SIMULASI DAN ANALISIS DETEKSI EMOSI MANUSIA DARI SUARA PERCAKAPAN BERBASIS *DISCRETE WAVELET TRANSFORM* DAN *LINEAR PREDICTIVE CODING*

Rita magdlena, IR.,MT[1],Ledya Novamizanti,SSi.,MT[2]
JurusanTeknikTelekomunikasi Universitas Telkom ivandy_chaniago@yahoo.co.id

Abtraksi

Emosi manusia adalah suatu hal yang terkadang hanya dapat diperkirakan melalui raut wajah dari seseorang saja, atau dari perubahan mimik wajahnya. Namun ternyata emosi manusia juga dapat dideteksi melalui suara yang diucapkannya. Emosi seseorang dalam keadaan tenang, marah, sedih atau senang dapat dideteksi melalui sinyal bicaranya. Pengembangan sistem pengenalan suara masih berjalan untuk sementara waktu ini. Sehingga pada penelitian ini dianalisis emose seseorang melalui sinyal bicaranya.

Pada tugas akhir yang akan dikerjakan ini, dirancang simulasi deteksi emosi manusia tersebut melalui sinyal bicara dengan melaksanakan ekstraksi ciri berbasis Discret Wavelet Transform(DWT) dan Linear Prediction Coding(LPC) untuk mendapatkan karakteristik dasar dari sinyal bicara. Kondisi emosi yang akan dideteksi tersebut nantinya akan menjadi state yang di dapat menggunakan metode Hidden Markov Model dan variabel ekstraksi ciri yang menjadi parameter penentu state.

Dari skenario pengujian terhadap paramater threshold didapat parameter terbaik yaitu 0.05. Setelah dilakukan pengujian terhadap klasifikasi 4 kelas emosi yaitu netral, marah, sedih, dan senang, akurasi tertinggi adalah 95% untuk jumlah 10 tiap-tiap kelas emosi, jumah data uji 5 tiap-tiap kelas emosi, nilai threshold crop 0.05, ukuran frame 512, nilai level DWT 2, nilai k dari KNN adalah 1.

Kata kunci : Deteksi Emosi, Suara percakapan, DWT, LPC.

Abstract

Human emotion is something that sometimes can only be estimated through the expression of a person alone, or of changes in her expression. But it turns out the human emotion can also be detected through the voice was saying. One's emotions in a state of calm, angry, sad or happy speech can be detected through the signal. The development of speech recognition system is still running for the time being. So in this study were analyzed emose someone through speech signals.

At the end of the task to be done is, designed the simulated detection of human emotion through speech signals by performing feature extraction based Discret Wavelet Transform (DWT) and Linear Prediction Coding (LPC) to obtain the basic characteristics of the speech signal. Emotional condition is detected the state will be able to use the method on hidden Markov models and variable feature extraction state that becomes the determining parameters.

Parameters of test scenarios to obtain the best parameter threshold is 0,05. After testing the 4-class emotion classification that is neutral, angry, sad, and happy, the highest accuracy was 95% for the number 10 of each class of emotion, the sheer number of test data 5 each emotion class, 0,05 threshold value crop, frame size 512 , the value of DWT level 2, the value of k of KNN is 1.

I. PENDAHULUAN

1.1 LatarBelakang

Perkembangan zaman yang semakin meningkat dengan cepat membuat pertumbuhan penduduk yang meningkat pula. Kemacetan, antrian sudah semakin sering terjadi di berbagai tempat seperti jalan, toko, atau tempat-tempat fasilitas umum. Hal tersebut juga dapat mengganggu emosi atau kejiwaan dari seseorang. Kesibukan yang berlipat ganda masalah yang datang bertubi-tubi akibat semakin banyak berinteraksi dengan orang pasti akan sangat mempengaruhi emosi dari seseorang. Terkadang kita hanya dapat melihat emosi seseorang dari raut wajahnya saja, namun sekarang kita dapat mendeteksi emosi atau kondisi kejiwaan seseorang dengan mendengarkan suaranya saja.

Suara manusia merupakan salah satu contoh dari sinyal analog yang berisikan informasi. Suara manusia juga unik, berbeda untuk masing-masing pribadi. Karakter suara seorang manusia ada 2 macam ada yang non akustik dan ada yang akustik. Non akustik contohnya adalah pulsa dan waktu sedangkan untuk akustik suara manusia terdiri dari pitch, formant, bandwidth formant, energy suara, dan durasi

pengucapan nya. Dari ciri akustik inilah kita dapat mengidentifikasi keadaan emosi sesorang apakah dia sedang merasakan senang, marah, atau sedih.

Di dalam tugas akhir ini akan dideteksi keadaan emosi seseorang melalui sinyal bicara manusia yang menggunakan ekstraksi ciri gabungan DWT dan LPC setelah itu akan diklasifikasikan menggunakan *classifier*. Metode ini sudah terkenal dalam *speech recognition*, karena sinyal bicara dapat di dikarakteristikan secara baik sebagai parameter proses acak dan dapat diestimasi secara akurat dengan perhitungan statistiknya.

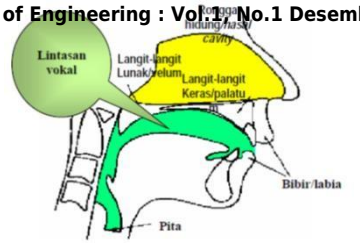
Dengan adanya deteksi emosi manusia ini maka diharapkan untuk kedepannya dapat memungkinkan kita menggunakan perangkat atau benda mati dalam melakukan segala hal. Deteksi emosi manusia ini diharapkan dapat menjadi salah satu alat komunikasi dengan benda mati seperti komputer atau robot yang didesain untuk melayani pemiliknya, sehingga terjadi pelayanan yang baik. Dalam hal lain juga deteksi ini diharapkan dapat membantu seseorang dalam berkonsultasi dengan psikolog. Mungkin akibat keterbatasan waktu atau kesibukan, seseorang tidak dapat berkonsultasi langsung dengan psikiater atau psikolog nya. Dengan deteksi ini maka cukup melalui telepon atau mendengar suara saja, psikolog dapat tetap menyelesaikan masalah si pasien.

