

Pengenalan Identitas Penutur Menggunakan Algoritma *Discrete Wavelet Transform* (DWT) dan *Hidden Markov Models* (HMM)

A.Asni B¹. Diah Patriana Setianingsih²

a_asni_b@uniba-bpn.ac.id

^{1,2}Teknik Elektro, Fakultas Teknologi Industri Universitas Balikpapan

Jln. Pupuk Raya Gn. Bahagia Balikpapan 76114 INDONESIA

ABSTRACT

This study aims to build a speaker recognition system using the Discrete Wavelet Transform (DWT) algorithm and Hidden Markov Models (HMM). Speech signals from each speaker were recorded using Indonesian words "kiri" and recorded 10 times. Five (5) data from the first record were recorded under normal conditions and the next 5 data were sourced from the nasal sounds produced by a pressed nose. The total data from 6 different speakers becomes 60 data

The results of the application of the Discrete Wavelet Transform (DWT) algorithm and the Hidden Markov Models (HMM) algorithm and the number of states tested 4 to 7 states, in this study have not provided optimal results. The identification of error rates is quite high, which is equal to 20% for the number of states 4 and 5, and reaches 30% for the number of states 6 and 7. This shows that the feature vector values generated from the DWT algorithm and then modeled and tested using the HMM algorithm has not optimal results yet. Further evaluation is needed to examine the opportunities of other algorithms that can be applied in DWT and to achieve high recognition accuracy

Keyword — *Speaker Recognition, Hidden Markov Model, Discrete Wavelet Transform.*

I. PENDAHULUAN

Pengenalan penutur (*speaker recognition*) adalah salah satu bidang pengenalan pola yang berkaitan dengan pemrosesan sinyal suara. Berbeda dengan *speech recognition* yang mengenali kata atau kalimat yang diucapkan, *speaker recognition* mengenali siapa penutur yang mengucapkan kata tersebut. Pada dasarnya setiap manusia memiliki sesuatu yang unik yang hanya dimiliki oleh dirinya sendiri.[1]

Metode DWT dengan menghitung energy tiap sub-band hingga level-3 menggunakan Wavelet Daubechies2 hingga diperoleh 8 vektor ciri. Pengenalan pola untuk verifikasi menggunakan

metode *Dynamic Time Warping* (DTW). Hasil yang dicapai dalam pengenalan isyarat tutur vocal Bahasa Indonesia mencapai hingga 100%[2].

Penelitian tentang pengenalan penutur dan tutur telah dilakukan oleh para peneliti[3], pengenalan pola *Hidden Markov Models* (HMM) memberikan tingkat akurasi yang lebih baik dibandingkan metode *Dynamic Time Warping* DTW yang diterapkan untuk pengenalan tutur dengan menggunakan data vector ciri dari hasil ekstraksi ciri yang menggunakan metode *Mel-Frekuensi Cepstral coefficients* (MFCC).

Penelitian yang lain menggunakan metode *Mel Frequency Cepstrum Coefficients* (MFCC) untuk ekstraksi fitur dan metode *Hidden Markov Model* (HMM) untuk mengenali penutur. Penelitian tersebut membangun aplikasi pengenalan penutur dan menerapkan metode *Hidden Markov Model* pada aplikasi pengenalan penutur. Hasil penelitian menunjukkan bahwa pengujian secara real time menggunakan mikrofon tingkat akurasinya 30%. Sedangkan pengujian dari file rekaman 100%[4].

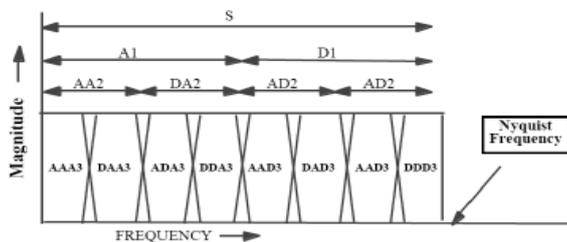
II. DISCRETE WAVELET TRANSFORM (DWT)

DWT diaplikasikan ke dalam data diskrit untuk menghasilkan keluaran diskrit yang selanjutnya 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 down sampling.

