

PENERAPAN KLASIFIKASI *VOICED* DAN *UNVOICED* PADA PENGENALAN TUTUR BEBERAPA KATA BAHASA INDONESIA

Heru Susanto¹⁾

¹⁾*Dosen Prodi Teknik Dirgantara, Sekolah Tinggi Teknologi Kedirgantaraan¹
hesa_3@yahoo.com*

*Teknik Dirgantara, Sekola Tinggi Teknologi Kedirgantaraan Yogyakarta
Jl.Parangtritis Km.4,5, Daerah Istimewa Yogyakarta 55143 INDONESIA*

Abstrak

*Penelitian ini bertujuan untuk mengembangkan dan menganalisis hasil akurasi kesesuaian jumlah kata proses klasifikasi *voiced* dan *unvoiced* berdasarkan algoritme baru untuk pengenalan kata pada tutur beberapa kata bahasa Indonesia. Data diperoleh dari responden melalui proses perekaman data tutur beberapa kata dari beberapa mahasiswa STTKD Yogyakarta prodi Aeronautika. Proses penelitian dimulai dengan perekaman tutur dan hasilnya disimpan dalam format wav. Selanjutnya data tutur beberapa kata diproses melalui beberapa tahap diantaranya adalah *End-Point-Detection*, *HPF 200 Hz*, *preemphasis* dan proses klasifikasi *voiced* dan *unvoiced*. Data akhir adalah hasil pemisahan tutur beberapa kata menjadi kata tunggal yang tersimpan dalam format wav. Hasil dari penelitian ini menunjukkan bahwa proses klasifikasi *voiced* dan *unvoiced* dapat digunakan sebagai klasifikasi tutur beberapa kata dengan menunjukkan hasil keberhasilan pengujian sebesar 89,7% sehingga sangat akurat dalam melakukan klasifikasi tutur beberapa kata.*

Kata kunci: *Pengenalan Tutur, Pengenalan Kata, Klasifikasi Voiced dan Unvoiced, Tutur Beberapa Kata, Tutur Bahasa Indonesia*

Pendahuluan

Sistem pengenalan tutur (*Speech Recognition*) pada aplikasi komputer sekarang sudah bukan merupakan hal yang baru. Banyak penelitian yang dilakukan untuk terus meningkatkan kemampuan pengenalan pembicaraan [1].

Pada dasarnya *speech recognition* adalah proses otomatis penggalan dan penentuan informasi linguistik yang disampaikan dengan sinyal tutur menggunakan komputer atau sirkuit elektronik. Metode pengenalan tutur yang telah diteliti selama bertahun-tahun memiliki aplikasi yang ditujukan untuk perwujudan transkripsi, sistem interaksi manusia dan komputer, verifikasi sistem keamanan, forensik dan aplikasi lainnya [2].

Selama ini pengenalan tutur masih banyak dilakukan dengan mengenali tutur kata tunggal, sehingga perlu juga dilakukan eksplorasi algoritme yang sudah ada menjadi algoritme yang dapat melakukan pengenalan tutur yang terdiri atas beberapa kata.

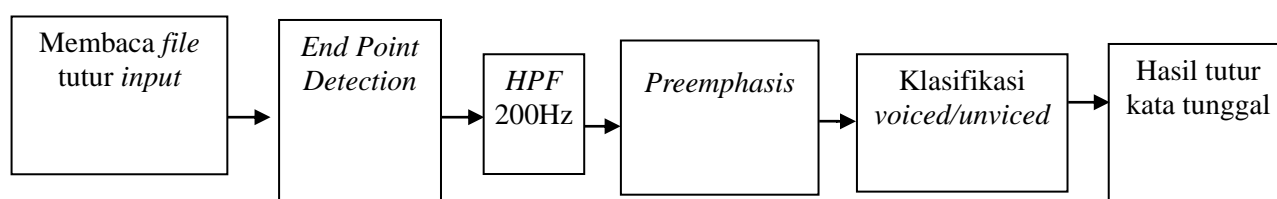
Penelitian ini akan dilakukan proses pengenalan kata pada suatu tutur yang terdiri atas beberapa kata dengan menggunakan klasifikasi *voiced* dan *voiced* berupa perhitungan nilai *zero crossing*, *energy*, dan *ink* setelah dilakukan pemrosesan awal pada tutur tersebut.