II. LANDASANTEORI

2.1 Sinyal Suara Manusia

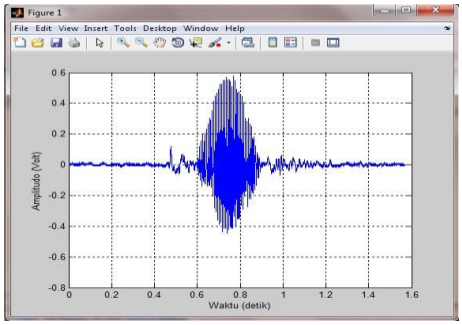
Suara manusia merupakan sinyal yang dihasilkan dari bergetarnya pita suara. Suara tersebut merupakan representasi dari pesan yang ingin disampaikan oleh otak kita. Pita suara manusia bergetar akibat adanya aliran udara dari paru-paru, dan getaran itu akan menghasilkan gelombang bunyi. Suara yang kita bunyikan akan bergantung dengan bagaimana kita meletakkan posisi lidah, gigi, dan rahang atau yang sering disebut dengan articulator, sehingga akan menghasilkan bunyi-bunyi vokal tertentu.

Pembangkitan sinyal suara terletak pada bentuk lintasan vokalnya (vocal tract). Lintasan vokal tersebut terdiri atas, dibawah katup tenggorokan (laryngeal pharynx), antara langit-langit lunak katub tenggorokan (oral pharynk), di atas velum dan diujung depan rongga hidung (nasal pharynx), dan rongga hidung (nasal cavity), seperti ditunjukkan pada gambar di bawah ini:



Gambar 2.1 lintasan vokal [digilib.its.ac.id]

Sinyal suara manusia merupakan sinyal yang berubah ubah secara periodik dengan kecepatan berubah yang relatif lama.

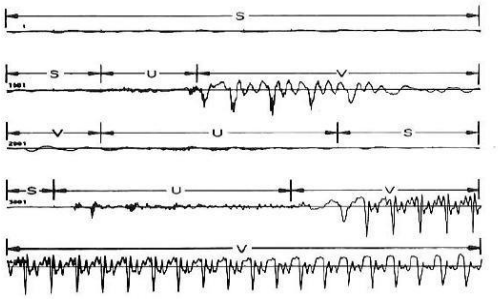


Gambar 2.2 contoh dari sinyal suara manusia

Sinyal suara memiliki frekuensi kerja antara 0 sampai 5 KHz. Ada komponen yang bisa diklasifikasikan dari sinyal suara manusia ini.

1. *Daerah silence,daerah disaat kita belum mengeluarkan suara apapun. Yang terekam hanya derau.*
2. *Daerah Unvoiced, daerah disaat pita suara belum bergetar, karena pita suara masih dalam keadaam lemas.*
3. *Daerah voiced, daerah disaat kita sudah mulai melakukan pengucapan huruf pertama dari kata yang akan kita ucapkan atau pita suara kita sudah bergetar dan menghasilkan bunyi.*

Sebagai contoh gambar dibawah ini :



Gambar 2.3 Cuplikan suara selama 500 ms [Rab93]

Gambar tersebut merupakan representasi sinyal suara yang dicuplik selama 100ms tiap gambar. S merupakan daerah silence, U merupakan daerah Unvoiced dan V merupakan daerah Voiced.

Namun ada juga daerah yang tidak dapat dikategorikan secara pasti termasuk dalam daerah yang mana, karena perubahan dari alat ucap manusia.

Sinyal suara manusia yang bisa didengar oleh telinga terletak di sekitaran 20Hz-20Khz, untuk suara dibawah atau diatas range tersebut tidak akan dapat didengar oleh telinga manusia. Suara manusia juga ada 2 macam, ada yang mono dan juga stereo. Suara mono adalah suara yang diproduksi dengan saluran tunggal, sehingga kualitas suara tidak cukup baik, jika direpresentasikan sebagai gambar maka suara mono itu seperti gambar grayscale yang hanya memiliki 1 layer dengan piksel yang sedikit. Sedangkan suara stereo merupakan suara yang dihasilkan dengan lebih dari satu saluran audio independent, sehingga suara yang didengar lebih natural.

2.2 Teori Emosi Manusia

Emosi merupakan hal yang dialami oleh setiap manusia. Pengekspresian emosi sesorang dapat berbeda-beda, bisa dengan perubahan raut wajah, perubahan nada bicara, maupun perubahan bahasa tubuh. Perbedaan emosi yang muncul dapat dipengaruhi oleh lingkungan sekitar dan orang-orang yang berada didalamnya.

Emosi berbeda dengan mood dan temparemen sesoarang. Ketiga hal tersebut merupakan kondisi psikologis yang dimililiki oleh manusia, perbedaannya emosi bisa dirasakan hanya sesaat atau pada waktu itu saja. Mood akan berlangsung dalam beberapa hari, sedangkan tempramen biasanya berlangsung seumur hidup manusia akan bisa dikatakan karakteristik dari manusia tersebut.

2.3 SuaraKarakteristik Manusia

2.3.1 Pitch

Pitch dapat diartikan sebagai nada dasar atau unsur bunyi terkecil dari suara manusia. Panjang Pitch sekitaran 10 ms. Pitch manusia berbeda tergantung pada usia dan jenis kelamin, karena pita suara perempuan dan laki-laki memiliki lebar yang berbeda maka akan menghasilkan pitch yang berbeda. Untuk laki-laki dewasa memiliki pitch yang lebih rendah dengan ukuran pita suara sekitar 17 mm sampai 25 mm. Sedangkan untuk perempuan 12.5 mm sampai 17.5 mm.

2.3.2 IntensitasEnergidanDurasiPengucapan

Dalam pengucapan suatu kalimat, biasanya tiap suku kata memiliki nada yang berbeda-beda. Terkadang ada saatnya nadanya harus rendah atau tinggi. Pelan atau kerasnya suara yang diucapkan oleh manusia biasa disebut dengan Intensitas Energi. Perbedaan nada tersebut biasanya untuk memberikan kesan terhadap pengucapan kalimat itu atau bisa diartikan sebagai keadaan emosi kita saat mengucapkan kata-kata tersebut.

Setiap manusia juga memiliki perbedaan waktu dalam mengatakan kata-kata atau kalimat tertentu. Kecepatan waktu yang diperlukan dalam pengucapan kata tersebut disebut dengan Durasi Pengucapan. Ada orang yang biasanya memerlukan waktu yang cepat dalam mengatakan sesuatu namun terkadang ada yang biasa-biasa saja, atau bahkan memerlukan waktu yang lama. Hal ini juga dipengaruhi oleh keadaan emosi seseorang tersebut.

