

Analisis Koefisien Cepstral Emosi Berdasarkan Suara

Ismail Mohidin ^{1*}, Frangky Tupamahu ^{2**}

* Teknik Informatika, Politeknik Gorontalo

** Teknik Informatika, Politeknik Gorontalo

is.mohidin@poligon.ac.id ¹, frangkytupamahu@poligon.ac.id ²

Article Info

Article history:

Received 12-10-2017

Revised 21-10-2017

Accepted 27-10-2017

Keyword:

Speech,
emotion,
MFCC,
classification

ABSTRACT

Abstract - The speech signal carries some sort of information, which consists of the intent to be conveyed, who speaks the information, and the emotional information that shows the emotional state of the utterance. One of the characteristics of human voice is the fundamental frequency. In this study the selection of features and methods of classification and recognition is important to recognize the emotional level (anger, sadness, fear, pleasure and neutral) contained in the dataset, this research proposes design through two main processes of training and introduction recognition). Experiments conducted using the Indonesian emotion voice dataset and the Mel-Frequency Cepstrum Coefficients (MFCC) algorithm were used to extract features from sound emotion. MFCC produces 13 cepstral coefficients of each of the sound emotion signals. This coefficient is used as an input of classification of emotional data from 250 data sampling, it is known that the Hidden Markov Model (HMM) method has a success rate of recognizing the average emotional percentage of recognition by using the number of training data is 85.44%, and the percentage of success of the introduction by using the number of test data is 44.25%.

Copyright © 2017 Journal of Applied Informatics and Computing.
All rights reserved.

I. PENDAHULUAN

Emosi suara berguna dalam menyampaikan sesuatu yang sedang dirasakan dengan menggunakan bahasa tubuh dan tidak terkecuali dengan berbicara [1]. Ketika seseorang berbicara, saat itulah sebenarnya seseorang tersebut menyampaikan beberapa informasi mengenai dirinya sendiri. Informasi tersebut dapat diketahui melalui fitur suara itu sendiri, diantaranya dengan *pitch* dan *formant*, dimana *pitch* adalah frekuensi *fundamental* dari sinyal suara yang dihasilkan karena getaran pita suara, dan *formant* adalah *frekuensi resonansi* akustik dari bidang suara manusia kedua *fitur* tersebut merupakan *fitur* suara yang sangat penting untuk mengenali keadaan emosi dari seseorang [2],[3]. beberapa *fitur* dari suara emosional yang banyak digunakan dalam pengenalan emosi, seperti *frekuensi* dasar [2], *formant*, *log energi* (amplitudo) [3], durasi, *koefisien cepstral* [4], *jitter* dan *shimmer* [5].

Mel Frequency Cepstral Coefficient atau MFCC adalah salah satu metode ekstraksi fitur yang digunakan dalam pengenalan suara untuk merepresentasikan suara menjadi sebuah *acoustic vector*. MFCC memiliki kelebihan dalam kompleksitas algoritma yang rendah dalam implementasi dari algoritma ekstraksi fitur. walaupun banyak sistem sukses menggunakan metode ini, namun dalam pengenalan emosi bahasa indonesia belum banyak

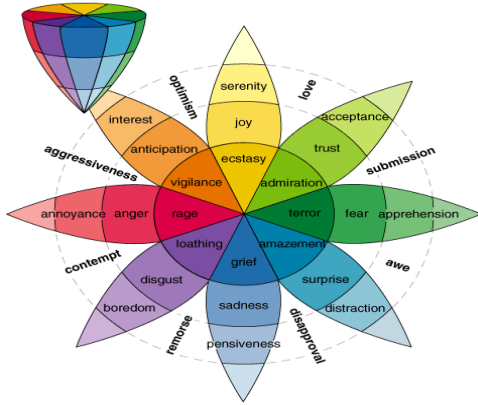
yang mengembangkan sehingga pada penelitian ini kami menggunakan MFCC sebagai metode ekstraksi fitur dalam pengenalan emosi dengan suara.

Beberapa penelitian mengklasifikasikan emosi diantaranya [6], mengklasifikasikan contoh suara Jepang menjadi lima emosi (marah, kebahagiaan, jijik, sedih, dan terkejut) dengan menggunakan *Bayes classifier* dan akurasi 51,25%. Klasifikasi suara Jerman menjadi enam emosi (marah, kebosanan, jijik, kecemasan, kebahagiaan dan kesedihan) dilakukan oleh [7]. menggunakan *classifier* SMO dan akurasi 92% [8]. mengklasifikasikan contoh suara indonesia menjadi empat emosi (marah, sedih, netral dan senang) dengan menggunakan MLP SMO dan Naive Bayes dan akurasi 97.78%, 97.22%, dan 57.22% untuk masing-masing classifier [9].

II. URAIAN PENELITIAN

Mendefinisikan kata emosi dengan “setiap kegiatan atau pergolakan pikiran, perasaan, nafsu, setiap keadaan mental yang hebat atau meluap-luap”. Emosi adalah suatu perasaan dan pikiran-pikiran khususnya, suatu keadaan biologis dan psikologis serta serangkaian kecenderungan untuk bertindak [7].

