

Good Morning to Good Night Greeting Classification Using Mel Frequency Cepstral Coefficient (MFCC) Feature Extraction and Frame Feature Selection

Klasifikasi Ucapan Selamat Pagi Sampai Selamat Malam Menggunakan Ekstraksi Ciri Mel Frequency Cepstral Coefficient (MFCC) dan Pemilihan Fitur Frame

Heriyanto

Informatika, Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia

heriyanto@upnyk.ac.id

*: *Penulis korespondensi (corresponding author)*

Informasi Artikel

Received: 30 December 2020

Revised: 25 January 2021

Accepted: 17 February 2021

Published: 28 February 2021

Keywords: extraction of features, features, frames, cepstral coefficient, linear

Kata kunci: ekstraksi ciri, fitur, frame, cepstral coefficient, linier

Abstract

Purpose:

Select the right features on the frame for good accuracy

Design/methodology/approach:

Extraction of Mel Frequency Cepstral Coefficient (MFCC) Features and Selection of Dominant Weight Normalized (DWN) Features

Findings/result:

The accuracy results show that the MFCC method with the 9th frame selection has a higher accuracy rate of 85% compared to other frames.

Originality/value/state of the art:

Selection of the appropriate features on the frame.

Abstrak

Tujuan:

Pemilihan fitur yang tepat pada frame untuk mendapatkan akurasi yang baik.

Perancangan/metode/pendekatan:

Ekstraksi Ciri Mel Frequency Cepstral Coefficient (MFCC) dan Pemilihan Fitur Normalisasi Bobot Dominan (NBD).

Hasil:

Hasil akurasi didapat bahwa metode MFCC dengan pemilihan frame yang ke-9 mempunyai tingkat akurasi yang lebih tinggi sebesar 85% dibandingkan dengan frame yang lain.

Keaslian/ state of the art:

Pemilihan Fitur yang tepat pada frame .

1. Pendahuluan

Selamat pagi, selamat siang, selamat sore dan selamat malam ucapan salam dalam bahasa Indonesia sehari-hari dan klasifikasi ucapan selamat tersebut dilakukan dengan pemilihan fitur frame. Pemilihan fitur frame sangat diperlukan dalam pengenalan ucapan, karena hanya pemilihan fitur yang tepat sangat berpengaruh terhadap akurasi dan tidak semua fitur terutama semua fitur frame dipakai untuk akurasi. Metode ekstraksi ciri umum dan banyak digunakan diantaranya menggunakan Mel Frequency Cepstral Coefficient (MFCC) dan Linier Predicted Code (LPC).

1.1. Metode ekstraksi ciri

Metode pengenalan suara yang menggunakan ekstraksi ciri sangatlah penting karena hasil dari ekstraksi ciri tersebut yang berupa fitur sangat berpengaruh terhadap hasil kecocokan dan pengecekan pengenalan pola. Penelitian yang menggunakan metode ekstraksi ciri di antaranya *Mel Frequency Ceptrum Coefficients* (MFCC) dan *Linear Predictive Code* (LPC) [1]. Kedua metode tersebut mempunyai kelemahan dan kelebihan di dalam ekstraksi ciri yang menghasilkan fitur.

MFCC memiliki kelemahan di antaranya terletak pada frekuensi rendah, kebisingan lingkungan, kepekaan, pola suara yang hampir mirip dan pengklasifikasian (Syafria dkk., 2014). Sedangkan, MFCC memiliki kelebihan di antaranya mampu menangkap karakteristik suara yang penting dalam pengenalan, menangkap informasi yang penting dalam suara, menghasilkan data seminimal mungkin tanpa menghilangkan informasi dan mereplikasi suara pendengaran manusia [3]. Selain itu, ekstraksi ciri menggunakan MFCC banyak digunakan untuk pengenalan suara dikarenakan lebih presisi dalam berbagai kondisi [4].

Metode ekstraksi ciri yang menggunakan *Linear Predictive Code* (LPC) mempunyai kelemahan di antaranya derau atau *noise*, frekuensi ucapan yang berubah-ubah dan pengklasifikasian (Irmawan dkk., 2014). Metode ini mempunyai kelebihan berupa autokorelasi [1], [6]. Penelitian ekstraksi ciri suara baik yang menggunakan MFCC dan LPC mempunyai kelemahan yang sama di antaranya derau atau *noise*, frekuensi ucapan yang hampir mirip, frekuensi yang sering berubah-ubah dan pengklasifikasian. Kelemahan kedua metode tersebut juga diungkapkan oleh Abriyono dan Harjoko [1] bahwa ekstraksi ciri yang menggunakan MFCC dan LPC tidak cocok untuk mengenali suara jumlah yang sangat besar sehingga diperlukan pengklasifikasian. Berdasarkan kelemahan dan kelebihan dari kedua metode tersebut baik ekstraksi ciri yang menggunakan MFCC dan maupun LPC, maka peneliti lebih memilih ekstraksi ciri yang menggunakan MFCC dikarenakan dari tingkat keakurasian lebih baik dibandingkan dengan LPC. Ekstraksi ciri MFCC di antara 58-75%. Selain itu, metode LPC menurut Widodo dkk., [9] lebih cocok untuk komputasi yang bersifat *linear*, sedangkan suara manusia pada dasarnya bersifat *nonlinier*.

Penelitian pengenalan suara yang menggunakan MFCC telah banyak dilakukan dalam semua bidang termasuk diterapkan dalam bidang bahasa. Penelitian pengenalan suara dalam bidang bahasa Arab oleh Chamidy [4] bahwa ekstraksi *Mel Frequency Cepstral Coeffisients* (MFCC) berupa fitur untuk mendapatkan nilai kesesuaian pada penutur Indonesia terhadap penutur aslinya diklasifikasi menggunakan *Hidden Markov Model* (HMM). MFCC diterapkan dalam bidang bahasa yang lain di antaranya bahasa Indonesia dengan mengidentifikasi sinyal ucapan menjadi kosakata yang menghasilkan *Phonem and Syllable Models* dan segmentasi [10].

Penelitian serupa dilakukan oleh Suyanto dan Putra [11] menggunakan *Mel Frequency Cepstral Coefficient* (MFCC) dan *Hidden Markov Model* (HMM) yang mampu mengenali segmentasi fonem dalam bahasa Indonesia. Penelitian sejenis tentang fonem juga dilakukan oleh Cahyarini dkk., [12] yang mampu mengidentifikasi ucapan jeda antara fonem.

Penelitian lain yang terkait dengan bahasa tentang pengenalan huruf hijaiyyah oleh Bethaningtyas [13] menggunakan MFCC dengan membandingkan penggunaan 3, 6, 9 dan 12 channel dari model data latih dan nilai deviasi. Penelitian lain terkait huruf hijaiyyah oleh Heriyanto [14] menggunakan metode *average energy* dan *deviasi wave* sebagai pembanding. Sedangkan penelitian lain terkait fonem huruf hijaiyyah oleh Subali dkk., [15] menggunakan metode LPC dan DTW menghasilkan frekuensi formant pembicara dalam pengucapan dan DTW mempunyai kelebihan berupa autokorelasi.

Penelitian MFCC yang lain dengan cara memodifikasi dilakukan oleh [16] pada bagian *windowing*. Penelitian lain juga dengan memodifikasi MFCC menghasilkan analisis *acoustic signal* dengan tahapan *preemphasis*, *frame blocking*, *hamming windowing*, *Fast Fourier Transform*, *Mel Filterbank*, *Discrete Cosine Transform* (DCT), *Delta energy* dan *delta spectrum*.

