Disain Sistem Pengenalan Suara sebagai Pengendali Dinamo *Starter* pada Otomobil

Rr. Sri Poernomo Sari[1] , Rezza Aditya[2]

[1,2]Jurusan Teknik Mesin, Fakultas Teknologi Industri
Universitas Gunadarma, Jakarta, Indonesia
Email: sri_ps@staff.gunadarma.ac.id

Abstrak

Teknologi pengenalan suara (s*peaker recognition*) merupakan aplikasi dalam pengolahan sinyal digital. Penelitian ini bertujuan merancang prototipe sistem pengenalan suara sebagai pengendali dinamo *starter* pada otomobil. Pengolahan sinyal yang dibahas adalah *speaker verification.* Metode yang digunakan adalah MFCC (*Mel Frequency Cepstrum Coefficients*) untuk proses ekstraksi ciri dari sinyal suara dan DTW (*Dynamic Time Warping*) untuk proses pencocokan. Data masukan suara adalah 1 buah kata sandi dan 1 buah kata perintah. Perancangan menggunakan modul Parallax Say It, mikrokontroller AVR ATMega16, LCD 16x2, motor *driver* dan baterai 12V. Hasil pengujian adalah tingkat akurasi paling rendah 70% nilai *threshold* 3,5 sedangkan tingkat akurasi tertinggi 82.5 % dengan nilai *threshold* 5,3.

Keywords: Pengenalan Suara, MFCC, DTW, Mikrokontroler, Parallax Say It

Pendahuluan

Komunikasi dilakukan manusia dengan sesama manusia dalam kehidupan sehari-hari, misalnya berbicara (*speech*). Berbicara (*speech*) memberikan informasi penting dan efektif antara lain tentang keadaan kesehatan, emosi, *gender* serta identitas pembicara.

Kemajuan teknologi bidang pengolahan sinyal digital (*Digital Signal Processing*) telah digunakan dalam berbagai aplikasi teknik pengenalan suara, kompresi sinyal berupa data dan gambar, televisi serta telepon digital [1].

Suara merupakan bagian dari tubuh manusia yang unik atau khas yang dapat dibedakan dengan mudah. Sistem biometrika suara memiliki karakteristik yaitu tidak dapat lupa, tidak mudah hilang dan tidak mudah untuk dipalsukan karena keberadaannya melekat pada diri manusia. Teknologi pengenalan suara *(speaker recognition)* merupakan teknologi biometrika yang tidak memerlukan biaya besar dan peralatan khusus.

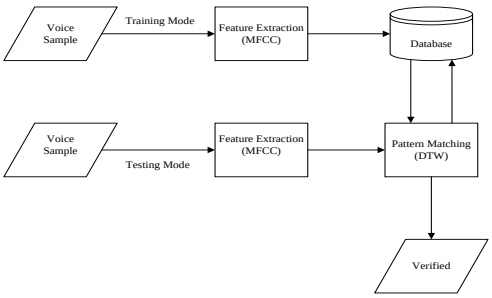
Tujuan penelitian ini adalah membuat prototipe sistem pengenalan suara untuk menggerakkan dinamo starter pada mobil. Prototipe ini memanfaatkan teknologi pengenalan suara (s*peaker recognition*) menggunakan modul *Parallax Say It* sebagai pemroses pengolahan sinyal suara yang diteruskan ke Mikrokontroler ATMega16. Sistem ini diharapkan mengenali suara pengguna yang hasilnya digunakan sebagai kata sandi dan perintah menggerakkan dinamo *starter* untuk menjalankan mobil.

Metode Penelitian

Penelitian dimulai dengan melakukan perancangan alat kemudian proses pembuatan prototipe. Mikrokontroler yang digunakan adalah jenis mikrokontroler AVR ATmega16 yang menggunakan bahasa C sebagai bahasa pemrograman dan diadaptasikan pada *software Code Vision* AVR. Analisis sinyal dilakukan dengan ekstraksi ciri menggunakan metode MFCC (*Mel-Frequency Ceptrums Coefficients*). Data masukan suara berupa 1 buah kata sandi dan 1 buah kata perintah. Proses pengenalan suara dilakukan dengan metode DTW (*Dynamic Time Warping*). Pengolahan sinyal menggunakan *software* Matlab. Dinamo *starter* bergerak selama 3 detik setelah sistem mengenali suara.

MFCC (*Mel Frequency Cepstrum Coefficients*) adalah metode yang digunakan dalam bidang *speech technology* baik *speaker recognition* maupun *speech recognition.*

DTW (*Dynamic Time Warping*) adalah metode untuk menghitung jarak antara dua data *time series*. Keunggulan DTW dari metode jarak yang lainnya adalah mampu menghitung jarak dari dua vektor data dengan panjang berbeda. Perancangan sistem pengenalan suara dilihat pada gambar 1.



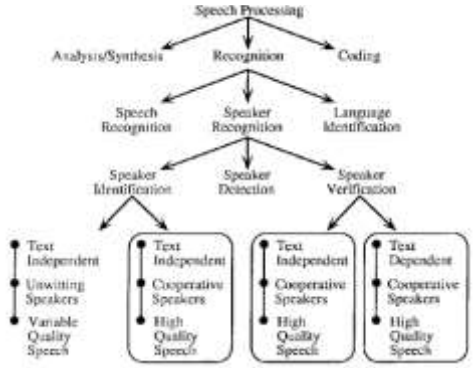
Gambar 1. Perancangan sistem pengenalan suara

Teori

A. Pengenalan Suara

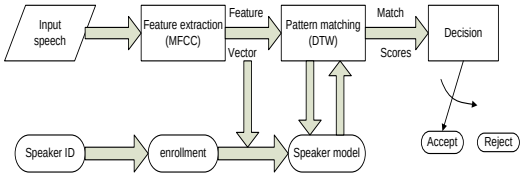
Pengenalan suara dikategorikan menjadi 3 bagian yaitu *speech recognition*, *speaker recognition* dan *language recognition* [1]. Penelitian ini membahas mengenai *speaker recognition* tentang *speaker verification* (*dependent*).

