

Perintah Menggunakan Sinyal Suara dengan Mel-Frequency Cepstral Coefficients (MFCC) dan K-Nearest Neighbor (KNN)

Gusti Putu Aryo Widano¹, Anak Agung Istri Ngurah Eka Karyawati²

^aInformatika, Universitas Udayana
Bali, Indonesia

¹aryawidana5013@gmail.com

²eka.karyawati@unud.ac.id

Abstract

Current technological developments allow every action that previously required more energy to complete can be more practical, input to give commands can not only be done with the mouse, keyboard, or touchscreen technology but can also be done with the help of voice. However, before it can be used as a command, the data will be retrieved using a computer first. In this study, two stages will be carried out, namely the voice feature extraction process using Mel-Frequency Cepstral Coefficients (MFCC) and the classification process using the K-Nearest Neighbor (KNN) method. Previously there have been similar studies that have been carried out, where the difference with this research lies in the classification method used previously, whereas previous research used Learning Vector Quantization (LVQ).

Keywords: Voice, Command, Mel-Frequency Cepstral Coefficients (MFCC), K-Nearest Neighbor (KNN), Learning Vector Quantization (LVQ)

1. Pendahuluan

Perkembangan teknologi di zaman ini memungkinkan setiap tindakan yang semulanya membutuhkan tenaga lebih untuk menyelesaikannya akan menjadi lebih praktis, salah satunya ketika sebuah masukan untuk memberikan sebuah perintah tidak hanya dapat dilakukan dengan *mouse*, *keyboard* ataupun teknologi layer sentuh, tetapi juga dapat dilakukan dengan suara dengan bantuan pemrosesan-pemrosesan data oleh komputer. Banyak teknologi yang memanfaatkan *input* suara untuk perintah, sebagai contoh *google assistant*, *Alexa*, *Bixby* dan masih banyak lagi teknologi yang memanfaatkan input suara sebagai perintah.

Terdapat beberapa penelitian yang mengangkat topik terkait suara atau *voice command* diantaranya perintah suara berdasarkan gender [3], pengenalan suara artikulasi p [1], perancangan *smart -tv* menggunakan perintah suara [2] sistem buka tutup pintu otomatis berbasis suara manusia [9].

Sebelum dapat digunakan sebagai sebuah perintah, suara terlebih dahulu melalui proses ekstraksi fitur suara dan pembelajaran mesin. Penelitian terdahulu menggunakan 180 data yang diperoleh dari enam naracoba, tiga diantaranya pria dan tiga lainnya adalah perempuan. Hasil akurasi tertinggi yang diperoleh pada penelitian sebelumnya adalah α 0.1 dan penurunan learning rate 0.1, dengan metode klasifikasi yang digunakan adalah *Learning Vector Quantization* (LVQ).

Pada penelitian ini, akan dilakukan pengujian akurasi terhadap data audio yang didapatkan dari enam naracoba, dengan data yang diperoleh sebanyak 120 data audio dengan format *.wav*. Dari data tersebut kata kunci yang digunakan adalah "buka" dan "tutup", masing-masing kata akan dilakukan 10 kali rekaman per orang.

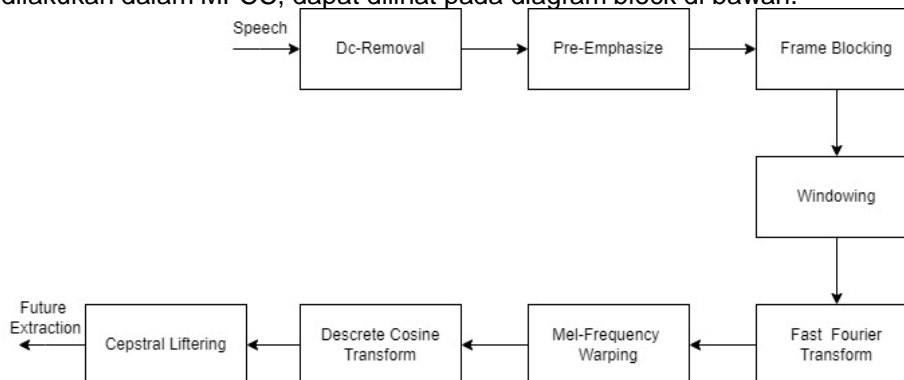
2. Metode Penelitian

2.1. Akusisi Data

Pengambilan data dilakukan secara online dengan menggunakan 2 kata kunci yaitu “buka” dan “tutup”, masing-masing naracoba akan melakukan 10 kali rekaman tiap 1 kata, rekaman melalui fitur *voice note* lewat *whatsapp*, dan artikulasi harus diucapkan dengan jelas sehingga mendapatkan data latih yang baik, jumlah data yang diperoleh sejumlah 120 data. Rata-rata durasi rekaman adalah selama 1 detik, dan format data yang digunakan adalah *.wav*.

