

# Perintah Menggunakan Sinyal Suara dengan Mel-Frequency Cepstrum Coefficients dan Learning Vector Quantization

Neneng Nurhamidah<sup>1</sup>, Esmeralda C. Djamil, Ridwan Ilyas

Jurusan Informatika, Fakultas MIPA  
Universitas Jenderal Achmad Yani  
Cimahi, Indonesia

<sup>1</sup>[nnurhamidah0@gmail.com](mailto:nnurhamidah0@gmail.com)

**Abstrak**—Perkembangan teknologi informasi mengarahkan perintah komputer tidak hanya dari keyboard, mouse ataupun touchscreen, tetapi dapat menggunakan suara atau pikiran. Persoalan utama dari instruksi suara adalah identifikasi kata yang diucapkan sehingga perlu proses ekstraksi yang tepat. Beberapa penelitian terdahulu mengidentifikasi suara dengan menggunakan Mel-Frequency Cepstrum Coefficients (MFCC) untuk mengoperasikan komputer, instruksi lampu otomatis dan penguncian pintu. Sementara penelitian lain menggunakan Linier Predictive Cepstral Coefficients sebagai pembanding MFCC dengan hasil akurasi MFCC lebih baik ketika mengenali suara dalam kondisi bising. MFCC memiliki kemiripan dengan sistem pendengaran manusia serta merupakan perhitungan yang kuat dan hemat biaya. Penelitian ini telah membuat sistem identifikasi kata. Identifikasi dibagi atas tiga kelas yaitu “Klasik”, “Dangdut” dan “Pop”, yang digunakan untuk aksi mengoperasikan tiga jenis lagu yang sesuai. Sinyal suara diekstraksi menggunakan MFCC yang kemudian diidentifikasi menggunakan Learning Vector Quantization (LVQ). Data latih dan data uji didapatkan dari enam naracoba dan 10 kali perulangan dari yang mengucapkan kata “Klasik”, “Dangdut” dan “Pop” secara terpisah. Kemudian sinyal suara yang direkam dilewatkan praproses menggunakan Histogram Equalization, DC Removal dan Pre-emphasize untuk mereduksi noise dari sinyal suara, yang kemudian diekstraksi menggunakan MFCC. Spektrum frekuensi yang dihasilkan dari MFCC diidentifikasi menggunakan LVQ setelah melewati proses pelatihan terlebih dahulu. Hasil penelitian diperoleh akurasi sebesar 92% untuk identifikasi data yang telah dilatih. Sedangkan pengujian data baru diperoleh akurasi sebesar 46%.

**Kata Kunci**—Identifikasi suara, Instruksi perangkat eksternal, Mel-Frequency Cepstrum Coefficients, Learning Vector Quantization

## I. PENDAHULUAN

Dewasa ini perkembangan teknologi informasi mengarahkan masukan untuk memberikan perintah tidak hanya dari keyboard, mouse ataupun touch screen, tetapi dapat menggunakan suara atau pikiran melalui Brain Computer Interface. Kata – kata yang direkam dalam bentuk suara melalui microphone diubah menjadi sinyal digital. Beberapa perangkat aplikasi telah memanfaatkan input suara untuk perintah. Di sisi lain, suara yang direkam dalam bentuk sinyal suara terdiri dari amplitudo, panjang gelombang, tempo, ritmik dan frekuensi mengandung informasi kata yang diucapkan, identitas pemilik

suara ataupun keduanya. Pengenalan suara di antaranya dapat dimanfaatkan untuk menyalakan lampu otomatis, absensi dan request lagu. Selain itu, identifikasi kata yang diucapkan dapat digunakan untuk menyimpan pembicaraan dalam bentuk teks ataupun dikembangkan untuk robot sahabat. Namun identifikasi kata yang diucapkan ataupun identitas dari sinyal suara tidaklah mudah. Oleh karena itu perlunya suatu metode komputasi agar dapat melakukan identifikasi kata yang diucapkan dengan tepat.

Beberapa penelitian mengenali pola suara untuk mengontrol nyala dan mati lampu secara otomatis [1], mengoperasikan kursor [2], sistem penguncian pintu [3], perintah untuk membuka aplikasi pada komputer [4], sistem kendali peralatan rumah tangga [5], serta dapat juga digunakan untuk identifikasi gender [6] [7] dan untuk analisis hipotiroidisme pada bayi melalui suara tangisan bayi [8].

