

Penerapan Metode *Discrete Wavelet Transform* (DWT) dan *Gaussian Mixture Model* (GMM) Sebagai Pengenal Penutur

Jeckson Sidabutar

Program Studi Rekayasa Keamanan Siber, Sekolah Tinggi Sandi Negara (STSN)

Jl. H. Usa, Putat Nutug, Ciseeng, Bogor, Jawa Barat 16120, Indonesia

E-mail: jeckson.sidabutar@stsn-nci.ac.id

Abstrak

Speaker Recognition adalah cabang dari pengolahan sinyal suara yang mempunyai ciri biometrik. Speaker Recognition dipengaruhi berbagai aspek seperti karakteristik bicara seseorang yang unik dalam bicaranya (tingkat suara karena waktu, kesalahan membaca atau pengucapan kata, kondisi kesehatan, tekanan emosional), bahasa, lingkungan, adanya noise dan lain sebagainya. Sehingga membuat penelitian tentang suara sampai saat ini masih tidak bekerja dengan baik. Penelitian ini membahas dan membangun sistem pengenalan penutur secara otomatis yang akan mengenali dan mengekstraksi kemiripan suara dari masing-masing manusia sehingga kita dapat mengetahui dan mengidentifikasi suara seseorang berdasarkan perbedaan karakteristik suara dari masing-masing manusia. Pada penelitian ini akan dilakukan kajian tentang proses ekstraksi ciri menggunakan metode Discrete Wavelet Transform (DWT) serta pengenalan pola dan clustering menggunakan metode Gaussian Mixture Model (GMM). Hasil percobaan yang telah dilakukan menunjukkan sistem dapat bekerja dengan baik berdasarkan banyaknya data training penutur, sehingga sistem menghasilkan akurasi di atas 90%. Hal ini dikarenakan pada teknik GMM pengenalan pola suara berdasarkan kemiripan suara yang telah disimpan dalam data training, dan melakukan pencocokan data testing melalui kemiripan ekstraksi suara yang ada pada data training.

Kata kunci: *speaker recognition, pengolahan sinyal suara, suara manusia, discrete wavelet transform, gaussian mixture model.*

Abstract

Speaker Recognition is a branch of speech processing that can be used for biometrics. Speaker Recognition relies on features influenced by several aspects such as; unique characteristics of human speaking style (sound level times, error reading or mispronounced words, health issues, emotional tension), languages, environment, and the presence of other kind of noise. Thus, the research on voice recognition is not working properly, that makes the study was conducted. This study discusses and builds automatic Speaker Recognition system that will recognize and extract the similarities of each human voice so that we can get to know and identity a human voice based on the type. This study was conducted by the process of feature extraction using Discrete Wavelet Transform (DWT) as well as pattern recognition and clustering using Gaussian Mixture Model (GMM). The results of this study showed that the system can work well by using speaker's training data, so the system has accuracy above 93%. This is because the voice pattern recognition GMM's technique is based on the similarity of sound that have been stored in the training data and perform test data matching based on the similarity of voice extraction in the training data.

Keywords: *Speaker Recognition, speech processing, human speaking, discrete wavelet transform, gaussian mixture model.*

1. PENDAHULUAN

Suara dalam definisi lainnya merupakan suatu besaran yang memenuhi syarat sebagai ciri biometrik yang efektif dan efisien. Suara adalah fenomena perpaduan multidimensi yang

dipengaruhi karakteristik pembicara (dimensi titik artikularis, emosi, kesehatan, umur, jenis kelamin, dealek), bahasa, dan lingkungan (*background* dan media transmisi), sehingga sistem yang telah dikembangkan hingga sekarang belum bisa bekerja dengan baik pada situasi *real*. Oleh karena itu, pemodelan sinyal bukanlah sesuatu yang mudah dan sangat menantang untuk dipelajari.

Voice Recognition adalah ilmu yang mempelajari tentang penutur suara, terbagi menjadi *Speech Recognition* dan *Speaker Recognition*. Dalam fungsinya *Speech Recognition* berfokus pada pengenalan teks ucapan penutur berdasarkan ekstraksi dari beberapa informasi pesan yang diucapkan, sedangkan *Speaker Recognition* menekankan pada analisis suara yang diucapkan oleh tiap penutur berdasarkan karakteristik bicara seseorang yang unik.

Penelitian ini memfokuskan tentang *Speaker Recognition* (Pengenalan Penutur), terdapat dua proses utama pada *Speaker Recognition* yaitu ekstraksi ciri dan pengenalan pola. Pada tahap ekstraksi ciri data masukan menjadi vektor ciri. Jika proses ekstraksi ciri dapat menghasilkan vektor ciri yang efektif mampu mencirikan obyek masukan tanpa terpengaruhi oleh adanya gangguan, maka proses pengenalan menjadi jauh lebih mudah. Oleh karena itu, untuk mengatasi gangguan *noise* yang muncul pada sistem pengenalan pembicara difokuskan pada bagian ekstraksi ciri [1]. Beberapa penelitian tentang ekstraksi ciri dan pengenalan pola tentang *voice recognition*, dapat dilihat pada Tabel 1.

