

# SISTEM *SPEAKER RECOGNITION* (PENGENAL PENGUCAP) UNTUK MENCARI KARAKTERISTIK UCAPAN SESEORANG DENGAN METODE *MEL FREQUENCY CEPTRUM COEFFICIENT (MFCC)* MENGGUNAKAN SOFTWARE *MATLAB*

Andriana, ST., MT.

Dosen Fakultas Teknik Elektro Universitas Nurtanio Bandung

[nana\\_zoel@yahoo.com](mailto:nana_zoel@yahoo.com)

## **Abstrak**

Sistem *Speaker Recognition* (Pengenal pengucap) 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 peng-ekstraksian sinyal ucapan adalah *Mel Frequency Cepstral Coefficients (MFCC)*. Identitas unik dari setiap orang dapat dikenali menggunakan model statistik *Hidden Markov Model (HMM)*.

## **Abstract**

System Speaker Recognition (Speaker Recognition) is a recognition process using MatLab that can identify a person with voice processing. The basic purpose of the study was to identify and classify the utterances of different people. Identification to know who is saying these words is by matching the characteristics of speech in the database with the input utterance. Characteristics of speech can be distinguished by extraction with a coding technique, a fundamental frequency (*pitch*), *formant* and *energy*. Coding techniques commonly used by the National Institute of Standards Technology (NIST) in the lawyer-ekstraksian speech signal is Mel Frequency Cepstral Coefficients (MFCC). Unique identity of each person can be identified using a statistical model of hidden Markov model (HMM).

**Keyword** : *Speaker Recognition, feature extraction, Mel Frequency Cepstral Coefficients (MFCC), Statistic Model, Hidden Markov Model (HMM).*

### Pendahuluan

Pengenalan pengucap 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 peng-ekstraksian sinyal ucapan adalah *Mel Frequency Cepstrum Coeffisiens (MFCC)*. 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 diucapkan oleh pengucap tidak harus sama dengan ucapan yang ada di dalam basis data. Sebaliknya, dalam sistem pengenalan pengucap Pada penelitian ini digunakan metode tidak berdasarkan teks.

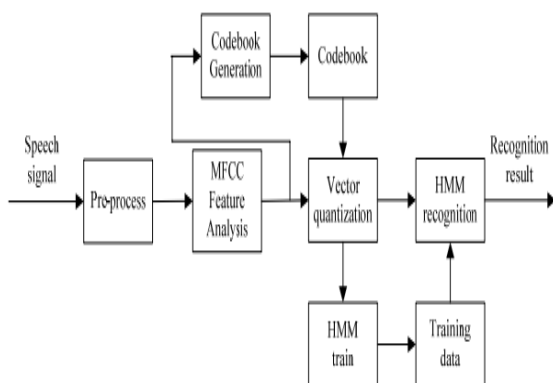
Pengenalan pengucap (*speaker recognition*) adalah sebuah proses pengenalan untuk mengetahui pengucap yang memberikan sinyal informasi berupa sinyal suara. Sinyal ucapan berisi beberapa karakteristik yang spesifik untuk setiap individu, yang beberapa diantaranya tidak dipengaruhi oleh pesan linguistik yang terkandung dalam suatu ucapan. Perbedaan karakteristik ucapan itulah

kemungkinan yang maksimum untuk proses pengenalan. Untuk Algoritma *MFCC* dan *HMM* digunakan software *MatLab*.

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

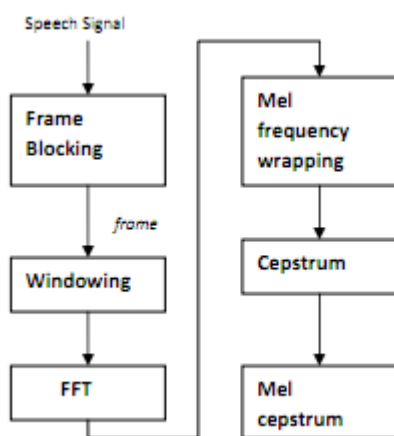
berdasarkan teks selain pengucap yang harus dikenali, ucapan masukan juga harus sesuai dengan ucapan yang ada di dalam basis data. yang menjadi dasar pengenalan pengucap melalui ucapannya. Teknik ini mengenali suara dari pengucap untuk memverifikasi identitas pengucap dan untuk layanan akses kontrol seperti *telephone-banking, voice dialing, telephone shopping, database access services*, layanan informasi, pesan suara, dan kontrol keamanan untuk area informasi yang sangat rahasia. Gambar 1. menunjukkan diagram blok proses identifikasi pengucap menggunakan *MFCC* dan *HMM*.

