EKSTRAKSI CIRI EMOSI MANUSIA BERDASARKAN UCAPAN MENGGUNAKAN *MEL-FREQUENCY CEPSTRAL COEFFICIENTS* (MFCC)

Siti Helmiyah[1*], Abdul Fadlil[2] dan Anton Yudhana[2]
[1]Magister Teknik Informatika, Pascasarjana Teknik Informatika
[2]Program Studi Teknik Elektro, Fakultas Teknologi Industri
Universitas Ahmad Dahlan Yogyakarta, Indonesia.
[*]Email : siti1708048022@webmail.uad.ac.id

Abstrak

Emosi merupakan perilaku manusia yang dapat diungkapkan dengan tingkah laku berupa raut wajah dan suara. Suara adalah suatu gelombang longitudinal yang merambat di udara. Pada kehidupan sehari-hari manusia berkomunikasi dengan menggunakan suara. Baik itu di dunia pendidikan, kantor, dan tempat umum lainnya. Ketika berkomunikasi seringkali seseorang tidak sadar dengan emosi yang sedang dirasakannya. Pengenalan emosi manusia berdasarkan suara ini merupakan permasalahan yang sulit untuk dipecahkan. Hal ini karena penilaian terhadap emosi sulit dikenali oleh peneliti secara objektif. Maka dari itu penelitian ini dilakukan bertujuan untuk mengetahui pengenalan dan klasifikasi emosi berdasarkan suara. Data diambil dari database Berlin Database of Emotional Speech. Data tersebut menggunakan suara aktor yang sudah dibedakan jenis kelamin, umur, emosinya serta kalimatnya. Ekstraksi ciri yang digunakan untuk pengenalan suara adalah Mel-Frequency Cepstral Coefficients (MFCC). Hasil pola emosi dari ucapan dengan frame bloking = 4, pre-emphasis dengan alpha =0.95, filterbank = 20, dan koefesien cepstral = 13 menunjukkan bahwa dengan kalimat yang sama, aktor, jenis kelamin, suara, dan umur yang berbeda tidak begitu terlihat banyak perbedaan polanya. Maka perlu dilakukan proses prepocessing untuk setiap data. Tapi tidak menutup kemungkinan hasil dari pola ini dapat dijadikan sebagai bahan untuk proses klasifikasi selanjutnya.

Kata kunci : *Ekstraksi Ciri, Emosi, Mel-Frequency Cepstral Coefficients (MFCC), Ucapan*

1. PENDAHULUAN

Emosi merupakan keadaan yang dirasakan pada setiap individu dalam intensitas yang tinggi terhadap sesuatu hal (N. H. Frieda, 1993). Emosi juga bisa disebut sebagai reaksi akibat timbal balik atas tindakan seseorang ataupun kejadian yang dialami pemilik emosi. Seringkali emosi mengakibatkan perubahan perilaku yang berakibat terganggunya hubungan dengan lingkungan. Emosi dapat dikategorikan menjadi emosi positif dan negatif dalam jenisnya. Beberapa kategori emosi positif adalah senang, kepedulian, ketertarikan, antusias, kebosanan dan keingintahuan. Beberapa kategori emosi negatif adalah marah, sedih, takut, iri dan kebencian.

Beberapa penelitian yang sudah dilakukan mengindikasikan bahwa ada beberapa parameter yang menunjukkan adanya hubungan yang kuat antara ucapan dengan emosi yang sedang dirasakan (B. Heuft, 1996). Parameter tersebut adalah pitch, energi, artikulasi dan bentuk spektral. Emosi sedih identik dengan kecepatan ucapan yang lambat dalam pitch rendah, sedangkan emosi marah identik dengan kecepatan ucapan dan pitch yang tinggi (A. Negueiras, 2001). Pengenalan emosi manusia melalui ucapan bukan hal yang patut dikesampingkan. Banyak penelitian yang membuktikan adanya kebutuhan pengenalan emosi dalam interaksi komputer dan manusia (Al-Talabani dkk., 2015).

Penelitian ini dilakukan untuk mengetahui pola emosi dari ekstraksi ciri ucapan manusia karena dilakukan untuk mengetahui ciri polanya (Gustina, dkk 2016). Ekstraksi ciri yang digunakan pada penelitian ini adalah *Mel-Frequency Cepstral Coefficients* (MFCC) karena merupakan metode ekstraksi ciri yang mendekati sistem pendengaran manusia (Riyanto and Sutejo, 2014). Harapannya adalah hasil dari pola emosi dapat digunakan untuk pemprosesan klasifikasi selanjutnya yaitu untuk pengenalan emosi manusia berdasarkan ucapan (Surya,dkk 2017).