2. Metode/Perancangan

Penelitian pengenalan suara menggunakan cara yang berbeda-beda menghasilkan keluaran yang berbeda-beda pula di antaranya dengan cara jaringan syaraf tiruan atau *neural network* [18], *Hidden Markov Model* (HMM) [4], pengenalan suara dengan *Dynamic Time Wrapping* (DTW) [19]. Pengenalan suara dengan metode DTW dilakukan untuk menghitung jarak antara dua data *time series*. Metode ini mempunyai keunggulan mampu menghitung jarak antara dua vektor data dengan panjang yang berbeda atau mengetahui nilai jarak terkecil pencocokan antara suara pembicara pemula dan pembicara ahli. DTW menurut Miftahuddin dan Hakim [19] merupakan algoritma sebagai *non-linear sequence alignment*, digunakan untuk mengukur kemiripan suatu pola dalam kawasan rentetan data bervariasi terhadap waktu dan lebih realistik.

DTW mempunyai kelemahan dari segi akurasi yaitu dengan hasil yang sangat bervariasi Novianto dan Yuliantari [21] dan masih menyamai tingkat akurasi HMM [4]. Sedangkan penggunaan metode HMM menurut Suyanto dan Putra [11] mempunyai kelemahan dari segi kurang tahan atau *robust*. Metode pengenalan suara yang lain menggunakan *Neural Network* (NN) mempunyai kelebihan pada sistem pembelajaran, akuisisi pengetahuan, klasifikasi dan generalisasi suatu pola [18]. NN menurut Martyna dan Sudaryanto [22] mempunyai kelemahan pada proses *training* memerlukan waktu yang cukup lama dengan jumlah data yang besar. Pernyataan yang sama oleh Aibinu dkk., [23] untuk mengidentifikasi ucapan angka satu sampai dengan angka sembilan mempunyai kendala pada saat proses *training* dengan data yang sangat besar membutuhkan waktu proses yang sangat lama.

Metode MFCC diperkenalkan pertama kali oleh Davis dan Mermelstein sekitar tahun 1980. MFCC salah satu metode yang cukup baik dalam pengenalan suara dalam bidang *speech recognition* [24]. MFCC merupakan ekstraksi ciri yang paling banyak digunakan dalam bidang *speaker recognition* dan *speech recognition*. MFCC merupakan ekstraksi ciri yang menghasilkan fitur atau ciri yang merupakan pembeda satu dengan yang lain berupa parameter

cepstral coefficient [1]. Ekstraksi ciri *Mel Frequency Cepstral Coefficient* (MFCC) mengubah gelombang suara menjadi beberapa tipe parameter seperti *cepstral coefficient* yang mempresentasikan *audio file* [4]. Selain itu MFCC menghasilkan fitur vektor yang mengonversi sinyal suara menjadi beberapa vektor untuk pengenalan fitur suara (Putra dkk., 2011).

MFCC mempunyai tahapan yaitu *pre-emphasis*, *frame blocking*, *windowing*, *Fast Fourier Transform* (FFT), *Mel Frequency Wrapping* (MFW), *Discrete Cosinus Transform* (DCT) dan *cepstral liftering* yang menghasilkan parameter sebagai fitur yaitu *frame* dan *cepstral coefficient*.

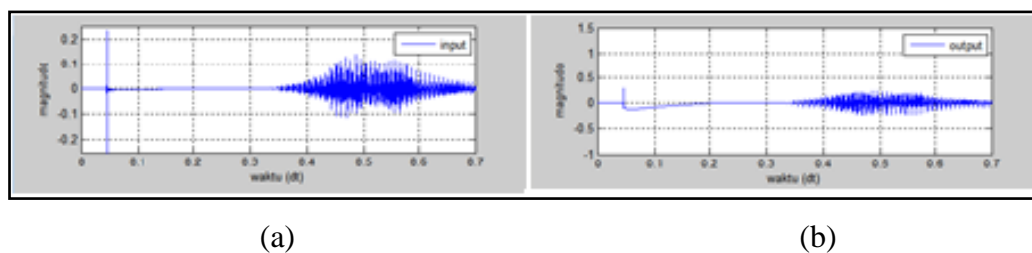
2.1 Pre-emphasis

Pre-emphasis menurut Chitode [25] melakukan penekanan pada komponen frekuensi tinggi, menyelaraskan pada frekuensi rendah dan frekuensi tinggi. *Pre-emphasis* menurut Tokunbo Ogunfunmi [26] mengurangi kebisingan untuk memperbaiki *Signal to Noise Rasio* (SNR) dan mengurangi suara yang tidak diinginkan. *Pre-emphasis* merupakan pemrosesan signal sederhana yang pada dasarnya adalah *filter linear* dan masih dalam ranah waktu [27].

Pre-emphasis merupakan tahap awal dalam proses MFCC. Tahap ini, dilakukan karena sinyal sering sekali mengalami gangguan *noise*, sehingga diperlukan untuk mengurangi *noise* atau derau. *Pre-emphasis* merupakan cara yang sangat sederhana dalam mengurangi *noise*. *Pre-emphasis* mempunyai tujuan agar level *base band* pada bagian frekuensi tinggi masih memiliki kualitas sinyal yang baik. Proses *pre-emphasis* menurut Proakis dan Manolakis [28] dengan nilai α di antara 0 sampai 1 atau di antara $0,9 \leq \alpha \leq 1,0$ menggunakan persamaan (1)

$$y(n) = s(n) - \alpha s(n - 1). \quad (1)$$

Dalam hal ini, $y(n)$ adalah signal hasil *pre-emphasis*, sedangkan $s(n)$ adalah signal sebelum *pre-emphasis*, simbol n merupakan nomor urut signal, α merupakan konstanta *filter pre-emphasis* di antara 0,9-1,0 dan s adalah signal. Proses *pre-emphasis* yang dilakukan seperti pada Gambar 1 menunjukkan bagian (a) merupakan suara masukan sebelum dilakukan *pre-emphasis*, sedangkan Gambar 1 bagian (b) adalah hasil *output* pengolahan signal setelah proses *pre-emphasis*.



Gambar 1. *Pre-emphasis* (a) sebelum dan (b) sesudah

Pengambilan signal ke- n pada *pre-emphasis* dilakukan sepanjang ucapan satu kata atau dua kata dengan waktu satu sampai tiga detik.

2.2 Frame blocking

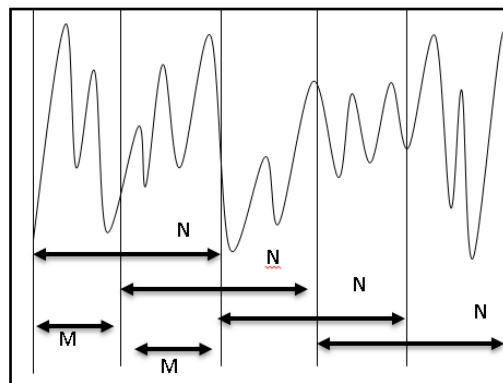
Sinyal yang telah melalui *pre-emphasis* selanjutnya dilakukan proses *frame blocking* yang diblok dalam *frame* dengan N sampel dan digeser sebesar M sampel sehingga $N=2M$

dengan $M < N$. Gambar 2 menunjukkan ilustrasi *frame blocking* [1]. Lebar *frame* dinotasikan dengan N , sedangkan lebar pergeseran setiap *frame* sebagai M . Lebar *overlap* dihitung selisih $N - M$.