Gambar 1. Diagram blok proses identifikasi pengucap.



**1. Mel Frequency Cepstrum Coefficiens (MFCC)**

Penelitian dimulai dengan Mel Frequency Cepstrum Coefisien (MFCC) untuk feature analysis pengucap sampai dihasilkan mel cestrum seperti pada gambar 2. di bawah ini:



Gambar 2. Proses ekstraksi sinyal suara dengan metode MFCC

Para metode MFCC, proses ini dibagi menjadi lima tahap atau blok. Pada bagian frame bloking bentuk gelombang ucapan dibagi menjadi beberapa frame, 1 frame sekitar 30

milidetik. Windowing meminimalkan diskontinuitas dari sinyal pada awal dan akhir dari setiap frame ke nol. Blok FFT mengubah setiap frame dari domain waktu ke domain frekuensi. Di blok mel frequency wrapping, sinyal diplot terhadap Mel spektrum untuk meniru pendengaran manusia. Penelitian telah menunjukkan bahwa pendengaran manusia tidak mengikuti skala linier melainkan skala Mel-spektrum yang merupakan jarak linier di bawah 1000 Hz dan skala logaritmik di atas 1000 Hz. Pada langkah terakhir, plot Mel spektrum diubah kembali ke domain waktu dengan menggunakan berikut persamaan:

$$Mel(f) = 2595 * \log_{10}(1 + f / 700)$$

Matriks yang dihasilkan disebut sebagai Mel-Frequency cepstrum Koefisien. Spektrum ini memberikan cukup sederhana namun unik representasi dari sifat spektral dari sinyal suara yang merupakan kunci untuk mewakili dan mengenali karakteristik suara dari speaker.

**2. 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 diidentifikasi 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.

### 3. 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 pengenalan 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 beberapa 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.

$$d = \sqrt{\sum_{i=1}^{\text{dim}} (x_i - y_i)^2}$$

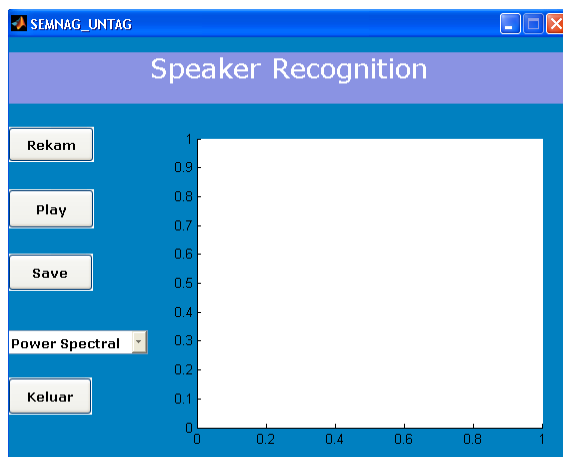
Persamaan di atas merupakan persamaan umum untuk menghitung jarak Euclidean yaitu persamaan yang digunakan untuk mengetahui jarak antara dua vektor.

#### 4. Hasil Pengujian

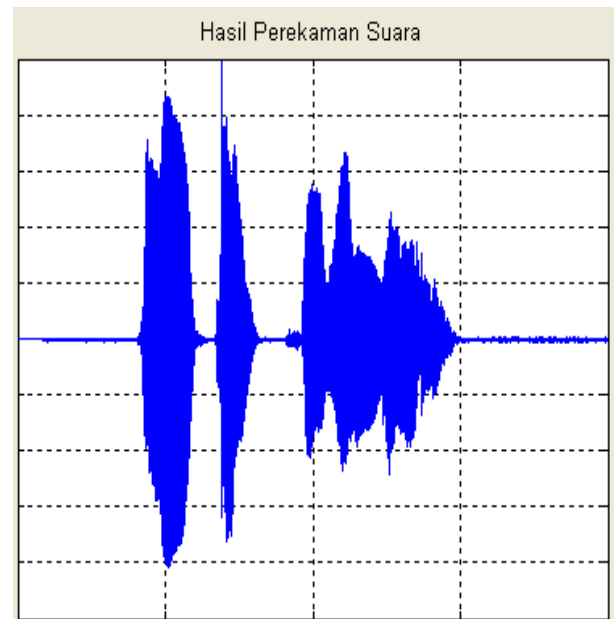
Pada Proses pengujian, grafik akan muncul pada interface seperti gambar 3. di bawah ini.

