

APLIKASI PENGENALAN UCAPAN KATA BAHASA INGGRIS MENGUNAKAN LINEAR PREDICTIVE CODING (LPC) DAN HIDDEN MARKOV MODEL (HMM)

Juniar Lestary

ne3a_azza@yahoo.com

Universitas Gunadarma

Jl. Margonda Raya No. 100 Pondok Cina Depok 16424

ABSTRAK

Pengenalan ucapan adalah suatu sistem yang berfungsi untuk mengubah kata-kata ucapan menjadi tulisan. Masukan untuk sistem adalah ucapan, selanjutnya sistem akan mengidentifikasi kata yang diucapkan dan hasil dari indentifikasi berupa teks yang sesuai dengan apa yang diucapkan.

Penulisan ini membahas mengenai pengenalan ucapan untuk menganalisis dan mengenali ucapan kata bahasa Inggris. Dalam proses pengenalan ucapan ini digunakan *Linear Predictive Coding* (LPC) dan *Hidden Markov Model* (HMM). LPC merupakan salah satu metode analisis sinyal suara yang menyatakan ciri-ciri penting dari sinyal suara dalam bentuk koefisien-koefisien LPC. Selanjutnya HMM merupakan salah satu bentuk pemodelan suara dimana sinyal suara dianalisis dan dicari nilai probabilitas yang maksimum sehingga bisa dikenali, dari hasil pemodelan tersebut akan didapatkan parameter yang selanjutnya digunakan dalam proses pengenalan kata. Kata yang dikenali merupakan kata yang memiliki *likelihood* yang paling maksimum.

Aplikasi pengenalan ucapan ini mampu untuk mengenali 10 kata bahasa Inggris seperti *one, two, three, four, five, six, seven, eight, nine*, dan *zero*. Hasil pengujian akurasi untuk responden yang memiliki basis data suara menunjukkan hasil sebesar 100% sedangkan untuk responden yang tidak memiliki basis data suara menunjukkan hasil sebesar 53,75%.

Kata Kunci : Pengenalan Ucapan, Kata Bahasa Inggris, LPC, HMM.

PENDAHULUAN

Perkembangan teknologi yang semakin pesat khususnya di bidang teknologi informasi telah memungkinkan manusia menciptakan perangkat-perangkat yang semakin canggih pula. Salah satu teknologi tersebut adalah *Speech Recognition* atau pengenalan ucapan. *Speech Recogniton* telah memungkinkan suatu perangkat untuk mengenali dan

memahami kata-kata yang diucapkan dengan cara digitalisasi kata dan mencocokkan sinyal digital dengan suatu pola tertentu. Kata-kata yang diucapkan diubah bentuknya menjadi sinyal digital dengan cara mengubah gelombang suara menjadi sekumpulan angka yang kemudian disesuaikan dengan kode-kode tertentu untuk mengidentifikasikan kata-kata tersebut. Hasil dari identifikasi kata yang diucapkan dapat ditampilkan dalam bentuk tulisan.