Frame blocking menurut Holmes [27] menganalisis sinyal ucapan ke dalam *frame* yang diwakili oleh fitur vektor tunggal digambarkan dalam spektrum rata-rata interval waktu. Waktu dalam *frame* diambil rata-rata antara 20-40 milidetik menurut Chamidy [4]. *Frame* diambil sepanjang mungkin untuk mendapatkan resolusi frekuensi yang baik, sedangkan waktu sependek mungkin dimaksudkan untuk mendapatkan ranah waktu yang terbaik. Perhitungan jumlah *frame blocking* menggunakan persamaan (2)

$$f_l(n) = y(Ml + n). \quad (2)$$

Dalam hal ini, $f_l(n)$ merupakan hasil *frame blocking*, simbol n adalah $0, 1, \dots, N-1$. Simbol N adalah jumlah sampel, M adalah panjang *frame*, l adalah $0, 1, \dots, L-1$. Simbol L adalah seluruh signal dan y adalah hasil *pre-emphasis*.



Gambar 2. Ilustrasi *Frame blocking* [1]

Gambar 2 menunjukkan M adalah *frame* pertama signal suara dalam formula disimbolkan f_l kemudian $M+M=N$.

2.3 Windowing

Windowing menurut Proakis dan Manolakis [28] mempunyai fungsi efek menghaluskan pada spektrum setelah melalui proses *frame blocking*. *Windowing* mempunyai tujuan untuk mengurangi efek *diskontinue* pada ujung-ujung *frame* yang dihasilkan oleh proses *frame blocking*. *Windowing* yang digunakan yaitu *Rectangular Window*, *Hamming Window* dan *Hanning Window* [4]. Dari ketiga fungsi *windowing* tersebut peneliti menggunakan *windowing Hanning* dikarenakan lebih halus dibandingkan dengan yang lain [29]. Representasi fungsi *windowing* menggunakan persamaan (3)

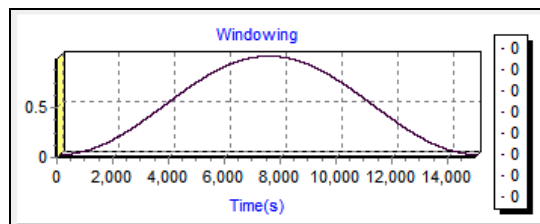
$$X(n) = f_l(n)w(n). \quad (3)$$

Dalam hal ini, fungsi $X(n)$ adalah signal hasil *windowing*, dengan f_l adalah hasil *frame blocking*, dengan n adalah $0, 1, \dots, N-1$. Simbol N adalah jumlah sampel dalam masing-masing

frame dan $w(n)$ adalah fungsi *window*. Sedangkan fungsi *windowing Hanning* menggunakan persamaan (4)

$$w(n) = 0,5 \left(1 - \cos \left(\frac{2\pi n}{M-1} \right) \right) \quad (4)$$

Dalam hal ini, $w(n)$ adalah fungsi *window* menggunakan *hanning*, dengan n adalah $0, 1, \dots, M-1$, M adalah *panjang frame*. Gambar 3.3 menjelaskan tentang hasil proses *windowing* menggunakan *Hanning window*.



Gambar 3. Ilustrasi *windowing* [28]

2.4 Fast Fourier Transform (FFT)

Fast Fourier Transform adalah pengembangan dari algoritma *Discrete Fourier Transform* (DFT) yang digunakan untuk mengubah sinyal digital pada ranah waktu ke ranah frekuensi [1]. Sinyal ini pada dasarnya merepresentasikan dekomposisi sinyal dalam hal komponen sinusoidal. Sinusoidal adalah sinusoid dari frekuensi yang sama tetapi amplitudo dan *fase* yang berbeda. FFT adalah algoritma yang dikembangkan oleh Cooley dan Turki merupakan proses signal dari ranah waktu menjadi ranah frekuensi.

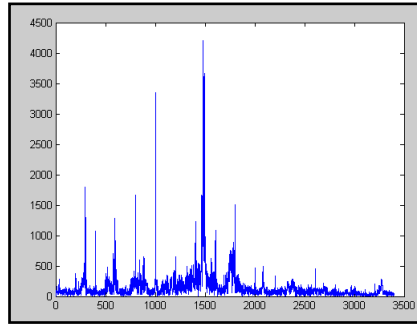
Transformasi *fourier* menurut Proakis dan Manolakis [28] salah satu dari beberapa alat matematika yang berguna dalam analisis dan desain sistem *Linier Time Invarian* (LTI) dan deret *fourier*. FFT adalah salah satu metode algoritma cepat untuk dapat mengimplementasikan *Discrete Fourier Transform* (DFT). DFT menurut Proakis dan Manolakis [28] alat komputasi yang memainkan peran yang sangat penting dalam banyak aplikasi pemrosesan sinyal digital, seperti analisis frekuensi, estimasi spektrum daya, dan *filter linear*. Waktu komputasi DFT terlalu lama dan tidak efisien kemudian FFT dapat melakukan efisiensi perhitungan. Seperti yang dikatakan menurut Proakis dan Manolakis [28] bahwa metode FFT digunakan sebagai cara yang efisien untuk dapat menghitung DFT. *Discrete Fourier Transform* (DFT) menggunakan persamaan (5)

$$d[m] = \sum_{n=0}^{N-1} X(n) e^{-j \frac{2\pi}{N} nm} ; m = 0, 1, 2, \dots, N-1. \quad (5)$$

Dalam hal ini, $d[k]$ adalah hasil perhitungan DFT, simbol $X(n)$ adalah hasil *windowing*. Simbol N adalah bilangan natural, N adalah jumlah sampel yang akan diproses ($N \in \mathbb{N}$). Simbol k adalah variabel *frequency discrete* bernilai ($m=N/2, m \in \mathbb{N}$). *Fast Fourier Transform* mempunyai tujuan mendekomposisi sinyal menjadi sinyal sinusoidal yang berupa unit *real* dan unit *imaginer*. *Fast Fourier Transform* menggunakan persamaan (6)

$$T(m) = \sum_{n=0}^{N-1} X(n) \cos\left(\frac{2\pi mn}{N}\right) - \sum_{n=0}^{N-1} X(n) \sin\left(\frac{2\pi mn}{N}\right) \quad (6)$$

Dalam hal ini, fungsi $T(m)$ adalah hasil perhitungan *Fast Fourier Transform* ke- m , simbol $X(n)$ adalah hasil perhitungan *windowing* ke- n . Simbol n adalah nomor urut sinyal. Simbol m adalah indeks dari frekuensi $(1, 2, \dots, N)$. Gambar 4 menunjukkan hasil spektrum dengan FFT.



Gambar 4. FFT menghasilkan spektrum [30]

2.5 Mel Frequency Wrapping (MFW)

Mel Frequency Wrapping (MFW) merupakan *filter* berupa *filterbank* untuk mengetahui ukuran energi dari *frequency band* tertentu dalam signal suara. MFW menurut Laha [31] mengubah frekuensi ke dalam mel.

