

Perbandingan Teknik Ekstrak Ciri Suara Pembicara Antara Metode MFCC Dan LPC Untuk Pengenalan Suara

*Comparison of Speaker Voice Feature Extract Techniques Between MFCC And LPC Methods
For Voice Recognition*

Budi Triandi¹, Herman Mawengkang*², Syahril Efendi³

¹ Fakultas Teknik dan Ilmu Komputer, FASIILKOM, FTIK-UPU, FASILKOM USU, Medan

^{2,3} Fakultas Ilmu Komputer, FASILKOM USU, Medan

e-mail: [1buditriandi@gmail.com](mailto:buditriandi@gmail.com), [*2mawengkang@usu.ac.id](mailto:mawengkang@usu.ac.id), [3syahnyata1@gmail.com](mailto:syahnyata1@gmail.com)

Abstrak

Suara pembicara memiliki informasi yang kompleks dalam bidang pengenalan suara, begitu juga halnya dalam bidang biometric yang menganalisa fisik dan perilaku manusia yang bertujuan untuk proses autentifikasi. Suara merupakan bagian dari karakteristik seseorang, terdapat banyak cara untuk mengenali suara seseorang yang hingga saat ini menjadi topik pembahasan yang sangat populer dalam penelitian, dalam metode pengenalan suara pembicara tahapan penting yang dilakukan adalah mengekstrak ciri dari suara tersebut sebagai inputan awal untuk dilakukan proses berikutnya, ada beberapa metode yang dikenal dalam bidang Speech Recognition (SR) seperti Mel Frequency Cepstral Calculation (MFCC) dan Linear Predictive Coding (LPC). Pekerjaan ini bertujuan untuk memberikan gambaran dari metode ekstraksi ciri yang efektif untuk identifikasi ciri yang mengarah ke pengenalan suara. Hasil yang didapat dari perbandingan kedua metode ini bahwa metode MFCC lebih baik jika dibandingkan dengan metode LPC dalam ekstrak ciri pada sinyal suara. Metode MFCC dan LPC memiliki tingkat keakuratan diatas 70%, kecepatan proses dan waktu ekstraksi ciri MFCC lebih unggul sedangkan untuk waktu training LPC lebih unggul dibandingkan, untuk klasifikasi menggunakan PCA sekaligus digunakan untuk mereduksi dan untuk menyederhanakan dataset.

Kata kunci— Perbandingan, Metode ekstrak ciri, MFCC, LPC, Speech Recognition

Abstract

The speaker's voice has complex information in the field of speech recognition, as well as in the field of biometrics which analyzes physical and human behavior for the purpose of the authentication process. Voice is part of a person's characteristics, there are many ways to recognize someone's voice which until now has become a very popular topic of discussion in research, in the speaker's voice recognition method the important step is to extract the characteristics of the voice as initial input for the next process, There are several methods known in the field of Speech Recognition (SR) such as Mel Frequency Cepstral Calculation (MFCC) and Linear Predictive Coding (LPC). This work aims to provide an overview of an effective feature extraction method for the identification of traits leading to speech recognition. The results obtained from the comparison of these two methods are that the MFCC method is better than the LPC method in extracting features of the sound signal. The MFCC and LPC methods have an accuracy rate above 70%, the process speed and time of feature extraction are superior to MFCC, while LPC training time is superior to that for classification using PCA as well as to reduce and simplify datasets.

