

Identifikasi Suara Pada Sistem Presensi Karyawan Dengan Metode Ekstraksi MFCC

Yuwono Fitri Widodo¹, Sunardi², Abdul Fadlil³

1,3 Program Studi Teknik Informatika (Universitas Ahmad Dahlan, Yogyakarta)

2 Program Studi Teknik Elektro (Universitas Ahmad Dahlan, Yogyakarta)

Jl. Prof. DR. Soepomo Sh, Warungboto, Umbulharjo, Yogyakarta 55164

email : yuwono1708048020@webmail.uad.ac.id

Abstract

Biometrics is the study of patterns of characteristics to recognize or identify humans based on one or more parts of the human body, both chemical, physical, and behavioral characteristics, such as faces, fingerprints, sounds, hand geometry, or iris. Nowadays technology has developed using sound to be used as an application that facilitates humans. Voice identification process is very necessary to know the accuracy of the sound based on the characteristics possessed, because some humans have similarities in saying. In this study aims to determine the sound pattern based on speech. The method used for voice identification using the Mel-Frequency Cepstral Coefficients (MFCC) feature extraction method, is a feature extraction method that approaches the human hearing system and is able to recognize speech patterns.

Keywords: *biometrics, identification, mfcc, voice*

Abstrak

Biometrik adalah ilmu yang mempelajari pola ciri-ciri untuk mengenali atau mengidentifikasi manusia berdasarkan satu atau lebih dari bagian tubuh manusia, baik ciri-ciri kimia, fisik, maupun tingkah laku, seperti wajah, sidik jari, suara, geometri tangan, ataupun iris mata. Saat ini sudah berkembang teknologi menggunakan suara untuk dijadikan sebuah aplikasi yang memudahkan manusia. Proses identifikasi suara sangat diperlukan untuk mengetahui keakuratan suara berdasarkan ciri yang dimiliki, karena beberapa manusia memiliki kemiripan dalam berucap. Pada penelitian ini bertujuan untuk mengetahui pola suara berdasarkan ucapan. Metode yang digunakan untuk identifikasi suara menggunakan metode ekstraksi fitur Mel-Frequency Cepstral Coefficients (MFCC), merupakan metode ekstraksi fitur yang mendekati sistem pendengaran manusia dan mampu untuk mengenali pola ucapan.

Kata kunci: *biometrik, identifikasi, mfcc, suara.*

1. PENDAHULUAN

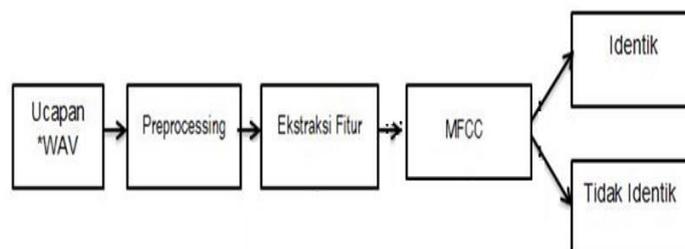
Dewasa ini perkembangan teknologi begitu pesat diberbagai bidang. Salah satu bidang dari kemajuan teknologi ini adalah kecerdasan buatan[1]. Dalam kegiatan sehari-hari dapat dijumpai alat-alat berteknologi dari perkembangan ilmu kecerdasan buatan yang digunakan untuk membantu pekerjaan manusia. Salah satu perkembangan kecerdasan buatan adalah biometrik pengenalan pola [2]. Biometrik adalah ilmu yang mempelajari pola ciri-ciri untuk mengenali [3] atau mengidentifikasi manusia berdasarkan satu atau lebih dari bagian tubuh manusia, baik ciri-ciri kimia, fisik, maupun tingkah laku, seperti wajah, sidik jari, suara[4], geometri tangan, ataupun iris mata. Menurut Kamus Besar Bahasa Indonesia Suara adalah bunyi yang dikeluarkan dari mulut manusia (seperti pada waktu bercakap-cakap,

menyanyi, tertawa, menangis, bunyi binatang, alat perkakas dan sebagainya). Menurut [5] suara merupakan suatu bentuk dari sebuah komunikasi. Manusia dapat menghasilkan suara untuk berkomunikasi [6] dengan menggunakan pita suara yang terdapat dirongga mulut dan suara yang dihasilkan mempunyai frekuensi yang berbeda-beda. Frekuensi yang dapat didengar oleh telinga manusia antara 20Hz sampai dengan 20.000Hz.