2.3.3 Discrete Wavelet Transform

Sinyal suara percakapan manusia adalah sinyal nonstasioner. Transformasi fourier tidak terlalu baik untuk analisis sinyal nonstasioner karena hanya memberikan informasi frekuensi tetapi tidak memberikan informasi saat waktu kapan frekuensi sinyal tersebut terjadi. Short Time Fourier Transform merupakan pengembangan selanjutnya dengan cara memotong sinyal menjadi beberapa bagian kecil dan bagian tersebut ditransformasi fourier. Dimana STFT memiliki nilai resolusi waktu yang tetap karena memiliki panjang jendela potongan yang sama pula. Transformasi Wavelet memiliki panjang jendela potongan waktu dan frekuensi yang fleksibel dan merupakan alat bantu yang baik untuk sinyal nonstasioner seperti sinyal percakapan manusia.

Transformasi wavelet mendekomposisi sinyal berdasarkan proses translasi dan dilatasi berdasarkan mother wavelet. Mother wavelet merupakan fungsi waktu yang memiliki energy terbatas dan delay yang cepat. Persamaan 2.1 menunjukan persamaan transformasi wavelet kontinu (TWK) dimana ψ(t), a, dan b merupakan mother wavelet, faktor skala, dan parameter translasi.

$$\text{[illegible]} = \frac{1}{\text{[illegible]}} \int_{-\infty}^{\infty} \text{[illegible]}\ \psi^{*}\left(\frac{\text{[illegible]}}{\text{[illegible]}}\right) \text{[illegible]} \quad (2.1)$$

Pada persamaan diatas terdapat dua parameter pada Transformasi Wavelet Kontinu yaitu a dan b. Analisis sinyal dengan menggunakan nilai angka yang kecil dari skala dan translasi parameter a = 2^j dan b = 2^jk merupakan Transformasi Wavelet Diskrit. Teori Transformasi Wavelet Diskrit terdiri dari dua fungsi yang terhubung disebut fungsi skala dan fungsi wavelet.

$$\Phi\ \text{[illegible]} = \text{[illegible]}\ \text{[illegible]}\ \overline{2}\,\Phi\ 2\text{[illegible]} \quad (2.2)$$

Dan

ψ = (2.3)

2.3.4 Linear Prediction Coding

LPC adalah koefisien dari sebuah model auto-regressive dari suara percakapan yang telah dilakukan proses framing. Semua pole dari Fungsi transfer dari komponen vokal adalah pada persamaan berikut

= (2.4)

2.3.5 K-Nearest Neighbour (KNN)

K-Nearest Neighbour (K-NN) adalah suatu metode yang menggunakan algoritma *supervised* dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada K-NN. Tujuan dari algoritma ini adalah mengklasifikasi objek baru berdasakan atribut dan sampel latih. Pengklasifikasian tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik uji, akan ditemukan sejumlah K objek (titik training) yang paling dekat dengan titik uji. Klasifikasi menggunakan *voting* terbanyak di antara klasifikasi dari K objek. Algoritma K-NN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sample uji yang baru. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak *Eucledian.*

Algoritma metode KNN sangatlah sederhana, bekerja berdasarkan jarak terpendek dari *sampel uji* ke *sampel latih* untuk menentukan KNN-nya. *Sampel latih* diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data.Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi *sampel latih*.Sebuah titik pada ruang ini ditandai dengan kelas c, jika kelas c merupakan klasifikasi yang paling banyak ditemui pada K buah tetangga terdekat dari titik tersebut. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan*Euclidean Distance* yang direpresentasikan sebagai berikut :

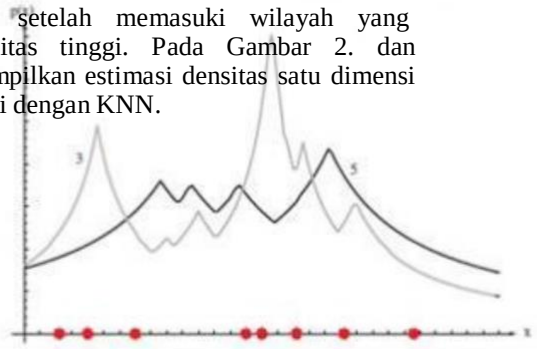
$$D(a,b) = \sqrt{\sum_{k=1}^{d} (a_k - b_k)^2},$$

(2.6)

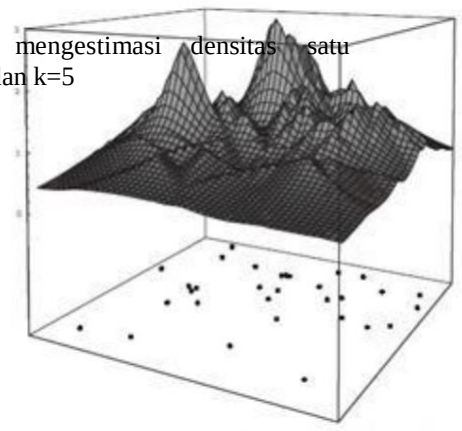
Dimana matriks D(a,b) adalah jarak skalar dari kedua vektor a dan b dari matriks dengan ukuran d dimensi.

Pada fase *latih*,algoritma ini hanya melakukan penyimpanan vektor-vektor fitur dan *data* (yang klasifikasinya tidak diketahui). Jarak dari vektor baru yang ini terhadap seluruh vektor *sampel latih* dihitung dan sejumlah *k* buah yang paling dekat diambil.Titik yang baru klasifikasinya diprediksikan termasuk pada klasifikasi terbanyak dari titik-titik tersebut.

Sebagai contoh, untuk mengestimasi *p(x)* dari *n sampel latih* dapat memusatkan pada sebuah sel disekitar *x* dan membiarkannya tumbuh hingga meliputi *k sampel*. *Sampel* tersebut adalah KNN dari *x*. Jika densitasnya tinggi di dekat *x*, maka sel akan berukuran relatif kecil yang berarti memiliki resolusi yang baik. Jika densitas rendah, sel akan tumbuh lebih besar, tetapi akan berhenti setelah memasuki wilayah yang memiliki densitas tinggi. Pada Gambar 2. dan Gambar 2.ditampilkan estimasi densitas satu dimensi dan dua dimensi dengan KNN.