Speaker recognition adalah suatu proses bertujuan mengenali siapa yang sedang berbicara berdasarkan informasi dalam *input* gelombang suara. *Speaker recognition* dibagi menjadi 2 bagian yaitu *speaker verification* dan *speaker identification. Taxonomy* pemrosesan suara terdapat pada gambar 2.



Gambar 2. *Taxonomi* pemrosesan suara

Speaker verification adalah proses verifikasi seorang pembicara dimana identitas pembicara tersebut telah diketahui sebelumnya berdasarkan data yang telah diinputkan. *Speaker verification* melakukan perbandingan *one to one* (1:1) artinya fitur-fitur suara dari seorang pembicara dibandingkan secara langsung dengan fitur-fitur seorang pembicara tertentu yang ada dalam sistem. Bila hasil perbandingan (skor) tersebut lebih kecil atau sama dengan nilai ambang (*threshold value*) maka pembicara tersebut diterima, bila tidak maka akan ditolak dengan asumsi semakin kecil skor berarti kedua sampel semakin mirip. Gambar 3 adalah blok diagram dari *speaker verification*.

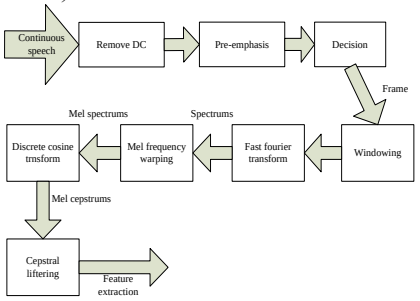


Gambar 3. Blok diagram s*peaker verification*

B. MFCC (Mel Frequency Cepstrum Coefficients)

Metode MFCC digunakan untuk *feature extraction* yaitu suatu proses yang mengkonversikan sinyal suara menjadi beberapa parameter [9],[10], [17], [18]. Gambar 4 merupakan blok diagram untuk MFCC.

MFCC *feature extraction* merupakan adaptasi dari sistem pendengaran manusia, dimana *signal* suara akan di-*filter* secara linear untuk frekuensi rendah (dibawah 1000 Hz) dan secara logaritmik untuk frekuensi tinggi (diatas 1000 Hz).



Gambar 4. Blok Diagram untuk MFCC

Remove DC

Bertujuan menghitung rata-rata data sampel suara dan mengurangkan nilai setiap sampel suara dengan nilai rata-rata tersebut untuk mendapatkan normalisasi dari *input* data suara.

$$yr[n] = xr[n] - \bar{x}, 0 \leq n \leq NL - 1 \quad (1)$$

Pre-Emphasis Filter

noise ratio pada sinyal dapat dikurangi sehingga meningkatkan kualitas sinyal dan menyeimbangkan spektrum dari *voice sound.* Persamaan (2) digunakan dalam *pre- emphasis filter*.

$$H(z) = 1 - \alpha z^{-1} \quad (2)$$

Dengan $0.9 \leq \alpha \leq 1.0$, dan $\alpha \in R$. Persamaan (2) dapat dijadikan sebagian *first order differentiator* pada persamaan 3.

$$yp[n] = sp[n] - \alpha\, s[n-1] \quad (3)$$

Frame Blocking

Sinyal suara terus mengalami perubahan akibat adanya pergeseran artikulasi dari organ produksi vokal, sinyal harus diproses secara *short segments (short frame).* Proses *frame* dilakukan terus sampai seluruh sinyal dapat diproses dan secara *overlapping* untuk setiap

frame. Panjang daerah *overlap* kurang lebih 30% sampai 50% dari panjang *frame*. *Overlapping* dilakukan untuk menghindari hilangnya ciri atau karakteristik suara pada perbatasan perpotongan setiap *frame*.

Windowing

Proses *framing* dapat menyebabkan terjadinya kebocoran spektral (*spectral leakage*) atau *aliasing* yaitu sinyal baru yang memiliki frekuensi berbeda dengan sinyal aslinya. Hasil dari proses *framing* harus melewati proses *window* untuk mengurangi kemungkinan terjadinya kebocoran spektral. Persamaan (4) adalah representasi dari fungsi *window* terhadap sinyal suara yang diinputkan.

$$x(n) \quad = x_t(n)w(n) \quad n = 0,1,\dots,N_f-1 \tag{4}$$

Fungsi *hamming window* sering digunakan dalam aplikasi *speaker recognition* dan terdapat pada persamaan (5).

$$w(n) = 0.54 - 0.4 \in cos\frac{2\pi n}{ML-1} \tag{5}$$

Analisis Fourier

Analisis *fourier* adalah metode untuk melakukan analisa terhadap *spectral properties* dari sinyal yang diinputkan [10].

- *Discrete Fourier Transform(DFT)*

DFT perluasan dari transformasi *fourier* untuk sinyal-sinyal diskrit dengan panjang yang terhingga. Semua sinyal periodik terbentuk dari gabungan sinyal-sinyal sinusoidal yang menjadi satu seperti pada persamaan (6).

$$S[kf] = \sum_{n=0}^{N-1} s[n]e^{-\int 2\pi n\frac{kf}{N}}, 0 \le kf \le N-1 \tag{6}$$

$$kf = {}^{N}/_{2}\ , kf \in N$$

- *Fast Fourier Transform (FFT)*

Perhitungan DFT secara langsung dalam komputerisasi menyebabkan proses perhitungan yang sangat lama. karena dibutuhkan N^2 perkalian bilangan kompleks. Hal itu dapat dilakukan dengan algoritma *fast fourier transform* (FFT) dimana FFT menghilangkan proses perhitungan yang kembar dalam DFT.