2.2. Mel-Frequency Cepstral Coefficients

MFCC merupakan metode yang digunakan untuk mengekstraksi fitur audio yang banyak digunakan dalam bidang *speech technology*. Terdapat beberapa tahapan yang dilakukan dalam MFCC, dapat dilihat pada diagram block di bawah.



Gambar 1. Ekstraksi menggunakan metode MFCC

2.2.1. Dc Removal

Proses ini dilakukan untuk mendapatkan nilai rata-rata serta untuk mengurangi nilai dari setiap sampel dengan nilai rata-rata yang didapat, tujuannya adalah untuk mengeliminasi data yang tidak digunakan. Berikut adalah persamaan dari proses *DC Removal*.

$$y[n] = x[n] - \bar{x}, 0 \leq n \leq N - 1 \quad (1)$$

Dimana:

$y[n]$ = *Sample signal* hasil *DC Removal*

$x[n]$ = *Sample signal* asli

\bar{x} = Nilai rata-rata *sample signal*

N = Panjang *signal*

2.2.2. Pre Emphasize

Proses ini dilakukan untuk meminimalisir *noise inputan* suara, dengan tujuan tingkat akurasi dari proses ekstraksi fitur dapat ditingkatkan. Berikut persamaan dari proses *pre emphasize*.

$$y[n] = s[n] - a \cdot s[n - 1], 0.9 \leq a \leq 1.0 \quad (2)$$

Dimana :

$y[n]$ = *Signal* hasil *pre-emphasize filter*

$s[n]$ = *Signal* sebelum *pre-emphasize filter*

a = nilai *alpha*

2.2.3. Frame Blocking

Proses ini digunakan untuk proses pembagian data menjadi beberapa *frame* atau *slot*. Berikut persamaan dari proses *frame blocking*.

$$\text{jumlah frame} = Ts/M \quad (3)$$

Dimana:

T_s = Durasi pengambilan suara (ms)
 M = Panjang *Frame* (ms)

2.2.4. Windowing

Proses ini digunakan untuk meminimalisir *discontinuitas* sinyal pada awal dan akhir setelah dilakukan proses *frame blocking*. Berikut persamaan dari proses *windowing*.

$$x(n) = x_i(n)w(n) \quad (4)$$

Dimana:

$x(n)$ = nilai *sample signal* hasil *windowing*

$x_i(n)$ = nilai *sample signal* dari *frame signal* ke- i

$w(n)$ = fungsi *window*

2.2.5. Fast Fourier Transform (FFT)

Proses ini digunakan untuk melakukan perubahan terhadap domain waktu menjadi domain frekuensi. Berikut persamaan dari proses FFT.

$$f(n) = \sum_{k=0}^{N-1} y_k e^{-2\pi jkn/N}, n = 0, 1, 2, \dots, N-1 \quad (5)$$

Dimana:

$f(n)$ = Frekuensi

N = Jumlah *sample* pada masing-masing *frame*

$k = 0, 1, 2, \dots, (N-1)$

j = Bilangan Imajiner ($\sqrt{-1}$)

2.2.6. Mel Frequency Warping

Proses ini dilakukan menggunakan *filterbank* yang dapat dilihat pada persamaan berikut.

$$Mel f = \frac{2595 * \log_{10}(1 + \frac{f}{700})}{\frac{Si}{2}} \quad (6)$$

Dimana:

Si = Sinyal awal hasil FFT

$f = f_0 - f_n$

2.2.7. DCT

Proses yang digunakan untuk menghitung *mel spectrum* untuk dapat menghasilkan representasi yang baik dari spektral suara. Berikut merupakan persamaan dari proses DCT.

$$c_n = \sum_{k=1}^K (\log S_k) \cos n \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{K} \right]; n = 1, 2, \dots, K \quad (7)$$

Dimana:

S_k = Keluaran dari proses *filterbank* pada *index* k

K = Jumlah Koefisien

2.2.8. Cepstral Filtering

Proses yang digunakan untuk meningkatkan kualitas pengenalan suara, sekaligus merupakan proses terakhir pada tahapan MFCC. Berikut merupakan persamaan dari proses *cepstral filtering*.