Saat ini sudah berkembang teknologi [7] menggunakan suara untuk dijadikan sebuah aplikasi yang memudahkan manusia [8]. Salah satunya adalah untuk presensi kehadiran [9] karyawan menggunakan suara. Kelebihan dari teknologi ini adalah memudahkan dalam proses absensi [10]. Selain keamanan sistem tinggi hal yang menjadi kekurangannya adalah apabila ada kemiripan suara atau perubahan suara dari pemilik maka sistem akan menolaknya. Proses identifikasi suara sangat diperlukan untuk mengetahui keakuratan suara berdasarkan ciri yang dimiliki, karena beberapa manusia memiliki kemiripan dalam berucap. Permasalahan ini menjadi dasar penulis untuk melakukan penelitian bagaimana mengidentifikasi suara ucapan yang identik berdasarkan ciri-ciri yang dimiliki. Adapun dalam ekstraksi ciri digunakan metode *Mel-Frequency Cepstrum Coefficient (MFCC)* [11]. Metode ini dilakukan dengan memasukkan data suara pengucap ke *software* Matlab.

2. METODOLOGI PENELITIAN

Proses dalam penelitian ini adalah melakukan identifikasi suara untuk menentukan suara yang identik atau tidak identik sesuai dengan suara seseorang. Teknik dan tahapan yang digunakan terdiri dari beberapa langkah seperti berikut:



Gambar 1. Alur poses Identifikasi

2.1. Data yang digunakan dalam penelitian ini menggunakan data suara karyawan

Basis data ini berisi rekaman data audio yang terdiri dari lima aktor, yaitu tiga pria dan dua wanita. Basis data ini berisi sekitar 250 sampel ucapan dari 5 peserta, sampel yang diambil adalah yang memiliki tingkat pengenalan manusia yang lebih baik dari 80% dan nilai kealamian lebih dari 60%.

2.2. Sampel Ucapan

Sampel pada penelitian ini sebanyak 50 data, dan akan dibagi menjadi dua yaitu untuk data latih dan data uji. Kalimat yang diucapkan adalah kata "Masuk" dengan suara yang diambil dari 5 aktor. Pada tahap ini, sampel ucapan dalam bentuk file *.wav dirubah menjadi sinyal ucapan dengan menggunakan aplikasi MATLAB R2013a. Sinyal ucapan ini akan menjadi deretan nilai dalam satuan matriks.

2.3. Pre-processing

Tahapan ini melakukan *pre-processing* pada rekaman yaitu dengan melakukan pemotongan rekaman suara sebagai sampling pada saat ekstraksi fitur. Tujuan dari sampling ini adalah agar pola dapat terlihat berbeda untuk masing-masing ucapan tertentu.

2.4. Sinyal Ucapan

Setelah tahapan pemotongan suara sebagai sampling tahapan selanjutnya adalah sinyal ucapan. Tahapan ini mengubah sampel ucapan bentuk *file .wav* menjadi sinyal ucapan yang dapat dibaca oleh komputer.

2.5. Ekstraksi Fitur

Ekstraksi ciri yang digunakan pada penelitian ini adalah menggunakan metode MFCC yang dimana metode ini adalah metode ekstraksi ciri yang mendekati sistem pendengaran manusia. MFCC sering digunakan untuk pemrosesan suara, karena dapat merepresentasikan sinyal dengan baik. Proses ekstraksi fiturnya adalah sebagai berikut:

2.5.1. Pre-emphasis

Proses pertama yaitu *pre-emphasi*, tujuannya adalah untuk *filtering* sinyal ucapan dan mengurangi nilai frekuensi sinyal suara masukkan sehingga hanya sinyal yang memiliki frekuensi tinggi yang dapat melewati proses filter. Selain itu juga dapat mengurangi noise-noise pada suara masukkan, sehingga hanya sinyal ucapan saja yang dapat ditangkap sistem. Persamaannya dapat dilihat pada rumus.

$$\rho(n) = s(n) - xs(n - 1) \quad (1)$$

dimana x adalah konstanta filter pre-emphasis, biasanya bernilai antara $0.9 < x < 1.0$

2.5.2. Framing

Pada proses ini input suara dipotong menjadi beberapa *frame* dengan durasi yang lebih pendek sebanyak matriks (M) yang disimpan di matriks Y dengan ukuran $M \times W$. Sinyal suara dilakukan segmentasi menjadi beberapa *frame* dengan cara tumpang tindih (*overlap*) agar tidak ada sinyal yang hilang. Proses ini terus berlanjut sampai seluruh sinyal masuk ke dalam satu atau lebih *frame*.