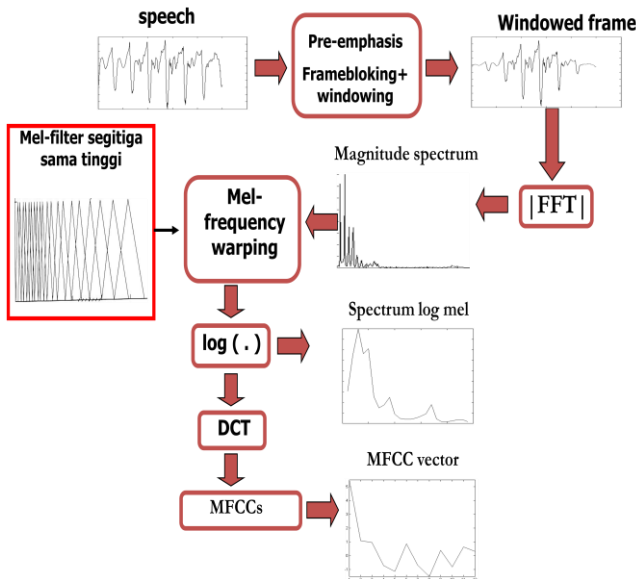
Plutchik mengemukakan idenya bahwa emosi dasar pada manusia merupakan sebuah blok bangunan utama dari emosi *derivatif*, emosi *sekunder* dan emosi *tersier*. Emosi *Derivatif* berada dalam campuran Negara, atau kombinasi dari emosi dasar majemuk. Hal ini mirip dengan prinsip warna pencampuran. Gambar 1. merupakan dasar model warna oleh *plutchik* [10].



Gambar 1. Dasar Warna Model Emosi (*Plutchik*)
 Sumber : (Plutchik, Robert , (2001). "Nature Of Emotions". *American*)

III. METODE PENELITIAN

Tahap Perancangan sistem ini terdiri atas proses pengambilan data, *Pre-processing*, ekstraksi ciri, pemodelan/inisialisasi, dan pelatihan dan pengujian model.



Gambar 2. Proses ekstraksi ciri MFCC

Proses *feature extraction* MFCC seperti dibawah

1. Melakukan proses *Pre-emphasize* untuk memperbaiki *envelope spectral* sinyal suara agar lebih halus.
2. Melakukan proses *Frame Blocking* bertujuan untuk membagi sampel signal menjadi beberapa *frame*.
3. Melakukan proses *Windowing* untuk mengurangi efek diskontinuitas pada ujung-ujung *frame* yang dihasilkan oleh proses *frame blocking*.
4. Melakukan proses FFT sehingga diperoleh sampel signal dalam *frekuensi domain*.
5. Melakukan proses *Fileterbank* untuk mengetahui ukuran energi dari setiap *frekuensi band*.
6. Melakukan proses DCT untuk mendapatkan *mel cepstrums*
7. Hasil dari proses DCT di sebut *Mel-Frequency Cepstrum Coefficients*

A. Pengambilan Data

Suara beremosi yang akan di klasifikasi haruslah berupa kata yang pendek yang dilatih untuk menirukan ucapan emosi, tetapi sudah menunjukkan emosi yang dikandung oleh pembicara, yang meliputi lima kelas emosi, seperti senang, sedih, marah, takut dan netral, dengan kata "tidak", Setiap kata diucapkan dengan intonasi yang menirukan suara emosional.

Pengambilan dataset suara dilakukan dengan laptop dan *mikrofon* untuk perekaman suara pada frekuensi sampel (F_s) 8000 Hz, 32 bit presisi resolusi, saluran mono dan dalam bentuk wav format *file*. Panjang setiap rekaman adalah 1,4 detik untuk setiap kata, Pengambilan data suara beremosi dilakukan dengan perekaman suara dengan lima orang responden pria untuk lima kelas emosi. Tiap responden melakukan perekaman sebanyak sepuluh kali untuk setiap kelas emosi suara dengan intonasi yang berbeda-beda. sehingga total *dataset* diperoleh 250 suara sampling dari lima kelas emosi. Hal ini bertujuan agar sistem mampu mengenali variasi suara kata yang diberikan.

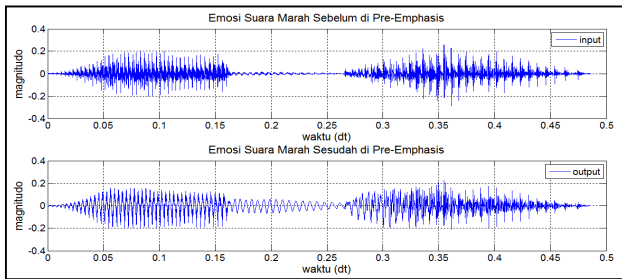
Untuk mengetahui tingkat keberhasilan pada penelitian ini nantinya dilihat dari *persentase* keakuratan pengenalan setiap emosi, yang didapatkan dari setiap *split rasio* jumlah suara yang dikenali benar dengan jumlah suara yang diujikan Antara lain :

1. 225 data *training* dengan *testing* dilakukan pada :
 - a. 225 data *training*
 - b. 25 data *testing*
2. 160 data *training* dengan *testing* dilakukan pada :
 - a. 160 data *training*
 - b. 90 data *testing*
3. 175 data *training* dengan *testing* dilakukan pada :
 - a. 175 data *training*
 - b. 75 data *testing*
4. 150 data *training* dengan *testing* dilakukan pada :
 - a. 150 data *training*
 - b. 100 data *testing*
5. 125 data *training* dengan *testing* dilakukan pada :

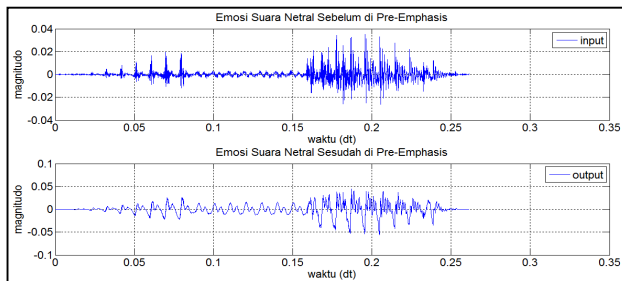
- a. 125 data *training*
- b. 125 data *testing*
- 6. 100 data *training* dengan *testing* dilakukan pada :
 - a. 100 data *training*
 - b. 150 data *testing*

