Proceedings of NATIONAL CONFERENCE ON COMPUTER SCIENCE & INFORMATION TECHNOLOGY 2007
January 29-30, 2007, Faculty of Computer Science, University of Indonesia

Pengembangan Model HMM Berbasis Maksimum Lokal Menggunakan Jarak Euclid Untuk Sistem Identifikasi Pembicara

Agus Buono[(a)] dan Benyamin Kusumoputro[b]

[a]Computational Intelligence Research Lab, Dept of Computer Science, Bogor Agriculture University Dramaga Campus, Bogor – West Java, Indonesia
Email: pudesha@yahoo.co.id; buono@ilkom.fmipa.ipb.ac.id
[b]Computational Intelligence Research Lab, Faculty of Computer Science, University of Indonesia Depok Campus, Depok 16424, PO.Box 3442, Jakarta, Indonesia
Email: kusumo@cs.ui.ac.id

ABSTRAK

Paper ini membahas aplikasi dari *Hidden Markov Model* (HMM) yang dimoifikasi pada distribusi observasi menggunakan jarak Euclid serta *Mel-Frequency Cepstrum Coefficients* (MFCC) sebagai ekstraksi ciri. Untuk menentukan distribusi lokal, maka state dari left-right HMM yang dikembangkan pada penelitian ini diklasterkan menggunakan Fuzzy C-meant clustering. Nilai keanggotaan dengan rentang [0,1] yang digunakan pada penelitian ini adalah berbanding terbalik dengan jarak Euclid. Nilai ini berikutnya dipergunakan untuk menduga nilai peluang observasi. Nilai peluang observasi dari observasi baru pada suatu state adalah sesuai dengan jarak terdekatnya terhadap klaster state tersebut.

Pada kasus suara tanpa dikondisikan dengan 10 pembicara, akurasi sistem mencapai 88% untuk data testing dan 96.7% untuk data training. Sementara itu, akurasi sistem tanpa pengklasteran adalah 54%. Nilai ini jauh di atas HMM standar yang dikembangkan menggunakan distribusi Normal yang memiliki akurasi sekitar 42%. Salah satu kelemahan HMM standar adalah seringkali ditemui masalah singularitas saat melakukan pembalikan matriks koragam. Sementara itu, pada sistem yang dikembangkan hal ini ditemui.

Kata Kunci : Hidden Markov Model, Mel-Frequency Ceptrum Coefficients, Fuzzy Clustering, Euclid Distance

1. PENDAHULUAN

Sistem Pengenalan Suara secara Otomatis mempunyai penerapan yang luas pada berbagai area, sehingga penelitian ini selalu diminati banyak peneliti. Sistem Pengenalan Pembicara, *Automatic Speaker identification* (ASI) System adalah salah satu sistem pengenalan suara yang mengidentifikasi orang atau dari kelompok apa orang tersebut berasal berdasar suara tanpa adanya klaim sebelumnya mengenai orang tersebut, [2].

Hidden Markov Model (HMM) telah banyak dipergunakan dalam sistem pemrosesan suara dengan hasil yang memuaskan. Namun kinerja sistem menjadi turun saat menghadapi situasi real, [1]. Oleh karena itu, masih diperlukan modifikasi model agar sesuai dengan situasi real, sehingga akurasi sistem tetap baik. Kelemahan tersebut disebabkan oleh beberapa hal, seperti [3] : (1)asumsi kenormalan distribusi observasi, (2)asumsi kebebasan antar kemunculan state dengan observasi periode sebelumnya, (3)asumsi kebebasan antar observasi, (4) munculnya singularitas matriks koragam yang dikarenakan keterbatasan data training ya.

