [6] A. Bala, “Voice Command Recognition System Based On MFCC and DTW,” *Int. J. Eng. Sci. Technol.*, vol. 2 (12), pp. 7335–7342, 2010.
- [7] A. Asni, “Ekstraksi Ciri Dan Pengenalan Tutur

- [8] Vokal Bahasa Indonesia Menggunakan Metode Discrete Wavelet Transform (DWT) dan Dynamic Time Warping (DTW),” in *Universitas Gadjah Mada*, 2014.
- [9] U. Sutisna, “Pengenalan Tutur Kata Terisolasi Menggunakan MFCC dan ANFIS 2013.,” Universitas Gadjah Mada, 2013.
- [10] N. S. Yaniar, “Perbandingan Ukuran Jarak pada Proses Pengenalan Wajah Berbasis Principal Component Analysis ( PCA ),” *Proceeding Semin. Tugas Akhir Jur. Tek. Elektro FTI- ITS*, pp. 1–6, 2011.
- [11] D. Nugraheny, “Metode Nilai Jarak Guna Kesamaan Atau Kemiripan Ciri Suatu Citra ( Kasus Deteksi Awan Cumulonimbus Menggunakan Principal Component Analysis ),” *Angkasa*, vol. 7, pp. 21–30, 2015.
- [12] H. M. El-Bakry, M. A. Abo-Elsoud, and M. S. Kamel, “Integrating Fourier descriptors and PCA with neural networks for face recognition,” in *National Radio Science Conference, NRSC, Proceedings*, 2000, p. C22.1-C22.8.
- [13] D. R. Wilson and T. R. Martinez, “Improved heterogeneous distance functions,” *J. Artif. Intell. Res.*, vol. 6, pp. 1–34, 1997.
- [14] M. D. Malkauthekar, “Analysis of euclidean distance and manhattan distance measure in face recognition,” *Third Int. Conf. Comput. Intell. Inf. Technol.*, vol. 1, no. 4, pp. 3–7, 2013.
- [15] I. MacLoughlin, “Applied Speech and Audio Processing With Matlab Examples,” in *New York: Cambridge University Press*, 2009, p. 2002.
- [16] K. R. Ghule and R. R. Deshmukh, “Feature Extraction Techniques for Speech Recognition: A Review,” *Int. J. Sci. Eng. Res.*, vol. 6, no. 5, pp. 2229–5518.
- [17] V. K. V R and B. A. P, “Features of Wavelet Packet Decomposition and Discrete Wavelet Transform for Malayalam Speech Recognition,” *Aceee*, vol. 1, no. 2, pp. 93–96, 2009.
- [18] H. Ali, N. Ahmad, X. Zhou, & K. Iqbal, and S. M. Ali, “DWT features performance analysis for automatic speech recognition of Urdu.,” *Springerplus*, vol. 3, pp. 1–10, 2014.
- [19] M. K. . G. and R. G. . Meegama, “Real time Translation of Discrete Sinhala Speech to Unicode Text Real-time Translation of Discrete Sinhala Speech to Unicode Text,” in *IEEE ICter*, 2015, no. AUGUST.
- [20] S. P. Nandyala and D. T. K. Kumar, “Real Time Isolated Word Speech Recognition System for Human Computer Interaction,” *Int. J. Comput. Appl.*, vol. 12, no. 2, pp. 1–7, 2010.
- [21] S. Hidayat, “Sistem pengenal tutur bahasa indonesia berbasis suku kata menggunakan mfcc, wavelet dan hmm,” *CITEE*, no. SEPTEMBER, 2015.