

Identifikasi Nama Surat Juz Amma dengan Perintah Suara Menggunakan MFCC dan Backpropagation

Fitri Nur Suciani¹, Esmeralda C.Djamal, Ridwan Ilyas
Jurusan Informatika, Fakultas Sains dan Informatika
Universitas Jenderal Achmad Yani
Cimahi, Indonesia
¹fitrinur440@gmail.com

Abstrak—Perintah suara dapat mengatasi keterbatasan seseorang dalam mengoperasikan perangkat eksternal seperti Al Quran elektronik. Penelitian ini telah membuat sistem identifikasi kata dalam sinyal suara untuk menampilkan surat Juz Amma. Kata yang diucapkan adalah salah satu dari 37 nama surat dalam Juz Amma. Sistem identifikasi dibangun dari pelatihan suara menggunakan 12 naracoba dengan lima kali perulangan pada setiap kata-kata yang diucapkan, sehingga digunakan 2220 set data latih. Sinyal suara setiap 2 detik dilewatkan praproses untuk mengurangi noise dan komponen DC yang kemudian diekstraksi menggunakan Mel – Frequency Cepstrum Coefficients (MFCC). Koefisien Ceptrum yang dihasilkan MFCC digunakan sebagai fitur dalam Jaringan Syaraf Tiruan menggunakan algoritma pembelajaran Backpropagation. Hasil penelitian menunjukkan sistem identifikasi menghasilkan akurasi sebesar 92% untuk data latih dan 36% untuk data baru. Hasil penelitian menunjukkan bahwa optimalisasi parameter memperoleh akurasi tertinggi menggunakan 496 hidden neuron dan 13 koefisien, dengan learning rate 0,01, minimum error 0,001 dan epoch 10000.

Kata kunci—identifikasi kata-kata, Mel-Frequency Cepstrum coefficients, Backpropagation

I. PENDAHULUAN

Perkembangan teknologi menjadikan Al Quran dapat berbentuk media elektronik yang dapat memberikan kemudahan untuk dibaca atau didengarkan. Bagi orang-orang yang memiliki kondisi fisik normal, menampilkan salah satu surat dalam Al Quran elektronik bukanlah hal yang sulit. Hal tersebut dapat dilakukan melalui perangkat masukan seperti *keyboard* atau *mouse*. Namun tidak begitu untuk orang-orang yang memiliki keterbatasan fisik. Salah satu cara untuk mengatasi keterbatasan fisik dalam mengoperasikan perangkat eksternal seperti Al Quran elektronik, yaitu dengan menggunakan perintah suara. Namun pengucapan kata untuk setiap bahasa memiliki perbedaan salah satunya dalam pengucapan menggunakan bahasa arab. Beberapa perbedaan antara pengucapan bahasa arab dan bahasa lainnya cenderung menjadi hambatan untuk orang-orang yang tidak terbiasa mengucapkan bahasa arab dalam kehidupan sehari-hari. Hambatan lainnya yaitu beberapa kata dalam bahasa arab memiliki pengucapan yang hampir serupa namun memiliki makna yang berbeda seperti halnya pengucapan nama surat

dalam Al Quran. Variasi dalam struktur bahasa tersebut memerlukan penelitian dengan pendekatan yang berbeda dalam melakukan pengenalan. Pengenalan suara sendiri sudah dikenal sejak puluhan tahun yang lalu menggunakan berbagai metode, baik untuk ekstraksi ataupun identifikasi. Namun penelitian terdahulu tidak menyebutkan secara spesifik metode terbaik untuk berbagai kasus.

Dalam pengenalan suara, sinyal suara diekstraksi terlebih dahulu. Beberapa penelitian menggunakan Mel-Frequency Cepstrum Coefficients (MFCC) dan Linear Predictive Coding (LPC). Sedangkan identifikasi suara dilakukan menggunakan pembelajaran mesin seperti Backpropagation, Learning Vector Quantization (LVQ), K-Nearest Neighbor (KNN) ataupun Hidden Markov Model (HMM). MFCC memiliki tingkat akurasi yang cukup tinggi terutama dalam kondisi bising yang memberikan kinerja yang lebih baik hingga 97% [1]. Penelitian lain menggunakan MFCC dan LVQ untuk mengoperasikan tiga jenis lagu dengan akurasi 75% [2], MFCC dan HMM sebagai pengontrol lampu otomatis [3], MFCC dan Backpropagation untuk membuka dan menutup aplikasi [4] dan MFCC-KNN untuk klasifikasi pengucapan surat Hijaiyyah [5].

