

DETEKSI SUARA UCAPAN SALAM BAHASA ARAB MENGUNAKAN MEL FREQUENCY CEPSTRAL COEFFICIENT (MFCC) DAN PEMILIHAN FITUR MIN MAX

Heriyanto ^{(1)*}

⁽¹⁾Jurusan Teknik Informatika Program Studi Informatika
UPN "Veteran" Yogyakarta
e-mail : heriyanto@upnyk.ac.id⁽¹⁾

Abstract

Arabic greeting sounds are used in everyday life for Muslims in Indonesia. Salam recognition is used to check how correct the pronunciation of Arabic greetings is for Indonesians. The first stage was to collect the sample of greeting readings as much as 50 records of male and female records in wav recordings. One person takes the source of greeting reading as a reference for reference. Retrieval of test data as much as 50 samples of test data. The second stage is to perform feature extraction with MFCC from cepstral coefficient and frame results. The third stage is testing by checking the suitability of the greeting reading with the calculation of min max. The result of checking the suitability of reading on the selection of the right features carried out by MFCC has a result of 60.25%. Meanwhile, MFCC with a minimum yield of 71.75.0%. This shows that the use of the min max test can improve accuracy because there are more unique cepstral and max and min coefficients with a significant difference of 11.5%.

Keywords : *checking, feature extraction, reference, features, speech*

Suara ucapan salam bahasa Arab digunakan dalam kehidupan sehari-hari bagi umat beragama Islam di Indonesia. Pengenal ucapan salam dilakukan untuk mengecek seberapa benar dalam pelafalan ucapan salam berbahasa Arab bagi orang Indonesia. Tahap pertama dilakukan pengambilan sampel bacaan salam sebanyak 50 data rakaman putra dan putri dalam rekaman wav. Pengambilan sumber bacaan salam diambil satu orang sebagai acuan untuk referensi. Pengambilan data uji sebanyak 50 sampel data uji. Tahap kedua adalah melakukan ekstraksi ciri dengan MFCC hasil cepstral coefficient dan frame. Tahap ketiga adalah pengujian dengan pengecekan kesesuaian bacaan salam dengan perhitungan min max. Hasil pengujian pengecekan kesesuaian bacaan terhadap pemilihan fitur yang tepat dilakukan dengan MFCC mempunyai hasil sebesar 60,25%. Sedangkan MFCC dengan min max hasil sebesar 71.75,0%. Hal tersebut menunjukkan bahwa penggunaan pengujian min max dapat meningkatkan akurasi karena terdapat cepstral dan coefficients max dan min lebih unik dengan selisih 11.5% cukup signifikan

Kata Kunci : pengecekan, ekstraksi ciri, referensi, fitur, ucapan

1. PENDAHULUAN

Suara ucapan bacaan salam dilakukan dengan berbahasa Arab. Pelafalan berbeda dengan bacaan bahasa Indonesia sehari-hari. Penelitian pengembangan suara (Heriyanto et al., 2018) menggunakan normalisasi bobot dominan untuk pengecekan suara hukum bacaan Alquran mempunyai akurasi 91.37% pada 9 frames dengan dominan yang ke-9.

Penelitian suara pengecekan angka satu sampai sepuluh (Heriyanto, 2019) hasil pengujian terhadap MFCC dan pemilihan fitur dengan normalisasi bobot dominan sebesar 70% sedangkan tanpa pemilihan fitur hanya sebesar 42%.

Penelitian ekstraksi ciri MFCC berupa fitur *frame* dan membuat rata-rata *cepstral coefficient* di setiap *frame* fitur langsung pengecekan dengan pembaca dengan referensi. Keakurasian sejauh ini menggunakan ekstraksi ciri MFCC yang mempunyai tingkat keakurasian sebesar 58% sampai dengan 75% (Aibinu, dkk., 2011).

