

Pengenalan Ucapan Kata Terisolasi dengan Metode *Hidden Markov Model* (HMM) melalui Ekstraksi Ciri *Linear Predictive Coding* (LPC)¹
(*Isolated Word Speech Recognition with Hidden Markov Model via LPC features extraction*)

Oleh:
Achmad Hidayatno dan Sumardi²

Abstrak

Penelitian yang intensif dalam bidang pengolahan sinyal menyebabkan teknologi komunikasi berkembang dengan pesat. Salah satunya adalah pengenalan ucapan (*speech recognition*). Pengenalan ucapan merupakan upaya agar manusia dan mesin dapat berkomunikasi dengan media suara. Salah satu contoh aplikasi dari pengenalan ucapan yaitu sebagai kata kunci (password) untuk kartu akses (misalnya kartu kredit) dan dapat juga digunakan sebagai kata sandi dalam sistem keamanan. Penelitian ini membahas mengenai pengenalan ucapan untuk menganalisis dan mengenali delapan ucapan kata Bahasa Indonesia yang memiliki korelasi tinggi antara kata yang satu dengan yang lainnya, yaitu: muka, muak, kamu, kaum, masuk, kumus, kusam dan sukma. Masing-masing kata diucapkan oleh 10 orang pria dan 10 orang wanita sebanyak 10 kali sehingga jumlah keseluruhan dari sampel data dalam penelitian ini sebanyak 1600 sampel data.

Dalam proses pengenalan digunakan algoritma ekstraksi ciri yang disebut *Linear Predictive Coding* (LPC), sedang pola pengenalan menggunakan metode *Hidden Markov Model* (HMM). LPC merupakan salah satu metode analisis sinyal suara yang menyatakan ciri-ciri penting dari sinyal suara tersebut dalam bentuk koefisien-koefisien LPC. Selanjutnya HMM digunakan untuk pelatihan dan pengenalan.

Hasil pengujian menunjukkan tingkat pengenalan tertinggi untuk data rekaman diperoleh pada state 18 dengan prosentase pengenalan sebesar 100% . Tingkat pengenalan tertinggi untuk pengujian on-line diperoleh pada state 18 dengan prosentase pengenalan sebesar 95%.

Abstract

Intensive research on the signal processing make communication technology growing faster, including speech recognition. Speech recognition is a method to create communication between human and machine. One of application on speech recognition used as password for access card and also codeword in security system. This final task study about speech recognition to analyse and recognize eight highly correlated Indonesian words, that is: muka, muak, kamu, kaum, masuk, kumus, kusam and sukma. Each of the words spoken by 10 man people and 10 women counted 10 times so that grand total of data sample in this final task counted 1600 data sample.

In recognize process will be use extraction algorithm that called Linear Predictive Coding (LPC), while for the pattern recognition using Hidden Markov Model (HMM). LPC is a signal analysis method, which representing the important features of speech signal in the form of coefficients LPC. The HMM was then applied for the training sessions as well as in the recognition phase.

The result show that the highest recognition level for the data of record obtained at 18th state with percentage of recognition equal to 100%. The highest recognition level for on-line examination obtained at 18th state with percentage of recognition equal to 95%.

¹ Penelitian Hibah Bersaing DIKTI Depdiknas dengan kontrak 319/SP3/PP/DP2M/II/2006

² Jurusan Teknik Elektro UNDIP Semarang

I. PENDAHULUAN

Pengenalan suara merupakan upaya agar manusia dan mesin dapat berkomunikasi dengan media suara. Salah satu contoh aplikasi dari pengenalan suara yaitu sebagai kata kunci (*password*) untuk kartu akses (misalnya kartu kredit). Pengenalan suara juga dapat digunakan sebagai kata sandi dalam sistem keamanan, pengoperasian komputer menggunakan perintah suara, maupun penekanan tombol telepon dengan suara.

Salah satu bentuk pemodelan suara adalah HMM (*Hidden Markov Model*). Dalam pemodelan ini suara dapat diasumsikan sebagai parameter acak yang dapat diperkirakan secara tepat. Sinyal suara dianalisis dan dicari nilai probabilitas yang maksimum sehingga bisa dikenali dalam pemodelan HMM. Hasil dari pemodelan yaitu akan didapatkan parameter yang selanjutnya digunakan dalam proses pengenalan.