Mel Frequency Wrapping

Mel Freqency Wrapping dilakukan dengan menggunakan *filterbank* untuk mengetahui ukuran energi dari *frequency band* tertentu dalam sinyal suara yang dapat diterapkan dalam domain waktu maupun frekuensi. Persamaan (7) digunakan dalam *filterbanks.*

$$Y[t] = \sum_{j=1}^{Ns} S[j]H_i[j] \tag{7}$$

Discrete Cosine Transform(DCT)

DCT merupakan langkah terakhir dari proses utama MFCC *feature extraction*. DCT mendekorelasikan *mel spectrum* sehingga menghasilkan representasi yang baik dari properti spektral lokal. Persamaan 8 digunakan untuk menghitung DCT.

$$C_n = \sum_{k=1}^{K}(logS_k)\cos\left[n\left(k-\frac{1}{2}\right)\frac{\pi}{K}\right];\quad n = 1,2,\dots,K \tag{8}$$

Koefisien ke nol dari DCT pada umumnya akan dihilangkan, walaupun sebenarnya mengindikasikan energi dari *frame* sinyal tersebut.

Cepstral Liftering

Hasil proses MFCC *feature extraction* memiliki beberapa kelemahan. *Low order* dari *cepstral coefficients* sangat sensitif terhadap *spectral slope*, sedangkan bagian *high order*-nya sangat sensitif terhadap *noise*. *Cepstral liftering* menjadi standar teknik yang diterapkan untuk meminimalisasi sensitifitas tersebut [14]. [15], [16], [19]. *Cepstral liftering* dapat dilakukan dengan mengimplementasikan fungsi *window* terhadap *cepstral features.*

$$W[n] - \left\{1 + \frac{L}{2}\sin\left(\frac{n\pi}{L}\right)\right. \quad n - 1,2,\dots,L \tag{9}$$

D. Pencocokan dengan Metode DTW (Dynamic Time Warping)

Perkembangan teknologi pengolahan sinyal wicara adalah memanfaatkan *dynamic-programming* yaitu DTW untuk mengakomodasi perbedaan waktu antara proses perekaman saat pengujian dengan yang tersedia pada *template* sinyal referensi. Metode pemrograman dinamis digunakan untuk menghitung DTW [9],[21],[22]. Jarak DTW dapat dihitung dengan persamaan (10):

$$D(U,V) = \gamma(m,n)$$

$$\gamma(m,n) = d_{base}(u_i,v_j) + min\begin{cases}\gamma(i-1,j)\\ \gamma(i-1)\\ \gamma(i,j-1)\end{cases} \tag{10}$$

E. Sistem Biometrika

Sistem biometrika adalah sistem pengenalan yang bekerja dengan mengambil data biometrika dari individu tertentu, mencari fitur dari data yang diperoleh dan membandingkan fitur ini dengan fitur yang ada dalam basis data [6].

Unjuk kerja sistem biometrika dinyatakan dengan rasio kesalahan keputusan (*decision error rate*), yaitu rasio

kesalahan penerimaan (*false acceptance rate*) dan rasio kesalahan penolakan (*false rejection rate*).

- Rasio Kesalahan Penerimaan

False Acceptance Rate (FAR) menunjukan kesalahan sistem dalam menerima input yang seharusnya ditolak. Data sampel uji dicocokkan dengan data sampel lain yang sebelumnya telah tersimpan dalam *template*. Apabila sistem menerima data sampel uji tersebut padahal kenyataannya data sampel tersebut tidak ada atau tidak sesuai dengan data sampel yang tersimpan dalam *template*. *False acceptance rate* disebut juga *false positive* [2],[3]. Rasio kesalahan penerimaan dihitung dengan persamaan (11):

$$FAR = \frac{Jumlah\ kejadian\ yang\ salah\ terima}{Jumlah\ keseluruhan\ kejadian} \times 100\% \quad (11)$$

- Rasio Kesalahan Penolakan

False Rejected Rate (FRR) menunjukan kejadian dimana sistem melakukan kesalahan dalam menolak masukan. Pengguna yang seharusnya diterima oleh sistem karena data sampel tersebut telah diregistrasi dan ada di dalam *template*, ternyata ditolak oleh sistem. *False rejected rate* disebut juga *False negative*. Rasio kesalahan penolakan dihitung dengan persamaan (12):

$$FRR = \frac{Jumlah\ kejadian\ yang\ salah\ tolak}{Jumlah\ seluruh\ kejadian} \times 100\% \quad (12)$$

- Nilai Ambang (*Threshold Value*)

Nilai ambang, dilambangkan dengan T. Nilai FAR/FRR tergantung pada besarnya nilai ambang yang digunakan. Nilai T akan dibandingkan dengan skor hasil dan bila memenuhi kondisi Skor ≤ T, maka pengguna dinyatakan sah, bila tidak, maka pengguna dinyatakan tidak sah. Penentuan nilai *threshold* menggunakan persamaan (13).

$$T = \text{skor } template \text{ referensi } \times 1.5 \quad (13)$$

Proses pengenalan dilakukan dengan satu nilai koefisien *cepstral* MFCC uji masuk dihitung jarak dengan dua nilai koefisien *cepstral* MFCC dari *template* referensi. Dari hasil proses perhitungan jarak diperoleh dua skor. Dua skor tersebut dijumlahkan lalu dibagi dua. Persamaan (14) menentukan skor untuk diuji.

$$\text{Skor} = \frac{\text{skor 1} + \text{skor2}}{2} \quad (14)$$

- *Grafik Receiver Operation Characteristics (ROC)*

Grafik Karakteristik Operasi Penerima (ROC) adalah grafik yang digunakan untuk menunjukkan unjuk kerja suatu sistem biometrika.