Penelitian suara dengan huruf hijaiyah menggunakan kombinasi energy (Yanto, 2015) menghasilkan akurasi 45%. Model pengembangan suara ucapan warna dengan pemilihan bobot dominan oleh (Heriyanto & Simanjuntak, 2017) dengan Linier Predictive Code (LPC) dengan akurasi 95%.

2. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini dengan MFCC. 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* (Davis & Mermelstein, 1980). MFCC merupakan ekstraksi ciri yang paling banyak digunakan dalam bidang *speaker recognition* dan *speech recognition*.

MFCC merupakan ekstraksi ciri yang menghasilkan fitur berupa parameter *cepstral coefficient* (Abriyono & Harjoko, 2012). Ekstraksi ciri *Mel Frequency Cepstral Coefficient* (MFCC) mengubah gelombang suara menjadi beberapa tipe parameter seperti *cepstral coefficient* yang mempresentasikan *audio file* (Chamidy, 2016). 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* (Putra, 2011).

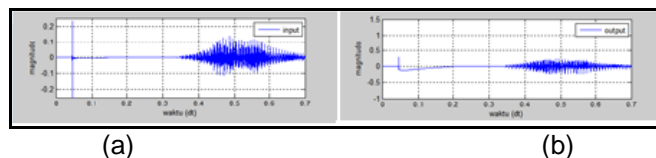
2.1. TAHAPAN EKSTRAKSI CIRI MFCC

Pre-emphasis

Pre-emphasis merupakan tahap awal dalam proses MFCC menurut Chitode (2010) melakukan penekanan pada frekuensi tinggi, menyelaraskan antara frekuensi tinggi dengan rendah. *Pre-emphasis* menurut Tokunbo Ogunfunmi (2015) bertujuan untuk mengurangi kebisingan dan memperbaiki *Signal to Noise Rasio* (SNR) serta mengurangi suara yang tidak diinginkan atau permasalahan *noise*. *Pre-emphasis* merupakan pemrosesan signal sederhana mempunyai tujuan agar level *base band* pada bagian frekuensi tinggi masih memiliki kualitas sinyal yang baik. Proses *pre-emphasis* menurut Proakis dan Manolakis (1996) 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, $y(n)$ merupakan signal hasil *pre-emphasis*, sedangkan $s(n)$ merupakan signal yang dilakukan 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 3.1 bagian (b) adalah hasil *output* pengolahan signal setelah proses *pre-emphasis*.



Gambar 1. *Pre-emphasis* (a) sebelum dan (b) sesudah (Putra, 2011)

Proses pengambilan signal ke- n pada *pre-emphasis* dengan bacaan dengan waktu satu sampai tiga detik perkata.

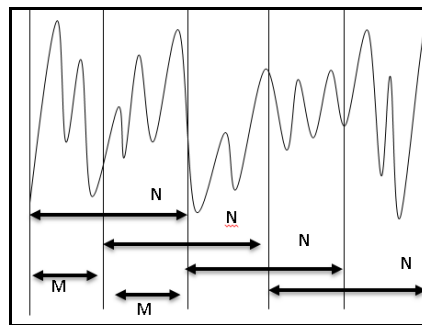
Frame blocking

Frame blocking dilakukan setelah melalui *pre-emphasis*. 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* (Abriyono & Harjoko, 2012). Lebar *frame* dinotasikan dengan N , sedangkan lebar pergeseran setiap *frame* sebagai M . Lebar *overlap* dihitung selisih $N-M$.

