

Sistem Pengenalan Kata Bahasa Indonesia Berbasis LabVIEW untuk Pengendalian Peralatan Ruang Perkuliahan

Indonesian Word Recognition System using LabVIEW for Controlling Equipment in the Lecture Hall

Muhammad Aria

Universitas Komputer Indonesia
Jl. Dipati ukur No 112, Bandung
Email : ariar554@gmail.com

Abstrak

Paper ini memaparkan tentang implementasi sistem pengenalan kata bahasa Indonesia menggunakan bahasa pemrograman LabVIEW 7.1 untuk mengendalikan beberapa peralatan umum pada ruang perkuliahan, yaitu lampu, proyektor, AC dan pintu. Sistem pengenalan kata terdiri dari dua subsistem, yaitu subsistem ekstraksi ciri dan subsistem pencocokan pola. Subsistem ekstraksi ciri akan mengkonversi sinyal suara masukan menjadi beberapa vektor ciri untuk proses pengenalan. Vektor ciri tersebut selanjutnya akan dibandingkan dengan vektor ciri yang tersimpan dalam basis data melalui subsistem pencocokan pola. Untuk memperoleh performansi yang baik, maka pada penelitian ini dibandingkan performansi beberapa metode ekstraksi ciri yaitu Mel Frequency Cepstrum Coefficients (MFCC), MFCC+Delta dan MFCC+Delta+Double Delta. Untuk metode pencocokan pola juga dilakukan perbandingan performansi metode Dynamic Time Warping (DTW) dan Vector Quantization (VQ). Kata yang digunakan terbatas pada 11 buah kata berbahasa Indonesia. Pembicara dalam penelitian ini dibatasi pada dua orang pembicara laki-laki dan dua orang pembicara perempuan. Penelitian berfokus pada akurasi pengenalan pada tingkat kata. Hasil penelitian menunjukkan bahwa pengenalan kata terbaik yaitu sebesar 98.2% diperoleh menggunakan metode MFCC+Delta+Double Delta sebagai ekstraksi ciri dan metode DTW sebagai pencocokan pola.

Kata Kunci : Dynamic Time Warping, LabVIEW 7.1, Mel Frequency Cepstrum Coefficients, Pengenalan Kata Berbahasa Indonesia, Vector Quantization

Abstract

This paper describes the implementation of the Indonesian word recognition system using LabVIEW 7.1 programming language for controlling some common equipment in the lecture hall, the lights, projectors, air conditioning and doors. Word recognition system consists of two subsystems, feature extraction subsystem and pattern matching subsystem. Feature extraction subsystem will convert input sound signal into a feature vector for recognition. The feature vector will then be compared with the characteristic vectors stored in a database through pattern matching subsystem. To obtain good performance, so in this study compared the performance of several feature extraction methods namely Mel Frequency Cepstrum Coefficients (MFCC), MFCC+Delta and MFCC+Delta+Double Delta. For pattern matching method also performed comparison of the performance of the Dynamic Time Warping (DTW) method and Vector Quantization (VQ) method. The word used is limited to 11 pieces of the Indonesian language. Speakers in this study is limited to two male speakers and two female speakers. The research focuses on the recognition accuracy at the word level. The results showed that the best word recognition is equal to 98.2% is obtained using the MFCC + Delta + Double Delta as feature extraction and DTW as pattern matching methods.

Keywords : Dynamic Time Warping, LabVIEW 7.1, Mel Frequency Cepstrum Coefficients, Indonesian language word recognition, Vector Quantization

I. PENDAHULUAN

Sistem pengenalan kata adalah salah satu cabang dari kecerdasan buatan yang memungkinkan komputer untuk dapat menerima

masukan berupa kata yang diucapkan. Teknologi ini memungkinkan suatu alat untuk mengenali ucapan dengan cara digitalisasi kata dan mencocokkan sinyal digital tersebut dengan basis data yang tersimpan dalam perangkat [1]. Hingga

saat ini, ilmuwan dari berbagai negara terus mengembangkan sistem pengenalan kata untuk bahasanya masing-masing [2], [3], [4]. Tetapi sistem pengenalan kata untuk bahasa Indonesia masih sangat minim dalam pengembangannya [5].

Penelitian ini bertujuan mengimplementasikan sistem pengenalan kata berbahasa Indonesia untuk mengendalikan beberapa peralatan umum pada ruang perkuliahan, yaitu lampu, proyektor, AC dan pintu. Adapun sistem pengenalan kata terdiri dari dua subsistem, yaitu subsistem ekstraksi ciri dan subsistem pencocokan pola. Maka untuk memperoleh performansi pengenalan kata yang baik telah dilakukan perbandingan beberapa metode ekstraksi ciri dan pencocokan pola.