B. Proses Extraction

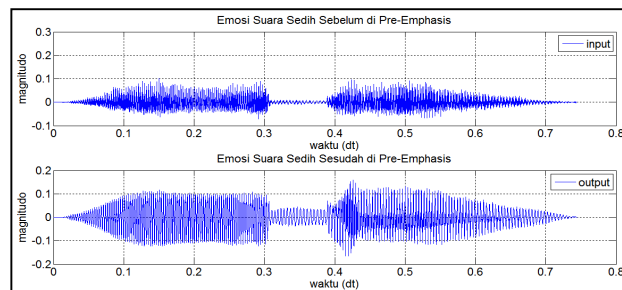
Preemphasis (penekanan sinyal) adalah suatu proses produksi suara manusia, radiasi pada bibir dan lidah ketika proses *phonation* mengakibatkan komponen frekuensi tinggi. Tujuan *preemphasis* adalah untuk memperhalus *envelope* agar meningkatkan sinyal pada pengolahan berikutnya. Default nilai α yang digunakan dalam proses ini adalah 0.97, Data sinyal baru adalah data sinyal sebelum proses *pre-emphasis* ditambah dengan data hasil *pre-emphasis*, berikut ini hasil *Preemphasis* lima kelas emosi:



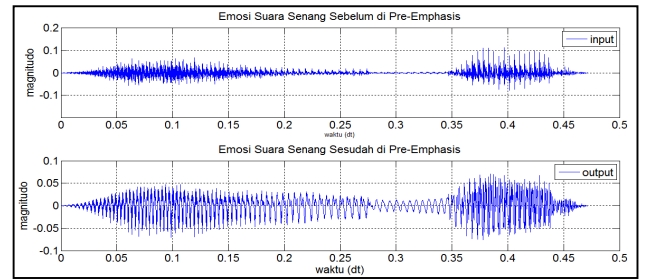
Gambar 3. Emosi suara marah



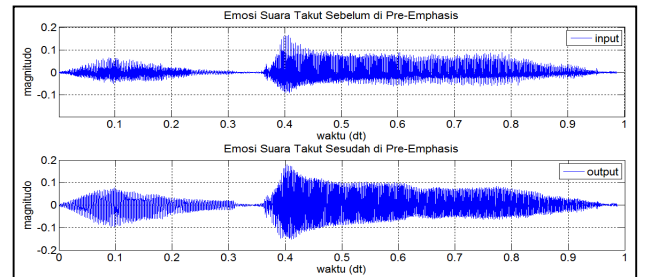
Gambar 4. Emosi suara netral



Gambar 5. Emosi suara sedih

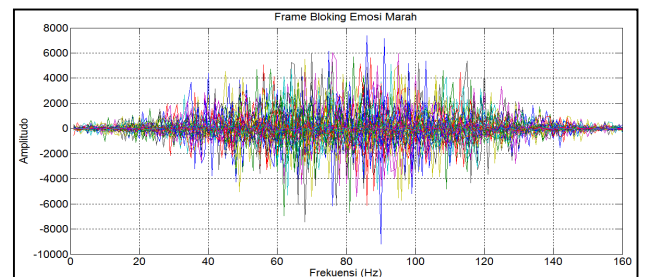


Gambar 6. Emosi suara senang

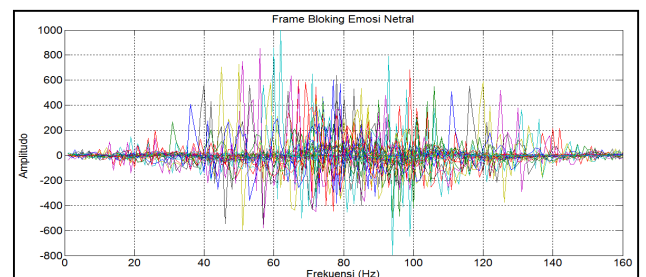


Gambar 7. Emosi suara takut

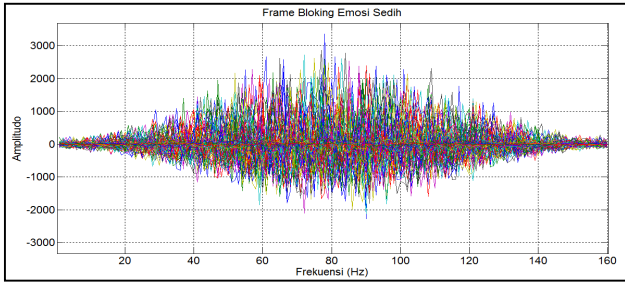
Frame Blocking pada proses ini sinyal suara disegmentasi menjadi beberapa *frame* yang saling tumpang tindih (*overlap*), hal itu dilakukan agar tidak ada sedikitpun sinyal yang hilang (*deletion*). Tujuan dari *frameblocking* adalah membagi sinyal kedalam sebuah matriks dengan interval panjang waktu yang diberikan. Berdasarkan asumsi sinyal dalam *frame* adalah 20 ms stasioner dan frekuensi sampling adalah 16000 Hz maka panjang tiap *frame* adalah 320 sampel. Pada *frameblocking* digunakan persentase *overlap* sebanyak 50%, memberikan faktor pemisah sepanjang 120 sampel. Berikut adalah hasil dari proses *frame blocking* untuk lima kelas emosi suara.