Proses yang akan diujikan adalah sbb:

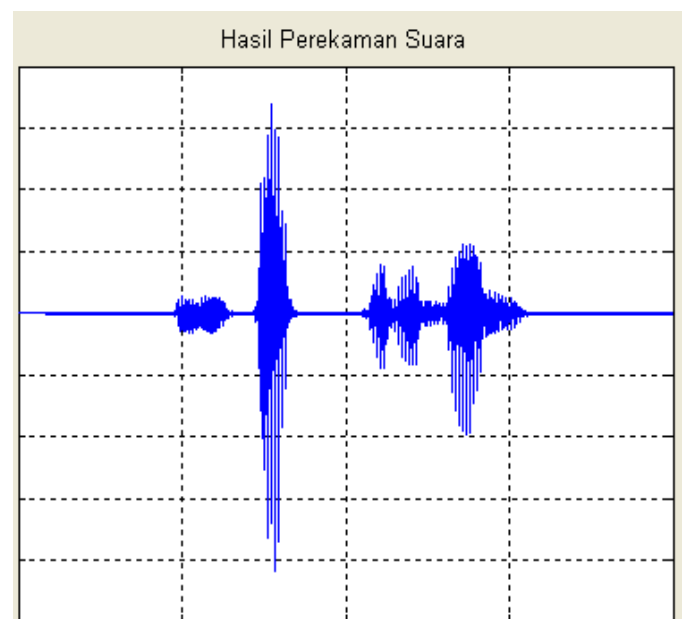
- Proses perekaman speaker 1 dan 2 dengan mengucapkan kata “Untag Cirebon”.
- Proses penyimpanan file .wav
- Proses ekstraksi sinyal speaker 1 dan 2, grafik output berupa pitch, formant, dan energy
- Proses membandingkan speaker 1 dan 2



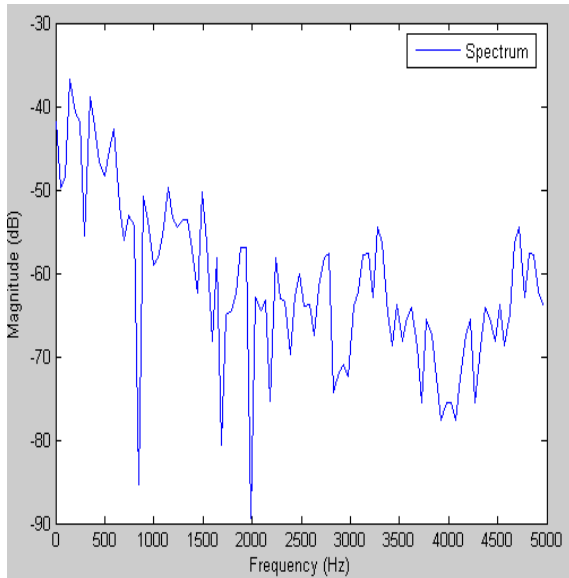
Gambar 3. Interface untuk proses pengujian speaker 1 dan 2