Terdapat beberapa metode ekstraksi ciri, antara lain *Linear Predictive Coding* (LPC), *Perceptual Linear Prediction* (PLP), *Zero Crossing Rate* (ZCR), *Short Time Energy* (STE) dan *Mel Frequency Cepstrum Coefficients* (MFCC). Tujuan dari ekstraksi ciri ini adalah untuk mengkonversi sinyal suara masukan menjadi beberapa vektor ciri untuk proses pengenalan. Vektor ciri tersebut selanjutnya akan dibandingkan dengan vektor ciri yang tersimpan dalam basis data melalui subsistem pencocokan pola. Pada penelitian ini akan digunakan metode ekstraksi ciri berbasis MFCC. MFCC dipilih karena merupakan teknik yang telah luas dipakai pada pemrosesan sinyal suara. Penggunaan metode ini memberikan pengenalan yang lebih baik dibandingkan dengan metode lain yang sudah ada [6]. Untuk memperoleh performansi yang baik, maka akan dibandingkan beberapa variasi MFCC yaitu MFCC, MFCC+Delta dan MFCC+Delta+Double Delta.

Adapun model pengenalan pola yang dapat digunakan untuk pengenalan suara antara lain adalah *Dynamic Time Warping* (DTW), *Vector Quantization* (VQ), *Hidden Markov Models* (HMM) dan *Artificial Neural Network* (ANN). Pada penelitian ini akan dibandingkan metode pengenalan pola berbasis DTW dan VQ. DTW adalah teknik yang cukup populer di awal perkembangan teknologi pengolahan sinyal wicara dimana DTW dapat melakukan pencocokan terhadap dua vektor data yang memiliki panjang yang berbeda. Adapun VQ adalah proses pemetaan vektor-vektor dari ruang vektor berukuran besar menjadi sebuah wilayah yang terbatas. Tujuan dari proses VQ dalam *voice processing* adalah untuk melakukan kompresi terhadap jumlah data yang ada sehingga dapat mengurangi jumlah dari proses komputasi yang ada.

Kata yang digunakan pada penelitian ini terbatas pada 11 buah kata berbahasa Indonesia. Pembicara dalam penelitian ini dibatasi pada dua orang pembicara laki-laki dan dua orang pembicara perempuan. Penelitian berfokus pada akurasi pengenalan pada tingkat kata.

Sistem pengenalan kata ini dibuat menggunakan bahasa pemrograman LabVIEW 7.1. Walaupun telah banyak peneliti yang membuat sistem pengenalan kata tetapi sepanjang yang penulis ketahui belum ada yang membuat sistem pengenalan kata menggunakan bahasa pemrograman LabVIEW.

Kontribusi dari penelitian ini adalah menyajikan alternatif metode interaksi manusia dengan komputer/alat menggunakan bahasa Indonesia dan juga sebagai salah satu usaha untuk mempermudah pengajar dalam mengendalikan peralatan yang ada di ruangan kelas. Selain itu juga dilakukan perbandingan performansi metode ekstraksi ciri antara MFCC, MFCC+Delta, MFCC+Delta+Double Delta dan WMFCC serta perbandingan performansi metode pencocokan pola antara DTW, DDTW dan IFDTW. Dan dari penelitian ini juga dirancang toolbox sistem pengenalan kata berbasis bahasa pemrograman LabVIEW.

Paper ini diorganisasikan sebagai berikut. Pada bagian 2 dijelaskan mengenai dasar teori Sistem Pengenalan Kata. Pada bagian 3 dijelaskan mengenai metode penelitian yang dilakukan. Implementasi alat pengenalan kata berbasis LabVIEW dijelaskan pada bagian 4. Pada bagian 5 disajikan hasil dan pembahasan. Dan Bagian 6 menyimpulkan makalah ini.

II. SISTEM PENGENALAN KATA

A. Prinsip Sistem Pengenalan Kata

Secara umum, sistem pengenalan kata terdiri dari dua subsistem, yaitu subsistem ekstraksi ciri dan subsistem pencocokan pola. Subsistem ekstraksi ciri melakukan proses transformasi sinyal input ke dalam satu set vektor ciri sebagai representasi dari sinyal suara. Adapun subsistem pencocokan pola melakukan identifikasi suara yang belum diketahui dengan cara membandingkan sinyal suaranya yang telah diekstrak ke vektor ciri dengan set vektor ciri dari "kata" yang telah diketahui dan tersimpan dalam sistem [10].

Dari aspek perancangan sistem, terdapat dua tahapan pada sistem pengenalan kata. Tahap pertama adalah tahap pelatihan. Pada tahap ini, sistem melakukan pengumpulan data untuk membangun kumpulan vektor ciri yang telah diketahui. Pada tahap kedua adalah pengujian, yaitu sinyal masukan yang diberikan pada sistem akan diubah menjadi vektor ciri dan dibandingkan dengan bank vektor ciri yang telah ada pada sistem. Keputusan untuk menentukan model kata yang diterima berdasarkan nilai terdekat vektor ciri yang ada pada sistem dengan vektor ciri masukan.