Keywords— comparison, feature extraction method, MFCC, LPC, Speech Recognition

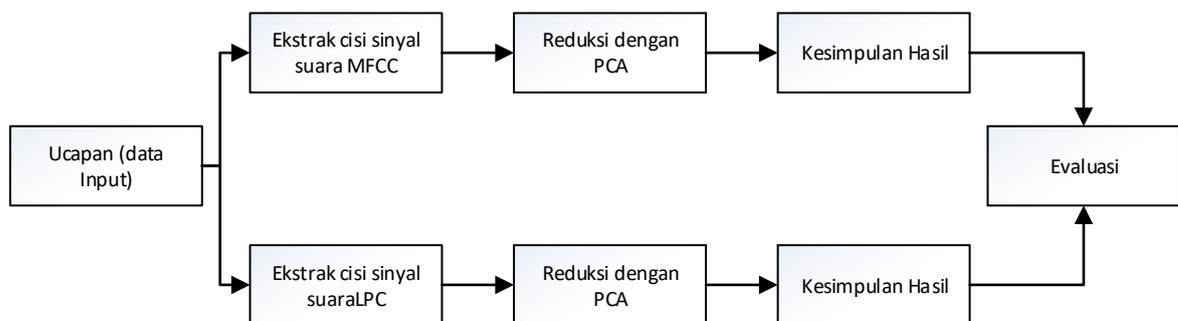
1. PENDAHULUAN

Karakteristik seseorang dapat dibedakan berdasarkan apa yang mereka miliki, seperti suara yang dimiliki masing-masing pembicara yang merupakan bagian dari karakteristik seseorang, manusia dapat mengenali seseorang melalui suara, seperti perilaku seseorang, kondisi seseorang dapat dibedakan melalui ucapan termasuk emosi seseorang tersebut [1]. Teknologi yang berkembang saat ini yaitu ASR (*Automatic Speech Recognition*) yang membuktikan bahwa perangkat komputer dapat mengenal suara manusia. Pengenalan suara merupakan hal yang sangat penting dalam melakukan interaksi antara manusia dengan mesin [1]. Ekstraksi fitur suara merupakan tahapan terpenting pada sistem pengenalan suara [2], yang mempunyai tujuan untuk memseleksi fitur mana yang memiliki makna. Ada beberapa kendala yang akan dihadapi dalam proses ekstraksi fitur. Kendala tersebut muncul dari variabilitas suara seperti kondisi seseorang seperti lingkungan yang menyebabkan tingginya noise seperti kebisingan, mikrofon, saluran transmisi dan gema [3]. Penelitian ini bertujuan untuk mengetahui bentuk dari pola suara yang diucapkan oleh pengguna agar dapat terlihat perbedaan dari setiap ciri suara sehingga dapat menghasilkan tingkat prediksi yang optimal dalam mengarah ke pengenalan ciri suara [4]. Hal paling penting dalam pengenalan suara adalah penggunaan fitur/ciri yang sesuai dalam mengekstraksi data dan informasi yang terkandung dalam sinyal suara [5,6,7].

Ekstraksi ciri suara bertujuan untuk menarik nilai koefisien dari ciri suara tersebut sehingga berguna untuk menentukan perbedaan pola, sehingga untuk menyeleksi kategori untuk klasifikasi itu menjadi lebih mudah [8]. Ada banyak metode ekstraksi ciri suara diantaranya yaitu metode *Linear Predictive Coding* (LPC), *Melf-Frequency Cepstral Coefficients* (MFCC), belakangan ini para peneliti mengatakan bahwa untuk hasil yang optimal dalam pengenalan suara bergantung terhadap teknik ekstraksi ciri dan kualitas ekstraksi ciri suara [9]. Maka berdasarkan hal tersebut penelitian ini melakukan studi perbandingan untuk metode ekstraksi ciri suara, MFCC merupakan metode dengan penerapan teknik ekstraksi ciri yang mendekati sistem pendengaran manusia. Sehingga metode ini masih lebih baik dibandingkan dengan metode yang lainnya.

