Sistem Speech Reqocnition dengan Metode Linier Predictive Coding (LPC) dan Hidden Markov Model (HMM) Menggunakan Matlab untuk Identifikasi Pembicara

Andriana, S.T., M.T.
Fakultas Teknik, Jurusan Elektro
Universitas Langlangbuana
Jl. Karapitan 116, Bandung
nana_zoel@yahoo.com

Zulkarnain, S.T., M.T.
Fakultas Teknik, Jurusan Elektro
Universitas Langlangbuana
Jl. Karapitan 116, Bandung
zoel_89@yahoo.com

Abstrak— Sistem *Speech Recognition* (Pengenalan Ucapan) adalah suatu proses pengenalan menggunakan *MatLab* yang dapat mengidentifikasi seseorang dengan mengolah suaranya. Tujuan dasar dari penelitian adalah untuk mengenali dan mengklasifikasikan ucapan-ucapan dari orang yang berbeda. Identifikasi untuk mengetahui siapa yang mengucapkan ucapan tersebut yaitu dengan cara mencocokkan karakteristik ucapan yang ada di dalam basisdata dengan ucapan masukan. Karakteristik ucapan dapat dibedakan melalui ekstraksi dengan suatu teknik pengkodean, berupa frekuensi dasar (*pitch*), *formant* dan *energy*. Teknik pengkodean yang umum digunakan oleh *National Institute of Standar Technology* (NIST) dalam pengekstraksian sinyal ucapan adalah LPC (*Linier Predictiv Coding*). Identitas unik dari setiap orang dapat dikenali menggunakan model statistik *Hidden Markov Model* (HMM).

Kata kunci— *Speaker Recognition, feature extraction, LPC (Linier Predictiv Coding), Statistic Model, Hidden Markov Model (HMM).*

I. PENDAHULUAN

Pengenalan ucapan adalah suatu proses pengenalan untuk mengetahui siapa yang mengucapkan suatu sinyal informasi dengan mencocokkan karakteristik ucapan yang ada di dalam basis data dengan ucapan masukan. Karakteristik ucapan dapat dibedakan melalui ekstraksi dengan suatu teknik pengkodean. Teknik pengkodean yang umum digunakan oleh *National Institute of Standar Technology* (NIST) dalam pengekstraksian sinyal ucapan adalah LPC (*Linier Predictive Coding*). Sedangkan Pengenalan Pola suara seseorang digunakan metode *Hidden Markov Model* (HMM). Dalam metode ini suara dianggap sebagai parameter acak yang dapat diperkirakan untuk dianalisa dan dicari nilai kemungkinan yang maksimum untuk proses pengenalan. Untuk Algoritma LPC dan HMM digunakan software *MatLab*.

Sistem *Speech Recognition* atau Sistem Pengenalan Ucapan adalah sistem yang berfungsi untuk mengubah bahasa lisan menjadi bahasa tulisan. Masukan sistem adalah ucapan manusia, selanjutnya sistem akan mengidentifikasikan kata atau kalimat yang diucapkan dan menghasilkan teks yang sesuai dengan apa yang diucapkan.

Sistem Speech Recognition dapat mengenali seluruh kata dalam suatu bahasa dan melakukan pengenalan untuk setiap unit bunyi pembentuk ucapan (*fonem*), selanjutnya mencoba mencari kemungkinan kombinasi hasil ucapan yang paling dapat diterima, dengan menggunakan metode *Hidden Markov Model (*HMM).

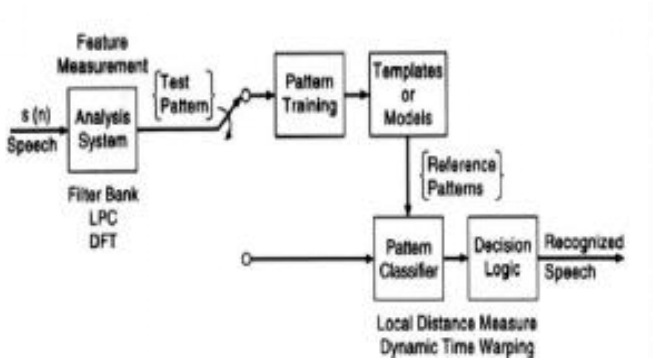
Sistem yang lebih sederhana adalah sistem yang hanya dapat mengenal sejumlah kata yang jumlahnya terbatas. Sistem ini biasanya lebih akurat dan lebih mudah dilatih, tetapi tidak dapat mengenal kata yang berada di luar kosa kata yang pernah diajarkan. Sistem ini menggunakan software *Matlab* sebagai pelengkap proses pelatihan dan pengenalan *fonem*.