$$W[n] = \left\{ 1 + \frac{L}{2} \sin \left(\frac{n\pi}{L} \right) \quad n = 1, 2, \dots, L \right\} \quad (8)$$

Dimana:

L = Jumlah *cepstral Coefficients*

N = *Index* dari *cepstral coefficients*

2.3. K-Nearest Neighbor

Algoritma K-Nearest Neighbor merupakan algoritma yang digunakan untuk melakukan proses klasifikasi data yang bekerja relatif menggunakan cara yang lebih sederhana dibandingkan dengan metode pengklasifikasian data lainnya. Metode ini akan melakukan proses klasifikasi terhadap data baru yang belum diketahui *class*-nya dengan cara memilih data sejumlah k yang letaknya paling dekat dengan data baru tersebut. Untuk menentukan *class* yang nantinya diprediksi untuk data baru adalah dengan cara mencari *class* terbanyak dari data terdekat sejumlah k .

Dengan demikian K-NN mengambil keputusan yaitu data baru d merupakan anggota dalam *class* C berdasarkan beberapa tetangga terdekat dari d . Berikut persamaan yang dapat digunakan untuk menentukan dekat atau jauhnya tetangga dari data yang dimiliki, yaitu.

$$d = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2} \quad (9)$$

$$d = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (10)$$

Klasifikasi *K-NN* dilakukan dengan mencari k -buah tetangga terdekat dan memilih kelas dengan k_i terbanyak pada kelas ω_i .

3. Hasil dan Pembahasan

Data yang digunakan dalam penelitian ini berjumlah 120 data yang diperoleh dari enam naracoba, kata yang di rekam adalah dengan kata kunci “buka” dan “tutup”, masing-masing kata dilakukan perekaman sebanyak 10 kali pengulangan. Rata-rata durasi perekaman adalah 1 detik, dan format file yang digunakan adalah .wav.

Pada penelitian sebelumnya jumlah data yang digunakan sebanyak 180 dengan 3 kata kunci, yaitu “klasik”, “dangdut”, “pop”, dan masing-masing kata akan dilakukan perekaman sebanyak 10 kali. Proses klasifikasi menggunakan algoritma LVQ dengan akurasi tertinggi yang diperoleh adalah α 0.1, dan penurunan *learning rate* 0.1, oleh karena itu parameter pelatihan LVQ yang digunakan yaitu α 0.1, dengan maksimum *epoch* 1000, minimum *learning rate* 0.0001, dan pengurangan α 0.1. Berikut adalah *table* dari hasil pengujian dari penelitian sebelumnya.

Table 1. Hasil Pengujian Sistem

No	Naracoba	Kelas	Jumlah Data	Jumlah Data Tepat Dikenali %	
				DL	DB
1	Naracoba 1	Klasik	10	100	90
2	Naracoba 1	Dangdut	10	100	90
3	Naracoba 1	Pop	10	100	70
4	Naracoba 2	Klasik	10	70	30
5	Naracoba 2	Dangdut	10	100	30
6	Naracoba 2	Pop	10	100	50
7	Naracoba 3	Klasik	10	90	60
8	Naracoba 3	Dangdut	10	100	50
9	Naracoba 3	Pop	10	70	50
10	Naracoba 4	Klasik	10	90	30

11	Naracoba 4	Dangdut	10	100	30
12	Naracoba 4	Pop	10	90	60
13	Naracoba 5	Klasik	10	100	30
14	Naracoba 5	Dangdut	10	100	20
15	Naracoba 5	Pop	10	70	30
16	Naracoba 6	Klasik	10	90	90
17	Naracoba 6	Dangdut	10	100	10
18	Naracoba 6	Pop	10	80	-
Rata-rata akurasi				92	46

4. Kesimpulan

Dari penelitian yang dilakukan oleh Nurhamidah, N., Djamal, E.C., dan Ilyas, R., dengan judul penelitian tentang Perintah Menggunakan Sinyal Suara dengan Mel-Frequency Cepstrum Coefficients dan Learning Vector Quantization, dapat ditarik kesimpulan dengan menggunakan maksimum *epoch* 1000, minimum *learning rate* 0.0001, α 0.1 dan terdapat pengurangan α 0.1 dari total 180 data uji yang telah dilatih dan menghasilkan akurasi sebesar 92% dengan hasil pengenalan 165 data dikenali dan 15 data tidak dikenali. Dan untuk pengujian data baru menghasilkan akurasi 46% dan data yang berhasil dikenali sebanyak 84 data dan 96 data lainnya tidak dikenali.