B. Digitalisasi Gelombang Suara

Gelombang suara yang diterima pada mikrofon masih berupa sinyal analog. Maka sinyal analog tersebut harus terlebih dahulu diubah menjadi sinyal digital menggunakan proses digitalisasi. Proses digitalisasi suara terdiri atas dua tahapan yaitu *sampling* dan kuantisasi [11]. *Sampling* adalah pengambilan/pengukuran amplitudo sinyal pada suatu indeks waktu tertentu. Sampling rate yang biasanya digunakan pada pengenalan suara ialah 8000 Hz sampai dengan 20k Hz. Hubungan panjang vektor yang dihasilkan, sampling rate dan panjang data suara yang digitalisasi dinyatakan dalam persamaan (1).

$$S = F_s \times T \quad (1)$$

dengan S = panjang vektor

F_s = *sampling rate* (Hz)

T = panjang suara (detik)

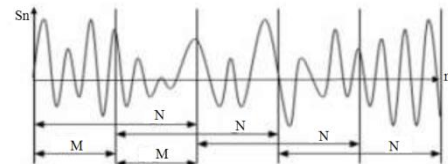
Setelah tahap *sampling* maka proses selanjutnya adalah proses kuantisasi. Proses ini menyimpan nilai amplitudo ke dalam nilai integer, yang dalam hal ini menggunakan representasi nilai 8 bit atau 16 bit. Setelah sinyal didigitalisasi, langkah berikutnya adalah menyekatnya ke dalam *frame* dan menkonversikannya menjadi *feature vector* yang selanjutnya menjadi masukan bagi tahap berikutnya. Fitur yang dipakai dalam penelitian ini adalah MFCC.

C. Ekstraksi Ciri Mel-Frequency Cepstrum Coefficients (MFCC)

Ekstraksi ciri merupakan proses untuk menentukan vektor yang dapat digunakan sebagai penciri objek atau individu. Ciri yang biasa digunakan adalah koefisien cepstral. MFCC

merupakan ekstraksi ciri yang menghitung koefisien cepstral yang didasarkan pada variasi dari frekuensi kritis telinga manusia. Filter diletakkan secara linear pada frekuensi rendah dan logaritmik pada frekuensi tinggi untuk mendapatkan karakteristik suara yang penting. Adapun tahapan-tahapan dalam MFCC adalah sebagai berikut [12].

1. *Frame Blocking*. Tahap ini sinyal suara analog dibagi menjadi beberapa frame yang terdiri dari N sample, masing-masing frame dipisahkan oleh M , dengan M adalah banyaknya pergeseran antar frame ($M < N$). Frame pertama berisi sampel N pertama. Frame kedua dimulai M sampel setelah permulaan frame pertama, sehingga frame kedua ini *overlap* terhadap frame pertama sebanyak $N-M$ sample. Selanjutnya, frame ketiga akan dimulai M sampel setelah frame kedua. Proses ini berlanjut sampai seluruh suara tercakup dalam frame. Hasil dari proses ini adalah matriks dengan N baris dan beberapa kolom sinyal $X[N]$. Proses ini ditunjukkan pada **Gambar 1**, S_n adalah nilai sampel yang dihasilkan dan n adalah urutan sampel yang akan diproses.



Gambar 1. Proses *frame blocking*

2. *Windowing*. Proses *framing* dapat menyebabkan terjadinya kebocoran spektral yaitu sinyal yang baru memiliki frekuensi yang berbeda dengan sinyal aslinya. Efek ini dapat terjadi karena rendahnya jumlah *sampling rate* ataupun karena proses *frame blocking* dimana menyebabkan sinyal menjadi *discontinue*. Untuk mengurangi kemungkinan terjadinya kebocoran spektral ini maka hasil dari proses *framing* harus melewati proses *windowing*. Konsep *windowing* adalah meruncingkan sinyal ke angka nol pada permulaan dan akhir setiap frame. Proses ini dilakukan dengan mengalikan antar frame dengan jenis window yang digunakan. Jika window didefinisikan sebagai $w(n)$, $0 \leq n \leq N - 1$, dengan N adalah jumlah sampel dalam tiap frame, maka proses *windowing* ini dapat dituliskan dalam persamaan (2) berikut.

$$y(n) = x(n)w(n), 0 \leq n \leq N - 1 \quad (2)$$

dengan

- $y(n)$ = sinyal hasil windowing sampel ke- n
- $x(n)$ = nilai sampel ke- n
- $w(n)$ = nilai window ke- n
- N = jumlah sampel dalam frame

Penelitian suara banyak menggunakan window hamming karena kesederhanaan formulanya dan nilai kerja window. Dengan pertimbangan tersebut, maka penggunaan window Hamming cukup beralasan. Persamaan window Hamming adalah :