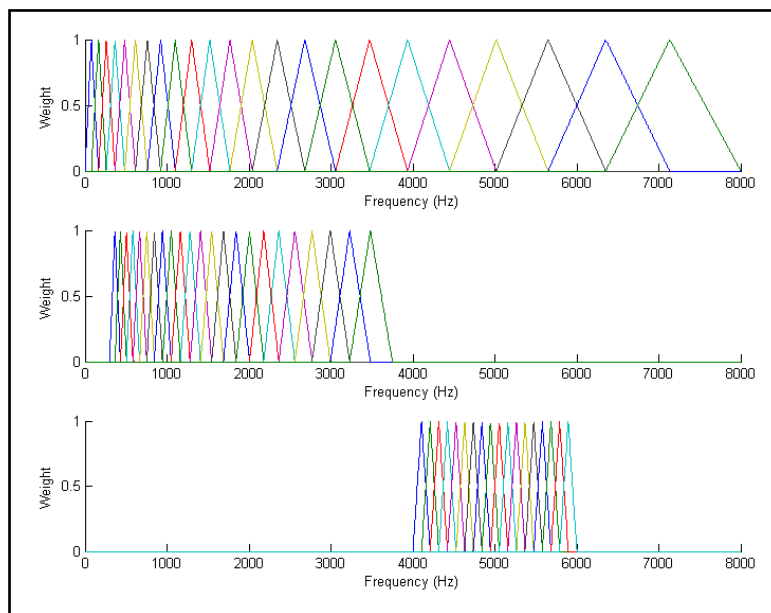
Sedangkan MFW menurut Tshilidzi Marwala [32] berisi *filterbank*, yang diberi jarak pada skala mel. *Filterbank* memiliki respon frekuensi lewat jalur segitiga yang jarak dan besarnya ditentukan oleh interval frekuensi yang konstan. Proses *output* yang diperoleh dari *filter* dikenal dengan spektrum mel. MFW mempunyai tujuan menghasilkan spektrum mel menggunakan persamaan (7)

$$Y[i] = \sum_{j=1}^G T[j]H_i[j] \quad (7)$$

Dalam hal ini, $Y[i]$ adalah hasil perhitungan *mel frequency wrapping* ke- i dengan G adalah jumlah *magnitude spectrum* ($G \in \mathbb{N}$), simbol $T[j]$ adalah hasil FFT, $H_i[j]$ adalah *coefficient filterbank* pada frekuensi j ($1 \leq i \leq E$), dan E adalah jumlah *channel* dalam *filterbank*. Pendekatan yang digunakan dalam bentuk mel menggunakan persamaan (8)

$$\text{mel}(f) = 2595 \log_{10} \left(1 + \frac{f}{700} \right). \quad (8)$$

Dalam hal ini, mel menggunakan frekuensi dengan skala mel, f sebagai *frequency*. MFW menghasilkan berupa spektrum mel. *Mel frequency scale* adalah *linear* frekuensi skala pada frekuensi di bawah 1.000 Hz dan merupakan *logarithmic* skala pada frekuensi di atas 1.000 Hz (Putra, 2011). Gambar 5 menunjukkan proses spektrum mel.



Gambar 5. Spektrum mel

Gambar 5 menunjukkan skala mel dibuat dari *filter bank* menggunakan jenis *filter triangular* dalam warna sedangkan *weight* dalam hal ini *db* merupakan *amplitude*.

2.6 Discrete Cosine Transform (DCT)

DCT menurut Smith [34] merupakan kerabat dari transformasi *fourier* terdekomposisi sinyal ke gelombang *cosinus*. DCT juga banyak dipakai untuk mengolah *image processing* misalkan file JPEG. Pada dasarnya konsep dari DCT memiliki kesamaan dengan *inverse fourier transform*. Namun hasil dari DCT adalah mendekati *Principle Component Analysis* (PCA). PCA adalah metode *static* klasik yang digunakan secara luas dalam analisis data dan kompresi. DCT diasumsikan menggantikan *inverse fourier transform* dalam proses ekstraksi ciri MFCC. *Discrete Cosine Transforms* (DCT) adalah anggota dari kelas transformasi unit sinusoidal (Britanak dkk., 2007). DCT mempunyai tujuan menghasilkan seprtrum mel untuk meningkatkan kualitas pengenalan. DCT menggunakan persamaan (9)

$$C_r = \sum_{k=1}^K (\log_{10} Y[i] \cos \left[r(i - \frac{1}{2}) \frac{\pi}{K} \right]); r = 1, 2, \dots, K. \quad (9)$$

Dalam hal ini, C_m adalah *Coefficient*, di mana $Y[i]$ adalah keluaran dari proses *filterbank* pada indek, r adalah banyaknya *coefficient* dan K merupakan jumlah *coefficient* yang diharapkan. Proses DCT menghasilkan seprtrum mel.

2.7 Cepstral Liftering

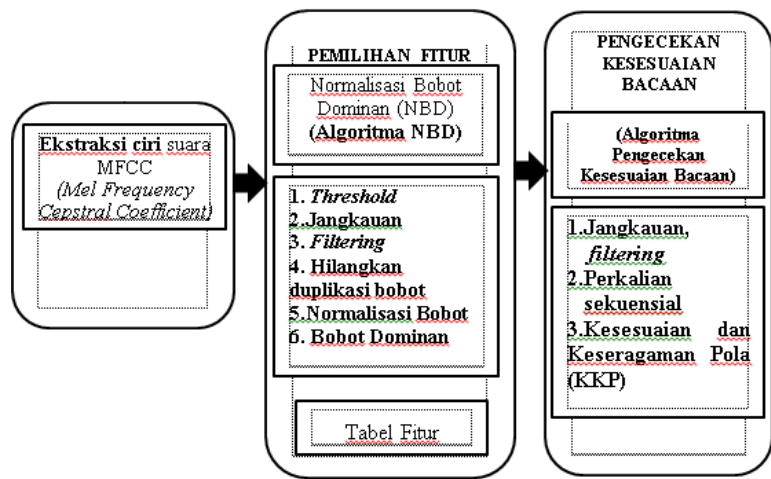
Cepstral coefficient menurut Proakis dan Manolakis [28] merupakan *the fourier series coefficients*. Seprtrum didefinisikan oleh Rabiner dan Schafer [36] adalah kebalikannya transformasi *fourier*. *Cepstral liftering* meningkatkan akurasi digunakan untuk *pattern matching*, baik *speaker recognition* maupun *speech recognition*. *Cepstral coefficient* menggunakan persamaan (10)

$$w(k) = 1 + \frac{C}{2} \sin\left(\frac{b\pi}{C}\right); b = 1, 2, \dots, C \quad (10)$$

Dalam hal ini, $w(k)$ adalah fungsi *window* terhadap *cepstral features*, C adalah *cepstral coefficients*, simbol k adalah index dari *cepstral coefficients*. Pengolahan *cepstral liftering* mempunyai hasil berupa *frame* dan *cepstral coefficients* selanjutnya diproses ke pemilihan fitur.

3. Model Pemilihan Fitur Normalisasi Bobot Dominan (NBD)

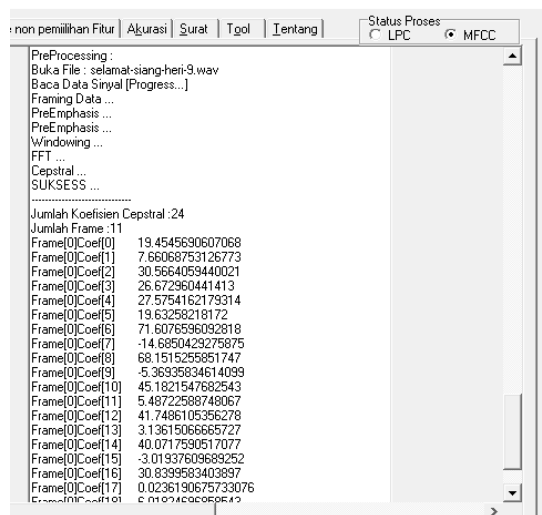
Pemilihan fitur diperlukan dalam hal analisis pengenalan suara agar mendapatkan tabel fitur yang tepat. Pemilihan fitur dan referensi yang tepat diperlukan dalam penelitian ini agar pengecekan dan pengujian kesesuaian ucapan menjadi lebih baik dan meningkat.