Pengenal pengucap adalah suatu proses pengenalan untuk mengetahui siapa yang mengucapkan sinyal informasi tersebut dengan mencocokkan karakteristik ucapan yang ada di dalam basis data dengan ucapan masukan. Karakteristik ucapan dapat dibedakan melalui ekstraksi dengan suatu teknik pengkodean. Teknik pengkodean yang umum digunakan dalam peng-ekstraksian sinyal ucapan adalah LPC (*Linear Predictive Coding*).

Metode pengenal pengucap dapat dibagi menjadi metode tidak berdasarkan teks (*Text-Independent*) dan metode berdasarkan teks (*Text-Dependent*). Dalam sistem pengenal pengucap tidak berdasarkan teks, ucapan masukan yang diucapkan oleh pengucap tidak harus sama dengan ucapan yang ada di dalam basis data. Sebaliknya, dalam sistem pengenal pengucap berdasarkan teks selain pengucap yang harus dikenali, ucapan masukan juga harus sesuai dengan ucapan yang ada di dalam basis data.

Sistem pengenal pengucap tidak berdasarkan teks yang dipaparkan di dalam Penelitian ini dapat diaplikasikan untuk sistem identifikasi orang yang berbicara pada telepon dan apa yang dibicarakannya. Sistem tidak hanya memeriksa siapa yang bicara tetapi juga memeriksa kebenaran kata yang diucapkannya. Sistem ini diharapkan akan mengenali suara, kemudian hasil dari pengenalan suara tersebut digunakan sebagai referensi untuk identifikasi pengenalan suara pengucap dan apa yang diucapkan.

Pada penelitian ini dibuat program simulasi dengan menggunakan bantuan *GUI Matlab* untuk mengenali dan menampilkan bentuk sinyal wicara seseorang. Bentuk sinyal wicara yang dikenali adalah energi (*power*), *pitch*, dan *formant*. Program ini akan membaca file suara dalam bentuk ekstensi *.wav dan menampilkannya berdasarkan *energy* (*power*), *pitch*, dan *formant*. Selain itu, pada proyek ini juga dibuat sebuah simulasi yang akan mengidentifikasi atau mengenali seseorang berdasarkan rekaman suara yang telah disimpan sebelumnya. Metode yang digunakan dalam simulasi tersebut adalah *autocorrelation* dan *Fast Fourier Transform* (FFT).



Gambar 1. Diagram Blok Sistem Pengenalan Ucapan

II. TINJAUAN PUSTAKA

A. *Linier Predictive Coding (LPC)*

Tujuan utama pemrosesan sinyal dalam sistem pengenal pengucap adalah untuk mendapatkan ciri yang sesuai dengan kandungan bahasa untuk semua pengucap, berbeda dengan sistem pengenal pengucap yang bertujuan untuk mendapatkan ciri yang sesuai dengan karakteristik masing-masing pengucap, yang bebas dari pengaruh kata yang diucapkan.

Analisis LPC yang biasanya digunakan pada sistem pengenalan ucapan juga bisa digunakan pada sistem pengenal pengucap. Langkah-langkah analisis LPC untuk mendapatkan koefisien LPC pada pengenal pengucap berdasarkan referensi adalah sebagai berikut :

1. *Preemphasis*

Preemphasis digunakan untuk mendatarkan spektral sinyal dan meningkatkan keaslian sinyal pada pemrosesan sinyal yang selanjutnya. Sistem *preemphasis* yang umum digunakan adalah sistem orde satu :

$H(z) = 1 - \tilde{a}z\text{-}1,\ 0{,}9 \mathbb{C} a \mathbb{C} 1$

Keluaran dari rangkaian *preemphasis* ,ŝ(n), adalah :

$\hat{s}(n) = s(n) - \tilde{a}s(n\text{-}1)$

Besarnya ã yang umum digunakan adalah 0,95 (untuk penggunaan yang teliti, biasanya digunakan nilai ã = 15/16 = 0,9375).

Perubahan sinyal yang telah dilakukan Preemphasis filter menggunakan LPF.