Frame blocking prinsipnya menganalisis sinyal ke dalam bentuk *frame* menurut Holmes (Holmes, 2003). Setiap *frame* diwakili oleh fitur vektor tunggal digambarkan dalam spektrum rata-rata untuk interval waktu dalam *frame* waktu antara 20-40 milidetik menurut Chamidy (Chamidy, 2016). *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* (Abriyono & Harjoko, 2012)

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

Windowing

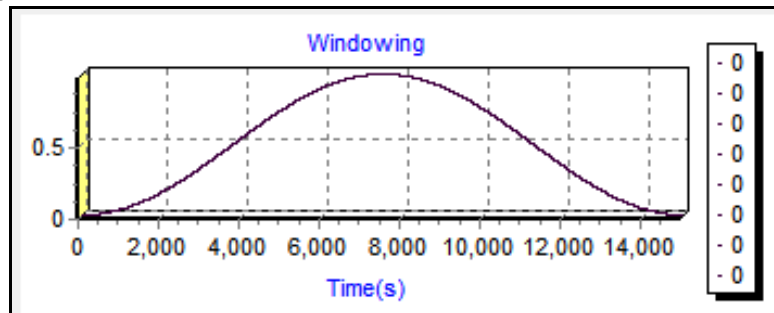
Windowing mempunyai efek dan fungsi untuk menghaluskan spectrum menurut Proakis dan Manolakis (Proakis & Manolakis, 1996). *Windowing* mempunyai tujuan untuk mengurangi efek *diskontinue* pada ujung-ujung *frame* yang dihasilkan oleh proses *frame blocking*. *Windowing* yang digunakan yaitu *Rectangular Window*, *Hamming Window* dan *Hanning Window* (Chamidy, 2016). Dari ketiga fungsi *windowing* tersebut peneliti menggunakan *windowing Hanning* dikarenakan lebih halus dibandingkan dengan yang lain (Putra, 2008). 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 menjelaskan tentang hasil proses *windowing* menggunakan *Hanning window*.



Gambar 3. Ilustrasi *windowing* (Proakis & Manolakis, 1996)

Fast Fourier Transform (FFT)

Fast Fourier Transform merupakan pengembangan dari algoritma *Discrete Fourier Transform* (DFT). FFT merupakan pengubah sinyal digital dari pada ranah waktu ke ranah frekuensi (Abriyono & Harjoko, 2012). 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 sinyal dari ranah waktu menjadi ranah frekuensi.

Transformasi *fourier* menurut Proakis dan Manolakis (Proakis & Manolakis, 1996) salah satu dari beberapa alat matematika yang berguna dalam analisis dan desain sistem *Linier Time Invariant* (LTI) dan deret *fourier*.

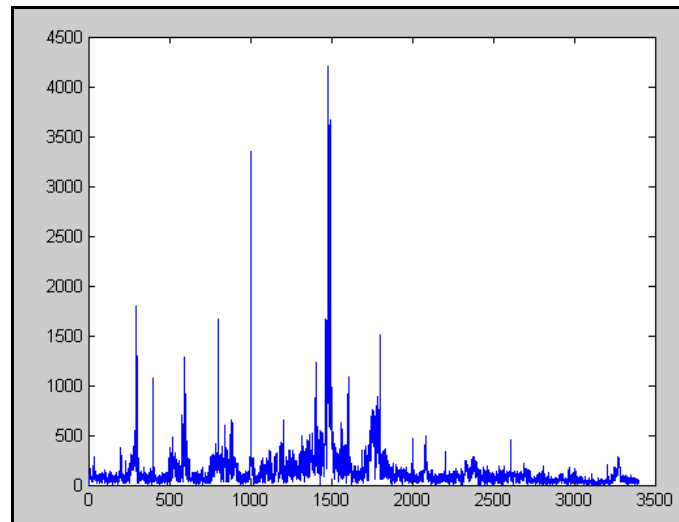
FFT adalah salah satu metode algoritma cepat untuk dapat mengimplementasikan *Discrete Fourier Transform* (DFT). DFT menurut Proakis dan Manolakis (Proakis & Manolakis, 1996) alat komputasi yang memainkan peran yang sangat penting dalam banyak aplikasi pemrosesan sinyal digital, seperti analisis frekuensi, 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 (Proakis & Manolakis, 1996) bahwa metode FFT digunakan sebagai cara yang efisien untuk dapat menghitung DFT. *Discrete Fourier Transform* (DFT) menggunakan persamaan (5).