Gambar 6. Pemilihan fitur, pengecekan dan pengujian [37]

Gambar 6 menunjukkan tahap pemilihan fitur, pengecekan dan pengujian kesesuaian ucapan Salam. Pemilihan fitur dilakukan untuk mendapatkan tabel fitur mulai dari ekstraksi ciri MFCC, pengecekan ucapan, sampai pengujian. Pengecekan kesesuaian ucapan Salam digunakan untuk pengujian dengan pemilihan referensi dan fitur yang tepat. Pengecekan terhadap referensi yang tepat dengan cara perulangan ucapan lebih dari satu kali, dipilih yang paling tepat dengan kesamaan jangkauan, *filtering*, perhitungan perkalian sekuensial dan KKP.

Pengujian terhadap pemilihan fitur yang tepat dilakukan terhadap jumlah *cepstral coefficient* dan jumlah *frame*, sedangkan pengujian MFCC tanpa NBD untuk membandingkan akurasi ucapan dengan model NBD. Pengujian yang lain juga dilakukan terhadap hukum ucapan untuk mencari pola kesesuaian ucapan dengan hukum ucapan.



Gambar 7. Output frame dan cepstral coefficient

Gambar 7 menunjukkan hasil MFCC *cepstral coefficient* sebanyak dua puluh empat dan sebanyak sebelas *frame*, sedangkan

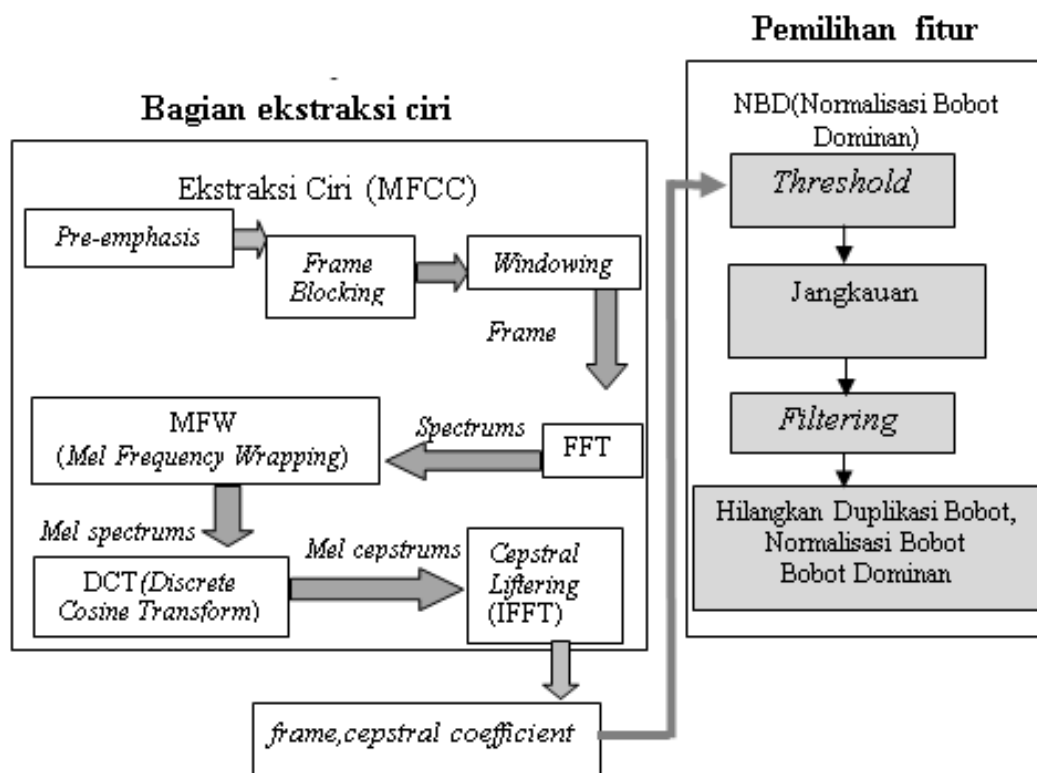
ID	Itrar/bacaan	Surat	PS	pola	HB	P1	P2	P3	P4	P5	P6	P7	pola	PF2	PF3	PF5	PF6	pPF2
128	0 selamatmalam	Selamat	1	REF		1	298	136	306	56	11		REF	298	136	56	11	0.1403674
129	1 selamatmalam	Selamat	1	REF		1	304	162	306	57	8		REF	0	162	57	8	0
131	2 selamatmalam	Selamat	1	REF			305	233	306	75	16		REF	0	233	75	16	0
132	3 selamatmalam	Selamat	1	REF			305	278	306	85	18		REF	0	278	85	0	0
133	4 selamatmalam	Selamat	1	REF			305	276	306	84	17		REF	305	276	84	17	0.1436646
134	5 selamatmalam	Selamat	1	REF			302	150	306	57	9		REF	302	150	57	9	0.14222515
135	6 selamatmalam	Selamat	1	REF			303	150	306	53	5		REF	303	150	53	5	0.1427225
136	7 selamatmalam	Selamat	1	REF			306	278	306	76	13		REF	306	278	76	13	0.1441356
137	8 selamatmalam	Selamat	1	REF			303	150	306	57	8		REF	303	150	57	8	0.1427225
138	9 selamatmalam	Selamat	1	REF			305	192	306	69	14		REF	305	192	69	14	0.1436646
130	10 selamatmalam	Selamat	1	REF		1	306	278	306	85	17		REF	0	278	85	17	0
195	0 selamatpagi	Selamat	2	REF		1	306	158	306	37	3		REF	306	158	37	3	0.1122935
196	1 selamatpagi	Selamat	2	REF		1	299	109	306	18	2		REF	299	109	0	0	0.1097247
198	2 selamatpagi	Selamat	2	REF			305	167	306	52	9		REF	0	0	52	9	0

Gambar 8. Hasil MFCC *frame* dan *cepstral coefficients*

Gambar 8 menunjukkan hasil MFCC ucapan “iqra” terdiri atas sebelas *frame* dan dua puluh empat *cepstral coefficient*.

3.1 Pemilihan Fitur

Pemilihan fitur yang diusulkan mempunyai enam tahap, yaitu menentukan *threshold* yang sama, membuat jangkauan yang sama, melakukan *filtering*, menghilangkan duplikasi bobot, normalisasi bobot dan bobot dominan [37]. Semua tahap tersebut diterapkan dalam algoritma pemilihan fitur Normalisasi Bobot Dominan (NBD) untuk menghasilkan tabel fitur.



Gambar 9. Ekstraksi ciri dan pemilihan fitur NBD [37]

Gambar 9 menunjukkan penelitian ekstraksi ciri suara menggunakan MFCC yang menghasilkan fitur berupa parameter *frame* dan *cepstral coefficient* dan diproses ke pemilihan fitur NBD mulai dari *threshold*, *jangkauan*, *filtering*, menghilangkan duplikasi bobot, normalisasi bobot, hingga bobot dominan.