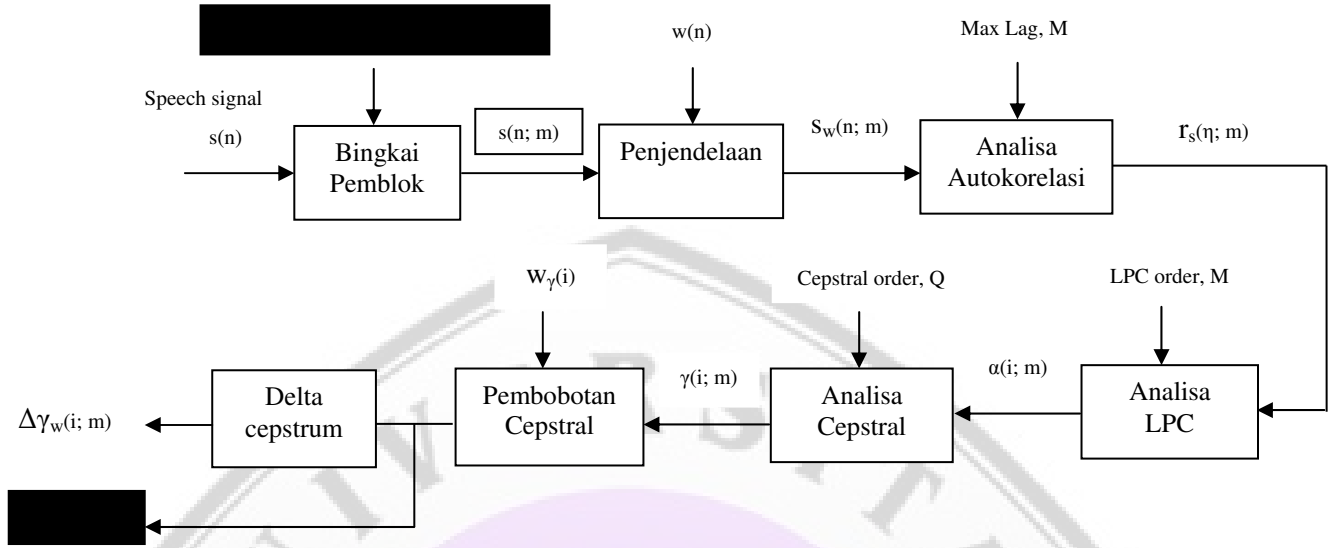
Salah satu metode analisis suara adalah *Linear Predictive Coding* (LPC). LPC memberikan parameter model yang tepat untuk sinyal suara yaitu menyatakan ciri-ciri penting dari sinyal suara dalam bentuk koefisien-koefisien LPC. [Rabiner 1993] Selanjutnya *Hidden Markov Model* (HMM) merupakan salah satu bentuk pemodelan suara dimana sinyal suara dianalisis dan dicari nilai probabilitas yang maksimum sehingga bisa dikenali, dari hasil pemodelan tersebut akan didapatkan parameter yang selanjutnya digunakan dalam proses pengenalan kata. [Rabiner 1989]

Penulisan ini membahas mengenai pengenalan ucapan untuk menganalisis dan mengenali ucapan kata bahasa Inggris. Dalam proses pengenalan ucapan ini dilakukan beberapa tahap yaitu input suara, proses ekstraksi ciri dari sinyal suara dengan menggunakan LPC untuk mendapatkan koefisien-koefisien LPC, melakukan vektor kuantisasi untuk *clustering* dan memperoleh *codebook*, dan proses pelatihan dan pengenalan dengan HMM.

Linear Predictive Coding (LPC)

Linear Predictive Coding (LPC) adalah salah satu teknik analisa suara yang paling kuat, dan salah satu metode yang paling berguna untuk pengkodean suara dengan kualitas yang baik pada bit rate rendah. LPC memberikan perkiraan yang sangat akurat pada parameter suara dan relatif efisien untuk perhitungan.

LPC membuktikan suatu model yang baik untuk pengenalan suara, yaitu memberikan parameter model yang tepat untuk sinyal suara, dapat dilihat pada spektrum koefisien peramalan yang mirip dengan spektrum sinyal aslinya. Blok diagram dari LPC seperti Gambar 1.



Gambar 1. Blok Diagram Analisis LPC

Langkah-langkah dari pemrosesan sinyal dengan LPC sebagai berikut [Rabiner 1993] :

1. Bingkai Pemblok

Pada tahap ini sinyal dikelompokkan ke dalam bingkai-bingkai dengan ukuran masing-masing bingkai sebesar N data.