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \quad (3)$$

dengan

$$n = 0, 1, \dots, N - 1$$

3. *Fast Fourier Transform* (FFT). Tahapan selanjutnya ialah mengubah setiap frame yang terdiri dari N samples dari domain waktu ke dalam domain frekuensi. Output dari proses ini disebut dengan nama spektrum atau periodogram. Sinyal dalam domain frekuensi dapat diproses dengan lebih mudah dibandingkan data pada domain waktu, karena pada domain frekuensi, amplitudo suara tidak terlalu berpengaruh. FFT adalah algoritma yang mengimplementasikan Discrete Fourier Transform (DFT) yang dioperasikan pada sebuah sinyal waktu diskrit yang terdiri dari N sampel menggunakan persamaan (4) dan (5) berikut.

$$Real\ X[k] = \sum_{i=0}^{N-1} x[i] \cdot \cos\left(\frac{2\pi ki}{N}\right) \quad (4)$$

$$Imaginer\ X[k] = - \sum_{i=0}^{N-1} x[i] \cdot \sin\left(\frac{2\pi ki}{N}\right) \quad (5)$$

dimana

N = jumlah data

$k = 0, 1, 2, \dots, \frac{N}{2}$

$x[i]$ = data pada titik ke i

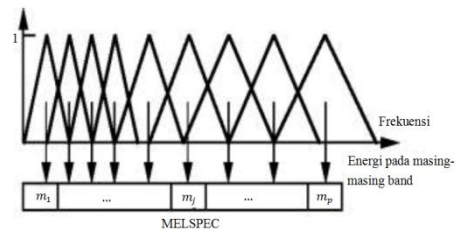
Proses selanjutnya adalah menghitung nilai magnitudo dari FFT. Persamaan yang digunakan adalah (6) berikut.

$$|X[k]| = \sqrt{(Real\ X[k])^2 + (Imaginer\ X[k])^2} \quad (6)$$

4. *Mel-Frequency Wrapping*. Persepsi sistem pendengaran manusia terhadap frekuensi sinyal suara ternyata 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 F_{HZ} , sebuah pola diukur dalam sebuah skala yang disebut ‘mel’ (berasal dari **Melody**). Skala ini didefinisikan oleh Stanley Smith, John Volkman dan Edwin Newman sebagai [13] :

$$F_{mel} = \begin{cases} 2595 \times \log_{10}\left(1 + \frac{F_{HZ}}{700}\right) & , F_{HZ} > 1000 \\ F_{HZ} & , F_{HZ} < 1000 \end{cases} \quad (7)$$

Sebuah pendekatan untuk simulasi spektrum dalam skala mel adalah dengan menggunakan filter bank yang diletakkan secara seragam dalam skala mel seperti yang ditunjukkan pada **Gambar 2** dimana setiap frame yang diperoleh dari tahapan sebelumnya difilter menggunakan M filter segitiga sama tinggi dengan tinggi satu.



Gambar 2. Contoh *mel-spaced filter bank*

Bila spektrum $F[N]$ adalah masukan proses ini, maka keluarannya adalah spektrum $M[N]$ yaitu spektrum $F[N]$ termodifikasi yang berisi *power output* dari filter-filter ini. Koefisien spektrum mel dinyatakan dengan K .

Dalam *mel-frequency wrapping*, sinyal hasil FFT dikelompokkan ke dalam berkas filter triangular ini. Proses pengelompokan tersebut adalah setiap nilai FFT dikalikan terhadap *gain filter* yang bersesuaian dan hasilnya dijumlahkan. Maka setiap kelompok mengandung sejumlah bobot energi sinyal sebagaimana dinyatakan sebagai $m_1 \dots m_p$ seperti ditunjukkan pada **Gambar 2** diatas. Proses *wrapping* terhadap sinyal dalam domain frekuensi dilakukan menggunakan persamaan (8) berikut.

$$X_i = \log_{10} \left(\sum_{k=0}^{N-1} X(k) \cdot H_i(k) \right) \quad (8)$$

dimana

X_i = nilai *frequency wrapping* pada filter $i = 1, 2, \dots, n$ (jumlah filter)

$X(k)$ = nilai magnitudo frekuensi pada k frekuensi

$H_i(k)$ = nilai tinggi filter i segitiga dan k frekuensi, dengan $k = 0, 1, \dots, N - 1$ (jumlah magnitudo frekuensi)

5. *Cepstrum*. *Cepstrum* adalah sebutan kebalikan untuk *spectrum*. *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 (9) berikut.