Dalam mengenali kata yang diucapkan dalam sinyal suara agar dapat digunakan sebagai perintah, diperlukan ekstraksi sinyal suara dan pembelajaran mesin. Penelitian terdahulu mengenali perintah suara dengan tingkat keberhasilan rata – rata 81% dengan menggunakan metode Mel-Frequency Cepstrum Coefficients (MFCC) setelah diekstraksi dengan metode Dynamic Time Warping (DTW) untuk pencocokan pola [3]. Penelitian lainnya membandingkan sinyal suara untuk proses verifikasi dengan menggunakan MFCC dan DTW dengan rata – rata akurasi yang didapatkan sebesar 96% [9]. Dalam kondisi bising sistem berbasis MFCC memberikan kinerja yang relatif kuat dibandingkan dengan LPCC, pada tingkat SNR 20dB sistem berbasis MFCC memiliki tingkat akurasi 97,03% [10].

Penelitian ini telah membuat sistem identifikasi kata dari perekaman dua detik sinyal suara menggunakan ekstraksi MFCC dan LVQ. Identifikasi kata dilakukan terhadap tiga kelas kata yang diucapkan, yaitu “Klasik”, “Dangdut” dan “Pop”. Hasil identifikasi digunakan untuk mengaktifkan lagu dengan genre yang sesuai dengan perintah yang dikenali. Suara yang diucapkan oleh naracoba akan dilewatkan praproses untuk memperoleh normalisasi dan memperbaiki dari noise, mengekstraksi menggunakan MFCC yang selanjutnya diidentifikasi jenis kata-kata yang diucapkan. Model komputasi yang dibangun sebagai dasar untuk perintah komputer melalui suara, termasuk pengembangan ke depan untuk robot sahabat, serta untuk konversi suara ke dalam teks.

Dalam penelitian ini pengenalan pola suara digunakan untuk memutar lagu tertentu sesuai kata kunci yang diucapkan, dengan menggunakan ekstraksi fitur MFCC dan klasifikasi Learning Vector Quantization (LVQ) yang sebelumnya dilakukan praproses untuk normalisasi ukuran data suara menggunakan metode Histogram Equalization, normalisasi data suara menggunakan DC Removal dan Pre-Emphasize filter untuk mengurangi gangguan *noise*. Data suara diambil dari enam naracoba dengan kata kunci sebanyak tiga kali dan 10 kali perulangan untuk setiap kata, sehingga jumlah data latih sebanyak 180 data. Pengujian dilakukan dengan menggunakan data yang digunakan sebagai pelatihan serta menggunakan data baru yang direkam dari naracoba yang sama dengan SNR yang berbeda. Data baru direkam sebanyak 10 kali perulangan untuk setiap kata. Sehingga jumlah data uji sebanyak 360 set.

## II. METODE

### A. Akuisisi Data

Pengambilan data suara dilakukan secara *offline* dengan tiga kata kunci yaitu “Klasik”, “Dangdut” dan “Pop”. Penentuan kata kunci tersebut dikarenakan ketiga kata tersebut memiliki pelafalan serta vokal akhir yang berbeda. Perekaman menggunakan microphone dengan SNR > 10dB, frekuensi sampling yang digunakan yaitu 8000Hz menggunakan channel mono dan resolusi 8bit. Durasi perekaman dilakukan selama 2 detik.

Data diambil melalui enam naracoba, masing – masing mengucapkan 3 kata dengan sepuluh kali perulangan untuk setiap kata, sehingga terdapat 180 data latih. Pengucapan harus dilakukan dengan artikulasi yang jelas serta durasi pengucapan harus mendekati 2 detik, hal ini dilakukan untuk dapat memperoleh data latih yang baik. Setiap satu perekaman menghasilkan data sampling sebanyak 16000, didasarkan pada spesifikasi perekaman yang telah dijelaskan, serta pengambilan data dilakukan setiap 0,2 detik.

### B. Perancangan Sistem Identifikasi

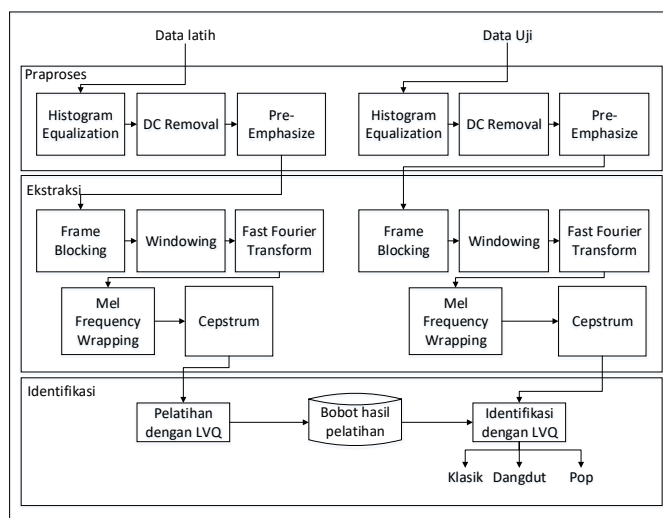
Input sistem dari penelitian ini yaitu data rekaman suara dari enam naracoba melalui *microphone* dengan frekuensi sampling 8000 Hz dan resolusi 8bit serta menggunakan *channel* mono. Data input dilakukan praproses melalui tiga tahap, yaitu Histogram Equalization, DC Removal dan Pre-Emphasize filter.