Penelitian terdahulu yang terkait dengan pengenalan ucapan bahasa arab telah dilakukan untuk pemeriksaan *qalqalah* dan *tajweed* dengan menggunakan MFCC dan Vector Quantization. Evaluasi dilakukan pada pembacaan surat Al Ikhlas dan sepuluh fonem dari semua huruf *qalqalah sughrah* dan *qubrha* menghasilkan akurasi 83,9% untuk pembicara laki-laki, 82,1% untuk pembicara perempuan dan 95,0% untuk pembicara anak-anak [6]. Pengenalan ucapan berbasis DWT untuk kata-kata bahasa arab terisolasi mencapai akurasi 98,5% dengan menambahkan delta dan koefisien percepatan agar meningkatkan hasil akurasi [7]. Penelitian lainnya melakukan identifikasi terhadap tiga huruf hijaiyyah yaitu sa, sya dan tsa dengan menggunakan MFCC dan Artificial Neural Network. Penelitian tersebut menghasilkan akurasi rata-rata 92,42% [8] dan pengenalan suara arabic dengan Bionic Wavelet dan MFCC menggunakan Multi Layer Perceptron mengusulkan sebuah teknik baru dalam pengenalan bahasa arab dengan mono-locutor dan pengurangan kosa kata menghasilkan akurasi 99,39% untuk mengenali 11 kata dalam bahasa arab [9].

Penelitian ini telah membangun sistem untuk identifikasi kata-kata dari sinyal suara menggunakan MFCC untuk proses

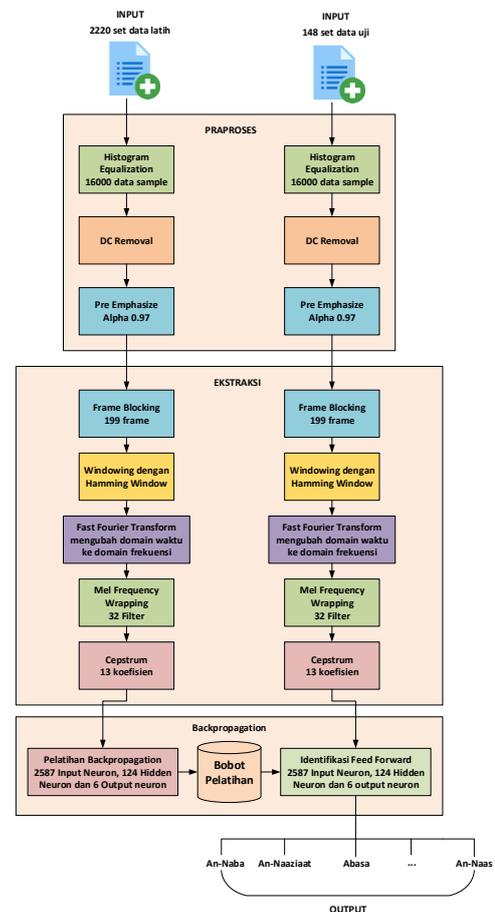
ekstraksi dan identifikasi menggunakan Backpropagation, untuk menampilkan surat Juz Amma. Sinyal suara dilakukan praproses terlebih dahulu dengan Histogram Equalization, DC Removal dan Pre Emphasize untuk mengurangi *noise* dan komponen DC, yang selanjutnya diekstraksi menggunakan MFCC. Agar dapat melakukan identifikasi kata yang diucapkan, dilakukan pembelajaran terlebih dahulu menggunakan sinyal suara yang telah diidentifikasi kelasnya yaitu 37 kelas atau sebanyak nama surat dalam Juz Amma. Pelatihan dilakukan menggunakan Backpropagation dengan 2220 set data latih yang diambil dari 12 naracoba dengan lima kali perulangan dari setiap kata yang diucapkan. Generalisasi dari pembelajaran mesin merupakan koefisien atau disebut bobot yang di simpan dalam *database*, dan kemudian digunakan untuk membuka salah satu surat dalam Juz Amma.

II. METODE

Tahap identifikasi kata-kata dari sinyal suara dapat dilihat pada Gambar 1. Sistem identifikasi sudah dilatih dengan 2220 set data latih, yang direkam menggunakan aplikasi Audacity dengan *frequency sampling* 8000 Hz, resolusi 16 bit dan *channel* mono selama 2 detik untuk setiap kata-kata yang diucapkan. Satu naracoba melakukan perekaman sebanyak 37 nama surat dalam Juz Amma dengan lima kali perulangan.

Sinyal suara diproses dengan Histogram Equalization untuk meratakan jumlah data sampel menjadi 16000 titik, DC removal untuk reduksi frekuensi DC yang menghasilkan normalisasi dari data masukan lalu dilakukan Pre Emphasize untuk mempertahankan frekuensi ciri. Ketiga proses tersebut disebut pra proses.