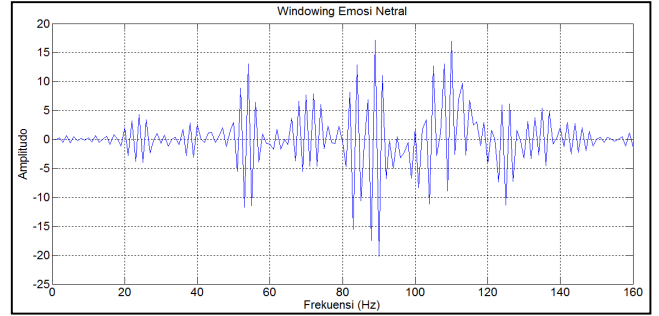
Gambar 8. Frame emosi marah



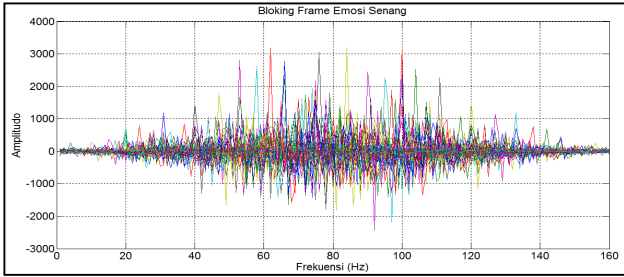
Gambar 9. Frame emosi netral



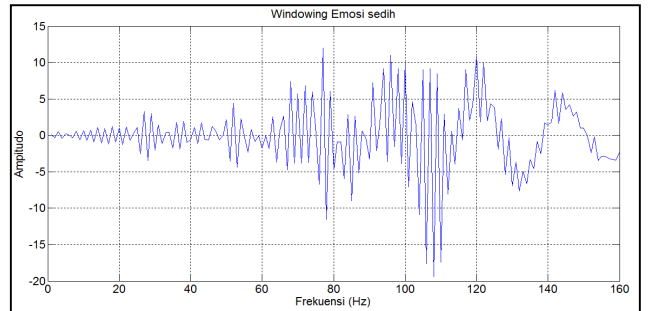
Gambar 10. Frame emosi sedih



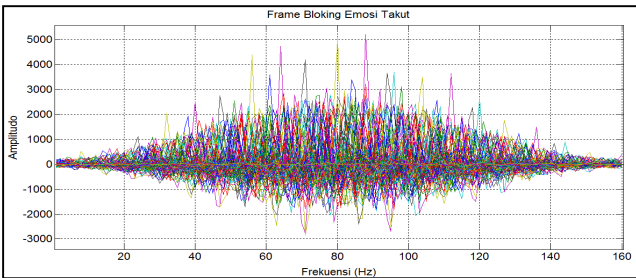
Gambar 14. windowing emosi suara netral



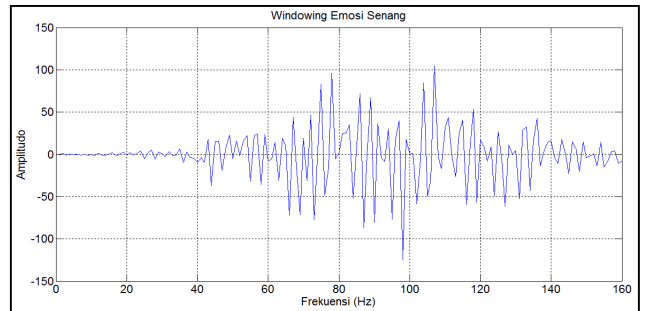
Gambar 11. Frame emosi senang



Gambar 15. windowing emosi suara sedih

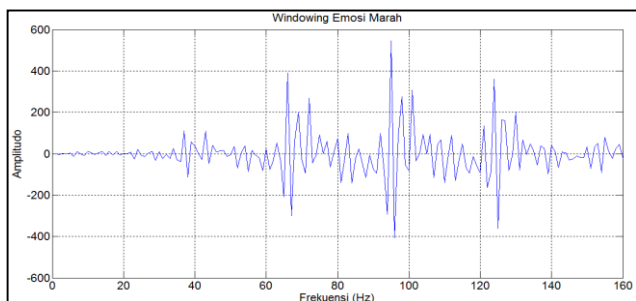


Gambar 12. Frame emosi takut

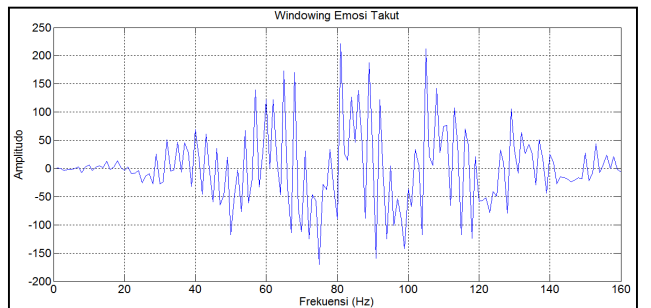


Gambar 16. windowing emosi suara senang

Setelah *frameblocking* dilakukan maka digunakan *Hamming window* pada tiap *frame*. Fungsi *window* ini digunakan untuk mengurangi diskontinu sinyal pada awal dan akhir tiap blok. *Windowing* Ada banyak fungsi *window* namun yang digunakan dalam penelitian ini adalah *hamming window* karena mempunyai hasil yang lebih baik dalam pembatasan sinyal yang akan dianalisa. Gambar dibawah ini adalah hasil dari proses *windowing* untuk lima kelas emosi.



