

PENGENALAN SUARA PEMBICARA MENGGUNAKAN METODE *DYNAMIC TIME WARPING*

Suzuki Syofian¹, Silvia Oktarina²

¹Dosen Teknik Informatika Universitas Darma Persada

²Jurusan Teknik Elektro Universitas Darma Persada

Abstract

The world of technology is vast and endless. Lately, it has been developing rapidly in many of its distinct fields, especially in the field of biometric. the software developed for this final project is just one fragment of an biometric application. it replicates the auditory system of a human being to indentify a speaker based on his/her voice, which is used as the input to the software.

This application was developed using LPC (Linear Predictive Coding), also supported by DTW (Dynamic Time Warping) method. LPC converts a sound signal so that needs many step in the speaker identification Pre-Emphasize Filter, Frame Blocking, Frame Windowing, Auto Corellation Analysis, LPC Analysis dan Cepstral Coeffisient into several data vector which provide useful information in the speaker identification process. therefore, the support the Dynamic Time Warping method for speaker identification process. the software uses the Object Oriented Programming (OOP) concept and programmed in Borland Delphi 7.0.

The application was tested on 10 people, each person was told to pronounce a set of 3 times predetermined words correctly, the test result showed that this application has accuracy rate of 65% to 99% in speaker identification. the parameter values used both LPC and Dynamic Time Warping methods highly contribute to affecting the aplication's accuracy rate.

Keywords: *Dynamic Time Warping, Linear Predictive Coding, Speaker Recognition*

I. PENDAHULUAN

Salah satu teknologi yang dikembangkan menggunakan sistem komputer adalah teknologi biometrik. Teknologi biometrik adalah sistem yang menggunakan bagian tubuh manusia untuk melakukan pengenalan . Jenis sistem biometrik yang sudah tersedia sekarang antara lain pengenalan sidik jari, pengenalan wajah, pengenalan retina, geometri lengan, geometri jari, pengenalan telapak tangan, pengenalan pembicara dan pengenalan tanda tangan. Salah satu bidang dalam sistem biometrik yang mengalami peningkatan adalah sistem pengenalan suara pembicara.

Teknologi pengenalan suara pembicara telah dikembangkan di dalam riset selama lebih dari empat dekade. Hal ini menjadi perhatian yang cukup besar bagi para ilmuwan sebab dengan teknologi ini manusia akan mampu berinteraksi dengan mesin sebagaimana halnya berinteraksi dengan sesama manusia.

Pengenalan suara pembicara atau *Speaker Recognition* dapat berfungsi sebagai *speaker identification* maupun *speaker verification*. *Speaker identification* adalah proses mengenali secara otomatis siapa saja yang berbicara dengan menggunakan informasi yang

spesifik dari suara pembicara termasuk gelombang suaranya. Tidak seperti *speaker verification* yang menerima atau menolak klaim identitas dari pembicara, sistem ini akan mengidentifikasi pembicara dari satu set model yang sudah tersedia.

Metode pengenalan suara pembicara mampu mengetahui identitas pembicaranya karena setiap manusia memiliki ciri suara yang berbeda-beda. Metode ini mengambil suara dari pembicara kemudian memproses dan mencocokkan faktor-faktor uniknya, sehingga akan dihasilkan suatu keputusan apakah suara tersebut dapat dikenali sebagai pembicara atau tidak. Salah satu cara pengenalannya adalah dengan menggunakan metode *Dynamic Time Warping*.

Pada metode *Dynamic Time Warping* (DTW), proses sistem pengenalan suara pembicara dilakukan dengan menggunakan sistem *text dependent*, yang memadukan beberapa pemrosesan yaitu proses *Linear Predictive Coding* (LPC) dan selanjutnya akan dilakukan pengenalan atau pencocokan dengan menggunakan algoritma DTW sehingga menghasilkan pemadanan suara yang berbasis pada jarak euclidiannya.