2. METODE PENELITIAN

Penelitian ini digunakan untuk mencari metode terbaik untuk ekstraksi ciri sinyal suara dengan melakukan uji coba, langkah awal yaitu menguraikan hasil perbandingan dari masing-masing metode ekstraksi fitur dari sinyal suara, yaitu metode MFCC dan LPC, langkah berikutnya melakukan perbandingan pengenalan suara melalui ciri sinyal suara, langkah terakhir membandingkan hasil dari ekstraksi ciri suara yang dihasilkan dari uji coba metode masing-masing dari segi waktu. Untuk mengatasi masalah dimensi vektor ciri yang tinggi, dalam uji coba nantinya menggunakan *Principle Component Analysis* (PCA)



Gambar 1. Blok Diagram Metodologi Penelitian Perbandingan Teknik Ekstraksi Ciri

2.1. Dataset Ucapan

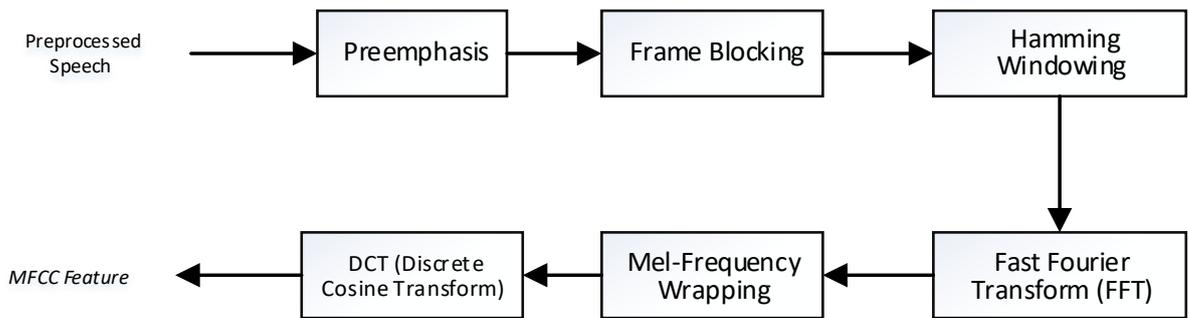
Dataset yang digunakan menggunakan RAVDESS diambil sampel data sebanyak 15 file suara dengan format file *.WAV, file diambil secara acak sebagai pengujian dengan suara ucapan dari actor yang berbeda

2.2. Estrak Fitur Suara

Data input yang dikumpulkan untuk pengenalan suara sering dirusak oleh noise selama fase penangkapan [10]. Karena gangguan ini, ekstraksi fitur dan klasifikasi menjadi kurang akurat. Ini berarti bahwa peningkatan data input merupakan langkah penting dalam sistem prediksi pada pengenalan seseorang[11].

2.3. Metode Ekstraksi Fitur MFCC

Metode MFCC *Melf-Frequency Cepstral Coefficients* sebuah metode ekstark fitur yang sering digunakan dalam mengekstrak ciri sinyal suara pembicara yang nantinya akan dikenali. Ekstraksi adalah representasi parametric dari sinyal akustik bertujuan agar kinerja pengenalan suara menghasilkan hasil yang lebih baik. MFCC mengadopsi konsep indra pendengaran seperti manusia[12], perbedaan dari frekuensi merupakan dasar dalam MFCC seperti panca indra pendengaran manusia, proses ekstraksi ciri MFCC dalam merepresentasikan sinyal ciri suara seperti layaknya indra pendengaran manusia. Berikut ini adalah proses MFCC dalam melakukan ekstrak ciri sinyal suara :



Gambar 2. Blok Melf-Frequency Cepstral Coefficients (Mfcc)

2.4.1. **Preemphasis** adalah tahapan proses filter sinyal suara tahapan ini dilakukan setelah pengolahan data sampel diperoleh. Proses ini dilakukan untuk mendapatkan *spectral* yang lebih baik, berikut adalah persamaan “(1)” dalam proses filter sinyal.

$$y(n)=s(n)- \alpha s (n-1) \dots\dots\dots(1)$$

2.4.2. **Frame Blocking** adalah proses segmentasi yang dilakukan menjadi beberapa *frame* sehingga saling bertumpang tindih (*overlap*), proses segmentasi dilakukan setelah sinyal suara didapatkan. berikut adalah persamaan “(2)” dalam proses frame blocking.

$$f_i(n) = y(Ml + n) \dots\dots\dots(2)$$

2.4.3. **Windowing** adalah suatu proses untuk menghaluskan spectrum setelah proses frame blocking, Windowing bertujuan mengurangi discontinue dari setiap ujung frame dari proses frame blocking. berikut adalah persamaan “(3)” dalam proses windowing.