GAR (*Genuine Acceptance Rate*) menyatakan tingkat kesuksesan pengenalan suatu sistem biometrika (bukan tingkat kesalahan), dan dapat dinyatakan sebagai:

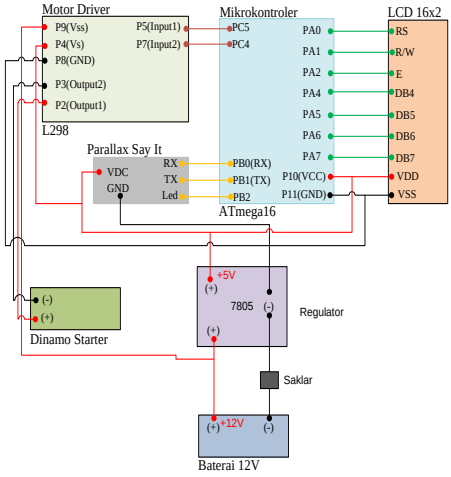
$$GAR = 1 - FRR \quad (15)$$

Hasil dan Pembahasan

Pembuatan prototipe membutuhkan modul pada tabel 1. Rangkaian sistem dari prototipe alat pengenalan suara ini terdapat pada gambar 5.

Tabel 1. Daftar Alat dan Modul

No	*Alat dan Modul*	*Jumlah*
1.	Parallax Say It	1
2.	Mikrokontroller AVR ATmega16	1
3.	LCD 16x2	1
4.	Motor driver	1
5.	Motor DC	1
6.	Baterai 12 V	1



Gambar 5. Rangkaian sistem pengenalan suara

Parallax Say It merupakan modul *voice recognition* multi-fungsi. Modul ini mendukung hingga 32 custom Speaker Dependet (SD) trigger atau perintah, yang dapat digunakan pada bahasa apapun. Komunikasi dengan perangkat lain menggunakan komunikasi serial antar muka UART (*Universal Asynchronous Receiver-Transmitter*). Protokol komunikasi menggunakan karakter ASCII.

Ketika saklar on ditekan maka mikrokontroler mengirimkan data karakter "b" untuk memanggil modul Parallax Say It. Jika koneksi berhasil, mikrokontroler membaca data karakter "o" dan LCD menampilkan pada baris atas "WELCOME" dan pada baris bawah "MR. REZZA". Selanjutnya LCD menampilkan "Say

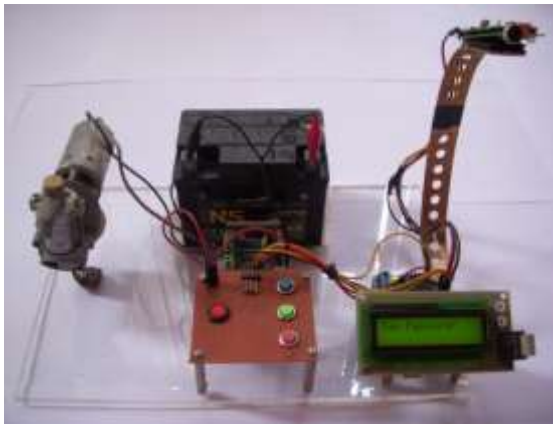
Password" dan mikrokontroler mengirimkan data "d" dan "A+0" ke Parallax Say It untuk proses pengenalan suara. Jika suara dikenali Parallax Say It mengirimkan data karakter "r" dan LCD menampilkan "Success". Selanjutnya LCD menampilkan "Say Command" dan mikrokontroler mengirimkan data "d" dan "A+1" ke parallax say it untuk proses pengenalan suara. Jika suara dikenali Parallax Say It mengirimkan data karakter "r", LCD menampilkan "Machine On" dan motor DC bergerak selama 3 detik.

Pengujian dilakukan menggunakan perangkat lunak Say It GUI 1.1.5, dengan menghubungkan modul Parallax Say It ke komputer melalui USB. Setelah terhubung melakukan *training* suara dan *testing* pengenalan suara.

Pengujian program menggunakan *virtual terminal* yang ada di perangkat lunak CodeVisionAVR, dengan cara menghubungkan modul mikrokontroler ke komputer melalui USB. Setelah terhubung melakukan pengiriman data karakter [7], [8].

Mekanisme pengiriman data akan disinkronisasi UART pada mikrokontroler dengan memanfaatkan bit "START" dan bit "STOP". Ketika jalur *transmitter* (Tx) dalam keadaan *idle* (tidak ada data yang ditransfer), maka output Tx dalam keadaan logika "1". Ketika data akan dikirim melalui Tx, maka *output* Tx akan diset lebih dulu ke logika "0" untuk selang waktu satu bit. Sinyal ini pada Rx akan dikenali sebagai sinyal "START" yang digunakan untuk mensinkronkan *fase clock*-nya sehingga sinkron dengan *fase clock* Tx. Selanjutnya, data akan dikirimkan secara serial dari bit paling rendah (bit 0 / LSB) sampai bit tertinggi (MSB). Bit data akan diikuti oleh sinyal "STOP" sebagai tanda akhir dari pengiriman data serial.

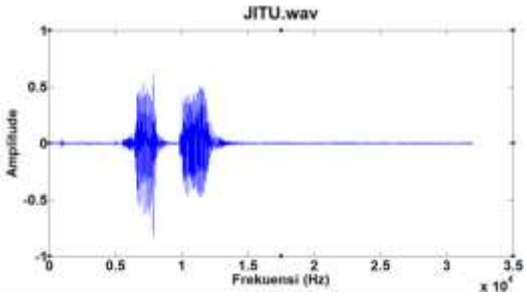
Prototipe sistem pengenalan suara terdapat pada gambar 6.



Gambar 6. Prototipe sistem pengenalan suara sebagai pengendali dinamo starter pada otomobil.

Tujuan proses pengolahan suara adalah untuk mendapatkan ciri atau parameter dari sinyal suara. Proses MFCC diimplementasikan dengan menggunakan *toolbox* yang telah tersedia, yaitu *speech and audio processing toolbox* yang dikembangkan . Adapun tahapan-tahapan proses MFCC yang dilakukan adalah: *voice recording*, *remove silent, remove dc*, *pre-emphasis*, *frame blocking*, *windowing*, *fast fourier transform*, *filterbank*, *discrete consine transform* dan *cepstral liftering* [6], [11], [12],[13].