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

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 ($k=N/2, k \in \mathbb{N}$). *Fast Fourier Transform* mempunyai tujuan mendekomposisi sinyal menjadi sinyal sinusoidal yang berupa unit *real* dan *unit imajiner*. *Fast Fourier Transform* menggunakan persamaan (6)

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

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



Gambar 4. FFT menghasilkan spektrum (Kumar, 2013)

Mel Frequency Wrapping (MFW)

Mel Frequency Wrapping (MFW) merupakan *filter* berupa *filter bank* untuk mengetahui ukuran energi dari *frequency band* tertentu dalam signal suara (Miftahuddin dan Hakim, 2017; Putra, 2011). MFW menurut Laha (2007) mengubah frekuensi ke dalam mel.

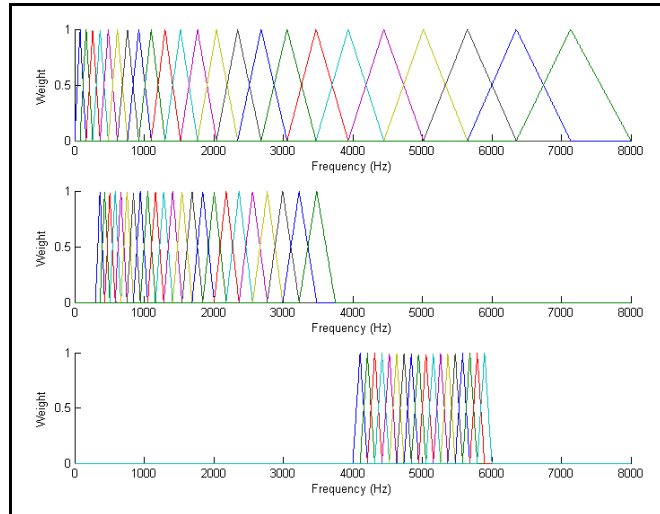
Sedangkan MFW menurut Tshilidzi Marwala (2012) berisi *filterbank*, yang diberi jarak pada skala mel. *Filterbank* memiliki respon frekuensi lewat jalur segitiga yang jarak dan besarnya ditentukan oleh interval frekuensi yang konstan. Proses *output* yang diperoleh dari *filter* dikenal dengan spektrum 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 (Hidayat dkk., 2015; Hassan dkk., 2007)

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

Discrete Cosine Transform (DCT)

DCT menurut Smith (2000) 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 (Putra, 2011). *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).

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

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

Cepstral Liftering

Cepstral coefficient menurut Proakis dan Manolakis (Proakis & Manolakis, 1996) merupakan *the fourier series coefficients*. Seprtrum didefinisikan oleh Rabiner dan Schafer (2007) adalah kebalikannya transformasi *fourier*. *Cepstral liftering* meningkatkan akurasi digunakan untuk *pattern matching*, baik *speaker recognition* maupun *speech recognition* (Putra, 2011). *Cepstral coefficient* menggunakan persamaan (10).

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

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

3. HASIL DAN PEMBAHASAN

Hasil Cepstral Liftering terlihat pada Gambar 6 dan Gambar 7. Hasil MFCC terlihat pada Gambar 6 pada ucapan assalam dengan frame 0-10 dan koefisien 0-23 sebanyak 24 koefisien.

The screenshot shows a web interface with a search bar containing 'assalam' and a 'Refresh' button. Below is a table with the following data:

ID	frame	bacaan	Surat	koefisien_ke	cepstral_coefficient
3471	0	assalam	Salamislam	0-23	40.1175008224153
3472	1	assalam	Salamislam	0-23	25.2541404760091
3473	10	assalam	Salamislam	0-23	40.0694784379923
3474	2	assalam	Salamislam	0-23	36.9887112006125
3475	3	assalam	Salamislam	0-23	22.3884556026527
3476	4	assalam	Salamislam	0-23	26.7398154440619
3477	5	assalam	Salamislam	0-23	33.797637901831
3478	6	assalam	Salamislam	0-23	38.3869195891244
3479	7	assalam	Salamislam	0-23	32.1754308590963
3480	8	assalam	Salamislam	0-23	36.9940724256858
3481	9	assalam	Salamislam	0-23	32.5126502952135

