

Pengenalan Pola Emosi Manusia Berdasarkan Ucapan Menggunakan Ekstraksi Fitur Mel-Frequency Cepstral Coefficients (MFCC)

Speech Based Emotion Pattern Recognition Using Mel-Frequency Cepstral Coefficients (MFCC) Feature Extraction

Siti Helmiyah^{*1}, Abdul Fadlil², Anton Yudhana³

¹Magister Teknik Informatika, Universitas Ahmad Dahlan, Yogyakarta

^{2,3}Program Studi Teknik Elektro, Fakultas Teknologi Industri, Universitas Ahmad Dahlan, Yogyakarta

e-mail: *1siti1708048022@webmail.uad.ac.id, 2fadlil@mti.uad.ac.id, 3eyudhana@ee.uad.ac.id

Abstrak

Pengenalan emosi manusia berdasarkan ucapan menjadi penting karena permasalahan ini sangat berguna dalam kehidupan sekarang yang memerlukan interaksi manusia dan teknologi. Pengenalan emosi merupakan masalah yang kompleks karena perbedaan adat dan dialek khusus yang ada pada etnis, wilayah dan komunitas yang berbeda. Masalah ini juga menjadi sulit karena penilaian secara objektif, emosi merupakan peristiwa yang terjadi di bawah alam sadar manusia. Penelitian ini dilakukan bertujuan untuk mengetahui pola pengenalan emosi berdasarkan ekstraksi fitur dari ucapan. Metode ekstraksi fitur yang digunakan pada penelitian ini adalah Mel-Frequency Cepstral Coefficients (MFCC) yang merupakan metode ekstraksi fitur yang mendekati sistem pendengaran manusia. Data yang digunakan diambil dari database Berlin Database of Emotional Speech (Emo-DB). Emosi yang digunakan dalam penelitian ini adalah bahagia, bosan, netral, sedih, dan marah. Masing-masing emosi, menggunakan tiga sample dari Emo-DB. Pola emosi ucapan berhasil terlihat dengan menggunakan nilai parameter MFCC seperti 25 untuk durasi frame, 10 untuk pembentukan frame, 0.97 untuk nilai koefisien preemphasis, 20 untuk nilai channels filterbank, 13 untuk cepstral coefficients. Fitur MFCC kemudian diekstraksi dan dihitung untuk menemukan nilai rata-rata dari parameter ini. Nilai-nilai ini kemudian diplot berdasarkan grafik kerangka waktu untuk menemukan pola spesifik yang muncul dari setiap emosi.

Kata kunci— Emosi, Ucapan, Mel-Frequency Cepstral Coefficients (MFCC).

Abstract

Human emotion recognition subject becomes important due to its usability in daily lifestyle which requires human and computer interaction. Human emotion recognition is a complex problem due to the difference within custom tradition and specific dialect which exists on different ethnic, region and community. This problem also exacerbated due to objectivity assessment for the emotion is difficult since emotion happens unconsciously. This research conducts an experiment to discover pattern of emotion based on feature extracted from speech. Method used for feature extraction on this experiment is Mel-Frequency Cepstral Coefficient (MFCC) which is a method that similar to the human hearing system. Dataset used on this experiment is Berlin Database of Emotional Speech (Emo-DB). Emotions that are used for this experiments are happiness, boredom, neutral, sad and anger. For each of these emotion, 3

samples from Emo-DB are taken as experimental subject. The emotion patterns are successfully visible using specific values for MFCC parameters such as 25 for frame duration, 10 for frame shift, 0.97 for preemphasis coefficient, 20 for filterbank channel and 12 for cepstral coefficients. MFCC features are then extracted and calculated to find mean values from these parameters. These mean values are then plotted based on timeframe graph to be investigated to find the specific pattern which appears from each emotion.

Keywords— Emotion, Speech, Mel-Frequency Cepstral Coefficients (MFCC).

1. PENDAHULUAN

Emosi adalah suatu perasaan yang dirasakan setiap individu dalam intensitas yang tinggi terhadap seseorang atau sesuatu keadaan. Emosi disebut juga sebagai reaksi timbal balik atas kejadian maupun tindakan dialami seseorang. Emosi seringkali mengakibatkan perubahan perilaku seseorang yang mengakibatkan terganggunya hubungan dengan lingkungan. Pengenalan emosi manusia adalah permasalahan yang sangat berguna dalam hal-hal yang memerlukan interaksi manusia dan teknologi. Misalnya, seorang customer care dapat memberikan pendekatan yang tepat terhadap emosi yang dirasakan pelanggan dengan mendeteksi emosi dari ucapan pelanggan. Sebuah sistem informasi dapat menyajikan informasi dan keterangan sesuai dengan mood yang dirasakan pengguna sistem dari cara pengguna tersebut mencari informasi, dan lain lain. Di sisi lain, pengenalan emosi merupakan masalah yang kompleks. Hal ini dikarenakan perbedaan adat dan dialek di setiap tempat dan penilaian terhadap emosi yang sulit dilakukan secara subjektif karena emosi merupakan peristiwa yang terjadi di bawah alam sadar manusia.