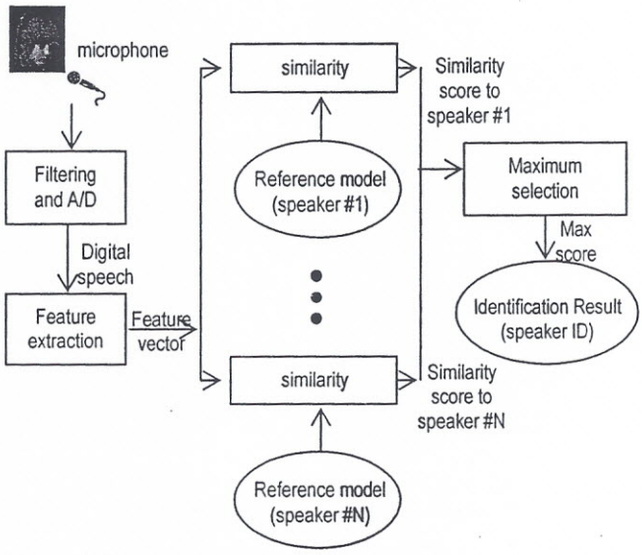
Penelitian yang dilakukan bertujuan untuk membangun model HMM sebagai *classifier* pada sistem pengenalan pembicara dengan ekstraksi ciri menggunakan *Mel-Frequency Ceptrums Coefficients* (MFCC) yang dikembangkan oleh Do, 1994, [4]. Inovasi yang dilakukan pada penelitian ini adalah bahwa peluang observasi yang sebelumnya menggunakan distribusi Normal pada HMM standar, digantikan dengan nilai keanggotaan yang didasarkan pada jarak Euclid. Pendekatan ini diharapkan mampu mengatasi masalah *non-Normality* serta *singularity* pada HMM standar.

1.1. Deskripsi Masalah

Suara manusia muncul karena adanya hembusan udara dari paru-paru, melewati suatu konfigurasi artikulator tertentu. Suara ini sebagai representasi pikiran yang diungkapkan mengikuti kaidah bahasa dan dipengaruhi oleh emosi, dialek serta medium antara. Oleh karena itu sinyal suara bersifat kompleks, selain faktor

ISSN: 0126-2866

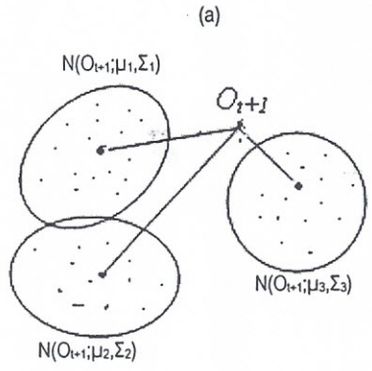
semantik, linguistik, artikulator dan akustik, juga variabilitas karena emosi, umur, kesehatan, jenis kelamin serta dialek. Secara generik, sistem identikasi pembicara disajikan pada Gambar 1, [4].



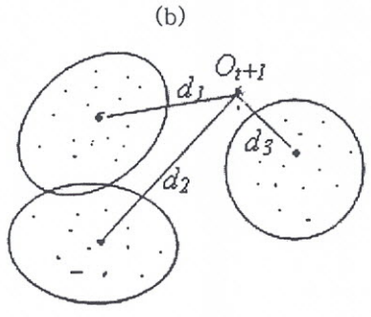
Gambar 1. Blok Diagram Sistem Identikasi Pembicara

Dalam HMM, sebuah ujaran dimodelkan dengan *Directed Acyclic Graph* (DAG) dengan setiap node merepresentasikan satu konfigurasi artikulator. Oleh karena kita tidak dapat mengamati langsung node ini, maka disebut dengan *hidden state*. *Link* antar state merepresentasikan transisi dari satu konfigurasi ke konfigurasi lain sesuai unit bunyi tertentu. Relasi ini dalam model HMM diparametisasi oleh matriks peluang transisi yang berdimensi NxN, dengan N adalah banyaknya kemungkinan hidden state tersebut. Dalam kenyataannya, yang diketahui adalah sinyal yang dihasilkan oleh setiap konfigurasi, sehingga sinyal ini sebagai *observable state*. Dalam Gaussian HMM (HMM standar), *observable state* adalah peubah acak dan diasumsikan berdistribusi Normal dengan vektor rataan μ_i dan matriks koragam Σ_i (i=1, 2, 3, ..., N. Oleh karena masalah variabilitas sumber suara, maka distribusi setiap state bersifat multimodal, yang dalam HMM standar dimodelkan dengan Distribusi Normal Campuran (Mixture Gaussian Distribution). Namun demikian tidak ada jaminan mengenai asumsi ini serta dengan model Gaussian, seringkali muncul masalah dari aspek komputasi.