Tabel 1
Penelitian Ekstraksi Ciri dan Pengenalan Pola

Referensi	Metode Ekstraksi Ciri	Metode Pengenalan Pola	Materi Penelitian	Akurasi Pengenalan dan Analisis
[2]	DWT	DTW	Data berasal dari 5 pembicara, dengan sampel suara isyarat vokal Indonesia.	Hasil terbaik pengenalan menggunakan DWT level 3 sebesar 80% secara <i>real time</i> .
[3]	MFCC	DTW & GMM	Pengenalan ucapan Bahasa Arab dari 3 pembicara.	Tingkat pengenalan teks-independen Bahasa Arab dari sistem mencapai 90%
(Andika <i>et. al.</i> , 2010)	<i>Gabor Wavelet</i>	<i>JST Back-propagation</i>	Data berasal dari 5 pembicara, 5 sampel suara sehingga jumlah total 25 suara latihan.	Tingkat keberhasilan sistem pengenalan suara 75% - 90% dan sistem belum dapat bekerja secara <i>real time</i>
[1]	HOS & MFCC	HMM	Data berasal dari 10 pembicara (8 Pria dan 2 Wanita) yang mengucapkan kata "PUDESHA" sebanyak 80 kali.	Sangat efektif untuk sinyal suara asli, namun dengan <i>noise</i> yang lebih tinggi sistem tidak bekerja dengan baik.
(Ali, 2007)	MFCC	<i>Vector Quantization</i>	Data berasal dari 10 pembicara dengan mengucapkan kata 'kampus'.	Pengujian penutur dalam <i>database</i> sangat efektif, akan tetapi ada sebagian orang yang belum dilatih dapat lolos.

Berdasarkan uraian pada Tabel 1 disimpulkan identifikasi dan verifikasi, teks-independen dan teks-dependen pada teknologi *Speaker Recognition* memiliki kelebihan dan kekurangan, dan memerlukan perawatan serta teknik yang berbeda. Sebagian besar peneliti menggunakan ekstraksi ciri MFCC. MFCC merupakan pendekatan yang berbasiskan Transformasi *Fourier*.

Pada penelitian ini akan dilakukan pendekatan lain yaitu Transformasi *Wavelet*. Transformasi *Wavelet* dapat menghasilkan resolusi waktu yang baik pada frekuensi tinggi dan mampu mengenali sinyal diskontinyu secara akurat (Chiyomi *et. al.*, 2001). Metode ekstraksi ciri *Discrete Wavelet Transform* (DWT) level 3 adalah teknik terpopuler dalam ekstraksi ciri sinyal suara dan terbukti mampu menghasilkan vektor ciri yang efektif, dan menghasilkan akurasi sistem secara *realtime* mencapai 80% [2]. Sedangkan pengenalan pola dan *clustering* menggunakan metode *Gaussian Mixture Model* (GMM) dan telah sukses memverifikasi suara pada beberapa NIST *Speaker Recognition Evaluations* (SRE) [4].

Dari penjelasan diatas disimpulkan bahwa tahapan ekstraksi ciri sangat diperlukan pada proses pengenalan suara dalam hal ini menggunakan metode DWT level 3. Output dari ekstraksi ciri menjadi masukkan dalam tahapan pengenalan pola yang dilakukan dengan menggunakan metode GMM. Untuk membangun konsep tersebut, peneliti membuat sebuah sistem “*Speaker Recognition*” menggunakan Matlab [5].

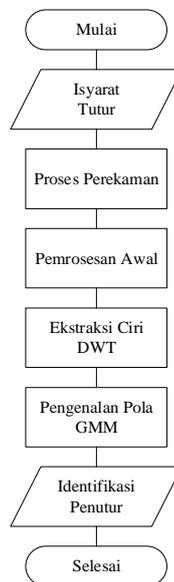
2. METODE PENELITIAN

2.1 Jalannya Penelitian

Diagram alir utama pada penelitian ini dibagi menjadi empat tahap, yaitu:

1. Perekaman data tutur;
2. Pemrosesan-awal (DC removal, normalisasi amplitudo, proses pembuang isyarat hening);
3. Ekstraksi ciri; dan
4. Pencocokan pola.

Isyarat tutur hasil rekaman pertama kali melalui tahap berikut diproses untuk ekstraksi ciri seperti pada Gambar 3.



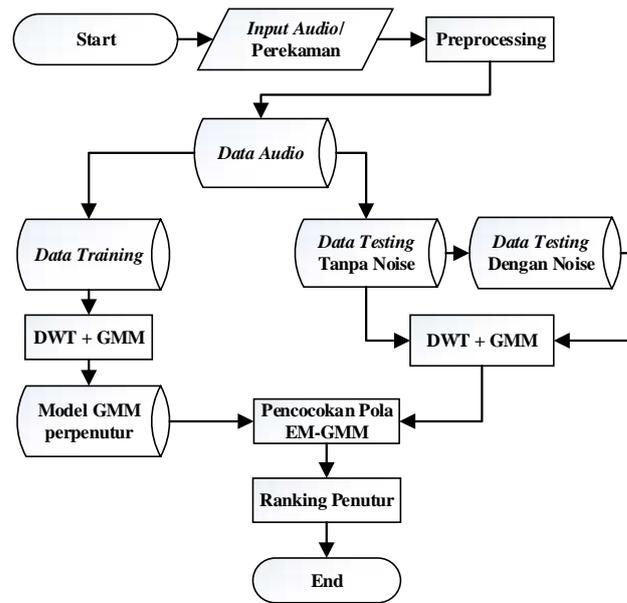
Gambar 1. Diagram Alir Secara Umum Proses Penelitian

2.2 Metode Pengumpulan Data dan *Flowchart* Pemodelan DWT dan GMM

Penelitian ini menggunakan data sampel suara dari 30 orang penutur, 15 orang pria dan 15 orang wanita dalam rentang usia 25-55 tahun.