2. Penjendelaan

Setiap frame kemudian dijendelakan (proses *windowing*) untuk meminimalkan diskontinuitas sinyal pada awal dan akhir bingkai. Jendela yang biasa digunakan untuk metode autokorelasi LPC adalah jendela Hamming dengan bentuk:

$$w(n) = 0,54 - 0,46 \cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \leq n \leq N-1 \quad (1)$$

3. Analisis Autokorelasi

Pada tahap ini masing-masing frame yang telah di *windowing* diautokorelasikan dengan nilai autokorelasi yang tertinggi adalah orde dari analisa LPC, biasanya orde LPC tersebut 8 sampai 16.

$$r_l(m) = \sum_{n=0}^{N-1-m} \tilde{x}_l(n) \tilde{x}_l(n+m) \quad (2)$$

4. Analisis LPC

Langkah selanjutnya adalah analisa LPC, dimana pada tahap ini nilai autokorelasi pada setiap frame diubah menjadi satu set LPC parameter yaitu koefisien LPC, koefisien pantulan (*reflection coefficient*), dan koefisien perbandingan daerah logaritmis (*log area ratio coefficient*).

5. Konversi parameter LPC menjadi koefisien cepstral

Koefisien cepstral ini merupakan koefisien transformasi Fourier yang merepresentasikan spektrum *log magnitude*.

$$c_m = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m} \right) c_k a_{m-k}, \quad 1 \leq m \leq p \quad (3)$$

$$c_m = \sum_{k=1}^{m-1} \left(\frac{k}{m} \right) c_k a_{m-k}, \quad m > p \quad (4)$$

6. Pembobotan Cepstral

Karena *cepstral coefficients orde* rendah sensitif terhadap kemiringan spektrum dan *cepstral coefficients orde* tinggi sensitif terhadap derau, maka dilakukan pembobotan *cepstral coefficient* dengan jendela penyadap sehingga meminimalkan sensitivitas tersebut.

7. Delta Cepstrum

Representasi *cepstral* dari spektrum suara memberikan representasi yang bagus atas sinyal spektrum lokal untuk analisis bingkai.

Hidden Markov Model (HMM)

Model Markov Tersembunyi atau lebih dikenal sebagai Hidden Markov Model (HMM) adalah sebuah model statistik dari sebuah sistem yang diasumsikan sebuah proses Markov dengan parameter yang tak diketahui, dan tantangannya adalah menentukan parameter-parameter tersembunyi (hidden) dari parameter-parameter yang dapat diamati.

Pada model Markov umum, state-nya langsung dapat diamati, oleh karena itu probabilitas transisi state menjadi satu-satunya parameter. Di dalam Model Markov yang Hidden (tersembunyi), state-nya tidak dapat diamati secara langsung, akan tetapi yang dapat diamati adalah variabel-variabel yang terpengaruh oleh state.

Suatu model HMM secara umum memiliki unsur-unsur sebagai berikut :

1. N, yaitu jumlah state dalam model. Secara umum state saling terhubung satu dengan yang lain, dan suatu state bisa mencapai semua state yang lain dan sebaliknya (disebut model ergodic). Namun hal tersebut tidak mutlak, terdapat kondisi lain dimana suatu state hanya bisa berputar ke diri sendiri dan berpindah ke satu state berikutnya, hal ini bergantung pada implementasi dari model.
2. M, yaitu jumlah simbol observasi secara unik pada tiap statenya, misalnya: karakter dalam alfabet, dimana state adalah huruf dalam kata.

3. Distribusi keadaan Transisi $A = \{a_{ij}\}$ dengan

$$a_{ij} = P[q_{t+1} = j \mid q_t = i], \quad 1 \leq i, j \leq N \quad (5)$$

4. Distribusi probabilitas simbol observasi $B = \{b_j(k)\}$ dengan

$$b_j(k) = P(O_t = V_k \mid q_t = j), \quad \text{untuk } 1 \leq j \leq N, 1 \leq k \leq M \quad (6)$$