Sinyal suara yang telah selesai dilakukan pra proses selanjutnya akan diekstraksi untuk mendapatkan fitur atau ciri pembeda yang terdapat pada sinyal. Tahap ekstraksi dalam penelitian ini dilakukan menggunakan Mel-Frequency Cepstrum Coefficients (MFCC). Penggunaan metode tersebut, dikarenakan memiliki keunggulan yang mampu menangkap informasi penting yang terkandung dalam sinyal suara, mereduksi sinyal tanpa menghilangkan informasi penting yang ada, dan mampu mereplikasi organ pendengaran manusia dalam melakukan persepsi sinyal. Terdapat 5 langkah dalam ekstraksi menggunakan MFCC yaitu Frame Blocking, Windowing, Fast Fourier Transform (FFT), Mel-Frequency Wrapping, dan Cepstrum. Setelah mendapatkan fitur dari hasil ekstraksi, langkah selanjutnya adalah melakukan pelatihan menggunakan Backpropagation. Pelatihan dimulai dengan komputasi setiap neuron *input* terhadap setiap neuron pada *hidden layer*, yang diteruskan pada setiap neuron dalam *output layer*. Pelatihan dilakukan dengan 2587 neuron *input*, 496 neuron *hidden* dan enam neuron *output* yang didapatkan dari pengkodean bit yang dapat menampung 37 kelas atau sejumlah nama surat dalam Juz Amma. Penelitian ini menggunakan fungsi aktivasi Sigmoid Biner untuk membuat *output* berada dalam rentang 0 sampai 1. Inisialisasi bobot dilakukan secara *random* pada rentang -0.5 sampai 0.5. Luaran dari tahap ini adalah bobot-bobot yang telah diperbaharui dan akan digunakan pada tahap identifikasi.



Gambar 1. Perancangan sistem identifikasi kata

A. Praproses

Praproses dilakukan melalui tiga tahap yaitu Histogram Equalization, DC Removal dan Pre Emphasize. Histogram Equalization digunakan untuk mendapatkan panjang data sampel yang sesuai dengan panjang data yang telah ditentukan. Beberapa penelitian lain meratakan panjang data sampel menjadi 16000[2][10].

Selanjutnya dilakukan DC Removal yang bertujuan untuk membuang sinyal DC agar sinyal tepat berada di tengah-tengah sumbu $x = 0$ volt. Apabila terdapat sinyal DC yang masuk maka sinyal tidak tepat berada di sumbu $x = 0$ volt. Hal tersebut akan menimbulkan adanya *noise* sehingga menyebabkan sinyal sedikit mendengung.

Untuk meningkatkan nilai akurasi dari proses ekstraksi ciri maka dilakukan Pre Emphasize yang merupakan salah satu jenis *filter* yang sering digunakan sebelum sinyal suara diproses lebih lanjut. *Filter* ini akan mempertahankan frekuensi ciri pada sebuah spektrum yang umumnya tereliminasi pada saat proses produksi suara.

B. Ekstraksi

MFCC merupakan cara yang paling sering digunakan untuk ekstraksi, karena menggunakan komputasi logaritmik sesuai dengan lingkup pendengaran manusia. Sehingga dapat

merepresentasikan parameter suara dengan baik [11]. MFCC juga mampu mereduksi sinyal, dengan menangkap bagian yang penting tanpa menghilangkan informasi dalam sinyal [12].

Penelitian terdahulu melakukan pencarian efektivitas penggunaan ekstraksi ciri MFCC dan Subband Based Cepstral (SBC) dengan ciri pembicara secara *real time* menghasilkan MFCC lebih efektif dibandingkan SBC secara akurasi dan kecepatan proses identifikasi baik *text-dependent* maupun *text-independent* [13]. MFCC tampaknya menjadi pilihan yang lebih baik untuk mengurangi *noise* dari pada LPC karena persentase deviasi dari koefisien yang serupa memiliki jenis kebisingan yang kurang di MFCC dibandingkan dengan LPC [14]. MFCC dapat dijadikan sebagai pilihan standar dalam beberapa penelitian mengenai suara, diantaranya pengenalan ucapan pada pengujian untuk sumber terlatih 83,33% dan untuk sumber tidak terlatih 80,25% [15].

Proses Frame Blocking dilakukan karena sinyal suara terus mengalami perubahan, akibat adanya pergeseran artikulasi dari organ produksi suara. Panjang *frame* yang biasanya digunakan untuk pemrosesan sinyal antara 10-30 ms. Proses ini umumnya dilakukan secara *overlapping* untuk setiap *frame*. Penelitian terdahulu menggunakan *overlapping* 50% dari panjang *frame* keseluruhan [12].

Windowing merupakan proses pembobotan terhadap setiap *frame* berfungsi untuk mengurangi efek diskontinuitas pada ujung *frame* yang dihasilkan oleh Frame Blocking. Proses ini dilakukan dengan menggunakan Persamaan 1.

