VOLUME 2 NO.1, Januari 2018 ISSN: 2580-4952

GENERATION JOURNAL

Departement Of Informatics Engineering

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS NUSANTARA PGRI KEDIRI
JL. KH. ACHMAD DAHLAN GG.1 MOJOROTO NO. 6 KEDIRI

Generation Journal

Vol. 2 No : 1 Januari 2018

e-ISSN : 2549-2233
p-ISSN : 2580-4952

Tim Redaksi Generation Journal

Teknik Informatika Fakultas Teknik Universitas Nusantara PGRI Kediri

Pembina	Dr. Suryanto, M.Si.
Penanggung Jawab	Ahmad Bagus S, ST.,M.M., M.Kom.
Pimpinanan Redaksi	Resty Wulanningrum, M.Kom
Penyunting Pelaksana	1) Fitroh Amaluddin, S.T., M.T. (Universitas Ronggolawe Tuban) 2) Diema Hernyka Satyareni, M.Kom. (Universitas Pesantren Tinggi Darul Ulum Jombang) 3) Danar Putra Pamungkas, M. Kom. (Universitas Nusantara PGRI Kediri) 4) Intan Nur Farida, M.Kom. (Universitas Nusantara PGRI Kediri) 5) Ratih Kumalasari N.,S.ST., M.Kom. (Universitas Nusantara PGRI Kediri)
Mitra Bestari	1) Dr. Kusrini, M.Kom. (Universitas AMIKOM Yogyakarta) 2) Rosida Vivin Nahari, S.Kom, MT.(Universitas Trunojoyo Madura)
Layout Desain Grafis	Fajar Rohman Hariri, M.Kom
Publikasi	Risky Aswi Ramadhani, M.Kom Ardi Sanjaya, M. Kom
Alamat	Jl. KH. Ahmad Dahlan Gg. 1 Mojoroto, Kota Kediri Kampus 2 Univ. Nusantara PGRI Kediri. Telp (0354)771576 Fax.771503 Kediri. Web : http://ojs.unpkediri.ac.id/index.php/gj Email : -

Kata Pengantar

Puji Syukur Alhamdulillah kepada Tuhan Yang Maha Esa atas limpahan berkat, rahmat dan kesehatan sehingga jurnal Program Studi Teknik Informatika "Generation Journal" Universitas Nusantara PGRI Kediri, **Vol 2 No. 1 Januari 2018** dapat terselesaikan dengan baik.

Dari buku jurnal ini diharapkan adanya tukar menukar informasi perihal perkembangan dan pemanfaatan, pengembangan kemampuan di bidang Teknologi Informasi, serta menjadi sebuah forum pertukaran informasi antar para pakar, peneliti dan pelaku industri.

Semoga penerbitan buku jurnal Program Studi Teknik Informatika "Generation Journal" ini dapat menjadi acuan informasi yang bermanfaat bagi seluruh staf pengajar khususnya, dan masyarakat pada umumnya.

Kami mengucapkan terima kasih yang sebesar- besarnya kepada semua pihak yang terlibat dalam penyelesaian jurnal ini.

Kediri, 25 Januari 2018

Daftar Isi

Voice Recognition untuk Sistem Keamanan PC Menggunakan Metode MFCC dan DTW

Destian Tri Handoko, Patmi Kasih
[1,2]Teknik Informatika, Fakultas Teknik, Universitas Nusantara PGRI Kediri
E-mail: *[1]*destian3handoko@gmail.com*, [2]fatkasih@gmail.com

Abstrak *- Teknologi berbasis ukuran pada tubuh manusia (disebut dengan istilah biometrik) seperti sidik jari, wajah, kornea mata dan lain-lain digunakan untuk keperluan keamanan, salah satunya untuk keamanan sistem PC. Sistem keamanan komputer merupakan upaya yang dilakukan untuk mengamankan kinerja, data, fungsi atau proses komputer. Sistem keamanan PC juga berguna untuk menjaga dari user yang tidak memiliki otoritas. Layaknya gembok kunci dalam rumah yang menjaga rumah dari pencuri masuk, sistem keamanan menggunakan suara (sistem speech recognition) untuk mengunci desktop dari orang yang tidak memiliki otoritas. Nilai amplitudo diambil dari sinyal suara masukan, sehingga didapatkan kumpulan angka real yang menjadi nilai masukan untuk ekstraksi ciri. Metode ekstraksi ciri yang digunakan dalam sistem ini adalah Mel Frequency Cepstral Coefficient (MFCC). Tahapan awal, MFCC memecah nilai amplitudo sinyal masukan menjadi frame-frame yang diolah dengan menggunakan mel-filterbak yang diadaptasi dari cara kerja pendengaran manusia. Hasil ekstraksi ciri kemudian dibuat menjadi vektor yang digunakan sebagai inputan simbol pada DTW (Dynamic Time Warping) untuk membandingkan hasil vector MFCC. Ketika pengujian ciri dari sinyal uji yang telah dikuantisasi kemudian dicocokan dengan data training yang telah dimasukan pada tahap penyimpanan, sehingga kata sandi dapat dikenali. Dari hasil pengujian, sistem dapat mengenali suara yang memiliki otoritas dengan kriteria dalam keaadaan noise 82% dan hening 86% dengan jumlah 10 data training dan diuji coba sebanyak 50x percobaan.*

Kata Kunci: *DTW, MFCC, Keamanan, PC, Voice Recognition.*

Abstract – *Size-based technology in the human body (called biometric terminology) such as fingerprint, face, cornea and others are used for security purposes, one for PC system security. Computer security system is an effort made to secure the performance, data, function or computer process. PC security systems are also useful for keeping from unauthorized users. Like a lock in a house lock that guards the house from thieves in, the security system uses the voice (speech recognition system) to lock the desktop from people who do not have the authority. The amplitude value is derived from the input sound signal, so that there is a collection of real numbers that become the input value for feature extraction. The method of feature extraction used in this system is Mel Frequency Cepstral Coefficient (MFCC). Initial stage, MFCC breaks the input signal amplitude into frames that are processed by using mel-filterbak adapted from the workings of human hearing. The characteristic extraction results are then made into vectors that are used as symbol inputs in DTW (Dynamic Time Warping) to compare MFCC vector results. When the characteristic test of the quantized test signal is then matched with the training data entered at the storage stage, the password can be identified. From the test results, the system can recognize sounds that have authority with the criteria in the absence of noise 82% and 86% silence with the number of 10 training data and tested as many as 50x experiments.*

Keywords: DTW, MFCC, Security, PC, Voice recognition.