5. Distribusi keadaan awal $\pi = \{\pi_i\}$

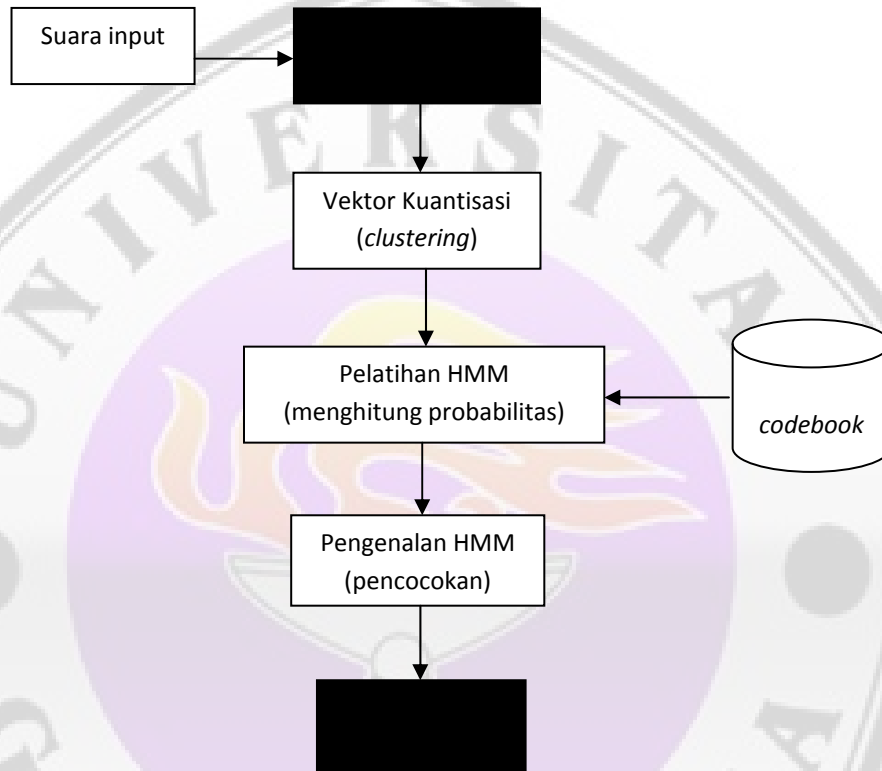
$$\pi_i = P[q_t = i], \quad 1 \leq i \leq N \quad (7)$$

Vektor Kuantisasi

Vektor Kuantisasi adalah teknik *clustering* sinyal *time series*, dalam kasus sinyal ucapan, ke sejumlah *cluster*. Masing-masing *cluster* merepresentasikan data milik suatu populasi tertentu dengan yang serupa (atau perbedaan minimum) karakteristik spektral. Pusat gravitasi masing-masing *cluster* ditugaskan untuk indeks tertentu dan dianggap sebagai wakil dari populasi *cluster* dalam proses pada sinyal. Urutan panjang contoh suara akan diwakili oleh aliran indeks mewakili bingkai (frame) dengan panjang jendela yang berbeda. Oleh karena itu, VQ dianggap sebagai suatu proses pemindahan redundansi, yang meminimalkan jumlah bit yang diperlukan untuk mengidentifikasi setiap bingkai sinyal suara. Vektor kuantisasi juga dapat digunakan untuk membuat sebuah *codebook*. Sebuah *codebook* dapat diperoleh dengan mengkuantisasi vektor pembobotan koefisien *cepstral* dari semua referensi kata.

Metodologi Penelitian

Secara garis besar, pengenalan kata di bagi menjadi 3 tahap, tahap pertama adalah proses ekstraksi ciri sinyal suara menggunakan LPC, tahap kedua adalah melakukan vektor kuantisasi, dan tahap ketiga adalah proses pelatihan dan pengenalan menggunakan HMM. Skema proses pengenalan kata menggunakan HMM dapat dilihat pada gambar 2.



Gambar 2. Skema proses pengenalan kata menggunakan HMM

Pengambilan Data Suara

Pengambilan data suara dilakukan dengan cara merekam suara melalui mikrofon yang dihubungkan dengan komputer. Perekaman suara tersebut menggunakan aplikasi *Sound Recorder Demo* dengan bantuan *software* Matlab. Suara yang direkam terdiri dari 10 kata bahasa Inggris yaitu *one, two, three, four, five, six, seven, eight, nine, dan zero*.