Adapun maksud dan tujuan dalam pembuatan Penelitian ini adalah untuk membuat aplikasi pengenalan suara pembicara menggunakan metode *Dynamic Time Warping*, yang dapat digunakan sebagai sistem keamanan yang berguna untuk meningkatkan keamanan suatu sistem dengan mengharuskan seseorang untuk mengucapkan suatu kata rahasia disamping mengetik kode atau menggunakan sidik jari. sehingga dengan program aplikasi pengenalan pembicara ini, penggunaan data oleh orang-orang yang tidak mempunyai wewenang dalam suatu perusahaan ataupun instansi-instansi pemerintahan dapat dicegah.

Penelitian ini dibatasi pada hal-hal sebagai berikut:

1. Bahasa pemrograman untuk algoritma pengenalan suara pembicara adalah Borland Delphi7.
2. Perekaman untuk pengambilan sampel-sampel suara dilakukan melalui mikrofon dengan menggunakan aplikasi Wavesurfer 1.8.5, karena menyediakan fasilitas pengeditan dan penyimpanan dalam bentuk *file*.
3. Sistem ini akan menerima *input* berupa suara yang hanya dibatasi untuk *file* suara dengan format wav saja.
4. *File* suara wav yang disediakan dalam data sebanyak sepuluh *file*, *file* tersebut berasal dari suara sepuluh orang berbeda yang mengucapkan namanya masing-masing.
5. Frekuensi sampling yang digunakan adalah 22050 Hz dengan *bit rate* adalah 8 dan *mono*.
6. Setiap periode perekaman maksimal 1 detik dan tidak lebih.

Linear Predictive Coding (LPC)

Linear Predictive Coding (untuk selanjutnya, disingkat sebagai LPC) adalah metode yang digunakan untuk

mengambil ciri dari suara karena kemudahan pengimplementasiannya pada perangkat lunak yang membutuhkan perhitungan matematis yang lebih singkat dari metode-metode yang dikenal sebelumnya.

LPC digunakan dalam proses penyadapan karakteristik suara untuk memperoleh cepstral suara. Cepstral adalah hasil perhitungan dari sinyal spektrum.

Dynamic Time Warping (DTW)

Dynamic Time Warping (DTW) merupakan cara untuk membandingkan pola suara dalam menentukan kesamaan jarak antara pola-pola yang berbeda . Untuk membandingkan dua pola suara akan diperlukan nilai koefisien cepstral dari setiap suara. Nilai tersebut didapatkan dari proses sebelumnya, yaitu *Linear Predictive Coding* (LPC).

Misalkan terdapat dua data Q dan C, dengan panjang masing - masing n dan m, dimana :

$$Q = q_1, q_2, \dots, q_i, \dots, q_n \tag{12}$$

$$C = c_1, c_2, \dots, c_j, \dots, c_m \tag{13}$$

Untuk membandingkan dua rangkaian data spektral menggunakan *Dynamic Time Warping* (DTW) maka harus diciptakan matrix N x M dengan (*i*th, *j*th) elemen matrix mengandung jarak $d(q_i, c_j)$ antara dua titik q_i dan c_j .

Untuk menghitung jarak tersebut digunakan *Euclidian Distance* (Eamonn J.Keogh dan Michael J.Pazzani: 2005):

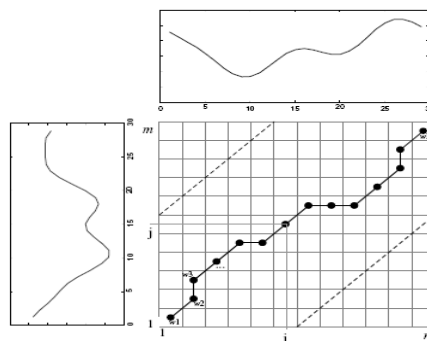
$$d(q_i, c_j) = (q_i - c_j)^2 \tag{14}$$

Elemen *k*th dari *W* didefinisikan sebagai $w_k = (i,j)_k$ sehingga :

$$W = w_1, w_2, \dots, w_k, \dots, w_K$$

W = *warping path*, yaitu jalur yang terbentuk pada matriks hasil penggabungan data Q dan C.