2. *Frame Blocking*

Setelah dipreemphasis, sinyal kemudian dipotong-potong dalam suatu frame dengan satu framenya terdiri dari N-sampel, dan setiap frame yang berdekatan berjarak M-sampel. Sinyal keluaran dari preemphasis ŝ(n), dipotong-potong kedalam suatu frame dengan persamaan :

$xl(n) = \hat{s}(Ml + n), n = 0,1,\ldots,N\text{-}1 ;\ l = 0,1,\ldots,L\text{-}1.$

dengan L merupakan jumlah frame. Besar N dan M adalah 300 dan 100 untuk sampling rate sinyal sebesar 6,67 kHz, yang berarti framenya sepanjang 45 ms dengan jarak pemisah antar frame adalah 15 ms.

3. *Windowing*

Windowing (penjendelaan) digunakan untuk menapis sinyal menjadi nol pada awal dan akhir *frame*. Setelah melalui proses *framing*, sinyal kemudian melewati proses *windowing* dengan persamaan :

$\tilde{x}_l(n) = x_l(n)w(n)$,0 ℂ n ℂ N-1

dengan *window* yang biasa digunakan adalah *Hamming window* yang mempunyai bentuk umum :

$w(n) = 0{,}54 - 0{,}46 \cos$,0 ℂ n ℂ N-1

Window lain yang juga bisa digunakan adalah *Hanning window*.

4. Analisis Autokorelasi

Setiap frame yang telah melalui *windowing*, kemudian melalui proses autokorelasi :

$r_l(m) = \sum_{n=0}^{N-1-m} \tilde{x}_l(n)\,\tilde{x}_l(n+m)$, m = 0,1,….,p. dengan nilai autokorelasi tertinggi p, adalah orde LPC. Nilai p biasanya adalah antara 8 sampai 16.

5. Analisis LPC

Proses selanjutnya adalah analisis LPC, yang mengubah setiap frame autokorelasi p+1 ke koefisien LPC.

B. *Hidden Markov Model (HMM)*

HMM adalah sebuah sistem pengenalan suara yang pada dasarnya mengasumsikan bahwa sinyal suara merupakan realisasi dari beberapa kode pesan yang berupa satu atau beberapa urutan simbol. Untuk mendapatkan simbol simbol itu, sinyal suara pertama kali diubah menjadi urutan vektor parameter diskrit dengan space yang sama. Vektor parameter diskrit ini diasumsikan membentuk representasi yang tepat terhadap sinyal suara dengan selang waktu selama kurang lebih 10 ms untuk satu vektornya, karena sinyal suara dapat

dianggap stasioner. Walaupun tidak sepenuhnya benar, tetapi hal itu adalah tafsiran yang rasional.

Dasar dari pengenalan adalah pemetaan antara rangkaian vektor suara dan rangkaian simbol yang diinginkan. Dua hal yang menjadi masalah yaitu :

1. Pemetaan dari simbol menjadi suara tidak satu per satu karena perbedaan simbol yang mendasar dapat mempengaruhi bunyi suara yang hampir sama.
2. Batasan antar simbol tidak dapat diidentifikasikan secara langsung pada sinyal suara. Oleh karena itu adalah tidak mungkin menganggap sinyal suara sebagai rangkaian gabungan pola pola statis.

Masalah kedua dapat diatasi dengan membagi sinyal menjadi simbol yang dikenali terpisah (*word isolated recognition*). Secara umum permasalahan yang terjadi pada sistem pengenalan suara seperti di atas dapat diselesaikan dengan menggunakan metode *hidden markov model* ini.

C. Vektor Quantization (VQ)

VQ merupakan salah satu metode pencocokan ciri (*feature matching*). VQ melakukan proses pemetaan vektor dari vektor yang berjumlah banyak menjadi vektor dengan jumlah tertentu, sehingga data yang ada dikompres dan tetap akurat. Vektor yang didapat dalam proses pengenal pengucap ini merupakan vektor ciri dari masing-masing pengucap yang terdapat pada basisdata. Dengan proses VQ, akan diperoleh representasi dari vektor ciri masing-masing pengucap dengan jumlah vektor yang lebih sedikit, vektor itu disebut sebagai *codebook* dari tiap-tiap pengucap. *Codebook* ini terdiri dari bebeberapa buah vektor *codeword* yang merupakan *centroid* dari sekumpulan vektor ciri.