2.5.3. Windowing

Proses *windowing* dilakukan dengan tujuan untuk memperoleh sampel sinyal yang tepat dalam waktu interval yang sangat singkat. Proses ini menghasilkan window $X(t)$ dimana $t = 1, 2, 3, \dots, T$ yang disebut *frame*. Pada kasus ini akan menggunakan persamaan *hamming window* dengan rumus.

$$w(n) = 0,54 + 0,46\cos 2\pi n/N - 1, 0 \leq n \leq N - 1 \quad (2)$$

dimana n adalah jumlah sampel dan N adalah jumlah frame 2.5.4. *Fast Fourier Transform (FFT)*. FFT adalah salah satu *metode* untuk mengkonversikan dari sinyal suara menjadi sinyal frekuensi. Proses ini akan dilakukan terhadap semua *frame* dari sinyal yang sudah di *windowing*. FFT merupakan algoritma cepat untuk menerapkan *Discrete Fourier Transform (DCT)* yang beroperasi pada sinyal diskrit yang terdiri dari N sampel, persamaannya dapat dilihat pada rumus.

$$f(n) = \sum_{k=0}^{N-1} w_k e^{-2\pi jkn/N}, 0 \leq n \leq N - 1 \quad (3)$$

dimana w adalah *windowing*.

2.5.4. Mel Filterbank

Mel-Filterbank merupakan *triangular* dari *filterbank*, yang membedakan adalah range frekuensi *linier* dari hasil FFT yang kemudian dikonversi ke skala *Mel-Frequency* untuk mendapatkan batas-batas *filterbank*. Persamaan *Mel Filterbank* ada pada rumus.

$$B(f) = 1125 \times \ln(1 + f/700) \quad (4)$$

Proses *Mel-Filterbank* yang perlu dilakukan adalah menentukan batas atas dan bawah dari filter. Kemudian bagi range batas atas dan bawah sesuai dengan jumlah filter yang dibuat dan dapat diketahui batas atas dan bawah untuk setiap *filterbank* dalam skala *mel*. Kedua batas tersebut dikonversi kembali ke skala frekuensi linier.

2.5.5. Discrete Cosine Transform

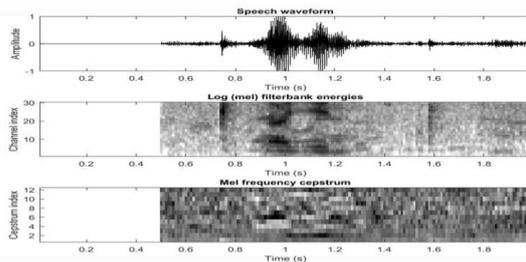
Proses terakhir adalah pengkonversian *domain* frekuensi ke *domain* waktu dengan menggunakan *Discrete Cosine Transform (DCT)*. Hasil log dari perkalian *domain* waktu menggunakan DCT menghasilkan *mel-frequency cepstrum coefficient (MFCC)*. Berikut adalah persamaan yang digunakan:

$$C_j = \sum_{i=0}^{M-1} x(i) \cos(j(i-1)/2\pi M) \quad (5)$$

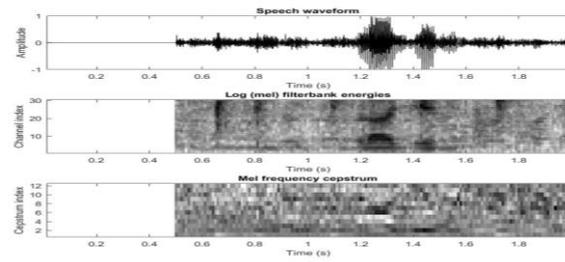
dimana $j = 1, 2, 3 \dots K$ adalah koefisien, dan M adalah jumlah *filter*. 3