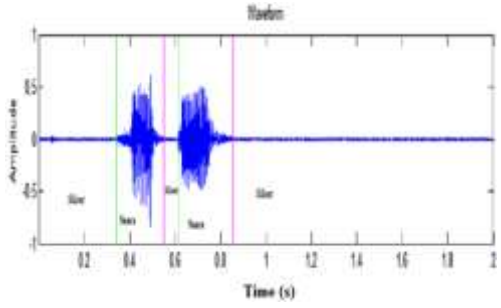
Pengambilan data suara dilakukan dengan perekaman suara pada frekuensi sampel (Fs) 16 KHz selama dua detik. Gambar 7 merupakan sinyal suara kata "jitu".



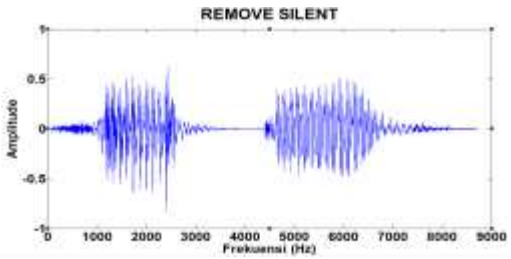
Gambar 7. Sinyal suara asli kata "jitu"

Tempat yang digunakan untuk proses pengambilan suara dilakukan pada kondisi ruangan dengan tingkat kebisingan yang rendah, karena bila *noise* yang terdapat pada ruangan terlalu besar dapat menyulitkan saat proses pembersihan data suara.

Proses *remove silent* untuk menghilangkan *frame-frame* yang mengandung *silent* seperti pada gambar 8. Proses yang dilakukan adalah mendeteksi mulai sinyal suara awal dan berakhir ketika sudah tidak diucapkan. Hasil data suara dari proses *remove silent* dapat dlihat pada gambar 9.

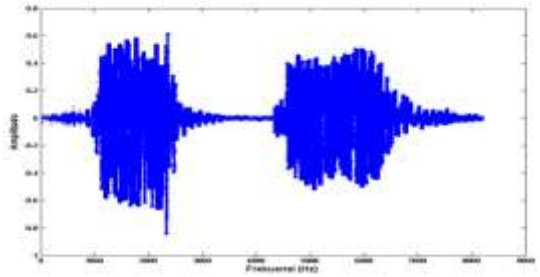


Gambar 8. Proses *remove silent*



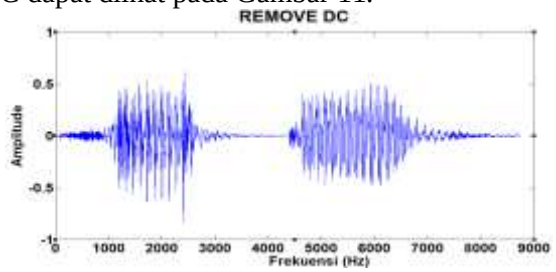
Gambar 9. Data suara setelah proses *remove silent*

Apabila *noise* yang terdapat pada suara terlalu besar, maka proses pembersihan data ini tidak dapat berjalan dengan optimal seperti yang terlihat pada gambar 10. Hal ini disebabkan sistem tidak mampu membedakan lagi antara gelombang suara dengan *noise* dari lingkungan. *Noise* juga dapat disebabkan dari gangguan distorsi pada gelombang sinyal listrik AC (*Alternate Current*) yang masuk melalui *power battery* atau *device* lain.



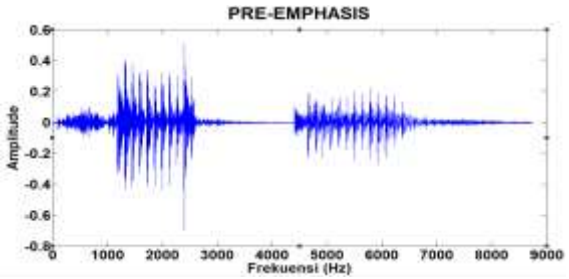
Gambar 10. Data suara dengan *remove silent* yang tidak optimal

Proses *remove DC* bertujuan melakukan normalisasi terhadap data sampel suara yang dimasukkan. Hasil data suara dari proses *remove DC* dapat dlihat pada Gambar 11.



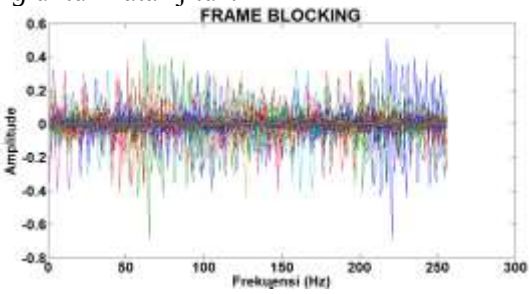
Gambar 11. Data suara setelah proses *remove DC*

Setelah melewati proses *remove DC,* selanjutnya data sampel suara memasuki proses *pre-emphasis filtering.* Gambar 12 merupakan hasil dari proses *pre-emphasis filtering* untuk kata "jitu".

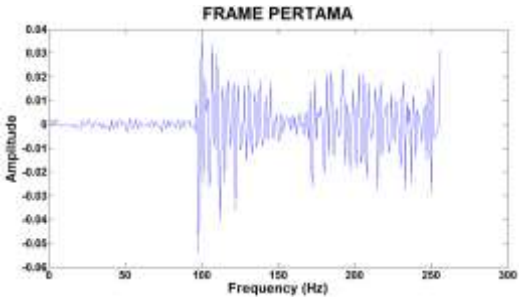


Gambar 12. Data suara setelah proses *Pre-Emphasis Filtering*

Pada penelitian ini sinyal suara dipotong sepanjang 256 Hz pada setiap pergeseran 128 Hz dengan frekuensi *sampling* sebesar 8737 Hz. Setiap potongan tersebut dinamakan *frame*. Jadi setiap satu *frame* terdapat 256 sampel dari 8737 sampel yang ada. Gambar 13 adalah hasil dari proses *frame blocking* untuk kata "jitu".