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

Proses Windowing dilakukan karena proses *framing* dapat menyebabkan terjadinya kebocoran spektral atau *aliasing*. Terdapat banyak fungsi *window*, salah satunya adalah Hamming Window. Fungsi tersebut menghasilkan *sidelobe* yang tidak terlalu tinggi dan *noise* yang dihasilkan pun tidak terlalu besar, sehingga banyak penelitian terdahulu menggunakan fungsi ini [2][3][16][17]. Dalam masalah ini Hamming Window digunakan karena memiliki hasil yang lebih baik dalam pembatasan sinyal yang akan dianalisis [18]. Fungsi Hamming Window didefinisikan pada Persamaan 2.

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

FFT merupakan algoritma yang digunakan untuk mengubah sinyal dalam domain waktu menjadi domain frekuensi, dimana pada domain waktu komponen dalam sinyal tersebut tidak dapat dilihat secara eksplisit. Fitur dalam domain waktu mencakup *zero-crossing rate*, energi rata-rata, dan amplitudo maksimum sedangkan domain frekuensi mencakup analisis spektral daya [19]. FFT dilakukan dengan menggunakan Persamaan 3.

$$X(m) = \sum_{n=0}^{N-1} x(n) \cdot \left[\cos \left(\frac{2\pi mn}{N} \right) - j \sin \left(\frac{2\pi mn}{N} \right) \right] \quad (3)$$

Mel – Frequency Wrapping digunakan untuk memetakan antara skala frekuensi sinyal ke skala logaritmik yang lebih tinggi dari 1 KHz. Hal ini membuat frekuensi sinyal sesuai dengan pendengaran manusia [20]. Skala ini didefinisikan

oleh Stanley Smith, John Volkman dan Edwin Newman pada Persamaan 4.

$$mel(f) = 1127 \ln \left(1 + \frac{f}{700} \right) \quad (4)$$

Tahap ini umumnya dilakukan dengan menggunakan *filterbank*, yang merupakan salah satu bentuk *filter* yang dilakukan dengan tujuan untuk mengetahui ukuran energi dari frekuensi *band* tertentu dalam sinyal suara. *Filterbank* dapat diterapkan baik pada domain waktu maupun domain frekuensi, tetapi untuk keperluan MFCC *filterbank* harus diterapkan dalam domain frekuensi. Terdapat beberapa jumlah *filter* yang digunakan diantaranya 12, 22, 32 dan 42. Penelitian terdahulu membandingkan jumlah *filter* dalam ekstraksi menggunakan MFCC, akurasi paling tinggi didapatkan pada jumlah *filter* 32 dengan tingkat akurasi 85% [18]. Proses *filter* tersebut dilakukan menggunakan Persamaan 5.

$$s[m] = \ln \left[\sum_{k=0}^{N-1} |X_a[k]|^2 H_m[k] \right], \quad 1 \leq m \leq M \quad (5)$$

Cepstrum merupakan tahap mengubah log Mel Sepktrum ke dalam domain waktu yang digunakan untuk memperoleh koefisien MFCC. Jumlah koefisien MFCC terdiri dari 1 sampai 23 koefisien, namun yang umumnya digunakan adalah 10-20 koefisien dan yang banyak digunakan dalam pengenalan suara adalah 13 koefisien [2][21][22].

C. Backpropagation

Backpropagation adalah salah satu algoritma dalam Jaringan Syaraf Tiruan (JST) yang menggunakan metode pelatihan *supervised*. Arsitektur MLP ditunjukkan pada Gambar 2. Beberapa penelitian menggunakan Backpropagation untuk analisis suara yaitu membuka dan menutup aplikasi dengan pengujian terbaik menggunakan *hidden* neuron 153[4], pengenalan digit 0 sampai 9 menggunakan dua *hidden layer* dengan jumlah neuron 35 buah [11], aplikasi pengenalan ucapan untuk buka dan tutup pintu menggunakan model neuron 100, 200, 1 dengan iterasi yang digunakan adalah 5000 [12], dan identifikasi bahasa otomatis dengan 2 *hidden layer* dan 30 neuron *hidden* [23]. Penelitian lainnya menggunakan 250 *neuron input*, satu *hidden layer* dengan 16 neuron *hidden*, dan 10 neuron *output* yang digunakan untuk mengenali 10 kata ucapan menghasilkan akurasi 96,332% untuk pembicara yang dikenal dan 92% untuk pembicara yang tidak dikenal [24].