Histogram equalization bertujuan untuk meratakan jumlah data sampel menjadi 16000 sampel untuk setiap perekaman, proses DC Removal dilakukan untuk menghilangkan komponen DC sehingga diperoleh normalisasi dari data input suara, Pre-Emphasize bertujuan untuk mempertahankan frekuensi – frekuensi tinggi pada spektrum sehingga dihasilkan data input suara dengan *noise* yang rendah.

Setelah dilakukan praproses kemudian dilakukan proses ekstraksi sinyal suara untuk mendapatkan fitur dari sinyal suara menggunakan MFCC, tahapan dalam MFCC terdiri dari lima langkah yaitu Frame Blocking, Windowing, Fast Fourier Transform, Mel-Frequency Wrapping dan Cepstrum. Frame Blocking dilakukan untuk membagi sinyal suara ke dalam bentuk *short segmen* atau *short frame*, agar sinyal suara mengalami perubahan dalam jangka waktu tertentu. Sinyal

suara selama dua detik, disampling setiap 0,02 detik. Oleh karena itu, satu perekaman menghasilkan 99 *frame*/detik. Windowing dilakukan untuk mengurangi diskontinuitas sinyal yang dihasilkan dari proses Frame Blocking, Windowing dilakukan menggunakan Hamming Window, hal ini dikarenakan Hamming Window menghasilkan *sidelobe* yang tidak terlalu tinggi akan tetapi menghasilkan *noise* yang tidak terlalu tinggi. FFT dilakukan untuk mengubah sinyal dalam domain waktu menjadi domain frekuensi, hal ini dilakukan untuk dapat melakukan proses selanjutnya yang memproses sinyal dalam domain frekuensi. Mel-Frequency Wrapping dilakukan melalui Mel-Filterbank yang memiliki sinyal dalam bentuk *triangular window*, jumlah filter yang digunakan sebanyak 32 filter, sehingga terdapat 34 titik untuk setiap *frame*. Cepstrum merupakan kebalikan dari spektrum, Cepstrum dilakukan untuk mendapatkan informasi yang terdapat dalam sinyal suara, koefisien yang digunakan untuk proses identifikasi adalah 13 koefisien MFCC untuk setiap *frame*.

Proses pelatihan dan pengujian dilakukan menggunakan LVQ. Pelatihan dilakukan terhadap data sinyal suara yang dihasilkan melalui proses ekstraksi, yang mana setiap *frame* memiliki 13 koefisien Cepstrum sehingga satu perekaman menghasilkan 2574 vektor sebagai input untuk pelatihan LVQ. Kelas yang digunakan sebanyak 3 kelas, yaitu Klasik, Dangdut dan Pop. Perancangan sistem Identifikasi Kata dapat dilihat pada Gambar 1.



Gambar 1. Perancangan sistem Identifikasi Kata

### C. Praproses

Sinyal suara yang didapat dari hasil perekaman memiliki lebar data yang berbeda, hal ini disebabkan oleh kata yang diucapkan oleh naracoba dalam setiap perulangan tidak tetap. Sehingga digunakan proses Histogram Equalization dengan menghitung distribusi kumulatif dari data sampel menggunakan Persamaan 1.

$$D'_n = D[n] + D[n - 1] \quad (1)$$

Setelah mendapatkan nilai distribusi kumulatif maka dilakukan perataan histogram sehingga jumlah data sampel adalah 16000 data menggunakan Persamaan 2.

$$h[v] = \left( \frac{D'[v] - \min(D')}{\max(D') - 1} \right) \cdot N + 1 \quad (2)$$

Selanjutnya dilakukan proses DC Removal dengan menghitung nilai rata – rata dari data sampel suara kemudian mengurangkannya dengan setiap data sampel untuk memperoleh normalisasi dari data input suara dengan Persamaan 3.

$$D[i] = s[i] - \frac{\sum_{i=1}^n s[i]}{n} \quad (3)$$

Penelitian terdahulu mengenali isyarat tutur vokal setelah dilakukan praproses menggunakan DC Removal untuk menghilangkan komponen DC, diperoleh akurasi sebesar 80% untuk tester trainer, dan 87% untuk tester acak [11].

Untuk menghilangkan gangguan *noise* dari data input suara maka digunakan proses Pre-Emphasize filter serta untuk mempertahankan frekuensi – frekuensi tinggi yang umumnya tereliminasi pada saat proses produksi suara, dengan menggunakan Persamaan 4, dimana  $\alpha$  yang digunakan adalah 0,97.

$$y[n] = s[n] - \alpha s[n - 1] \quad (4)$$

#### D. Mel-Frequency Cepstrum Coefficients

Terdapat lima tahapan dalam proses ekstraksi menggunakan MFCC, yaitu Frame Blocking, Windowing, FFT, Mel-Frequency Wrapping dan Cepstrum.