Gambar 2.4KNN mengestimasi densitas satu dimensi dengan k=3 dan k=5



klasifikasi data *sampel latih.* Pada fase klasifikasi, fitur-fitur yang sama dihitung untuk *testing*

Gambar 2.5KNN mengestimasi densitas dua dimensi dengan k=5

Nilai *k* yang terbaik untuk algoritma ini tergantung pada data. Secara umum, nilai *k*yang tinggi akan mengurangi efek *noise* pada klasifikasi, tetapi membuat batasan antara setiap klasifikasi menjadi semakin kabur. Nilai *k* yang bagus dapat

dipilih dengan optimasi parameter, misalnya dengan menggunakan *cross-validation*. Kasus khusus dimana klasifikasi diprediksikan berdasarkan *data latih* yang paling dekat (dengan kata lain, *k = 1*) disebut algoritma *nearest neighbor*.

Ketepatan algoritma KNN sangat dipengaruhi oleh ada atau tidaknya fitur-fitur yang tidak relevan atau jika bobot fitur tersebut tidak setara dengan relevansinya terhadap klasifikasi.Riset terhadap algoritma ini sebagian besar membahas bagaimana memilih dan memberi bobot terhadap fitur agar performa klasifikasi menjadi lebih baik.

KNN memiliki beberapa kelebihan yaitu ketangguhan terhadap *data latih* yang memiliki banyak *noise* dan efektif apabila *pelatihandatanya* besar. Sedangkan, kelemahan KNN adalah KNN perlu menentukan nilai dari parameter *k* (jumlah dari tetangga terdekat), *pelatihan* berdasarkan jarak tidak jelas mengenai jenis jarak apa yang harus digunakan dan atribut mana yang harus digunakan untuk mendapatkan hasil terbaik, dan biaya komputasi cukup tinggi karena diperlukan perhitungan jarak dari tiap *data uji* pada keseluruhan *sampel latih*.

Nearest Neighbor didasarkan pada suatu asumsi bahwa sekumpulan sesuatu ***d***yang mirip (= dekat) mestinya merupakan satu kelas yang sama. Ukuran kedekatan dalam distribusi data adalah jarak-jarak geometrik yang sering dikenal dengan *Euclidean Distance, Mahalanobis Distance,* dan sebagainya. Berikut ini langkah-langkah untuk mengklasifikasi sebuah vector baru ***d***, jika diberikan data latih ***(D$_i$, C$_i$),*** i=1,2, …p, ***D$_i$*** merupakan data ke-I pada ***C$_i$,***kelas ke-i.

Pertama hitung jarak ***s***antara ***d*** dengan setiap anggota ***D,*** misalnya menggunakan jarak *Euclidean* seperti diberikan oleh persamaan di bawah ini. Kemudian tentukan vektor ***D$_i$*** yang paling dekat dengan ***d,*** dalam hal ini ***s$_i$*** adalah minimum dalam semua ***s.*** Terakhir nyatakan bahwa ***d*** termasuk dalam kelas ***C$_i$***. Dalam hal terdapat beberapa jarak yang sama *(equidistant),* maka kelas dengan anggota *equidistant* terbanyak akan menjadi kelas bagi ***d.***

$s = \sqrt{(\square^2 + (\square^2}$ (2.7)

Perluasan dari *nearest neighbor* adalah *K-Nearest Neighbor*. Dalam *K-Nearest Neighbor* klasifikasi diawal dengan membangun *hypersphere* mengelilingi data atau vektor yang akan diklasifikasikasi. Pada gambar diatas, lingkaran pertama merupakan *nearest neighbor* sehingga data "?" akan dimasukkan sebagai kelas 1. Tapi dengan menggunakan 3 *nearest neighbor* terdapat dua buah 2 dan sebuah 1, sehingga data "?" diklasifikasikan sebagai data kelas 2.

K-Nearest Neighbor sesuai namanya mengambil keputusan bahwa data baru ***d*** termasuk dalam kelas ***C*** berdasarkan beberapa tetangga terdekat dari ***d.*** Jika digunakan jarak *Euclidean* sebagai ukuran kedekatan maka ***d*** akan menjadi pusat *hypersphere* dengan jari-jari **r** sama dengan jarak *Euclidean* tersebut. Yang dilakukan adalah menaikkan **r** sehingga *hyperspere* memuat **K** data. Kelas untuk data ***d*** diberikan berdasar jumlah anggota kelas terbanyak yang muncul dalam *hyperspere* tersebut. Pada k-NN terdapat beberapa aturan jarak yang dapat digunakan, yaitu :

1. *Euclidean Distance*, dengan rumus :

$$l_2 \square\square = \sqrt{\square^1(\square - \square)^2} \quad (2.8)$$

2. *City block* atau *manhattan distance*, dengan rumus :

$$\square_1 \square\square = \sum_{\square=1} \square - \square \quad (2.9)$$

3. *Cosine*, dengan rumus :

$$\cos \square\square = \frac{\square\square\square\square}{\sqrt{\square\square^2}\ \sqrt{\square\square^2}} \quad (2.10)$$

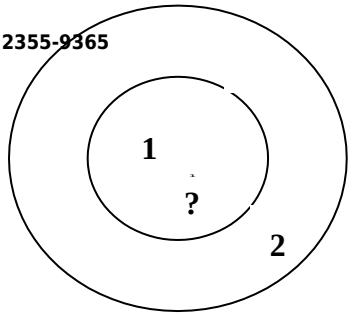
4. *Correlation*, dengan rumus :

$$\square\square = \frac{\sum_{\square=1} \square\square - \square\ (\square\square - \square)}{\left[\sum_{\square=1}(\square\square - \square)^2\ \sum_{\square=1}(\square\square - \square)^2\right]^{\frac{1}{2}}} \quad (2.11)$$

dimana

$$\square = \frac{1}{\square} \sum_{\square=1} \square\square \qquad \square\square\square\square = \frac{1}{\square} \sum_{\square=1} \square\square \quad (2.12)$$

Sangat jelas bahwa jika K bertambah besar



Gambar 2.6 Ilustrasi K-NN

maka hasil klasifikasi akan konvergen kepada kelas yang jumlah anggotanya terbanyak. Dengan demikian perlu dipilih K > 1 namun K < p, sehingga klasifikasi memberikan hasil benar yang maksimal. Untuk melakukan itu perlu diujikan beberapa data masukkan yang sudah diketahui kelasnya namun bukan anggota data latih. Untuk setiap K yang dipilih, banyaknya