$$X(n) = f_l (n)w(n) \dots\dots\dots(3)$$

2.4.4. **Fast Fourier Transform (FFT)** adalah pengembangan dari algoritma Discrete Fourier Transform (DFT) yang digunakan untuk mengubah sinyal digital pada ranah waktu ke ranah frekuensi (Abriyono dan Harjoko, 2012). berikut adalah persamaan “(4)” dalam proses FFT

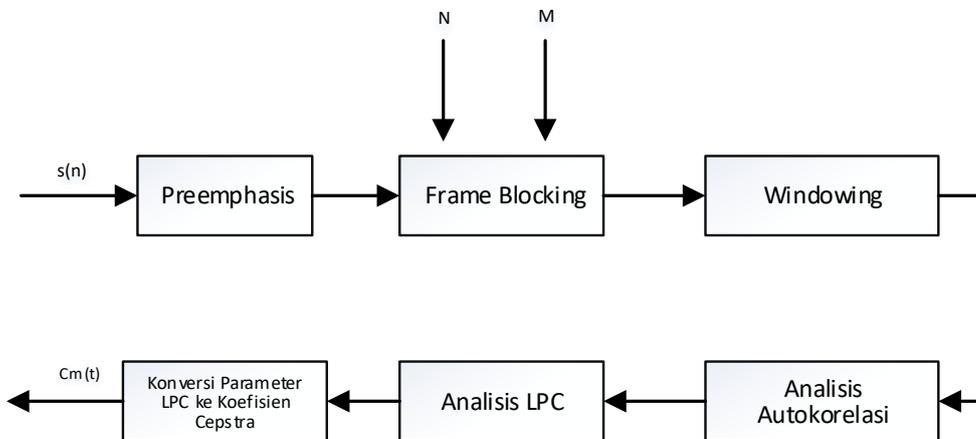
$$d[k] = \sum_{n=0}^{N-1} X(n)e^{-j\frac{2\pi}{N}nk}; k = 0,1,2, \dots N - 1 \dots(4)$$

2.4.5. **Mel-Frequency Wrapping** merupakan filter berupa filterbank untuk mengetahui ukuran energi dari frequency band tertentu dalam signal suara (Miftahuddin dan Hakim, 2017; Putra, 2011). Persepsi cara pendengaran pada manusia terhadap sebuah frekuensi suara yang terdengar tidak bisa diukur kedalam skala linear. berikut adalah persamaan “(5)” dalam proses Mel-Frequency Wrapping

$$y[i] = \sum_{j=i}^G T[j]H_i[j] \dots\dots\dots(5)$$

2.4. **Linier Predictive Coding (LPC)**

Terdapat beberapa penelitian yang menggunakan metode LPC karena menurut (Rabiner, Juang, 1993) LPC memiliki kelebihan yaitu sederhana dan dapat diterapkan pada perangkat lunak dan keras. LP memiliki dua komponen utama yaitu *Encoding* dan *Decoding* [15]. Bagian *encoding* digunakan untuk memeriksa dan membagi sinyal suara menjadi beberapa bagian, sedangkan *decoding* digunakan untuk mengubah suara yang telah diproduksi.



GAMBAR 2. BLOK LINIER PREDICTIVE CODING (LPC)

2.4.1. Pre-emphasis dalam PCA adalah proses untuk menghilangkan noise dari suara sebagai masukan. Sinyal noise dalam suara memiliki perbedaan signifikan dari sinyal suara lainnya, berikut ini adalah persamaan “(6)” dalam tahapan per-emphasis pada PCA.

$$y(n)=s(n)- \alpha s (n-1) \dots\dots\dots(6)$$

2.4.2. *Frame Blocking* dalam PCA bersifat sama seperti halnya MFCC untuk membagi sinyal suara menjadi beberapa block suara dan menggunakan persamaan “(7)” *frame blocking*.