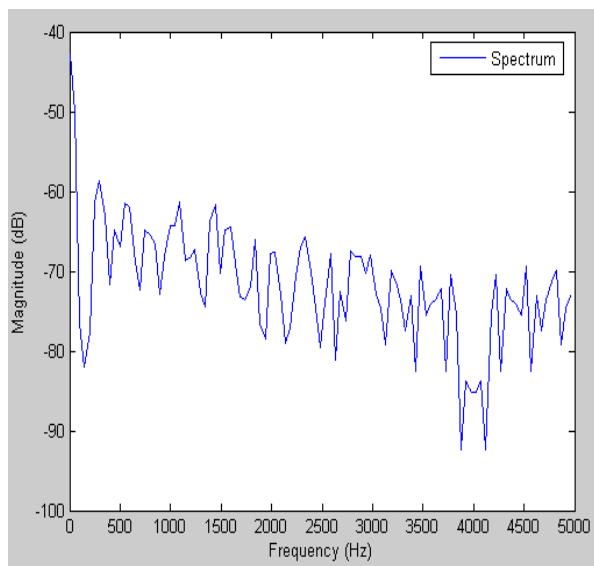
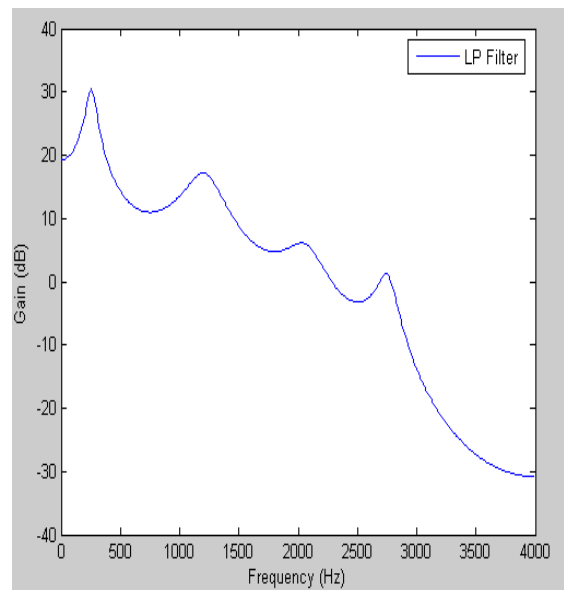
Gambar 4. Hasil rekaman suara speaker 1 Pada pengucapan kata “Untag Cirebon”



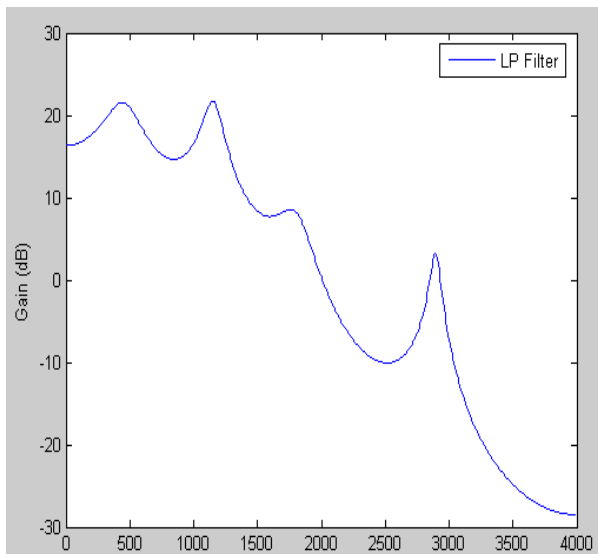
Gambar 5. Hasil rekaman suara speaker 2 Pada pengucapan kata “Untag Cirebon”



Gambar 6. Hasil spectrum speaker 1



Gambar 7. Hasil spectrum speaker 2



Gambar 9. Hasil formant speaker 2

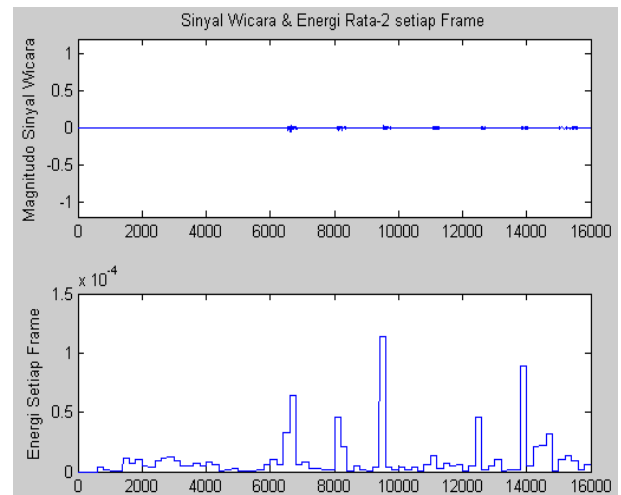
Hasil formant speaker 2 ada di frekuensi sbb :