Tujuan penelitian ini adalah membuat program simulasi untuk menganalisis dan mengenali bentuk sinyal ucapan yang terdiri dari kata-kata yang mempunyai korelasi tinggi antara satu dengan yang lainnya dengan memanfaatkan suatu sistem pengenalan pola. Dengan sistem pengenalan pola yang menggunakan LPC untuk proses ekstraksi ciri dan HMM untuk proses klasifikasinya.

II. TINJAUAN PUSTAKA

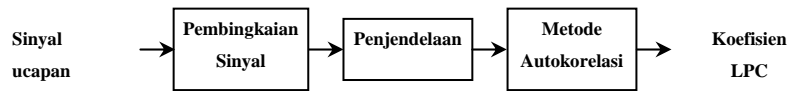
Pengenalan suara adalah upaya agar suara dapat diidentifikasi sehingga dapat dimanfaatkan. Pengenalan suara secara umum dapat dibagi menjadi beberapa tahap yaitu: ekstraksi ciri, pemodelan dan pengenalan.

2.1 Ekstraksi Ciri

Ciri sinyal ucapan sangat berguna pada sistem pengenalan suara. Salah satu metode yang digunakan untuk proses ekstraksi ciri adalah *Linear Predictive Coding* (LPC). Prinsip dasar dari pemodelan sinyal dengan menggunakan LPC adalah bahwa contoh sinyal ucapan $s(n)$ pada waktu ke- n dapat diperkirakan sebagai kombinasi linier p sampel sinyal ucapan sebelumnya yaitu

$$s(n) \approx a_1 s(n-1) + a_2 s(n-2) + \dots + a_p s(n-p)$$

dengan koefisien a_1, a_2, \dots, a_p diasumsikan bernilai konstan pada suatu *frame* analisis ucapan. Prosedur untuk mendapatkan koefisien LPC diperlihatkan pada Gambar 1.



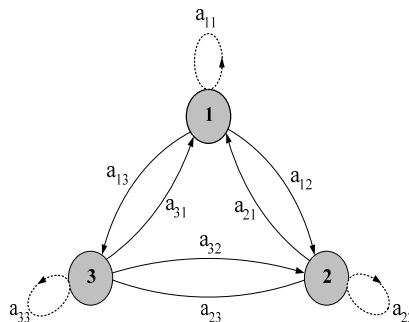
Gambar 1 Blok diagram analisis LPC.

2.2 Pemodelan dengan HMM

HMM adalah analisis statistika yang memodelkan sinyal suara dan mencari bentuk kata yang paling sesuai. HMM berkembang dengan sangat cepat karena pemodelan ini sangat kaya dalam struktur matematika dan mengacu pada fungsi probabilitas rantai markov.

2.2.1 Rantai Markov

Algoritma HMM didasari oleh model matematik yang dikenal dengan rantai markov. Rantai markov secara umum ditunjukkan pada Gambar 2.



Gambar 2 Rantai Markov.

Beberapa hal yang dapat dijelaskan tentang rantai markov yaitu:

- ♣ Transisi keadaan dari suatu keadaan tergantung pada keadaan sebelumnya.

$$P[q_t = j | q_{t-1} = i, q_{t-2} = k, \dots] = P[q_t = j | q_{t-1} = i]$$

- ♣ Transisi keadaan bebas terhadap waktu.

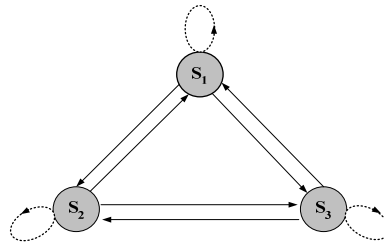
$$a_{ij} = P[q_t = j | q_{t-1} = i]$$

2.2.2 Tipe HMM

HMM dibagi menjadi dua tipe dasar yaitu HMM ergodic dan HMM Kiri-Kanan.

