

PENGENALAN UCAPAN KATA BERKORELASI TINGGI

Yultrisna¹, Andi Syofian², Rahmat¹

¹Jurusan Teknik Elektronika Politeknik Negeri Padang

²Jurusan teknik Elektro Institute Teknologi Padang

ABSTRAK

Penelitian pengenalan ucapan dengan metode stokastik sudah mulai dipopulerkan sejak tahun 1970-an. Salah satu bentuk metode pendekatan stokastik adalah Hidden Markov Model. Penelitian ini menggunakan model tersebut untuk menganalisis dan mengenali delapan ucapan kata bahasa Indonesia yang memiliki korelasi tinggi antara kata yang satu dengan kata yang lainnya, yaitu: muka, muak, kamu, kaum, masuk, kamus, kusam dan sukma.

Dengan topik aplikasi HMM, analisis LPC (Linear Predictive Coding) digunakan untuk mencari ciri-ciri penting setiap sinyal suara kata-kata terucap. Selanjutnya HMM (Hidden Markov Model) digunakan untuk pelatihan dan pengenalan.

Hasil penelitian menunjukkan bahwa rata-rata pengenalan seluruh kata cukup signifikan yaitu di atas 97%. Tingkat pengenalan runtun observasi pengucap perempuan terhadap pelatihan runtun observasi pengucap laki-laki hanya 80% relatif lebih rendah dari pengenalan runtun observasi pengucap laki-laki terhadap pelatihan runtun observasi pengucap perempuan yang mencapai angka 92%. Hal ini dapat disebabkan oleh karena harmoniks spektral suara perempuan terpisahkan 2 kali lebih besar dari harmoniks spektral suara laki-laki.

Kata kunci : Ucapan kata berkorelasi tinggi, pelatihan, pengenalan, runtun observasi, LPC, HMM.

1. PENDAHULUAN

Proses pengenalan ucapan oleh manusia mulai terbentuk sejak balita yaitu ketika sudah dapat mendengar dan mampu mengeluarkan bunyi. Proses ini tanpa disadari dilakukan melalui proses pembelajaran, yaitu belajar mengenal ucapan yang didengar. Pada manusia tidaklah begitu sulit untuk mengenali suatu ucapan yang didengarnya, karena manusia mempunyai sistem informasi yang sangat canggih dan memiliki kemampuan pengenalan pola yang sangat baik.

Penelitian pengenalan suara telah banyak dilakukan, diantaranya penelitian pengenalan ucapan kata-kata bahasa Indonesia, namun masih terbatas pada kata-kata tertentu yang tidak mempunyai hubungan satu sama lain baik dari segi kemiripan bunyi, maupun huruf-huruf yang membentuknya. Penelitian ini adalah untuk mengetahui tingkat kemampuan pengenalan suatu sistem pengenalan terhadap kata-kata yang mempunyai kesamaan huruf-huruf pembentuknya. Diharapkan dengan menggunakan model LPC dan model HMM, dapat dibangun suatu sistem pengenalan yang mampu mengenali kata-kata yang menjadi bahan penelitian. Kata-kata yang dimaksud adalah *muka, muak, kamu, kaum, masuk, kamus, kusam dan sukma*.

2. TINJAUAN PUSTAKA

Di tahun 1970-an dan khususnya pada tahun 1980-an, penelitian tentang suara diarahkan ke pendekatan stokastik untuk memodelkan suara ditujukan untuk berbagai masalah, khususnya dalam sistem berskala besar. Istilah pendekatan stokastik