1. PENDAHULUAN

Setiap manusia mempunyai ciri khas yang berbeda, salah satu ciri khas manusia adalah tipe atau warna suara. Banyak teknologi yang berbasis ukuran yang ada pada tubuh manusia (istilah: biometrik) seperti sidik jari, wajah, kornea mata dimanfaatkan dalam bidang keamanan. Dalam perkembangannya, komputer (‘*to compute’* yang berarti berhitung) bukanlah semata-mata sebagai alat hitung saja tetapi adalah suatu alat hitung dengan konstruksi elektronika yang mempunyai tempat penyimpanan (*storage internal*) dan bekerja dengan bantuan sistem operasi (*operating system*) menurut program-program yang diberikan kepadanya. Sistem keamanan komputer

merupakan sebuah upaya yang dilakukan untuk mengamankan kinerja, data, fungsi atau proses komputer. Sistem keamanan komputer dirasa masih sangat lemah. Seseorang yang memiliki privasi yang tinggi memerlukan tingkat keamanan yang lebih terhadap perangkat komputer yang dimiliki, maka diperlukan suatu sistem yang mampu mengidentifikasi pemilik komputer. Salah satunya dengan cara pengenalan suara (*voice recognition).*

2. METODE PENELITIAN

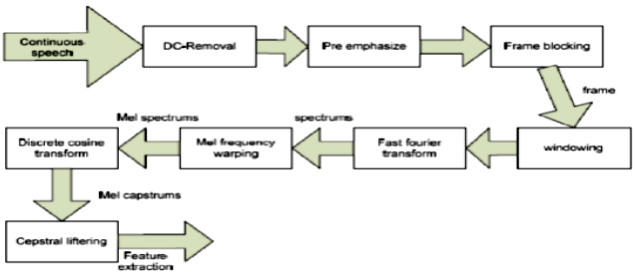
Secara umum metode penelitian yang dilakukan adalah tahapan-tahapan untuk memperoleh data dan informasi yang berhubungan dengan sistem *voice recognition*.

2.1. Analisa Algoritma

Untuk mengetahui dan memastikan cara kerja algoritma yang akan digunakan dalam pembuatan sistem, maka dilakukan analisa terhadap metode (algoritma) tersebut.

2.1.1 Mel Frequency Cepstrum Coefficients (MFCC)

Metode MFCC adalah salah satu metode yang digunakan untuk melakukan *feature extraction* (ekstraksi ciri), sebuah proses yang mengkorversikan sinyal suara menjadi beberapa parameter. Adapun tahap-tahap pengolahan suara menggunakan metode MFCC ini dapat dilihat gambar berikut.



Gambar 1. Blok diagram MFCC

1. *DC Removal*, berfungsi menghitung nilai rata-rata dari data sample suara, dan mengurangkannya untuk setiap sample pada window. Tujuannya adalah untuk memperoleh normalisasi dari data suara *input*.

   Rumus:

$$D[i] = s[i] - \frac{\sum_{i=1}^{n} s[i]}{n} \quad \text{..... (1)}$$

   Keterangan:

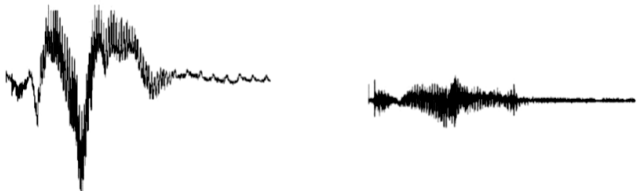
   D[i] = hasil signal ke –i setelah dilakukan DC Removal
   S[i] = signal awal ke-i
   N = jumlah sample, n > 0

2. *Pre-emphasis*, merupakan salah satu jenis filter yang sering digunakan sebelum sebuah signal diproses lebih lanjut. Filter ini mempertahankan frekuensi-frekuensi tinggi pada sebuah spektrum, yang umumnya tereliminasi pada saat proses produksi suara. Tujuan dari preemphasis filter ini adalah:
   a) Mengurangi noise ratio pada sinyal, sehingga dapat meningkatkan kulitas sinyal.
   b) Untuk mendapatkan bentuk spektral frekuensi sinyal wicara yang lebih halus.
   c) Menyeimbangkan spektrum dari sinyal suara.

   Pada saat memproduksi sinyal suara, glotis manusia menghasilkan sekitar -12 dB octave slope. Namun ketika energi akustik tersebut dikeluarkan melalui bibir, terjadi peningkatan sebesar +6 dB. Sehingga sinyal yang terekam oleh microphone adalah sekitar -6 dB octave slope. Dimana bentuk spektral yang relatif bernilai tinggi untuk daerah rendah dan cenderung turun secara tajam untuk daerah fekuensi diatas 2000 Hz. Filter preemphasis didasari oleh hubungan *input/output* dalam domain waktu yang dinyatakan dalam persamaan sebagai berikut:

$$y(n) = s(n) - \alpha s(n-1) \quad \text{..... (2)}$$

Dimana:
y(n) = Sinyal hasil preemphasis
s(n) = Sinyal sebelum preemphasis
α merupakan konstanta filter preemphasis, biasanya bernilai 0.97. Dalam bentuk dasar operator s sebagai unit filter, persamaan diatas akan memberikan sebuah transfer *function filter preemphasis* seperti berikut:



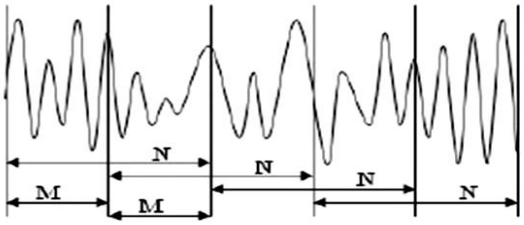
Gambar 2. Sebelum dan sesudah Preemphasis

3. Frame blocking

Frame Blocking adalah pembagian suara menjadi beberapa frame yang nantinya dapat memudahkan dalam perhitungan dan analisa suara, satu frame terdiri dari beberapa sampel tergantung tiap berapa detik suara akan disampel dan berapa besar frekuensi samplingnya. Panjang frame yang digunakan, sangat mempengaruhi keberhasilan dalam analisa spektral. Di satu sisi, ukuran dari frame harus sepanjang mungkin untuk dapat menunjukkan resolusi frekuensi yang baik. Tetapi di lain sisi, ukuran frame juga harus cukup pendek untuk dapat menunjukkan resolusi waktu yang baik. Refresentasi fungsi *frame blocking* sebagai berikut:

$$\text{Jumlah frame} = ((I-N)/M)+1) \quad \text{..... (3)}$$

Dimana: I = Sample rate,
N= Sample point (Sample rate * waktu framing (s))
M = N/2



Gambar 3. Bentuk sinyal yang di frame blocking

Proses framing ini dilakukan terus sampai seluruh signal dapat terproses. Selain itu, proses ini umumnya dilakukan secara overlapping untuk setiap frame-nya. Panjang daerah overlap yang umum digunakan adalah kurang lebih 30% sampai 50% dari panjang frame.

4. Windowing

Suara yang di potong-potong menjadi beberapa frame membuat data suara menjadi discontinue, hal ini mengakibatkan kesalahan data proses fourier transform. Agar tidak terjadi kesalahan data pada proses fourier transform maka sampel suara yang telah dibagi menjadi beberapa frame perlu dijadikan suara kontinu dengan cara mengalikan tiap frame dengan window tertentu. Jenis window yang dipakai pada proses ini adalah window hamming. Berikut ini adalah representasi fungsi window terhadap signal suara yang dimasukan.

$$x(n) = x(n)w(n)$$

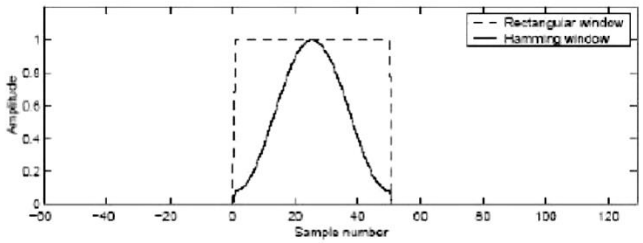
rumus window hamming sebagai berikut:

$$w[n] = 0.54 - 0.46 * \cos(2*\pi*n/(N-1)) \quad \text{..... (4)}$$

dimana: n = 1,2,3,…
π = 3.14

N = panjang frame

Fungsi window yang paling sering digunakan dalam aplikasi voice recognition adalah Window Hamming. Fungsi window hamming ini menghasilkan sidelobe level yang tidak terlalu tinggi (kurang lebih -43 dB), selain itu noise yang dihasilkan pun tidak terlalu besar (kurang lebih 1.36 BINS).



Gambar 4. Bentuk gelombang dari window hamming

5. *Fast Fourier Transform (FFT)*

Fast fourier transform adalah suatu metode yang sangat efisien untuk menyelesaikan transformasi fourier diskrit yang banyak dipakai untuk keperluan analisa sinyal seperti pemfilteran, analisa korelasi, dan analisa spektrum. Ada 2 (dua) jenis algoritma FFT yaitu algoritma fast fourier transform decimation in time (FFT DIT) dan algoritma fast fourier transform decimation in frequency (FFT DIF). Di mana pada FFT DIT masukan disusun/dikelompokan menjadi kelompok ganjil dan kelompok genap. Sedangkan pada FFT DIF masukan tetap, tetapi *output* disusun/dikelompokkan menjadi kelopmpok ganjil dan kelompok genap.

Bentuk rumusannya sebagai berikut:

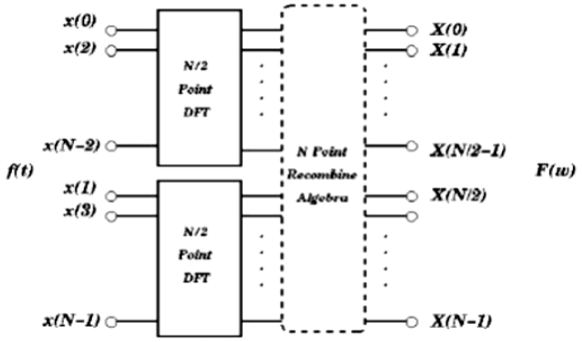
$$X[k]=\sum_{n=1}^{N=1} x(N)w_N^{kn} \quad \text{..... (5)}$$

Dimana :
- X[k] = Merupakan magnitude frekuensi
- x(n) = Nila sampel sinyal
- W = cos(2*Pi*k*n/N)-jSin(2*Pi*k*n/N)
- N = jumlah sinyal yang akan diproses
- k = sinyal yang diproses

Gunakan rumus :

$$||[R^2 + I^2]^{\frac{1}{2}}| \quad \text{..... (6)}$$

FFT dilakukan dengan membagi N buah titik pada transformasi fourier diskrit menjadi 2, masing-masing (N/2) titik transformasi. Proses memecah menjadi 2 bagian ini diteruskan dengan membagi (N/2) titik menjadi (N/4) dan seterusnya hingga diperoleh titik minimum.