Tabel 1. Hasil MFCC ucapan selamat malam

kodfile	frame_ke	ucapan	koefisien_ke	nilai_koefisien_cepstral
1214269	0	selamatmalam	0	20.38
1214270	0	selamatmalam	1	-3.08
1214271	0	selamatmalam	2	33.05
1214272	0	selamatmalam	3	15.86
1214273	0	selamatmalam	4	36.08
1214274	0	selamatmalam	5	15.03
1214275	0	selamatmalam	6	66.50
1214276	0	selamatmalam	7	9.77
1214277	0	selamatmalam	8	53.25
1214278	0	selamatmalam	9	28.77
1214279	0	selamatmalam	10	47.97
1214280	0	selamatmalam	11	11.74
1214281	0	selamatmalam	12	42.65
1214282	0	selamatmalam	13	1.83
1214283	0	selamatmalam	14	40.99
1214284	0	selamatmalam	15	-8.42
1214285	0	selamatmalam	16	38.52
1214286	0	selamatmalam	17	8.41
1214287	0	selamatmalam	18	7.38
1214288	0	selamatmalam	19	0.25
1214289	0	selamatmalam	20	8.79
1214290	0	selamatmalam	21	2.98
1214291	0	selamatmalam	22	3.00
1214292	0	selamatmalam	23	-0.12
1214293	1	selamatmalam	0	18.23
1214294	1	selamatmalam	1	0.20
1214295	1	selamatmalam	2	29.91
	...			
	10			

Tabel 1 menunjukkan hasil ekstraksi ciri menggunakan MFCC ucapan “*selamat malam*”. Hasil ekstraksi ciri berupa *frame* yang terdiri atas sebelas *frame* dan dua puluh empat *cepstral coefficient*. Proses selanjutnya hasil *frame* dan *cepstral coefficient* tersebut dilakukan pemilihan fitur dengan pemilihan fitur *frame* dengan algoritma pemilihan fitur menggunakan Normalisasi Bobot Dominan.

Gambar 10 menunjukkan algoritma NBD yang dimulai dari langkah pertama dan kedua pengambilan suara ucapan, langkah ketiga melakukan proses *threshold*. Kemudian, pada langkah keempat dan seterusnya dibuat jangkauan, *filtering*, menghilangkan duplikasi bobot, dilakukan normalisasi bobot dan bobot dominan.

Algoritma Pemilihan Fitur Menggunakan Normalisasi Bobot Dominan (NBD)

0. Mulai
1. Ambil suara ucapan
2. Fungsi MFCC dengan $frame=f$, $cepstral\ coefficient\ (c)=w(k)$
3. Menentukan **threshold 1** sampai **threshold 6** (b_1 - b_6)
 - a. Buat $b_1 = \min(w(k))$,
 - b. buat $b_2 = \frac{\min(w(k)) + \left(\frac{\min(w(k)) + \max(w(k))}{2}\right)}{2}$
 - c. buat $b_3 = rata - rata(w(k))$,
 - d. buat $b_4 = \frac{\min(w(k)) + \max(w(k))}{2}$
 - e. buat $b_5 = \frac{\left(\frac{\min(w(k)) + \max(w(k))}{2}\right) + \max(w(k))}{2}$,
 - f. buat $b_6 = \max(w(k))$.
4. Membuat jangkauan
Jangkauan dengan cek kondisi di setiap $frame$ dengan $cepstral\ coefficient(c)=w(k)$
Bobot=1
 - a. Rule₁ adalah if (b_1) $\min=w(k)$ then p_{i1} =bobot elseif bobot=0
 - b. Rule₂ adalah if (b_1) $\geq w(k)$ and ($w(k) < b_2$) then p_{i2} =bobot elseif bobot=0
 - c. Rule₃ adalah if (b_2) $\geq w(k)$ and ($w(k) < b_3$) then p_{i3} =bobot elseif bobot=0
 - d. Rule₄ adalah if (b_3) $\geq w(k)$ and ($w(k) < b_4$) then p_{i4} =bobot elseif bobot=0
 - e. Rule₅ adalah if (b_4) $\geq w(k)$ and ($w(k) < b_5$) then p_{i5} =bobot elseif bobot=0
 - f. Rule₆ adalah if (b_5) $\geq w(k)$ and ($w(k) < b_6$) then p_{i6} =bobot elseif bobot=0
 - g. Rule₇ adalah if (b_6) $\max= w(k)$ then p_{i7} =bobot elseif bobot=0
5. **Filtering**
Hasil filtering berupa p_{i1} sampai p_{i7}
Hitung jumlah p_{i1} sampai p_{i7} dengan $G_j = \sum_{i=0}^F p_{ij}$
Hitung Jumlah seluruh pola dengan $U = \sum_{j=0}^a G_j$
6. Menghilangkan duplikasi bobot
Tetapkan bahwa $Q_{ij}=p_{ij}$ dan mencari kesamaan menghilangkan duplikasi.
if $Q_{i0}=p_{i1}$ then $Q_{ij}=0$, if $Q_{i1}=p_{i2}$ then $Q_{ij}=0$, if $Q_{i2}=p_{i3}$ then $Q_{ij}=0$
if $Q_{i0}=p_{i2}$ then $Q_{ij}=0$, if $Q_{i1}=p_{i3}$ then $Q_{ij}=0$, if $Q_{i2}=p_{i4}$ then $Q_{ij}=0$
if $Q_{i0}=p_{i10}$ then $Q_{ij}=0$, if $Q_{i1}=p_{i10}$ then $Q_{ij}=0$, if $Q_{i2}=p_{i10}$ then $Q_{ij}=0$
- g. Perhitungan Jumlah dengan $Z_j = \sum_{i=0}^F Q_{ij}$
Normalisasi Bobot
Perhitungan Normalisasi Bobot dengan $S_j = \sum_{i=0}^F \frac{Q_{ij}}{Z_j}$
- h. Bobot dominan
Urutkan nilai terbesar S_j di tabel fitur ($npf_2, npf_3, npf_5, npf_6$) menjadi variabel B_j
- i. Simpan tabel fitur perhitungan jumlah bobot= Z ,
Normalisasi bobot = S_j , keseluruhan pola= U
10. selesai

Gambar 10. Algoritma pemilihan fitur dengan NBD [37]

2.1. Pengumpulan Data

Pengambilan sampling suara dilakukan sebanyak 296 suara reaktan dan diujikan terhadap 296 suara rekaman. Penggunaan cepstral coefficient yang digunakan mulai dari 0 sampai dengan 23 sebanyak 24 cepstral coefficient dan 0 sampai 10 frame. Frekuensi yang digunakan menggunakan 44,100 kHz stereo 16 bit dengan merekam suara laki-laki dan suara perempuan.

Tabel 2. Data rekaman suara Training

NO	Jenis Kelamin	Ucapan	Jumlah Sampling
1	Laki-Laki	Selamat Pagi	37
2	Laki-Laki	Selamat Siang	37
3	Laki-Laki	Selamat Sore	37
4	Laki-Laki	Selamat Malaam	37
1	Perempuan	Selamat Pagi	37
2	Perempuan	Selamat Siang	37
3	Perempuan	Selamat Sore	37
4	Perempuan	Selamat Malaam	37
			296