Gambaran umum dalam penelitian ini adalah melakukan proses pemrosesan awal sinyal tutur beberapa kata yang telah direkam dan disimpan dalam format wav. Pemrosesan awal dilakukan

terhadap sinyal tutur beberapa kata dimulai dengan menentukan *end-point*, melakukan tapis lolos atas dengan frekuensi *cut-off* 200 Hz, dan preemphasis. Hasil dari pemrosesan awal ini selanjutnya dilakukan klasifikasi *voiced* dan *unvoiced* sehingga akan terlokalisasi isyarat yang masuk dalam kelompok tutur (*voiced*) dan bukan tutur (*unvoiced*). Isyarat tutur ini selanjutnya akan disimpan kembali dalam format wav. Hasil dari penyimpanan isyarat ini selanjutnya jumlahnya akan di cari kesesuaiannya dengan jumlah tutur kata awal.

Metode Penelitian

Penelitian ini adalah penelitian rancang bangun sebuah sistem yang dapat mengenali kata-kata tertentu pada tutur beberapa kata menjadi bentuk kata tunggal. Sinyal tutur terdiri dari proses perekaman ini disimpan dalam bentuk *file wav*. Sehingga alur sistem yang dibangun dapat digambarkan secara diagram alir seperti terdapat pada Gambar 1 berikut.



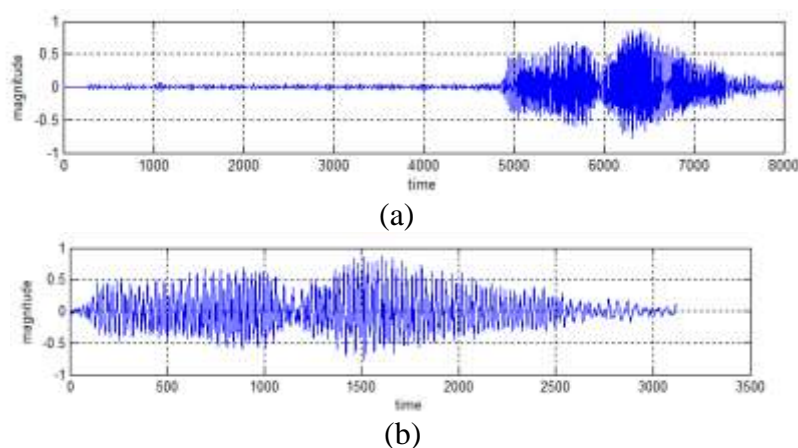
Gambar 1. Diagram alir perancangan sistem

a. Perekaman Tutur

Perekaman tutur dilakukan dengan menggunakan frekuensi *sampling* $F_s = 8000$, dan *bit* = 16. Perekaman dilakukan dua tahap yaitu perekaman berupa tutur beberapa kata dengan waktu 4 detik dan hasilnya disimpan dalam bentuk *file.wav*. Sinyal tutur inilah yang akan digunakan sebagai *template*. Sedangkan perekaman kedua adalah perekaman tutur berupa kata tunggal yang menjadi *input* dari salah satu kata yang terdapat pada sinyal tutur *template*. Begitupula hasil perekaman kata disimpan dalam bentuk *file .wav*.

b. End Point Detection

Proses *End-Point-Detection* (EPD) bertujuan untuk menentukan letak awal dan akhir dari sinyal tutur *template* maupun sinyal tutur *input* yang telah disimpan dalam bentuk *file.wav*. Gambar 2 menunjukkan hasil proses EPD dari sinyal tutur asli “kiri”.



Gambar 2.(a) Sinyal tutur asli “kiri”, (b) Tutur hasil EPD

c. HPF 200Hz

Sinyal tutur yang telah dilakukan proses *EPD* selanjutnya di buang deraunya menggunakan tapis frekuensi tinggi (*High Pass Filter/HPF*) 200Hz. Tujuannya adalah untuk membuang komponen dc, membuang *hum* frekuensi rendah, dan membuang *noise* yang mungkin ditimbulkan oleh sinyal tutur[3]. HPF 200Hz terdiri dari dua *poles* dan dua *zeros* danpersamaan fungsi alihnya dapat dinyatakan sebagai berikut.

$$H(z) = \frac{1 - 2z^{-1} + z^{-2}}{1 - 2e^{-aT} \cos(bT) z^{-1} + e^{-2aT} z^{-2}}$$