Proses Frame Blocking digunakan untuk membagi sinyal audio ke dalam *frame*, untuk menentukan ukuran *frame* biasanya menggunakan “the power of two”, akan tetapi terdapat cara lain untuk menentukan panjang *frame* dengan menggunakan lapisan nol untuk mendekati pada aturan “the power of two” [12]. Waktu pengambilan data *sampling* ( $T_s$ ) setiap 20ms dikarenakan karakteristik sinyal berubah dalam selang waktu tertentu untuk merefleksikan suara yang berbeda pada orde 0,2 detik atau lebih. Dengan Persamaan 5 maka didapat jumlah *frame* adalah 99 *frame*/detik;

$$\left( \frac{(I-N)}{M} + 1 \right) \quad (5)$$

$$\text{Dengan } I = \text{sample rate} : \frac{F_s}{T_s} = \frac{16000}{2} = 8000$$

$$N = \text{sample point} = 8000 \times 0,02 = 160$$

$$M = \frac{N}{2} = \frac{160}{2} = 80$$

$$f = \frac{8000 - 160}{80} + 1 = 99 \text{ frame}$$

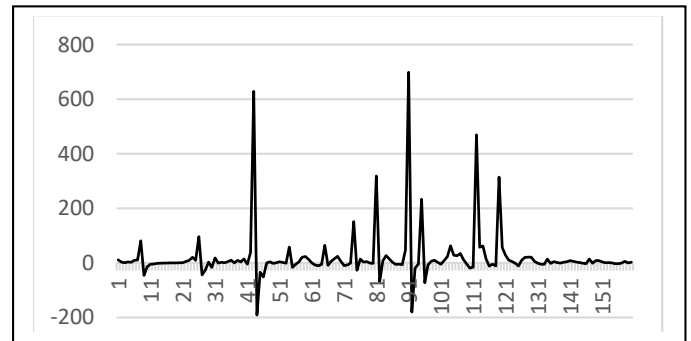
Windowing digunakan untuk meminimalkan diskontinuitas sinyal menggunakan Persamaan 6 dengan metode *windowing* yang digunakan yaitu metode Hamming Window.

$$w(n) = 0,5 - 0,46 \cos \left[ \frac{2\pi n}{N-1} \right], 0 \leq n \leq N - 1 \quad (6)$$

Dimana  $w(n)$  adalah nilai windowing ke- $n$  maka hasil dari proses ini adalah sinyal yang didapat dari Persamaan 7.

$$y(n) = x(n)w(n), 0 \leq n \leq N - 1 \quad (7)$$

Sehingga didapat sinyal hasil proses Windowing pada Gambar 2.



Gambar 2. Windowing

FFT mengkonversi setiap *frame*  $N$  sampel dalam domain waktu ke domain frekuensi. panjang FFT yang digunakan sebanyak 160, atau sama dengan panjang data setiap *frame*.

Mel-Frequency Wrapping dilakukan melalui Mel-Filterbank yang terdiri dari rangkaian Triangular Window yang saling *overlap*. Nilai – nilai mel tidak dipengaruhi oleh pilihan dasar logaritma, mengingat skala Mel menggunakan logaritma natural atau desimal.

Terdapat variasi jumlah *filter triangular* yaitu 12, 22, 32 dan 42. Akan tetapi terlalu sedikit atau banyak *filter* yang digunakan tidak akan memberikan akurasi yang tinggi. Penelitian terdahulu membandingkan jumlah *filter* dalam ekstraksi menggunakan MFCC, persentase yang paling tinggi didapat pada jumlah *filter* sebanyak 32 dengan tingkat akurasi sebesar 85% [13].

Jumlah *filter* yang digunakan sebanyak 32, yang artinya Mel-Filterbank memiliki *triangular window* sebanyak 32 buah dengan 34 titik. Batas bawah dan batas atas nilai Mel ditentukan antara 0 Hz dan 4000 Hz yang diubah ke dalam nilai Mel menggunakan Persamaan 8.

$$\text{mel}(f) = 2595 \times \log \left( 1 + \frac{f}{700} \right) \quad (8)$$

Untuk mendapatkan nilai frekuensi atau *inverse* Mel dihitung dengan Persamaan 9.

$$\text{mel}^{-1}(f) = 700 \left( e^{\left( \frac{\text{mel}(f)}{1127} \right)} - 1 \right) \quad (9)$$

Nilai frekuensi dikonversikan ke dalam nilai FFT terdekat untuk mendapatkan nilai Filterbank. Proses Filtering dilakukan untuk mendapatkan nilai *log energy* pada setiap *filter*.