Pengenalan emosi melalui ucapan bukanlah hal baru dalam dunia teknologi. Beberapa penelitian telah dilakukan menggunakan sistem cerdas maupun cara konvensional. Penelitian-penelitian tersebut menggunakan dataset ucapan yang tersedia sebagai standar seberapa baik sebuah sistem melakukan pengenalan emosi. Banyak database emosi yang disediakan oleh beberapa lembaga diluar negeri seperti SUSAS (Speech Under Simulated and Actual Stress), Berlin database of emotional speech (Emo-DB), Surrey Audio-Visual Expressed Emotion (SAVEE) Database, The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) dan masih banyak lagi.

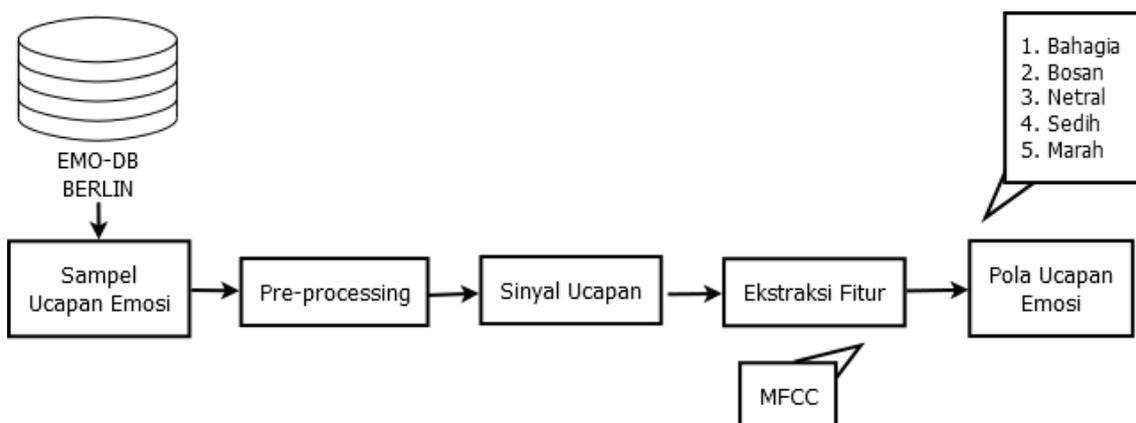
Beberapa penelitian yang sudah dilakukan mengindikasikan bahwa ada beberapa parameter yang menunjukkan adanya hubungan yang kuat antara ucapan dengan emosi yang sedang dirasakan. Parameter tersebut adalah pitch, gelombang, artikulasi dan bentuk spektral. Pada penelitian yang lain, emosi sedih identik dengan kecepatan ucapan yang lambat dalam pitch rendah, sedangkan emosi marah identik dengan kecepatan ucapan dan pitch yang tinggi [1]. Adapun beberapa penelitian yang dilakukan untuk mengenali emosi diantaranya, penelitian untuk pengenalan emosi melalui ucapan yang didasarkan dengan objek pengujian Emo-DB [2] menggunakan fitur energi, entropi, MFCC (Mel-Frequency Cepstral Coefficients), ZCC, dan pitch dengan algoritma K-NN (K-Nearest Neighbor) sebagai pengklasifikasi mencapai akurasi 86.02%. Kemudian penelitian serupa juga dilakukan [3] untuk mengembangkan sistem pengenalan emosi menggunakan beberapa metode dan mencapai akurasi tertinggi sebesar 92% menggunakan GMM (Gaussian Mixture Model), dan 72% menggunakan K-NN untuk emosi marah. Sedangkan untuk emosi lain selain marah, akurasi yang didapat masih dibawah 80%. Penelitian [1] melakukan pengenalan emosi menggunakan algoritma HMM (Hidden Markov Model) dengan fitur MFCC. Sistem ini memiliki akurasi rata-rata 86.66%. Banyak penelitian yang membuktikan adanya kebutuhan pengenalan emosi dalam interaksi komputer dan manusia [4].