References

- [1] A. Anggoro, S. Herdjunanto dan R. Hidayat, "MFCC dan KNN untuk Pengenalan Suara Artikulasi P", *Avitec*, vol. 2, no. 13-17, 2020.
- [2] A. K. Saputro, M. Ulum, A. Karim and R. Alfita, "Perancangan *Smart-TV* Menggunakan Perintah Suara dengan Metode *Hidden Markov Model*", *Jurnal Teknik Elektro dan Komputer Triac*, vol. 7, no.2, p. 1-3, 2020.
- [3] C. Adipradana, T. S. Pamungkas and A. W. Hidayat, "Analisis Spektrum Perintah Suara Berdasarkan Gender menggunakan Algoritma K-Nearest Neighbors", *Tecnoscienza*, vol. 4, no. 1, p. 142-146, 2019.
- [4] E. Susanti1, S. M. A. Sasongko2 and I. G. P. Suta, "Klasifikasi Suara Berdasarkan Usia Menggunakan *Mel-Frequency Cepstral Coefficient (MFCC)* dan *K-Nearest Neighbor (K-NN)*", *Dielektrika*, vol. 4, no.2, p. 120-126, 2017.
- [5] F. D. Septria, J. Raharjo and N. Ibrahim, "Klasifikasi Emosi Berdasarkan Sinyal Suara Manusia Menggunakan Metode *K-Nearest Neighbor (K-NN)*", *e-Proceeding of Engineering*, vol.6, No. 2, p. 6-8, 2019.
- [6] F. N. Suciani, E. C. Djamal, R. Ilyas. "Identifikasi Nama Surat Juz Amma dengan Perintah Suara Menggunakan MFCC dan Backpropagation", *Seminar Nasional Aplikasi Teknologi Informatika (SNATI)*, p. 19-20, 2018.
- [7] I. G. Harsemandi, M. Sudarma and N. Pramaita, "Implementasi Algoritma K-Nearest Neighbor pada Perangkat Lunak Pengelompokan Musik untuk Menentukan Suasana Hati", *Teknologi Elektro*, vol. 16, no.1, p. 17-19, 2017
- [8] I. S. Permana, Y. I. Nurhasanah, A. Zulkarnain. "Implementasi Metode MFCC dan DTW untuk Pengenalan Jenis Suara Pria dan Wanita". *MIND Jurnal*, vol. 3, No. 1, p. 49-63, 2018.
- [9] L. Fitria, K. Muttaqin and M. S. Nasution. "Implementasi Speech Recognition pada Kata Kerja Dasar Menggunakan Metode MFCC", *Jurnal Informatika dan Teknologi Komputer*, vol. 2, No. 1, p. 43-50, 2021.
- [10] . Nurhamidah and E. C. Djamal, "Perintah Menggunakan Sinyal Suara dengan Mel-Frequency Cepstrum Coefficients dan Learning Vector Quantization" *Seminar Nasional Aplikasi Teknologi Informasi, ISSN: 1907 – 5022*, p. 11-14, 2017.

- [11] . Hartayu, C. Anam and D. A. Aziz, "Simulasi Kestraksi Fitur Suara Menggunakan Mel-Frequency Cepstrum Coefficient", *Jurnal Sains dan Informatika*, vol. 8, no. 1, p. 83-85, 2022.
- [12] . Ariyanti, S. S. Adi and S. Purbawanto, "Sistem Buka Tutup Pintu Otomatis Berbasis Suara Manusia" *Electronics, Informatics, and Vocational Education (ELINVO)*, vol. 3, no. 1, p. 83-91, 2018.
- [13] Safriadi. 2020. "Klasifikasi Gender Berdasarkan Suara dengan Naïve Bayes dan Mel Frequency Cepstral Coefficient". *Vocational Education and Tehnology (VOCATECH)*, vol. 2, No.1, pp. 19-26, 2020.
- [14] W. Muldayani, A. R. Chaidir, G. D. Kalando and C. S. Sarwono. "Pengendali Robot Lengan Berbasis Perintah Suara Menggunakan MFCC dan ANN". *Seminar Nasional Teknik Elektro*, 2017.
- [15] Y. Fatman and Islamiyati. "Pengenalan Suku Kata Bahasa Indonesia Menggunakan LPC dan Backpropogation Neural Network". *Journal of Information Technology and Computer Science (Jointecs)*. Vol. 5, No. 3, pp. 155-166, 2020.