EKSTRAKSI CIRI DAN PENGENALAN TUTUR VOKAL BAHASA INDONESIA MENGGUNAKAN METODE *DISCRETE WAVELET TRANSFORM* (DWT) DAN *DYNAMIC TIME WARPING* (DTW) SECARA *REALTIME*

Risky Via Yuliantari*, Risanuri Hidayat , Oyas Wahyunggoro
Jurusan Teknik Elektro dan Teknik Informasi, Fakultas Teknik,
Universitas Gadjah Mada
Jln. Grafika 2 Yogyakarta 55281
Email:* rviayuliantari@gmail.com

Abstrak

Pada paper ini disajikan tentang pengembangan sebuah sistem pengenalan isyarat vokal bahasa Indonesia secara realtime. Pada pengenalan isyarat vokal bahasa Indonesia secara konvensional memberikan akurasi yang tinggi. Oleh karena itu, pada penelitian ini dilakukan pengembangan proses pengenalan dilakukan secara realtime yang diterapkan pada vokal bahasa Indonesia. Metode ekstrasi ciri yang digunakan adalah Discrete Wavelet Transform (DWT) level 3 dan Dynamic Time Wraping (DTW) sebagai metode pengenalan isyarat vokal bahasa Indonesia. Pada metode ekstrasi ciri Discrete Wavelet Transform (DWT) level 3 didapatkan 8 buah ciri. Sedangkan metode pengenalan menggunakan Dynamic Time Wraping (DTW) dilakukan dengan menghitung diskriminasi jarak terkecil dan tanpa adanya pelatihan terlebih dahulu. Hasil pengenalan menggunakan metode DWT level 3 menunjukkan akurasi terbaik sebesar 80 %. Dari hasil pengenalan tersebut dilakukan pengujian terhadap 5 penutur yang berbeda secara bergantian sebagai data referensi, sehingga diperoleh 500 pasang data pengukuran. Hasil persentase rata-rata pengenalan dengan akurasi terbaik dari pengujian 5 penutur yang berbeda mencapai 87,2% dari 500 pasang data yang diperoleh secara realtime.

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

1. PENDAHULUAN

Sistem identifikasi satu vokal pengenalan isyarat tutur dengan menggunakan algoritme pembelajaran pada mesin pengenalan, belum mampu memberikan pengenalan selayaknya 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 Wraping* (DTW) sebagai metode pengenalan isyarat tutur digunakan untuk mengoptimalkan hasil pengenalan suara tanpa harus mengurangi komputasi (Sakoe & Chiba 1978). Sedangkan *Discrate Wavelet Transform* (DWT) sebagai metode ekstrasi ciri digunakan untuk mengatasi isyarat tutur yang mengandung derau dan sinyal yang tidak stasioner, serta mengurangi panjang sinyal input (R.C. Guido, J.F.W. Slaets, R. Koberle & Almeida 2006).

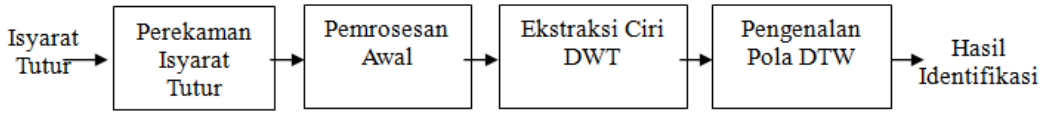
Banyak penelitian yang dilakukan untuk meningkatkan kemampuan pengenalan tutur. Sehingga sistem pengenalan tutur menggunakan aplikasi komputer bukan merukapan hal yang baru (R. Adipranata 2003). Beberapa contoh penerapan pengenalan isyarat tutur diantaranya tentang penambahan filter median pada metode DTW untuk menyamai tingkat akurasi pengenalan pola *Hidden Markov Model* (HMM) (Yuxin & Miyanaga 2011). Penerapan algoritma *Shape Averaging* (SA) pada DTW untuk peningkatan akurasi pengenalan (Ratanamahatana 2009). Penerapan ekstrasi ciri *Mel Frequency Cepstral Coefficient* (MFCC) untuk pengenalan kata terisolasi angka menggunakan bahasa Inggris (Bala 2010). Pengenalan isyarat tutur bahasa Indonesia secara konvensional untuk vokal (Asni 2014) dan beberapa kata terisolasi menggunakan *Mel Frequency Cepstral Coefficient* (MFCC) secara otomatis (Sutisna 2013).