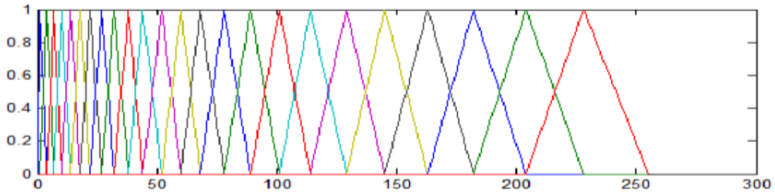
Gambar 5. Pembagian Sinyal Suara Menjadi Dua Kelompok

6. *Mel Frequency Warping*

Persepsi sistem pendengaran manusia terhadap frekuensi sinyal suara tidak hanya bersifat linear. Penerimaan sinyal suara untuk frekuensi rendah (<1k Hz) bersifat linear, dan untuk frekuensi tinggi (>1k Hz) bersifat logaritmik. Jadi, untuk setiap nada dengan frekuensi sesungguhnya , sebuah pola diukur dalam sebuah skala yang disebut "mel" (berasal dari Melody). Skala ini didefinisikan sebagai :

$$H_i = \begin{Bmatrix} 2595 \times log_{10}\left(1 + \frac{F_{Hz}}{700}\right), F_{Hz} > 1000 \\ F_{Hz} \qquad\qquad , F_{Hz} < 1000 \end{Bmatrix} \quad \text{..... (7)}$$

Sebuah pendekatan untuk simulasi spektrum dalam skala mel adalah dengan menggunakan *filter bank* dalam skala mel seperti yang ditunjukkan pada gambar di bawah ini dimana setiap frame yang diperoleh dari tahapan sebelumnya difilter menggunakan M filter segitiga sama tinggi dengan tinggi satu.



Gambar 6. Mel Filter Bank dengan 24 buah filter

Dalam mel-frequency wrapping, sinyal hasil FFT dikelompokkan ke dalam berkas filter triangular ini. Proses pengelompokan tersebut adalah setiap nilai FFT dikalikan terhadap filter yang bersesuaian dan hasilnya dijumlahkan. Proses wrapping terhadap sinyal dalam domain frekuensi dilakukan menggunakan persamaan berikut.

$$Y[i]=\sum_{j=1}^{N} s(j).H_i(j) \quad \text{..... (8)}$$

Dimana: N = jumlah magnitude spectrum
S[j] = magnitude spectrum pada frekuensi j
$H_i(j)$ = koefisien *filterbank* pada frekuensi j($1 \leq i \leq M$)
M= jumlah Channel dalam filterbank

Untuk mendapatkan Hi digunakan rumus pada nomor 6.

7. *Discrete Cosine Transform (DCT)*

Cepstrum biasa digunakan untuk mendapatkan informasi dari suatu sinyal suara yang diucapkan oleh manusia. Pada tahap terakhir pada MFCC ini, spektrum log mel akan dikonversi menjadi domain waktu menggunakan *Discrete Cosine Transform (DCT)* menggunakan persamaan berikut.

$$c_j = \sum_{i=1}^{M} X_i.\cos\left(\frac{j(i-1)}{2}.\frac{\pi}{M}\right) \quad \text{..... (9)}$$

Dimana: Ci = nilai koefisien $Ckej$
j = 1,2,… jumlah koefisien yang diharapkan
Xi = nilai X hasil *mel-frequency wrapping* pada frekuensi
i = 1,2,…,n (jumlah wrapping)
M= jumlah filter

Hasil dari proses ini dinamakan *Mel-Frequency Cepstrum Coefficients (MFCC).*

8. *Cepstral liftering*

Hasil dari fungsi DCT adalah *cepstrum* yang sebenarnya sudah merupakan hasil akhir dari proses *feature extraction*. Tetapi, untuk meningkatkan kualitas pengenalan, maka cepstrum hasil dari DCT harus mengalami *cepstral liftering*

$$w[n] = \left\{1 + \frac{L}{2}\sin\left(\frac{n\pi}{L}\right)\right\} \quad \text{..... (10)}$$

L = jumlah cepstral coefficients
N = index dari cepstral coefficients

9. Remove Silence

Dengan demikian selesailah proses Mel Frequency Cepstrum Coefficients (MFCC) feature extraction. Namun hasilnya masih mengandung frame-frame silence. Frame semacam ini dapat menggangu pengenalan kata. Oleh karena itu sebaiknya frame-frame ini harus dihilangkan. Frame energy sendiri didapatkan dari frame ke-0 dari hasil feature extraction. Energy tersebut akan dijadikan acuan untuk apakah sebuah frame tergolong sebagai suatu noise. Apabila energi suatu frame kurang dari nilai rata-rata, maka tergolong sebagai noise dan akan segera dihilangkan. Perhitungan noise removal ini adalah sebagi berikut:

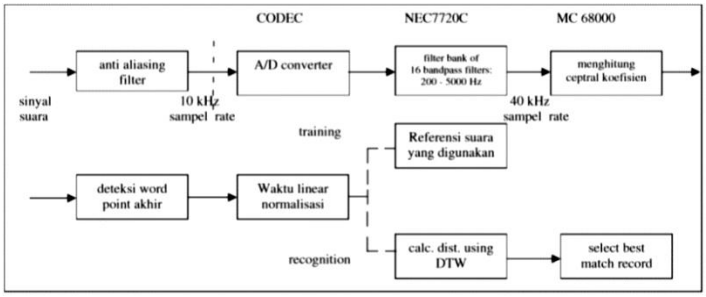
$$Noise = \frac{\sum_{i-1}^{n} s[i]}{n} \quad \text{..... (11)}$$

Noise = rata-rata energy frame, S[i] = signal frame ke-i

2.1.2 Dynamic Time Warping (DTW)

Satu masalah yang cukup rumit dalam pengenalan suara adalah proses perekaman yang terjadi seringkali berbeda durasinya, meskipun kata atau kalimat yang diucapkan sama. Bahkan untuk satu suku kata/ vokal yang sama seringkali terjadi dalam durasi yang berbeda. Sebagai akibatnya proses pencocokan antara sinyal uji dengan sinyal record suara di database seringkali tidak menghasilkan nilai yang optimal.