Cepstrum didefinisikan sebagai kebalikan dari logaritma spektrum sinyal yang seringkali digunakan untuk mendapatkan informasi dari suatu sinyal ucapan. Proses Cepstrum berfungsi untuk mengkonversikan log mel spektrum ke dalam domain waktu, dengan menggunakan Persamaan 10.

$$c_n = \sum_{k=1}^K (S[k]) \cos \left[ \frac{\pi n (m + \frac{1}{2})}{K} \right], n = 1, 2, \dots, K \quad (10)$$

Jumlah koefisien cepstrum yang digunakan adalah sebanyak 13 koefisien untuk setiap *frame*. Penelitian terdahulu yang terkait memilih koefisien MFCC sebanyak 13 dikarenakan rentang yang sangat umum digunakan dalam MFCC adalah 10 – 20 koefisien [14].

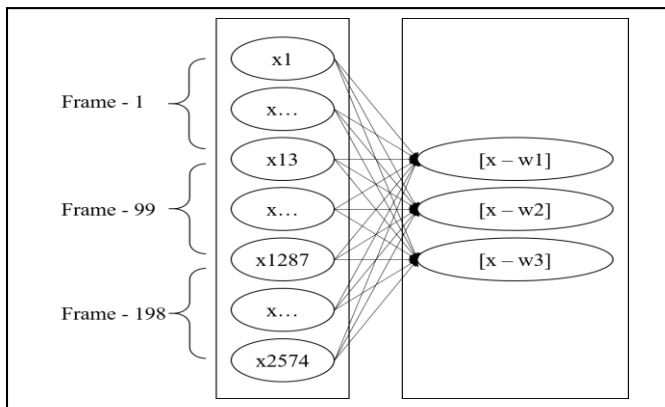
Penelitian terdahulu membandingkan efisiensi MFCC dan Linear Predictive Coding (LPC) untuk sistem pengenalan suara, didapat hasil bahwa ekstraksi menggunakan MFCC dan VQ menunjukkan akurasi yang lebih baik dibandingkan dengan menggunakan LPC dan VQ [14]. Penelitian lainnya melakukan evaluasi terhadap kinerja MFCC dan LPC untuk identifikasi bahasa otomatis, identifikasi tertinggi diperoleh dari hasil ekstraksi menggunakan MFCC [15]. Kesuksesan MFCC dikombinasikan dengan perhitungan yang kuat dan hemat biaya, hal ini menjadikan MFCC sebagai pilihan standar dalam beberapa penelitian pengenalan suara [16], diantaranya analisis hipotiroidisme pada bayi yang menghasilkan analisis bahwa suara tangisan bayi dengan hipotiroidisme memiliki frekuensi di antara 300 dan 600Hz [8], pengenalan Sistem ASR dengan akurasi dari sistem yang diusulkan lebih besar dari 96% [17], pengenalan suara menggunakan KNN dan Double Distance [18], serta verifikasi biometrika suara dengan rata – rata hasil pengujian terbaik mencapai 88% [19].

*E. Learning Vector Quantization*

LVQ merupakan suatu metode untuk melakukan pembelajaran pada lapisan kompetitif dan merupakan gabungan dari terbimbing (*supervised*). Suatu lapisan kompetitif akan secara otomatis belajar untuk mengklasifikasikan vektor-vektor *input*. Jika dua vektor *input* mendekati sama, maka lapisan kompetitif akan meletakkan kedua vektor *input* tersebut ke dalam kelas yang sama.

LVQ memiliki kesederhanaan dalam generalisasi sehingga digunakan untuk mempercepat proses komputasi

Learning Vector Quantization digunakan untuk pelatihan dengan jumlah neuron *input* sebanyak 2574 neuron dan 3 neuron *output layer*. Arsitektur jaringan LVQ untuk identifikasi kata dapat dilihat pada Gambar 3.



Gambar 3. Arsitektur LVQ

Proses pelatihan dilakukan dengan mencari nilai bobot yang kemudian digunakan untuk identifikasi vektor – vektor ke dalam kelas tujuan pada lapisan kompetitif. Tahap awal dalam proses pelatihan menggunakan LVQ yaitu menentukan *learning rate* serta maksimum *epoch*. Penelitian terdahulu melakukan klasifikasi abstrak tesis menggunakan LVQ dengan menggunakan *learning rate* dari 0.1 – 0.9, dihasilkan nilai akurasi terbaik dengan rata – rata 90% dengan menggunakan *learning rate* antara 0.1 – 0.5 [20]. Penelitian lainnya membandingkan LVQ dengan Backpropagation untuk

pengenalan wajah dengan hasil akhir yang didapatkan adalah LVQ lebih baik dibandingkan dengan Backpropagation [21]. LVQ diterapkan dalam identifikasi citra sidik jari kotor dengan tingkat keberhasilan mencapai 87% [22] dan untuk mengklasifikasi status gizi anak menggunakan LVQ1 dan LVQ3 dihasilkan akurasi sebesar 95% pada LVQ3 dan 88% pada LVQ1 [23].