Contoh gambar *template* dan *testing* dapat dilihat pada Gambar dibawah ini;



warping path antara pola dilihat pada Gambar

Gambar 1. *Warping Path*

Beberapa langkah yang dibutuhkan untuk membentuk jalur pelengkungan:

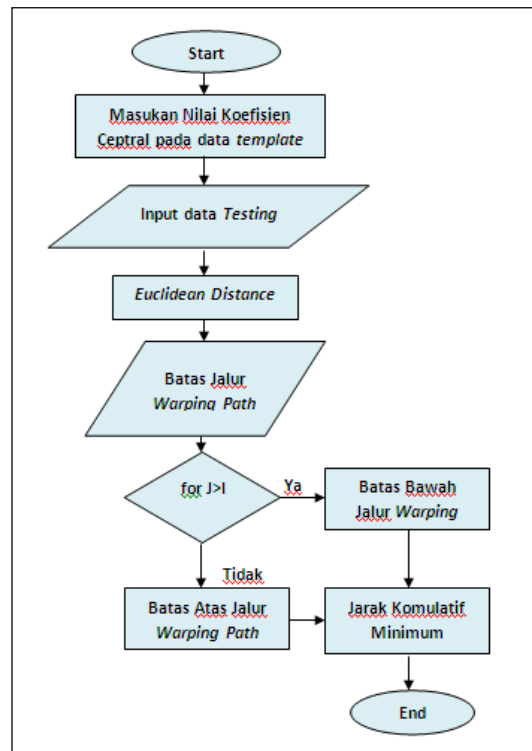
- *Boundary conditions* : $w_1 = (1,1)$ dan $w_k = (m,n)$, jalur kelengkungan harus dimulai dan diakhiri dalam diagonal yang terbentuk dalam matrix.
- *Continuity* : diberikan $w_k = (a,b)$ maka $w_{k-1} = (a',b')$ dengan $a-a' \leq 1$ dan $b-b' \leq 1$.
- *Monotonicity* : diberikan $w_k = (a,b)$ maka $w_{k-1} = (a',b')$ dengan $a-a' \geq 0$ dan $b-b' \geq 0$.

Dynamic Programming

Dynamic Programming (DP) digunakan untuk memecahkan masalah deteksi pengurutan, kemampuan penggunaan pola pengenalan wicara dan masalah waktu penjarangan dan normalisasi. Ada dua tipe masalah penggunaan Dynamic Programming, yaitu :

1. Masalah jalan optimal, dengan menganggap sekumpulan titik mulai dari 1 sampai dengan N, yang dihubungkan dengan pasangan setiap dari titik (i,j). Representasi perpindahan secara langsung dari titik ke 1 ke titik lainnya. Rentetan perpindahan tersebut tidak mempunyai bilangan yang ditetapkan dari transisi satu titik ke titik lainnya, hal ini disebut rentetan keputusan asinkron.
2. Masalah keputusan rentetan sinkronisasi, terdapat perbedaan dari satu sinkronisasi proses

Maka dapat yang paling minimum, paling mirip dengan menentukan bahwa dikenali atau tidak persentase kesalahan ditentukan.



yang beraturan dari keputusan. ditentukan dari jarak pola manakah yang *voice input*. Untuk suara yang diuji akan dibatasi oleh minimum yang sudah

Gambar 2. Flowchart proses testing Dynamic Time Warping

Pengujian Kotak Hitam (*Black-Box*)

Pengujian *black-box* adalah pengujian yang dilakukan untuk antarmuka perangkat lunak, pengujian ini dilakukan untuk memperlihatkan bahwa fungsi-fungsi bekerja dengan baik dalam arti masukkan yang diterima dengan benar dan keluaran yang dihasilkan benar-benar tepat, pengintegrasian dari eksternal data berjalan dengan baik (*file/table*).