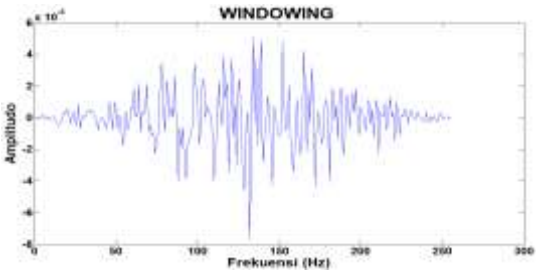


Gambar 13. Data suara setelah proses *frame blocking*



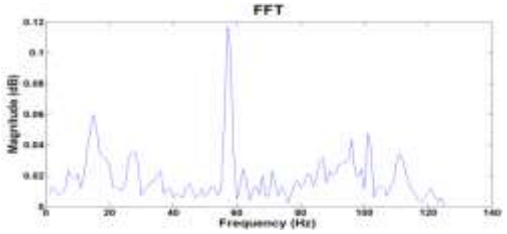
Gambar 14. Data Suara setelah proses *frame blocking* untuk *frame* pertama

Proses *windowing* dilakukan untuk mengurangi efek diskontinuitas dari proses *frame blocking* terutama pada ujung-ujung *frame.* Gambar 15 adalah hasil dari proses *windowing* untuk kata "jitu".



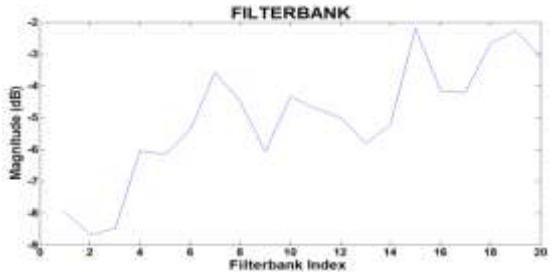
Gambar 15. Data suara setelah proses *Windowing* untuk *frame* pertama

Proses FFT *(Fast Fourier Transform)* akan mengubah sinyal suara ke dalam domain frekuensi dengan 256 titik. Gambar 16 merupakan hasil dari proses FFT untuk kata "jitu".



Gambar 16. Data suara setelah proses FFT untuk *frame* pertama

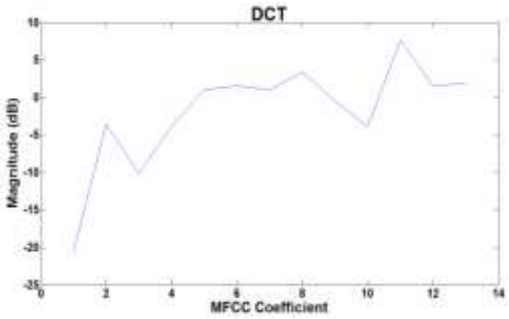
Konsep pendengaran telinga manusia terhadap suara atau bunyi adalah dalam skala linear pada frekuensi kurang dari 1 KHz dan logaritmik diatas Skala frekuensi *filterbank* adalah sama dengan konsep pendengaran manusia sehingga sering dijadikan parameter ekstraksi dalam pengolahan sinyal suara. Panjang *filterbank* adalah 20 setiap *frame*. Gambar 17 adalah hasil dari proses *filterbank* untuk kata "jitu".



Gambar 17. Data suara setelah proses *filterbank* untuk *frame p*ertama

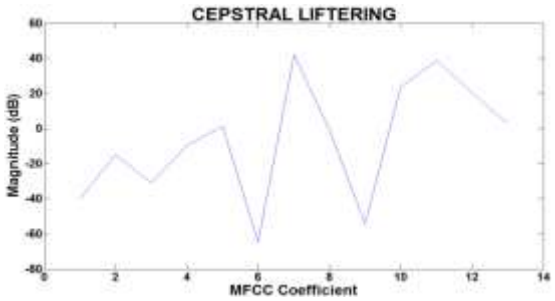
Hasil proses DCT adalah *mel frequency cepstrum coefficients* yang merupakan hasil proses MFCC. Gambar 18 adalah hasil proses DCT untuk kata "jitu".

Panjang berikut adalah data koefisien MFCC untuk kata "jitu" dengan jumlah koefisien MFCC sebanyak 13 koefisien untuk masing-masing *frame*.



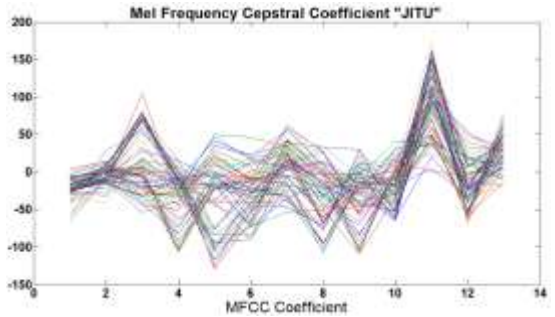
Gambar 18. Data suara setelah proses DCT untuk *frame* pertama

Cepstral liftering berfungsi untuk menghaluskan spektrum hasil dari proses MFCC sehingga diharapkan dapat meningkatkan akurasi program dalam melakukan pengenalan. Gambar 19 adalah hasil dari proses *cepstral liftering* untuk kata "jitu".



Gambar 19. Data suara setelah proses *cepstral liftering* (*frame* pertama)

Gambar 20 adalah hasil keseluruhan ekstraksi ciri pada 67 *frame* ucapan kata "jitu".