Pada tahap pelatihan menggunakan LVQ ditetapkan *learning rate* ( $\alpha$ ) sebesar 0,1 dengan pengurangan setiap satu kali iterasi  $0,1 \cdot \alpha$ , minimum *learning rate* 0,0001 dan maksimum *epoch* 1000 serta bobot awal ditentukan secara acak yang mewakili setiap kelas. Jarak data latih terhadap bobot dapat dihitung dengan Persamaan 11.

$$j = \sqrt{(x_{11} - w_{11})^2 + \dots + (x_{1n} + w_{1n})^2} \tag{11}$$

Dimana  $x$  adalah data latih dan  $w$  adalah data bobot awal. Jika bobot yang memiliki jarak terkecil sama dengan kelas data latih maka bobot baru dihitung dengan Persamaan 12.

$$w_{j(\text{baru})} = w_{j(\text{lama})} + \alpha(x - w_{j(\text{lama})}) \tag{12}$$

Sebaliknya jika bobot yang memiliki jarak terkecil tidak sama dengan kelas data latih maka bobot baru dihitung dengan Persamaan 13.

$$w_{j(\text{baru})} = w_{j(\text{lama})} - \alpha(x - w_{j(\text{lama})}) \tag{13}$$

III. HASIL DAN DISKUSI

Data latih yang digunakan dalam penelitian ini sebanyak 180 data yang diperoleh dari enam naracoba tiga diantaranya laki – laki dan tiga lainnya perempuan, dari keenam naracoba tersebut terdapat naracoba tua, muda dan anak – anak, masing – masing naracoba mengucapkan tiga kata kunci yang telah ditentukan sebanyak 10 kali perulangan. Pengujian dilakukan dengan menggunakan data yang telah dilatih dan data yang tidak dilatih. Data baru didapatkan dari hasil perekaman naracoba yang sama dengan menggunakan SNR <10db.

Pelatihan terhadap data latih menggunakan LVQ dengan  $\alpha$  0.01 - 0.03 dengan pengurangan *learning rate* 0.01, serta menggunakan  $\alpha$  0.10 dengan penurunan *learning rate* 0.10, maksimum *epoch* 1000, dan minimum *learning rate* 0.0001. Analisis laju pembelajaran pelatihan LVQ dapat dilihat pada Tabel I.

TABEL I. ANALISIS PARAMETER LAJU PEMBELAJARAN LVQ

No	$\alpha$	Penurunan $\alpha$	Akurasi (%)	
			Data Latih	Data Baru
1	0.01	0.01	90	41
2	0.03	0.01	90	42
3	0.05	0.01	88	45
4	0.10	0.10	92	46

Hasil analisis pada Tabel I menunjukkan bahwa akurasi tertinggi diperoleh dengan  $\alpha$  0.1 dan penurunan *learning rate* 0.1, sehingga parameter pelatihan LVQ yang digunakan yaitu  $\alpha$  0.1, maksimum *epoch* 1000, minimum *learning rate* 0.0001, dan pengurangan  $\alpha$  0.1. Hasil pengujian terhadap 180 data latih dapat dilihat pada Tabel II.

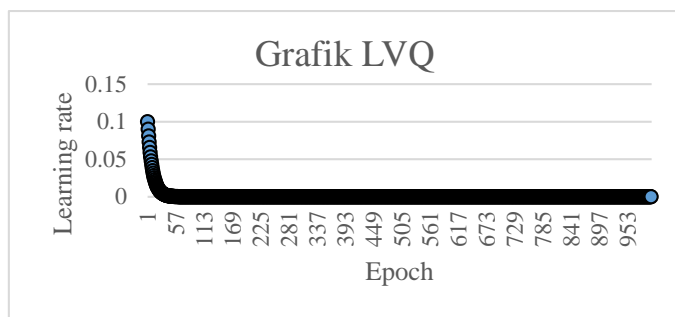
TABEL II HASIL PENGUJIAN SISTEM

No	Naracoba	Kelas	Jumlah Data	Jumlah Data Tepat Dikenali %	
				DL	DB
1	Naracoba 1	Klasik	10	100	90
2	Naracoba 1	Dangdut	10	100	90
3	Naracoba 1	Pop	10	100	70
4	Naracoba 2	Klasik	10	70	30
5	Naracoba 2	Dangdut	10	100	30
6	Naracoba 2	Pop	10	100	50
7	Naracoba 3	Klasik	10	90	60
8	Naracoba 3	Dangdut	10	100	50
9	Naracoba 3	Pop	10	70	50
10	Naracoba 4	Klasik	10	90	30
11	Naracoba 4	Dangdut	10	100	30
12	Naracoba 4	Pop	10	90	60
13	Naracoba 5	Klasik	10	100	30
14	Naracoba 5	Dangdut	10	100	20
15	Naracoba 5	Pop	10	70	30
16	Naracoba 6	Klasik	10	90	90
17	Naracoba 6	Dangdut	10	100	10
18	Naracoba 6	Pop	10	80	-
Rata – rata akurasi				92	46