Data yang digunakan dalam penelitian adalah berupa *sample* suara dari 30 orang penutur, 15 penutur pria dan 15 penutur wanita berusia antara 25-55 tahun. Setiap penutur mengucapkan kata kanan dan kiri sebanyak 15 kali sehingga total masing-masing penutur berjumlah 30 ucapan. Dalam tahap pengujian ini akan digunakan tiga *set database data training* sebagai bahan perbandingan, yaitu: 5 ucapan, 10 ucapan dan 15 ucapan. Sedangkan untuk data *testing* akan menggunakan 2 set data *testing*, yaitu dengan *audio* asli dan *audio* asli yang ditambahkan *noise* menggunakan *gaussian noise*.

Percobaan pengenalan penutur dilakukan dengan mengikuti *flowchart* pada Gambar 4.



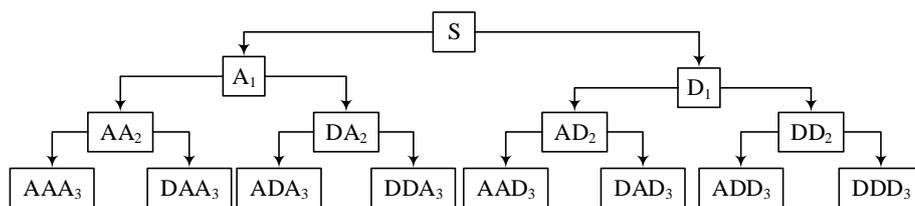
Gambar 2. Flowchart Pengenal Penutur dengan Metode DWT dan GMM

3. LANDASAN TEORI

3.1 Discrete Wavelet Transform (DWT)

Pada dasarnya prinsip dari *Discrete Wavelet Transform* adalah bagaimana cara memperoleh representasi waktu dan skala dari sebuah sinyal menggunakan operasi *sub-sampling* atau *down-sampling* dan teknik penyaringan digital pada jangkauan frekuensi yang lebih luas dengan melakukan dekomposisi *full binary Wavelet Packet Transform* (WPT).

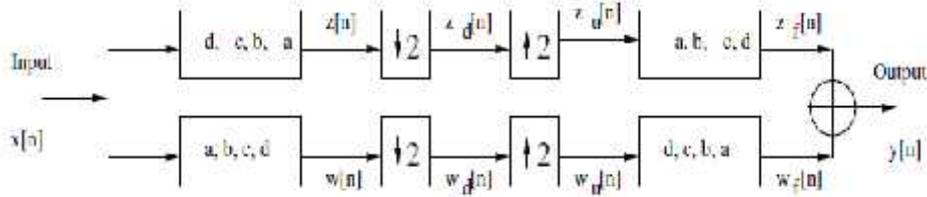
WPT adalah variasi dari DWT dengan dekomposisi pada dua sisi filter melalui dekomposisi aproksimasi untuk frekuensi rendah melalui *Low Pass Filter* (LPF) dan detail yang frekuensi tinggi menggunakan *High Pass Filter* (HPF). Proses dekomposisi dilakukan hingga level ke-3 ditunjukkan pada Gambar 1. Jumlah frekuensi *sub-band* pada level 3 adalah $2^3 = 8$.



Gambar 3. Dekomposisi *Full Binnary Paket Wavelet Level-3*

Proses dekomposisi dan rekontruksi menggunakan Fast DWT merupakan proses konvolusi antara isyarat dan koefisien filter. Hasil konvolusi kemudian diseleksi menggunakan faktor 2 untuk proses *down sampling* seperti pada Gambar 2.

Persamaan proses dekomposisi:



Gambar 4. Proses Dekomposisi dan Rekonstruksi pada DWT

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

h_n = koefisien LPF wavelet ke-n ($n=0,1,2,\dots$)

$2k$ = dyadic (*down sampling* dengan mengeliminasi nilai pada runtun ke- $2k$ ($k=0,1,2,\dots$))

j = level dekomposisi

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

g_n = koefisien HPF wavelet ke-n ($n=0,1,2,\dots$)

$2k$ = dyadic (*down sampling* dengan mengeliminasi nilai pada runtun ke- $2k$ ($k=0,1,2,\dots$))

j = level dekomposisi

Persamaan proses rekonstruksi:

$$a_k^{(j)} = \sum_{n=-\infty}^{\infty} h_{k-2n} \tilde{a}_n^{(j+1)} + \sum_{n=-\infty}^{\infty} g_{k-2n} a_n^{(j+1)}$$

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

Dengan:

$$\tilde{a}_k^{(j+1)} = \begin{cases} a_p^{(j+1)} & \text{if } k = 2p \\ 0 & \text{if } k = 2p + 1 \end{cases}$$

dan

$$a_k^{(j+1)} = \begin{cases} a_p^{(j+1)} & \text{if } k = 2p \\ 0 & \text{if } k = 2p + 1 \end{cases}$$

$\tilde{a}^{(j+1)}$ dan $a^{(j+1)}$ adalah koefisien aproksimasi dan detail pada level $j+1$ yang nilainya berasal dari $a_k^{(j+1)}$ dan $\tilde{a}_k^{(j+1)}$ yang melalui operasi dyadupsampling seperti pada persamaan (3-7), yaitu menambahkan nilai nol diantara 2 titik interval, jika interval ganjil akan diisi dengan nol, kemudian hasilnya akan dikonvolusikan dengan koefisien filter h_k sisi LPF dan koefisien g_k pada sisi HPF.