Sebuah teknik yang cukup popular di awal perkembangan teknologi pengolahan sinyal pengenalan suara (*voice recognition*) adalah teknik *Dynamic Time Warping (DTW)* yang juga lebih dikenal sebagai *dynamic programming*. Teknik ini ditujukan untuk mengakomodasi perbedaan waktu antara proses perekaman saat pengujian dengan suara yang tersedia pada *record* suara di *database*. Prinsip dasarnya adalah dengan memberikan sebuah rentang 'steps' dalam ruang (dalam hal ini sebuah *frame-frame* waktu dalam sampel, *frame-frame* waktu dalam *database*) dan digunakan untuk mempertemukan lintasan yang menunjukkan *local match* terbesar (kemiripan) antara *time frame* yang lurus. Total *`similarity cost'* yang diperoleh dengan algoritma ini merupakan sebuah indikasi seberapa bagus sampel suara dan *record* suara di *database* ini memiliki kesamaan, yang selanjutnya akan dipilih *best-matching record* suara di *database*.



Gambar 7. Blok Diagram dari Sistem *Recognition* dengan Metode DTW.

2.1.3 Dynamic Time Warping Algorithm

Dynamic Time Warping algorithm adalah algortima yang menghitung optimal warping path antara dua waktu. Algoritma ini menghitung baik antara nilai warping path dari dua waktu dan jaraknya. Misalnya, memiliki dua barisan numerik:

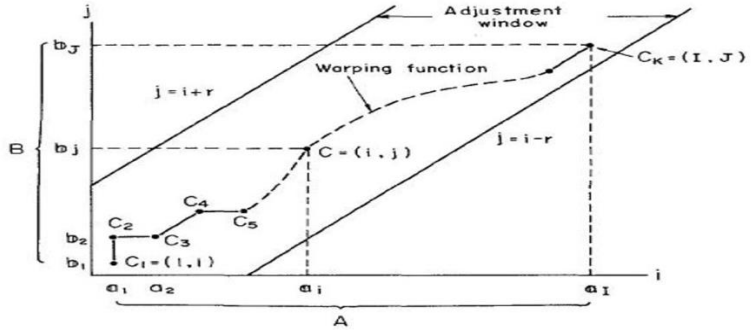
$$\{A\} = (a1, a2, \ldots, an)$$
$$\{B\} = (b1, b2, \ldots, bn) \quad \text{..... (12)}$$

Dengan pemisalan ini, maka dapat dikatakan bahwa panjang dua barisan ini bisa saja berbeda. Dimana masing-masing barisan A dan B dikembangkan sepanjang sumbu i dan sumbu j. Dimana pola suara ini adalah dari pola suara yang sama, perbedaan waktu antara barisan A dan B dapat digambarkan berdasarkan urutan point c= (i,j) :

$$F = c(l), c(2), \ldots, c(k) \quad \text{..... (13)}$$

Dimana: c(k) = (i(k), j(k))

Urutan ini dapat dianggap mewakili suatu fungsi pemetaan dari waktu barisan A dan waktu barisan B, hal itu disebut warping path sebuah fungsi ketika tidak ada perbedaan waktu.



Gambar 8. Fungsi Warping dan Definisi Pengaturan Windows

Barisan fungsi warping bertepatan dengan diagonal baris j = i, ini menyimpang jauh dari garis diagonal sebagai perbedaan waktunya.

$$[i(k) - j(k)] \leq r \quad \text{..... (14)}$$

Dimana r adalah bilangan bulat positif yang tepat disebut signal-window, k adalah jumlah titik sinyal warping path. Kondisi ini sesuai dengan fakta bahwa fluktuasijarak waktu dalam kasus-kasus biasa tidak pernah ada perbedaan jarak waktu terlalu berlebihan. Algoritma ini memulai dengan penghitungan jarak lokal antara elemen dari barisan menggunakan tipe jarak yang berbeda. Frekuensi yang paling banyak menggunakan metode untuk penghitungan jarak adalah jarak absolut antar nilai dua elemen. Jika dalam matriks maka dapat ditulis dengan memiliki i garis dan j kolom, secara umum:

$$d(c) = d(i,j) = |ai - bj| \quad \text{..... (15)}$$

Mulai dengan matrik jarak lokal, kemudian minimum jarak pada warping F menjadi fungsi sebagai berikut :

$$E(F)=\sum_{k=1}^{K} d\big(c(k)\big).w(k) \quad \text{..... (16)}$$

Dimana w(k) adalah koefesien bobot non-negatif yang diperkenalkan untuk memungkinkan E (F) mengukur fleksibel karakteristik) dan merupakan jarak yang minimum untuk fungsi warping F. Pencapaian nilai yang minimum saat F fungsi warping ditentukan perbedaan waktu menjadi optimal. Nilai jarak minimum ini dapat dianggap sebagai jarak antara barisan A dan B, masih tersisa setelah eliminasi perbedaan waktu antara barisan A dan B, dan tentu saja diperkirakan waktu akan stabil dengan sumbu fluktuasi. Berdasarkan pertimbangan ini, waktu normalisasi jarak antara dua barisan suara A dan B didefinisikan sebagai berikut:

$$D(A,B) = \min_{F}\left[\frac{\sum_{K=1}^{K} d(c(k)).w(k)}{\sum_{K=1}^{K} w(k)}\right] \quad \text{.....(17)}$$

Dimana penyebut ∑w(k) yang digunakan untuk mengkompensasi pengaruh K (jumlah titik pada fungsi F warping). Persamaan (9) tidak lebih dari sebuah fundamental definisi jarak waktu normal. Efektif karakteristik ini sangat tergantung pada pengukuran spesifikasi fungsi warping path dan definisi koefisien pembobotan. Pengukuran karakteristik dari normalisasi jarak waktu akan bervariasi, menurut sifat barisan suara (terutama waktu sumbu ekspresi barisan suara) yang harus diselesaikan. Oleh karena itu, masalah sekarang terbatas pada kasus yang paling umum di mana ada dua kondisi sebagai berikut:

1. Kondisi barisan suara adalah waktu sampel yang umum dan periode sampling konstan.
2. Kondisi dimana pengetahuan tentang bagian dari barisan suara berisi informasi yang penting.