1. HMM *ergodic*.

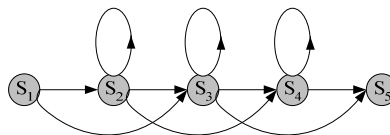
Pada HMM model *ergodic* perpindahan keadaan satu ke keadaan yang lain semuanya memungkinkan, hal ini ditunjukkan pada Gambar 3.



Gambar 3 HMM model *ergodic*.

2. HMM Kiri-kanan

Pada HMM kiri-kanan perpindahan keadaan hanya dapat berpindah dari kiri kekanan, perpindahan keadaan tidak dapat mundur ke belakang, hal ini ditunjukkan pada Gambar 4.



Gambar 4 HMM model kiri-kanan.

2.2.3 Elemen HMM

Elemen yang terdapat pada HMM yaitu :

Elemen yang terdapat pada HMM yaitu :

1. N , jumlah keadaan (*state*) dalam model.
2. M , jumlah simbol observasi yang berbeda tiap keadaan.
3. Distribusi keadaan transisi $A=\{a_{ij}\}$ dengan
$$a_{ij} = P[q_{t+1} = j \mid q_t = i], \quad 1 \leq i, j \leq N$$
4. Distribusi probabilitas simbol observasi, $B=\{b_j(k)\}$

dengan

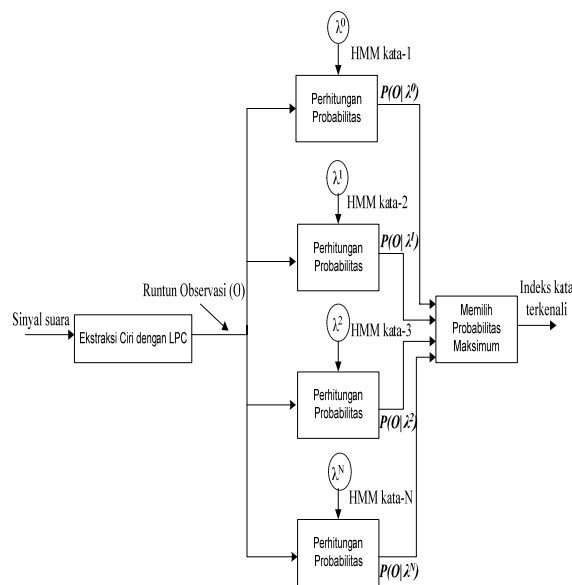
$$b_j(k) = P(o_t = v_k | q_t = j), \quad \text{untuk } 1 \leq j \leq N, \quad 1 \leq k \leq M$$

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

$$\pi_i = P[q_t = i] \quad 1 \leq i \leq N$$

2.3 Pengenalan Kata

Kata berkorelasi tinggi adalah masukan suara berupa kata-kata yang terdiri dari huruf-huruf yang hampir sama dengan susunan yang hanya dibolak-balik saja. Diagram blok pengenalan kata berkorelasi tinggi ditunjukkan pada Gambar 5.

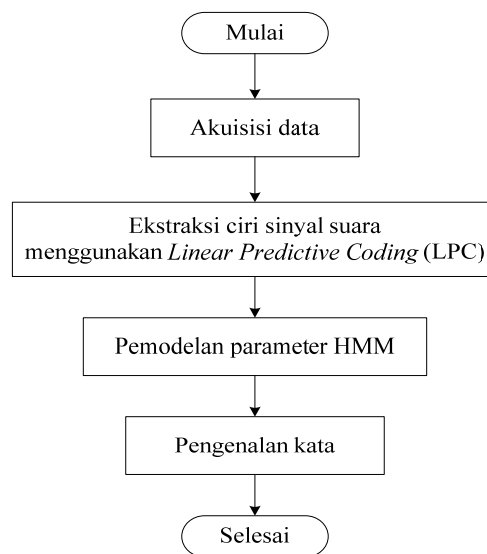


Gambar 5 Diagram blok pengenalan kata.

Sinyal suara $s(n)$ dimasukkan ke dalam ekstraksi ciri dengan menggunakan LPC sehingga didapatkan vektor runtun observasi O . Kemudian dihitung probabilitas dari runtun observasi terhadap model HMM untuk tiap-tiap. Hasil perhitungan dipilih probabilitas yang paling maksimum untuk kemudian ditetapkan sebagai kata terkenal dengan keluaran berupa teks.