Ekstraksi ciri menggunakan metode DWT, menerapkan metode *full binary DWT* level 3 persamaan-persamaan yang digunakan dalam proses ekstraksi ciri sebagaimana telah dilakukan dalam penelitian sebelumnya[1]. Namun dalam penerapan

metode pengenalan HMM sebagai pengenalan pola identitas penutur, menggunakan vector ciri dari nilai energi yang dihitung dari masing-masing *sub-band* tanpa perlu dinormaslisasi. Tiap isyarat tutur dipecah menjadi 2, dengan asumsi untuk memisahkan suku kata “ki” dan “ri” pada kata “kiri” yang direkam sehingga diperoleh 16 ciri dari masing-masing data rekaman isyarat tutur .

Proses pembentukan vektor ciri dengan menghitung energi dari hasil rekonstruksi nilai-nilai koefisien pada masing-masing frekuensi *sub-band*. Frekuensi *sub-band* pada dekomposisi *full binary* memiliki lebar yang sama pada masing-masing level dekomposisi ke-*j*, seperti pada pada Gambar 1[5].



Gambar 1 Alokasi Frekuensi pada Full Binary Paket Wavelet Level-3[5].

Total Energi yang ada pada semua frekuensi *sub-band* dihitung dengan persamaan (1)

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

E_i = energi pada frekuensi *sub-band*
 $X_i(k)$ = nilai pada runtun ke-*k* pada frekuensi *sub-band*

Isyarat tutur yang telah direkonstruksi merupakan nilai yang mewakili isyarat asli yang masih berada dalam domain waktu, dan untuk mendapatkan vektor ciri, diperlukan proses lebih lanjut yaitu menghitung energi total hasil rekonstruksi isyarat pada masing-masing *sub-band* frekuensi menggunakan persamaan (2).

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

E_{tot} = energi total frekuensi *sub-band* pada dekomposisi level ke-*j*
 I = jumlah frekuensi *sub-band* pada dekomposisi level ke-*j*

III. HIDDEN MARKOV MODELS (HMM)

Hidden Markov Model (HMM) merupakan pemodelan probabilitas suatu system dengan mencari parameter-parameter *Markov* yang tidak diketahui untuk memperoleh analisis system tersebut. Kita harus menentukan parameter-parameter tersembunyi (*state*) dari parameter-parameter yang dapat diamati. Parameter-parameter yang ditentukan kemudian dapat digunakan untuk analisis yang lebih jauh, misalnya untuk aplikasi *pattern recognition*[6][7].

Suatu HMM dengan observasi system diskret terdiri atas elemen-elemen yang dikarakteristikan sebagai berikut[6]:

1. N , banyaknya *state* dalam model. Masing-masing *state* diberi indeks $\{1, 2, \dots, N\}$. Pada waktu-waktu diskret yang berjarak teratur, model tersebut mengalami perubahan *state* menurut suatu himpunan probabilitas yang sesuai dengan *state* tersebut. Waktu perubahan *state* dinyatakan sebagai t dan *state* pada waktu t sebagai q_t .
2. M , banyaknya simbol observasi berbeda pada setiap *state* dan berhubungan dengan keluaran fisik dari sistem yang akan dimodelkan. Simbol-simbol observasi tersebut dinyatakan sebagai $V = \{v_1, v_2, \dots, v_m\}$.
3. $A = \{a_{ij}\}$, distribusi probabilitas transisi *state*, dengan:

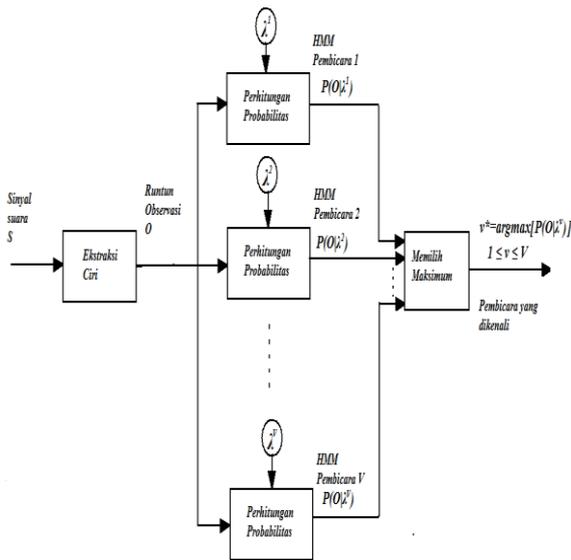
$$a_{ij} = P[q_{t+1} = j | q_t = i], \quad 1 \leq i, j \leq N \quad (3)$$
4. $B = \{b_j(k)\}$, distribusi probabilitas observasi dengan:

$$b_j(k) = P[o_t = v_k | q_t = j], \quad 1 \leq k \leq m \quad (4)$$
5. $\pi = \{\pi_i\}$, distribusi *state* awal, dengan :

$$\pi_i = P[q_t = i], \quad 1 \leq i \leq N \quad (5)$$

Pada penelitian ini, untuk setiap jenis ekstraksi ciri dengan jumlah elemen yang berbeda akan dibangun sistem pengenalan pembicaranya dengan

jumlah pembicara yang akan dikenali 6 orang. Gambar 4.3 memperlihatkan blok diagram sistem pengenalan pembicara yang dibangun. Untuk sistem pengenalan pembicara HMM dengan jumlah pembicara yang akan dikenali 2 orang, maka jumlah model HMM yang dibangun adalah 2 buah, untuk jumlah pembicara 3 orang maka model HMM yang dibangun 3 buah, begitu seterusnya sampai dengan jumlah 6 orang pembicara yang akan dikenali maka dibangun 6 buah model HMM[6][7].



Gambar 2 Blok diagram sistem pengenalan pembicara HMM [7].

Untuk masing masing pembicara dibangun sebuah model HMM dengan parameter-parameter sebagai berikut.

$$\lambda = (A, \pi, \mu) \tag{6}$$

dengan :

- a. A merupakan matriks probabilitas transisi *state*
- b. π merupakan distribusi probabilitas *state* awal.
- c. Rata-rata (μ) dicari dengan merata-ratakan runtun observasi dari 5 suara pertama yang diucapkan oleh masing-masing pembicara

Prosedur dalam membangun model HMM untuk 6 penutur yang berbeda dan pengujian model-model HMM menggunakan algoritma Baum-Welch, serta penentuan hasil pengenalan menggunakan algoritma Viterbi. mengacu pada langkah-langkah yang dilakukan peneliti [6].

Uruan proses pengenalan identitas penutur/pembicara sebagai berikut:

- a. Setiap data isyarat suara yang akan ditujukan pada system harus dilakukan ekstraksi ciri untuk mendapatkan runtun observasi yang disusun menjadi matriks data.
- b. Menghitung probabilitas observasi, untuk setiap runtun observasi dicari matriks probabilitas observasinya dengan menggunakan persamaan sebagai berikut:

$$b_j(o_t) = \frac{1}{1+d_j(o_t)} \tag{7}$$

dengan

$$d(o_t, \mu_j) = \sqrt{\sum_{k=1}^M (o_t, \mu_j)^2} \tag{8}$$

$d(o_t, \mu_j)$, merupakan jarak eucliden antar runtun observasi pada waktu t dengan rata-rata runtun observasi pada state j. Dan M adalah jumlah elemen runtun observasi pada state j.