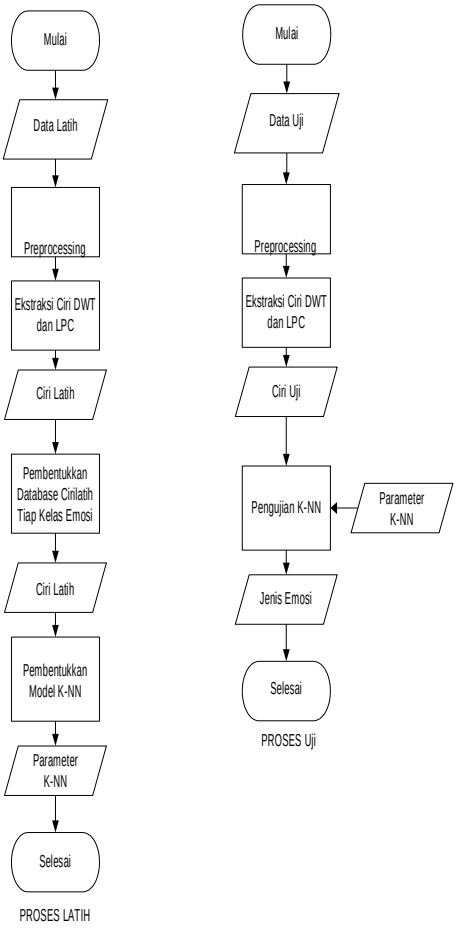
klasifikasi benar akan berbeda-beda. Nilai K yang memberikan hasil benar terbanyak ini merupakan K optimal yang akan dipakai untuk klasifikasi-klasifikasi selanjutnya dalam sistem yang sama.

Dibandingkan metode klasifikasi yang lain, metode ini memiliki keakuratan yang cukup tinggi karena data yang masuk akan diklasifikasikan berdasarkan kemiripan ciri yang ada dari data sebelumnya yang sudah diklasifikasikan. Namun pada algoritma K-NN perlu menentukan nilai dari parameter K (jumlah dari tetangga terdekat).

III. PERANCANGAN DAN SIMULASI SISTEM

3.1 Perancangan Sistem

Sistemdeteksiemosimanusia dari suara percakapan yang dirancang terdiri dari 2 proses yaitu proses latih dan proses uji. Alur kerja sistem dalam tugas akhir ini dapat dilihat dari gambar di bawah ini.



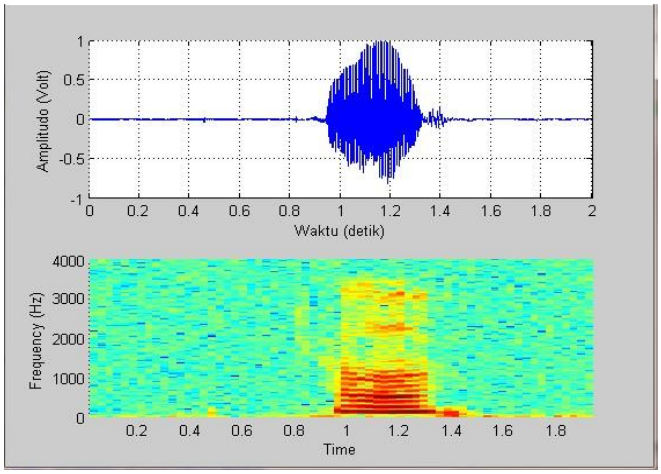
Gambar 3.1 PerancanganSistem (a) Proses Latih (b) Proses Uji

Proses latihmerupakan proses pembentukkan sistem klasifikasi berdasarkan data latih sebagai acuan. Sedangkan proses uji merupakan proses sesungguhnya sistem yang telah dirancang pada proses latih untuk mengklasifikasi jenis emosi dari data uji yang dipilih pada proses uji. Proses latih dan uji secara garis besar sama hanya saja pada proses latih berarti membangun sistem klasifikasi menggunakan K-NN yang menghasilkan model klasifikasi K-NN yaitu parameter K-NNyang digunakan pada proses uji klasifikasi.

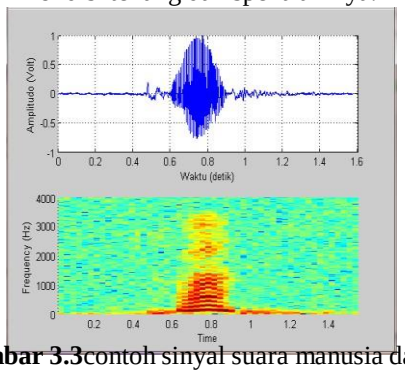
3.1.1 Proses Sinyal Suara

Dalam proses ini dilakukan pemotongan rekaman data suara manusia yang telah didapat dari *Emotional Prosody Speech and Transcript Library* menggunakan *Cool Edit Pro 2.1*, karena perekaman yang dilakukan sekaligus dalam satu file .wav untuk semua emosi. Sinyal suara yang sudah dipotong tadi adalah sinyal dari suara aktor dengan setiap suara merepresentasikan masing-masing emosi yaitu netral, marah, sedih, dan senang. Sinyal suara tersebut kemudian disimpan dalam format .wav dan dikelompokkan berdasarkan emosinya.

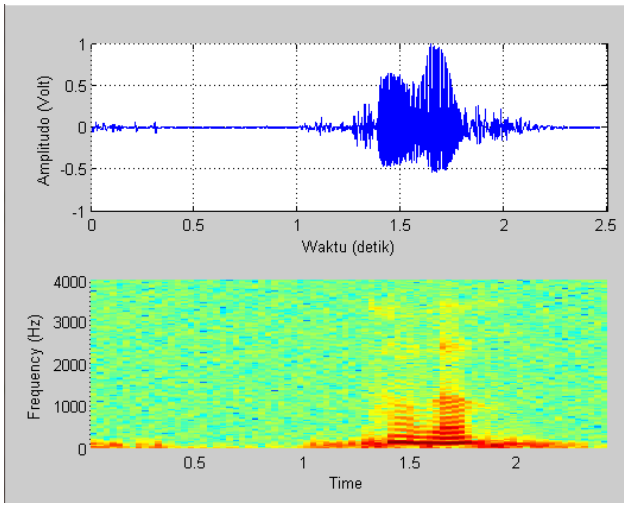
Berikut adalah contoh gambar dari sinyal hasil perekaman yang merepresentasikan masing-masing emosi manusia :



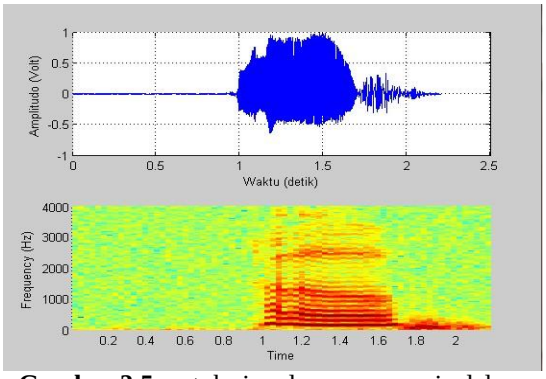
Gambar 3.2contoh sinyal suara manusia dalam kondisi tenang dan spektrumnya.