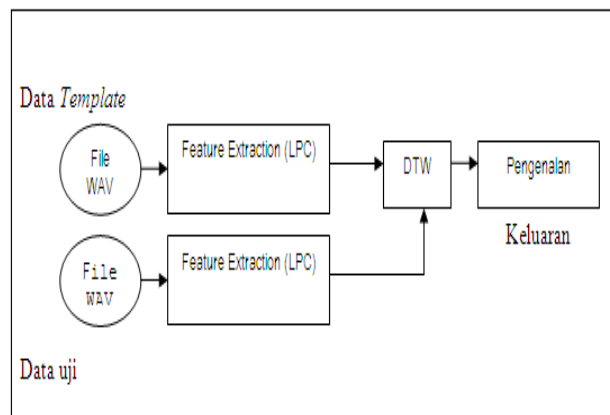
Pengujian kotak hitam berusaha menemukan kesalahan dalam kategori sebagai berikut (R. Pressman, 2002:551) :

- Fungsi-fungsi yang tidak benar atau hilang.
- Kesalahan antarmuka (*interface*).
- Kesalahan dalam struktur data atau akses basis data eksternal.
- Kesalahan kinerja.
- Inisialisasi dan kesalahan terminasi.

II. ANALISIS SISTEM

Tahapan analisis merupakan tahapan yang paling awal dalam membuat sebuah perangkat lunak. Pada tahapan ini dilakukan pengenalan sidik jari menggunakan metode *Dynamic Time Warping* adalah membandingkan suara yang masuk dengan suara yang telah disimpan sebelumnya, atau disebut sebagai data *template*. Suara yang baru masuk diproses terlebih dahulu sebelum dibandingkan dengan *template* yang sudah ada.

Sistem yang dirancang pada program ini adalah suatu program aplikasi pengenalan pembicara yang dapat mengenali suara seseorang yang telah direkam dalam format *.wav dan selanjutnya diekstrak dengan menggunakan *Linear Predictive Coding* (LPC). Kemudian hasil dari proses LPC diproses menggunakan metode *Dynamic Time Warping* (DTW) untuk melakukan pengenalan. Sistematika sistem yang dirancang dapat dilihat pada gambar dibawah ini.



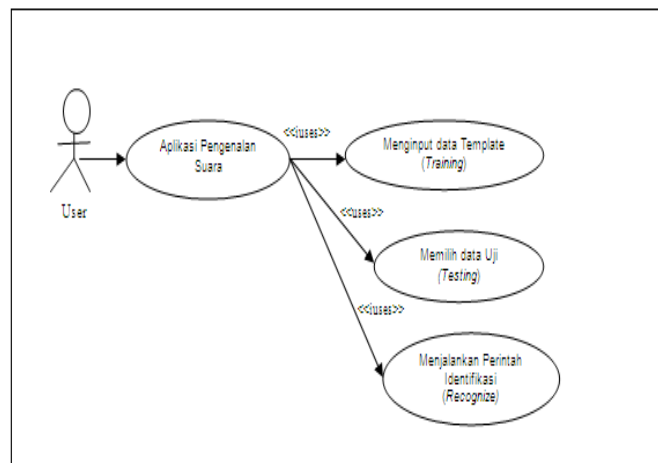
Gambar 3. Sistematika pengenalan suara

Pada gambar di atas dijelaskan bahwa kedua file yaitu data *template* dan data uji diproses dengan *Linear Predictive Coding* (LPC) untuk mendapatkan koefisien cepstral yang kemudian diproses oleh *Dynamic Time Warping* (DTW) untuk didapatkan jarak

antara data *template* dengan data uji. Pada proses pengenalan, akan diambil jarak paling minimum antara data uji dan data *template* yang ditetapkan dan diputuskan apakah data uji tersebut cocok dengan salah satu data *template* yang ada dengan menggunakan ambang batas yang telah ditentukan