digunakan untuk menunjukkan bahwa model yang diberlakukan bersifat memberi karakter beberapa keanekaragaman dalam suara. Salah satu bentuk pendekatan stokastik adalah HMM (Hidden Markov Model). Sejarah HMM dalam bidang pengenalan suara diawali dengan kerja sama Baker pada Universitas Mellon di Carnegie (Baker, 1975a, 1975b) dan Jelinek dan kawan-kawan di IBM (Jelinek et al., 1975, 1976). Selanjutnya Levinson (1985) dan Poritz (1988) mempopulerkan HMM untuk pekerjaan pemrosesan sinyal suara. Pendekatan yang mereka lakukan adalah memodelkan masalah sebagai sebuah proses stokastik ganda dimana data yang diobservasi merupakan hasil proses yang telah dilewatkan melalui sebuah sensor dan menghasilkan proses kedua (yang terobservasi). Kedua proses telah dikarakteristikkan namun hanya satu yang dapat diobservasi. Algoritma identifikasi ini dikenal dengan algoritma estimate-maximize (EM) (Dempster, 1977). Pada tahun 1966 dan permulaan tahun 1970-an, Baum dan kawan-kawan membangun algoritma forward-backward yang mempertimbangkan algoritma EM kedalamnya untuk sebuah kasus HMM. Algoritma ini dikenal dengan nama algoritma reestimasi Baum-Welch yang ditujukan untuk estimasi parameter HMM dan pembacaan kode dalam waktu yang linear disepanjang rentetan observasi.

2.1 Analisis LPC

Secara garis besarnya, prosesor model LPC dibagi kedalam 8 bagian seperti diperlihatkan pada

Gambar-1. Bagian-bagian tersebut adalah seperti berikut.

- b. *Preemphasis*; sinyal suara digital $s(n)$, dimasukkan kedalam sistem digital orde rendah (biasanya berupa filter FIR orde satu) untuk mereratakan spektrum sinyal.
- c. *Bingkai pemblok*; sinyal suara hasil *preemphasis* diblok kedalam beberapa bingkai N sampel suara.

- d. *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:

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \leq n \leq N-1, \quad (01)$$

- e. *Analisis autokorelasi*; setiap bingkai sinyal yang telah dijendelakan kemudian dianalisis autokorelasinya untuk memberikan:

$$r_t(m) = \sum \tilde{x}_t(n) \tilde{x}_t(n+m), \quad m = 0, 1, \dots, p, \quad (02)$$

dengan nilai autokorelasi tertinggi, p , adalah orde analisis LPC.

- f. *Analisis LPC*; pada tahap ini setiap bingkai dengan autokorelasi ke-($p+1$) akan dikonversi ke bentuk parameter-parameter LPC yang dapat berupa koefisien LPC, koefisien refleksi, *cepstral coefficients* atau transformasi yang lainnya sesuai kebutuhan.
- g. *Konversi parameter LPC menjadi koefisien cepstral*; LPC adalah koefisien *cepstral*, $c(m)$, yang ditentukan secara rekursif sebagai berikut:

$$c_0 = \ln \sigma^2, \quad (03)$$

$$c_m = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) c_k a_{m-k}, \quad 1 \leq m \leq p \quad (04)$$

$$c_m = \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) c_k a_{m-k}, \quad m > p, \quad (05)$$

dengan σ^2 adalah *gain* dalam model LPC.

- h. *Pembobotan parameter*; 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.

- i. *Turunan cepstral terhadap waktu*; representasi *cepstral* dari spektrum suara memberikan representasi yang bagus atas sinyal spektrum lokal untuk analisis bingkai. Peningkatan representasi dapat diperoleh melalui aproksimasi $\partial c_m(t) / \partial t$ dengan polinomial ortogonal pada jendela dengan panjang berhingga, yaitu :

$$\frac{\partial \hat{c}_m(t)}{\partial t} = \Delta c_m(t) \approx \mu \sum_{t=K}^K k c_m(t+k), \quad (06)$$