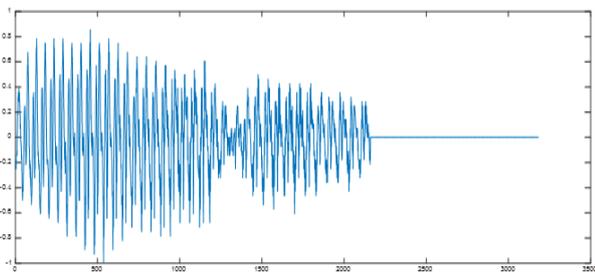
- c. Pembicara yang dikenali pada proses pengujian adalah model pembicara yang memiliki probabilitas maksimum.

Persentase total hasil pengenalan dihitung berdasarkan persamaan (9), dan dalam penelitian ini akan diamati pengaruh perubahan jumlah state dalam meningkatkan persentase akurasi pengenalan

$$\%Pengenalan = \frac{\text{Jumlah hasil sesuai target}}{\text{Total data yang diidentifikasi}} \times 100\% \tag{9}$$

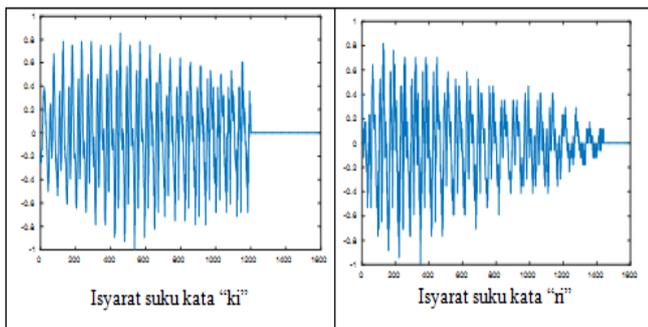
IV. HASIL DAN PEMBAHASAN

Isyarat tutur kata “kiri” direkam dengan menggunakan software bantu perekam dan file isyarat yang direkam disimpan dalam format ‘.wav’. isyarat tutur yang direkam sebelum melalui pemrosesan awal memiliki panjang yang bervariasi antara 18000 hingga 35000 sampel, yang kemudian dibaca dan ditampilkan untuk diproses lebih lanjut ke tahap pemrosesan awal. gambar 3 merupakan contoh dari hasil dari pemrosesan awal isyarat kata “kiri” Gambar 2 merupakan gambar isyarat yang dibagi menjadi 2 bagian untuk masing-masing suku kata “ki” dan “ri”.



Gambar 4 Isyarat tutur “kiri”

Gambar 5 merupakan gambar isyarat yang dibagi menjadi 2 bagian untuk masing-masing suku kata “ki” dan “ri”



Gambar 6. Isyarat suku kata “ki” dan “ri”

A. Hasil ekstraksi Ciri dengan Discrete Wavelet Transform

Pada tahapan ini menghasilkan vektor ciri untuk masing-masing isyarat tutur kata kiri yang disimpan pada format “.mat”. Vektor ciri yang diperoleh pada metode penelitian ini berjumlah 16 seperti yang terlihat pada Tabel 5.1. masing-masing adalah nilai-nilai energi pada frekuensi sub-band yang berada pada level 3, jumlah masing-masing 8 ciri mewakili suku kata “ki” dan “ri” isyarat tutur yang sebelumnya memiliki apanjang 4000 sampel yang tidak di normalisasi.

Tabel 1 Perbandingan vektor ciri pada kelas yang berbeda dengan modifikasi suara sengau.