Gambar 6. Hasil Cepstral coefficients ucapan assalam

Hasil Cepstral Liftering terlihat pada Gambar 7. Hasil MFCC terlihat pada gambar 3.7 pada ucapan mualaikum dengan frame 0-10 dan koefisien 0-23 sebanyak 24 koefisien. Pemilihan fitur dilakukan dengan mencari nilai min dan nilai max hasil tersebut akan di bagi dengan dua dan hasil tersebut dilakukan penyimpanan terhadap referensi.

The screenshot shows a web interface with a search bar containing 'mualaikum' and a 'Refresh' button. Below is a table with the following data:

ID	frame	bacaan	Surat	koefisien_ke	cepstral_coefficient
3482	0	mualaikum	Salamislam	0-23	56.2982573909314
3483	1	mualaikum	Salamislam	0-23	47.4601797327836
3484	10	mualaikum	Salamislam	0-23	37.8818110235881
3485	2	mualaikum	Salamislam	0-23	49.286875986744
3486	3	mualaikum	Salamislam	0-23	51.513822961461
3487	4	mualaikum	Salamislam	0-23	61.0996603955581
3488	5	mualaikum	Salamislam	0-23	53.1728707829612
3489	6	mualaikum	Salamislam	0-23	45.591471553977
3490	7	mualaikum	Salamislam	0-23	47.5417955092898
3491	8	mualaikum	Salamislam	0-23	56.3588510613234
3492	9	mualaikum	Salamislam	0-23	42.5929951658549

Gambar 3.6 Hasil cepstral coefficients ucapan mualaikum

Setelah diketahui hasil cepstral koefisien dan juga frame maka selanjutnya dilakukan pengujian terhadap pemilihan fitur dengan min max dan di bagi dua.

Pemilihan fitur Min Max

Pengecekan dengan pemilihan fitur dilakukan untuk menghitung pendekatan antara referensi yang dipilihah fitur dengan min max dengan pengujian pada persamaan (11).

$$C = \frac{(\min + \max)}{2} \quad (11)$$

Dalam hal ini C adalah hasil pemilihan fitur, sedangkan min adalah nilai minimum dan max adalah nilai maksimum.

Pengujian Pemilihan Fitur Min Max

Pengecekan bacaan salam dilakukan terhadap tabel referensi dengan perhitungan min dan max bagi dua. Perhitungan hasil kesesuaian bacaan salam tersebut diharapkan mendekati dengan hasil yang ada pada tabel referensi sehingga persentase kesesuaian ucapan salam menjadi meningkat. Tahap ini, dilakukan pengambilan sampel bacaan salam sebanyak 50 data rekaman putra dan putri dalam rekaman wav. Pengambilan sumber bacaan salam diambil satu orang sebagai acuan untuk referensi. Pengambilan data uji sebanyak 50 sampel data uji.

Kesamaan fitur min dan max dipilih berdasarkan kedekatan dan mendekati nilai rata-rata. Jangkauan min dan max dilakukan pengecekan terhadap tabel referensi berupa sekuensial satu per satu setiap hasil min dan max ucapan pada *frame* dan *coeffisient*. Hasil perhitungan tersebut dilakukan pengecekan terhadap tabel referensi yang ada.