3.2 Gaussian Mixture Models (GMM)

GMM adalah sebuah tipe *destiny model* yang terdiri dari komponen fungsi-fungsi Gaussian. Komponen fungsi ini terdiri dari *threshold* yang berbeda untuk menghasilkan *multi-model destiny* (Douglas et. al, 1995) . *Mixture model* juga merupakan metode yang efektif untuk perubahan dari model yang bergerak lambat karena merupakan semi parameter alternatif untuk histogram tanpa parameter dan memberikan fleksibilitas lebih serta ketelitian dalam memodelkan statistik dari data,

dengan kata lain untuk memvisualkan suatu *dynamic scene*. GMM adalah model statistik dari distribusi probabilitas yang didapatkan dari nilai bobot setiap distribusi *Gaussian* sehingga GMM merupakan metode yang sangat tepat untuk perhitungan, baik dengan parameter maupun tidak. Bila model telah dihasilkan, syarat peluang dapat dihitung dan GMM juga dapat ditampilkan sebagai bentuk fungsi hubungan dasar *network*, seperti berikut ini (Ari, 2010).

1. *Mixture Models*

Data $D = \{x_1, \dots, x_N\}$ adalah vector berukuran d . Diasumsikan data digunakan untuk meningkatkan kerapatan $\rho(x)$. Lebih lanjut $\rho(x)$ didefinisikan sebagai *mixture models* dengan K komponen,

$$\rho(x|\Theta) = \sum_{k=1}^K \alpha_k P(x|z_k, \theta_k)$$

Dimana:

$p_k(x|z_k, \theta_k)$ adalah komponen campuran, $1 \leq k \leq K$. Masing-masing adalah kerapatan atau distribusinya didefinisikan sebagai $p(x)$, dengan parameter θ_k .

$z = (z_1, \dots, z_K)$ adalah sebuah K vektor variabel indikator biner yang saling terkait dan lengkap (yaitu satu dan hanya satu z_k sama dengan 1, dan yang lain adalah 0). Z adalah sebuah K -array variabel acak yang mewakili identitas komponen *mixture* yang mengeluarkan nilai x . Hal ini sesuai untuk *mixture models* untuk mewakili z sebagai vektor dari indikator K variabel.

$\alpha_k = p(z_k)$ adalah bobot *mixture*, mewakili probabilitas bahwa x dipilih secara acak dihasilkan oleh k komponen, dimana $\sum_{k=1}^K \alpha_k = 1$.

Parameter untuk *mixture models* dengan komponen K adalah;

$$b = \{\alpha_1, \dots, \alpha_K, \theta_1, \dots, \theta_K\}$$

2. Bobot Keanggotaan

Bobot keanggotaan dapat di hitung dari data titik x dalam kluster k , diberikan parameter Θ sebagai:

$$S_{ik} = p(z_{ik} = 1 | x_i, b) = \frac{p_k(x_i | z_k, \theta_k) \cdot \alpha_k}{\sum_{m=1}^K p_m(x_i | z_m, \theta_m) \cdot \alpha_m}$$

Hal ini langsung mengikuti aturan Bayes, bobot keanggotaan di atas menunjukkan ketidaktentuan. Nilai x_i dan Θ , tentang K komponen yang dihasilkan vektor x_i . Perhatikan bahwa pembentukan *mixture model* dari tiap x_i yang dihasilkan oleh komponen tunggal sehingga probabilitas ini mencerminkan ketidakpastian dengan adanya nilai x_i , tidak ada pencampuran dalam proses pembangkitan.

3. *Gaussian Mixture Models*

Untuk nilai $x \in R^d$, ditetapkan sebagai sebuah GMM dengan membuat tiap K komponen kerapatan *gaussian* dengan parameter μ_k dan setiap komponen adalah sebuah kerapatan *gaussian* yang bervariasi.

$$p_k(x | \theta_k) = \frac{1}{(2\pi)^{d/2} |\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1}(x - \mu_k)}$$

Dengan parameter $\theta_k = \{\mu_k, \Sigma_k\}$

4. Algoritme *Expectation Maximization* (EM)

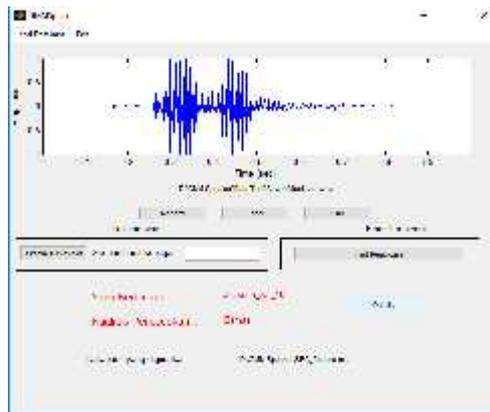
Penulis menempatkan algoritma EM (*Expectation-Maximization*) untuk GMM, sebagai Algoritma melakukan proses iterasi yang dimulai dengan inialisasi awal tujuan dari Θ (sebagai

contoh dengan cara acak), hasil dari proses iterasi memperbaiki nilai dari Θ hingga mencapai nilai konvergen. Tiap proses iterasi terdiri dari satu langkah-E dan satu langkah-M.