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

dengan

C_j = nilai koefisien C ke j

$j = 1, 2, \dots$ jumlah koefisien yang diharapkan

X_i = nilai X hasil *mel-frequency wrapping* pada frekuensi $i = 1, 2, \dots, n$ (jumlah wrapping)

M = jumlah filter

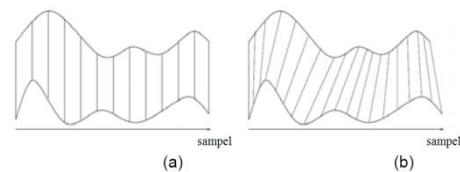
Hasil dari proses ini dinamakan *Mel-Frequency Cepstrum Coefficients* (MFCC).

D. Pengenalan Pola berbasis Dynamic Time Warping

Salah satu metode yang cukup populer di awal perkembangan teknologi pengolahan sinyal wicara untuk membandingkan sinyal uji dengan sinyal referensi adalah dengan menggunakan sebuah teknik *dynamic programming* yang juga lebih dikenal sebagai *Dynamic Time Warping* (DTW). Teknik ini ditujukan untuk mengakomodasi perbedaan waktu antara proses perekaman saat pengujian dengan yang tersedia pada basis data sinyal referensi. Prinsip dasarnya adalah dengan memberikan sebuah rentang 'steps' dalam ruang (dalam hal ini sebuah frame-frame waktu dalam sampel dan frame-frame waktu dalam basis data) 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 suatu indikasi mengenai seberapa dekatnya sample dengan basis data tersebut. Selanjutnya akan dipilih basis data yang memiliki kemiripan terbaik.

Keunggulan DTW adalah kemampuan untuk menghitung jarak dari dua vektor data dengan panjang yang berbeda. Jarak DTW diantara dua vektor dihitung berdasarkan jalur pembengkokan optimal dari kedua vektor tersebut. Ilustrasi dari metode DTW ini ditunjukkan pada **Gambar 3** dibawah ini. Penjelasan lengkap dari DTW dapat ditemukan pada [14].



Gambar 3. Pencocokan *sequence*
(a) alignment asli dari 2 *sequence* (b) alignment dengan DTW

Salah satu teknik untuk menghitung jarak DTW adalah dengan menggunakan rumus berikut.

$$D(U, V) = \gamma(m, n) \quad (10)$$

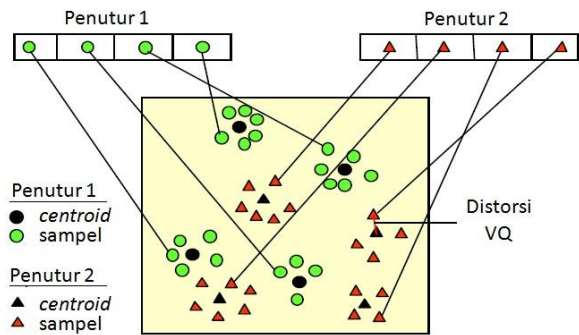
$$\gamma(m, n) = d_{base}(u_i, v_j) + \min \begin{cases} \gamma(i-1, j) \\ \gamma(i-1, j-1) \\ \gamma(i, j-1) \end{cases} \quad (11)$$

E. Pengenalan Pola berbasis Vector Quantization

Vector Quantization (VQ) adalah proses pemetaan vektor dari ruang berukuran besar menjadi sebuah wilayah yang terbatas. Masing-masing daerah tersebut disebut cluster dan dapat direpresentasikan oleh pusat (*centroid*) yang disebut *codeword*. Kumpulan dari semua *codeword-codeword* disebut *codebook*. Tujuan dari proses VQ dalam *voice processing* adalah untuk melakukan kompresi terhadap jumlah sampel/data yang ada sehingga dapat mengurangi jumlah dari proses komputasi yang ada. VQ akan mengompresi sinyal dalam domain frekuensi sesuai dengan ukuran *codebook*.

Dalam **Gambar 4** menunjukkan konsep diagram untuk ilustrasi proses pengenalan menggunakan VQ ini. Untuk menyederhanakan, hanya dua penutur dan dua dimensi dari ruang akustik yang ditunjukkan. Lingkaran menunjukkan vektor dari penutur 1 dan kotak menunjukkan vektor dari penutur 2. Pada tahap pelatihan, tiap vektor penutur akan dikelompokkan dengan meng-*cluster* tiap-tiap

vektornya. Jarak antara vektor yang dekat dengan *codeword* disebut sebagai *distortion*.



Gambar 4. Formasi VQ *codebook* antara penutur 1 dan penutur 2

Pada tahap pengenalan, sinyal masukan akan di vektor-kuantisasi menggunakan semua *trained codebook* dan selanjutnya akan dihitung nilai total *distortion*nya. Total *distortion* yang paling kecil antara *codeword* dari salah satu sinyal dalam *database* dengan VQ *codebook* dari sinyal masukan akan dipilih sebagai hasil identifikasi.