Pada penelitian ini disajikan tentang pengembangan sebuah sistem pengenalan isyarat vokal bahasa Indonesia secara *realtime.* Pada pengenalan isyarat vokal bahasa Indonesia secara konvensional memberikan akurasi yang tinggi. Oleh karena itu, pada penelitian ini dilakukan pengembangan proses pengenalan dilakukan secara *realtime* yang diterapkan pada vokal bahasa

Indonesia menggunakan metode *Discrete Wavelet Transform* (DWT) level 3 sebagai metode ekstrasi ciri dan metode *Dynamic Time Wraping* (DTW) sebagai metode pengenalan isyarat vokal bahasa Indonesia. Pada metode ekstrasi ciri *Discrete Wavelet Transform (*DWT) level 3 didapatkan 8 buah ciri. Sedangkan metode pengenalan menggunakan *Dynamic Time Wraping* (DTW) dilakukan dengan menghitung diskriminasi jarak terkecil dan tanpa adanya pelatihan terlebih dahulu.

2. METODOLOGI

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

2.1. Perekaman Isyarat Tutur

Pengumpulan data isyarat tutur dilakukan melalui proses perekaman secara langsung yang dapat tersimpan dengan ektensi *.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* (I. MacLoughlin 2009).

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

$$fs \geq 2f \ , \tag{1}$$

fs = frekuensi *sampling* (diskrit) (Hz)
f = frekuensi isyarat (analog) (Hz)

2.2. Pemrosesan Awal

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

2.2.1. 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)).

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

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

ISBN 978-602-99334-5-1

2.2.2. Normalisasi amplitudo

Normalisasi amplitudo dilakukan untuk mengatasi tingkat energy yang tidak konsisten antara tiap isyarat.Sehingga kualitas ciri dapat ditingkatkan dan semua data memiliki standard pengukuran yang sama.

Proses normaliasi amplitude 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):

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

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

2.2.3. 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).

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

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

2.3. 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 (Asni 2014). 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 (Ghule & Deshmukh n.d.).

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* (R & P 2009)(Ali et al. 2014). Perhitungan DWT dapat dilakukan dengan menkonvolusi koefisien LPF (h) dan HPF (g) (Asni 2014) yang ditunjukan pada persamaan (5) dan (6).

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

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

2.4. 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 = { $w_1, w_2, \ldots, w_k$ } sebagai berikut :

1. *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 posisis kiri atas dan berakhir pada posisi kanan bawah.
2. *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
3. *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 :

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

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 dingunakan 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).

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

Adapun rumus perhitungan akurasi pengenalan ditunjukkan pada persamaan (9) berupa persentase pengenalan terbaik (Meegama 2015).

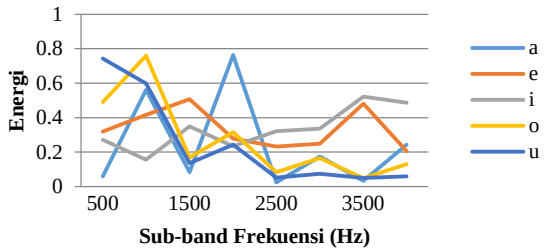
$$\% \text{ Pengenalan} = \frac{\Sigma \text{ keluaran yang dikenali}}{\Sigma \text{ total keluaran yang diuji}} \text{ x } 100 \% \quad (9)$$

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 (Nandyala & Kumar 2010) sebesar 88%, (Hidayat 2015) sebesar 77,2 %, dan (Meegama 2015) sebesar 86.2%.

3. HASIL DAN PEMBAHASAN

Penentuan metode pengenalan terbaik berdasarakan metode ekstrasi ciri *Discrete Wavelet Transform* (DWT) level 3 secara *realtime* menghasilkan 20 dari 25 pengenalan terbaik dengan persentase 80%. Vektor ciri isyarat vokal yang dihasilkan menggunakan metode ekstrasi ciri *Discrete Wavelet Transform* (DWT) level 3 ditunjukkan pada gambar 2.

ISBN 978-602-99334-5-1



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

Berdasarkan metode pengenalan terbaik maka dilakukan pengujian pada 5 penutur yang berbeda tanpa dilakukan pelatihan terlebih dahulu secara bergantian sebagai data referensi Hasil persentase rata-rata pengenalan yang diperoleh secara *realtime* sebesar 87,2% dari 500 pasang data terbaik.