Pendekatan yang diajukan dalam penelitian ini untuk menangani masalah tersebut adalah sebagai berikut : pertama, untuk mengakomodai distribusi multimodal, maka setiap *hidden state* diklasterkan. Kedua, peluang observasi didekati dengan nilai keangggotaan yang didefinisikan menggunakan jarak Euclid. Nilai keanggotaan dihitung untuk setiap klaster state, dan dipilih nilai terbesar sebagai peluang observasi untuk state tersebut. Oleh karena itu, pendekatan ini diharapkan robust terhadap ketidaknormalan dan mengatasi masalah komputasi. Gambar 2 menyajikan pembandingan dari kedua metode tersebut.



$$b_j(O_{t+1}) = \sum_{i=(1,2,3)} c_i N(O_{t+1}, \mu_i, \Sigma_i)$$



$$b_j(O_{t+1}) = \mu_j(O_{t+1}) = \frac{1}{1 + \min_{k\in\{1,2,3\}} \{d_j(O_{t+1}, k)\}}$$

Gambar 2. Pembandingan Penghitungan Peluang Observasi antara HMM standar (a) dengan HMM-Euclid Distance (b).

1.2. HMM Untuk Identifikasi Pembicara

Proses stokastik adalah [5] adalah keluarga peubah acak, $\{X_t\}_{t=1}^{\infty}$, dengan t parameter indeks. Proses stokastik yang bersifat bahwa : jika diketahui nilai X_t, maka nilai X_s untuk s>t tidak

tergantung dari X(u) untuk u<t, disebut sebagai rantai Markov. Untuk kondisi ruang state (himpunan yang berisi semua kemungkinan nilai peubah acak X_t) adalah *finite* atau tercacah dan nilai indeks adalah bilangan cacah, maka dikenal dengan Rantai Markov Waktu Diskret, *discrete time Markov Chain*. Jika nilai peubah acak X_t hanya tergantung pada X_{t-1} dan ketergantungan ini bebas dari indeks t, maka dikenal sebagai Rantai Markov Stationer Orde Satu. Jika kita tidak dapat mengamati barisan peubah acak X_t tersebut secara langsung, namun pengamatan dilakukan pada sinyal sebagai cerminan dari X_t, maka proses ini dikenal dengan nama *Hidden Markov Model*, HMM.

Suatu HMM secara lengkap dispesifikasikan oleh tiga komponen, yaitu distribusi awal state, Π; Matriks peluang transisi, A; dan matriks peluang observasi, B. HMM tersebut dinotasikan sebagai $\lambda = (A,B,\Pi)$, dengan :

A:matriks transisi berdimensi NxN dengan entri $a_{ij}=P(X_{t+1}=j|X_t=i)$, N adalah banyaknya hidden state

B:matriks observasi dengan entri $b_{jk}=P(O_{t+1}=v_k|X_t=j)$, k=1, 2, 3, ..., M; M adalah banyaknya kemungkinan state terobservasi

Π : vektor ditribusi awal state berdimensi Nx1, dengan entri $\pi_i=P(X_1=i)$

Untuk HMM Gaussian, matriks B berisi vektor rataan dan matriks koragam untuk setiap state, μ_i dan Σ_i, i=1, 2, 3, ..., N. Formula untuk $b_j(O_{t+1})$ adalah $N(O_{t+1},\mu_j,\Sigma_j)$, yang dirumuskan sebagai :

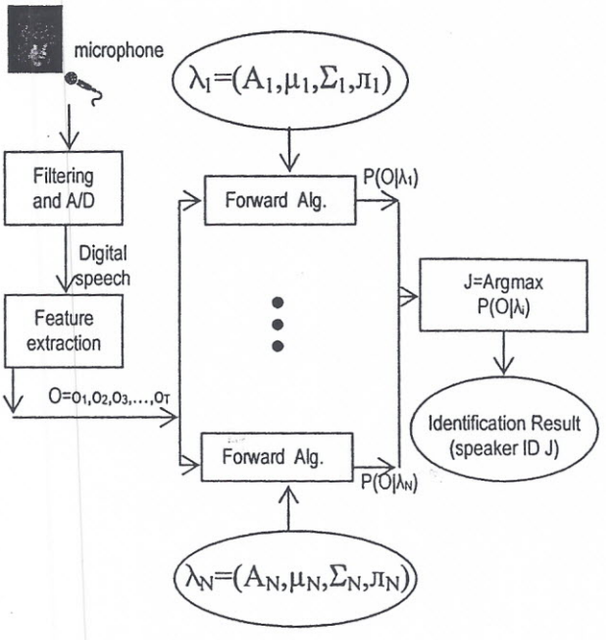
$$N(\mu_j,\Sigma_j)=\frac{1}{(2\pi)^{d/2}|\Sigma_j|^{1/2}}\exp\left[-\frac{1}{2}(O_{t+1}-\mu_j)\Sigma_j^{-1}(O_{t+1}-\mu_j)'\right] \quad (1)$$