Dalam pembentukan *codebook* untuk iterasi dalam memperbaiki VG digunakan algoritma yang dikenal sebagai algoritma LBG (Linde, Buzo, dan Gray) [15]. Algoritma tersebut dapat diimplementasikan dengan prosedur rekursif berikut.

1. Penentuan vektor *codebook* yang merupakan *centroid* dari keseluruhan vektor training
2. Pengandaan ukuran dari *codebook* dengan membagi *codebook* saat ini (y_n) menggunakan persamaan berikut

$$y_n^+ = y_n(1 + \epsilon) \tag{12}$$

$$y_n^- = y_n(1 - \epsilon) \tag{13}$$

Dengan n berubah dari 1 ke ukuran *codebook* sekarang dan ϵ adalah parameter pembagi. ($\epsilon = 1$)

3. *Nearest Neighbour Search*. Yaitu pengelompokan training vektor yang mengumpul pada daerah tertentu. Selanjutnya menentukan *codeword* dalam *current codebook* yang terdekat dan memberikan tanda vektor yaitu *cell* yang diasosiasikan dengan *codeword-codeword* yang terdekat.
4. *Centroid Update*. Yaitu menentukan *centroid* baru yang merupakan *codeword* yang baru pada masing-masing *cell* dengan menggunakan *training vector* pada *cell* tersebut.

5. Iterasi 1 : mengulangi langkah 3 dan 4 hingga jarak rata-rata dibawah nilai ambang
6. Iterasi 2 : mengulangi langkah 2, 3 dan 4 hingga *codebook* berukuran M .

III. METODE PENELITIAN

Metode penelitian yang dilakukan adalah sebagai berikut.

1. Perancangan perangkat lunak pengenalan kata berbasis LabVIEW 7.1. Perangkat lunak menggunakan metode Mel Frequency Cepstrum Coefficients (MFCC), MFCC+Delta dan MFCC+Delta+Double Delta untuk proses ekstraksi ciri dan metode Dynamic Time Warping (DTW) serta Vector Quantization (VQ) sebagai proses pencocokan. Perangkat keras yang digunakan adalah Intel Core B950 2.1 GHz, 2 GB RAM.s.
2. Perakaman sinyal referensi sebagai basis data. Penelitian ini dilakukan dengan mengambil data suara dari satu orang dengan mengucapkan suatu kata sebanyak 9 kali. Setiap suara direkam dengan rentang waktu 1,5 detik dengan *sampling rate* 8000 Hz. *Cepstral coefficient* yang digunakan sebanyak 22. Terdapat sebelas kata uji yang digunakan, yaitu AC, lampu, depan, belakang, pintu, proyektor, salah, mati, hidup, tutup dan buka.
3. Pengujian basis data yang telah dibuat. Pengujian dilakukan untuk memeriksa apakah sistem mengenali perintah-perintah berikut dari pembicara : “AC mati”, “AC hidup”, “lampu mati”, “lampu hidup”, “lampu depan mati”, “lampu depan hidup”, “lampu belakang mati”, “lampu belakang hidup”, “lampu belakang mati”, “pintu kunci”, “pintu buka”, “proyektor mati”, “proyektor hidup”, “salah” (untuk reset). Terdapat tiga metode ekstraksi ciri (MFCC, MFCC+Delta dan MFCC+Delta+Double Delta) dan dua metode pengenalan pola (DTW dan VQ) yang dibandingkan. Perhitungan nilai akurasi dilakukan dengan membandingkan banyaknya hasil kata yang benar dengan kata yang diuji. Persentasi tingkat akurasi dihitung menggunakan persamaan (14) berikut.

$$\text{Hasil} = \frac{\sum \text{data yang benar}}{\sum \text{data yang diuji}} \times 100\% \tag{14}$$

4. Rancang bangun perangkat keras untuk menerima sinyal dari hasil sistem pengenalan kata pada komputer. Perangkat keras yang

dibuat meliputi *interfacing* antara komputer dengan mikrokontroller menggunakan komunikasi serial, pemrograman mikrokontroller, pembuatan rangkaian driver menggunakan relay untuk mengendalikan tegangan AC yang akan dikirim ke lampu.

IV. PERANCANGAN SISTEM

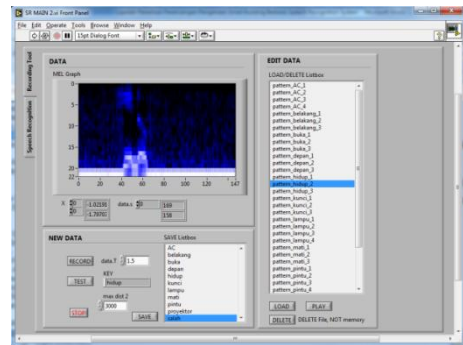
F. Perancangan Perangkat Lunak

Pada penelitian ini telah dibuat beberapa program berbasis LabVIEW untuk mendukung proses pengenalan kata ini. Perangkat lunak pengenalan kata ini menggunakan metode Mel Frequency Cepstrum Coefficients (MFCC) untuk proses ekstraksi ciri dan metode Dynamic Time Warping (DTW) sebagai proses pencocokan. Adapun program yang telah dibuat yaitu sebagai berikut.