Gambar 3.3contoh sinyal suara manusia dalam kondisi marah dan spektrumnya



Gambar 3.4contoh sinyal suara manusia dalam kondisi sedih dan spektrumnya



Gambar 3.5contoh sinyal suara manusia dalam kondisi senang dan spektrumnya

3.1.2 Pre Processing

Pre processing merupakan proses permulaan sinyal suara yang akan diolah atau dicari ciri nya. Dalam *Pre processing* ini ada 3 tahapan yang akan dilakukan yaitu :

1. Resample dan Convert to mono

   Data yang semula adalah kanal *stereo* dan memiliki frekuensi sampling sebesar 44100 Hz dirubah menjadi kanal *mono* dengan cara mengubah data yang dua kanal menajdi satu kanal atau satu kolom matriks datanya saja dengan frekuensi sampling sebesar 8000 Hz. Tahap ini berguna untuk mempersingkat waktu kerja sistem karena mengolah data yang lebih sedikit. Hal ini tidak menghilangkan informasi karena matriks data kolom satu tidak jauh berbeda dengan matriks kolom dua

2. Normalisasi dan Cropping

   Sinyal suara manusia merupakan sinyal analog yang tidak terbatas oleh waktu. Sinyal suara tersebut juga memiliki besaran amplitude yang berbeda-beda untuk masing-masing karakter suara manusia. Oleh karena itu sebelum diproses lebih jauh maka sinyal tersebut di normalisasi terlebih dahulu.

Normalisasi adalah membagi nilai masing-masing amplitudo dengan nilai maksimumnya. Hal ini dilakukan agar rentang dari semua amplitudo untuk semua karakter suara menjadi sama yaitu -1 sampai 1, sehingga dalam proses selanjutnya tidak dipengaruhi oleh nilai amplitudo yang sangat besar atau sangat kecil. Selanjutnya dilakukan cropping diawal suara dengan menggunakan nilai *threshold*. Sehingga suara diam yang terjadi di awal suara akan dihilangkan.

3.1.3 Ekstraksi Ciri

Dalam tahap ini digunakan DWT dan LPC per *frame* untuk mengambil nilai ciri per frame pada *input* sinyal. Vektor ciri yang diperoleh adalah hasil rata-rata dari ciri perframe sehingga menghasilkan satu vector cirri yang memliki ukuran yang sama walaupun jumlah frame berbeda pada panjang suara yang berbeda. Hal tersebut dapat dilakukan dengan beberapa tahap berikut ini:

1. Tentukan ukuran window dalam milisekon. Contoh: 40ms.
2. Hitung banyak data sampel dalam satu frame dengan rumus:

   $X_{data} = (\text{Fs})(\text{ms})\ w_n$ (3.1)

   ket.
   w_n= window
   Fs = Frekuensi Sampling
   Contoh: X_{data}= (0.04) (8000) = 320 (data sampel)
3. Hitung banyak frame dengan rumus:

   $$\text{banyak frame} = \frac{\text{panjang data}}{X_{data}}$$ (3.2)

   Contoh Berikut apabila duarasi suara 10 detik:

   banyak frame= 800000/320 = 250 frame
4. Untuk Setiap Frame akan dilakukan proses:
   a. DWT dengan level tertentu
   b. Masing-masing subband dari DWT dilakukan pengambilan koefisien LPC.

5. Terkahir diambil rata-rata cirri tiap frame sehingga menghasilkan 1 vektor ciri.

3.2 Simulasi

Sistem yang telah dibuat akan diuji parameter-parameternya dengan beberapa data uji. Kemudian akan diukur tingkat akurasi sistem dalam melakukan deteksi emosi serta waktu kerja sistem.

Untuk perhitungan akurasi digunakan rumus sebagai berikut:

$$\text{Akurasi} = \frac{\text{jumlah data benar}}{\text{jumlah } \square\square\square} \text{ x } 100\% \quad (3.3)$$

IV. ANALISIS DAN PENGUJIAN

Pada bab ini dilakukan beberapa pengujian terhadap sistem yang telah dirancang. Setelah dilakukan pengujian maka hasil pengujian tersebut dianalisis dan disimpulkan hasilnya.

Untuk mengetahui performansi sistem yang telah dirancang, maka dilakukan pengujian terhadap sistem dengan beberapa skenario pengujian yaitu:

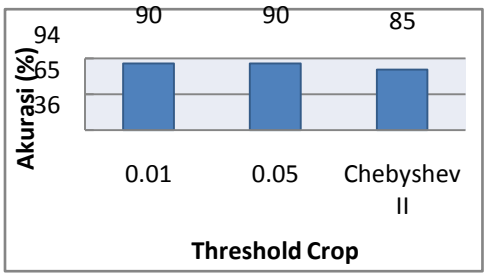
1. Pengujian dan analisis pengaruh Threshold Crop pada preprocessing terhadap akurasi *output* sistem.
2. Pengujian dan analisis pengaruh ukuran frame dan overlap terhadap akurasi *output* sistem.
3. Pengujian dan analisis pengaruh level DWT dan orde pada LPC terhadap akurasi *output* sistem.
4. Pengujian dan analisis pengaruh nilai iterasi pada pelatihan HMM terhadap akurasi *output* sistem.

4.1 Pengaruh Threshold Crop Terhadap Akurasi *Output* Sistem

Untuk menganalisis pengaruh threshold crop terhadap akurasi *output* sistem terdapat beberapa skenario yang diujikan. Dalam skenario ini dilakukan pengujian tiga nilai threshold yaitu 0.01, 0.05, dan0.1. Dalam pengujian digunakan data latih sebanyak 10 data suara dan 5 data suara tiap-tiap kelas emosi sebagai data uji sehingga total data latih sebanyak 40 data suara dan total data uji sebanyak 20 data suara . Dari hasil pengujian, dilakukan analisis akurasi menggunakan rumus 3.3.