Ada tiga masalah terkait dengan HMM, [1], yaitu : Masalah Evaluasi, $P(O|\lambda)$; Masalah Dekoding, $P(Q|O,\lambda)$; dan Masalah Pelatihan Model untuk menduga nilai parameter model, A, B, and Π. Blok diagram sistem pengenalan pembicara menggunakan HMM sebagai classifier disajikan pada Gambar 3.

HMM untuk setiap pembicara dilatih dengan algoritma K-segmental yang dikembangkan oleh Rakesh, 1996 [6]. Evaluasi terhadap observasi baru menggunakan algoritma Forward untuk menghitung $P(O|\lambda_i)$, i=1, 2, 3, ..., n (n adalah banyaknya pembicara, yang dalam penelitian ini adalah 10). Dengan HMM Gaussian, peluang observasi $b_j(O_{t+1})=P(O_{t+1}|X_t=j)$ dihitung menggunakan formula (1).

II. PERBAIKAN METODE

Paling sedikit ada 2 masalah dengan HMM Gaussian. Pertama, tidak adanya jaminan bahwa asumsi kenormalan dipenuhi. Kedua, masalah singularitas matriks koragam, terutama untuk dimensi yang besar dan terbatasnya data training. Untuk mengatasi hal ini, metode yang diajukan adalah nilai peluang observasi, $b_j(O_{t+1})=P(O_{t+1}|X_t=j)$, didekati dengan nilai keanggotaan, $\mu_j(O_{t+1})$. Dalam penelitian ini nilai $\mu_j(O_{t+1})$ dihitung sebagai kebalikan jarak Euclid. Untuk mengantisipasi adanya distribusi multimodal, maka sebelumnya dilakukan pengklasteran terhadap setiap state menggunakan *Fuzzy C-Mean Clustering* (FCM).



Gambar 3. Blok Diagram Sistem Pengenalan Pembicara Menggunakan HMM

Dalam pendekatan ini, komponen HMM untuk setiap pembicara adalah vektor distribusi awal state, Π; matriks peluang transisi, A; serta matriks P_j, yaitu matriks dengan entri vektor pusat klaster k untuk state j, j=1, 2, 3, ..., N. Banyaknya klaster adalah c serta dimensi data d, maka matriks P_j dituliskan sebagai :

$$P_j=\begin{bmatrix}p_1\\p_2\\\cdots\\p_c\end{bmatrix}=\begin{bmatrix}x_{11} & x_{12} & \cdots & x_{1d}\\x_{21} & x_{22} & \cdots & x_{2d}\\\cdots & \cdots & \cdots & \cdots\\xc1 & x_{c2} & \cdots & x_{cd}\end{bmatrix} \quad j=1, 2, 3, \ldots, N \quad (2)$$

Untuk observasi baru, O_{t+1}, dihitung :

$$D_j(O_{t+1},k)=\sum_{i=1}^{d}(x_{ki}-o_{t+1,i})^2 \quad (3)$$

ISSN: 0126-2866

untuk k=1, 2, 3, ..., c. Nilai keanggotaan O_{t+1} pada state j, $\mu_j(O_{t+1})$, dihitung dengan formula (j=1, 2, 3, ..., N) :

$$b_j(O_{t+1}) = \mu_j(O_{t+1}) = \frac{1}{1 + \min_{k \in \{1,2,..c\}} \{d_j(O_{t+1}, k)\}} \quad (4)$$