$$f_i(n) = y(Ml + n) \dots\dots\dots(7)$$

2.4.3. Windowing adalah proses untuk pengecekan dan menghapus diskontinu sinyal yang sudah dibagi menjadi beberapa segmen.

$$w(n) = 0,54 - 0,46 \cdot \cos\left(\frac{2\pi n}{M-1}\right), 0 < n \leq N - 1$$

\dots\dots\dots(8)

2.4.4. Analisis Autokorelasi tahapan ini merupakan upaya mengkorelasikan sinyal tahapan ini dilakukan setelah mendapatkan hasil segmentasi dari proses windowing.

$$R_n(i - k) = \sum_{m=0}^{N-1-(i-k)} S_n(m)S_n(m + i - k) \dots\dots(9)$$

2.4.5. Analisis LPC tahapan ini penentuan nilai parameter LPC dengan mengubah nilai dari autokorelasi menjadi koefisien LPC, pantulan dan perbandingan logaritmis.

$$\sum_{k=1}^p R_n(|i - k|) \widehat{a}_k = R_n(i) \leq i \leq p \dots\dots\dots(10)$$

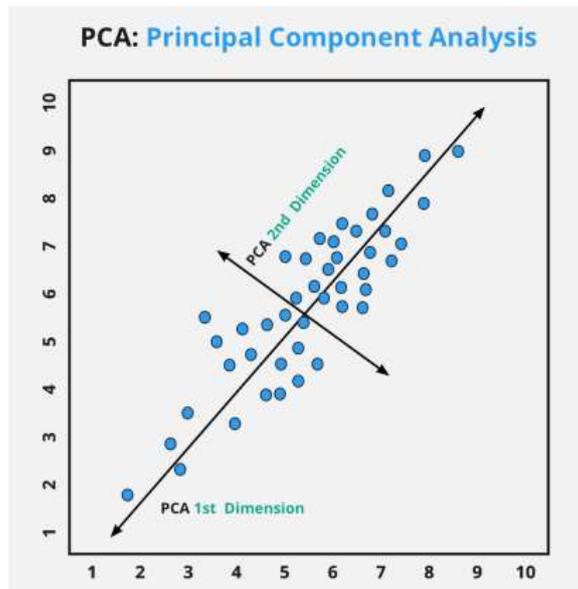
2.4.6. Konversi parameter LPC pada tahapan ini bertujuan untuk mendapatkan koefisien spectral berdasarkan representasi spectrum transformasi fourier.

$$c_m = a_m \sum_{k=1}^{m-1} \binom{k}{m} c_k a_{m-k}, 1 \leq m \leq p \dots\dots\dots(11)$$

$$c_m = a_m \sum_{k=m-p}^{m-1} \binom{k}{m} c_k a_{m-k}, m > p \dots\dots\dots(12)$$

2.5. *Principal Component Analysis (PCA)*

Analisis komponen utama bertujuan untuk mereduksi data dalam penelitian ini PCA digunakan untuk mereduksi vektor ciri sinyal suara dalam dimensi yang tinggi, dengan mempertahankan informasi dari kumpulan data sehingga perhitungan tetap dapat dilakukan secara optimal [13]. Konsep utama dalam PCA menghapus redundansi pada dimensi dari korelasi sehingga menghasilkan korelasi utuh.



Gambar 2. Principal Componen Analysis

Principal Components Analysis (PCA) teknik yang digunakan untuk menyederhanakan data set, tergolong dalam teknik statistic multivariate, untuk melakukan perhitungan pada komponen utama, dengan menghitung nilai eigen dan vektor eigen, dari matrik kovarians dapat dilakukan perhitungan vektor eigen n variable yang saling berhubungan dengan kondisi $n \times n$. komponen utama didefenisikan berdasarkan eigen vektor, bobot dari egien vektor berupa nilai koefisien. Untuk *gen i* didefenisikan oleh komponen sepanjang sumbu [2] :

$$a'_{ij} = \sum_{t=1}^n a_{it} v_j \dots\dots\dots(13)$$

3. HASIL DAN PEMBAHASAN

3.1. Perbandingan Ekstrak Ciri Suara Dari Kedua Metode

Pada proses perbandingan membandingkan ekstrak ciri menggunakan metode MFCC dan LPC, dan dengan menggunakan PCA sebagai klasifikasi untuk menentukan akurasi, dari perbandingan ekstrak ciri sinyal suara dapat dilihat dari hasil berikut ini :