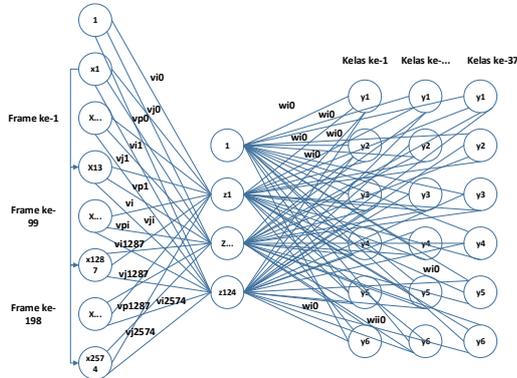
Dalam JST terdapat fungsi aktivasi yang digunakan untuk menentukan keluaran suatu neuron. Fungsi aktivasi yang digunakan harus memenuhi beberapa syarat yaitu kontinu, terdiferensial dengan mudah dan fungsi yang tidak turun [11]. Fungsi yang memenuhi syarat tersebut adalah Sigmoid Biner yang ditunjukkan pada Persamaan 12.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (12)$$

Pelatihan menggunakan Backpropagation terdapat 3 fase. Fase 1 adalah *feed forward* yaitu setiap neuron *input* akan mengirimkan sinyal masukan pada *hidden layer*. Masing-masing neuron di *hidden layer* dikalikan dengan bobot dan dijumlahkan dengan *bias* seperti ditunjukkan pada Persamaan 13 dan menghitung bobot neuron *hidden* dengan Persamaan 14.

$$z_{net_j} = v_{j0} + \sum_{i=1}^m x_i v_{ki} \quad (13)$$

$$z_j = f(z_{net_j}) = \frac{1}{1 + e^{-z_{net_j}}} \quad (14)$$



Gambar 2. Arsitektur Multilayer Perceptron

Masing-masing neuron *output* ($y_k, k = 1,2,3,\dots,m$) dikalikan dengan bobot dan dijumlahkan dengan *bias* seperti pada Persamaan 15 dan menghitung bobot neuron *output* dengan Persamaan 16.

$$y_{net_k} = w_{k0} + \sum_{j=1}^p z_j w_{kj} \quad (15)$$

$$y_k = f(y_{net_k}) = \frac{1}{1 + e^{-y_{net_k}}} \quad (16)$$

Perhitungan kuadrat *error* pada *ouput layer* menggunakan Mean Square Error (MSE). Nilai MSE merepresentasikan rata-rata kesalahan *error* absolut antara hasil identifikasi dengan nilai sebenarnya (target). Jika nilai diperoleh nol maka tidak ada lagi kesalahan, nilai yang lebih rendah dari nilai yang telah ditetapkan berarti hasilnya lebih baik [25]. Menghitung MSE dilakukan dengan menggunakan Persamaan 17.

$$\nabla E = \sum_{k=0}^m (t_k - y_k)^2 \quad (17)$$

Fase 2 adalah *back propagation*, masing-masing neuron *output* ($y_k, k = 1,2,3,\dots,m$) menerima pola target yang telah dicapai sesuai dengan pola masukan saat pembelajaran lalu hitung perubahan *bias* dengan menggunakan Persamaan 20.

$$\delta_k = (t_k - y_k) f'(y_{net_k}) = (t_k - y_k) y_k (1 - y_k) \quad (18)$$

$$\Delta W_{kj} = \alpha \delta_k z_j; k = 1,2,3,\dots,m; j = 0,1,\dots,p \quad (19)$$

$$\Delta W_{k0} = \alpha \delta_k \quad (20)$$

Pada setiap *hidden layer* dilakukan perhitungan kesalahan *hidden layer* kemudian informasi kesalahan tersebut

digunakan untuk menghitung besar koreksi bobot dan *bias* antara *input layer* dan *hidden layer*.

$$\delta_{net_j} = \sum_{k=1}^m \delta_k w_{kj} \quad (21)$$

$$\delta_j = \delta_{net_j} f'(\delta_{net_j}) = \delta_{net_j} z_k (1 - z_j) \quad (22)$$

Hitung suku perubahan bobot pada *hidden layer* dengan menggunakan Persamaan 23.

$$\Delta v_{ji} = \alpha \delta_j x_i \quad (23)$$

Hitung suku perubahan *bias* pada *hidden layer* dengan menggunakan Persamaan 24.

$$\Delta v_{j0} = \alpha \delta_j \quad (24)$$

Fase 3 adalah modifikasi bobot yaitu setelah semua faktor δ dihitung, semua bobot dimodifikasi bersamaan.

$$W_{kj} = W_{kj}(\text{lama}) + \Delta W_{kj} \quad (25)$$

$$V_{ji} = V_{ji}(\text{lama}) + \Delta V_{ji} \quad (26)$$

III. HASIL DAN DISKUSI

Sistem identifikasi kata-kata menggunakan sinyal suara telah dilakukan uji coba dengan menggunakan 2220 data latih dan 740 data uji. Pengujian sistem dibagi menjadi empat bagian yaitu pengaruh optimalisasi parameter pelatihan, jumlah koefisien ekstraksi Mel – Frequency Cepstrum Coefficients (MFCC), pengujian pada setiap naracoba dan pengujian sistem.