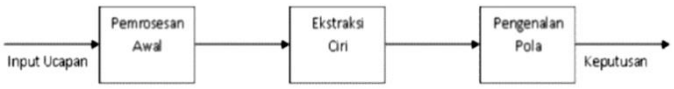
Dalam hal ini, adalah wajar untuk mempertimbangkan setiap bagian dari barisan suara mengandung jumlah informasi linguistik yang sama. Dimana w(k) adalah elemen yang dimiliki warping path dan k adalah jumlahnya. Penghitungannya dibuat untuk dua barisan diperlihatkan pada gambar dibawah dan warping path diberi highlight.

	-2	10	-10	15	-13	20	-5	14	2
3	5	12	25	37	53	70	78	89	90
-13	16	28	15	43	37	70	78	105	104
14	32	20	39	16	43	43	62	62	74
-7	37	37	23	38	22	49	45	66	71
9	48	38	42	29	44	33	47	50	57
-2	48	50	46	46	40	55	36	52	54

Gambar 9. Warping path

2.2 Desain Sistem

Sistem yang dirancang untuk pengamanan PC dengan pengenalan suara *(voice recognition)* ini dirancang dengan konsep yang cukup sederhana. *Voice recognition* (*speaker recognition*) adalah suatu proses untuk mengenali seseorang dengan mengenali suara dari orang tersebut. Dengan memanfaatkan fasilitas *audio speaker* pada PC untuk menerima suara yang akan diproses dalam sistem yang telah ditanam dalam PC, yaitu *Automatic speaker recognition* merupakan penggunaan sebuah mesin yang mengenali seseorang dari sebuah frasa yang diucapkan. Sistem ini dapat berfungsi dalam dua buah mode yaitu mengenali seseorang yang khusus atau membuktikan identitas yang diklaim oleh seseorang. Secara sederhana sistem *voice recognition* untuk keamanan PC ini dapat digambarkan dengan diagram pada gambar 10.

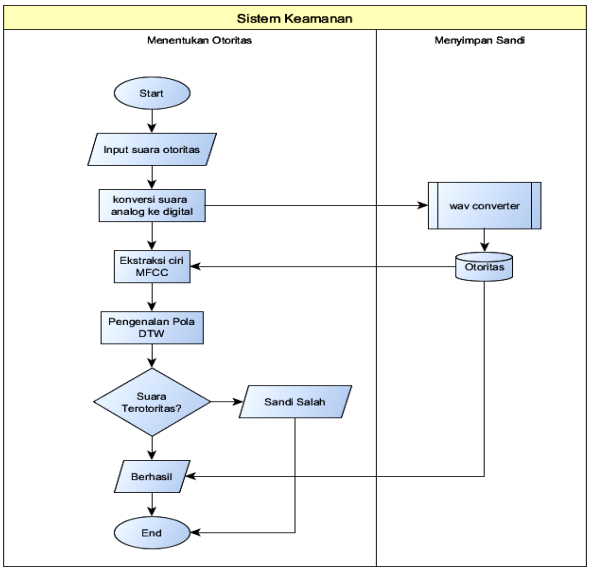


Gambar 10. Tahapan dalam Vaoice Recognition.
Sumber: Rachman (2006)

2.2.1 Rancangan Input

Setelah aplikasi terpasang pada sistem operasi, maka perlu dilakukan pengaturan data otoritas, dimana suara pemiliki PC/ sistem direkam menggunakan microphone dengan cara menekan tombol tertentu melalui interface yang sudah disediakan dan disimpan di database. Setelah aplikasi sistem keamanan terpasang dan diatur data otoritasnya, maka sistem keamanan siap diguanakan. Pada saat ingin digunakan, perangkat pc perlu dilakukan restart atau sleep dan dinyalakan lagi. Setelah itu masuk interface keamanan sistem dimana orang yang tidak terotoritas dan terotoritas ingin masuk. Pada gambar 5.1 akan dijelaskan proses *input* guna menentukan otoritas pemilik sitem.

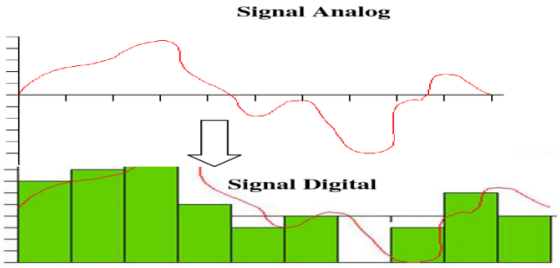
2.2.2 Rancangan Output



Gambar 11. Flowchart Menentukan Otoritas Sistem Dan Menyimpan Sandi

Pada saat *input* suara, suara masih berupa sinyal analog dimana sistem tidak dapat mengidentifikasinya, maka perlu dilakukan konversi suara menjadi digital dalam bentuk format ".WAV".

Suara digital/ audio digital merupakan versi digital dari suara analog. Pengubahan suara analog menjadi suara digital membutuhkan suatu alat yang disebut *Analog To Digital Converter* (ADC), yaitu pengubah amplitudo sebuah gelombang analog ke dalam waktu interval (sampling) sehingga menghasilkan representasi digital dari suara. Sampling adalah melakukan pencuplikan amplitudo gelombang suara pada tiap satu satuan waktu.



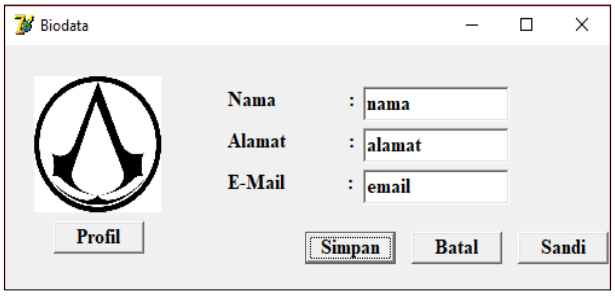
Gambar 12. Gelombang suara analog menjadi digital

Audio yang dihasilkan dari ADC akan berformat WAV. WAV (*Wave form audio*) adalah format audio standar Microsoft dan IBM untuk PC. WAV biasanya menggunakan coding PCM *(Pulse Code Modulation*). WAV adalah data tidak terkompres sehingga seluruh sampel audio disimpan semuanya di harddisk. Sofware yang dapat menciptakan WAV dari analog sound misalnya windows sound recorder. WAV jarang digunakan di internet karena ukurannya yang relative besar. Maksimum ukuran file WAV adalah 2GB.