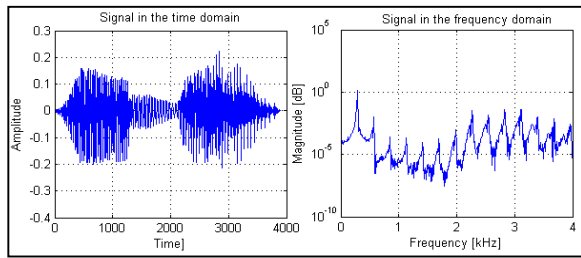
Gambar 13. windowing emosi suara marah



Gambar 17. windowing emosi suara takut

Untuk mengkonversi setiap *frame* dari *N sampel* dari domain waktu ke dalam domain frekuensi maka digunakan *FFT (Fast Fourier Transform)*. Karena panjang sampel adalah 320, maka *zero padding* digunakan untuk membuat panjang *sampel* menjadi kelipatan 2, yaitu 512. *Zero padding* adalah teknik untuk menambahkan nilai 0 pada sebuah barisan sehingga mendapatkan panjang barisan

yang diinginkan. Berikut hasil FFT pada salah satu sinyal suara untuk lima kelas emosi.



Gambar 18. Hasil Plot FFT emosi suara marah

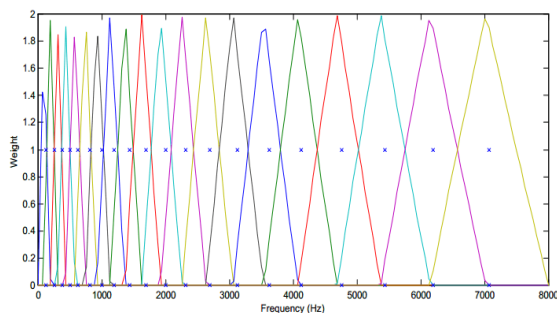
Tahap ini merupakan proses pengfilteran dari *spektrum* setiap *frame* yang diperoleh dari tahapan sebelumnya, menggunakan sejumlah M filter segitiga dengan tinggi satu. Filter ini dibuat dengan mengikuti persepsi telinga manusia dalam menerima suara. Persepsi ini dinyatakan dalam skala 'mel' (berasal dari Melody) yang mempunyai hubungan tidak linear dengan frekuensi suara, Dalam hal ini skala mel-frequency adalah linear untuk frekuensi kurang dari 1000 Hz dan logaritmik untuk frekuensi di atas 1000 Hz. Satu relasi antara frekuensi bunyi (dalam Hz) dengan *skala mel*. *Magnitude* hasil dari proses FFT selanjutnya akan melalui tahap *filterbank*. perhitungan *Filterbank* :

Diketahui :

$$S_0 = 0.4545$$

$$H_0 = \frac{2595 * \log(1 + \frac{0}{700})}{\frac{0.4545}{2}} = 4400,37$$

Maka didapat $S_0 = 0.4545 * 4400.37 = 1999,96$ Dengan cara yang sama maka didapatkan sinyal hasil *filterbank* adalah : (1999.96 , 246.4 , 418.7, 316.0 , 211.4, 216.2 , 117.3 ,181.1)



Gambar 19. Hasil Proses Mel-filterbank

Pada tahap akhir, ekstraksi MFCC yang diperoleh dari *Mel-Filterbank* harus dikonversi kembali. Dikarenakan *spektrum mel* merupakan bilangan real, maka *spektrum log mel* akan dikonversi menjadi domain waktu digunakan *Discrete Cosine Transform (DCT)*. Tujuannya adalah untuk mendapatkan informasi dari suatu sinyal emosi suara yang diucapkan oleh manusia. Hasil dari proses ini dinamakan *Mel-Frequency Cepstrum Coefficients (MFCC)*. Berikut ini adalah perhitungan untuk proses DCT:

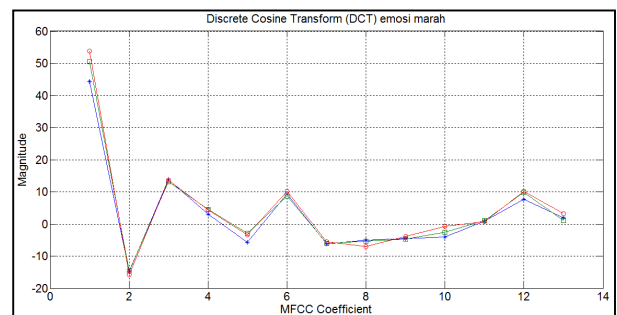
Koef (k) = 8, Untuk $n=0$ maka,

$$\begin{aligned} n_0 &= \log(1999,9) \cos \left[0 \left(0 - \frac{1}{2} \right) \frac{3,14}{8} \right] + \log(246,4) \cos \left[0 \left(1 - \frac{1}{2} \right) \frac{3,14}{8} \right] + \\ &\log(418,7) \cos \left[0 \left(2 - \frac{1}{2} \right) \frac{3,14}{8} \right] + \log(316,1) \cos \left[0 \left(3 - \frac{1}{2} \right) \frac{3,14}{8} \right] + \\ &\log(211,4) \cos \left[0 \left(4 - \frac{1}{2} \right) \frac{3,14}{8} \right] + \log(216,2) \cos \left[0 \left(5 - \frac{1}{2} \right) \frac{3,14}{8} \right] + \\ &\log(117,3) \cos \left[0 \left(6 - \frac{1}{2} \right) \frac{3,14}{8} \right] + \log(181,1) \cos \left[0 \left(7 - \frac{1}{2} \right) \frac{3,14}{8} \right] \\ &= 19,8 \end{aligned}$$