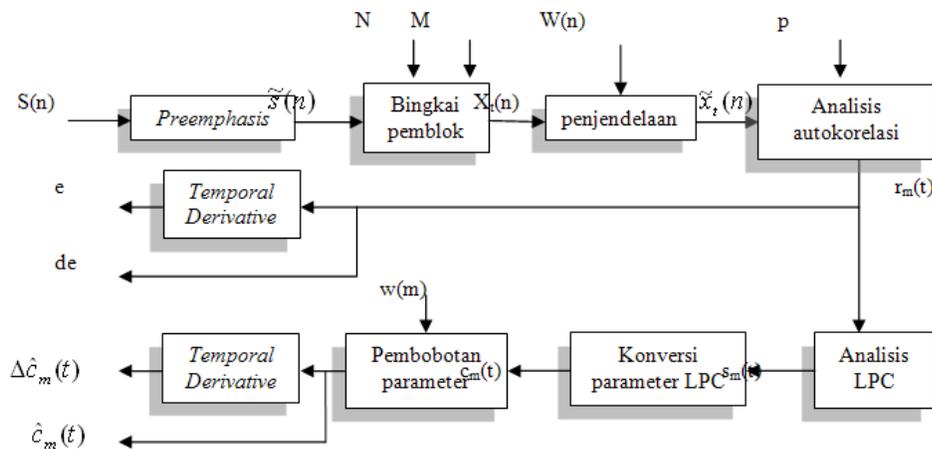
dengan μ adalah konstanta normalisasi. Untuk setiap bingkai t , hasil analisis LPC adalah vektor koefisien cepstral terbobot Q , yaitu :

$$Q = [\hat{c}_1(t); \hat{c}_2(t); \dots; \hat{c}_p(t)] \quad (07)$$

dengan o_t adalah komponen-komponen vektor 2Q dan menunjukkan matriks transpose.

Hasil analisis prosesor LPC berupa besaran parameter-parameter koefisien *cepstral* ($\hat{c}_m(t)$), turunan koefisien *cepstral* ($\Delta \hat{c}_m(t)$), besaran energi bingkai (e), dan besaran turunan energi frame (de) akan menjadi runtun observasi bagi HMM, dituliskan dalam bentuk:

$$o = [\hat{c}_m(t) \Delta \hat{c}_m(t) e] \quad (08)$$



Gambar-1. Diagram blok ekstraksi ciri dengan analisis LPC

2.2 Model Markov Tersembunyi

Model markov tersembunyi adalah salah satu bentuk model Markov dengan observasi yang merupakan sebuah fungsi probabilistik keadaan, artinya, model yang dihasilkan adalah sebuah proses stokastik yang disimpan secara berganda. Proses stokastik tidak dapat diobservasi langsung, tetapi dapat diobservasi **hanya melalui rangkaian proses stokastik yang menghasilkan runtun observasi.**

Suatu model HMM dinyatakan dengan simbol $\lambda = (A, B, \pi)$ yang menetapkan probabilitas untuk O, $P(O|\lambda)$, seperti berikut.

a. Distribusi probabilitas transisi keadaan $A = \{a_{ij}\}$ dengan:

$$a_{ij} = P[q_{t+1} = j | q_t = i], \quad 1 \leq i, j \leq N \quad (09)$$

Untuk kasus khusus bila setiap state dapat menjangkau setiap state lainnya pada satu langkah tunggal, maka $a_{ij} > 0$ untuk semua i, j .

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

$$b_j(k) = P[o_k = v_k | q = j], \quad 1 \leq k \leq M \quad (10)$$

menetapkan distribusi simbol dalam keadaan $j, j = 1, 2, \dots, N$.

c. Distribusi state inisial $\pi = \{\pi_i\}$, dengan:

$$\pi_i = P[q_1 = i], \quad 1 \leq i \leq N \quad (11)$$

3. METODE PENELITIAN

Bahan yang akan diteliti adalah ucapan kata-kata yang berkorelasi tinggi yang akan diucapkan oleh banyak penutur. Sampel kata adalah 8 buah kata bahasa Indonesia yaitu : *muka, muak, kamu, kaum, masuk, kamus, kusam, sukma*. Jumlah penutur sebanyak 85 orang dengan spesifikasi dinyatakan dalam Tabel-1.