Tabel 4.1Pengujian Threshold Crop

Threshold	AkurasiUji	WaktuAmbilCiriLatih	Waktu Proses Rata-Rata
0.01	90	78.92	3.85
0.05	90	72.04	3.60
0.1	85	68.66	3.29



Gambar 4.1 PengaruhTipe Filter TerhadapAkurasi*Output* Sistem

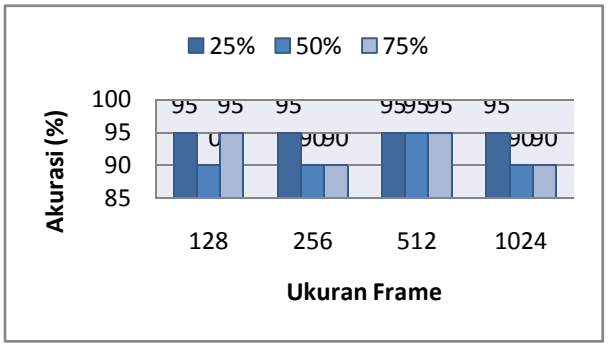
Dari gambar 4.1 didapat akurasi threshold crop terbaik yaitu 90% pada threshold crop dengan nilai 0.01 dan 0.05. Namun dipilih nilai threshold 0.05 karena waktu proses threshold 0.05 lebih cepat daripada threshold crop 0.01 walaupun menghasilkan nilai akurasi yang sama.

4.2 Pengaruh ukuran frame dan overlap Terhadap Akurasi *Output* Sistem

Untukmenganalisispengaruhukuran frame terhadapakurasi*output*sistemterdapatbeberapaskenario yang diujikan. Dalamskenarioinidilakukanpengujianempatukuran frame yaitu128, 256, 512, dan 1024 dengannilai overlap yaitu 25%, 50%, dan 75%. Dalampengujiandigunakan data latihsebanyak 10 data suaradan 5 data suaratiap-tiapkelasemosisebagai data ujisehingga total data latihsebanyak 40 data suaradan total data ujisebanyak 20 data suara. Dari hasilpengujian, dilakukananalisisakurasimenggunakanrumus 3.3.

Tabel 4.2PengujianUkuran Frame dan Overlap

Panjang Frame	Overlap	AkurasiUji	WaktuAmbilCiriLatih	Waktu Proses Rata-Rata
128	0.25	95	95.43	2.44
	0.5	90	138.87	3.61
	0.75	95	269.20	7.02
256	0.25	95	48.99	1.27
	0.5	90	70.83	1.83
	0.75	90	137.15	3.57
512	0.25	95	27.55	0.71
	0.5	95	38.21	0.98
	0.75	95	70.13	1.82
1024	0.25	95	15.94	0.41
	0.5	90	21.32	0.55
	0.75	90	37.40	0.97



Gambar 4.2 PengaruhUkuran Frame TerhadapAkurasi*Output* Sistem

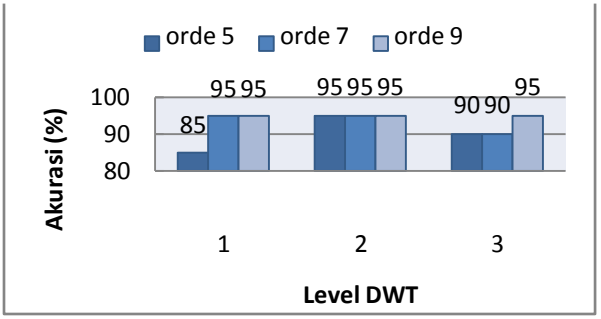
Dari gambar 4.2 didapatukuran frame terbaikyaitu512 untuksemua overlap. Semakinbesarnilaiukuran frame makaakurasiakansemakinmenurundisebabkan detail ciridarisebuah frame untuknilai frame yang terlalubesarmenjaditidakbaik. Untuk overlap semakinbesarmakaakurasisemakinbaikdikarenakanse makinbesarnilai overlap makasemakin detail nilaiciridalamsuatu frame.

4.3 Pengaruh Level DWT dan Orde pada LPC Terhadap Akurasi *Output* Sistem

Untuk menganalisis pengaruh perameter ekstraksi ciri terhadap akurasi *output* sistem terdapat beberapa skenario yang diujikan. Dalam skenario ini dilakukan pengujian tiga level DWT yaitu 1, 2, dan 3 dengan nilai orde LPC yaitu 5, 7, dan 9. Dalam pengujian digunakan data latih sebanyak 10 data suara dan 5 data suaratiap-tiap kelas emosisebagai data uji sehingga total data latih sebanyak 40 data suara dan total data uji sebanyak 20 data suara. Dari hasil pengujian, dilakukan analisis akurasi menggunakan rumus 3.3.

5 Tabel 4.3 Pengujian Parameter Ekstraksi Ciri

Level	Orde	AkurasiUji	WaktuCiriLatih	Waktu Proses Rata-Rata
1	5	85	13.7278	0.3386
	7	95	13.0597	0.3389
	9	95	13.1720	0.3394
2	5	95	13.3086	0.3442
	7	95	13.3490	0.3428
	9	95	13.3324	0.3452
3	5	90	13.5442	0.3503
	7	90	13.5386	0.3515
	9	95	13.6744	0.3487



Gambar 4.3 Pengaruh Level DWT dan Orde LPC

Dari gambar 4.3 didapat level DWT terbaik yaitu 2 untuk semua orde LPC. Untuk orde LPC 5, 7,

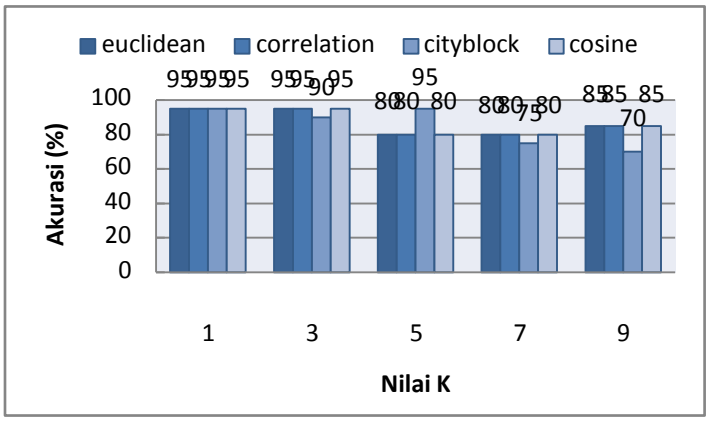
dan 9 berpengaruh terhadap banyaknya koefisien filter LPC yang didapat. Sehingga digunakan nilai orde yang paling kecil yaitu 5. Untuk orde semakin besar maka akurasi semakin baik dikarenakan semakin besar nilai orde maka semakin detail nilai ciri dalam suatu frame dan filter LPC bekerja lebih baik dalam mensintesis sinyal yang menjadi masukan LPC.