Perhitungan jarak penyimpangan dilakukan dengan membandingkan antara koefisien *cepstrum* dari sinyal ucapan yang akan dikenali dan *codebook* dari tiap-tiap pengucap pada basisdata.

Jarak *Euclidean* dari ke dua *vector* dapat dituliskan dengan persamaan.

$$d_E(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{i=1}^{\dim} (x_i - y_i)^2}$$

Persamaan di atas merupakan persamaan umum untuk menghitung jarak *Euclidean* yaitu persamaan yang digunakan untuk mengetahui jarak antara antara dua vektor.

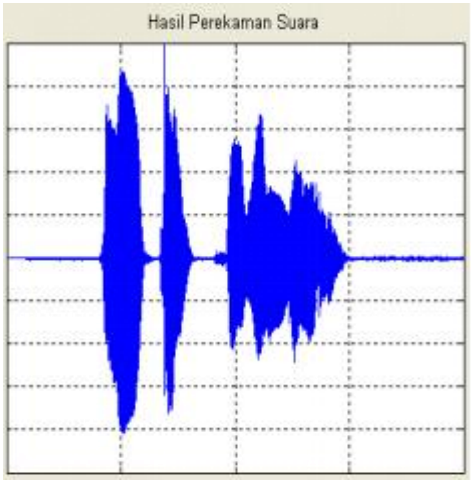
Hasil Akhir dari membandingkan speaker 1 dan 2 di data base, dengan speaker 1 dan 2 pada proses pengujian adalah sbb:

Speaker 1 matches with speaker 1

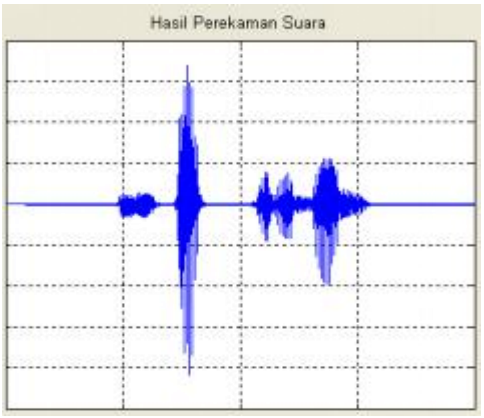
Speaker 2 matches with speaker 2

III. HASIL PENGUJIAN

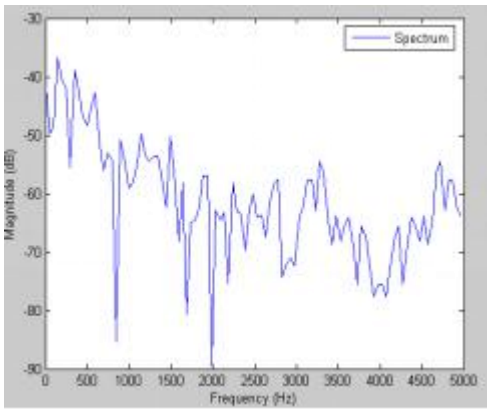
Tampilan Awal Proses Pengujian



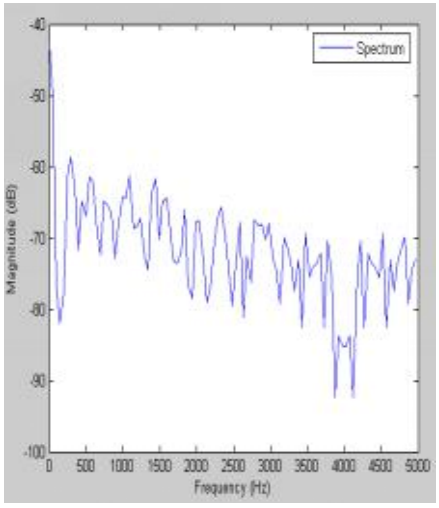
Gambar 2. *Hasil rekaman suara speaker 1*



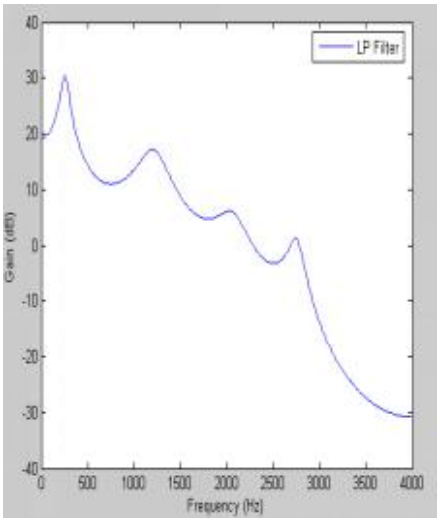
Gambar 3. *Hasil rekaman suara speaker 2*