Penelitian ini dilakukan untuk mencari pengenalan pola emosi berdasarkan ucapan dengan melakukan ekstraksi fitur ucapan manusia untuk mengetahui ciri polanya [5]. Ekstraksi ciri yang digunakan pada penelitian ini adalah *Mel-Frequency Cepstral Coefficients* (MFCC)

karena merupakan metode ekstraksi fitur yang mendekati sistem pendengaran manusia. Harapannya adalah hasil dari pengenalan pola emosi ini sistem dapat melakukan pengenalan emosi manusia berdasarkan ucapan [6].

2. METODE PENELITIAN

Proses penelitian yang dilakukan untuk pengenalan emosi berdasarkan ucapan untuk teknik pengenalan polanya secara umum dapat dilihat pada Gambar 1 yang terdiri dari beberapa tahapan seperti berikut:



Gambar 1 Alur Pengenalan Emosi

2.1. Berlin database of emotional speech (Emo-DB Berlin)

Data yang digunakan dalam penelitian ini menggunakan Berlin Emotion Speech Database yaitu basis data yang paling sering digunakan dalam komunitas pemrosesan ucapan [3]. Basis data ini mudah didapatkan, bebas digunakan dan banyak yang menggunakannya sebagai penelitian [7]. Emo-DB adalah salah satu basis data emosi ucapan yang direkam oleh Technical University of Berlin in Germany yang didanai oleh German Research Community. Emosi ucapan ini direkam menggunakan mikrophone Sennheiser pada frekuensi 16 kHz, yang disuarakan oleh sepuluh aktor profesional yaitu lima orang laki-laki dan lima orang perempuan. Kesepuluh aktor tersebut diminta untuk mensimulasikan tujuh emosi yaitu marah, bosan, jijik, takut, bahagia, sedih dan netral. Mereka mensimulasikannya dengan menggunakan sepuluh kalimat, diantaranya lima kalimat pendek dan lima kalimat yang lebih panjang. Kalimat ini merupakan kalimat netral yang dapat digunakan dalam komunikasi sehari-hari, dan juga dapat dikatakan dengan semua emosi.

Sekitar 816 ucapan yang disimpan. Kemudian ucapan tersebut akan dinilai oleh dua puluh hakim untuk mendengarkan ucapan dalam urutan acak, di depan monitor komputer. Mereka mendengarkan masing-masing sampel hanya sekali, sebelum mereka memutuskan keadaan emosi yang dimiliki pembicara. Setelah pemilihan, database berisi total 535 file ucapan [8]. EMO-DB adalah database yang tidak seimbang karena semua emosi tidak memiliki jumlah sampel rekaman yang sama, jumlah sampel terbanyak adalah emosi kemarahan (127) dan jumlah emosi terkecil adalah emosi jijik (46). Semua informasi mengenai basis data ini dapat diakses melalui internet.

2.2. Sampel Ucapan Emosi

Penelitian ini menggunakan tiga data ucapan untuk masing-masing sampel emosi yaitu emosi bahagia, bosan, netral, sedih dan marah, jadi total keseluruhan ada 15 data. Sampel ucapan emosi yang diambil adalah sampel ucapan dari kalimat “Das will sie am Mittwoch abgeben.”. Kalimat tersebut diucapkan oleh suara aktor yang berbeda-beda.

2.3. Pre-processing

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

2.4. Sinyal Ucapan

Setelah tahapan pemotongan suara sebagai sampling, tahapan selanjutnya adalah mengubah rekaman suara menjadi sinyal ucapan. Tahapan ini mengubah sampel ucapan dalam bentuk file .wav menggunakan software Matlab. Perubahan ini berguna untuk tahapan ekstraksi fitur.

2.5. Ekstraksi Fitur

Tujuan dari ekstraksi fitur adalah untuk menajamkan perbedaan-perbedaan pola sehingga dapat memudahkan dalam pemisahan emosi pada proses klasifikasi [9]. Ekstraksi fitur yang digunakan pada penelitian ini menggunakan metode *Mel-Frequency Cepstral Coefficients* (MFCC) yang dimana metode ekstraksi fitur ini merupakan metode yang mendekati sistem pendengaran manusia. Mel-Frequency Cepstral Coefficients (MFCC) sering digunakan untuk ekstraksi fitur dalam pengolahan ucapan. Teknik penggunaan skala Mel di derivasi koefisien cepstrum diperkenalkan. Tujuan utama ekstraksi fitur adalah untuk mengekstrak karakteristik penting dari sinyal ucapan yang unik untuk setiap kata untuk membedakan antara serangkaian kata-kata yang berbeda [10]. MFCC dianggap sebagai metode standar untuk ekstraksi fitur dalam pengenalan suara. MFCC sering digunakan untuk pemrosesan suara, karena dapat merepresentasikan sinyal dengan baik. Proses ekstraksi fiturnya [11] adalah sebagai berikut:

2.5.1. Pre-emphasis

Proses pertama yaitu pre-emphasi, tujuan dari proses ini adalah untuk mem-filter sinyal ucapan untuk 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 wicara saja yang dapat ditangkap sistem. Persamaannya [12] dapat dilihat pada rumus (1).