Gambar 20. Hasil ekstrasi ciri ucapan "jitu" metode MFCC

Pengujian terhadap sistem verifikasi suara yang dibuat dalam penelitian ini dilakukan menggunakan metode DTW dengan melakukan proses perhitungan jarak yaitu membandingkan dua buah sampel yang diperoleh dari proses ekstraksi ciri. Jarak yang dihitung adalah jarak antara nilai koefisien *cepstral* MFCC yang ada di *template* referensi dan menghitung jarak *template* referensi dengan nilai koefisien *cepstral* MFCC dari suara uji yang masuk. Dari proses DTW ini akan diperoleh suatu nilai atau skor hasil perbandingan antara dua buah sampel.

Pengujian dilakukan menggunakan data sampel dari 5 orang, dengan komposisi 1 orang laki-laki sebagai pengguna, 3 orang laki-laki dan 1 orang perempuan. Masing-masing orang mengucapkan satu buah kata "jitu". Pengguna diambil data sebanyak 12 data sampel dengan 2 data sampel sebagai *template* referensi dan 10 data sampel sebagai data uji. Untuk bukan pengguna setiap orang diambil data sebanyak 10 data sampel sebagai data uji, sehingga jumlah sampel yang ada 2+10+(4x10) = 52 data sampel. Tabel II adalah skor hasil pencocokan data uji dengan *template* referensi menggunakan metode DTW dan grafik distribusi probabilitas skor pengguna (pengguna sah dan pengguna tidak sah.

Setiap hasil pengujian akan ditampilkan grafik unjuk kerja sistem (FRR dan FAR) atau disebut juga grafik ROC. Hasil pengujian akan disajikan dalam bentuk tabel ataupun grafik untuk mempermudah analisa.

Tabel 2. Skor Pencocokan Data Uji

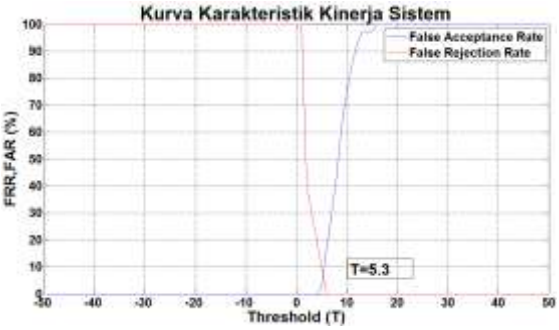
HASIL	DATA				
	Genuine	Impostor 1	Impostor 2	Impostor 3	Impostor 4
Skor 1	1,6913	5,3943	8,1767	5,5926	10,6204
Skor 2	1,6146	8,4347	7,2369	6,9271	12,2062
Skor 3	2,6549	8,4952	5,8032	7,9476	10,6081
Skor 4	5,7401	5,1787	7,2224	4,2621	11,9408
Skor 5	1,8683	9,3358	7,6255	7,411	10,0287
Skor 6	1,8552	8,9372	8,1221	6,9705	10,3974
Skor 7	3,5367	6,9181	8,4562	5,4202	9,9539
Skor 8	1,9279	5,0273	6,8766	9,0546	11,8979
Skor 9	1,9785	11,7513	12,9389	7,385	9,4879
Skor 10	4,1628	6,7385	9,9548	8,3607	15,5555

Gambar 21. Grafik hasil skor pengguna asli dan palsu

Ada dua pengujian ucapan kata "jitu" yang dilakukan dalam penelitian ini, diantaranya adalah:

a. Menentukan nilai *threshold* menggunakan grafik unjuk kerja sistem (FRR dan FAR).
b. Menentukan nilai *threshold* menggunakan persamaan 13.

Pengujian ini bertujuan untuk mengetahui nilai *threshold* yang akan digunakan untuk sistem dalam melakukan verifikasi. Hasil pengujian dengan melihat titik pertemuan antara FRR dan FAR. Gambar 22 adalah hasil grafik unjuk kerja sistem (FRR dan FAR).



Gambar 22. Kurva karakterisitik kinerja sistem (FRR dan FAR)

Dalam pengujian ini nilai threshold adalah 5.3 dengan tingkat kesalahan 7 %.

Penelitian ini menemukan bahwa hasil terbaik untuk menentukan nilai *threshold* menggunakan persamaan 13, diperoleh ketika dua ucapan yang sama ("jitu") digunakan untuk *template* referensi sistem untuk setiap pengguna. Dari dua ucapan yang sama tersebut dilakukan proses pencocokan menggunakan metode DTW. Skor hasil pencocokan adalah 2.3412.

Nilai *threshold* ditentukan dari skor hasil pencocokan *template* referensi dikali 1.5. Jadi nilai *threshold*-nya adalah 3.5118 yang berarti bahwa jika skor data uji $\leq$ 3.5118 maka pengguna dinyatakan sah, bila tidak, maka pengguna dinyatakan tidak sah.

Dari hasil penelitian dengan menentukan nilai *threshold* pencocokan, maka akan didapatkan *False Aceptance Rate* (FAR) dan *False Reject Rate* (FRR). Nilai FAR akan naik apabila *threshold* dinaikkan, sedangkan nilai FRR akan turun. Tabel III adalah nilai FRR, FAR, GAR dan akurasi sistem dari hasil *threshold* yang digunakan pada penelitian ini.

Tabel 3. Nilai FRR, FAR, GAR dan Akurasi Sistem

Threshold	FRR (%)	FAR (%)	GAR (%)	Akurasi Sistem (%)
3,5118	30	0	80	70
5,3	10	7.5	90	82.5

Kesimpulan

1. Metode *Mel Frequency Cepstrums Coefficients* (MFCC) adalah metode yang baik untuk ekstraksi fitur dalam pengenalan suara karena mampu untuk menangkap karakteristik suara yang sangat penting bagi pengenalan suara, menghasilkan data seminimal mungkin dan mereplikasi organ pendengaran manusia dalam melakukan persepsi terhadap sinyal suara.
2. Proses pengenalan suara sensitif terhadap kebisingan karena dapat mempengaruhi proses ekstraksi fitur sinyal suara.
3. Metode *Dynamic Time Warping* (DTW) dapat digunakan untuk membandingkan dua buah fitur suara hasil dari proses MFCC.
4. Tingkat keberhasilan sistem verifikasi tergantung nilai *threshold* yang digunakan.
5. Hasil pengujian adalah tingkat akurasi paling rendah 70% nilai *threshold* 3,5 sedangkan tingkat akurasi tertinggi 82.5 % dengan nilai *threshold* 5,3.