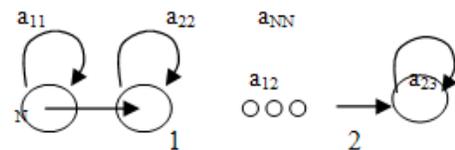
Tabel-1 Spesifikasi pengucap sebagai bahan penelitian

No.	Spesifikasi	Jumlah	Total Jumlah
1	Berdasarkan jenis kelamin	39	85
	▪ Laki-laki	46	
	▪ Perempuan		
2	Berdasarkan umur	43	85
	▪ umur < 20	31	
	▪ 20 ≤ umur ≤ 30	9	
	▪ 30 < umur ≤ 40	2	
	▪ umur > 40		

Adapun tahapan pelaksanaan penelitian adalah seperti berikut.

- a. Tahap pertama, pengambilan data suara dengan proses perekaman menggunakan perangkat keras mikrofon dan PC serta perangkat lunak perekaman yang sebelumnya telah dibuat dengan bantuan GUI pada program Matlab6P1. Hasil perekaman disimpan dalam berkas dengan format .mat
- b. Tahap kedua, melakukan proses ekstraksi ciri menggunakan fungsi-fungsi program yang sebelumnya telah dibuat di Matlab6P1. Proses ekstraksi ciri menggunakan model LPC dengan parameter-parameter $p = 8, N = 400$ (jumlah sampel per-bingkai), dan $M = 100$ (jarak antara bingkai yang berturutan).
- c. Tahap ketiga, melakukan proses pelatihan sebagai berikut:
 - o Inialisasi model $\lambda_0 = (A_0, \pi_0, \mu_0, U_0)$: Fungsi `init_pi_A` untuk inialisasi A_0 , dipilih $a_{ii} = 0,75; a_{i,i+1} = 0,25; a_{NN} = 1$, untuk $1 \leq i \leq N$, dan awal $\pi_i = 1$ untuk $i = 1$ dan $\pi_i = 0$ untuk $i \neq 1$. Fungsi `hmm_mint` untuk inialisasi μ_0 dan U_0 dengan cara mengelompokkan masing-masing runtun observasi (hasil ekstraksi ciri) kedalam N kelompok sesuai cacah keadaan HMM, dengan vektor rerata menjadi nilai awal μ_0 dan matriks kovarians menjadi U_0 .
 - o Reestimasi model $\lambda_i = (A_i, \pi_i, \mu_i, U_i)$: Fungsi `hmm_mest` untuk reestimasi A_i , fungsi `mix_par` untuk reestimasi parameter μ_0 dan U_0 .

Type pelatihan adalah terbimbing yaitu runtun observasi pelatihan akan diproses secara berturutan sesuai dengan urutan jenis kata terucapkan yang disimpan di dalam satu berkas (file) . mat. Model HMM yang digunakan adalah model kiri-kanan, seperti tampak pada Gambar-2.



Gambar-2 HMM kiri-kanan dengan N cacah keadaan

- d. dan menghitung persentasi keberhasilan pengujian, dan mengamati pada cacah keadaan berapa dicapai tingkat keberhasilan tertinggi.

4. HASIL PENELITIAN

Tabel-2 Koefisien korelasi antara kata yang menjadi bahan penelitian

	muka	muak	kamu	kaum	masuk	kamus	kusam	sukma
muka	1	0,5788	0,9420	0,8009	0,7161	0,9299	0,9379	0,7584
muak	0,5788	1	0,6581	0,8770	0,7643	0,7258	0,5205	0,7235
kamu	0,9420	0,6581	1	0,9069	0,8170	0,9787	0,9249	0,8948
kaum	0,8009	0,8770	0,9069	1	0,8934	0,9255	0,7581	0,9098
masuk	0,7161	0,7643	0,8170	0,8934	1	0,8718	0,7271	0,8497
kamus	0,9299	0,7258	0,9787	0,9255	0,8717	1	0,8992	0,8835
kusam	0,9379	0,5205	0,9249	0,7581	0,7271	0,8992	1	0,8161
sukma	0,7584	0,7235	0,8948	0,9098	0,8497	0,8835	0,8161	1

Tabel-3 Persentase rata-rata pengenalan kata seluruh pengucap dengan variasi cacah keadaan; (o = [c_m dc_m e de])