III. METODE PENELITIAN

Pembuatan program simulasi pengenalan kata di bagi menjadi 5 tahap, tahap pertama adalah akuisisi data, tahap kedua adalah proses ekstraksi ciri sinyal suara menggunakan LPC, tahap ketiga adalah proses pemodelan parameter HMM dan tahap yang keempat adalah pengenalan angka. Secara umum pembuatan program simulasi ini mengikuti alur sesuai yang ditunjukkan pada Gambar 6.



Gambar 6 Alur program pengenalan kata berkorelasi tinggi.

3.1 Akuisisi Data

Data berupa sinyal suara diperoleh dengan cara merekam suara melalui mikropon yang dihubungkan dengan komputer pribadi. Perekaman suara dilakukan dengan bantuan program aksesories windows yaitu *sound recorder* dengan frekuensi *sampling* 8000 Hz, 8 bit, mono.

Suara diucapkan oleh 20 orang yang terdiri dari sepuluh orang pria dan sepuluh orang wanita dimana untuk setiap kata diulang sebanyak 10 kali. Dengan *sound recorder* suara direkam dan berkas suara dibuka kembali dengan Cool Edit Pro 2.0 untuk dipotong kemudian disimpan dalam berkas dengan format penamaan ABC.wav. Huruf A menunjukkan nama responden, huruf B

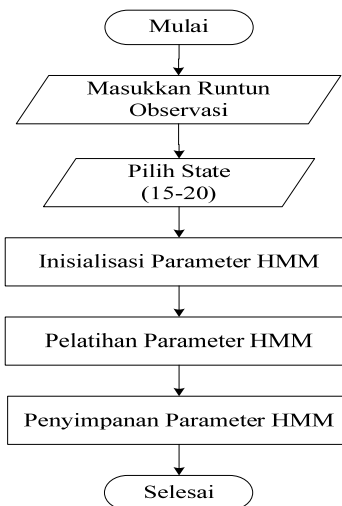
menunjukkan kata yang diucapkan dan huruf C menunjukkan suara yang ke (1-10).

3.2 Ekstraksi Ciri

Setelah melalui proses perekaman, sinyal suara akan dibaca atau dipanggil kembali dengan fungsi `wavread`. Hasil pembacaan data untuk keseluruhan sinyal suara digunakan untuk proses selanjutnya yaitu ekstraksi ciri. Proses ekstraksi ciri digunakan untuk mencari nilai koefisien-koefisien LPC dari sinyal suara.

3.3 Pemodelan Parameter HMM

Untuk mendapatkan parameter HMM melalui lima tahap, yaitu : memasukkan runtun *observasi* hasil dari proses ekstraksi ciri, memilih *state*, inialisasi parameter HMM, pelatihan HMM, pelatihan parameter HMM dengan tujuan untuk mendapatkan parameter yang lebih baik dan penyimpanan. *State* untuk pemodelan parameter HMM dapat dipilih dari *state* 15 sampai 20. Proses untuk mendapatkan parameter HMM ditunjukkan pada Gambar 7.



Gambar 7 Bagan pemodelan parameter HMM.

3.4 Pengenalan Kata

Proses pengenalan kata dibagi menjadi 5 tahap, tahap pertama memasukkan sinyal suara $s(n)$, tahap kedua adalah proses ekstraksi ciri sinyal suara, tahap ketiga menghitung probabilitas dari runtun observasi hasil proses ekstraksi ciri, tahap keempat memilih hasil probabilitas maksimum dan tahap kelima menetapkan kata terkenali dengan keluaran berupa teks. Diagram proses pengenalan kata ditunjukkan pada Gambar 8.



Gambar 8 Bagan proses pengenalan kata.

Sinyal suara $s(n)$ dimasukkan ke dalam ekstraksi ciri sehingga didapatkan vektor ciri, dari vektor ciri ini kemudian ditentukan probabilitas terhadap semua kata dan dipilih nilai probabilitas maksimum yang akhirnya nilai maksimum tersebut diubah menjadi keluaran berupa teks.