Ucapan Terima kasih

Terimakasih kepada saudara Rezza Aditya yang telah banyak membantu dalam pengujian dan pengambilan data pada penelitian ini.

Nomenklatur

$yr[n]$ sampel sinyal hasil proses *remove DC*
$xr[n]$ sampel sinyal asli
$\bar{x}$ nilai rata-rata sampel sinyal asli
NL panjang sinyal
$yp[n]$ sinyal hasil *pre-emphasis filter*
$sp[n]$ sinyal sebelum *pre-emphasis filter*

$x(n)$ nilai sampel sinyal hasil *windowing*
$x_t(n)$ nilai sampel dari *frame* sinyal ke i
$w(n)$ fungsi *window*
N_{f} *frame* size, merupakan kelipatan 2
M_L panjang *frame*
N jumlah sampel yang akan diproses
$s[n]$ nilai sampel sinyal
kf variabel frekuensi diskrit
N_{s} jumlah *magnitude spectrum*
$S[j]$ *magnitude spectrum* pada frekuensi j
$H_i[j]$ koefisien *filterbank* pada frekuensi j
S_k keluar dari proses *filterbank* pada *index* k.
k jumlah koefisien yang diharapkan.
L jumlah *cepstral coefficients.*

Referensi

[1.] Joseph P. Campbell, JR., Speaker Recognition: A Tutorial. Proceedings of The IEEE, Vol. 85, No. 9, September 1997.
[2.] S. K. Singh, Features and Techniques for Speaker Recognition, M. Tech. Credit Seminar Report, Electronic Systems Group, EE Dept, IIT, 2003
[3.] Bojan Imperl, .Speaker recognition techniques., Laboratory for Digital Signal, Processing, Faculty of Electrical Engineering and Comp. Sci., Smetanova 17, 2000 Maribor,Slovenia.
[4.] D. A. Reynolds, .An Overview of Automatic Speaker Recognition Technology., Proc. IEEE, pp. 4072-4075, 2002.
[5.] Gunawan, Juwono, F. Hilman, *Pengolahan Sinyal Digital Dengan Pemrograman Matlab,* Graha Ilmu, Yogyakarta, 2012.
[6.] D. Putra, *Sistem Biometrika*, Andi, Yogyakarta, 2009.
[7.] S. Rangkuti, *Mikrokontroller Atmel AVR: Simulasi dan Praktek Menggunakan ISIS Proteus dan CodeVisionAVR*, Informatika, Bandung, 2011.
[8.] L. Willa, *Teknnik Digital Mikroprosesor dan Mikrokomputer,* Informatika, Bandung, 2010.
[9.] B.R.Wildermorth, Text-Independent Speaker Recognition Using Source Based Features, Griffith University Australia, 2001
[10.] T. Kinnunen, Spectral Features for Automatic, Text-Independent Speaker Recognition, Thesis, University of Joensuu, Finlandia, 2003
[11.] C. Cornaz, U. Hunskeler, An Automatic Speaker Recognition System Digital Signal Processing, Ecole Polytechnique Federal de Laussane, Switzerland, February, 2003
[12.] J.-S Roger Jang, Audio Signal Processing and Recognition, 1996
[13.] L. Feng, Speaker Recognition, Thesis, Technical University of Denmark, 2004
[14.] Sirko Molau, Michael Pitz, Ralf Schluter, and Hermann Ney, Computing Mel Frequency Cepstral Coefficients on The Power Spectrum, Proceedings ICASSP, 2001
[15.] Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi, Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques, Journal of Computing, Volume 2, Issue 3, March 2010, ISSN-2151-9617.
[16.] Xinhui Zhou, Daniel Garcia-Romero, Ramani Duraiswami, Carol Espy-Wilson, Shihab Shamma, Linear versus Mel Frequency Cepstral Coefficients for Speaker Recognition, ASRU 2011
[17.] Zbynìk Tychtl and Josef Psutka, Speech Production Based on the Mel-Frequency Cepstral Coefficients, University of West Bohemia, Department of Cybernetics,
[18.] Amelia C.Kelly and Christer Gobl, A comparison of mel frequency cepstral coefficient (MFCC) calculation techniques, Journal of Computing, Volume 3, Issue 10, October 2011, ISSN 2151-961.
[19.] Priyanka Mishra, Suyash Agrawal, Recognition Of Voice Using Mel Cepstral Coefficient & Vector International Journal of Engineering Research and Applications (IJERA) Vol. 2, Issue 2,Mar-Apr 2012, pp.933-938, ISSN: 2248-9622.
[20.] Ben J. Shannon, Kuldip K. Paliwal, A Comparative Study of Filter Bank Spacing for Speech Recognition, Microelectronic Enginnering Research Coference 2003.
[21.] D.E.Riedel, S.Venkatesh, W.Liu,Threshold Dynamic Time Warping for Spatial Activity Reecognitopn, International Journal of Information and System Sciences, Volume 1, Number 1, Pages 1–14, 2004.
[22.] Selina Chu, Shrikanth Narayanan, C.-C. Jay Kuo, Efficient Rotation Invariant Retrieval of Shapes using
Dynamic Time Warping with Applications in Medical Databases, 19th IEEE International Symposium on Computer-Based Medical Systems (CBMS 2006), pg 673-678.