Setelah data sudah terkonversi maka perlu di ekstraksi ciri menggunakan metode MFCC dan kemudian dicocokan dengan data otoritas yang ada di database menggunakan metode DTW. Setelah itu ditampilkan ke-otoritasanya.

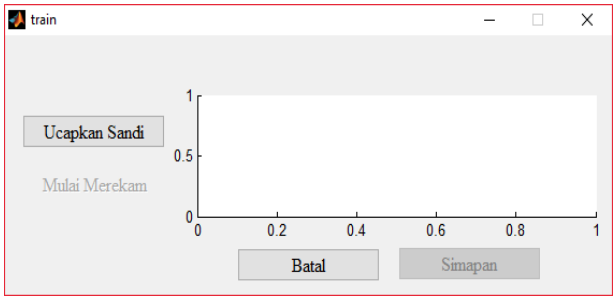
3. HASIL DAN PEMBAHASAN

Sistem keamanan ini dibuat dengan penyimpanan data untuk setiap user yang akan menggunakan sistem/ PC. Setiap user harus mengisi serangkaian data terlebih dahulu untuk pengenalan identitas hingga pengenalan suara yang akan dikonfirmasi sebagai user pengguna PC/ sistem.



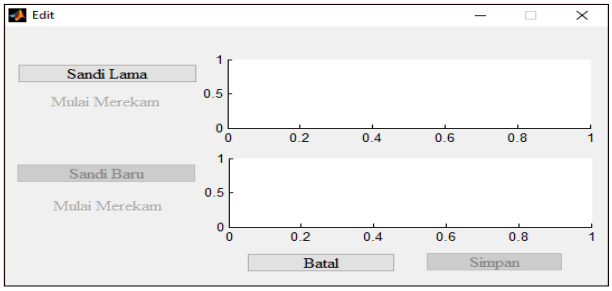
Pada form/ lembar kerja pada gambar 13, user dapat melakukan isi biodata, isi profil dan simpan sandi.
Pada sistem juga disediakan form untuk melakukan ubah profil dan juga ubah sandi.

Gambar 13. Layar Kerja Biodata User



Form pada gambar 14 bertujuan untuk mengatur sandi yang diinginkan user, setelah button ucapkan sandi di klik maka user di haruskan mengucapkan sandi yang diinginkan, dimana batas waktu untuk mengucapkan sandi selam 3 detik.

Gambar 14. Form untuk Memasukan Sandi



Gambar 15. Form untuk Edit Sandi

Sistem mempersiapkan form untuk mengubah kata sandi bagi user. Untuk mengubah kata sandi, user diharuskan mengucapkan sandi yang pernah ada guna melakukan ferifikasi *user* terotoritas. Jika sandi salah maka button merekam sandi baru tidak akan aktif, jika benar mak button tersebut akan aktif.

Untuk uji coba sistem digunakan contoh 20 suara dari 20 orang. Uji coba dilakukan dengan data *training* suara tidak terotoritas dan suara terotoritas. Uji coba ini dilakukan dalam keadaan hening dan bising/ *noise*. Selanjutnya dilakukan hitung nilai recall dan precition, dengan matrik seperti tabel 1, sebagai berikut:

Tabel 1. Matrik *Recall* dan *Pricision*

	Relevan	Tidak Relevan	Total
Ditemukan	a *(hits)*	b (*noise*)	a + b
Tidak ditemukan	c (*misses*)	d (*rejected*)	c + d
Total	a + c	b + d	a + b + c + d

Untuk mengukur tingkat akurasi sistem dalam menentukan seseorang memiliki otoritas tidaknya dapat menggunakan rumus pada tabel 1. Berdasarkan tabel tersebut, rumus *recall* dan *pricision* menjadi :

Recall = [a/ (a+c)] x 100
Pricision = [a/ (a+b)] x 100 ... (18)

Tabel 2. Hasil Precision Uji Coba

Skenario ke-	**Relevan**	**Tidak Relevan (b)**	**Total (a+b)**	**Total (a+c)**	**Pricision [a/(a+b)]x100%**
1	7	13	20	7	35%
2	3	17	20	3	15 %
3	9	11	20	9	45%
4	4	16	20	4	20%
					29%

Setelah dilakukan hitung nilai recall dan precision didapatkan hasil seperti tabel 2, yang dapat disimpyulkan sebagai berikut:

1. Sistem masih dapat diakses oleh rata-rata 29% orang. Semakin banyak data training maka semakin sedikit pula seseorang yang tidak memiliki otoritas dapat masuk.
2. Tingkat kebisingan atau noise dapat mempengaruhi tingkat keberhasilan seseorang dapat membuka sistem keamanan. Semakin sedikit persentase keberhasilan orang tidak memiliki otoritas membuka sistem keamanan, maka sistem keamanan tersebut semakin baik.
3. Hasil uji coba yang dilakukan user yang memiliki otoritas dengan keadaan hening dengan jumlah 4 data training sebanyak 50x percobaan didapat hasil 88% suara yang dapat membuka sistem keamanan pc.
4. Hasil uji coba yang dilakukan user yang memiliki otoritas dengan keadaan noise dengan jumlah 4 data training sebanyak 50x percobaan didapat hasil 74% suara yang dapat membuka sistem keamanan pc.
5. Hasil uji coba yang dilakukan user yang memiliki otoritas dengan keadaan hening dengan jumlah 10 data training sebanyak 50x percobaan didapat hasil 86% suara yang dapat membuka sistem keamanan pc.
6. Hasil uji coba yang dilakukan user yang memiliki otoritas dengan keadaan noise dengan jumlah 10 data training sebanyak 50x percobaan didapat hasil 82% suara yang dapat membuka sistem keamanan pc Kondisi terbaik untuk user untuk dapat membuka keamanan adalah semakin hening lokasi membuka sistem maka persentase keberhasilan semakin besar.
7. Jika data training semakin besar, tingkat keberhasilan semakin menurun. Agar sistem dapat di implementasikan pada perngkat pc, maka diperlukan pengaturan, antara lain meletakan file hasil compiler gui di directory C:\New_folder kemudian membuat shortcut file password.exe ke directory C:\Users\ user\ AppData\ Roaming\ Microsoft\ Windows\ Start Menu\ Programs\ Startup. Pada bagian yang bergaris bawah menyesuaikan terhadap sistem masing-masing.