Tabel.1. Hasil Uji Ekstrak ciri menggunakan MFCC Feature Extraction

n-dimension data	MFCC Ekstrak Ciri									
	PCA	1	2	3	4	5	6	7	8	9
2		86	86.6	87	87.4	87.8	88.2	88.6	89	89.4
4		85.4	86	86.4	86.8	87.2	87.6	88	88.4	88.8
6		85	85.6	86	86.4	86.8	87.2	87.6	88	88.4
8		84.29	84.89	85.29	85.69	86.09	86.49	86.89	87.29	87.69
10		85.71	86.31	86.71	87.11	87.51	87.91	88.31	88.71	89.11
12		86.7	87.3	87.7	88.1	88.5	88.9	89.3	89.7	90.1
14		80.4	81	81.4	81.8	82.2	82.6	83	83.4	83.8
16		72.71	73.31	73.71	74.11	74.51	74.91	75.31	75.71	76.11
18		72.14	72.74	73.14	73.54	73.94	74.34	74.74	75.14	75.54
20		73.5	74.1	74.5	74.9	75.3	75.7	76.1	76.5	76.9
22		74.14	74.74	75.14	75.54	75.94	76.34	76.74	77.14	77.54
24		73.3	73.9	74.3	74.7	75.1	75.5	75.9	76.3	76.7

Tabel.2. Hasil Uji Ekstrak ciri menggunakan LPC Feature Extraction

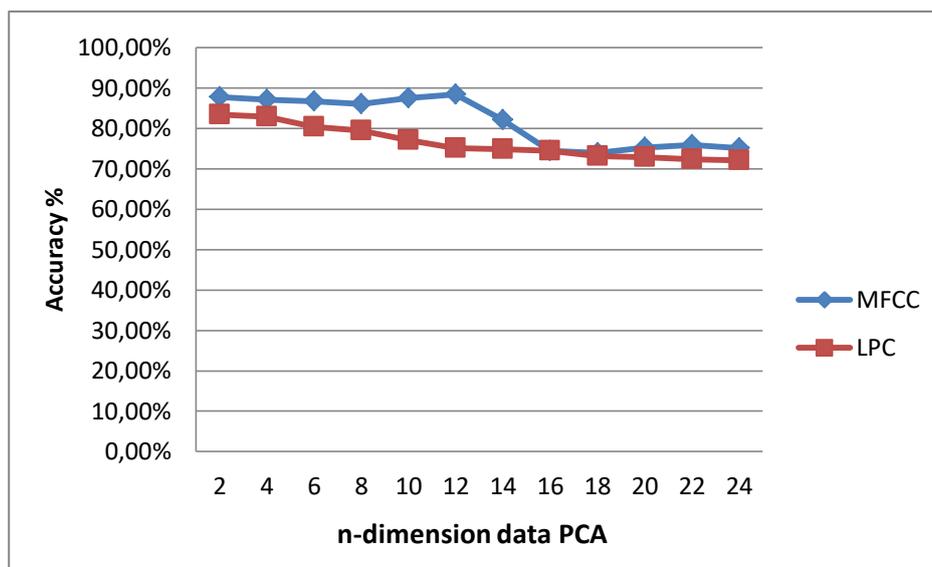
n-dimension data	LPC Ekstrak Ciri									
	PCA	1	2	3	4	5	6	7	8	9
2		81.67	82.27	82.67	83.07	83.47	83.87	84.27	84.67	85.07
4		81.1	81.7	82.1	82.5	82.9	83.3	83.7	84.1	84.5
6		78.64	79.24	79.64	80.04	80.44	80.84	81.24	81.64	82.04
8		77.74	78.34	78.74	79.14	79.54	79.94	80.34	80.74	81.14
10		75.3	75.9	76.3	76.7	77.1	77.5	77.9	78.3	78.7
12		73.4	74	74.4	74.8	75.2	75.6	76	76.4	76.8
14		73.1	73.7	74.1	74.5	74.9	75.3	75.7	76.1	76.5
16		72.64	73.24	73.64	74.04	74.44	74.84	75.24	75.64	76.04
18		71.4	72	72.4	72.8	73.2	73.6	74	74.4	74.8
20		71.1	71.7	72.1	72.5	72.9	73.3	73.7	74.1	74.5
22		70.6	71.2	71.6	72	72.4	72.8	73.2	73.6	74
24		70.3	70.9	71.3	71.7	72.1	72.5	72.9	73.3	73.7