1.1. Ekstraksi Cri Menggunakan MFCC

MFCC adalah ekstraksi ciri yang sering digunakan pada pemrosesan suara, karena dapat merepresentasikan sinyal dengan baik. MFCC memiliki cara kerja yang didasarkan pada perbedaan frekuensi yang sesuai dengan pendengaran manusia sehingga dapat merepresentasikan sinyal suara

seperti manusia merepresentasikannya. Proses ekstraksi ciri MFCC (Azizah dkk, 1996) adalah sebagai berikut:

a. Pre-emphasis

Tujuan dari pre-emphasis ini adalah untuk memfilter sinyal ucapan yang dengan mengurangi nilai frekuensi sinyal tersebut sehingga hanya sinyal memiliki frekuensi tinggi yang dapat melewati proses filter dan dapat mengurangi noise pada input suara sehingga hanya data sinyal wicara saja yang dapat ditangkap sistem. Persamaannya (Sanjaya dan Salleh, 2014) dapat dilihat pada rumus (1).

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

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

b. Framing

Pada proses ini input suara dipotong menjadi frame-frame dengan durasi yang lebih pendek sebanyak matriks (M) yang disimpan di matriks Y dengan ukuran MxW. 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.

c. 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,….T yang disebut frame. Pada kasus ini akan menggunakan persamaan hamming window dengan rumus (2).

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

dimana *n* adalah jumlah sampel dan *N* adalah jumlah frame

d. Fast Fourier Transform (FFT)

FFT adalah salah satu metode untuk mengkonversikan dari sinyal suara menjadi sinyal frekuensi (anton). 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 (3).

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

dimana w adalah windowing.

e. 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 (4).

$$B(f) = 1125\, x\, ln\,(1 + \frac{f}{700}) \quad (4)$$

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.

f. Discrete Consine Trasform

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

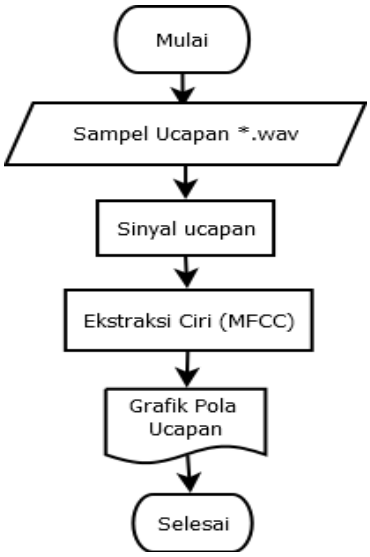
$$C_j = \sum_{i=1}^{M} X_i \cos\left(j(i-1)/2\frac{\pi}{M}\right) \quad (5)$$

dimana j = 1, 2, 3... K adalah koefesien, dan M adalah jumlah filter.

ISBN 978-602-99334-9-9

2. METODOLOGI

Penelitian ini dilakukan untuk mengatahui hasil dari ekstraksi ciri MFCC dari ucapan emosi manusia. Penelitian ini mengunakan *software Matlab R2013a* dengan menggunakan sampel dari ucapan aktor di *database Berlin Database of Emotional*. Bahasa yang digunakan adalah bahasa Berlin. Metode MFCC dalam penelitian ini digunakan untuk mengekstraksi ciri sampel ucapan dari aktor. Berikut alur tahapan penelitan dapat dilihat pada Gambar 2.



Gambar 1. Alur Tahapan Penelitian

Berikut adalah langkah-langkah tahapan penelitian:

a. Sampel ucapan *.wav, tahapan ini adalah mengambil beberapa sampel dari *database Berlin Database of Emotional* sebagai masukkan.
b. Sinyal ucapan, tahapan ini adalah mengubah sampel ucapan bentuk file .wav menjadi sinyal ucapan yang dapat dibaca oleh komputer.
c. Ektraksi ciri menggunakan MFCC adalah proses yang dilakukan untuk menghasilkan pola emosi dari sinyal ucapan.
d. Grafik pola ucapan, adalah hasil dari ekstraksi ciri MFCC mengetahui pola emosi yang ditampilkan dalam bentuk grafik plot.