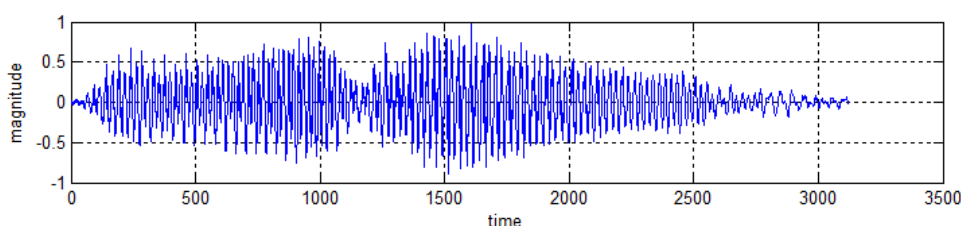
Dimana:

$$a = 130 * 2\pi$$

$$b = 200 * 2\pi$$

$$T = 10^{-4}$$

Gambar 3 menunjukkan hasil proses *HPF* 200Hz dari sinyal tutur keluaran dari proses *EPD* untuk tutur “kiri”.



Gambar 3.Sinyal tutur “kiri” setelah proses *HPF* 200Hz

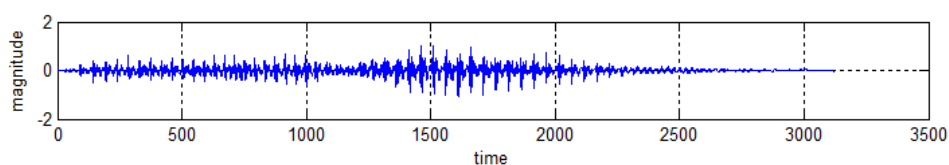
d. Preemphasis

Proses *preemphasis* adalah proses yang didesain untuk mengurangi efek tidak baik dari transmisi dan gangguan suara latar. Perhitungan proses *preemphasis* dilakukan pada saat sampel sinyal digital suara berada pada domain waktu[4].

Adapun perumusan untuk proses *preemphasis* dituliskan sebagai berikut.

$$y(n) = x(n) - \alpha x[n-1]$$

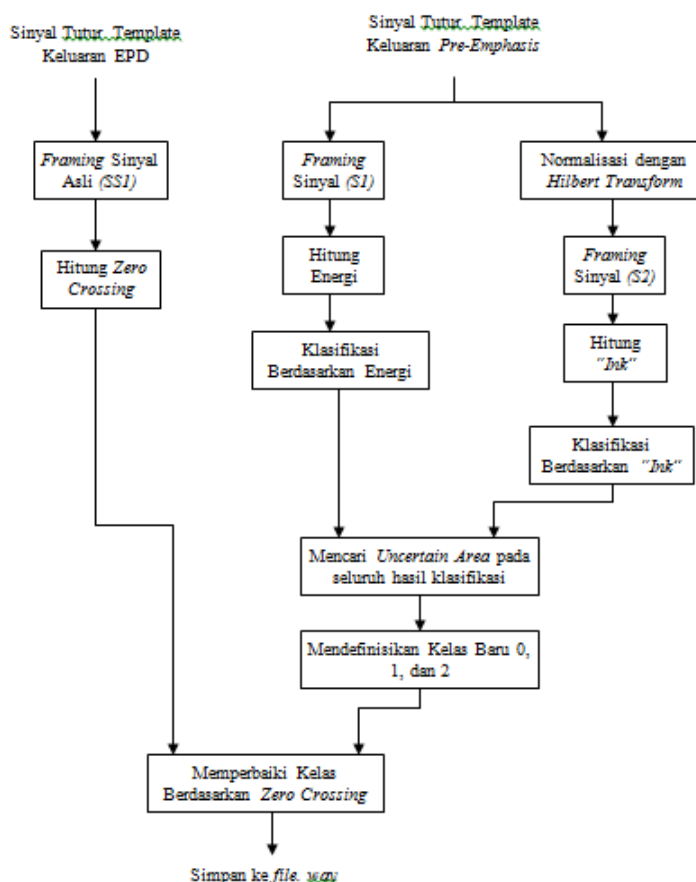
di mana, x adalah nilai sinyal digital sebelum proses *preemphasis*, y adalah nilai sinyal setelah proses *preemphasis*, dan α adalah nilai koefisien *preemphasis* yang berkisar $0.95 \leq \alpha \leq 1$. Penelitian ini menggunakan nilai $\alpha = 0.9375$. Gambar 4 menunjukkan hasil dari proses *preemphasis* dengan tutur “kiri” keluaran dari *HPF* 200Hz.