Tabel 1. Hasil ucapan salam MFCC dan MFF Fitur Min dan Max

No	Ucapan	MFCC	MFCC fitur Min dan Max
1	assalam	50	76
2	mualaikum	60	73
3	warohmatullahi	63	68
4	wabarokatuh	68	70
	Rata-rata	60,25 %	71,75 %

Tabel 1 menunjukkan hasil pengecekan kesesuaian ucapan salam dengan MFCC dengan tabel referensi dimulai dari suara bacaan "assalam" sampai "wabarokatuh", kemudian ekstraksi ciri menggunakan MFCC menghasilkan komponen *frame* dan *cepstral coefficient* dengan akurasi rata-rata sebesar 60,25%. Pada MFCC pemilihan fitur pengecekan ucapan pembaca dimulai dengan ekstraksi ciri MFCC yang dilakukan pemilihan fitur mempunyai hasil yang berbeda lebih meningkat menghasilkan komponen *frame* dan *cepstral coefficient* dan akurasi sebesar 71,75%. Hasil tersebut terlihat mempunyai selisih yang signifikan sebesar 11,5%.

Hasil MFCC dan Pemilihan Fitur

Penelitian ini juga dilakukan dengan pengujian membandingkan MFCC dan MFCC pemilihan fitur min dan max. Perbandingan tersebut dilakukan untuk mengetahui perbedaan dampak menggunakan model MFCC dengan MFCC pemilihan fitur ternyata mempunyai selisih sebesar 11,5%. Hasil selisih rata-rata MFCC dan MFCC pemilihan fitur min dan max berdampak pada kesesuaian ucapan salam dan akurasinya dikarenakan terdapat banyak kesalahan fitur *frame* dan *cepstral coefficient* yang belum unik, sehingga berpengaruh pada ketepatan dengan fitur yang ada pada suara pembaca.

KESIMPULAN

Pengujian MFCC dan MFCC dengan pemilihan fitur mempunyai hasil rata-rata sebesar 60,25% lebih rendah dibandingkan dengan MFCC pemilihan fitur min dan max sebesar 71,75%. Penelitian persentase MFCC dengan pemilihan fitur min max mempunyai selisih perbedaan sebesar 11,5%. Hal tersebut menunjukkan keuntungan bahwa penggunaan pemilihan fitur min dan max dapat meningkatkan akurasi dengan *frame* dan *cepstral coefficient* yang lebih unik kalau dilakukan pemilihan fitur Min dan Max. Untuk penelitian lebih lanjut, disarankan untuk melakukan pengambilan referensi dilakukan dengan satu pakar untuk memilih referensi yang tepat dan juga mengatasi kelemahan penggunaan model ini yaitu tabel referensi yang digunakan untuk referensi dengan pakar yang lain perlu mengambil referensi ulang.