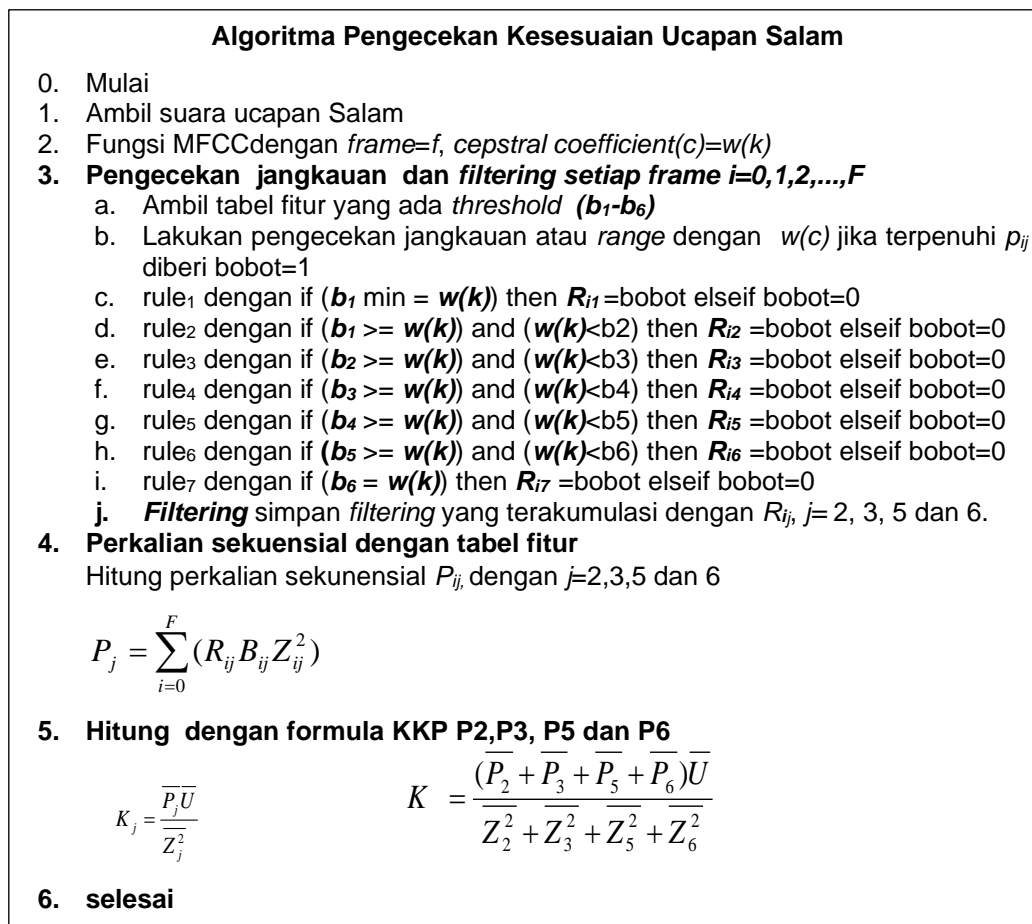
Tabel 3. Data rekaman Uji

NO	Jenis Kelamin	Ucapan	Jumlah Sampling
1	Laki-Laki	Selamat Pagi	37
2	Laki-Laki	Selamat Siang	37
3	Laki-Laki	Selamat Sore	37
4	Laki-Laki	Selamat Malaam	37
1	Perempuan	Selamat Pagi	37
2	Perempuan	Selamat Siang	37
3	Perempuan	Selamat Sore	37
4	Perempuan	Selamat Malaam	37
			296

Tabel 2 dan tabel 3 pengambilan sampling suara untuk masing-masing suara ucapan bacaan 37 sampling baik laki-laki maupun perempuan. Total pengambilan suara sebanyak 296 suara.

3. Hasil dan Pembahasan (Heading 1)

Pengecekan kesesuaian ucapan Salam dilakukan terhadap tabel fitur dengan perhitungan KKP. Perhitungan hasil kesesuaian ucapan tersebut diharapkan mendekati dengan hasil yang ada pada tabel fitur sehingga persentase kesesuaian ucapan menjadi meningkat. Pengecekan terhadap tabel fitur yang ada *threshold* dengan kesamaan jangkauan dan *filtering*, perhitungan sekuensial dan perhitungan KKP menggunakan persamaan seperti dalam algoritma. Pengecekan ucapan tersebut dilakukan terhadap pemilihan referensi yang tepat dan pemilihan fitur yang tepat.



Gambar 11. Algoritma pengecekan kesesuaian ucapan salam

Gambar 11 menunjukkan algoritma pengecekan kesesuaian ucapan dengan tabel 2 hasil fitur dimulai dari suara ucapan salam, kemudian ekstraksi ciri menggunakan MFCC menghasilkan komponen *frame* dan *cepstral coefficient* di langkah kedua. Langkah ketiga, algoritma pengecekan ucapan pembaca dimulai dengan ekstraksi ciri MFCC yang menghasilkan komponen *frame* dan *cepstral coefficient*. Berikut hasil pengujian dapat dilihat pada Tabel 2.

Tabel 4. Hasil Ucapan Selamat

No	UCAPAN	Samplng	MFCC	Frame										
				0	1	2	3	4	5	6	7	8	9	10
1	Selamat Pagi	74	60	10	10	20	40	30	40	30	40	50	60	40
2	Selamat Siang	74	90	30	30	50	80	50	60	50	40	70	90	70
3	Selamat Sore	74	100	50	50	70	60	80	50	60	50	80	100	90
4	Selamat Malam	74	90	50	50	60	50	60	70	60	50	80	90	90
		296	85	35	35	50	58	55	55	50	45	70	85	73

Berdasarkan tabel 2 dengan pemilihan fitur frame yang terbaik pada frame ke-9 dengan nilai rata-rata 85% lebih baik dibandingkan dengan frame yang lain.

4. Kesimpulan dan Saran

Penelitian ini untuk menggunakan MFCC dan pemilihan fitur frame yang tepat untuk ucapakan selamat pagi, selamat siang, selamat sore dan selamat malam yang menghasilkan frame dan cepstral coefficient. Penggunaan cepstral coefficient yang digunakan mulai dari 0 sampai dengan 23 sebanyak 24 cepstral coefficient. Pengambilan sampling suara sebanyak 296 suara rekaman diujikan terhadap 296 suara rekaman. Frekuensi yang digunakan menggunakan 44,100 kHz stereo 16 bit. Hasil akurasi didapat bahwa metode MFCC dengan pemilihan fitur frame yang ke-9 mempunyai tingkat akurasi yang lebih tinggi sebesar 85% dibandingkan dengan frame yang lain

Daftar Pustaka

- [1] 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.
- [2] F. Syafria, A. Buono, and B. I. B. P. Silalahi, "Pengenalan Suara Paru - Paru dengan MFCC sebagai Ekstraksi Ciri dan Backpropagation sebagai Classifier," vol. 3, 2014.
- [3] H. S. Manunggal, "Perancangan dan Pembuatan Perangkat Lunak Pengenalan Suara Pembicara Dengan Menggunakan Analisa MFCC Feature Extraction.," *Tugas Akhir Sarj. pada Jur. Tek. Inform. Fak. Teknol. Ind. Univ. Kristen Petra Surabaya*, 2005.
- [4] T. Chamidy, "Metode Mel Frequency Cepstral Coeffisients (MFCC) Pada klasifikasi Hidden Markov Model (HMM) Untuk Kata Arabic pada Penutur Indonesia," *Matics*, vol. 8, no. 1, pp. 36–39, 2016.
- [5] Irmawan, H. Hikmarika, D. W. Sari, and M. C. Tammimi, "Pengenalan Kata dengan Metode Linear Predictive Coding dan Jaringan Syaraf Tiruan Pada Mobile Robot," no. December, 2014.
- [6] H. S. Thiang, "Sistem Pengenalan Kata dengan Menggunakan Linear Predictive Coding dan Nearest Neighbor Classifier," *Univ. Kristen Petra*, vol. 5, no. September, pp. 19–24, 2005.
- [7] A. M. Aibinu, M. J. E. Salami, A. R. Najeeb, J. F. Azeez, and S. M. A. K. Rajin, "Evaluating the effect of voice activity detection in isolated Yoruba word recognition system," *2011 4th Int. Conf. Mechatronics Integr. Eng. Ind. Soc. Dev. ICOM'11 - Conf. Proc.*, no. May, pp. 17–19, 2011, doi: 10.1109/ICOM.2011.5937134.
- [8] S. Hidayat, R. Hidayat, and T. B. Adji, "Sistem Pengenal Tutur Bahasa Indonesia Berbasis Suku Kata Menggunakan MFCC, Wavelet Dan HMM," *Conf. Inf. Technol. Electr. Eng.*, no. September, pp. 246–251, 2015.
- [9] S. M. Widodo, E. Siswanto, and O. Sudjana, "Penerapan Metode Mel Frequency Ceptral Coefficient dan Learning Vector Quantization untuk Text-Dependent Speaker Identification," vol. 11, no. 1, pp. 15–20, 2016.