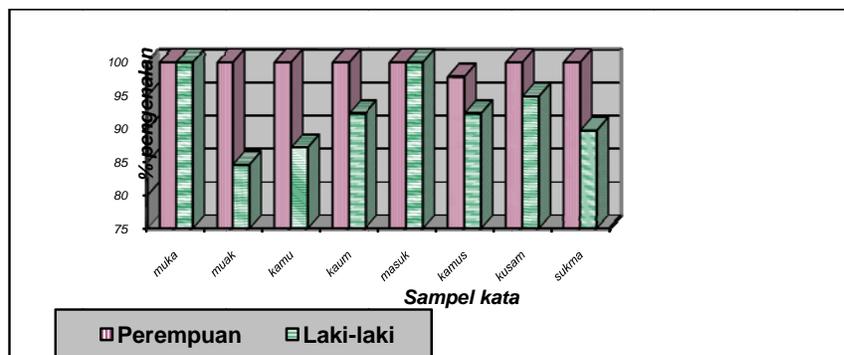
No.	Cacah State	Sampel Kata								Total Hasil
		muka	muak	Kamu	Kaum	masuk	kamus	kusam	sukma	
1.	5	100	97.65	97.65	95.29	100	94.12	98.82	96.47	97.50
2.	6	100	96.47	97.64	96.47	100	94.12	98.82	97.65	97.65
3.	7	100	98.82	100	100	100	96.47	98.82	100	99.27
4.	8	100	98.82	100	98.82	100	94.12	98.82	97.65	98.53

Tabel-4 Persentase rata-rata pengenalan kata seluruh pengucap dengan variasi cacah keadaan; (o = [c_m dc_m])

No.	Cacah State	Sampel Kata								Total Hasil
		Muka	muak	Kamu	Kaum	masuk	kamus	kusam	sukma	
1.	5	98.84	98.84	100	96.51	97.67	91.86	98.84	100	97.82
2.	6	98.84	97.67	98.84	95.51	98.84	93.02	98.84	100	97.82
3.	7	100	100	97.67	98.84	98.84	95.35	100	100	98.84
4.	8	100	98.82	97.65	100	100	95.29	100	100	98.97

Tabel-5 Rata-rata pengenalan runtun observasi ucapan laki-laki terhadap pelatihan runtun observasi ucapan perempuan (o = [cm dcm e de])

No.	Kelompok Pengucap	Sampel Kata								Total Hasil
		muka	muak	Kamu	kaum	masuk	kamus	kusam	sukma	
1.	Laki-laki	100	84.62	87.18	92.31	100	92.31	94.87	89.74	92.63
2.	Perempuan	100	100	100	100	100	97.83	100	100	99.73



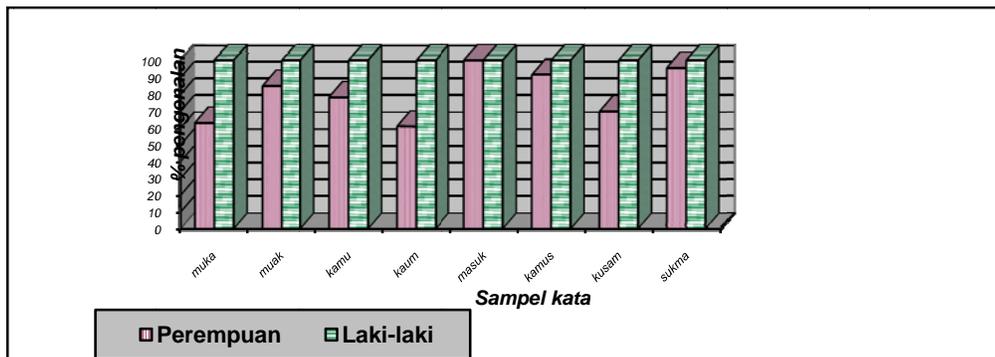
Gambar-3 Grafik rata-rata pengenalan runtun observasi ucapan laki-laki terhadap pelatihan runtun observasi ucapan perempuan pada cacah state 7