Akurasi didapatkan dari menghitung persentase data yang dikenali dari semua data dengan Persamaan 14.

$$\text{akurasi} = \frac{\text{jumlah data dikenali} \times 100}{\text{total semua data}} \quad (14)$$

Perubahan learning rate untuk setiap iterasi dapat dilihat pada Gambar 4.



Gambar 4. Grafik perubahan learning rate

Analisis pengaruh praproses terhadap akurasi pengujian sistem dapat dilihat pada Tabel III.

TABEL III ANALISIS PENGARUH PRAPROSES

$\alpha$	Minimum Learning Rate	Dengan Praproses		Tanpa Praproses	
		Akurasi (%)		Akurasi (%)	
		DL	DB	DL	DB
0.1	0.0010	91	44	87	62
0.1	0.0001	92	46	86	62

Hasil analisis pengaruh praproses pada Tabel III menunjukkan bahwa pengujian data uji yang dilatih dengan menggunakan praproses memiliki akurasi yang lebih tinggi dibandingkan dengan pelatihan tanpa menggunakan praproses. Namun pengujian data baru tanpa menggunakan praproses memiliki akurasi yang lebih tinggi dibandingkan dengan pengujian data baru dengan menggunakan praproses.

## IV. KESIMPULAN

Penelitian ini telah membangun sistem identifikasi kata dari sinyal suara menggunakan Mel-Frequency Cepstrum Coefficients untuk ekstraksi sinyal dan menggunakan Learning Vector Quantization untuk pelatihan dan pengujian yang sebelumnya dilakukan praproses untuk menyamakan lebar data sampel menggunakan Histogram Equalization, DC Removal untuk normalisasi sinyal dan Pre-Emphasize untuk mengurangi noise dari data input suara. LVQ memiliki kesederhanaan dalam generalisasi sehingga digunakan untuk mempercepat proses komputasi. Hasil pengujian menggunakan maksimum epoch 1000, minimum learning rate 0.0001,  $\alpha$  0.1 dan pengurangan  $\alpha$  0.01 dari 180 data uji yang dilatih menghasilkan akurasi sebesar 92% dengan hasil pengenalan 165 data dikenali dan 15 data tidak dikenali. Sedangkan pengujian data baru menghasilkan akurasi 46% dengan hasil pengenalan 84 data dikenali dan 96 lainnya tidak dikenali. Penggunaan praproses mempengaruhi akurasi pengujian terhadap data latih. Hasil pengujian data uji yang dilatih dengan menggunakan praproses memiliki akurasi 6% lebih tinggi dibandingkan dengan pelatihan tanpa menggunakan praproses.

## UCAPAN TERIMAKASIH

Terimakasih kepada Direktorat Jenderal Pembelajaran dan Kemahasiswaan Kementerian Riset, Teknologi dan Pendidikan Tinggi atas pemberian dana Program Kreativitas Mahasiswa Bidang Penelitian Eksakta tahun 2017.