Proses Ekstraksi Ciri Menggunakan LPC

Kebutuhan utama dari pengenalan ucapan adalah ekstraksi ciri dari sinyal suara. Suara biasanya disaring (*filter*) sebelum sampai pada tahap pengenalan atau *recognition*. Penyaringan tersebut dilakukan untuk mengurangi *noise* yang diakibatkan dari lingkungan. Proses ekstraksi ciri ini menghasilkan koefisien-koefisien LPC dari setiap sinyal suara yang diinput.

Proses Vektor Kuantisasi

Setelah proses ekstraksi ciri dilakukan, langkah selanjutnya adalah melakukan proses *clustering* dengan vektor kuantisasi. Proses *clustering* ini dilakukan menggunakan algoritma K-Means. Algoritma ini berdasarkan pada dua langkah yaitu penyebaran vektor observasi dan melakukan *clustering* pada daerah dimana penyebaran yang lebih tinggi dengan memisahkan *cluster* yang paling tersebar. Vektor kuantisasi juga digunakan untuk memperoleh sebuah *codebook*. *Codebook* tersebut nantinya akan dipakai untuk proses pelatihan dan pengenalan kata.

Proses Pelatihan Menggunakan HMM

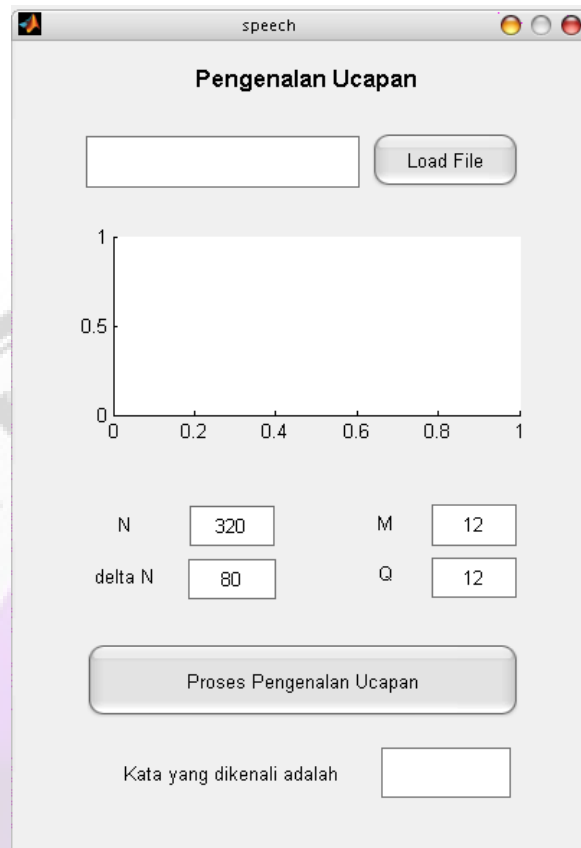
Suatu bagian penting dari proses pengenalan suara adalah pelatihan. Pada tahap pelatihan, sistem melatih setiap kata yang diambil dari *codebook* menggunakan algoritma *Forward-Backward* untuk memperoleh *log-likelihood* dari setiap kata yang dilatih.