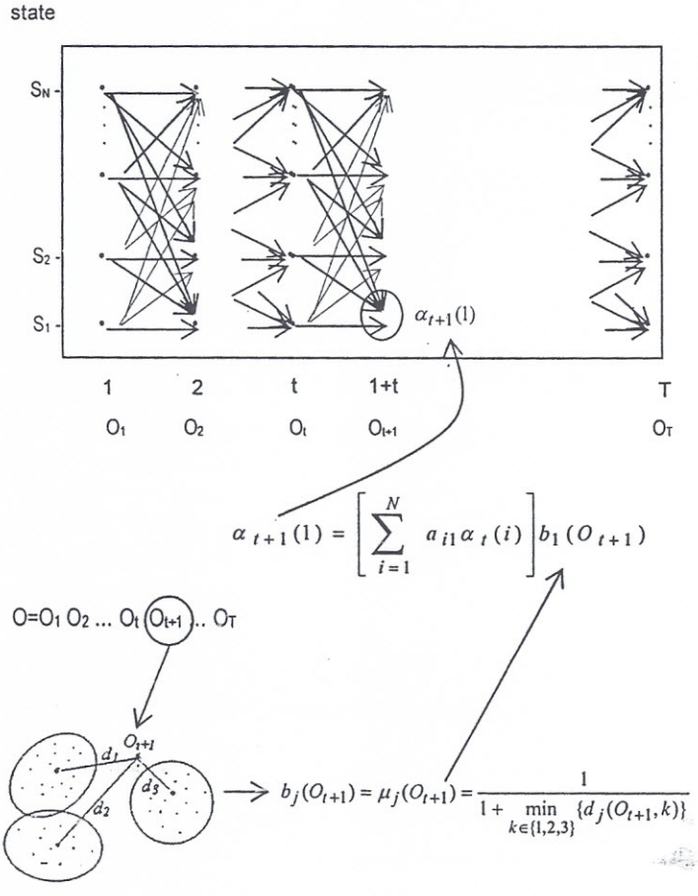
Dengan pendekatan ini, step 3 dan 4 pada algoritma K-segmental untuk melatih HMM yang dikembangkan oleh Rakesh [6] menjadi :

step 3 : Hitung matriks pusat P_j, j=1, 2, 3, ..., N

a. Klasterkan setiap observasi dengan label j ke dalam c klaster menggunakan FCM, j=1, 2, 3, ..., N
b. Assign pusat klaster k, (k=1, 2, 3, ..., c), dari state j ke baris ke k, dari matriks P_j, (j=1, 2, 3, ..., N).

step 4 : Hitung peluang observasi, $b_j(O_{t+1})=\mu_j(O_{t+1})$ using formula (4)

Gambar 4 menyajikan visualisasi proses komputasi dari metode yang diajukan.



Gambar 4. Visualisasi Proses Penghitungan Dalam HMM Berbasis Jarak Euclid

III. EXPERIMENTS AND RESULT

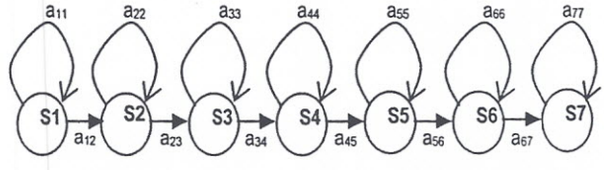
3.1. Data

Data suara diperoleh dari 10 pembicara yang mengucapkan kata 'PUDESHA' dalam kondisi yang **tidak** dibatasi panjang-pendek serta tekanannya. Masing-maisng pembicara mengucapkan sebanyak 40 kali. Data suara disampling dengan sampling rate 11 kHz dan durasi waktu 1.28 detik. Oleh karena itu diperoleh total 400 suara, 300 suara sebagai data training (masing-maisng pembicara 30 suara) dan sisanya sebagai data testing.

Selanjutnya pada setiap suara dilakukan praproses yang terdiri dari windowing dengan lebar window 30 ms tanpa overlaping, sehingga diperoleh 43 observasi (window) untuk setiap suara. Berikutnya adalah menghitung koefisien MFCC untuk setiap setiap observasi (window), yang dalam hal ini diambil 13 dan 26 koefisien. Oleh karena itu untuk setiap suara dikonversi menjadi 43 barisan observasi berdimensi 13 dan 26 sesuai banyaknya koefisien yang diambil.