## DAFTAR PUSTAKA

- [1] M. S. W.S dan Z. Salleh, "Implementasi Pengenalan Pola Suara Menggunakan Mel-Frequency Cepstrum Coefficients dan Adaptive Neuro-Fuzzy Inference System Sebagai Kontrol Lampu Otomatis," *ALHAZEN Journal of Physics*, vol. 1, no. 1, pp. 43-54, 2014.
- [2] A. Setiawan, A. Hidayatno dan R. R. Isnanto, "Aplikasi Pengenalan Ucapan dengan Ekstraksi Mel-Frequency Cepstrum Coefficients Melalui Jaringan Syaraf Tiruan Learning Vector Quantization untuk Mengoperasikan Kursor Komputer," *TRANSMISI*, vol. 13, no. 3, pp. 82-86, 2011.
- [3] Z. Effendi, F. T. Erlina dan R. Aisuwarya, "Pengenalan Suara Menggunakan Metode Mel Frequency Cepstrum Coefficients dan Dynamic Time Warping untuk Sistem Penguncian Pintu," dalam *Seminar Nasional Teknologi Informasi dan Komunikasi Terapan (SEMANTIK)*, 2015.
- [4] A. D. Andriana, "Perangkat Lunak Untuk Membuka Aplikasi Pada Komputer Dengan Perintah Suara Menggunakan Metode Mel Frequency Cepstrum Coefficients," *Jurnal Ilmiah Komputer dan Informatika*, vol. 2, no. 1, pp. 22-26, 2013.
- [5] S. R. Jaybhaye dan P. K. Srivastava, "Dynamic Speech Recognition System to Control Home Appliances," *IJCSR*, vol. 2, no. 5, pp. 27-32, 2015.
- [6] D. Abdullah, C. I. Erliana dan I. Kamal, "Identifikasi Gender Melalui Suara Dengan Metode Statistik Ciri Orde Pertama," *Indonesian Journal on Networking and Security*, vol. 4, no. 22, pp. 49-53, 2015.
- [7] M. S. Ali, M. S. Islam dan M. A. Hossain, "Gender Recognition System Using Speech Signal," *IJCSEIT*, vol. 2, no. 1, 2012.
- [8] A. Zabidi, W. Mansor, L. Y. Khuan, R. Sahak dan F. Y. Abd Rahman, "Mel-Frequency Cepstrum Coefficient Analysis of Infant Cry with Hypothyroidism," *International Colloquium on Signal Processing & Its Applications*, pp. 204-208, 2009.
- [9] A. H. Mansour, G. Z. A. Salh dan K. A. Mohammed, "Voice Recognition using Dynamic Time Warping and Mel-Frequency Cepstral Coefficients

- Algorithm,” *International Journal of Computer Applications*, vol. 116, no. 2, pp. 34-41, 2015.
- [10] U. Bhattacharjee, “A Comparative Study Of LPCC And MFCC Features For The Recognition Of Assamese Phonemes,” *International Journal of Engineering Research & Technology*, vol. 2, no. 1, 2013.
- [11] R. V. Yuliantari, R. Hidayat dan O. Wahyunggoro, “Ekstraksi Ciri dan Pengenalan Tutar Vokal Bahasa Indonesia Menggunakan Metode Discrete Wavelet Transform dan Dynamic Time Warping Secara Realtime,” dalam *SNST ke-7*, 2016.
- [12] R. Jang, “Speech Feature MFCC,” National Taiwan University, 2013. [Online]. Available: <http://www.mirlab.org>. [Diakses 19 February 2017].
- [13] V. Tiwari, “MFCC and its Applications in Speaker Recognition,” *International Journal on Emerging Technologies*, vol. 1, no. 1, pp. 19-22, 2010.
- [14] S. Farah dan A. Shamim, “Speaker Recognition System Using Mel-Frequency Cepstrum Coefficients, Linear Prediction Coding and Vector Quantization,” *IEEE*, 2013.
- [15] E. M. Mohammed, M. S. Sayed, A. M. Moselhy dan A. A. Abdelnaiem, “LPC and MFCC Performance Evaluation with Artificial Neural Network for Spoken Language Identification,” *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 6, no. 3, pp. 55-66, 2013.
- [16] T. D. Ganchev, “Speaker Recognition,” 2005.
- [17] A. Maesa, F. Garzia, M. Scarpiniti dan R. Cusani, “Text Independent Automatic Speaker Recognition System Using Mel-Frequency Cepstrum Coefficient and Gaussian Mixture Models,” *Journal of Information Security*, vol. 3, pp. 335-340, 2012.
- [18] Ranny, “Voice Recognition using k Nearest Neighbor and Double Distance Method,” dalam *ICIMSA*, 2016.
- [19] D. Putra dan A. Resmawan, “Verifikasi Biometrika Suara Menggunakan MFCC dan DTW,” *Lontar Komputer*, vol. 2, no. 1, pp. 8-21, 2011.
- [20] F. R. Hariri, E. Utami dan A. Ambrowati, “Learning Vector Quantization untuk Klasifikasi Abstrak Tesis,” *CITEC Journal*, vol. 2, no. 2, pp. 128-143, 2015.
- [21] M. D. Wuryandari dan I. Afrianto, “Perbandingan Metode Jaringan Syaraf Tiruan Backpropagation dan Learning Vector Quantization Pada Pengenalan Wajah,” *Jurnal Komputer dan Informatika (KOMPUTA)*, vol. 1, no. 1, pp. 45-51, 2012.
- [22] M. Nasir dan M. Syahroni, “Pengujian Kualitas Citra Sidik Jari Kotor Menggunakan Learning Vector Quantization (LVQ),” *Jurnal Litek*, vol. IX, no. 1, pp. 65-69, 2012.
- [23] E. Budianita dan W. Prijodiprodo, “Penerapan Learning Vector Quantization (LVQ) untuk Klasifikasi Status Gizi Anak,” *IJCCS*, vol. VII, no. 2, pp. 155-166, 2013.