IV. HASIL DAN ANALISIS

4.1 Analisis Korelasi

Dalam penelitian ini digunakan sampel data 8 kata Bahasa Indonesia, yaitu kamu, kaum, muka, muak, kamus, kusam, masuk dan sukma. Pengujian terhadap 8 kata tersebut dilakukan melalui analisis korelasi terhadap spektrumnya dengan

koefisien FFT sebesar 256 untuk melihat apakah benar terdapat korelasi yang tinggi antara kata yang satu dengan yang lain.

Hasil perhitungan korelasi antara 2 kata secara keseluruhan dapat dilihat pada Tabel 1.

Tabel 1 Koefisien korelasi 8 kata Bahasa Indonesia.

Kata Masukan	Dikenali Sebagai Kata	Nilai Korelasi	Kata Masukan	Dikenali Sebagai Kata	Nilai Korelasi
Muka	muka	1	Masuk	muka	0,89
	muak	0,86		muak	0,94
	kamu	0,89		kamu	0,83
	kaum	0,93		kaum	0,86
	masuk	0,89		masuk	1
	kamus	0,89		kamus	0,89
	kusam	0,96		kusam	0,90
	sukma	0,95		sukma	0,93
Muak	muka	0,86	Kamus	muka	0,89
	muak	1		muak	0,92
	kamu	0,89		kamu	0,97
	kaum	0,89		kaum	0,95
	masuk	0,94		masuk	0,89
	kamus	0,92		kamus	1
	kusam	0,92		kusam	0,93
	sukma	0,90		sukma	0,92
Kamu	muka	0,89	Kusam	muka	0,96
	muak	0,89		muak	0,92
	kamu	1		kamu	0,94
	kaum	0,94		kaum	0,94
	masuk	0,83		masuk	0,90
	kamus	0,97		kamus	0,93
	kusam	0,94		kusam	1
	sukma	0,90		sukma	0,96
Kaum	muka	0,93	Sukma	muka	0,95
	muak	0,89		muak	0,90
	kamu	0,94		kamu	0,90
	kaum	1		kaum	0,90
	masuk	0,86		masuk	0,93
	kamus	0,95		kamus	0,92
	kusam	0,94		kusam	0,96
	sukma	0,90		sukma	1

Dari Tabel 1 dapat dilihat bahwa ada korelasi yang tinggi antara kata yang satu dengan kata yang lainnya. Nilai korelasi tertinggi diperoleh antara kata kamu dan kamus yaitu sebesar 0,97 sedangkan nilai korelasi terendah diperoleh antara kata kamu dan masuk yaitu sebesar 0,83.

4.2 Pengujian dan Analisis Hasil Pengujian

Pengujian program dilakukan dalam dua tahap, tahap pertama yaitu pengujian dengan menggunakan data rekaman dan tahap kedua adalah pengujian secara *on-line*. Pada pengujian dengan data rekaman dilakukan dengan menggunakan suara yang telah direkam, sedangkan pada pengujian *on-line* dilakukan dengan menggunakan suara yang direkam pada saat itu juga.

4.2.1 Hasil Pengujian dengan Data Rekaman

Pengujian dengan data rekaman dilakukan oleh 20 responden. Keseluruhan data yang ada akan melalui proses pengujian dengan menggunakan model yang telah didapatkan untuk setiap kata. Pengujian dilakukan dengan mengubah masukan sinyal suara dan *state*.

Dari pengujian yang telah dilakukan kemudian dihitung prosentase pengenalan kata masukan terhadap kata keluaran dengan persamaan (1) berikut ini.

$$\frac{\text{Jumlah ucapan yang dikenali}}{\text{Jumlah ucapan total}} \times 100 \% \quad (1)$$

dengan jumlah ucapan adalah 10 untuk masing-masing responden sehingga jumlah ucapan total pada pengujian dengan data rekaman adalah 200. Prosentase pengenalan untuk keseluruhan kata pada *state* 15 ditunjukkan oleh Tabel 2.

Tabel 2 Prosentase pengenalan data rekaman untuk *state* 15.