Tabel-6 Kata terkenali dari pengenalan ucapan laki-laki terhadap pelatihan ucapan perempuan berdasarkan pola referensi yang ada

Kata Terucap	Kata terkenali sebagai								
	muka	muak	kamu	kaum	masuk	kamus	kusam	sukma	X
muka	100	-	-	-	-	-	-	-	-
muak	5.13	84.62	-	-	-	2.56	2.56	5.13	-
kamu	-	-	87.18	2.56	-	7.69	-	2.56	-
kaum	-	-	-	92.31	-	-	-	7.69	-
masuk	-	-	-	-	100	-	-	-	-
kamus	-	-	5.13	-	2.56	92.31	-	-	-
kusam	2.57	-	-	-	-	-	94.87	2.57	-
sukma	-	-	-	-	-	-	5.13	89.74	-

Tabel-7 Rata-rata pengenalan runtun observasi ucapan perempuan terhadap pelatihan runtun observasi ucapan laki-laki (o = [cm dcm e de])

No.	Kelompok Pengucap	Sampel Kata								Total Hasil
		muka	Muak	Kamu	kaum	Masuk	Kamus	Kusam	sukma	
1.	Laki-laki	100	100	100	100	100	100	100	100	100
2.	Perempuan	63.04	84.78	78.26	60.87	100	91.30	69.57	95.65	80.43



Gambar-4 Grafik rata-rata pengenalan ucapan perempuan terhadap pelatihan ucapan laki-laki

Tabel-8 Kata terkenali dari pengenalan ucapan perempuan terhadap pelatihan ucapan laki-laki berdasarkan pola referensi yang ada

Kata Terucap	Kata terkenali sebagai								
	muka	muak	kamu	kaum	masuk	kamus	kusam	sukma	X
muka	63.02	2.17	-	-	-	-	2.17	26.10	6.52
muak	-	84.78	-	-	2.17	-	-	13.05	-
kamu	-	-	78.26	-	-	21.74	-	-	-
kaum	-	-	28.26	60.87	-	8.69	2.17	-	-
masuk	-	-	-	-	100	-	-	-	-
kamus	-	-	2.17	-	6.53	91.30	-	-	-
kusam	2.17	-	8.69	-	10.87	-	69.57	8.69	-
sukma	-	-	-	-	1.68	-	-	96.65	1.68

Tabel-9 Pengaruh orde analisis LPC terhadap pengenalan ucapan

No.	Orde LPC	Sampel Kata								Total Hasil
		Muka	muak	Kamu	kaum	masuk	kamus	kusam	Sukma	
1.	8	100	98.82	100	100	100	96.47	98.82	100	99.26
2.	9	100	98.82	100	95.29	100	94.18	98.82	96.47	97.94
3.	10	100	98.82	100	97.65	100	94.12	98.82	100	98.68

Tabel-10. Persentase pengenalan 24 ucapan laki-laki terhadap pelatihan 15 ucapan laki-laki

Kata Terucap	Kata terkenal sebagai								
	muka	muak	kamu	kaum	masuk	kamus	kusam	sukma	X
muka	91.67	-	-	-	-	-	-	4.17	4.17
muak	8.33	79.17	-	-	4.17	-	-	8.33	-
kamu	-	-	45.83	-	4.17	50	-	-	-
kaum	-	-	4.17	79.17	4.17	12.50	-	-	-
masuk	-	-	-	-	95.83	-	-	4.17	-
kamus	-	-	-	-	4.17	91.67	-	-	4.17
kusam	-	-	-	-	4.17	-	1.67	-	4.17
sukma	-	-	-	-	8.33	-	-	91.67	-

5. KESIMPULAN & SARAN

5.1. Kesimpulan

Dari hasil penelitian diatas dapat disimpulkan yang berikut.