Gambar 4.Sinyal tutur “kiri” setelah proses *preemphasis*

e. Klasifikasi *Voiced/Unvoiced*

Proses segmentasi *voiced/unvoiced* bertujuan untuk melokalisasi tutur yang termasuk dalam tutur *voiced* dan *unvoiced* pada tutur beberapa kata keluaran dari *preemphasis*. Output dari segmentasi *voiced/unvoiced* adalah sinyal tutur *voiced* yang terpisah-pisah sesuai jumlah tutur kata yang terdapat pada tutur beberapa kata. Salah satu langkah yang digunakan adalah dengan mengenali sinyal tutur beberapa kata tersebut berdasarkan kategori *voiced* dan *unvoiced*. Adapun proses segmentasi *voiced/unvoiced* dapat dijelaskan berdasarkan diagram alir seperti Gambar 5 berikut ini.



Gambar 5. Diagram alir segmentasi *voiced/unvoiced*

Berdasarkan diagram alir Gambar 5 bahwa segmentasi *voiced/unvoiced* dilakukan dengan melakukan tiga macam pengukuran yaitu pengukuran energi, *ink*, dan *zero crossing rate*. Parameter energi adalah ukuran dari amplitudo sinyal pada masing-masing *framen* yang di tentukan dengan persamaan sebagai berikut.

$$E_n = \sum_{k=1}^{160} s^2(k, n)$$

Dalam sebagian besar kasus, bahwa *voiced speech* memiliki amplitudo yang lebih besar dari pada *unvoiced speech* dan termasuk bagian *silent* yang didefinisikan sebagai kondisi tanpa adanya amplitudo.^[5] Sehingga parameter energi sangat cocok digunakan untuk klasifikasi awal sinyal tutur

proses *voiced/unvoiced*.

Parameter *ink* digunakan untuk meningkatkan proses klasifikasi awal secara iteratif. Parameter *ink* ini mengukur besarnya *optical density* dari plot sinyal tutur dan mencerminkan *uncertain*, yang merupakan hasil dari jumlah frekuensi luar dan jumlah *zero-crossing*.^[3] Hasil dari parameter *ink* ditentukan dengan persamaan berikut.

$$Ink_n = \sum_{k=1}^{160} \sqrt{[y(k) - y(k-1)]^2 + 1}$$

Setelah menghitung besarnya energy dan *ink*, maka langkah selanjutnya pada proses klasifikasi ini adalah mengatur atau menentukan besarnya nilai ambang batas (*thresholding*). Untuk energi, dua *threshold* yang ditetapkan, sedangkan parameter "Ink" dibandingkan dengan satu *threshold*[5].

Pada penelitian ini hanya digunakan dua klasifikasi sinyal saja berdasarkan parameter energi untuk ditentukan besarnya nilai *thresholding* yaitu *voiced* dengan nilai 2 dan *unvoiced* dengan nilai 1.

Nilai *threshold* yang sesuai ditentukan dengan melakukan tes percobaan *codec* tutur menggunakan skema *encoding* yang berbeda untuk *voiced*, *unvoiced* dan *silent*[5].

Hasil terbaik dicapai dengan rata-rata logaritma dari energi secara keseluruhan sinyal tutur, $\overline{\log_{10} E}$, sebagai ambang batas atas untuk teks yang diucapkan terus-menerus. Namun, jika banyak terjadi kondisi *silence*, batas atas kemudian harus lebih rendah. Semakin rendah ambang batas akan mendefinisikan batas antara *unvoiced* dan *silence*, harus ditetapkan pada nilai kurang lebih $2 \sqrt{\log_{10} E}$.

Threshold untuk parameter *ink* hanya ditetapkan satu saja. Alasannya adalah untuk normalisasi dari daerah *silence* yang menyebabkan sisa *backgroundnoise* untuk mendapatkan amplitudo yang sama seperti pada daerah *voiced* dan daerah *unvoiced*. Normalisasi *background noise* disalahartikan sebagai *unvoiced* karena *uncertain*. Dengan demikian, parameter *ink* hanya membedakan antara segmen *voiced* dan *unvoiced*. Sedangkan untuk klasifikasi *silence*, hanya parameter energi saja yang diperhitungkan. Nilai dari $\overline{\log_{10}(Ink)}$ telah terbukti menjadi nilai *threshold* yang berguna untuk parameter *ink* [5].