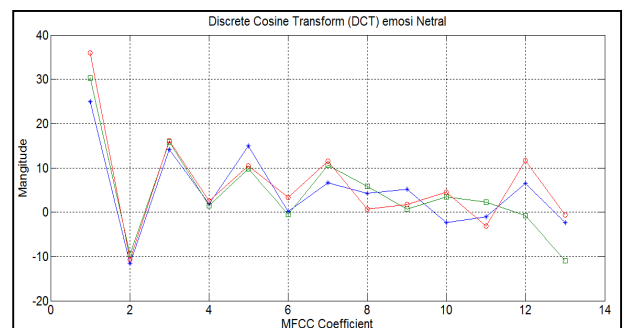
Jadi $n_0 = 19,8$

A. HASIL EKSPERIMEN

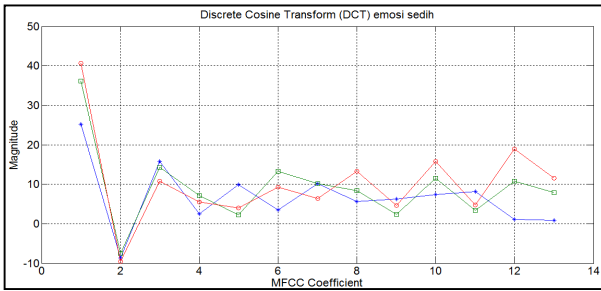
Proses DCT merupakan langkah terakhir dari proses utama MFCC. Hasil dari proses ini adalah *mel frekuensi cepstrum koefisien* yang merupakan hasil dari proses MFCC. data *koefisien* MFCC untuk lima kelas emosi dengan jumlah *koefisien* MFCC sebanyak 13 *koefisien* untuk masing-masing *frame*. Dengan demikian setiap data menjadi *matriks* 13 x 66. Hasil ini merupakan masukan untuk pelatihan HMM, yaitu berupa $O=O_1, O_2, \dots, O_{66}$. Seperti gambar dibawah ini :



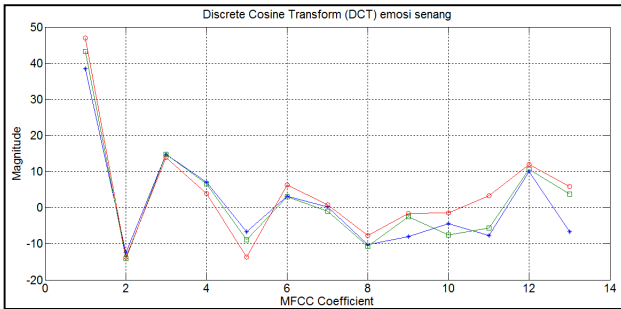
Gambar 20. DCT emosi marah



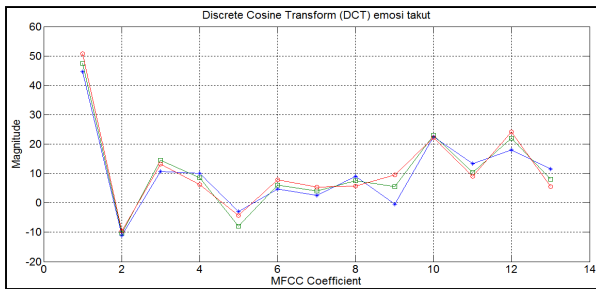
Gambar 21. DCT emosi netral



Gambar 22. DCT emosi sedih



Gambar 23. DCT emosi senang



Gambar 24. DCT emosi takut

Hasil ekstraksi emosi suara dari masing-masing lima kelas emosi menghasilkan nilai fitur yang berbeda. Hasil ekstraksi fitur didapat dari nilai rata-rata *magnitude* dimana nilai tersebut merupakan hasil dari nilai analisis *frame duration* 20 ms, nilai analisis *frame shift* 10 ms, nilai *preemphasis coefficient* 0,97, nilai *filterbank* 20, nilai *cepstral coefficient* 13, nilai *cepstral sine filter* parameter 22. Hasil ekstraksi fitur dari masing-masing kelas emosi dapat dilihat pada table data uji masing-masing kelas.

B. DATA UJI

Pada uji coba yang akan diproses, menggunakan dataset dari setiap emosi (marah, sedih, senang, takut dan netral) dengan kata “tidak, buka, tutup, diam”. Sebelum data diproses dalam HMM data diolah terlebih dahulu dengan menggunakan *Mel-Frequency Cepstrum Coefficients* (MFCC) untuk mendapatkan fitur emosi suara. Kemudian fitur emosi suara digunakan sebagai input *Hidden Markov Models* (HMM). Dibawah ini adalah table hasil pencocokan data uji.

Tabel berikut adalah nilai hasil dari pengujian emosi marah terhadap model emosi yang berbeda.

Tabel 1. Data Uji Emosi Suara Marah

No	Nilai Emosi Suara				
	Marah	Netral	Sedih	Senang	Takut
1	3.5395	6.8254	4.5843	6.1076	8.1838
2	6.6932	13.5455	6.3583	5.2628	6.8130
3	4.4692	8.8695	4.5105	5.7320	4.4335
4	5.2580	9.2067	9.1231	9.0289	6.8801
5	4.2491	6.3712	6.6977	4.7622	4.8529
6	7.1062	8.9012	7.3292	6.9262	9.7652
7	6.4876	9.4378	7.6712	6.2188	5.7118
8	6.6892	11.4637	12.5326	5.8583	8.1329
9	13.1965	13.0724	11.4799	10.1627	9.4095
10	13.6042	15.0755	14.0143	6.5749	7.4377
11	12.4347	10.7491	11.0984	5.7254	6.9389
12	8.6501	10.9310	13.0848	6.2796	4.5888
13	8.2920	10.6681	4.6445	5.4486	6.6387

Tabel berikut adalah nilai hasil dari pengujian emosi Netral dan sedih terhadap model emosi yang berbeda.