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

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

2.5.2. 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 $M \times W$. 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 frame atau bahkan lebih dari satu frame.

2.5.3. Windowing

Proses windowing ini dilakukan bertujuan 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 penelitian ini teknik windowing yang digunakan adalah teknik hamming windowing, persamaan dari teknik ini dapat dilihat pada 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

2.5.4. Fast Fourier Transform (FFT)

Fast Fourier Transform (FFT) adalah salah satu metode pengkonversian dari sinyal suara menjadi sinyal frekuensi [13]. Pada proses FFT ini dilakukan terhadap semua frame dari sinyal yang sudah di windowing. FFT adalah salah satu algoritma yang cepat untuk menerapkan Discrete Fourier Transform (DCT) yang beroperasi pada sinyal diskrit. Sinyal diskrit ini terdiri dari N sampel, dengan 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.

2.5.5. Mel Filterbank

Proses selanjutnya yaitu, mel-filterbank sama dengan triangular dari filterbank, yang membedakan adalah dari range frekuensi linier hasil FFT kemudian dikonversi ke skala Mel-Frequency untuk mendapatkan batas-batas filterbank, rumus mel-filterbank dapat dilihat pada rumus (4).

$$B(f) = 1125 \times \ln \left(1 + \frac{f}{700} \right) \quad (4)$$

Tahapan mel-filterbank yang dilakukan bertujuan untuk 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.

Pembuat mel-filterbank tahapan pertama yang dilakukan adalah menentukan batas atas dan bawah dari filter. Jadi nilai yang berada diluar batas tidak dimasukkan kedalam filter batas atas dan batas bawah. Lalu filter tersebut dikonversi kedalam skala mel, kemudian bagi range kedua batas tersebut sesuai dengan jumlah filter yang dibuat. Maka akan diketahui range dari batas atas dan bawah dari setiap filterbank dalam skala mel. Lalu batas-batas tersebut dikonversi lagi menjadi skala frekuensi linier, karena range tersebut tidak dapat merepresentasikan dalam bin FFT. Langkah terakhir adalah membuat filter triangular berdasarkan batas-batas yang sudah ditentukan. Hasil dari FFT pada tahap sebelumnya kemudian dikalikan dengan Mel-Filterbank.

2.5.6. Discrete Cosine Transform

Proses terakhir adalah *Discrete Cosine Transform* (DCT), yaitu proses perubahan konversi dari domain frekuensi ke domain waktu dengan menggunakan *Discrete Cosine Transform* (DCT). Proses ini dilakukan untuk mendapatkan nilai koefisien dari hasil perkalian mel-filterbank yang sudah dikonversi ke domain waktu. Proses ini akan menghasilkan log dari perkalian DCT yang sudah diubah ke domain waktu. Hasil log perkalian domain waktu ini menghasilkan *mel-frequency cepstrum coefficient* (MFCC). Berikut adalah persamaan yang digunakan:

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

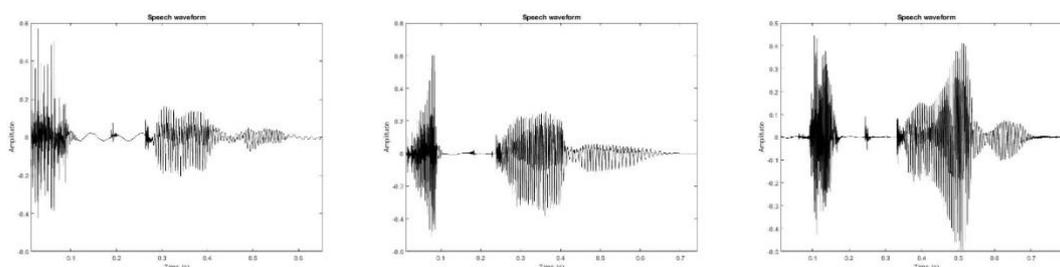
dimana $j = 1, 2, 3 \dots K$ adalah koefisien, dan M adalah jumlah filter.