Proses Pengenalan Menggunakan HMM

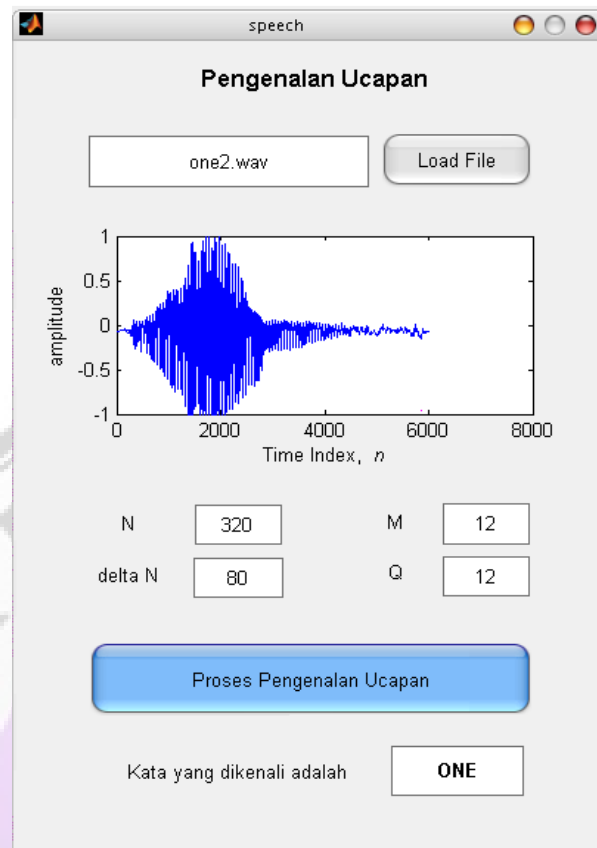
Pengenalan adalah proses pencocokan suara yang diinputkan dengan suara yang terdapat di dalam *codebook*. Pada tahap pelatihan HMM sebelumnya setiap kata telah memperoleh probabilitas *log-likelihood* maka pada tahap pengenalan kata ini setiap probabilitas *likelihood* tersebut dicocokkan ke setiap kata. Kata yang memiliki probabilitas *likelihood* maksimum adalah kata yang dikenali.

Uji Coba Aplikasi

Berikut merupakan tampilan awal aplikasi pengenalan ucapan.



Gambar 3. Tampilan aplikasi pengenalan ucapan



Gambar 4. Tampilan hasil pengenalan ucapan kata *one*

Pengujian Kehandalan Sistem

1. Pengujian terhadap hasil koefisien LPC

Tabel 1.

Jumlah koefisien LPC dari kata *one* sampai *zero* yang diucapkan oleh responden laki-laki

Nama File	Jumlah Koefisien LPC
one.mat	708
two.mat	504
three.mat	480
four.mat	408
five.mat	588
six.mat	864

seven.mat	780
eight.mat	360
nine.mat	648
zero.mat	768

Tabel 2.

Jumlah koefisien LPC dari kata *one* sampai *zero* yang diucapkan oleh responden perempuan

Nama File	Jumlah Koefisien LPC
one3.mat	636
two3.mat	564
three3.mat	660
four3.mat	660
five3.mat	480
six3.mat	900
seven3.mat	1056
eight3.mat	408
nine3.mat	684
zero3.mat	936

Kata *one* sampai *zero* yang diucapkan 1 kali oleh responden laki-laki dan perempuan menghasilkan jumlah koefisien LPC yang berbeda-beda karena setiap kata yang diucapkan oleh seseorang (laki-laki atau perempuan) memiliki karakteristik yang berbeda pada cara pengucapannya yaitu tergantung pada panjang atau pendek kata yang diucapkan.

2. Pengujian terhadap basis data suara

A. Pengujian terhadap responden yang memiliki basis data suara

Tabel 3.

Hasil pengujian oleh responden yang memiliki basis data suara

Nama File	Hasil Pengujian	Nama File	Hasil Pengujian
digits.one1.mat	dikenali	digits.one2.mat	dikenali
digits.two1.mat	dikenali	digits.two2.mat	dikenali
digits.three1.mat	dikenali	digits.three2.mat	dikenali
digits.four1.mat	dikenali	digits.four2.mat	dikenali
digits.five1.mat	dikenali	digits.five2.mat	dikenali
digits.six1.mat	dikenali	digits.six2.mat	dikenali
digits.seven1.mat	dikenali	digits.seven2.mat	dikenali
digits.eight1.mat	dikenali	digits.eight2.mat	dikenali
digits.nine1.mat	dikenali	digits.nine2.mat	dikenali
digits.zero1.mat	dikenali	digits.zero2.mat	dikenali

Rata-rata hasil pengujian terhadap kata *one* sampai *zero* yang diucapkan oleh 2 responden laki-laki yang memiliki basis data suara dapat dikenali dengan sangat baik oleh sistem.

