EKSTRASI CIRI DAN PENGENALAN SUARA VOKAL BAHASA INDONESIA BERDASARKAN JENIS KELAMIN SECARA REAL TIME

Risky Via Yuliantari[1*], Risanuri Hidayat[2] , Oyas Wahyunggoro[3]
[1,2,3]Jurusan Teknik Elektro dan Teknik Informasi, Fakultas Teknik, Universitas Gadjah Mada
Jl. Grafika No.2, Daerah Istimewa Yogyakarta 55281
Email: rviayuliantari@gmail.com

Abstrak

Suara manusia memiliki ciri yang beraneka ragam, sehingga dapat dijadikan media komunikasi yang efektif. Oleh karena itu banyak penelitian yang berkaitan dengan suara manusia dilakukan untuk meningkatkan pengenalan suara. Proses pengembangan pengenalan suara dilakukan secara realtime berdasarkan jenis kelamin untuk menghasilkan akurasi yang tepat dalam batas waktu yang telah ditentukan. Metode Discrete Wavelet Transform (DWT) level 3 dan Dynamic Time Wraping (DTW) digunakan sebagai metode ekstrasi ciri dan metode pengenalan suara. Pada metode ekstrasi ciri Discrete Wavelet Transform (DWT) level 3 didapatkan 8 buah ciri. Sedangkan metode pengenalan suara menggunakan Dynamic Time Wraping (DTW) dilakukan dengan menghitung diskriminasi jarak terkecil antara dua ciri yang berbeda tanpa dilakukan pelatihan terlebih dahulu. Pengenalan suara diujikan pada 6 orang penutur pria dan 6 orang penutur wanita secara bergantian dengan masing-masing data pengukuran 900 pasang. Hasil persentase rata-rata pengenalan akurasi terbaik mencapai 54,6% dari pengujian terhadap 6 orang penutur pria secara bergantian dan 54,17 % dari pengujian terhadap 6 orang penutur wanita secara bergantian dari masing-masing pasangan 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.

Isyarat tutur yang berupa suara manusia memiliki ciri yang beraneka ragam dapat menyampaikan sebuah informasi berupa sinyal yang diterima telinga manusia (Gultom et al. n.d.). 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).

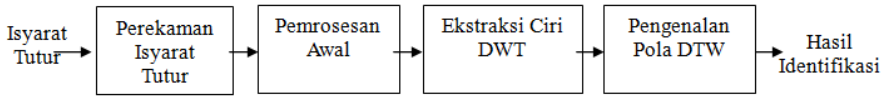
Banyak penelitian yang dilakukan untuk meningkatkan kemampuan pengenalan isyarat tutur menggunakan aplikasi komputer maka dikatakan bukan merupakan hal yang baru (R. Adipranata 2003). Dengan demikian diharapkan adanya interaksi cerdas yang efektif antara manusia dan komputer (Kumar & Rangababu 2015). 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 pengenalan isyarat vokal bahasa Indonesia secara konvensional memberikan akurasi yang tinggi dan bervariasi. Oleh karena itu, pengembangan sebuah sistem pengenalan suara vokal

bahasa Indonesia secara realtime berdasarkan jenis kelamin dilakukan dengan 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 untuk menghasilkan akurasi yang tepat dalam batas waktu yang telah ditentukan.

2. METODOLOGI

Proses pengenalan isyarat tutur digambarkan pada Gambar 1, yang terdiri dari; tahap perekaman isyarat tutur berupa suara vocal pria dan wanita dengan batas waktu yang ditentukan, 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 berupa suara vokal yang dilakukan melalui proses perekaman dengan batas waktu tertentu 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 48000 Hz tergantung pada spesifikasi soundcard komputer yang mempengaruhi kecepatan sampling (I. MacLoughlin 2009).

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

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

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

Perekaman isyarat tutur berupa suara vokal “a”, “i”, “u”, “e”, “o” dilakukan secara *realtime* yang merupakan penggunaan perangkat keras berupa komputer dan perangkat lunak berupa *software* dengan batasan waktu yang telah ditentukan (Wikipedia n.d.)