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan beberapa sampel suara dari Berlin Emotion Speech Database (Emo-DB) dengan mengambil sampel kalimat “Das will sie am Mittwoch abgeben.” Rekaman suara tersebut kemudian di pre-processing dengan mengambil sinyal ucapan “abgeben” untuk setiap rekaman. Pada penelitian ini digunakan tiga sampel suara untuk masing-masing emosi yaitu emosi bahagia, bosan, netral, sedih, dan marah. Ketiga sampel suara tersebut diucapkan oleh aktor yang berbeda. Sampel-sampel suara tersebut di cari polanya dengan menggunakan ekstraksi fitur MFCC, dengan beberapa parameter 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, $C = 13$; nilai cepstral coefficients, dari parameter tersebut dihasilkan fitur MFCC. Fitur tersebut kemudian dicari nilai rata-ratanya untuk dikenali cirinya. Hasil dari ekstraksi fitur MFCC menggunakan ciri mean dapat dilihat pada Gambar 3 untuk pola emosi bahagia, Gambar 5 untuk pola emosi bosan, Gambar 7 untuk pola emosi netral, Gambar 9 untuk pola emosi sedih, dan Gambar 11 untuk pola emosi marah. Setiap gambar menampilkan masing-masing tiga sampel untuk setiap emosi.

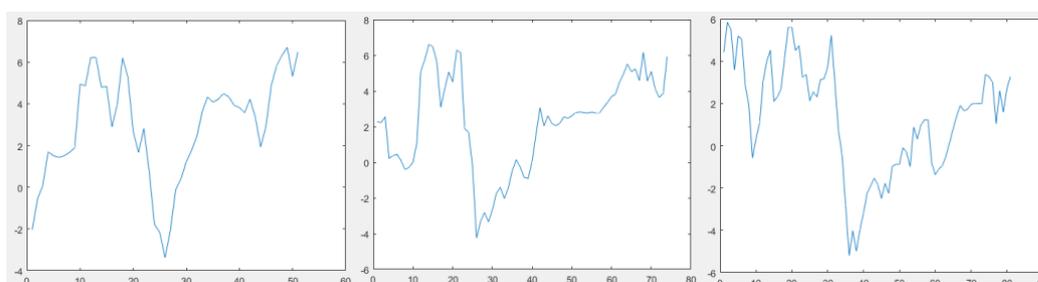
3.1. Pola Emosi Bahagia

Pada emosi bahagia pola sinyal ucapan yang belum di ekstraksi fitur dapat dilihat pada Gambar 2. Sinyal ucapan pada ketiga sampel emosi bahagia dibagian awal ucapan berwarna hitam pekat yang menunjukkan gelombang atau energi bahagia dari pemilik ucapan.



Gambar 2 Sinyal Ucapan Bahagia

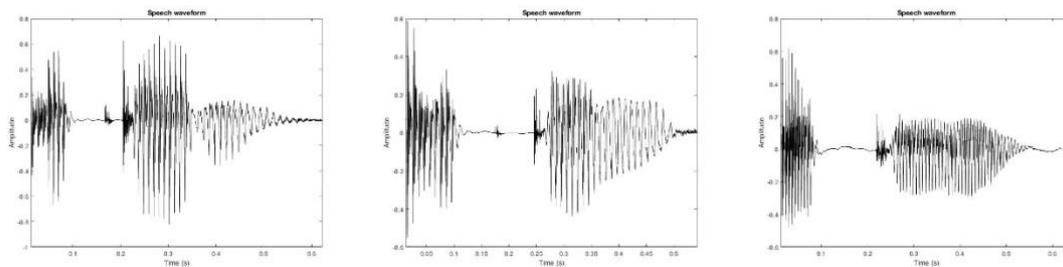
Ekstraksi fitur emosi bahagia dapat dilihat pada Gambar 3, pada gambar tersebut terlihat pola emosi bahagia cenderung memiliki gelombang tinggi pada bagian awal ke bagian akhir ucapan.



Gambar 3 Pola Emosi Bahagia

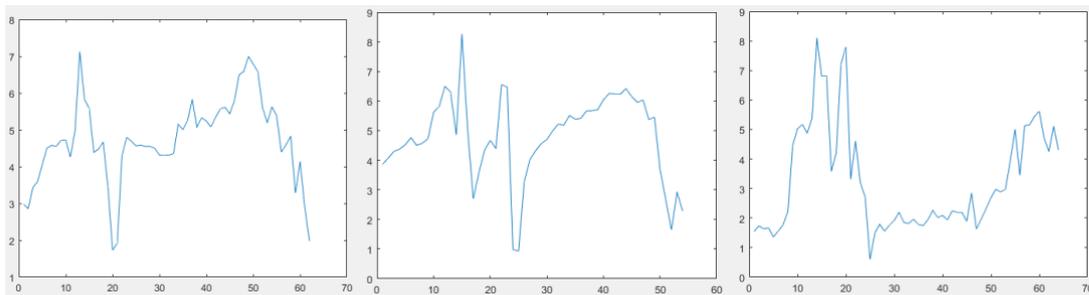
3.2. Pola Emosi Bosan

Pada emosi bosan pola sinyal ucapan yang belum di ekstraksi fitur dapat dilihat Gambar 4. Sinyal ucapan pada ketiga sampel emosi bosan memiliki energi yang sedikit terlihat dari warna hitam samar-samar. Warna hitam samar ini menunjukkan gelombang bosan dari pemilik ucapan yang menunjukkan bahwa emosi bosan tidak banyak mengeluarkan energi atau getaran.