Gambar 4. *Hasil spektrum speaker 1*



Gambar 5. *Hasil spektrum speaker 2*



Gambar 6. *Hasil formant speaker 1*

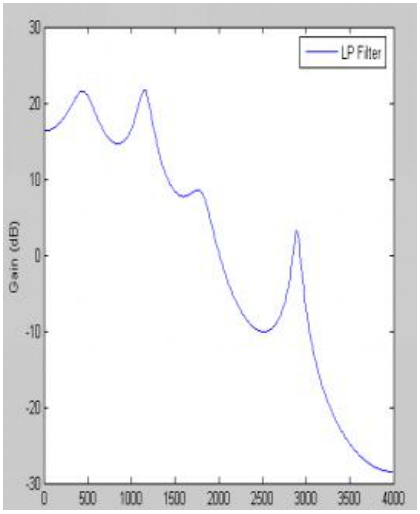
Formant 1 Frequency 183.5

Formant 2 Frequency 687.6

Formant 3 Frequency 1375.3

Formant 4 Frequency 2397.6

Formant 5 Frequency 2848.7



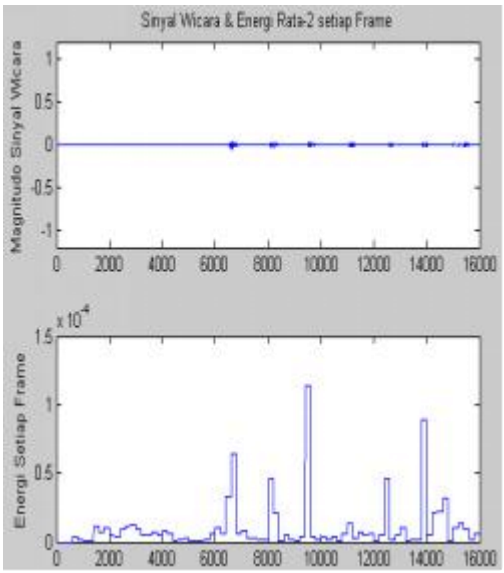
Gambar 7. *Hasil formant speaker 2*

Formant 1 Frequency 453.2

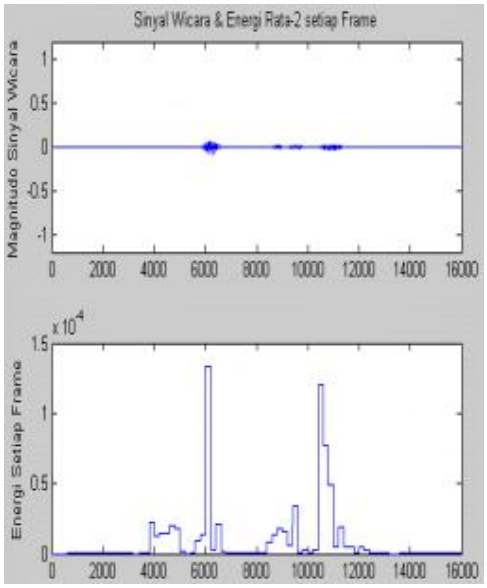
Formant 2 Frequency 1155.9

Formant 3 Frequency 1808.7

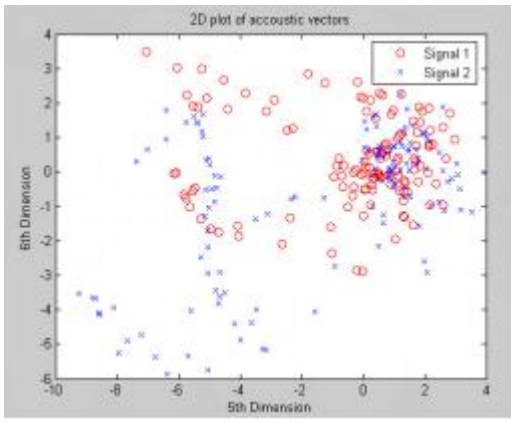
Formant 4 Frequency 2895.3



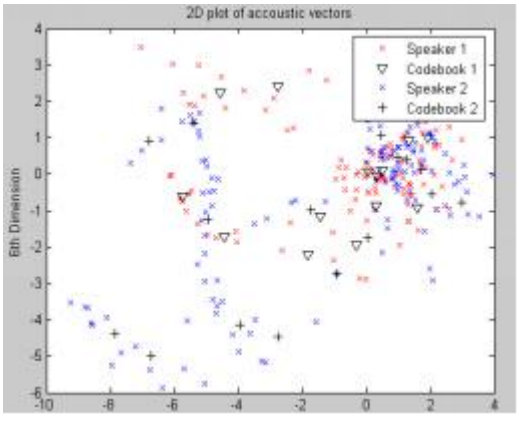
Gambar 8. *Sinyal Wicara dan Energi Rata-rata setiap Frame 1*



Gambar 9. *Sinyal Wicara dan Energi Rata-rata setiap Frame 2*



Gambar 10. *2D plot accoustic vectors 1*

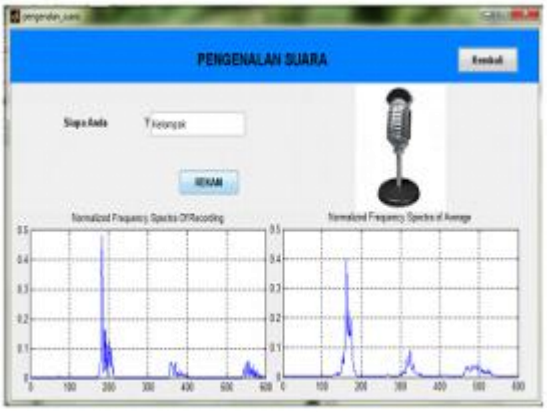


Gambar 11. *2D plot accoustic vectors 2*

Hasil Akhir dari membandingkan speaker 1 dan 2 di data base, dengan speaker 1 dan 2 pada proses pengujian adalah sbb:

Speaker 1 matches with speaker 1

Speaker 2 matches with speaker 2



Gambar 12. *Pengenalan Suara*

Apabila cepstrum suara hasil rata-rata perekaman pada data base dan hasil pengetesan sama yang dibuktikan oleh kemungkinan jarak euclidean yang dekat, maka identitas pengucap dapat dikenali.

IV. KESIMPULAN

Dari pengujian sistem yang telah dilakukan dapat diambil beberapa kesimpulan sebagai berikut :

1. Setiap orang mempunyai karakteristik suara yang berbeda dalam pitch, formant dan energy.