- Ucapan kata-kata yang berkorelasi tinggi dapat dikenali oleh komputer setelah terlebih dahulu dilakukan proses ekstraksi ciri menggunakan model LPC dan proses klasifikasi menggunakan model HMM, dengan tingkat pengenalan cukup signifikan yaitu di atas 97%.
- Runtun observasi yang hanya mengandung koefisien cepstral dan koefisien derivatif cepstral sebagai hasil analisis model LPC dapat menjadi ciri-ciri besaran yang akan dikenali. Dalam penelitian ini, pengenalan untuk dua ciri besaran tersebut memberikan hasil rata-rata di atas 97%.
- Pengenalan mencapai tingkat keberhasilan tertinggi pada cacah keadaan 7 untuk runtun observasi dengan empat ciri-ciri besaran yaitu koefisien cepstral dan turunannya serta koefisien energi dan turunannya dimana rata-rata keberhasilan pengenalan untuk seluruh kata adalah 99%. Hampir seluruh kata dapat dikenali pada cacah keadaan 7 ini, kecuali untuk kata *muak*, *kamus* dan *kusam* dikenali masing-masing 98%, 96% dan 98%.
- Untuk pengenalan berdasarkan jenis kelamin ternyata tingkat pengenalan ucapan perempuan terhadap pelatihan ucapan laki-laki lebih rendah dibandingkan tingkat pengenalan ucapan laki-laki terhadap pelatihan ucapan perempuan, dimana pengenalan untuk masing-masing kata ada yang hanya 60% sedangkan pada ucapan laki-laki pengenalan terendah masing-masing kata adalah di atas 80%. Hal ini dapat disebabkan oleh karena harmoniks spektral suara perempuan terpisahkan 2 kali lebih besar dari harmoniks spektral suara laki-laki.
- Perubahan orde analisis LPC tidak memberikan pengaruh yang signifikan terhadap tingkat pengenalan ucapan kata berkorelasi tinggi dimana pada penelitian ini untuk orde LPC dari 8 sampai 10 tingkat pengenalan rata-rata di atas 97%.
- Korelasi antara kata yang sangat tinggi akan memberikan pengaruh terhadap tingkat pengenalan kata tersebut. Tingkat pengenalan

kata terucap kamu dikenali 50% sebagai kata kamus, hal ini dikarenakan korelasi antara kedua kata yang cukup tinggi yaitu 0,9787.

- Jumlah data pelatihan akan mempengaruhi besarnya tingkat pengenalan ucapan kata. Makin besar jumlah data untuk pelatihan, tingkat pengenalanpun akan menjadi lebih baik.

5.2 Saran-Saran

Adapun saran-saran yang dapat diusulkan sehubungan dengan pelaksanaan penelitian pengenalan pola ucapan adalah :

- Perlu menetapkan sampel kata-kata yang tepat sebelum mulai melakukan penelitian agar bahan penelitian yang dipilih dapat mewakili kelompok kata sejenis.
- Perlu dicari upaya yang tepat untuk proses pengumpulan data sehingga kesulitan dalam pengumpulan data bisa teratasi, misalnya mencoba perekaman kata terucap melalui tape recorder.
- Untuk proses pengumpulan data terucap perlu diperhatikan keseimbangan jumlah variasi kelompok pengucap, baik itu berdasarkan kelompok umur maupun berdasarkan kelompok jenis kelamin.

DAFTAR PUSTAKA

- Bristow, G., 1976, "Electronic Speech Synthesis", McGraww-Hill Book Company
- Deller, J. R., 1993, "Discrete-Time Processing of Speech Signals", Macmilon Publishing Company, a division of Macmillan, Inc.
- Fallside, F., dan Woods A. W., 1985, "Computer Speech Processing", Prentice-Hall, Inc.
- Rabiner, L., 1993, "Fundamentals of Speech Recognition", Prentice-Hall International, Inc.
- Schalkoff, R.J., 1992, "Pattern Recognition : Statistical, Structural and Neural Approach", John Willey & Sons Inc. Canada.
- Program Pasca Sarjana UGM.
- Wilpon, J.G., 1990, "Automatic Recognition of Keywords in Unconstrained Speech Using Hidden Markov Models", IEEE Transaction on Acoustic, Speech, and Signal Processing, vol. 38. No. 11.