- [10] Suyanto and S. Hartati, "Design of Indonesian LVCSR using Combined Phoneme The Approaches of LVCSR," *Icts*, pp. 191–196, 2013, doi: 10.12962/p9772338185001.a33a33.
- [11] S. Suyanto and A. E. Putra, "Automatic Segmentation of Indonesian Speech into Syllables using Fuzzy Smoothed Energy Contour with Local Normalization, Splitting, and Assimilation," *J. ICT Res. Appl.*, vol. 8, no. 2, pp. 97–112, 2014, doi: 10.5614/itbj.ict.res.appl.2014.8.2.2.
- [12] R. Cahyarini, U. L. Yuhana, and A. Munif, "Rancang Bangun Modul Pengenalan Suara Menggunakan Teknologi Kinect," *J. Tek. Pomits*, vol. 2, no. 1, pp. 1–5, 2013.
- [13] M. K. A. S. . dan S. Hertiana Bethaningtyas, "Pengenalan Huruf Hijayyah Berbasis Pengolahan Sinyal Suara dengan Metode MFCC," *Momentum, Vol.13, No. 2 Oktober 2017 ISSN 2406-9329 Hal 49-52 Fak. Tek. WAHID HASYIM SEMARANGe-*, vol. 13, no. 2, pp. 49–52, 2017.
- [14] Heriyanto, "Analisa Deteksi Huruf Hijaiyah Melalui Voice Recognition Menggunakan Kombinasi Energy," *Telematika*, vol. 12, no. 01, pp. 11–22, 2015.
- [15] M. Subali, M. Andriansyah, and C. Sinambela, "ANALISIS FREKUENSI DASAR DAN FREKUENSI FORMANT DARI FONEM HURUH HIJAIYAH UNTUK PENGUCAPAN MAKHRAJ DENGAN METODE DTW," *Pros. PESAT (Psikologi, Ekon. Sastra, Arsit. & Teknik Sipil) Vol. 6, Oktober 2015 ISSN 1858-2559 Univ. Gunadarma - Depok - 20-21 Oktober 2015 Hal S-60-72*, vol. 6, pp. S60-72, 2015.
- [16] C. G. K. Leon, "Robust computer voice recognition using improved MFCC algorithm," *Proc. - 2009 Int. Conf. New Trends Inf. Serv. Sci. NISS 2009*, pp. 835–840, 2009, doi: 10.1109/NISS.2009.12.
- [17] L. Muda, M. Begam, and I. Elamvazuthi, "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques," vol. 2, no. 3, pp. 138–143, 2010, [Online]. Available: <http://arxiv.org/abs/1003.4083>.
- [18] M. W. . Sanjaya and Z. Salleh, "Implementasi Pengenalan Pola Suara Menggunakan Mel-Frequency Cepstrum Coefficients (Mfcc) Dan Adaptive Neuro-Fuzzy Inferense System (Anfis) Sebagai Kontrol Lampu Otomatis," *Al-HAZEN J. Phys.*, vol. 1, no. 1, pp. 1–19, 2014.
- [19] Y. Miftahuddin and M. R. Hakim, "COEFFICIENT DAN DYNAMIC TIME WARPING UNTUK PENGENALAN NADA PADA ALAT MUSIK BELLRYA," pp. 120–127, 2017.
- [20] A. R. Darma Putra, "Verifikasi Biometrika Suara Menggunakan Metode MFCC dan DTW," *LONTAR Komput. Biometrika, Univ. Udayana ISSN2088-1541*, vol. 2, no. 1, pp. 8–21, 2011.
- [21] D. Novianto and R. V. Yuliantari, "Pengenalan Isyarat Tutar Vokal Bahasa Indonesia Menggunakan Metode Dynamic Time Wrapping (Dtw) Berbasis Fungsi Jarak," no. 1, pp. 1–4, 2017.
- [22] S. Martyna and S. Sudaryanto, "Penerapan Metode Particle Swarm Optimization pada Artificial Neural Network Backpropagation untuk Peramalan Penjualan Furniture pada CV. Octo Agung," pp. 1–9, 2011.

- [23] A. M. Aibinu, M. J. E. Salami, A. R. Najeeb, J. F. Azeez, and S. M. A. K. Rajin, "Evaluating the effect of voice activity detection in isolated Yoruba word recognition system," *2011 4th Int. Conf. Mechatronics Integr. Eng. Ind. Soc. Dev. ICOM'11 - Conf. Proc.*, no. May, pp. 17–19, 2011, doi: 10.1109/ICOM.2011.5937134.
- [24] S. B. Davis and P. Mermelstein, "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," *IEEE Trans. Acoust.*, vol. 28, no. 4, pp. 357–366, 1980, doi: 10.1109/TASSP.1980.1163420.
- [25] D. J. . Chitode, "01.Communication Theory." 2010.
- [26] R. T. M. (Sim) narasimha Tokunbo Ogunfunmi, *Speech and Audio Processing and Recognition*, no. part 1. springer, 2015.
- [27] J. H. and W. Holmes, *Speech Synthesis and Recognition, Second Edition*. 2003.
- [28] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, algorithms, and applications*. 1996.
- [29] A. E. Putra, "Frekuensi Cuplik pada FFT," *Tan Li, Process. Digit. Signal*, vol. 1, 2008.
- [30] A. A. Kumar, *Digital signal processing*, vol. 23, no. 4. Prentice-Hall of India Pvt.Ltd, 2013.
- [31] D. Laha, *Handbook of Computational Intelligence in Manufacturing and Production Manajemen*. 2007.
- [32] Tshilidzi Marwala, *Condition Monitoring Using Computational Intelligence Methods*. 2012.
- [33] T. Hassan, A. Wassim, and M. Bassem, "Analysis and Implementation of an Automated Delimiter of " Quranic " Verses in Audio Files using Speech Recognition Techniques," *Robust Speech Recognit. Underst.*, no. June, pp. 352–362, 2007, doi: 10.5772/4759.
- [34] S. W. Smith, *Digital signal processing*, vol. 17, no. 2. 2000.
- [35] K. R. R. Vladimir Britanak, Patrick C.Yip, *Discrete Cosine and Sine Transform*. 2007.
- [36] L. R. Rabiner and R. W. Schafer, *Introduction to digital speech processing*, vol. 1, no. 1. 2007.
- [37] H.- Heriyanto, S. Hartati, and A. E. Putra, "Evaluation of Suitability of Voice Reading of Al-Qur'an Verses Based on Tajwid Using Mel Frequency Cepstral Coefficients (MFCC) and Normalization of Dominant Weight (NDW)," *Adv. Image Video Process.*, vol. 6, no. 2, 2018, doi: 10.14738/aivp.62.4268.