2. Perbedaan karakteristik ini dapat digunakan untuk mengenali identitas seseorang dengan bantuan software MatLab menggunakan metode LPC dan metode HMM.

3. Apabila cepstrum suara hasil rata-rata perekaman pada data base dan hasil pengetesan sama yang dibuktikan oleh kemungkinan jarak euclidean yang dekat, maka identitas pengucap dapat dikenali.

DAFTAR PUSTAKA

[1] Lawrence, Rabiner, & Biing-Hwang, Juang. (1999). Fundamentals of speech recognition. Beijing: Prentice-Hall International, Inc.

[2] Lawrence, R, Rabiner. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE, VOL.77, NO 2, February.

[3] Zhou, Haitao. (2009). Design and Implementation of Speech Recognition System

Based on Field Programmable Gate Array. The research is financed by Applied Program of Basic Research of Tianjin (08JCYBJC14700)

[4] Woodland, P.C., Odell, J.J., Valtchev, V. & Young, S.J. Large vocabulary continuous speech recognition using HTK. ICASSP '94, 2, pp.125-128.

[5] Young, S., A review of large-vocabulary continuous-speech recognition. IEEE Signal Processing Magazine, 13, No.5, 1996, pp.45-57.

[6] S J Melnikoff, S F Quigley & M J Russell, Implementing a Simple Continuous Speech Recognition System on an FPGA, Proceedings of the 10

th Annual IEEE Symposium on Field-Programmable Custom Computing Machines (FCCM'02), 2002.

[7] Lintang Y.B., Adaptasi Sistem Pengenalan Ucapan Bahasa Inggris ke dalam Sistem Pengenalan Ucapan Bahasa Indonesia Baku Menggunakan Pendekatan Bootstraping Termodifikasi.

[8] Arman.A.A. Proses Pembentukan dan Karakteristik Sinyal Ucapan. Jakarta.2006.