3. HASIL DAN PEMBAHASAN

3.1. Sampel Ucapan

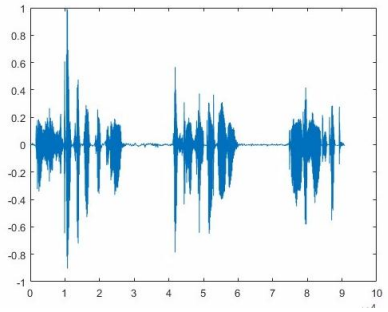
Sampel ucapan diambil dari *database Berlin Database of Emotional* berupa emosi marah, senang, sedih, bosan, dan cemas/takut. Bahasa yang digunakan adalah bahasa Berlin dengan kalimat yang sama dan aktor yang berbeda.

3.2. Sinyal Ucapan

Sampel ucapan dalam bentuk file *.wav sebagai *inputan* direpresentasikan menjadi sinyal ucapan dalam bentuk matriks dengan cara memberikan perintah *audioread* di *Matlab* seperti berikut pada

```
audioread( wav_file );
```

Hasil dari sampel ucapan menjadi sinyal suara dari perintah di atas dapat dilihat pada Gambar 2.



Gambar 2. Representasi .wav dalam grafik

Gambar 2 adalah hasil dari salah satu sampel suara yang direpresentasika menjadi sinyal ucapan dan di tampilkan bentuk grafik.

3.3. Ektrasksi Ciri menggunakan MFCC

Ekstraksi ciri MFCC memiliki beberapa tahapan untuk itu maka perlu dibuat dan diatur variabel-variabelnya dan *function* untuk menghasilkan pola emosi yang dapat dilihat seperti berikut:

```
Tw = 25;           % analysis frame duration (ms)
Ts = 1000;         % analysis frame shift (ms)
alpha = 0.95;      % preemphasis coefficient
M = 20;            % number of filterbank channels
C = 13;            % number of cepstral coefficients
L = 22;            % cepstral sine lifter parameter
LF = 300;          % lower frequency limit (Hz)
HF = 3700;         % upper frequency limit (Hz)
wav_file = 'suara/Sad.wav';  % input audio filename
```

Kemudian masukkan perintah berikut di *Matlab*

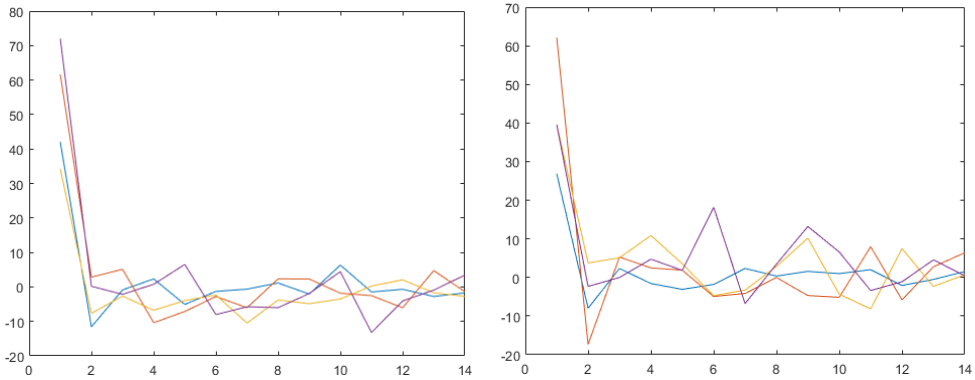
```
% Feature extraction (feature vectors as columns)
[ MFCCs, FBEs, frames ] = ...
  mfcc( speech, fs, Tw, Ts, alpha, @hamming, [LF HF], M, C+1, L );
% Generate data needed for plotting
[ Nw, NF ] = size( frames );                % frame length and number of
frames
time_frames = [0:NF-1]*Ts*0.001+0.5*Nw/fs;  % time vector (s) for frames
time = [ 0:length(speech)-1 ]/fs;           % time vector (s) for signal
samples
logFBEs = 20*log10( FBEs );                 % compute log FBEs for plotting
logFBEs_floor = max(logFBEs(:))-50;         % get logFBE floor 50 dB below
max
logFBEs( logFBEs<logFBEs_floor ) = logFBEs_floor; % limit logFBE dynamic
range
```

Perintah diatas digunakan untuk mencari pola pada masing-masing emosi ucapan.