3. HASIL DAN PEMBAHASAN

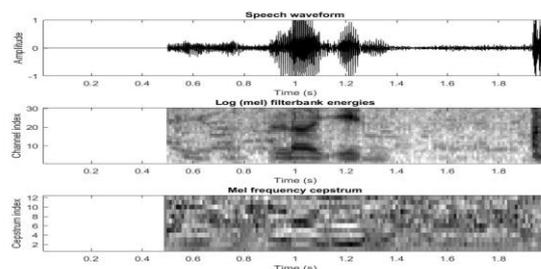
Penelitian ini menggunakan beberapa sampel suara dari karyawan dengan mengambil sampel kalimat “Masuk” Rekaman suara tersebut kemudian di *pre-processing*. Pada penelitian ini digunakan lima sampel suara untuk masing-masing percobaan. Sampel suara tersebut juga diucapkan oleh aktor yang berbeda. Sampel-sampel suara tersebut di cari polanya dengan menggunakan ekstraksi fitur MFCC dengan beberapa parameter yaitu $T_w = 25$; sebagai durasi masing-masing frame, $T_s = 10$; sebagai pembentukan *frame*, $\alpha = 0.97$; sebagai nilai koefisien *preemphasis*, $M = 20$; sebagai nilai *filterbank* pada setiap *channels*, $C = 12$; nilai *cepstral coefficients*, dari parameter tersebut dihasilkan fitur MFCC. Hasil ekstraksi suara menggunakan MFCC dari tiap sampel dapat dilihat dilihat pada Gambar.2.1 untuk sampel suara ke-1, Gambar.2.2 untuk sampel suara ke-2, Gambar.2.3 untuk sampel suara ke-3, Gambar.2.4 untuk sampel suara ke-4 dan Gambar.2.5 untuk sampel suara ke-5. Seluruh sampel suara yang sudah diekstraksi kemudian dianalisis dengan normalisasi sehingga pola suara dapat dilihat seperti pada Gambar.3. yang telah dinormalisasi dapat dilihat pada Gambar Proses perhitungan normalisasi dari ekstraksi ciri dilakukan untuk menyamakan nilai pola yang-dapat dilihat pada tabel 1. Setiap gambar menampilkan sampel untuk setiap ciri suara yang telah dianalisis