Gambar 4 Sinyal Ucapan Bosan

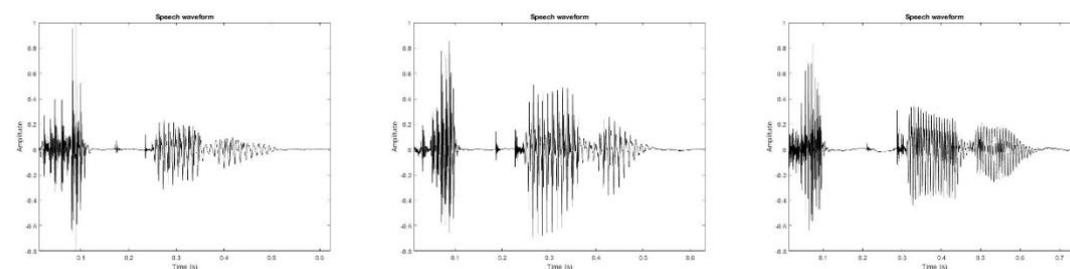
Ekstraksi fitur emosi bosan dapat dilihat pada Gambar 5, pada tiga gambar tersebut terlihat pola emosi bosan untuk masing-masing sampel cenderung dimulai dan diakhiri dengan gelombang turun.



Gambar 5 Pola Emosi Bosan

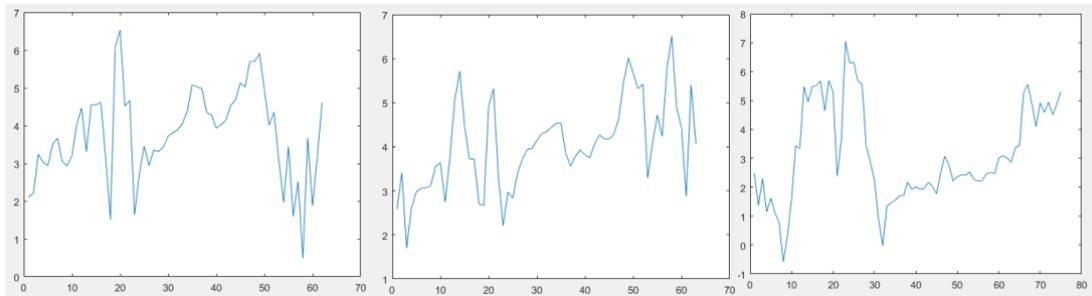
3.3. Pola Emosi Netral

Pada emosi netral pola sinyal ucapan dapat dilihat pada Gambar 6. Sinyal ucapan pada ketiga sampel emosi netral terlihat warna hitam lebih cenderung berwarna samar-samar dan gelombongnya tidak begitu rapat. Hal ini menunjukkan bahwa energi yang dikeluarkan pada saat emosi netral tidak begitu besar.



Gambar 6 Sinyal Ucapan Netral

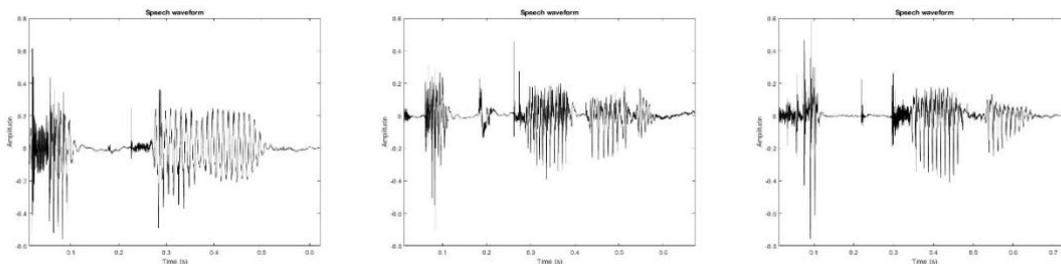
Ekstraksi fitur emosi netral dapat dilihat pada Gambar 7, pada gambar tersebut terlihat pola emosi netral cenderung memiliki gelombang turun pada bagian awal ke bagian akhir ucapan dan cenderung tidak beraturan.