4.4 Pengaruh Nilai k dan fungsi jarak pada KNN Terhadap Akurasi *Output* Sistem

Untuk menganalisis pengaruh parameter KNN terhadap akurasi *output* sistem terdapat beberapa skenario yang diujikan. Dalam skenario ini dilakukan pengujian lima nilai k yaitu 1, 3, 5, 7, dan 9 dengan fungsi jarak yaitu euclidean, correlation, cityblock, dan cosine. Dalam pengujian digunakan data latih sebanyak 10 data suara dan 5 data suaratiap-tiap kelas emosisebagai data uji sehingga total data latih sebanyak 40 data suara dan total data uji sebanyak 20 data suara. Dari hasil pengujian, dilakukan analisis akurasi menggunakan rumus 3.3.

Tabel 4.4 Pengujian Parameter Ekstraksi Ciri

K	Jenis Distance	Akurasi Uji	Waktu Ambil Ciri Latih	Waktu Proses Rata-Rata
1	euclidean	95	17.79	0.44
	correlation	95	13.09	0.34
	cityblock	95	13.10	0.34
	cosine	95	13.01	0.34
3	euclidean	95	12.83	0.33
	correlation	95	12.74	0.33
	cityblock	90	12.81	0.33
	cosine	95	12.78	0.33
5	euclidean	80	12.79	0.33
	correlation	80	12.73	0.33
	cityblock	95	12.87	0.33
	cosine	80	12.85	0.33
7	euclidean	80	12.76	0.33
	correlation	80	12.81	0.33
	cityblock	75	12.79	0.33
	cosine	80	12.74	0.33
9	euclidean	85	12.81	0.33
	correlation	85	12.88	0.33
	cityblock	70	12.85	0.33
	cosine	85	12.93	0.33



Gambar 4.4 Pengaruh Level DWT dan Orde LPC

Dari gambar 4.4 didapat nilai k terbaik yaitu 1 untuk semua fungsi jarak. Sehingga digunakan nilai k = 1. Untuk nilai k semakin besar maka akurasi semakin turun.

V. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil implementasi, pengujian, dan analisis yang telah dilakukan, maka dapat ditarik kesimpulan sebagai berikut :

1. Perancangan simulasi deteksi emosi dari suara percakapan manusia dengan ekstraksi ciri DWT dan LPC sudah dirancang. Dari hasil pengujian dapat disimpulkan bahwa metode ini dapat digunakan untuk simulasi yang dirancang.
2. Setelah dilakukan pengujian terhadap klasifikasi 4 kelas emosi yaitu netral, marah, sedih, dan senang, akurasi tertinggi adalah 95% untuk jumlah data latih 10 tiap-tiap kelas emosi, jumah data uji 5 tiap-tiap kelas emosi, nilai threshold crop 0.05, ukuran frame 512, nilai level DWT 2, nilai k dari KNN adalah 1.

5.2. Saran

Saran yang dapat digunakan untuk perkembangan penelitian Tugas Akhir selanjutnya, yaitu :

1. Pengembangan metode ekstraksi ciri lain yang dapat meningkatkan akurasi.
2. Tugas akhir ini juga dapat dikembangkan menggunakan TMS untuk proses yang lebih *real time*.

V. DaftarPustaka

[1] Adhi, Pribadi Mumpuni dkk.2011.”Analisis Karakteristik Akustik Suara Manusia”. Bandung : Institut Teknologi Bandung.

[2] Arman,Arry Akhmad.2003.”Proses Pembentukan Karakteristik Sinyal Ucapan”. Bandung:Departemen Teknik Elektro,Institut Teknologi Bandung.

[3] Gerhard,David.2003.”*Pitch Extraction and Fundamental Frequency : History and Current Technology.* Canada :University of Regina.

[4] Mcloughlin, Ian.2009.”Applied Speech and Audio Processing : with Matlab Example”. New York: Cambridge University Press

[5] Mandasari,Miranti Indar.2008.”Studi Pengenalan Emosi Manusia Berbasis Ciri Akustik Suara Ucap”. Bandung : Institut Teknologi Bandung.

[6] Mao,Xia dkk.2007.”*Speech Emotion Recognition Based on Hybrid of HMM/ANN*”. Beijing: School of Electronic and Information Engineering Beihang University.

[7] Maulidia,Nia.2009.”Pembuatan Program Aplikasi Untuk Menampilkan Ciri Sinyal Wicara dengan Matlab”. Surabaya : Politeknik Elektronika Negeri Surabaya, Institut Teknologi Sepuluh November.

[8] Munawar, Badri.2010.”Pengidentifikasian Kata dengan Menggunakan Metode Hidden Markov Model (HMM) Melalui Ekstraksi Ciri Linear Predictive Coding (LPC)”. Bandung: Indonesian Computer University.

[9] Nilsson, Mikael.2002.”Speech recognition Using Hidden Markov Model : Performancy Evaluation In Noisy Environment”.Swedia : Departement Of Technology and Signal Processing Blekinge Institute Of Technology.

[10] Purnama, Budi Surya dan Drs. Miftahul Huda MT.2006.”Pembuatan aplikasi Untuk Pengenalan Jenis Kelamin Berbasis Sinyal Suara”. Surabaya: Politeknik Elektronika Negeri Surabaya, Institut Teknologi Sepuluh November.

[11] Rabiner, Lawrence dan Biing-Hwang Juang.1993.”Fundamental Of Speech Recognition Prentice-hall International”.Mexico

[12] Rong, Jia dkk.2008.”Acoustic Feature Selection For automatic Emotion Recognition from Speech”. Melbourne : School of Engineering and Information Technology Deakin University.

[13] Sutikyo, Prabowo Hadi.2009.”Pengolahan Suara Berdasarkan Usia dengan Menggunakan Metode K-Means”. Surabaya : Politeknik Elektronika Negeri Surabaya, Institut Teknologi Sepuluh November.

[14] Verividis, Dimitrios dan Constantine Kotropoulus.2006.”Emotional Speech Recognition : Resources, Feature, and Methods”.Thessaloniki : Artificial Intelligence and Information Analysisi Lab Departement of Informatics, Aristotle University of Thessaloniki.