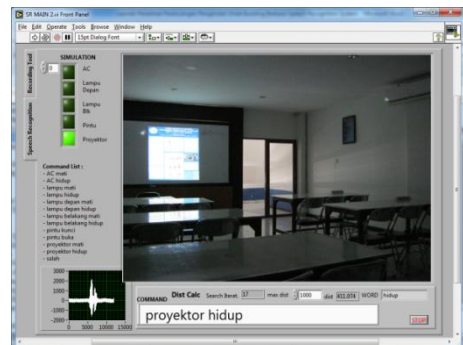
- subSR melFilterMatrix.vi : menghitung mel filterbank dengan masukan jumlah *Cepstral coefficient*, *sampling rate* dan panjang window FFT
- subSR mel2freq.vi : mengkonversi frekuensi dari skala mel ke skala linear
- subSR freq2mel.vi : mengkonversi frekuensi dari skala linear ke skala mel
- subSR loglimit.vi : menghitung nilai $\log(x)$ jika $x < \text{limit}$, jika $x > \text{limit}$ maka akan menghitung nilai $\log(\text{limit})$
- subSR computeSpectrum.vi : menghitung *log power spectrum* dengan masukan panjang FFT dan sinyal suara
- subSR computeMelSpectrum.vi : menghitung *log mel power spectrum* dengan masukan koefisien mel filter dan sinyal suara
- subSR dp_asym.vi : mengeksekusi algoritma DTW untuk mengukur jarak/kemiripan diantara dua vektor X dan Y
- subSR vqlbg.vi : mengeksekusi algoritma VQ menggunakan Linde, Buzo, Gray (LBG)
- SR RecTest.vi : membandingkan sinyal masukan dengan sinyal referensi dan mengeluarkan nilai sinyal frekuensi yang memiliki kemiripan terbaik
- subSR plotData.vi : menampilkan mel frekuensi spektrum
- subSR RecordMelMatrix.vi : merekam sinyal suara dan mengkonversikannya

- menjadi koefisien mel spectral dengan masukkan jumlah delta yang digunakan
- subSR wavplay simple.vi : mengirim file suara ke sound card
- SR MAIN 2 : file utama untuk melakukan pelatihan, pengujian dan simulasi sistem

Tampilan dari program utama untuk melakukan pelatihan/pengenalan kata ditunjukkan pada **Gambar 5**, adapaun tampilan program utama untuk melakukan pengujian dan simulasi sistem ditunjukkan pada **Gambar 6**.



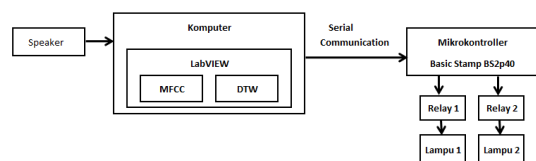
Gambar 5. Program utama untuk melakukan pelatihan/pengenalan kata



Gambar 6. Program utama untuk melakukan pengujian dan simulasi sistem

G. Perancangan Perangkat Keras

Blok diagram perangkat keras yang dibangun ditampilkan pada **Gambar 7** berikut. Adapun foto alat hasil perancangan ditunjukkan pada **Gambar 8**.



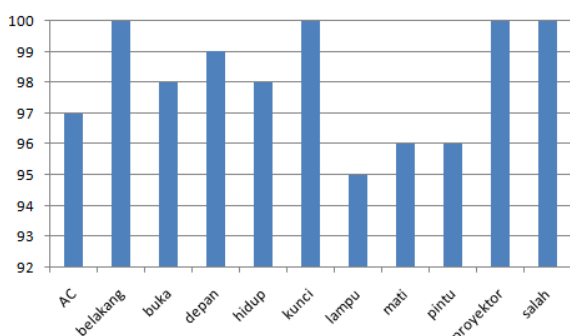
Gambar 7. Blok diagram perangkat keras yang dirancang



Gambar 8. Foto alat hasil perancangan

V. HASIL DAN PEMBAHASAN

Gambar 9 menunjukkan akurasi pengenalan oleh sistem yang dibuat menggunakan metode MFCC+Delta+Double Delta sebagai ekstraksi ciri dan DTW sebagai metoda pengenalan pola. Tabel I membandingkan akurasi pengenalan yang diperoleh menggunakan beberapa teknik.