3.2. Struktur HMM

Pada penelitian ini, struktur HMM yang dipilih adalah struktur *left-right* dengan 7 hidden state. Struktur left-right dipilih karena kesesuaian alamiahnya dengan sinyal suara. Sedangkan pemilihan 7 state didasarkan pada beberapa percobaan terdahulu bahwa HMM dengan jumlah state 7 memberikan hasil yang optimal. Selain itu juga bahwa kata yang dipilih, yaitu "PUDESHA" terdiri dari 7 unit bunyi. Left-right HMM dengan 7 state disajikan dalam Gambar 5.



Gambar 5. Struktur Left right HMM dengan 7 State untuk model kata "pudesha"

Matriks transisi,A, dan vektor peluang awal state, л, untuk HMM di atas adalah :

$$A = \{a_{ij}\} = \begin{pmatrix} a_{11} & a_{12} & 0 & 0 & 0 & 0 & 0 \\ 0 & a_{22} & a_{23} & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{33} & a_{34} & 0 & 0 & 0 \\ 0 & 0 & 0 & a_{44} & a_{45} & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{55} & a_{56} & 0 \\ 0 & 0 & 0 & 0 & 0 & a_{66} & a_{67} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \text{ dan } \pi = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Dengan struktur ini, berikutnya adalah melatih model untuk setiap pembicara dengan data training yang terdiri dari 30 suara. Pelatihan ini dilakukan dengan algoritma K-segmental yang dikembangkan oleh Rakesh [6] yang dimodifikasi sesuai dengan pendekatan yang diajukan dalam penelitian ini (kecuali untuk HMM Gaussian).

3.3. Result

Hasil percobaan disajikan dalam Tabel 1.

Table 1. Perbandingan Tingkat Akurasi Sistem (%) untuk 10 Pembicara

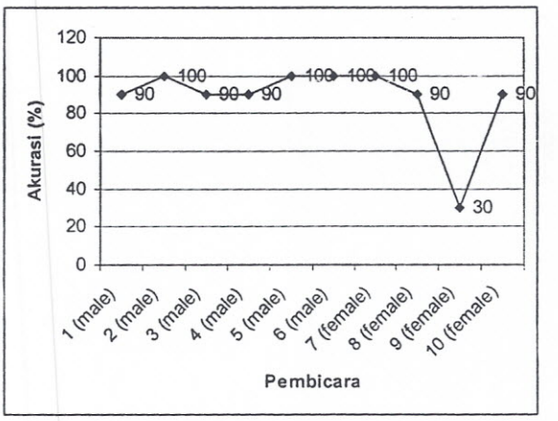
Methods	Data Pelatihan		Data Testing	
	d=13	d=26	d=13	d=26
HMM's Gaussian	50	Fail	42	Fail
HMM Euclidean tanpa state Clustering	70	70	54	54
HMM Euclidean dengan state Clustering	94.7	96.7	88	85

Terlihat bahwa HMM berbasis jarak Euclid mampunyai akurasi jauh di atas akurasi dari HMM standar. Selain itu juga terlihat bahwa pengklasteran setiap state mampu meningkatkan akurasi sangat nyata, yaitu sekitar 30 %. Untuk data uji, akurasi terbaik adalah 88% dan untuk data pelatihan, akurasi mencapai 96.7%.

Dari tabel di atas juga terlihat bahwa pemilihan banyaknya koefisien dari 13 menjadi 26 tidak memberikan pengaruh yang nyata. Bahkan untuk model HMM berbasis Euclidean, model dengan jumlah koefisien 26 memberikan akurasi (85 %) di bawah akurasi dari model dengan jumlah koefisien 13 (88 %). Juga terlihat bahwa pemilihan jumlah koefisien sebanyak 26 menyebabkan HMM standar gagal melalukan pelatihan model yang dikarenakan masalah singularitas matriks koragam.

Dengan 13 koefisien MFCC, tingkat akurasi untuk HMM standar hanya 50 % and 42 % masing-masing untuk data pelatihan dan data testing. Dengan HMM berbasis jarak Euclid tanpa pengklasteran state, tingkat akurasi meningkat lebih dari 10% dibandingkan dengan HMM standar. Untuk meningkatkan akurasi, dilakukan pengklasteran setiap state menjadi 3 klaster, dan peluang observasi pada indeks t+1 dari state j dihitung sesuai dengan jarak Euclid terkecil dari observasi tersebuut ke klaster state j. Dengan pendekatan ini akurasi naik dari 42 % (standard HMM) ke 88 % untuk data testing, dari 50 % to 94.7 % untuk data training. Keuntungan lain dari metode ini adalah bahwa metode ini bebas dari ketidaknormalan dan mampu berjalan pada kondisi data training yang terbatas serta dimensi yang tinggi.