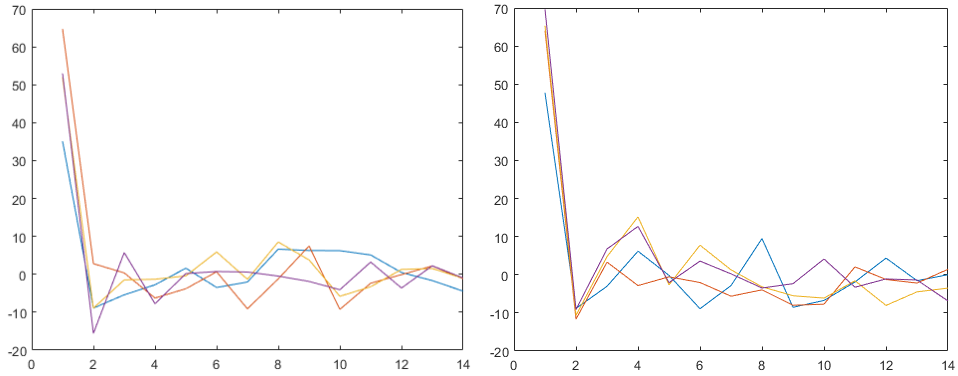
3.4. Hasil Pola Emosi dalam Bentuk Grafik

Setelah proses ekstraksi ciri selesai dilakukan pada tiap-tiap ucapan dan emossi yang berbeda maka menghasilkan pola emosi yang dapat dilihat pada Gambar 3-7.