B. Pengujian terhadap responden yang tidak memiliki basis data suara

Tabel 4.

Hasil pengujian oleh responden yang tidak memiliki basis data suara

Nama File	Hasil Pengujian	Nama File	Hasil Pengujian
one.wav	tidak dikenali	nine.wav	dikenali
two.wav	tidak dikenali	zero.wav	dikenali
three.wav	dikenali	one2.wav	dikenali
four.wav	tidak dikenali	two2.wav	tidak dikenali
five.wav	dikenali	three2.wav	dikenali
six.wav	tidak dikenali	four2.wav	dikenali
seven.wav	tidak dikenali	five2.wav	dikenali

eight.wav	tidak dikenali	six2.wav	tidak dikenali
seven2.wav	tidak dikenali	seven.mat	tidak dikenali
eight2.wav	tidak dikenali	eight.mat	tidak dikenali
nine2.wav	dikenali	nine.mat	dikenali
zero2.wav	dikenali	zero.mat	dikenali
one3.wav	dikenali	one2.mat	dikenali
two3.wav	tidak dikenali	two2.mat	tidak dikenali
three3.wav	tidak dikenali	three2.mat	dikenali
four3.wav	dikenali	four2.mat	dikenali
five3.wav	dikenali	five2.mat	dikenali
six3.wav	tidak dikenali	six2.mat	tidak dikenali
seven3.wav	tidak dikenali	seven2.mat	dikenali
eight3.wav	tidak dikenali	eight2.mat	tidak dikenali
nine3.wav	tidak dikenali	nine2.mat	dikenali
zero3.wav	dikenali	zero2.mat	dikenali
one4.wav	dikenali	one3.mat	dikenali
two4.wav	tidak dikenali	two3.mat	tidak dikenali
three4.wav	tidak dikenali	three3.mat	tidak dikenali
four4.wav	tidak dikenali	four3.mat	dikenali
five4.wav	dikenali	five3.mat	dikenali
six4.wav	tidak dikenali	six3.mat	tidak dikenali
seven4.wav	tidak dikenali	seven3.mat	tidak dikenali
eight4.wav	dikenali	eight3.mat	tidak dikenali
nine4.wav	dikenali	nine3.mat	dikenali
zero4.wav	tidak dikenali	zero3.mat	dikenali
one.mat	tidak dikenali	one4.mat	dikenali
two.mat	tidak dikenali	two4.mat	dikenali
three.mat	dikenali	three4.mat	tidak dikenali
four.mat	dikenali	four4.mat	tidak dikenali

five.mat	dikenali	five4.mat	dikenali
six.mat	tidak dikenali	six4.mat	dikenali
seven4.mat	dikenali	nine4.mat	dikenali
eight4.mat	dikenali	zero4.mat	dikenali

Rata-rata hasil pengujian terhadap kata *one* sampai *zero* yang diucapkan oleh 2 responden laki-laki dan 2 responden perempuan yang tidak memiliki basis data suara dapat dikenali dengan baik oleh sistem namun hasilnya belum maksimal. Hal tersebut dapat disebabkan karena perbedaan cara pengucapan kata oleh setiap responden.

Analisis Hasil Pengujian Akurasi

Dari pengujian yang telah dilakukan pada setiap kata *one*, *two*, *three*, *four*, *five*, *six*, *seven*, *eight*, *nine*, dan *zero*. Kemudian dihitung persentase tingkat akurasi dengan menggunakan rumus di bawah ini.

$$\text{Pengujian Akurasi} = \frac{\text{Jumlah kata yang dikenali}}{\text{Jumlah kata yang diuji}} \times 100\%$$

1. Pengujian akurasi terhadap responden yang memiliki basis data suara
Hasil rata-rata pengujian 20 *file* suara dari 2 responden laki-laki yang memiliki basis data suara dengan jumlah kata yang dikenali sebanyak 20 kata menunjukkan tingkat akurasi sebesar 100%.
2. Pengujian akurasi terhadap responden yang tidak memiliki basis data suara
Hasil rata-rata pengujian 80 *file* suara dari 2 responden laki-laki dan 2 responden perempuan yang tidak memiliki basis data suara dengan jumlah kata yang dikenali sebanyak 43 kata menunjukkan tingkat akurasi sebesar 53,75%.