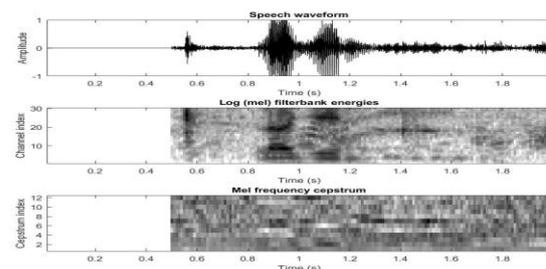
Gambar 1. Sampel Suara ke-1



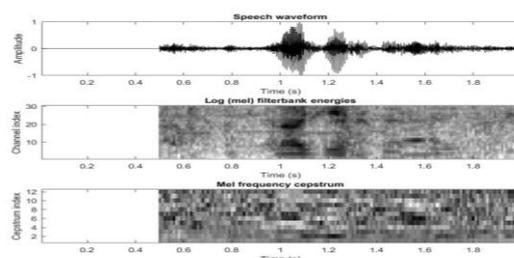
Gambar 2. Sampel Suara ke-2



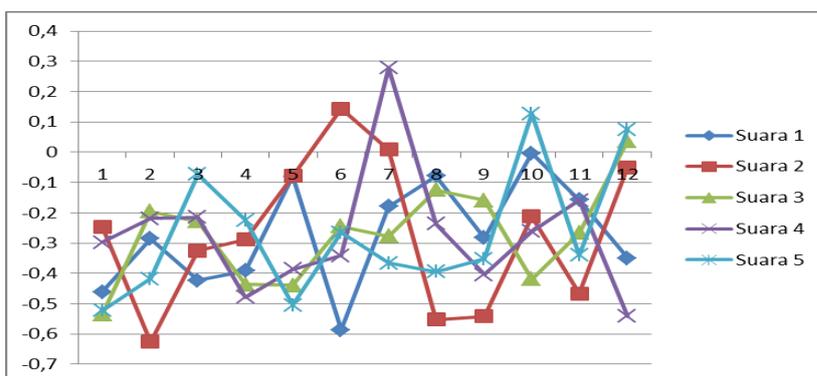
Gambar 3. Sampel Suara ke-3



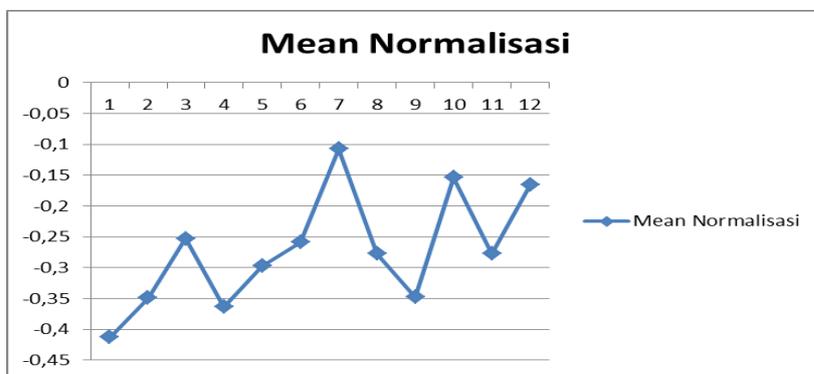
Gambar 4. Sampel Suara ke-4



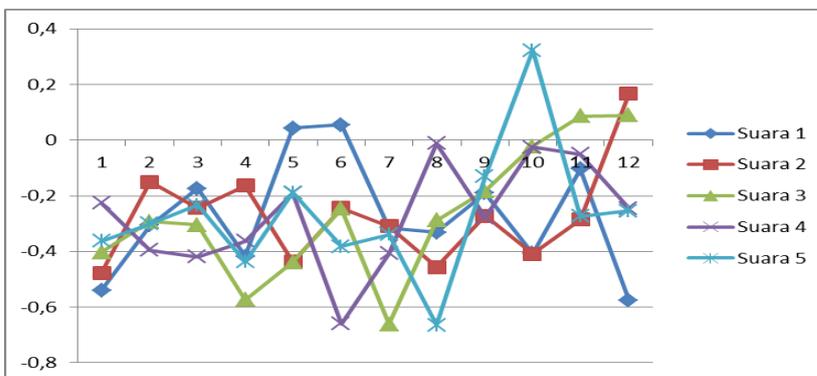
Gambar 5. Sampel Suara ke-5



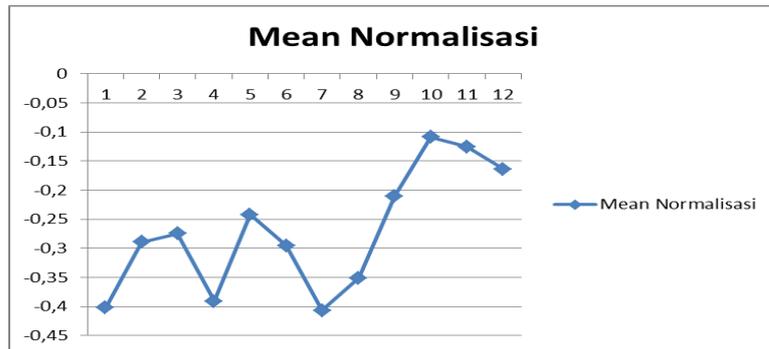
Gambar 6. Pola suara orang pertama



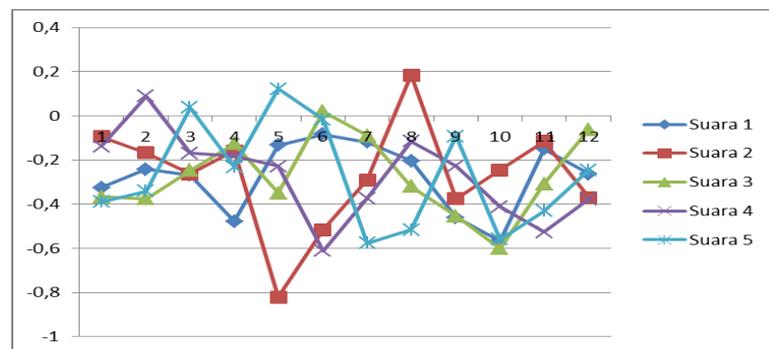
Gambar 7. Pola suara orang pertama yang telah dinormalisasi



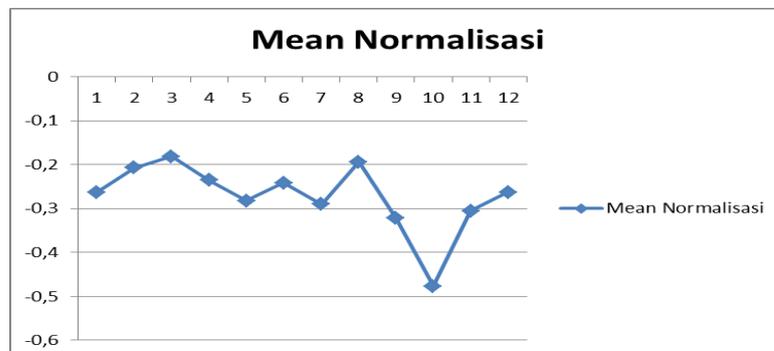
Gambar 8. Pola suara orang kedua



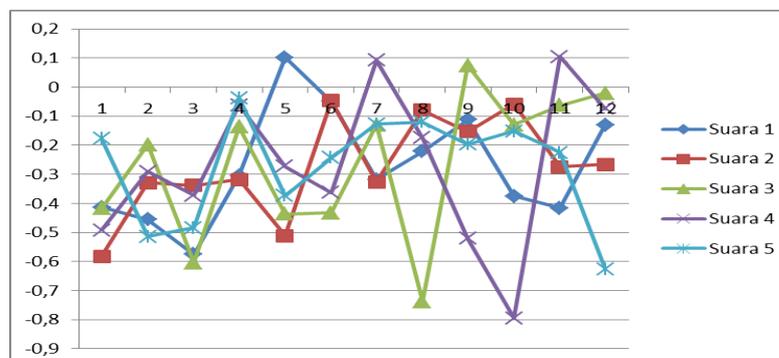
Gambar 9. Pola suara orang kedua yang telah dinormalisasi