Gambar 7 Pola Emosi Netral

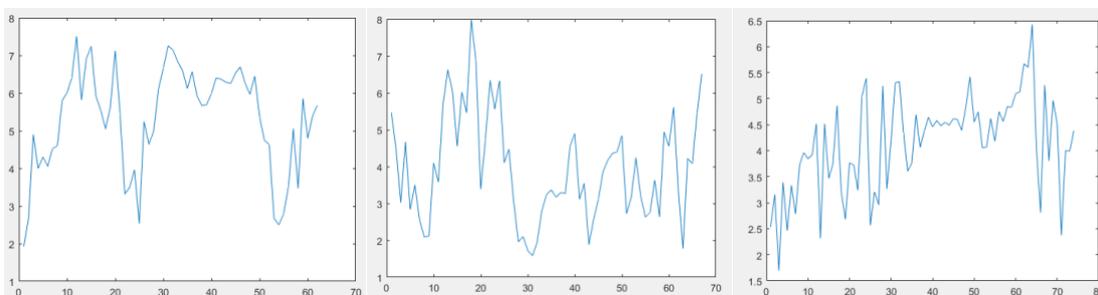
3.4. Pola Emosi Sedih

Pada emosi sedih pola sinyal ucapan dapat dilihat pada Gambar 8. Sinyal ucapan pada ketiga sampel emosi sedih gelombang yang dihasilkan tidak rapat yang terlihat dari warna hitam samar. Gelombang ini menunjukkan bahwa energi yang dikeluarkan suara ketika sedang sedih sangat sedikit.



Gambar 8 Sinyal Ucapan Sedih

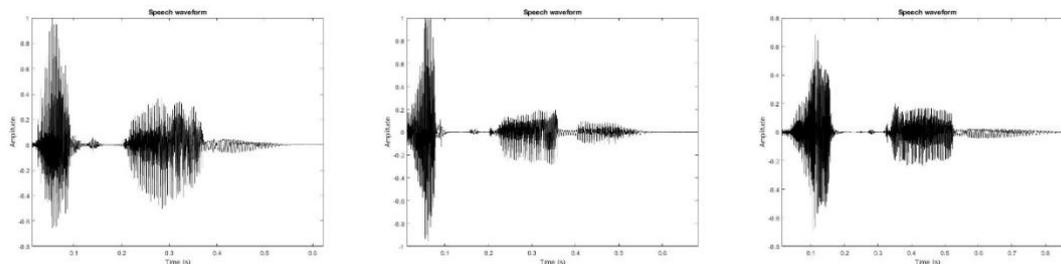
Ekstraksi fitur emosi sedih dapat dilihat pada Gambar 9, pada gambar tersebut terlihat pola emosi sedih cenderung memiliki gelombang naik pada bagian awal ke bagian akhir ucapan dan pola untuk emoi sedih terlihat sangat tidak beraturan.



Gambar 9 Pola Emosi Sedih

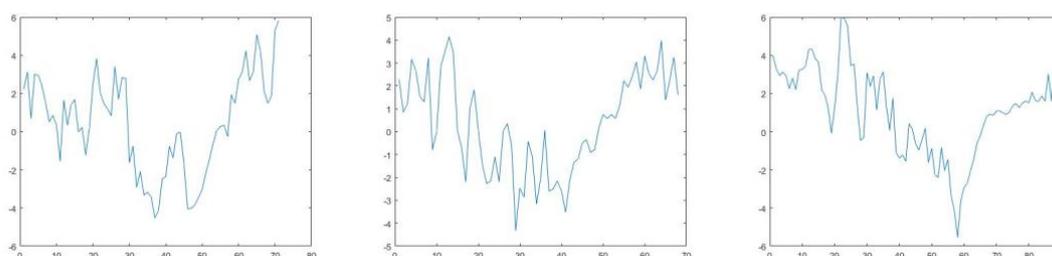
3.5. Pola Emosi Marah

Pada emosi marah pola sinyal ucapan dapat dilihat pada Gambar 10. Sinyal ucapan pada ketiga sampel emosi marah terlihat jelas gelombang yang dihasilkan memiliki warna hitam pekat. Gelombang tersebut menandakan bahwa suara yang dikeluarkan pada saat marah memiliki energi yang besar.



Gambar 10 Sinyal Ucapan Marah

Ekstraksi fitur emosi marah dapat dilihat pada Gambar 11, pada gambar tersebut terlihat pola emosi marah cenderung memiliki gelombang yang selalu naik dan tinggi. Selain itu, juga terlihat gelombang tersebut dimulai dan diakhiri dengan gelombang tinggi dan naik.