Gambar 9. Akurasi pengenalan kata menggunakan MFCC+Delta+Double Delta dan DTW

Tabel I. Perbandingan akurasi menggunakan beberapa teknik

	Features	VQ	DTW
MFCC	13	87.9	90.2
MFCC + Delta	26	93.9	95.6
MFCC + Delta + Double Delta	39	95.6	98.1

Dari hasil diatas menunjukkan bahwa ekstraksi ciri menggunakan MFCC+Delta+Double Delta dan teknik pengenalan DTW lebih baik daripada kombinasi metode lainnya. Secara umum metode DTW memberikan hasil yang lebih baik daripada VQ.

VI. KESIMPULAN

Dari pembahasan hasil penelitian pengenalan kata yang telah dilakukan ini, dapat diambil beberapa kesimpulan sebagai berikut..

1. Perancangan perangkat lunak pengenalan kata berbasis LabVIEW menggunakan metode MFCC, MFCC+Delta, MFCC+Delta+Double Delta untuk proses ekstraksi ciri dan metode DTW dan VQ sebagai proses pengenalan telah dibuat.
2. Tingkat keberhasilan pengenalan kata yang diucapkan rata-rata adalah 93.6 %. Pengenalan kata terbaik yaitu sebesar 98.2% diperoleh menggunakan metode MFCC+Delta+Double Delta sebagai ekstraksi ciri dan metode DTW sebagai pencocokan pola.
3. Perancangan perangkat keras yang berfungsi menghubungkan keluaran pengenalan kata dari komputer ke mikrokontroller menggunakan komunikasi serial dan diteruskan untuk menghidupkan atau mematikan lampu telah dibuat.

DAFTAR PUSTAKA

- [1] Angga S, Achmad H., R. Rizal (2011), "Aplikasi Pengenalan Ucapan dengan Ekstraksi Mel-Frequency Cepstrum Coefficients (MFCC) Melalui Jaringan Syaraf Tiruan (JST) Learning Vector Quantization (LVQ) untuk Mengoperasikan Cursor Komputer", ejournal undip TRANSMISI, 13(3), pp 82 - 86
- [2] Z. Hachkar, A. Farchi, B. Mounir, J.EL Abbadi (2011), "A Comparison of DHMM and DTW for Isolated Digits Recognition System of Arabic Language", International Journal on Computer Science and Engineering (IJCSSE), Vol.3, no. 3, pp 1002 - 1008
- [3] A. Thakur, N. Singla, V.V. Patil (2011), "Design of Hindi Key Word Recognition System for Home Automation System Using MFCC and DTW", International Journal of Advanced Engineering Sciences and Technologies (IJAEEST), vol 11, no 1, pp 177 - 182.
- [4] B.W. Gawali, S. Gaikwad, P Yannawar, S.C.Mehrotra (2011), "Marathi Isolated Word Recognition System using MFCC and DTW Features", ACEEE International Journal on Information Technology, Vol 01, no 01, pp 21 - 24.
- [5] E. Kurniasari (2005), "Hybrid Jaringan Syaraf Tiruan dan Model Markov Tersembunyi Untuk Pengenalan Bilangan Berbahasa Indonesia Menggunakan CSLU Toolkit", Skripsi Departemen Ilmu Komputer, Institut Pertanian Bogor
- [6] T.D. Ganchev, (2005), Ph.D. Thesis. Wire Communications Laboratory, Department of Computer and Electrical Engineering, University of Patras, Greece
- [7] V. Tiwari (2010), "MFCC and Its Applications in Speaker Recognition", International Journal on Emerging Technologies 1(1), pp 19-22
- [8] B. Sung, M.B. Jung, I. Ko (2008), "A Feature Based Music Content Recognition method using Simplified MFCC", International Journal of Principles and Applications of Information Science and Technology, Vol. 2, No. 1, pp 13 - 23

- [9] S.V. Chapaneri (2012), "*Spoken Digits Recognition using Weighted MFCC and Improved Features for Dynamic Time Warping*", International Journal of Computer Applications, Vol 40, No. 3, pp 6 – 12
- [10] Cornaz, C. Dan U. Hunkeler. "*An Automatic Speaker Recognition System*". Mini-project
- [11] Jurafsky D, Martin JH. (2000). "*Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*". New Jersey: Prentice Hall
- [12] Do MN. (1994). "*Digital Signal Processing Mini-Project : An Automatic Speaker Recognition System*." Audio Visual Communications Laboratory, Swiss Federal Institute of Technology, Lausanne, Switzerland
- [13] Dugad R, Desai UB. (1996), *A Tutorial on Hidden Markov Models*. Technical Report, Department of Electrical Engineering, Indian Institute of Technology – Bombay, India
- [14] Eamonn Keogh (2002), *Exact Indexing of Dynamic Time Warping*. Proceedings of the 28th VLDB Conference, Hong Kong, China
- [15] Furui. S (1989), "*Digital Speech Processing, Synthesis and Recognition*", Marcel Dekker Inc. New York