TABEL I. PENGARUH OPTIMALISASI PARAMETER PELATIHAN

α	Hidden Neuron	Minimum Error	MSE	Akurasi (%)	
				DL	DB
0.01	124	0.001	0.0946	61	24
		0.010	0.0962	60	21
		0.100	0.0999	57	18
	248	0.001	0.0497	83	28
		0.010	0.0325	84	30
		0.100	0.0999	39	15
	496	0.001	0.0145	92	36
		0.010	0.0166	91	34
		0.100	0.0992	37	15

Pada Tabel I menunjukkan bahwa penambahan jumlah *hidden* neuron pada saat pelatihan membuat akurasi data latih dan data baru mengalami peningkatan. Akurasi tertinggi diperoleh pada saat menggunakan 496 *hidden* neuron dengan *learning rate* 0.01, minimum *error* 0.001 dan *epoch* 10000.

TABEL II. PENGARUH JUMLAH KOEFISIEN MFCC

Koefisien MFCC	Waktu (detik)	MSE	Akurasi (%)	
			DL	DB
7	781.675	0.0245	83	23
10	821.211	0.0212	89	29
13	921.709	0.0167	92	36

Jumlah koefisien dalam MFCC cukup memiliki pengaruh dalam meningkatkan nilai akurasi. Hal ini dapat dilihat seperti hasil pengujian yang terdapat pada Tabel II. Dalam penelitian ini semakin banyak jumlah koefisien yang digunakan maka semakin tinggi nilai akurasi yang didapatkan. Hal tersebut

dikarenakan proses dalam sinyal suara memerlukan ciri pembeda yang besar. Oleh karena itu, akurasi tertinggi diperoleh pada 13 koefisien dengan nilai akurasi untuk data baru 36% dan untuk data latih 92%.

TABEL III. PENGUJIAN DATA LATIH SETIAP NARACOBA

Naracoba	Akurasi DL (%)
Naracoba 1	84
Naracoba 2	97
Naracoba 3	84
Naracoba 4	97
Naracoba 5	89
Naracoba 6	81
Naracoba 7	86
Naracoba 8	86
Naracoba 9	81
Naracoba 10	84
Naracoba 11	89
Naracoba 12	92

TABEL IV. PENGUJIAN DATA BARU SETIAP NARACOBA

Naracoba	Akurasi DB (%)
Naracoba 1	33
Naracoba 2	41
Naracoba 3	32
Naracoba 4	47
Naracoba 5	37
Naracoba 6	46

Hasil pengujian data pada setiap naracoba memiliki akurasi tertinggi yaitu 47% untuk data baru dan 97% untuk data latih. Hal ini disebabkan karena jumlah naracoba pada data latih yang digunakan hanya 12 untuk masing-masing kelas. Sehingga data latih kurang bervariasi maka saat dilakukan pengujian terhadap data baru hasil data yang dikenali tidak terlalu banyak. Sehingga hal tersebut dapat berpengaruh terhadap akurasi sistem yang dihasilkan. Hasil pengujian untuk data latih pada setiap naracoba dapat dilihat pada Tabel III dan pengujian untuk data uji pada setiap naracoba dapat dilihat pada Tabel IV.

TABEL V. HASIL PENGUJIAN SISTEM

No	Kelas	Akurasi (%)	
		DL	DB
1.	An-Naba'	100	48
2.	An-Naazi'aat	84	42
3.	'Abasa	100	92
4.	At-Takwiir	92	82
5.	Al-Infithaar	100	58
6.	Al-Muthaffifin	100	25
7.	Al-Insyqaq	100	25
8.	Al-Buruuj	92	42
9.	Ath-Thaariq	100	25
10.	Al-A'laa	100	42
11.	Al-Ghaasyiyah	100	25
12.	Al-Fajr	100	50
13.	Al-Balad	100	58
14.	Asy-Syams	92	23
15.	Al-Lail	100	8
16.	Adh-Dhuha	100	25
17.	Al-Insyirah	75	17
18.	At-Tiin	83	55
19.	Al-'alaq	92	25

No	Kelas	Akurasi (%)	
		DL	DB
20.	Al-Qadr	92	32
21.	Al-Bayyinah	92	33
22.	Al-Zalzalah	92	17
23.	Al-'adhiyaat	83	25
24.	Al-Qaari'ah	92	35
25.	At-Takaasur	100	33
26.	Al-'ashr	92	42
27.	Al-Humazah	92	25
28.	Al-Fiil	100	17
29.	Al-Quraisy	100	67
30.	Al-Maa'uun	100	45
31.	Al-Kautsar	100	8
32.	Al-Kaafiruun	58	17
33.	An-Nashr	75	8
34.	Al-Lahab	92	28
35.	Al-Ikhlash	67	8
36.	Al-Falaq	75	18
37.	An-Naas	83	17
Rata-rata akurasi		92	36