Kata Masukan	Dikenali Sebagai Kata	Tingkat Pengenalan (%)	Kata Masukan	Dikenali Sebagai Kata	Tingkat Pengenalan (%)
Muka	muka	100	Kamu	muka	0
	muak	0		muak	1
	kamu	0		kamu	99
	kaum	0		kaum	0
	masuk	0		masuk	0
	kamus	0		kamus	0
	kusam	0		kusam	0
	sukma	0		sukma	0
Muak	muka	0,5	Kaum	muka	0
	muak	99,5		muak	0
	kamu	0		kamu	1
	kaum	0		kaum	99
	masuk	0		masuk	0
	kamus	0		kamus	0
	kusam	0		kusam	0
	sukma	0		sukma	0
Masuk	muka	0	Kusam	muka	0
	muak	0		muak	0
	kamu	0		kamu	0
	kaum	0		kaum	0
	masuk	100		masuk	0
	kamus	0		kamus	0
	kusam	0		kusam	100
	sukma	0		sukma	0
Kamus	muka	0	Sukma	muka	0
	muak	0		muak	0
	kamu	0		kamu	0
	kaum	0		kaum	0
	masuk	0		masuk	0
	kamus	100		kamus	0
	kusam	0		kusam	0
	sukma	0		sukma	100

Prosentase hasil pengenalan kata yang benar dan sesuai dengan yang diinginkan pada *state* 15 ditunjukkan pada Tabel 2 dengan kotak warna biru. Prosentase pengenalan tertinggi dicapai oleh kata muka, masuk, kamus, kusam dan sukma dengan prosentase pengenalan sebesar 100%.

4.2.2 Hasil Pengujian *On-line*

Selain menggunakan data rekaman, dilakukan juga pengujian secara *on-line*, yaitu program mengenali kata langsung dari seorang penutur. Prosentase pengenalan untuk keseluruhan kata pada *state* 15 ditunjukkan oleh Tabel 3.

Tabel 3 Prosentase pengenalan *on-line* untuk *state* 15.

Kata Masukan	Dikenali Sebagai Kata	Tingkat Pengenalan (%)	Kata Masukan	Dikenali Sebagai Kata	Tingkat Pengenalan (%)
Muka	muka	88	Masuk	muka	0
	muak	4		muak	0
	kamu	0		kamu	0
	kaum	0		kaum	0
	masuk	0		masuk	84
	kamus	0		kamus	4
	kusam	0		kusam	0
	sukma	8		sukma	12
Muak	muka	16	Kamus	muka	0
	muak	68		muak	0
	kamu	0		kamu	0
	kaum	0		kaum	0
	masuk	0		masuk	8
	kamus	0		kamus	68
	kusam	8		kusam	24
	sukma	8		sukma	0
Kamu	muka	4	Kusam	muka	0
	muak	0		muak	0
	kamu	48		kamu	0
	kaum	24		kaum	0
	masuk	16		masuk	0
	kamus	0		kamus	0
	kusam	4		kusam	72
	sukma	4		sukma	28
Kaum	muka	0	Sukma	muka	0
	muak	4		muak	0
	kamu	4		kamu	0
	kaum	80		kaum	0
	masuk	12		masuk	4
	kamus	0		kamus	0
	kusam	0		kusam	0
	sukma	0		sukma	96

Prosentase hasil pengenalan kata yang benar dan sesuai dengan yang diinginkan pada *state* 15 ditunjukkan pada Tabel 3 dengan kotak warna biru. Prosentase pengenalan tertinggi dicapai oleh kata sukma dengan prosentase pengenalan sebesar 96%. Pengenalan terendah dicapai oleh kata kamu dengan prosentase pengenalan sebesar 48%. Kesalahan yang terjadi disebabkan karena kata masukan yang bukan ucapan kata muka, muak, kamu, kaum, masuk, kamus, kusam dan sukma dikenali sebagai kata tersebut.

4.3 Rata-rata Prosentase Pengenalan Hasil Pengujian

Prosentase pengenalan yang diperoleh dari hasil pengujian baik dengan data rekaman maupun secara *on-line* untuk masing-masing kata pada setiap *state* kemudian dihitung rata-rata prosentase pengenalannya dengan persamaan (2). (%)
 Pengenalan =

$$\frac{\text{Jumlah prosentase pengenalan tertinggi tiap kata}}{\text{Jumlah kata}} \dots\dots(2)$$

dimana jumlah kata adalah 8. Rata-rata prosentase pengenalan keseluruhan *state* baik untuk pengujian dengan data rekaman maupun pengujian *on-line* ditunjukkan oleh Tabel 4.