3.2. Perbandingan Akurasi pengenalan ciri suara dengan 12 parameter PCA

Dari data uji coba dengan 9 ciri suara yang mengarah pada nilai yang membedakan tingkat akurasi pengenalan suara dengan 12 dimensi data, menunjukkan perbedaan pada vektor ciri bahwa teknik ekstrak ciri sangat mempengaruhi tingginya tingkat akurasi pengenalan suara dapat dilihat pada data berikut ini.

Tabel3. Perbandingan Hasil Uji Tingkat Akurasi metode MFCC dan LPC

n-dimension data PCA	Tingkat Akurasi MFCC	Tingkat Akurasi LPC
2	87.78%	83.45%
4	87.18%	82.88%
6	86.78%	80.42%
8	86.07%	79.52%
10	87.49%	77.08%
12	88.48%	75.18%
14	82.18%	74.88%
16	74.49%	74.42%
18	73.92%	73.18%
20	75.28%	72.88%
22	75.92%	72.38%
24	75.08%	72.08%



Gambar 3. Grafik Hasil Uji Perbandingan Kedua Metode Mfcc Dan Lpc

3.3. Estimasi Waktu Ekstrak Ciri Dan Waktu Training

Meskipun LPC dikenal sebagai metode yang sederhana dalam mengekstrak sinyal suara, namun pada kenyataannya dalam pengujian membutuhkan waktu yang lebih lama jika dibandingkan dengan MFCC, sedangkan untuk waktu training LPC lebih cepat dibandingkan MFCC. Berikut ini adalah perbandingan estimasi waktu yang dibutuhkan dalam melakukan ekstrak ciri dan waktu training mengambil rata-rata dari uji coba masing-masing parameter dimensi PCA.

Tabel4. Perbandingan Ekstimasi Waktu Ekstrak ciri dan Waktu Training

Metode Ekstrak Ciri	Waktu Ekstrak Ciri	Waktu Training
MFCC	0,089	0,348
LPC	0,149	0,216

4. KESIMPULAN

Perbandingan dua metode ekstrak ciri suara dalam penelitian ini, memberikan gambaran bahwa untuk mendapatkan tingkat akurasi yang tinggi dalam pengenalan suara, sebaiknya mempertimbangkan penggunaan teknik ekstrak ciri sinyal suara, meski pun banyak teknik yang dapat digunakan namun ada kelebihan dan kekurangan dari teknik yang digunakan, menentukan metode ekstrak ciri yang akan digunakan dalam penelitian pengenalan suara merupakan langkah awal yang terpenting. Dalam perbandingan teknik ekstrak ciri MFCC merupakan metode terbaik jika dibandingkan dengan LPC dengan tingkat akurasi mencapai 87,78% sedangkan LPC 83,45% dengan n dimensi bernilai 2, sedangkan dari segi estimasi ekstrak ciri meski LPC lebih sederhana tetapi dalam penggunaan waktu Ekstrak LPC lebih lama jika dibandingkan dengan MFCC, sedangkan untuk estimasi waktu dalam melakukan training MFCC lebih lama jika dibandingkan dengan LPC. Dalam hal ini vektor ciri yang panjang dengan variable yang berlebihan mengakibatkan hasil ekstrak yang buruk. Penggunaan PCA dapat mereduksi data sehingga dapat meminimalisir eigen vektor ciri yang berlebihan.