Rata-rata akurasi diperoleh dari menghitung persentase data yang dikenali dari semua data dengan persamaan 27.

$$Akurasi = \frac{Jumlah\ data\ dikenal \times 100}{total\ semua\ data} \tag{27}$$

Tabel V merupakan hasil pengujian terhadap data latih dan data baru menggunakan *learning rate* 0.01, minimum *error* 0.001 dan *epoch* 10000. Akurasi rata-rata dari data latih 92% dan untuk data baru 36%. Hasil pengujian untuk data baru tidak terlalu baik, hal ini dikarenakan kurang bervariasinya data latih yang digunakan dalam pelatihan. Sehingga pada saat sistem diuji dengan data baru yang belum dikenali, diperoleh nilai akurasi yang rendah. Selain itu, naracoba yang digunakan tidak semua fasih dalam Bahasa Arab sehingga terdapat beberapa kata yang diucapkan tidak sesuai dengan pengucapan kata yang seharusnya. Sedangkan untuk rendahnya akurasi terhadap data latih, hal tersebut dapat disebabkan karena pada saat pelatihan nilai MSE yang diperoleh masih jauh dari nilai minimum *error* yang ditetapkan sehingga kemungkinan masih banyak bobot yang belum dikoreksi dengan benar dan masih jauh dari nilai target.

IV. KESIMPULAN

Penelitian ini telah menghasilkan sebuah sistem identifikasi perintah suara untuk menampilkan surat Juz Amma menggunakan Mel-Frequency Cepstrum Coefficients (MFCC) dan Backpropagation. Penelitian ini telah menunjukkan bahwa jumlah *hidden* neuron yang digunakan memiliki pengaruh yang cukup besar untuk hasil akurasi yang lebih baik. Hasil pengujian menggunakan *learning rate* 0.01, minimum *error* 0.001 dan *epoch* 10000 dengan 496 *hidden* neuron dan 13 koefisien menghasilkan akurasi tertinggi 92% untuk data latih yang telah dikenali sebelumnya oleh sistem dan 36% untuk data baru yang belum dikenali oleh sistem. Hal tersebut dipengaruhi karena data diambil dari naracoba yang tidak semua fasih dalam membaca Al Quran. Sehingga terdapat beberapa kata yang diucapkan tidak sesuai dengan pengucapan kata yang seharusnya. Namun, dengan banyaknya

jumlah *hidden* neuron yang digunakan maka pelatihan dan pengujian membutuhkan waktu komputasi yang cukup lama.