Gambar 11 Pola Emosi Marah

4. KESIMPULAN

Hasil dari penelitian yang telah dilakukan dapat disimpulkan bahwa ekstraksi fitur dengan menggunakan MFCC dapat dibedakan polanya untuk setiap emosi 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 emosi. Ektstaksi fitur ini dapat digunakan untuk proses pengklasifikasian emsoi selanjutnya.

5. SARAN

Penelitian selanjutnya untuk pengenalan emosi berdasarkan ucapan dengan menggunakan ekstraksi fitur MFCC dapat dilakukan dengan mengganti nilai pada parameter-parameter yang ada di fitur MFCC dan dapat menambah ciri lain seperti nilai minimum, maksimum, standar deviasi, dan ciri lainnya. Selain itu, dapat pula dilakukan ekstraksi fitur dengan metode lain untuk menghasilkan ekstraksi fitur yang lebih baik. Penelitian ini diharapkan bisa dilanjutkan pada tahap klasifikasi pengenalan emosi.

DAFTAR PUSTAKA

- [1] B. H. Prasetio, W. Kurniawan, M. Hannats, H. Ichsan, F. I. Komputer, and U. Brawijaya, "Pengenalan Emosi Berdasarkan Suara Menggunakan Algoritma HMM," vol. 4, no. 3, pp. 168–172, 2017.
- [2] A. Bombatkar, G. Bhoyar, K. Morjani, S. Gautam, and V. Gupta, "Emotion Recognition Using Speech Processing Using K- Nearest Neighbor Algorithm," *Int. J. Eng. Res. Appl.*, pp. 2248–9622, 2014.
- [3] R. B. Lanjewar, S. Mathurkar, and N. Patel, "Implementation and Comparison of Speech Emotion Recognition System Using Gaussian Mixture Model (GMM) and K- Nearest Neighbor (K-NN) Techniques," *Procedia Comput. Sci.*, vol. 49, pp. 50–57, Jan. 2015.
- [4] A. Al-Talabani, H. Sellahewa, and S. A. Jassim, "Emotion recognition from speech: tools and challenges," vol. 9497, p. 94970N, 2015.
- [5] S. Gustina, A. Fadlil, and R. Umar, "Identifikasi Tanaman Kamboja menggunakan Ekstraksi Ciri Citra Daun dan Jaringan Syaraf Tiruan," vol. 2, no. 1, pp. 128–132, 2016.
- [6] R. A. Surya, A. Fadlil, and A. Yudhana, "Ekstraksi Ciri Metode Gray Level Co-Occurrence Matrix (GLCM) dan Filter Gabor untuk Klasifikasi citra Batik Pekalongan," *J. Inform. J. Pengemb. IT*, vol. 2, no. 2, pp. 23–26, 2017.
- [7] I. Idrisa, M. S. H. Salamb, and M. S. Sunarc, "Speech Emotion Classification Using SVM and MLP on Prosodic and Voice Quality Features," *J. Teknol.*, vol. 78, 2015.
- [8] F. Burkhardt, A. Paeschke, M. Rolfes, W. F. Sendlmeier, and B. Weiss, "A database of German emotional speech," in *Ninth European Conference on Speech Communication and Technology*, 2005.
- [9] S. Gustina, A. Fadlil, and R. Umar, "Sistem Identifikasi Jamur Menggunakan Metode Ekstraksi Ciri Statistik Orde 1 dan Klasifikasi Jarak," *Techno. Com*, vol. 16, no. 4, pp. 378–386, 2017.
- [10] S. N. Zaini, H. Zaini, S. Sunardi, G. Kamarul Hawari, and T. Saiful Nizam, "Application of Speech Recognition for Swiftlet Vocalizations," 2013.
- [11] R. S. Azizah, D. Nurjanah, and F. D. Sari, "Sistem Automatic Speech Recognition Menggunakan Metode MFCC dan HMM Untuk Deteksi Kesalahan Pengucapan Kata Bahasa Inggris," *eProceedings Eng.*, vol. 2, no. 3, 2015.
- [12] M. W. . Sanjaya and Z. Salleh, "Implementasi Pengenalan Pola Suara Menggunakan Mel-Frequency Cepstrum Coefficients (Mfcc) Dan Adaptive Neuro-Fuzzy Inferense System (Anfis) Sebagai Kontrol Lampu Otomatis," *Al-HAZEN J. Phys.*, vol. 1, no. 1, pp. 1–19, 2014.
- [13] A. Yudhana, S. Sunardi, J. Din, S. Abdullah, and R. B. R. Hassan, "Turtle Hearing Capability Based on ABR Signal Assesment," *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 8, no. 2, pp. 187–194, 2010.