5. SARAN

Untuk penelitian berikutnya dapat difokuskan pada teknik prediksi yang lebih baik lagi, sehingga bukan hanya sebatas untuk pengenalan suara saja melainkan untuk eksplorasi untuk pengenalan suara kearah yang lebih luas.

DAFTAR PUSTAKA

- [1] Umar, R., Riadi, I., & Hanif, A. (2019). Analisis Bentuk Pola Suara Menggunakan Ekstraksi Ciri Mel-Frequency Cepstral Coefficients (MFCC). *CogITo Smart Journal*, 4(2), 294-304.
- [2] Khalil, R. A., Jones, E., Babar, M. I., Jan, T., Zafar, M. H., & Alhussain, T. (2019). Speech emotion recognition using deep learning techniques: A review. *IEEE Access*, 7, 117327-117345.
- [3] Eko, R. (2013). *SISTEM PENGENAIAN PENGUCAP MANUSIA DENGAN EKSTRAKSI CIRI MFCC DAN ALGORITMA JARINGAN SARAF TIRUAN PERAMBATAN BAIK SERAGAI PENGENAIANNYA PADA SISTEM KEHADIRAN* (Doctoral Dissertation, Diponegoro University).
- [4] Adriana, A. D. (2013). Perangkat Lunak Untuk Membuka Aplikasi Pada Komputer Dengan Perintah Suara Menggunakan Metode Mel Frequency Cepstrum Coefficients. *Komputa: Jurnal Ilmiah Komputer dan Informatika*, 2(1).
- [5] Prasetyo, B. H., Kurniawan, W., & Ichsan, M. H. H. (2017). Pengenalan emosi berdasarkan suara menggunakan algoritma hmm. vol, 4, 168-172.
- [6] Lugović, S., Dunder, I., & Horvat, M. (2016, May). Techniques and applications of emotion recognition in speech. In 2016 39th international convention on information and communication technology, electronics and microelectronics (mipro) (pp. 1278-1283). IEEE.

- [7] Koduru, A., Valiveti, H. B., & Budati, A. K. (2020). Feature extraction algorithms to improve the speech emotion recognition rate. *International Journal of Speech Technology*, 23(1), 45-55.
- [8] Gustina, S., Fadlil, A., & Umar, R. (2017). Sistem Identifikasi Jamur Menggunakan Metode Ekstraksi Ciri Statistik Orde 1 dan Klasifikasi Jarak. *Techno. com*, 16(4), 378-386.
- [9] 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.
- [10] Han, J., Zhang, Z., Ringeval, F., & Schuller, B. (2017, March). Reconstruction-error-based learning for continuous emotion recognition in speech. In *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 2367-2371). IEEE.
- [11] Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2008). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE transactions on pattern analysis and machine intelligence*, 31(1), 39-58.
- [12] Sharma, G., Umapathy, K., & Krishnan, S. (2020). Trends in audio signal feature extraction methods. *Applied Acoustics*, 158, 107020.
- [13] Winursito, A., Hidayat, R., & Bejo, A. (2018, March). Improvement of MFCC feature extraction accuracy using PCA in Indonesian speech recognition. In *2018 International Conference on Information and Communications Technology (ICOIACT)* (pp. 379-383). IEEE.
- [14] Rabiner, L., & Juang, B. H. (1993). *Fundamentals of speech recognition*. PTR Prentice-Hall. Inc., New Jersey.
- [15] Jain, U., Nathani, K., Ruban, N., Raj, A. N. J., Zhuang, Z., & Mahesh, V. G. (2018, October). Cubic SVM classifier based feature extraction and emotion detection from speech signals. In *2018 International Conference on Sensor Networks and Signal Processing (SNSP)* (pp. 386-391). IEEE.