4. SIMPULAN

Berdasarkan hasil ujicoba sistem dan analisa hasil yang dilakukan, diperoleh kesimpulan sebagai berikut:

1. Dalam mengenali suara, sistem yang dibuat merubah suara yang direkam oleh microfon selama 3 detik menjadi sinyal digital, kemudaian melalui tahap MFCC guna mengekstrasi ciri sinyal tersebut, setelah itu dicocokan dengan suara yang berada didatabase yang berformat file .wav menggunakan metode DTW. Kondisi ini hanya berlaku jika saat sistem keamanan yang dibuat aktif.
2. Kondisi yang paling baik sistem ini diimplementasikan pada saat keadaan hening dengan akurasi 86% dengan jumlah 10 data training. Apabila persentase keberhasilan seseorang (tidak memiliki otoritas) yang dapat membuka sistem keamanan yang dibuat dengan suara yang tidak terotoritas besar, maka sistem memerlukan perbaikan dan masih jauh dari keberhasilan. Begitu pula sebaliknya, apabila persentase keberhasilan seseorang (tidak memiliki otoritas) yang dapat membuka sistem keamanan yang dibuat dengan suara yang tidak terotoritas kecil, maka sistem sudah mendekati keberhasilan. Sistem yang dibuat dapat diimplementasikan pada system operasi windows 8.1 64 bit dengan spesifikasi ram 3 GB dan prosesor dualcore.

5. SARAN

Saran dari penulis untuk pembangunan sistem keamanan menggunakan suara ini lebih lanjut adalah:

1. Sistem diharapkan dapat terkoneksi dengan perangkat smart phone melaui jaringan internet, jadi apabila ada seseorang yang tidak bertanggungjawab mengakses sistem, maka sistem akan mengirim notifikasi berupa pesan.
2. Sistem diharapkan mampu diterapkan pada sistem operasi lain.

DAFTAR PUSTAKA

[1] Effendi, Zulham. dan Firdaus, 2015. Pengenalan Suara Menggunakan Metode MFCC *(Mel Frequency Cepstrum Coefficients)* dan DTW *(Dynamic Time Warping)* untuk Sistem penguncian Pintu. *Seminar Nasional Teknologi Informasi dan Komunikasi Terapan (SEMANTIK)*, 2 (1), tersedia: http://komputa.if.unikom.ac.id, diunduh 25 Januari 2017.

[2] Chamidy, Totok. 2015. *Metode Mel Frequency Cepstral Coeffisients* (MFCC) Pada klasifikasi *Hidden Markov Model* (HMM) Untuk Kata *Arabic* pada Penutur Indonesia. *Jurnal MATICS*, 2 (1), tersedia: http://download.portalgaruda.org, diunduh 25 Januari 2017.

[3] Darmadi, F., Rizal, A., dan Sunarya, Unang. 2015. Deteksi *Sleep Apnea* Melalui Analisis Suara Dengkuran Dengan *Metode Mel Frekuensi Cepstrum Coefficient. e-Proceeding of Engineering*, 2 (2), tersedia: https://repository.telkomuniversity.ac.id, diunduh 25 Januari 2017.

[4] Kadir, Abdul (2001). *Dasar Pemrograman Delphi.* Yogyakarta: C.V. ANDI OFFSET.

[5] Manunggal, HS. 2005. Perancangan dan Pembuatan Perangkat Lunak Pengenalan Suara Pembicara dengan Analisa MFCC Feature Extraction. Surabaya : Universitas Kristen Petra.

[6] Muttaqin, I., Usman, K. dan Atmaja, R.T. 2013. *Simulasi dan Analisis Identifikasi Alat Musik Tradisional Berdasarkan Nada Bunyi Dengan Metode Mel Frequency Cepstrum Coefficient(MFCC) dan Support Vector Machine (SVM).* (Online). tersedia: https://openlibrary.telkomuniversity.ac.id, diunduh 25 Januari 2017.

[7] Nursyeha, M. A., Rivai, M. dan Suwito. 2015. *Pengenalan Suara Burung Menggunakan Mel Frequency Cepstrum Coefficient dan Jaringan Syaraf Tiruan pada Sistem Pengusir Hama Burung*. (Online). tersedia: http://digilib.its.ac.id/, diunduh 25 Januari 2017.

[8] Pangeran, A. A. (2010). *SISTEM OPERASI.* Yogyakarta: C.V ANDI OFFSET.

[9] Puspitasari, M., Supardi, J. dan Sazaki, Y. 2013. *Pengenalan Suara Menggunakan Mel Frequency Cepstral Coefficients Dan Self Organizing Maps.* (Online). tersedia: http://eprints.unsri.ac.id, diunduh 25 Januari 2017.

[10] Putra, Darma. dan Resmawan, Adi. 2011. Verifikasi Biometrika Suara Menggunakan Metode MFCC Dan DTW. *LONTAR KOMPUTER*, (Online), 2 (1): 8-21, tersedia: http://www.it.unud.ac.id, di unduh 22 Januari 2017.

[11] Riyanto, Eko, 2013, Sistem Pengenalan Pengucap Manusia Dengan Ekstraksi Ciri MFCC Dan Algoritma Jaringan Saraf Tiruan Perambatan Balik Sebagai Pengenalanya, JSIB.

[12] Tanudjaja, Harlianto. 2007. Pengolahan Sinyal Digital dan Sistem Pemrosesan Digital. Andi Offset : Yogyakarta.