4. HASIL DAN PEMBAHASAN

4.1 Desain User Interface Aplikasi

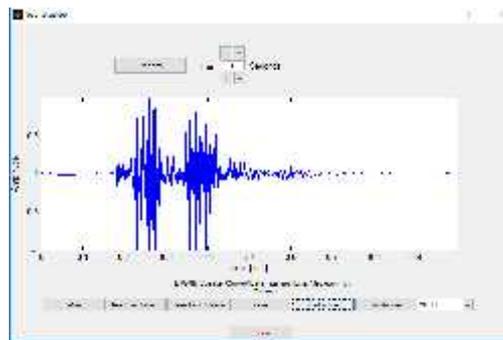
GUI (*Graphical User Interface*) aplikasi yang digunakan pada jendela utama ada lima buah *push button*, *record*, *load*, *play*, ekstraksi *wavelet*, dan *find* pembicara. *Design user interface* sistem pengenalan penutur ini disajikan pada Gambar 5.



Gambar 5. *Design User Interface* Aplikasi

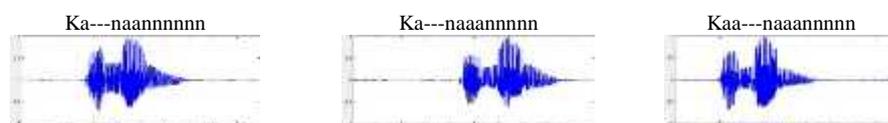
4.2 Proses Perekaman Data Suara

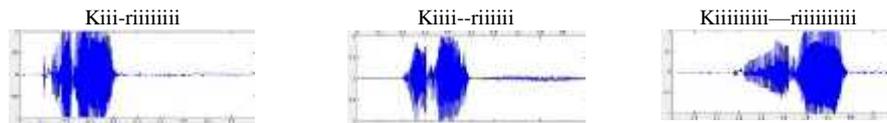
Waverecord berfungsi untuk merekam data audio dari pembicara. Setelah itu audio yang sudah terekam masuk dalam tahap *preprocessing audio* dengan cara menghilangkan *DC offset* dan normalisasi *amplitude* dari -1 sampai 1. Selanjutnya file *audio* diplot. Proses perekaman data suara dapat dilihat pada Gambar 6.



Gambar 6. Proses Perekaman Data Suara

Data *input audio* menggunakan 30 orang penutur. Masing-masing orang diambil 30 file audio yang terdiri dari 15 file dengan suara “kanan” dan “kiri”. Semua penutur mengucapkan suara tanpa dikontrol cara pengucapannya, hal ini dimaksudkan agar sistem dapat mengenali suara dengan beberapa pengucapan (tinggi-rendah dan panjang-pendek) nada suara. Beberapa sinyal suara kanan dan kiri cara pengucapannya tanpa dikontrol disajikan pada Gambar 7.

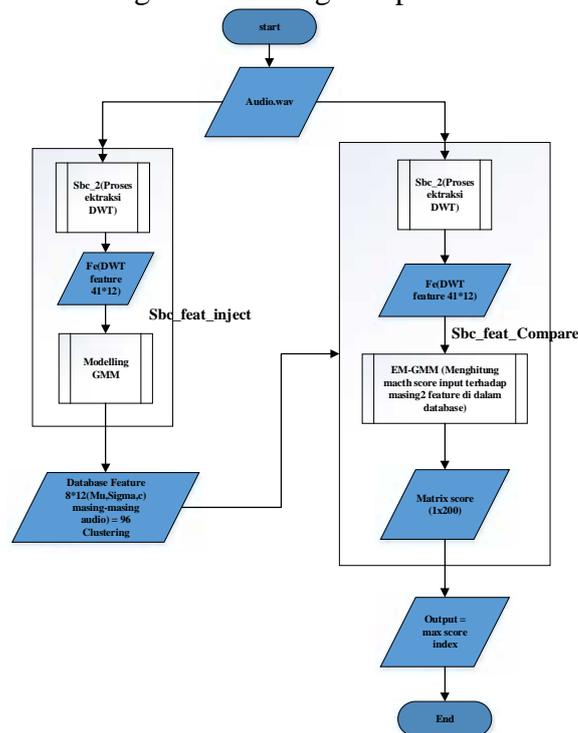




Gambar 7. Perbandingan Suara Tanpa di Kontrol Pengucapannya

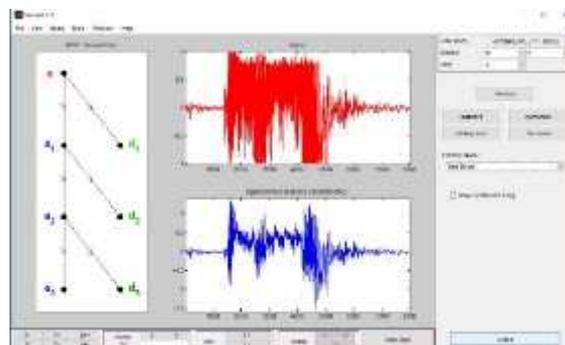
4.3 Proses Ekstraksi Ciri DWT dan Modelling GMM (Data Training)

Pada level tertinggi, semua *Speaker Recognition* berisi dua modul utama yaitu: ekstraksi fitur dan pencocokan pola. Ekstraksi fitur adalah proses yang mengekstrak sejumlah kecil data dari sinyal suara yang nantinya dapat digunakan untuk mewakili masing-masing pembicara. Pencocokan pola melibatkan prosedur yang sebenarnya untuk mengidentifikasi pembicara yang tidak diketahui dengan membandingkan dan mengelompokkan ekstraksi fitur dari inputan suara.