Tabel 4 Rata-rata prosentase pengenalan semua *state*.

State	Tingkat Pengenalan (%)	
	Pengujian Data Rekaman	Pengujian <i>On-line</i>
15	99,69	75,5
16	99,75	81
17	99,75	87,5
18	100	95
19	99,94	92
20	99,81	94,5

Dari Tabel 4 didapatkan bahwa untuk pengujian dengan data rekaman, tingkat pengenalan terendah terjadi pada *state* 15 yaitu sebesar 99,69% sedangkan tingkat pengenalan tertinggi terjadi pada *state* 18 yaitu sebesar 100%. Pada pengujian secara *on-line*, tingkat pengenalan terendah terjadi pada *state* 15 yaitu sebesar 75,5% sedangkan tingkat pengenalan tertinggi terjadi pada *state* 18 yaitu sebesar 95%.

Pengenalan *on-line* memiliki prosentase pengenalan yang lebih rendah jika dibandingkan dengan pengenalan dengan data rekaman. Hal ini disebabkan karena pada saat pengenalan *on-line*, suara yang diucapkan oleh responden tidak mengalami proses pelatihan data terlebih dahulu sehingga memiliki perbedaan karakteristik sinyal ucapan dengan basisdata yang telah disimpan. Perbedaan tersebut mempengaruhi terhadap hasil pengenalan yang didapat.

KESIMPULAN

Dari hasil penelitian dan pembahasan dapat disimpulkan bahwa :

1. Dengan menggunakan ekstraksi ciri LPC dan pemodelan HMM akan memberikan akurasi berkisar pada angka 99,82% untuk pengenalan ucapan kata yang berkorelasi tinggi pada pengujian dengan data rekaman. Pengenalan pada pengujian *on-line* memberikan akurasi berkisar pada angka 87,58%.
2. Jumlah data pelatihan mempengaruhi besarnya tingkat pengenalan ucapan kata. Makin besar jumlah data untuk pelatihan didapatkan tingkat pengenalan yang lebih baik.

UCAPAN TERIMAKASIH

Penulis mengucapkan banyak terimakasih atas kesempatan yang diberikan oleh Dirjen DIKTI melalui Direktorat Penelitian dan Pengabdian Masyarakat lewat Lembaga Penelitian Universitas Diponegoro, sehingga penulis dapat melakukan penelitian di bidang pengolahan suara digital dalam bentuk Penelitian Hibah Bersaing.

DAFTAR PUSTAKA

- [1] Abdulla, W. H. and Kasabov, N. K., *The Concept of Hidden Markov Model in Speech Recognition*, Knowledge Engineering Lab. Information Science Department University of Otago, New Zealand, 1999.
- [2] Cappe, O., *H2M : A Set of Matlab/Octave Functions for The EM Estimation of Mixtures and Hidden Markov Model*, ENST dpt. TSI/LCTI (CNRS-URA 820), Paris, 2001.
- [3] Furui, S., *Digital Speech Processing, Synthesis, and Recognition*, Marcel Dekker, Inc., New York, 1989.
- [4] Gold, B., and Morgan, N., *Speech and Audio Signal Processing : Processing and Perception of Speech and Music*, John Wiley & Sons, Inc., New York, 1999.
- [5] Kanungo, T., *Hidden Markov Models*, <http://www.cfar.umd.edu/~kanungo>, Maret 2005.
- [6] Little, J.N. and Shure, L., *Signal Processing Toolbox User's Guide*, The MathWorks, Inc., Natick, MA., 1992.
- [7] Rabiner, L., *A Tutorial on Hidden Markov Model and Selected Application in Speech Recognition*, vol. 77, no. 2, pp. 257-286, IEEE, 1989.
- [8] Rabiner, L., Biing-Hwang Juang. *Fundamentals Of Speech Recognition*, New Jersey: Prentice Hall, 1993.