No	Suku Kata	Frekuensi sub-band	Vektor ciri dengan menurunkan suara sengau					
			Vj_n_kr	S_n_kr	R_n_kr	F_n_kr	As_n_kr	A_kr_n
1	"ki"	500	29.49723	10.81918	8.987022	9.065187	25.96406	11.32695
2		1000	2.988787	7.225631	3.604223	4.239972	3.788602	6.543531
3		1500	1.160904	1.810925	1.653256	1.748778	1.047519	1.713109
4		2000	1.275526	2.072917	1.151511	0.956268	1.062128	1.52206
5		2500	0.196889	0.692345	0.526144	0.807772	0.24681	0.529996
6		3000	0.343269	0.955345	0.605221	0.869525	0.27132	0.781809
7		3500	0.876245	0.689273	1.199375	1.548509	0.375641	0.673766
8		4000	0.46138	0.960346	0.859916	0.946864	0.277751	0.766093
9	"ri"	500	37.20277	9.335761	8.814926	7.977014	43.16293	8.338341
10		1000	4.176461	6.234708	4.268328	3.134797	4.901494	3.973439
11		1500	1.238343	1.560986	1.8347	1.461319	1.481568	1.031797
12		2000	1.474982	1.576742	1.117247	0.813634	1.238296	1.036944
13		2500	0.262947	0.588729	0.773684	0.467059	0.34133	0.390764
14		3000	0.368798	0.677628	0.728602	0.482331	0.330612	0.453533
15		3500	0.659704	0.58866	0.89169	1.083254	0.66421	0.449388
16		4000	0.432717	0.648065	0.763956	0.606939	0.357261	0.436916

B. Hasil Pencarian Identitas Penutur

Isyarat tutur setelah diekstraksi ciri dengan DWT akan dibandingkan dengan masing-masing data vector ciri penutur yang berbeda. Data vector ciri tiap penutur diberi label berbeda untuk suara normal dan sengau, sebagai contoh label ” 1 “ untuk pembicara 1 yang bersuara normal dan lebel “10” untuk pembicara 1 dengan suara sengau. Tabel 2. adalah hasil simulasi pencarian identitas menggunakan algoritma HMM.

Tabel 2 Hasil Pengujian untuk jumlah State =4

No	Identitas Pembicara	Inisial Target Pembicara	Hasil Pengujian HMM					Target sesuai
1	Pembicara 1	Suara Normal (SN) =1	1	1	1	1	1	10
2		Suara Sengau (SS) = 10	10	10	10	10	10	
3	Pembicara 2	Suara Normal (SN) =2	2	2	2	6	2	9
4		Suara Sengau (SS) = 20	20	20	20	20	20	
5	Pembicara 3	Suara Normal (SN) =3	30	20	60	30	30	6
6		Suara Sengau (SS) = 30	20	30	2	30	30	
7	Pembicara 4	Suara Normal (SN) =4	4	40	40	20	40	9
8		Suara Sengau (SS) = 40	40	40	40	40	40	
9	Pembicara 5	Suara Normal (SN) =5	50	20	1	50	40	5
10		Suara Sengau (SS) = 50	10	50	50	50	1	
11	Pembicara 6	Suara Normal (SN) =6	2	6	60	6	60	9
12		Suara Sengau (SS) = 60	60	60	60	60	60	
Total			60					48
% Pengenalan								80%

Proses pencarian pembicara menggunakan algoritma HMM diperoleh hasil pengujian Tabel 2 .Proses pencarian target pembicara mampu mengidentifikasi suara sengau dan normal berdasarkan vector ciri DWT yang diberikan melalui pemodelan HMM, adapun tingkat persentase tingkat pengenalan identitas 80 % untuk jumlah state sama dengan 4. Nilai persentase pengenalan dihitung menggunakan persamaan (9).

Perbandingan nilai persentase hasil pengenalan dengan mengubah jumlah state (4,5,6,7) dalam hal ini jumlah state akan memberi pengaruh dalam penentuan indeks runtun observasi pada data yang akan dimodelkan dan diuji pada proses HMM system pencarian identitas pembicara yang dirangkum dalam Tabel 3.