Gambar 8. Detil *Flowchart* Pemodelan DWT dan GMM

Sebelum proses penambahan data suara ke *database*, *audio* dalam format *.wav* akan di ekstraksi ciri dan di *import* ke dalam Matlab menjadi *variabel* dalam bentuk matrik. Proses Ekstraksi Ciri ini menggunakan DWT Level 3, jika diplot akan terlihat seperti Gambar 9 dan Gambar 10.



Gambar 9. Proses Ekstraksi Ciri DWT Level 3

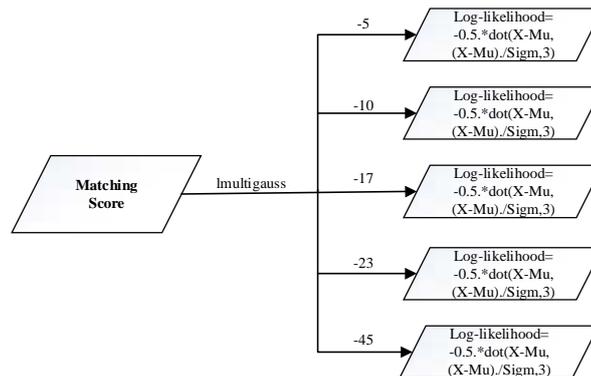


Gambar 10. Variabel Matriks Ekstraksi Ciri DWT Level 3

Setelah ekstraksi *feature* dengan DWT, data *audio* akan di proses menggunakan metode GMM. Satu *audio* terdiri dari 3 buah matrix: 8x12 matrix Mu (nilai rata-rata distribusi normal gauss), 8x12 nilai sigma (*variance*), dan 12x1 nilai *coefisien*. Hasil dari ketiga proses ini disimpan ke dalam *database*. *file feature* akan disimpan ke dalam *database Matlab* dengan format **.mat*.

4.4 Proses Pencarian Pembicara (Data Testing)

Setelah semua suara disimpan ke dalam *database audio* (data *training*), maka proses selanjutnya akan dilakukan data uji suara dengan melakukan ekstraksi ciri DWT dan pengenalan pola GMM dan menghitung kemungkinan pola yang sama dengan *matching* skor menggunakan metode *log-likelihood* (EM-GMM). Secara detail dapat dilihat alur pengenalan pola GMM dengan *matching score* pada Gambar 11.



Sumber: Olahan Peneliti

Gambar 11. Matching Score EM-GMM

Pada saat perhitungan *matching score*, hasil keluaran pencocokan berupa *match_value* dari (-100 0). *File feature* yang mendapatkan skor paling tinggi dianggap sebagai penutur yang paling sesuai dengan *audio* inputan. Hasil skor perhitungan ini dapat dilihat pada Gambar 12.

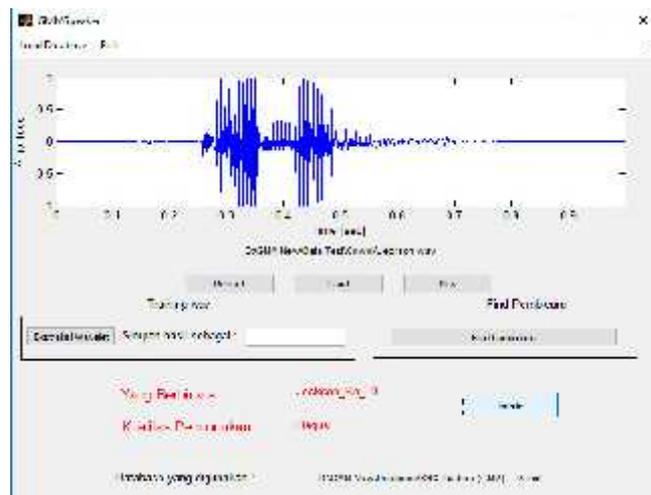


Gambar 12. Detailed Score Pengenalan Pola Suara

Audio inputan setelah di ekstraksi ciri dengan DWT dan pengenalan pola menggunakan GMM akan dibandingkan dengan masing-masing file *feature* yang ada di dalam database menggunakan metode *Expectation Maximition - Gaussian Mixture Models* (EM-GMM). Hasil keluaran pencocokan berupa *macth_value* dari (-100 0). File *feature* yang mendapatkan *score* paling tinggi dianggap sebagai Penutur yang paling sesuai dengan *audio* inputan.

4.5 Hasil Pengujian Aplikasi

Pada Gambar 13. hasil pengujian aplikasi sistem pengenal penutur menggunakan ekstraksi ciri suara dengan metode DWT dan pengenalan pola dengan metode GMM akan dilakukan beberapa pengujian, yaitu: Hasil pengujian pengenal penutur dengan suara asli, hasil pengujian pengenal penutur dengan suara asli ditambahkan *noise* dan analisa hasil pengujian dari masing-masing penutur.