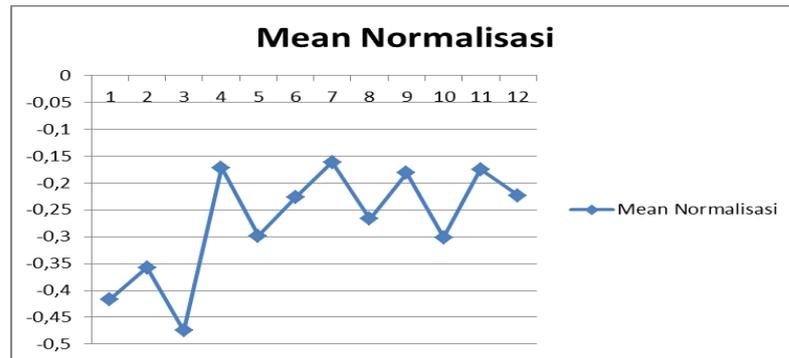
Gambar 10. Pola suara orang ketiga



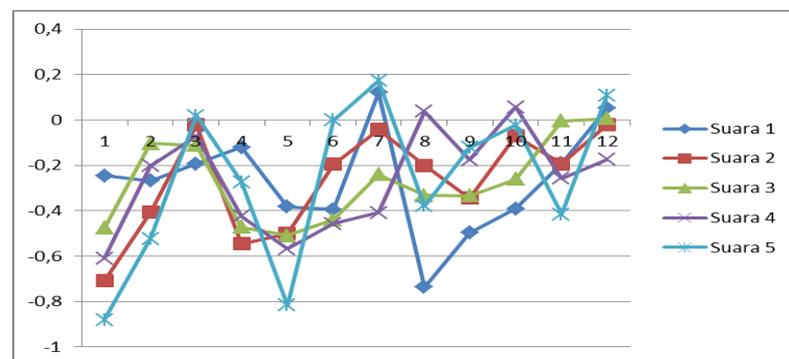
Gambar 11. Pola suara orang ketiga yang telah dinormalisasi



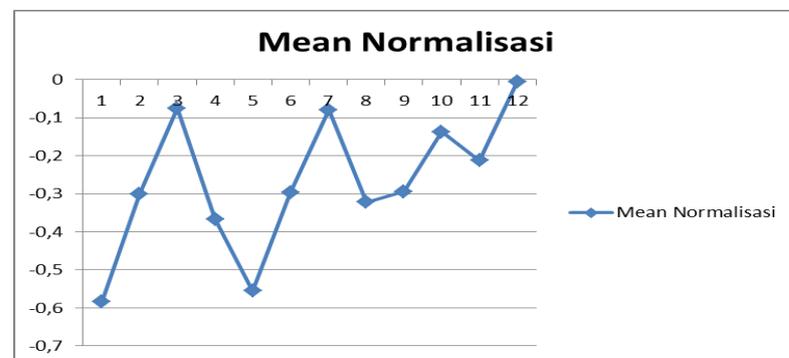
Gambar 12. Pola suara orang keempat



Gambar 13. Pola suara orang keempat yang telah dinormalisasi



Gambar 14. Pola suara orang kelima



Gambar 15. Pola suara orang kelima yang telah dinormalisasi

Tabel 1. Tabel Normalisasi Sampel Suara

Aktor 1

Ciri	N1	N2	N3	N4	N5	Mean Normalisasi
MFCC1	3,280	3,235	3,288	3,262	3,259	3,265
MFCC2	-0,462	-0,247	-0,536	-0,297	-0,522	-0,413
MFCC3	-0,285	-0,627	-0,196	-0,219	-0,418	-0,349
MFCC4	-0,424	-0,325	-0,228	-0,215	-0,074	-0,253
MFCC5	-0,392	-0,288	-0,436	-0,478	-0,224	-0,364
MFCC6	-0,079	-0,078	-0,439	-0,386	-0,505	-0,297
MFCC7	-0,587	0,144	-0,245	-0,341	-0,266	-0,259
MFCC8	-0,180	0,009	-0,278	0,279	-0,366	-0,107