Tabel 3 Perbandingan tingkat akurasi berdasarkan jumlah state yang berbeda

No	Identitas Pembicara	Jumlah State			
		4	5	6	7
1	Pembicara 1	10	10	10	10
2	Pembicara 2	9	9	9	9
3	Pembicara 3	6	6	5	5
4	Pembicara 4	9	9	8	8
5	Pembicara 5	5	6	5	2
6	Pembicara 6	9	8	9	8
% Pengenalan		80%	80%	77%	70%

Berdasarkan data yang diperoleh pada Tabel 5.4, tingkat akurasi yang diperoleh untuk jumlah state 4 dan 5 sebesar 80%, sementara jumlah state 6 dan 7 secara berurutan hanya 77% dan 70% . Tingkat akurasi yang hanya mencapai 80 % menunjukkan tingginya tingkat kesalahan dalam pengenalan identitas target dari data vector ciri yang diberikan. Beberapa faktor penyebab tidak optimalnya hasil yang dicapai dalam penelitian ini diantaranya; karakteristik vektor ciri dari kata yang sama memiliki tingkat kesamaan yang cukup tinggi hingga nilai diskriminasi antara masing - masing data pembicara yang berbeda sangat kecil.

Sebagai bahan evaluasi penelitian berikutnya akan dilakukan dua cara untuk mengoptimalkan hasil penelitian yang diperoleh dalam penelitian ini; pertama pemilihan metode pencarian vector ciri DWT yang tepat dan mampu mengoptimalkan hasil pengenalan identitas penutur menggunakan metode HMM. Kedua mencari metode pengenalan yang tepat untuk diterapkan bersama metode HMM agar dapat mengoptimalkan nilai vector ciri yang dihasilkan dari metode DTW

IV. KESIMPULAN (PENUTUP)

Hasil penerapan metode *Discrete Wavelet Transform* (DWT) dan *Hidden Markov Models* (HMM) yang telah dilakukan dalam penelitian ini belum memberikan hasil yang optimal. Hal ini menunjukkan bahwa nilai-nilai vector ciri yang dihasilkan dari algoritma DWT jika dimodelkan dan diuji menggunakan algoritma HMM untuk tujuan pengenalan identitas dengan jumlah state yang diuji adalah 4 sampai dengan 7 buah state, akan diperoleh nilai persentase tingkat kesalahan pengenalan cukup tinggi hingga 20 % untuk jumlah state 4 dan 5, dan kesalahan pengenalan meningkat hingga 30 % untuk jumlah state 6 dan

Penyebab tingginya nilai presentase kesalahan pengenalan perlu dikaji lebih lanjut agar diperoleh metode dan algoritma yang tepat untuk meningkatkan persentase pengenalan dan dapat diterapkan bersama metode HMM dan DWT.

REFERENSI

- [1] B. . Asni, R. Hidayat, and N. A. Setiawan, "Pengenalan Tutur Vokal Bahasa Indonesia Menggunakan Metode DWT dan DTW," *Pros. ReTII ke-9*, vol. 1, pp. 37–42, 2014.
- [2] S. C. Sajjan and C. Vijaya, "Comparison of DTW and HMM for Isolated Word Recognition," *IEEE Proceeding Int. Conf. Pattern Recognition, Informatics Med. Eng.*, no. 1, pp. 466–470, 2012.
- [3] S. Hidayat, R. Hidayat, and T. B. Adji, "Sistem Pengenal Tutur Bahasa Indonesia Berbasis Suku Kata Menggunakan MFCC, Wavelet Dan HMM," *7Th Natl. Conf. Inf. Technol. Electr. Eng.*, no. September, pp. 246–251, 2015.
- [4] M. Gultom and D. Alamsyah, "Penerapan Hidden Markov Model (HMM) pada Pengenalan Penutur," *Jatisii*, *iISSN240-4322*, vol. 1, no. 1, pp. 111–122, 2014.
- [5] N. Trivedi, "Speech Recognition by Wavelet Analysis," vol. 15, no. 8, pp. 27–32, 2011.
- [6] L. R. Rabiner, "Hmm_Paper_Rabiner.Pdf," *PROceedings IEEE*, vol. 77, no. 2, pp. 259–585, 1989.
- [7] B. Darmawan and S. Ariessaputra, "Sistem Pengenalan Dan Verifikasi Pembicara HMM," *CITEE*, *ISSN 2085-6350*, pp. 24–26, 2018.