Gambar 13. Hasil Pengujian Suara

4.6 Hasil Pengujian Pengenal Penutur dengan Suara Asli dan *Noise*

Pengujian pengenal penutur dilakukan dengan sinyal suara asli dan ditambahkan *noise*, dilakukan untuk menguji kemampuan akurasi metode DWT dan GMM dalam pengenalan pola suara dari masing-masing suara penutur terhadap gangguan *noise* (*gaussian*).

Hasil pengenalan penutur dengan ekstraksi ciri DWT dan pengenalan pola menggunakan metode GMM disajikan pada Lampiran 2 dan Lampiran 3, dari pengujian tersebut terlihat bahwa teknik yang dipakai bekerja dengan baik mengenali penutur, yaitu sekitar 93% untuk data asli pada proporsi data *training* 15 kata. Jika suara asli ditambahkan dengan *noise*, akurasi sistem turun menjadi 60% untuk proporsi data *training* 15 kata. Hal ini disajikan pada Tabel 2.

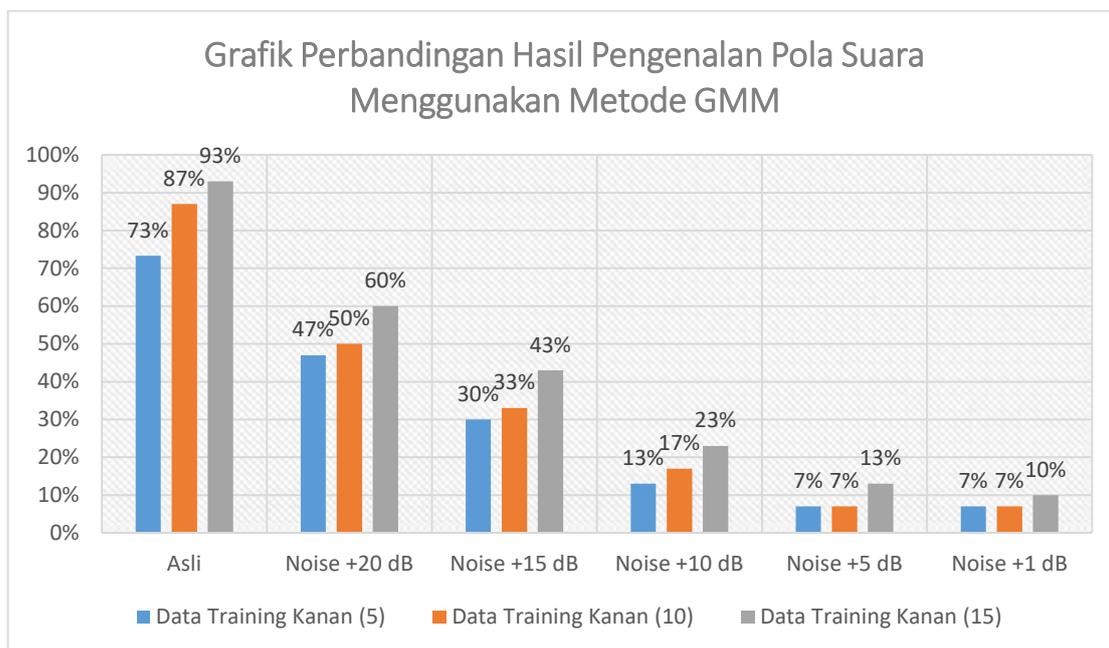
Tabel 2
Akurasi Pengujian Menggunakan Suara Asli dan Noise
Akurasi Pengenal Penutur

No	Data Training	Akurasi Pengenal Penutur					
		Asli	+Noise 20 dB	+Noise 15 dB	+Noise 10 dB	+Noise 5 dB	+Noise 1 dB
1	Data (5)	73 %	47 %	30 %	13 %	7 %	7 %
2	Data (10)	87 %	50 %	33 %	17 %	7 %	7 %
3	Data (15)	93 %	60 %	43 %	23 %	13 %	10 %

Sumber: Olahan Peneliti

Hal ini menunjukkan bahwa untuk memberikan hasil yang optimal, maka teknik GMM memerlukan jumlah data pelatihan yang banyak. Dengan 15 data pelatihan, membuat banyak data

pelatihan dengan tempo dan nada yang berbeda dan bervariasi sehingga membuat akurasi sistem menjadi lebih baik. Grafik data ini disajikan pada Gambar 14.