MFCC9	-0,079	-0,553	-0,125	-0,237	-0,395	-0,278
MFCC10	-0,282	-0,542	-0,158	-0,404	-0,353	-0,348
MFCC11	-0,004	-0,212	-0,419	-0,262	0,127	-0,154
MFCC12	-0,157	-0,467	-0,266	-0,159	-0,340	-0,278
MFCC13	-0,349	-0,050	0,038	-0,542	0,075	-0,166

Aktor 2

Ciri	N1	N2	N3	N4	N5	Mean Normalisasi
MFCC1	3,264	3,280	3,241	3,272	3,246	3,261
MFCC2	-0,540	-0,477	-0,403	-0,227	-0,363	-0,402
MFCC3	-0,306	-0,150	-0,293	-0,396	-0,300	-0,289
MFCC4	-0,174	-0,244	-0,305	-0,419	-0,233	-0,275
MFCC5	-0,417	-0,163	-0,574	-0,363	-0,437	-0,391
MFCC6	0,044	-0,436	-0,439	-0,190	-0,189	-0,242
MFCC7	0,055	-0,243	-0,247	-0,659	-0,382	-0,295
MFCC8	-0,316	-0,310	-0,661	-0,407	-0,338	-0,406
MFCC9	-0,332	-0,457	-0,289	-0,012	-0,665	-0,351
MFCC10	-0,190	-0,272	-0,182	-0,278	-0,131	-0,211
MFCC11	-0,408	-0,409	-0,023	-0,025	0,322	-0,109
MFCC12	-0,105	-0,286	0,086	-0,051	-0,273	-0,126
MFCC13	-0,576	0,167	0,089	-0,245	-0,254	-0,164

Aktor 3

Ciri	N1	N2	N3	N4	N5	Mean Normalisasi
MFCC1	3,291	3,234	3,279	3,270	3,240	3,263
MFCC2	-0,325	-0,094	-0,367	-0,137	-0,389	-0,262
MFCC3	-0,242	-0,168	-0,376	0,090	-0,341	-0,207
MFCC4	-0,268	-0,263	-0,247	-0,168	0,037	-0,182
MFCC5	-0,477	-0,159	-0,126	-0,183	-0,232	-0,236
MFCC6	-0,133	-0,821	-0,350	-0,228	0,122	-0,282
MFCC7	-0,084	-0,516	0,020	-0,611	-0,016	-0,241
MFCC8	-0,119	-0,291	-0,089	-0,374	-0,577	-0,290
MFCC9	-0,204	0,185	-0,320	-0,115	-0,516	-0,194
MFCC10	-0,460	-0,376	-0,454	-0,228	-0,093	-0,322
MFCC11	-0,567	-0,246	-0,601	-0,412	-0,560	-0,477
MFCC12	-0,150	-0,115	-0,307	-0,526	-0,431	-0,306
MFCC13	-0,262	-0,370	-0,062	-0,377	-0,245	-0,263

Aktor 4

Ciri	N1	N2	N3	N4	N5	Mean Normalisasi
MFCC1	3,270	3,284	3,231	3,220	3,276	3,256
MFCC2	-0,414	-0,582	-0,418	-0,492	-0,177	-0,417
MFCC3	-0,455	-0,329	-0,199	-0,291	-0,514	-0,358
MFCC4	-0,576	-0,338	-0,604	-0,372	-0,484	-0,475
MFCC5	-0,304	-0,318	-0,137	-0,062	-0,038	-0,172
MFCC6	0,101	-0,511	-0,436	-0,273	-0,374	-0,298
MFCC7	-0,050	-0,047	-0,432	-0,363	-0,244	-0,227
MFCC8	-0,318	-0,327	-0,130	0,092	-0,127	-0,162
MFCC9	-0,221	-0,080	-0,738	-0,175	-0,121	-0,267
MFCC10	-0,113	-0,153	0,075	-0,521	-0,196	-0,182
MFCC11	-0,375	-0,059	-0,130	-0,794	-0,151	-0,302
MFCC12	-0,416	-0,275	-0,063	0,105	-0,225	-0,175
MFCC13	-0,130	-0,266	-0,020	-0,074	-0,626	-0,223