PENUTUP

Dari hasil pembahasan dan pengujian dapat disimpulkan bahwa :

1. Dalam proses pengenalan ucapan dilakukan beberapa tahap yaitu input suara, proses ekstraksi ciri dari sinyal suara dengan menggunakan *Linear Predictive Coding* (LPC) untuk mendapatkan koefisien-koefisien LPC, melakukan *clustering* dengan vektor kuantisasi, dan proses pelatihan dan pengenalan dengan *Hidden Markov Model* (HMM).
2. Hasil pengujian terhadap koefisien LPC dengan menggunakan $N=320$, delta $N=80$, $M=12$, dan $Q=12$ dari kata *one* sampai *zero* yang diucapkan oleh responden laki-laki dan perempuan menghasilkan jumlah koefisien LPC yang berbeda-beda karena setiap kata yang diucapkan oleh seseorang (laki-laki atau perempuan) memiliki karakteristik yang berbeda pada cara pengucapannya yaitu tergantung pada panjang atau pendek kata yang diucapkan.
3. Hasil pengujian akurasi terhadap 20 data *file* suara oleh responden yang memiliki basis data suara adalah sebesar 100%, sedangkan hasil pengujian akurasi terhadap 80 data *file* suara oleh responden yang tidak memiliki basis data suara adalah sebesar 53,75%.
4. Dari hasil pengujian akurasi dapat dilihat bahwa basis data suara sangat berpengaruh pada akurasi pengenalan, semakin banyak basis data yang digunakan maka peluang pengenalan kata akan semakin besar.
5. Diharapkan dilakukan penambahan terhadap basis data suara agar aplikasi pengenalan ucapan ini dapat semakin baik dalam mengenali kata. Untuk pengembangan selanjutnya diharapkan aplikasi ini dapat mengenali ucapan secara *realtime*.

DAFTAR PUSTAKA

Abdulla, W. H. dan Kasabov, N. K., 1999. *The Concept of Hidden Markov Model in Speech Recognition*. Knowledge Engineering Lab Information Science Department University of Otago. New Zealand.

Away, Gunaidi Abdia. 2006. *The shortcut of MATLAB programming*, Bandung : Informatika.

IT University of Copenhagen Multimedia Technology Speech and IT Systems. *Speech Coding and Recognition Course*. <http://www.itu.dk/courses/TKG/E2005/exercises.html>.

Murugan, M. T. Bala dan Balaji, M., 2006. *SOPC-Based Speech-to-Text Conversion*. NIOS II Embedded Processor Design Contest—Outstanding Designs. National Institute of Technology. Trichy.

Nilsson, Mikael dan Ejnarsson, Marcus. 2002. *Speech Recognition using Hidden Markov Model : performance evaluation in noisy environment*. Departement of Telecommunications and Signal Processing. Blekinge Institute of Technology.

Petrushin, Valery A., 2000. *Hidden Markov Models : Fundamentals and Applications*. Center for Strategic Technology Research. Accenture.

Rabiner, Lawrence. 1989. *A Tutuorial on Hidden Markov Model and Selected Application in Speech Recognition*. vol. 77. no. 2. pp. 257-286. IEEE.

Rabiner, Lawrence. 1993. *Fundamental of Speech Recognition*. New Jersey : Prentice Hall.

Rabiner, Lawrence dan Juang, B. H., 1991. *Hidden Markov Models for Speech Recognition*. vol. 33. no. 3. pp. 251-272. TECHNOMETRICS.

Sugiharto, Aris. 2006. *Pemrograman GUI dengan MATLAB*. Yogyakarta : ANDI.