DAFTAR PUSTAKA

- Abriyono & Harjoko, A. (2012). Pengenalan Ucapan Suku Kata Bahasa Lisan Menggunakan Ciri LPC, MFCC, dan JST. *Indonesian Journal of Computing and Cybernetics Systems*, 6(2), 23–34.
- Aibinu, A. M., Salami, M. J. E., Najeeb, A. R., Azeez, J. F., & Rajin, S. M. A. K. (2011). Evaluating the effect of voice activity detection in isolated Yoruba word recognition system. *2011 4th International Conference on Mechatronics: Integrated Engineering for Industrial and Societal Development, ICOM'11 - Conference Proceedings, May, 17–19*. <https://doi.org/10.1109/ICOM.2011.5937134>
- Chamidy, T. (2016). Metode Mel Frequency Cepstral Coeffisients (MFCC) Pada klasifikasi Hidden Markov Model (HMM) Untuk Kata Arabic pada Penutur Indonesia. *Matics*, 8(1), 36–39.
- Chitode, D. J. . (2010). *01.Communication Theory*.
- Darma Putra, A. R. (2011). Verifikasi Biometrika Suara Menggunakan Metode MFCC dan DTW. *LONTAR KOMPUTER Biometrika, Universitas Udayana ISSN:2088-1541*, 2(1), 8–21.
- Davis, S. B., & Mermelstein, P. (1980). Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(4), 357–366. <https://doi.org/10.1109/TASSP.1980.1163420>
- Hassan, T., Wassim, A., & Bassem, M. (2007). Analysis and Implementation of an Automated Delimiter of " Quranic " Verses in Audio Files using Speech Recognition Techniques. *Robust Speech Recognition and Understanding*, June, 352–362. <https://doi.org/10.5772/4759>
- Heriyanto, H.-, Hartati, S., & Putra, A. E. (2018). 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). *Advances in Image and Video Processing*, 6(2). <https://doi.org/10.14738/aivp.62.4268>
- Heriyanto, H. (2019). Deteksi Ucapan Angka Satu Sampai Sepuluh Bahasa Palembang Menggunakan Mfcc Dan Bobot Dominan. *Telematika*, 16(1), 52. <https://doi.org/10.31315/telematika.v16i1.3024>
- Heriyanto, H., & Simanjuntak, O. S. (2017). Identifikasi Ucapan Warna Menggunakan LPC (Linier Predictive Code) Dan Kelompok Pemilihan Bobot. *Telematika*, 14(01), 68–73. <https://doi.org/10.31315/telematika.v14i01.1968>
- Hidayat, S., Hidayat, R., & Adji, T. B. (2015). Sistem Pengenal Tutur Bahasa Indonesia Berbasis Suku Kata Menggunakan MFCC, Wavelet Dan HMM. *Conference on Information Technology and Electrical Engineering (CITEE)*, September, 246–251.
- Holmes, J. H. and W. (2003). *Speech Synthesis and Recognition, Second Edition*. <https://doi.org/10.1145/1185448.1185459>
- Kumar, A. A. (2013). Digital signal processing. In *Published by Asoke K. Ghosh, PHI Learning Private Limited, M-97, Connaught Circus, New Delhi-110001 and Printed by Rajkamal Electric Press, Plot No. 2, Phase IV, HSIDC, Kundli-131028, Sonapat, Haryana (Vol. 23, Issue 4)*. Prentice-Hall of India Pvt.Ltd. <https://doi.org/10.1109/TASSP.1975.1162707>
- Laha, D. (2007). *Handbook of Computational Intelligence in Manufacturing and Production Manajemen*.
- Miftahuddin, Y., & Hakim, M. R. (2017). *COEFFICIENT DAN DYNAMIC TIME WARPING UNTUK PENGENALAN NADA PADA ALAT MUSIK BELLRYA*. 120–127.
- Proakis, J. G., & Manolakis, D. G. (1996). Digital Signal Processing: Principles, algorithms, and applications. In *Digital Signal Processing: Principles, algorithms, and applications*.
- Putra, A. E. (2008). Frekuensi Cuplik pada FFT. *Tan Li, Processing, Digital Signal*, 1.
- Rabiner, L. R., & Schafer, R. W. (2007). Introduction to digital speech processing. In *Foundations and Trends in Signal Processing (Vol. 1, Issue 1)*. <https://doi.org/10.1561/2000000001>
- Smith, S. W. (2000). Digital signal processing. In *The Scientist and Engineer's Guide to Digital Signal Processing (Vol. 17, Issue 2)*. <https://doi.org/10.1109/79.826412>
- Tokunbo Ogunfunmi, R. T. M. (Sim) narasimha. (2015). *Speech and Audio Processing and Recognition (Issue part 1)*. springer.
- Tshilidzi Marwala. (2012). *Condition Monitoring Using Computational Intelligence Methods*. <https://doi.org/10.1007/978-1-4471-2380-4>

- Vladimir Britanak, Patrick C.Yip, K. R. R. (2007). *Discrete Cosine and Sine Transform*.
- Yanto, H. Y. (2015). Analisa Deteksi Huruf Hijaiyah Melalui Voice Recognition Menggunakan Kombinasi Energy. *Telematika*, 12(1), 11–22.
<https://doi.org/10.31315/telematika.v12i1.523>
- Ian Ucapan Suku Kata Bahasa Lisan Menggunakan Ciri LPC, MFCC, dan JST. *Indonesian Journal of Computing and Cybernetics Systems*, 6(2), hal.23–34.