Use Case Diagram Sistem Pengenalan Suara

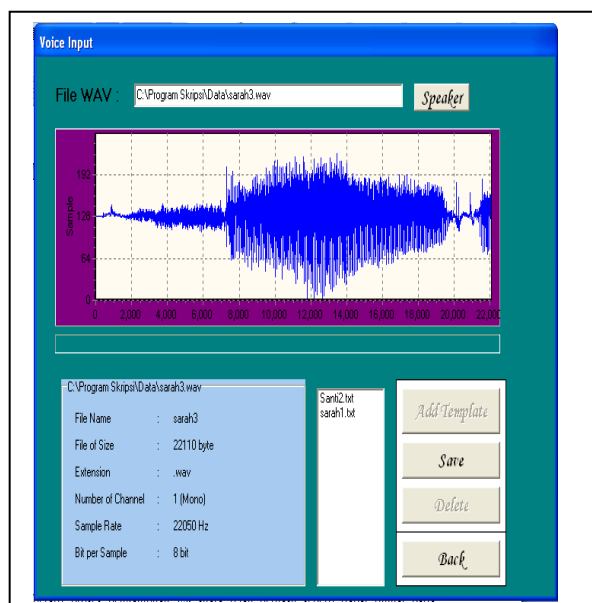
Use case diagram mendeskripsikan interaksi tipikal antara para pengguna sistem dengan sistem itu sendiri, dengan memberi sebuah narasi tentang bagaimana sistem tersebut digunakan. Diagram ini menunjukkan fungsionalitas suatu sistem atau kelas dan bagaimana sistem berinteraksi dengan dunia luar. *Use case diagram* sistem pengenalan suara menggunakan metode *Dynamic Time Warping (DTW)* secara umum dapat dilihat pada gambar 6 berikut ini :



Gambar 5. *Use case Diagram*

Pelatihan *Linear predictive Coding*

Pada tahap pelatihan merupakan suatu tahapan terpenting dalam menginput data *template*.suara yang diinput dari *file directory* kemudian *feature extraction* dengan *Linear predictive Coding (LPC)* merupakan proses mengekstraksi data hasil akuisisi sehingga dihasilkan data yang berdimensi lebih kecil, yang nantinya digunakan untuk merepresentasikan tiap-tiap pembicara.

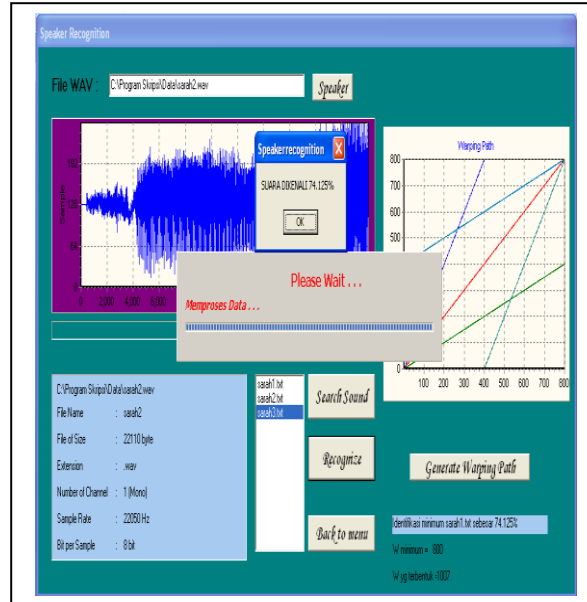


Gambar 4. Proses *Training*

Gambar Di atas merupakan proses pengambilan suara yang akan dijadikan data template, yang kemudian secara otomatis akan tersimpan dalam *template list*.

Pengujian Dengan Metode *Dynamic Time Warping*

Proses pengujian merupakan tahap akhir dari proses menggunakan yang ada di dalam sistem pengenalan suara pembicara menggunakan *Dynamic Time Warping* (DTW). Pada tahap pengujian membandingkan data template yang sudah tersimpan dalam *template list* dan kemudian dibandingkan dengan data uji yang diambil dari *file directory*, dan dengan menggunakan metode DTW akan menghasilkan gambar *Warping Path* antara data template dengan data uji. Berikut hasil pengujian dari data suara yang diuji.



Gambar 7. Proses Testing

III. HASIL

Hasil pengujian didapatkan dari hasil pengenalan suara pada bidang masukan.