Tabel 2. Data Uji Emosi Suara Netral

No	Nilai Emosi Suara				
	Marah	Netral	Sedih	Senang	Takut
1	3.1379	5.2908	3.5671	2.1122	2.0578
2	2.9628	5.2798	7.1194	4.3856	3.0985
3	3.3686	7.6714	5.3405	3.0825	4.0510
4	6.7057	8.8156	6.0854	5.0211	4.1014
5	4.2210	6.3809	5.3664	3.3382	3.8223
6	5.4276	5.5969	5.1759	4.5492	4.8817
7	4.2713	5.9168	5.8310	4.8222	5.1185
8	4.6620	7.5208	7.1703	4.5936	5.3287
9	5.6986	5.4640	8.4423	5.6992	6.8055
10	5.1057	6.2179	6.1133	6.2716	4.0966
11	5.2193	4.8469	5.2663	4.7329	4.4689
12	4.9613	5.2450	4.4045	3.6676	3.9990
13	8.5162	3.5573	4.4201	2.8407	3.7509

Tabel 3. Data uji Emosi Suara Sedih

No	Nilai Emosi Suara				
	Marah	Netral	Sedih	Senang	Takut
1	3.4931	9.4064	3.0452	3.5229	4.3350
2	4.0423	8.4009	4.9257	5.0682	5.6000
3	5.0527	7.2876	4.1982	4.2617	4.9985
4	7.9266	12.9527	5.7833	3.5741	8.2337
5	4.9408	6.7091	6.2450	4.7717	6.2830
6	7.4644	8.8805	6.8210	5.6880	7.0998

7	4.7604	11.1824	8.5586	7.7432	8.5950
8	5.9429	9.2189	6.3036	6.8600	9.7288
9	6.9205	13.0060	7.9987	5.3055	7.3489
10	8.0664	14.3539	7.2302	6.3521	7.4332
11	9.9693	15.7958	12.0398	5.8330	9.3523
12	10.0793	12.4091	10.5838	4.9546	5.0800
13	6.9271	7.0917	3.2788	3.3953	8.2382

Tabel berikut adalah nilai hasil dari pengujian emosi senang dan takut terhadap model emosi yang berbeda.

Tabel 4. Data uji Emosi Suara Senang

No	Nilai Emosi Suara				
	Marah	Netral	Sedih	Senang	Takut
1	3.8966	4.7240	3.3412	2.4932	2.9039
2	6.4841	7.3940	5.0471	4.2910	5.6393
3	4.1799	8.3014	3.3268	2.5739	4.5214
4	5.9789	7.3337	6.3926	3.5264	5.9714
5	4.4815	7.6204	5.0057	5.3153	5.8344
6	6.1363	7.8633	7.5192	6.2126	6.2182
7	4.7013	7.7633	5.2625	4.2596	5.3052
8	6.5322	9.6570	6.0440	4.9075	6.9932
9	8.5262	8.5035	5.5365	4.3557	6.6166
10	9.2210	7.9531	6.1852	4.7137	8.3595
11	12.1852	6.1033	6.2565	4.0696	5.1313
12	11.3578	7.4072	5.3791	5.2553	4.8685
13	5.3390	6.5380	3.6171	3.3632	6.6273

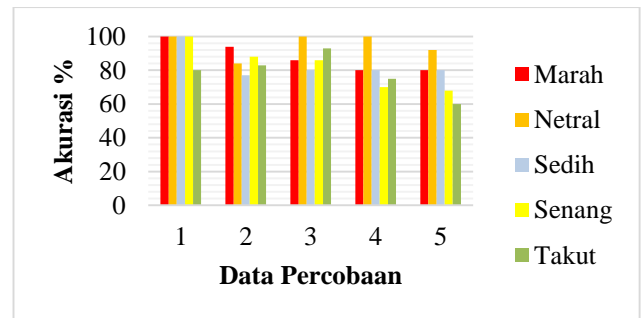
Tabel 5. Data uji Emosi Suara Takut

No	Nilai Emosi Suara				
	Marah	Netral	Sedih	Senang	Takut
1	5.2034	3.6715	3.3607	3.3745	2.9013
2	5.5209	8.5100	4.7655	3.9736	4.6058
3	7.1871	6.2387	5.7135	5.8822	7.9860
4	5.9099	5.7535	7.8044	5.7147	4.6678
5	5.0803	7.2642	6.4619	5.2330	5.1371
6	5.3193	7.9394	6.3937	7.8716	7.1743
7	6.0878	8.6506	5.5586	6.8849	6.4725
8	7.5572	14.9237	6.6076	7.1053	8.2163
9	11.1846	12.8952	6.5125	8.8901	6.9820
10	13.3020	13.7550	7.8191	7.1261	6.4065
11	15.6365	15.0941	8.4564	6.6995	3.8718
12	11.1109	15.7097	8.9237	5.9403	4.4177
13	5.0906	4.8844	3.9116	4.9591	5.1362