Gambar 14. Grafik DWT dan GMM dengan Suara Asli dan ditambahkan *Noise*

4.7 Analisa Hasil Pengujian Data *Training* dengan Data *Testing*

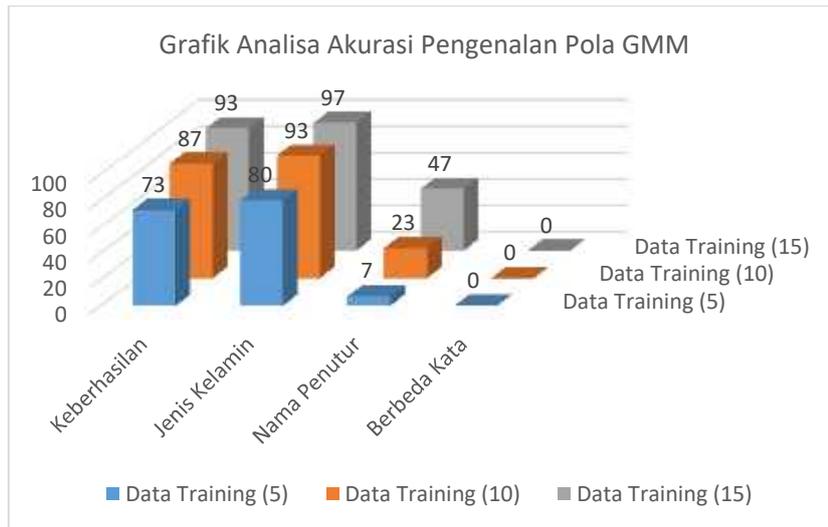
Berdasarkan hasil pengujian yang sudah kita lakukan dengan menggunakan data penutur yang berjumlah 30 (tiga puluh) orang, suara penutur dapat dikenali berdasarkan karakteristik suara dengan menggunakan pengenalan pola suara pada masing-masing penutur melalui ekstraksi suara yang telah kita simpan kedalam *database*. Hal ini dikarenakan kinerja pengenalan penutur sangat tergantung pada besarnya data training dan data testing. Sinyal suara memiliki ciri yang istimewa, berdasarkan waktu dan tinggi rendah nada suara yang dikeluarkan. Hasil pengujian ini dapat dilihat pada Tabel 3.

Tabel 3
Analisa Akurasi Rangkang Terbaik menggunakan GMM

No	Kecocokan Pola Suara	Akurasi Pengenalan Pola		
		Data Training (5)	Data Training (10)	Data Training (15)
1	Tingkat Keberhasilan	73 %	87 %	93 %
2	Sama Jenis Kelamin	80 %	93 %	97 %
3	Sama Nama Penutur	7 %	23 %	47 %
4	Berbeda Kata	0 %	0 %	0 %

Sumber: Olahan Peneliti

Dari tabel tersebut terlihat bahwa tingkat keberhasilan kecocokan pola berdasarkan banyaknya data training audio setiap penutur, sehingga penilaian skor pada audio berdasarkan kemiripan sinyal suara yang masuk. Jika sinyal suara yang diuji sama dengan komposisi sinyal suara yang ada dalam *database*, maka sistem pengenal penutur akan sesuai dengan nama penutur yang terletak pada *database audio* tersebut. Hal ini dapat dilihat grafiknya pada Gambar 16.



Gambar 15. Grafik Analisa Akurasi Pengenalan Pola GMM

Dari Gambar 16. dapat di lihat pengenalan penutur dengan ekstraksi ciri DWT dan pengenalan pola dengan GMM dapat mengenali suara penutur lebih dari 93% untuk data training 15 dan dapat mengenali suara jenis kelamin laki-laki dan perempuan dengan tingkat akurasi sampai 97 % untuk data training 15, sedangkan untuk akurasi perbedaan ucapan kiri dan kanan sistem ini menghasilkan akurasi 0% dikarenakan kata yang di ucapkan berbeda (kanan-kiri).

Beberapa faktor yang mempengaruhi karakteristik dari pengenalan penutur, diantaranya adalah: nada penutur saat perekaman tidak konsisten, ada jeda suara pada saat proses perekaman audio, dan jumlah data training masih sedikit.

5. KESIMPULAN

Pengujian pada sistem DWT dan GMM, mempunyai tingkat keberhasilan di atas 90% untuk sinyal asli, sedangkan untuk sinyal *bernoise* 20 dB tingkat keberhasilannya mencapai 60 %. Sistem mampu mengenali sinyal suara tanpa *noise* maupun dengan penambahan *noise*, secara visual ciri suara masih terlihat dengan baik di atas 10 dB, serta kinerja pengenalan penutur bergantung pada besarnya data *testing* dan data *training*. Hal ini dikarenakan pada teknik GMM pengenalan pola suara berdasarkan kemiripan suara yang telah disimpan dalam data *training*.

DAFTAR PUSTAKA

- [1] A. Buono, "Representasi Nilai Hos dan Model MFCC sebagai Ekstraksi Ciri pada Sistem Identifikasi Pembicara di Lingkungan ber-Noise," *Disertasi FASULKOM UI*, pp. 1–209, 2009.
- [2] R. V. Yuliantari, R. Hidayat, and O. Wahyunggoro, "Ekstraksi Ciri dan Pengenalan Tutar Vokal Bahasa Indonesia Menggunakan Metode Discrete Wavelet Transform (DWT) dan Dynamic Time Warping (DTW) Secara Realtime," in *Prosiding SNST ke-7*, 2016, pp. 173–178.
- [3] S. Chen, J. Zhao, and R. Yang, "Improved Design of DTW and GMM Cascaded Arabic Speaker Verification," vol. 6, no. 2, pp. 39–44, 2013.
- [4] U. Susilawati, "Penerapan Metode Penggerombolan Berdasarkan GMM dengan Menggunakan Algoritma EM," 2011.
- [5] G. S. Kumar, K. A. P. Raju, M. Rao, and P. Satheesh, "Speaker Recognition Using GMM," vol. 2, no. 6, pp. 2428–2436, 2010.