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

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):

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

$S_{nor}(n)$ = 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).

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

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).

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

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

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; ip-1 ≤ ip ≤ ip+1 dan jp-1 ≤ jp ≤ jp+1 , ∀p∈ [1,k]. Kondisi dimana tidak ada pengulangan jalur pada ciri isyarat yang sama untuk mempertahankan waktu pemesanan,, konsekuensi, dan sebab akibatnya*
3. *Step size; ip- ip-1 ≤ 1 dan jp- jp-1 ≤ 1 , ∀p∈ [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* untuk menghasilkan akurasi yang tepat dalam batas waktu yang telah ditentukan. Hasil pengukuran diperoleh dari jarak minimum yang digunakan dalam pengenalan pola untuk mengambil keputusan seperti pada persamaan (8).

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

Pengenalan isyarat tutur dikembangkan dengan berbagai macam metode secara konvensional maupun *realtime* dengan tingkat akurasi yang tinggi dan bervariasi. Adapun rumus perhitungan akurasi pengenalan ditunjukkan pada persamaan (9) berupa persentase pengenalan terbaik (Meegama 2015).

$$\%\,\textbf{Pengenalan} = \frac{\Sigma\,\textbf{keluaran yang dikenali}}{\Sigma\,\textbf{total keluaran yang diuji}}\ \textbf{x}\ 10 \quad (9)$$

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 tutur berupa suara vokal yang dihasilkan menggunakan metode ekstrasi ciri *Discrete Wavelet Transform* (DWT) level 3 ditunjukkan pada gambar 2.

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

Hasil persentase rata-rata pengenalan yang diperoleh secara *realtime* sebesar 54,6% dari pengujian terhadap 6 orang penutur pria secara bergantian ditunjukkan pada tabel 1.

Tabel 1. Hasil persentase pengenalan DTW dengan sumber referensi penutur pria secara bergantian

No	Nama penutur sebagai referensi	Pengenalan (%)
1	Ari	58,3
2	Aris	47,2
3	Cipto	52,8
4	Didi	63,9
5	Jeki	50,0
6	Novian	55,6
Rata- rata pengenalan (%)		54,6

Hasil persentase rata-rata pengenalan yang diperoleh secara *realtime* sebesar 54,2% dari pengujian terhadap 6 orang penutur wanita secara bergantian ditunjukkan pada table 2.

Tabel 2. Hasil persentase pengenalan DTW dengan sumber referensi penutur wanita secara bergantian

No	Nama penutur sebagai referensi	Pengenalan (%)
1	Aya	47,2
2	Sofi	47,2
3	Tia	55,6
4	Yanti	58,3
5	Yuan	55,6
6	Yuli	61,1
Rata- rata pengenalan (%)		54,2

Dari hasil persentase rata-rata pengenalan yang diperoleh secara *realtime* antara penutur pria dan wanita terdapat selisih sebesar 0,4 %.

4. KESIMPULAN

Ekstrasi ciri isyarat tutur berupa suara vokal dengan menggunakan metode *Discrete Wavelet Transform* (DWT) level 3 terdapat 8 ciri suara vokal sebanyak 900 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 secara realtime dengan waktu yang telah ditentukan maka diperoleh rata-rata pengenalan terbaik sebesar 54,6% dan 54,2% dari pengujian terhadap 6 orang penutur pria dan 6 orang penutur wanita.

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.

Gultom, M., Alamsyah, D. & Teori, L., Rancang Bangun Aplikasi Pengenal Penutur menggunakan Metode Hidden Markov Model (HMM). , pp.1–9.

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

Kumar, S.S. & Rangababu, T., 2015. Emotion and Gender Recognition of Speech Signals Using SVM. International Journal of Engineering Science and Innovative Technology (IJESIT), 4(3), pp.128–137.

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.

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.

Wikipedia, Komputasi Waktu Nyata. Available at: https://id.wikipedia.org/wiki/Komputasi_waktu_nyata [Accessed July 20, 2016].

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.