Aktor 5

Ciri	N1	N2	N3	N4	N5	Mean Normalisasi
MFCC1	3,244	3,247	3,275	3,252	3,136	3,231
MFCC2	-0,245	-0,708	-0,476	-0,609	-0,880	-0,584
MFCC3	-0,268	-0,406	-0,104	-0,202	-0,523	-0,301
MFCC4	-0,193	-0,022	-0,111	-0,071	0,018	-0,076
MFCC5	-0,120	-0,545	-0,472	-0,424	-0,273	-0,367
MFCC6	-0,382	-0,500	-0,509	-0,567	-0,816	-0,555
MFCC7	-0,395	-0,194	-0,440	-0,457	-0,002	-0,298
MFCC8	0,121	-0,043	-0,239	-0,408	0,172	-0,079
MFCC9	-0,735	-0,201	-0,332	0,038	-0,380	-0,322
MFCC10	-0,496	-0,343	-0,333	-0,175	-0,123	-0,294
MFCC11	-0,390	-0,069	-0,260	0,057	-0,023	-0,137
MFCC12	-0,192	-0,194	-0,005	-0,257	-0,415	-0,213
MFCC13	0,052	-0,021	0,007	-0,175	0,107	-0,006

4. SIMPULAN

Hasil dari penelitian yang telah dilakukan dapat disimpulkan bahwa ekstraksi fitur dengan menggunakan MFCC dapat dibedakan polanya untuk setiap suara dengan beberapa parameter yaitu $T_w = 25$; sebagai durasi masing-masing *frame*, $T_s = 10$; sebagai pembentukan *frame*, $\alpha = 0.97$; sebagai nilai koefisien *preemphasis*, $M = 20$; sebagai nilai *filterbank* pada setiap *channels*, $C = 12$; nilai *cepstral coefficients*, dari parameter tersebut dihasilkan fitur MFCC, hasil dari matriks MFCC tersebut kemudian dicari nilai rata-ratanya (*mean*) untuk cirinya. Hasil dari ciri *mean* tersebut menghasilkan pola-pola yang menjadikan ciri masing-masing suara. Setiap suara memiliki pola dan besaran spektrum yang berbeda-beda. Perbedaan spektrum dikarenakan pengaruh kualitas suara. Dengan identifikasi suara menggunakan MFCC ekstraksi ciri ini dapat digunakan untuk proses pengklasifikasian suara selanjutnya.

DAFTAR PUSTAKA

- [1] 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
- [2] Pengendali, S. and Elektronik, P. (2016) 'Aplikasi Pengenalan Suara Menggunakan Microsoft Sapi sebagai Pengendali Peralatan Elektronik', (January 2007).
- [3] Bhaskoro, S. B., Pelatihan, F. and Pengenalan, F. (2012) 'Aplikasi pengenalan gender menggunakan suara', 2012(Snati), pp. 15-16.
- [4] Nim, G. M., & Uta, S. (2008). Pencocokan Pola Suara (Speech Recognition) Dengan Algoritma FFT Dan Divide And Conquer
- [5] Fadlisyah. Bustami.M.Ikhwanus. (2013) Pengolahan Suara.Graha Ilmu.

- [6] Prodi, S. et al. (2017) 'Perancangan Sistem Pengenalan Suara Dengan Metode Linear Predictive Coding Speech Recognition System Design Using Linear Predictive Coding Method', 4(1), pp. 404–411.
- [7] Setiawan, A., Hidayatno, A. and Isnanto, R. R. (2011) 'Aplikasi Pengenalan Ucapan dengan Ekstraksi Mel-Frequency Cepstrum Coefficients (MFCC) Melalui Jaringan Syaraf Tiruan (JST) Learning Vector Quantization (LVQ) untuk Mengoperasikan Cursor Komputer', 13(3), pp. 82–86.
- [8] Wijaya, R. F. and Utomo, R. B. (2018) 'Kamus Terjemahan Bahasa Indonesia Ke Bahasa Inggris Dengan Penerapan Speech To Text Berbasis', 5(1), pp. 44–47.
- [9] Fadlil, A. et al. (2008) 'Pengembangan Sistem Basis Data Presensi Perkuliahan Dengan Kartu Mahasiswa Ber-Barcode', Telkomnika, 6(1), pp. 65–72.
- [10] Yudhana, A., Fadlil, A. and Prianto, Eko Teori, L. (2017) 'Kepegawaian Dan Kegiatan Berbasis Android', pp. 63–68.
- [11] Mega Tiara Nur Azizah (2017) 'Aplikasi Pengenal Pengucap Berbasis Identifikasi Suara Dengan Ekstraksi Ciri Mel-Frequency Cepstrum Coefficients (MFCC) Dan Kuantisasi Vekto', Transient, 6.