Pada saat terjadi transisi antara sinyal tutur satu ke yang lainnya sering terjadi adanya daerah tidak tentu atau *uncertain interval*. Sehingga harus dipastikan bahwa daerah tersebut tidak masuk ke dalam parameter energi dan juga *ink* sehingga perlu dilakukan perbaikan atau koreksi. Perbaikan dapat dilakukan dengan mencari koefisien *capstral* dengan persamaan berikut.

$$c(i) = a(i) + \frac{1}{i} \sum_{k=1}^{i-1} k c(k) a(i-k), \quad 0 < i \leq N_p$$

Dengan $\alpha(i)$ adalah koefisien *LPC* dan N_p sebagai orde tapis dari *linear prediction*. Berikutnya adalah menentukan jarak *capstral* dengan persamaan berikut.

$$d_{cep} = 10 \log_{10}(e) \sqrt{2 \sum_{i=1}^{N_p} (c(i) - c'(i))^2} \quad [dB]$$

$$0 < i \leq N_p$$

Jarak *capstral* mengukur *spectral similarity* dari dua buah *frame*. Sehingga dari sini *uncertain interval* dapat dicari dengan rerata dari jarak *capstral* dan perkecualian perbaikan *uncertain interval*.

Untuk mendapatkan pemisahan yang baik maka dalam penelitian ini juga dilakukan perhitungan *zero crossing* dari sinyal tutur beberapa kata *original*. Kata *original* mengacu pada sinyal tutur yang belum mengalami *pre-processing* sehingga diambil dari keluaran *EPD*. Hasil *zero crossing* selanjutnya akan menjadi sinyal perbaikan terhadap hasil dari proses kalsifikasi energi dan *ink*.

Selanjutnya hasil perbaikan terakhir menghasilkan klasifikasi berupa *voiced* dipisah-pisahkan menjadi sinyal tutur kata yang akhirnya juga disimpan menggunakan *file.wav*.

Metode Pengumpulan Data

Data diperoleh dengan melakukan perekaman tutur dari responden berupa kalimat pembicaraan. Perekaman dilakukan secara *real-time* menggunakan laptop VAIO dengan frekuensi *sampling* 8000Hz. Besarnya waktu perekaman untuk data kalimat adalah 4 detik dan data berupa kata selama 1 detik.

Selanjutnya hasil perekaman isyarat tutur ini disimpan dalam format *file.wav*. Adapun jenis kalimat dan kata yang direkam sebagai data dari seseorang dapat dilihat pada Tabel 1 berikut.

Tabel 1. Tutur beberapa kata yang direkam sebagai *template*

No.	Jenis tutur yang direkam
1	HITAM PUTIH KUNING
2	SATU DUA TIGA EMPAT LIMA
3	MAJU MUNDUR KANAN KIRI
4	BUMI ARI PLANET
5	MERAH KUNING HIJAU
6	BAPAK IBU ANAK
7	JAUH DEKAT LAMA CEPAT
8	BANDUNG JOGJA SOLO
9	ROBOT MESIN MOBIL
10	PAGI SIANG SORE

Metode Analisis Data

Ada dua tahapan proses dalam pengenalan kata ini yaitu tahapan pengenalan dan tahap pengujian. Tahapan pelatihan adalah tahapan untuk melatih atau mengajari sistem untuk mengenali ciri umum dari sinyal tutur. Dalam hal ini digunakan ekstraksi ciri *wavelet daubechies orde 4 (db4)*. Sedangkan tahapan pengujian adalah tahapan untuk mengetahui kemampuan pengenalan kata yang dapat dilakukan oleh sistem berdasarkan tahapan pelatihan yang dilakukan. Dalam hal ini digunakan *DTW* sebagai proses pengenalan kata dengan mencari jarak minimum dari *DTW*.

Hasil dan Pembahasan

Pengukuran unjuk kerja dari sistem pengenalan kata pada tutur beberapa kata dilakukan dengan mencari persentase kesalahan pencocokan yang menyatakan probabilitas terjadinya kesalahan pada sistem yang telah dibuat.

Pada pengujian ini penulis menggunakan 3 orang responden yang masing-masing mengucapkan berupa tutur beberapa kata sebagai *input* sesuai Tabel 1. Pengujian pada proses klasifikasi *voiced* dan *unvoiced* tutur beberapa kata.