Tabel 1 Perbandingan persentase pengenalan DTW dengan sumber referensi penutur yang berbeda

No	Nama penutur sebagai referensi	Pengenalan (%)
1	Ari	92%
2	Didi	84%
3	Novi	80%
4	Lia	92%
5	Yuli	88%
Rata- rata pengenalan (%)		87,2%

4. KESIMPULAN

Setelah dilakukan ekstrasi ciri isyarat vokal dengan menggunakan metode *Discrete Wavelet Transform* (DWT) level 3 secara *realtime* maka diperoleh rata-rata pengenalan terbaik sebesar 80% dengan jumlah pengukuran 225 pasang vektor ciri. Kemudian dilakukan pengujian pada 5 penutur yang berbeda tanpa dilakukan pelatihan terlebih dahulu secara bergantian sebagai data referensi sehingga menghasilkan rata-rata pengenalan terbaik sebesar 87,2% dengan jumlah pengukuran 500 pasang vektor ciri. Analisis dapat dikembangkan dan diujikan untuk data yang lebih besar.

DAFTAR PUSTAKA

Ali, H. et al., 2014. DWT features performance analysis for automatic speech recognition of Urdu. *SpringerPlus*, 3, pp.1–10.

Asni, A., 2014. Ekstraksi Ciri Dan Pengenalan Tutur Vokal Bahasa Indonesia Menggunakan Metode Discrete Wavelet Transform (DWT) dan Dynamic Time Warping (DTW). In *Universitas Gadjah Mada*.

Bala, A., 2010. VOICE COMMAND RECOGNITION SYSTEM BASED ON MFCC AND DTW. *International Journal of Engineering Science and Technology*, 2 (12), pp.7335–7342. Available at: http://www.waset.org/publications/4967.

Ghule, K.R. & Deshmukh, R.R., Feature Extraction Techniques for Speech Recognition: A Review. *International Journal of Scientific & Engineering Research*, 6(5), pp.2229–5518.

Hidayat, S., 2015. Sistem pengenal tutur bahasa indonesia berbasis suku kata menggunakan mfcc, wavelet dan hmm. *CITEE*, (SEPTEMBER).

I. MacLoughlin, 2009. Applied Speech and Audio Processing With Matlab Examples. In *New York: Cambridge University Press*. p. 2002.

Meegama, M.K.. G. and R.G.., 2015. Real time Translation of Discrete Sinhala Speech to Unicode

Text Real-time Translation of Discrete Sinhala Speech to Unicode Text. In *IEEE ICter*.

Nandyala, S.P. & Kumar, D.T.K., 2010. Real Time Isolated Word Speech Recognition System for Human Computer Interaction. *International Journal of Computer Applications*, 12(2), pp.1–7.

R, V.K. V & P, B.A., 2009. Features of Wavelet Packet Decomposition and Discrete Wavelet Transform for Malayalam Speech Recognition. *Aceee*, 1(2), pp.93–96.

R. Adipranata, A.N., 2003. Implementasi Sistem Pengenalan Suara Menggunakan SAPI 5.1 dan Delphi 5. , 4, pp.107–114.

R.C. Guido, J.F.W. Slaets, R. Koberle, L.O.B. & Almeida, J.C.P., 2006. New Technique to construct a wavelet transformmatching a specified signal with applications to digital, real-time, spike and overlap pattern recognition. In *Digital Signal Processing, Elsevier, v. 16, n. 1*. pp. 24–44,.

Ratanamahatana, D.S. and C.A., 2009. Efficient Time Series Classification under Template Matching using Time Warping Alignment. *IEEE Int. Conf. Comput. Sci. Converg. Inf. Technol.*, (2009), pp.685–690.

Sakoe, H. & Chiba, S., 1978. Dynamic Programming Algorithm Optimization for Spoken Word Recognition. In *IEEE Transactions on Acoustic Speech and Signal Processing*. pp. 43–49.

Sutisna, U., 2013. *Pengenalan Tutur Kata Terisolasi Menggunakan MFCC dan ANFIS 2013.* Universitas Gadjah Mada.

Yuxin, Z. & Miyanaga, Y., 2011. An improved dynamic time warping algorithm employing nonlinear median filtering. *2011 11th International Symposium on Communications Information Technologies ISCIT*, pp.439–442.

ISBN 978-602-99334-5-1