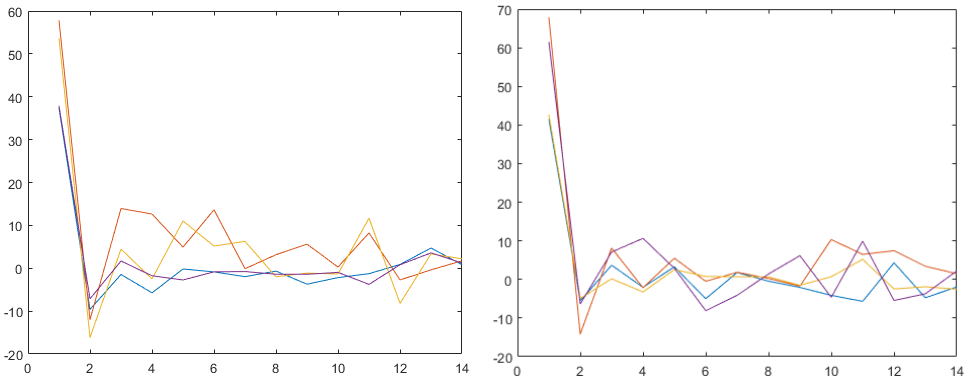
ISBN 978-602-99334-9-9



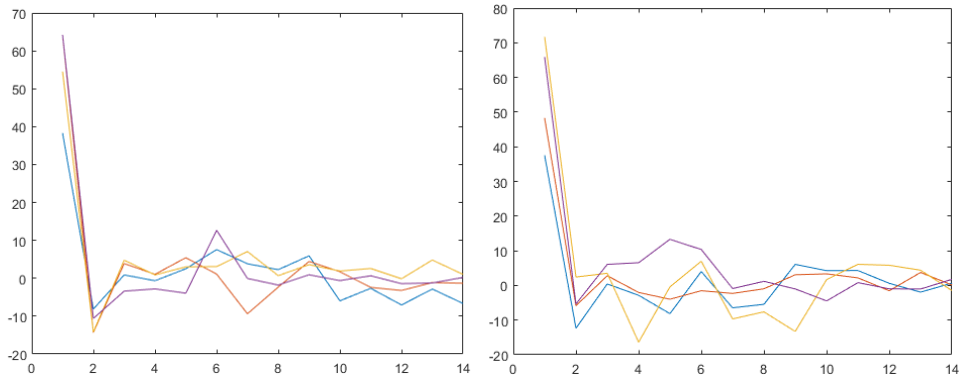
Gambar 3. Grafik Emosi Marah



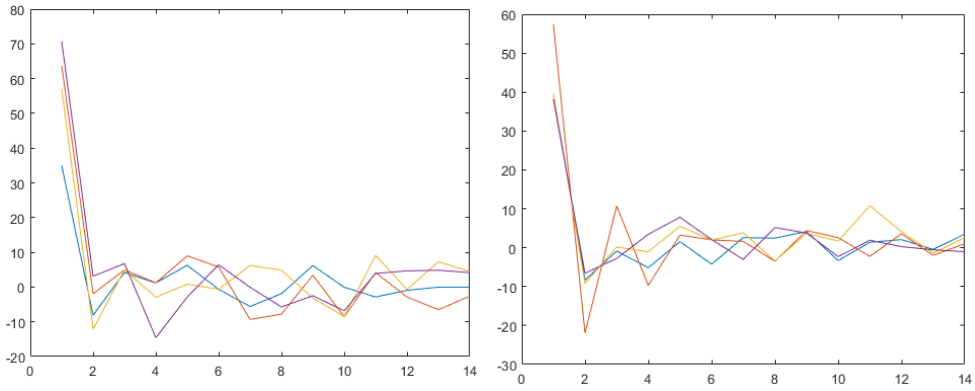
Gambar 4. Grafik Emosi Senang



Gambar 5. Grafik Emosi Sedih



Gambar 6. Grafik Emosi Bosan



Gambar 7. Grafik Emosi Cemas/Takut

Gambar 3 sampai 7 dengan frame bloking = 4, pre-emphasis dengan alpha =0.95, filterbank = 20, dan koefesien cepstral = 13 menghasilkan pola emosi yang tidak begitu berbeda. Maka perlu dilakukan prepocessing untuk setiap data dan perlu bisa juga dilakukan klasifikasi untuk mengenali emosi pada ucapan manusia.

4. KESIMPULAN

Kesimpulan dari penelitian ini menghasilkan pola emosi dari proses ekstraksi ciri menggunakan metode MFCC. Hasil pola emosi dari ucapan dengan frame bloking = 4, pre-emphasis dengan alpha =0.95, filterbank = 20, dan koefesien cepstral = 13 menunjukkan bahwa dengan kalimat yang sama, aktor, jenis kelamin, suara, dan umur yang berbeda tidak begitu terlihat banyak perbedaan polanya. Maka perlu dilakukan proses prepocessing untuk setiap data. Tapi tidak menutup kemungkinan hasil dari pola ini dapat dijadikan sebagai bahan untuk proses klasifikasi selanjutnya.

DAFTAR PUSTAKA

Al-Talabani, A., Sellahewa, H. & Jassim, S. A., 2015. Emotion Recognition from Speech: Tools and Challenges. *Mobile Multimedia/Image Processing, Security, and Applications*, 9497(Mobile Multimedia/Image Processing, Security, and Applications), p. 94970N.

Azizah, R. S., Nurjanah, D. & Sari, F. D., 2015. *Sistem Automatic Speech Recognition Menggunakan Metode MFCC dan HMMs Untuk Deteksi Kesalahan Pengucapan Kata Bahasa Inggris.* Bandung, eProceedings of Engineering.

Firdausy, K. & Yudhana, A., 2006. Aplikasi Watemarking Untuk Perlindungan Haki Pada Citra Digital Dalam Domain Frekuensi Menggunakan Fast Fourier Transform (FFT). *Telkomnika (Telecommunication Computing Electronics And Control),*, Volume 4 (2), pp. 123-126..

Fridja, N. H., 1993. *Moods, emotion episodes, and emotions*, New York: Guilford Press.

Heuft, B., Portele, T. & Rauth, M., 1996. Emotions in Time Domain SynthesisSpoken Language. *ICLS 96 Proceedings Fourth International Conference on*, Volume 3, pp. 1974-1977.

Nogueiras, A., Moreno, A., Bonafonte, A. & Marino, J. B., 2001. Speech Emotion Recognition Using Hidden Markov Models. *Seventh European Conference on Speech Communication and Technology*, 2nd(Eurospeech), pp. 2679-2682.

Riyanto, E. & Sutejo, 2014. Perbandingan Metode Ekstraksi Ciri Suara MFFCC, ZCPA, dan LPC. Volume 10.

Sanjaya, M. & Salleh, Z., 2014. Implementasi Pengenalan Pola Suara Menggunakan Mel-Frequency Cepstrum Coefficients (MFCC) dan Adaptive Neuro-Fuzzy Inferense System (ANFIS) sebagai Kontrol Lampu Otomatis. *ALHAZEN Jurnal of Physics*, Volume 1(1), pp. 43-54.

Surya, R. A. A. F. a. A. Y., 2017. Ekstraksi Ciri Metode Gray Level Co-Occurrence Matrix (GLCM) dan Filter Gabor untuk Klasifikasi citra Batik Pekalongan. *JURNAL INFORMATIKA: Jurnal Pengembangan IT*, Volume 2.2, pp. 23-26.

ISBN 978-602-99334-9-9