Formant 1 Frequency 453.2

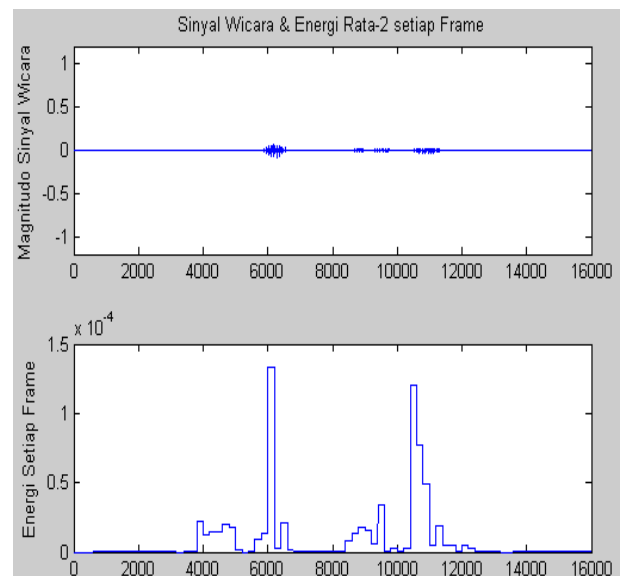
Formant 2 Frequency 1155.9

Formant 3 Frequency 1808.7

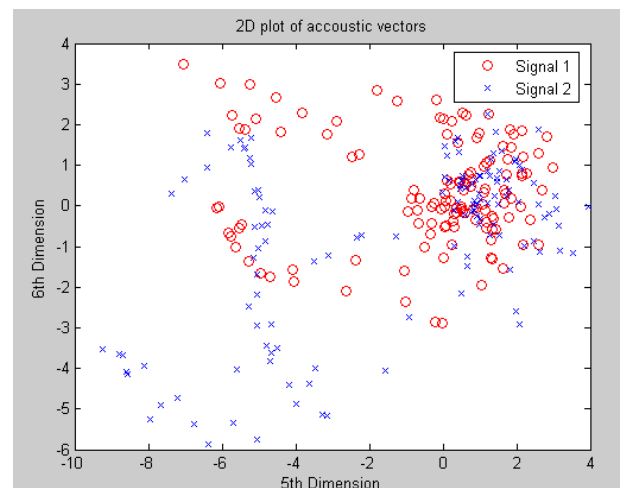
Formant 4 Frequency 2895.3



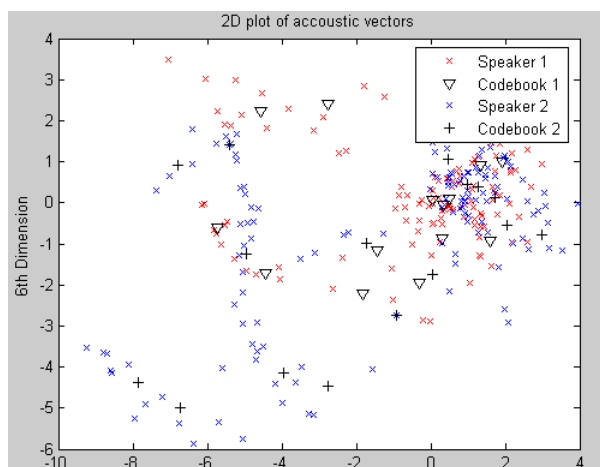
Gambar 10. Hasil Energy speaker 1



Gambar 11. Hasil energy speaker 2



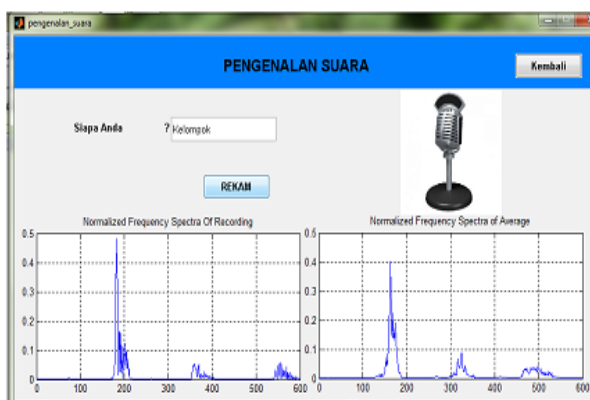
Gambar 12. Vektor 2 dimensi dari speaker 1 dan 2



Gambar 13. Code book 2 dimensi dari 2 speaker

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 14. Hasil spektrum pengujian data base dan hasil pengujian rekaman

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.

### 5. 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. Perbedaan karakteristik ini dapat digunakan untuk mengenali identitas seseorang dengan bantuan software MatLab, metode MFCC dan metode HMM.
2. Apabila cepstrum suara hasil rata-rata perekaman pada data base dan hasil pengetesan sama, 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 *Bootstrapping* Termodifikasi.
- [8] Arman.A.A. Proses Pembentukan dan Karakteristik Sinyal Ucapan. Jakarta.2006.