C. PEGUJIAN

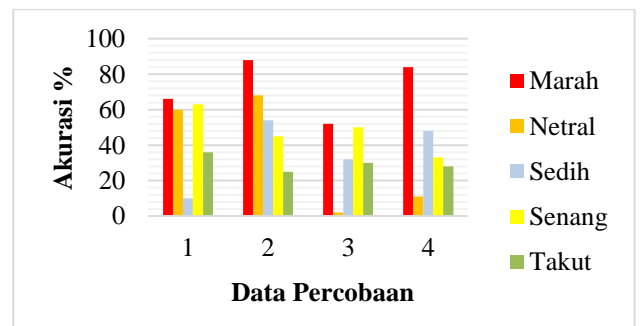
Pengujian sistem dilakukan menggunakan *algoritme Forward* dengan menghitung peluang data *testing* terhadap model HMM. Sistem yang telah dibuat, digunakan untuk

melakukan pengujian dalam pengukuran tingkat keakuratan klasifikasi suara berdasarkan emosi, dimana sistem ini memiliki proses pembelajaran terhadap fitur-fitur yang terdapat pada setiap kelas emosi. *Dataset* yang digunakan dalam penelitian ini adalah 250 data *sampling*. Untuk mengetahui tingkat keberhasilan pada penelitian ini dilihat dari persentase keakuratan identifikasi setiap emosi, yang didapatkan dari setiap rasio jumlah suara yang diidentifikasi benar dengan jumlah suara yang diujikan. Perbandingan tingkat akurasi berdasarkan jumlah banyaknya data yang *di training*kan dapat dilihat pada Gambar dibawah.



Gambar 25. Grafik Perbandingan Tingkat Akurasi Identifikasi Emosi Berdasarkan Banyaknya Data *Training*

dapat diketahui bahwa tingkat akurasi identifikasi emosi dengan data *training* lebih banyak memiliki nilai yang paling baik, dari beberapa kali percobaan. hal ini dikarenakan rasio data *training* lebih banyak dibandingkan dengan data *testing*. Sedangkan perbandingan tingkat akurasi berdasarkan jumlah banyaknya data yang di *testing* dapat dilihat pada Gambar dibawah ini.



Gambar 25. Grafik Perbandingan Tingkat Akurasi Identifikasi Emosi Berdasarkan Banyaknya Data *Testing*

Dari Gambar diatas terlihat bahwa tingkat akurasi rata-rata menurun drastis dari beberapa percobaan yg di ujikan. Di sisi lain pada pada proses pengujian untuk kelas emosi marah masih menunjukkan akurasi yang cukup baik dibandingkan dengan kelas emosi yang lain.hal ini dikarenakan rasio data *training* lebih kecil dibandingkan dengan data *testing*.

V. KESIMPULAN

1. Metode *Hidden Markov Model* (HMM) dapat digunakan untuk pengenalan emosi suara.
2. Hasil pengujian pengenalan emosi berdasarkan suara menunjukkan bahwa tingkat akurasi di luar *data train* menghasilkan akurasi 90%.
3. Data set emosi yang digunakan untuk klasifikasi dengan metode *Hidden Markov Model* (HMM) terdapat 200 data sampling dari lima kelas emosi, untuk data latih 200 dan 50 data testing.
4. Nilai-nilai parameter MFCC yang digunakan sangat mempengaruhi baik buruknya hasil dari proses MFCC itu sendiri, sehingga berpengaruh terhadap tingkat kesuksesan saat pengenalan.
5. Metode Mel Frequency Cepstrum Coefficients (MFCC) adalah metode yang baik untuk ekstraksi fitur dalam pengenalan emosi suara.
6. Nilai-nilai parameter MFCC yang digunakan sangat mempengaruhi baik buruknya hasil dari proses MFCC itu sendiri, sehingga berpengaruh terhadap tingkat kesuksesan saat pengenalan.
7. Jumlah data dalam pelatihan ternyata berpengaruh terhadap tingkat akurasi pengenalan yang dihasilkan. Semakin banyak jumlah data dalam pelatihan maka semakin tinggi tingkat akurasi yang dihasilkan.

DAFTAR PUSTAKA

- [1]. J. Or, (2008). *Affective Computing, Emotion Expression, Synthesis and Recognition*. I-Tech Education and Publishing.
- [2]. C. Busso, S. Lee, and S. Narayanan, (2009). "Analysis of Emotionally Salient Aspects of Fundamental Frequency for Emotion Detection," *IEEE Transactions on Audio, Speech, and Language Processing*.
- [3]. Y. Shi and W. Song, (2010). "Speech emotion recognition based on data mining technology," in *Sixth International Conference on Natural Computation (ICNC)*, Aug., vol. 2, pp. 615–619.
- [4]. T. Iliou and C.-N. Anagnostopoulos, (2010). "SVM-MLP-PNN Classifiers on Speech Emotion Recognition Field - A Comparative Study.
- [5]. I. Luengo, E. Navas, and I. Hernáez, (2010). "Feature Analysis and Evaluation for Automatic Emotion Identification in Speech," *IEEE Transactions on Multimedia*, vol. 12, no. 6, pp. 490–501.
- [6]. J. Cho and S. Kato, (2011). "Detecting emotion from voice using selective Bayesian pairwise classifiers," in *2011 IEEE Symposium on Computers Informatics (ISCI)*, pp. 90–95.
- [7]. S. Casale, A. Russo, G. Scebba, and S. Serrano, (2008). "Speech Emotion Classification Using Machine Learning Algorithms," in *IEEE International Conference on Semantic Computing*, pp. 158–165.
- [8]. J. Wang, Z. Han, and S. Lung, (2011). "Speech emotion recognition system based on genetic algorithm and neural network," in *International Conference on Image Analysis and Signal Processing (IASP)*, Oct., pp. 578–582.
- [9]. Muljono, Arifin, (2013). "Emotion Classifications of Indonesian Speech Data Set. 978-1-4673-6278-8/13/\$31.00 ©2013IEEE
- [110]. Plutchik, Robert , (2001). "Nature Of Emotions". *American*