Hasil Klasifikasi *Voiced* dan *Unvoiced*

Pengujian pada proses klasifikasi *voiced* dan *unvoiced* dilakukan untuk mengetahui banyaknya kata-kata dari tutur beberapa kata yang diucapkan. Data dari tutur beberapa kata dari responden diambil dan dihitung hasil ketepatan sistemnya dalam memisahkan kata-kata yang terdapat dalam tutur beberapa kata tersebut. Persentase responden menunjukkan selisih absolut antara jumlah kata yang seharusnya terdapat pada tutur beberapa kata dari responden dengan jumlah kata yang dapat dideteksi oleh sistem. Sebagai contoh pada tutur "HITAM PUTIH KUNING" terdapat 3 kata yaitu "HITAM", "PUTIH", dan "KUNING" sehingga seharusnya pemotongannya terdapat 3 kata tersebut, tidak boleh kurang dan tidak boleh lebih.

Akurasi data uji diperoleh dengan mencari perbandingan antara data seharusnya dengan data hasil pengujian dan dikalikan dengan persentase 100% sehingga diperoleh persentase akurasi data dan berikut hasil pengujiannya seperti pada Tabel 2.

Tabel 2. Hasil pengujian klasifikasi *voiced* dan *unvoiced*

No.	Kalimat	Responden			Akurasi Data Uji (%)
		A (%)	B (%)	C (%)	
1	Pertama	33,3	66,7	100	66,7
2	Kedua	100	100	100	100
3	Ketiga	50	100	100	83,3
4	Keempat	100	100	100	100
5	Kelima	100	33,3	100	77,8
6	Keenam	100	100	100	100
7	Ketujuh	100	75	50	91,7
8	Delapan	100	100	100	100
9	Sembilan	100	100	66,7	88,9
10	Sepuluh	100	100	66,7	88,9

Sehingga keberhasilan pengujian klasifikasi *voiced* dan *unvoiced* adalah 89.7%. Sehingga dapat disimpulkan bahwa tingkat keberhasilan dari sistem klasifikasi *voiced* dan *unvoiced* dalam memotong *input* menjadi kata-kata dalam penelitian ini adalah akurat dan metode yang digunakan dalam proses klasifikasi *voiced* dan *unvoiced* dapat berjalan dengan sangat baik.

Kesimpulan

Berdasarkan uraian Pada penelitian ini telah dilakukan pengembangan algoritme berupa klasifikasi *voiced* dan *unvoiced* untuk melakukan klasifikasi tutur beberapa kata bahasa Indonesia. Hasil pengujian menggunakan 10 isyarat tutur beberapa kata dan 10 isyarat tutur kata tunggal dari 3 responden yang berbeda, Nilai persentase akurasi klasifikasi *voiced* dan *unvoiced* adalah 89,7% sehingga sangat akurat dalam melakukan segmentasi tutur beberapa kata.

Daftar Pustaka

- [1] Agustinus Noertjahyana dan Rudy Adipranata. "Implementasi Sistem Pengenalan Suara Menggunakan SAPI 5.1 dan DELPHI 5", Skripsi Teknik Informatika, Universitas Kristen Petra, 2003.
- [2] Nitin Trivedi, Dr. Vikesh Kumar, Saurabh Singh, Sachin Ahuja, Raman Chadha. "Speech Recognition by Wavelet Analysis". *International Journal of Computer Applications*, February 2011.
- [3] Atal Bishnu S and Rabiner Lawrence R. "A Pattern Recognition Approach to Voiced-Unvoiced-Silence Classification with Application to Speech Recognition", *IEEE Journal*, 1976.
- [4] Lukman Achmad, "Klasifikasi Nyamuk Berdasarkan Suaranya Dengan Metode *Mel Frequency Cepstral Coefficient* dan Jaringan Syaraf Tiruan". Thesis, Jurusan Ilmu Komputer dan Elektronika, UGM, 2013..
- [5] Hoelper C, Frankort A, Erdmann C, and Vary P. "A Novel Voiced / Unvoiced / Silence Classification Scheme for Offline Speech Coding", Aachen University of Technology, *Institute of Communication Systems and Data Processing*.
- [6] L. Chruszczyk, "Wavelet Transform in Fault Diagnosis of Analogue Electronic Circuits ", *InTech-Advances in Wavelet Theory and Their Applications in Engineering, Physics and Technology* p. 197, 2012
- [7] M. Misiti, Y. Misiti, G. Oppenheim, and J.-M. Poggi, "Wavlet Toolbox User's Guide", The MathWorks, Inc., 1996.
- [8] Mohan Bhadragiri Jagan and N Ramesh babu. "Speech Recognition using MFCC and DTW, VIT University Vellore India.