Pengujian yang dilakukan menggunakan sidk jari dengan metode *Dynamic Time Warping*. Berikut adalah tabel hasil pengujian:

Tabel 1. Tabel hasil pengujian

No	Voice Template (* .txt)	Nilai persentase kecocokkan (%)	Keputusan	File suara yang diuji (* .wav)
1	Agvie2	99.625	Dikenali	Agvie1
2	Fitri2	67.5	Dikenali	Fitri3
3	Irene1	66.875	Dikenali	Irene2
4	Lendra2	99.375	Dikenali	Lendra1
5	Luna3	85.5	Dikenali	Luna2
6	Resa3	54.75	Tidak Dikenali	Lendra3
7	San2	70.5	Dikenali	San3
8	Santi2	60.5	Tidak Dikenali	Reza1
9	Sarah1	70.375	Dikenali	Santi3
10	Silvie3	61.75	Tidak Dikenali	Sarah2

Berdasarkan data dari tabel pengujian di atas dengan 10 *file* suara yang diucapkan oleh 10 orang yang berbeda, dapat dilihat bahwa 7 dari 10 orang pembicara menghasilkan persentase kecocokan lebih dari 65%. Dari hasil pengujian tersebut, dapat dikatakan bahwa program aplikasi telah berfungsi dengan baik, karena terbukti program aplikasi ini telah dapat mengenali *file* suara yang diuji.

Sehingga dapat disimpulkan bahwa kegunaan dari program aplikasi pengenalan pembicara ini adalah untuk mengenali suara seseorang dengan cara mengucapkan kata yang sama dengan menggunakan metode *Dynamic Time Warping* (DTW). Namun, berdasarkan hasil pengujian secara keseluruhan, didapatkan keterbatasan dari program aplikasi pengenalan suara pembicara ini, yaitu terbatas pada format *file* suara *.wav yang panjangnya 1 detik dan tidak semua kata yang diucapkan oleh pembicara dapat dikenali dengan baik.

IV. KESIMPULAN

Berdasarkan pembahasan dan pemaparan sebelumnya, maka didapat kesimpulan

dari Aplikasi Pengenalan Suara Pembicara Menggunakan Metode *Dynamic Time Warping* , sebagai berikut.

1. Metode *Dynamic Time Warping* cukup berhasil dalam mengidentifikasi suara, sehingga dapat digunakan sebagai sistem keamanan yang berguna untuk meningkatkan keamanan suatu sistem dengan mengharuskan seseorang untuk mengucapkan suatu kata rahasia disamping mengetik kode atau menggunakan sidik jari.
2. Pada aplikasi ini dibutuhkan tahap pelatihan (*training*) sebelum tahap identifikasi, karena jika tidak program akan meminta suara yang akan diidentifikasi melalui tahap pelatihan (*training*) terlebih dahulu.
3. Sebuah sistem pengenalan suara dapat dibangun dengan menggunakan analisis LPC sebagai pengekstraksi *feature* dan *Dynamic Time Warping* sebagai proses pengenalan atau pencocokan, dengan batas persentase keakuratan suara minimum adalah 65% yang berasal dari suara sepuluh orang berbeda yang mengucapkan namanya masing-masing.
4. Keadaan lingkungan dan nada pengucapan seseorang pada saat dilakukan perekaman suara sangat mempengaruhi tingkat keberhasilan dalam mengidentifikasi suara. Semakin berisik keadaan lingkungan pada saat perekaman suara maka tingkat keberhasilan pengenalan semakin rendah.

V. DAFTAR PUSTAKA

1. MIT.DynamicTimeWarping&Search.
<http://www.twocw.net/MIT/NR/rdonlyres/electrical-engineering-and-computer-science/6-345Automatic-Speech-RecognitionSpring2003/3FA11507-BE7B-429A-A374-D95F79DA7570/0/lecture9.pdf>,06 Desember 2006.
2. Nugroho, Eko. 2008."Biometrika Mengenal Sistem Indentifikasi Masa Depan". Yogyakarta : Penerbit Andi.
3. Prototipe(<http://www.asephs.web.ugm.ac.id/Artikel/RPL/RPL.pdf>) Keogh, Eamonn J. dan Pazzani, Michael J. 2005. "*Derivative Dynamic Time Warping*."http://www.siam.org/meetings/sdm01/pdf/sdm01_01.pdf.
4. Picone, Joseph. 2005. *ECE 8463: Fundamentals Of Speech Recognition*.
http://www.isip.msstate.edu/resources/courses/ece_8463.html
5. Thiang, Hadi Saputra. 1999. "Sistem Pengenalan Kata dengan Menggunakan Linear Predictive Coding dan Nearest Neighbor Classifier". Jakarta: Universitas Kristen Petra.