REFERENSI

- [1] Y. Yuan, P. Zhao, and Q. Zhou, "Research of speaker recognition based on combination of LPCC and MFCC," *Proc. - 2010 IEEE Int. Conf. Intell. Comput. Syst. ICIS 2010*, vol. 3, pp. 765–767, 2010.
- [2] E. C. Djamal, N. Nurhamidah, and R. Ilyas, "Spoken Word Recognition Using MFCC and Learning Vector Quantization," in *EECSI, 2017*, September, pp. 19–21.
- [3] A. K. Munggaran, E. C. Djamal, and R. Yuniarti, "Identifikasi Suara Pengontrol Lampu Menggunakan Mel-Frequency Cepstral Coefficients dan Hidden Markov Model," in *Seminar Nasional Aplikasi Teknologi Informasi (SNATi) 2017*, 2017, pp. 17–22.
- [4] Z. Ramadhan and S. N. Endah, "Perintah Suara Berbahasa Indonesia untuk Membuka dan Menutup Aplikasi dalam Sistem Operasi Windows Menggunakan Metode Mel Frequency Cepstrum Coefficient dan Metode Backpropagation," in *Seminar Nasional Ilmu Komputer (SNIK 2016)*, 2016, Snik, pp. 33–41.
- [5] M. N. Aulia, M. S. Mubarak, U. N. W, and F. Nhita, "A Comparative Study of MFCC-KNN and LPC-KNN for Hijaiyyah Letters Pronunciation Classification System," in *2017 fifth International Conference on Information and Communication Technology (ICoICT)*, 2017, vol. 0, no. c, pp. 2–6.
- [6] A. Ismail, M. Y. I. Idris, N. M. Noor, Z. Razak, and Z. M. Yusoff, "MFCC-VQ approach for Qalqalah Tajweed rule checking," *Malaysian J. Comput. Sci.*, vol. 27, no. 4, pp. 275–293, 2014.
- [7] K. A. Darabkh, A. F. Khalifeh, B. A. Bathech, and S. W. Sabah, "Efficient DTW-Based Speech Recognition System for Isolated Words of Arabic Language," *Int. J. Comput. Electr. Autom. Control Inf. Eng. Vol7, No5, 2013*, vol. 7, no. 5, pp. 586–593, 2013.
- [8] E. S. Wahyuni *et al.*, "Arabic Speech Recognition Using MFCC Feature Extraction and ANN Classification," *icitisee*, vol. 2, no. 2, pp. 22–25, 2015.
- [9] M. Ben Nasr, M. Talbi, and A. Cherif, "Arabic Speech Recognition by Bionic Wavelet Transform and MFCC using a Multi Layer Perceptron," *2012 6th Int. Conf. Sci. Electron. Technol. Inf. Telecommun. SETIT 2012*, pp. 803–808, 2012.
- [10] A. Wahidah *et al.*, "Makhraj Recognition for Al-Quran Recitation using MFCC," *Int. J. Intell. Inf. Process.*, vol. 4, no. 2, pp. 45–53, 2013.
- [11] S. Amalia, "Pengenalan Digit 0 Sampai Digit 9 Menggunakan Ekstraksi Ciri MFCC dan Jaringan Syaraf Tiruan Backpropagation," *Tek. Elektro ITP*, vol. 6, no. 1, pp. 1–14, 2011.
- [12] A. Suroso, Y. Fitri, and S. Fitriana, "Aplikasi Pengenalan Ucapan Dengan Ekstraksi Ciri Mel- Frequency Cepstrum Coefficients (MFCC) dan Jaringan Syaraf Tiruan (JST) Propagasi Balik Untuk Buka dan Tutup Pintu," vol. 1, no. 2, pp. 121–132, 2015.
- [13] P. Prasetyawan, "Perbandingan Identifikasi Pembicara menggunakan MFCC dan SBC dengan Ciri Pencocokan LBG-VQ," in *Seminar Nasional Teknologi Informasi dan Komunikasi 2016 (SENTIKA 2016)*, 2016, vol. 2016, no. Sentika, pp. 18–19.
- [14] S. S. Khalid, S. Tanweer, D. A. Mobin, and D. A. Alam, "A comparative Performance Analysis of LPC and MFCC for Noise Estimation in Speech Recognition Task," *Int. J. Electron. Eng. Res.*, vol. 9, no. 3, pp. 377–390, 2017.
- [15] Fitrilina, R. Kurnia, and S. Aulia, "Pengenalan Ucapan Metoda MFCC-HMM untuk Perintah Gerak Robot Mobil Penjejak Identifikasi Warna," *J. Nas. Tek. Elektro*, vol. 2, no. 1, pp. 31–40, 2013.
- [16] T. B. Adam and M. Salam, "Spoken English Alphabet Recognition with Mel Frequency Cepstral Coefficients and Back Propagation Neural Networks," *Int. J. Comput. Appl.*, vol. 42, no. 12, pp. 21–27, 2012.
- [17] M. P. Sarma and K. K. Sarma, "Assamese Numeral Speech Recognition using Multiple Features and Cooperative LVQ - Architectures," *English*, vol. 5, no. 9, pp. 67–77, 2011.
- [18] V. Tiwari, "MFCC and its applications in speaker recognition," *Int. J. Emerg. Technol.*, vol. 1, no. 1, pp. 19–22, 2010.
- [19] P. D, D. S, and D. Devi A, "Pathological Voice Recognition for Vocal Fold Disease," *Int. J. Comput. Appl.*, vol. 47, no. 13, pp. 31–37, 2012.
- [20] C. Ittichaichareon, S. Suksri, and T. Yingthawornsuk, "Speech Recognition using MFCC," *Int. Conf. Comput. Graph. Simul. Model.*, pp. 135–138, 2012.
- [21] Abriyono and A. Harjoko, "Pengenalan Ucapan Suku Kata Bahasa Lisan Menggunakan Ciri LPC, MFCC, dan JST," *Indones. J. Comput. Cybern. Syst.*, vol. 6, no. 2, pp. 23–34, 2012.
- [22] S. Sarma and A. Barman, "MULTILINGUAL SPEECH IDENTIFICATION USING," vol. 5, no. 1, pp. 1–6, 2015.
- [23] E. Mansour, M. S. Sayed, A. M. Moselhy, and A. A. Abdelnaiem, "LPC and MFCC Performance Evaluation with Artificial Neural Network for Spoken Language Identification," *Int. J. Signal Process. Image Process. Pattern Recognit.*, vol. 6, no. 3, pp. 55–66, 2013.
- [24] A. H. Md, M. R. Md, U. Kumar Prodhan, and F. K. Md, "Implementation of back-propagation neural Network for isolated bangla speech recognition," *Int. J. Inf. Sci. Teh.*, vol. 3, no. 4, pp. 1–9, 2013.
- [25] M. Singh and K. Verma, "Speech recognition using neural networks," *Int. J. Technol. Eng. Syst.*, vol. 2, no. 1, pp. 108–110, 1995.