Gambar 6 menyajikan tingkat akurasi pada setiap pembicara untuk data testing.



Gambar 6. Tingkat Akurasi untuk Data Testing padaSetiap Pembicara

Dari Gambar 6 terlihat bahwa pada hampir semua pembicara akurasi mecapai di atas 90 %, bahkan ada 4 pembicara dengan akurasi 100%. Kalau dilihat sesuai jenis kelamin, terlihat bahwa secara rata-rata akurasi untuk pembicara laki-laki adalah 96.11 % dan 70 % untuk perempuan.

Tabel 2 menyajikan klasifikasi untuk 10 pembicara untuk data tetsing menggunakan metode HMM berbasis jarak Euclid.

ISSN: 0126-2866

Table 2. Hasil Klasifikasi untuk Data Testing Menggunakan HMM Berbasis Jarak Euclid

Pembicara	Dikenali sebagai Pembicara ke									
	1	2	3	4	5	6	7	8	9	10
1	9	0	0	0	1	0	0	0	0	0
2	0	10	0	0	0	0	0	0	0	0
3	0	0	9	0	1	0	0	0	0	0
4	0	0	0	8	2	0	0	0	0	0
5	0	0	0	0	10	0	0	0	0	0
6	0	0	0	0	0	10	0	0	0	0
7	0	0	0	0	0	0	10	0	0	0
8	0	0	0	0	0	0	0	9	1	0
9	0	0	0	0	0	0	0	6	3	1
10	0	0	0	0	1	0	0	0	0	9

Dari Tabel 2 dan Gambar 6 terlihat bahwa akurasi sistem drop pada pembicara ke 9. Kesalahan klasifikasi terjadi karena terdeteksi sebagai pebmicara ke 8. Pembicara 8 dan 9 adalah dua bersaudara, yang berumur 12 tahun dan 9 tahun. Dari data yang ada, secara alamiah kedua pembicara tersebut sulit dibedakan oleh telinga manusia. Pembicara ke 10 juga bersaudara dengan pembicara 8 dan 9. Namun umur pembicara 10 relatif berbeda jauh dengan kedua pembicara tersebut, yaitu 5 tahun. Data menunjukkan bahwa sistem mampu mendeteksi pembicara ke 10 dengan cukup baik. Ini mengatakan bahwa sistem masih harus dikembangkan lagi untuk situasi dimana terdapat pembicara yang saling berhubungan keluarga dan dengan umur relatif sama (<4 tahun).

IV. KESIMPULAN

Model left-right HMM sebagai *classifier* dengan tujuh state dan MFCC sebagai ekstrksi ciri dengan mengambil 13 koefisien dapat diterapkan pada sistem identifikasi pembicara, dan memberikan akurasi yang memadai untuk situasi pembicara yang tidak dibatasi panjang-pendek dan tekanan dalam pengucapan.

Hasil percobaan memperlihatkan bahwa sistem mampu mendeteksi dengan akurasi terbaik 88 % untuk data testing dan 96.7% untuk data training. Sementara HMM standar hanya memberikan akurasi 42 % dan 50%. Akurasi sistem meningkat lebih dari 30 % setelah dilakukan pengklasteran pada setiap state. Juga terlihat bahwa sistem kurang baik dalam mengidentifikasi pembicara yang saling bersaudara dengan perbedaan umur yang tidak besar (<4 tahun)

REFERENCES

[1] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", Proceeding IEEE, Vol 77 No. 2, pp 257-289, 1989.

[2] J. Campbell, "Speaker Recognition: A Tutorial", Proc. of the IEEE, Vol 85, No. 9, pp 1437-1462, 1997.

[3] Farbod H. and M. Teshnehlab, "Phoneme Classification and Phonetic Transcription Using a New Fuzzy Hidden Markov Model", WSEAS Transactions on Computers, Issue 6, Vol. 4, 2005.

[4] Do MN. *Digital Signal Processing Mini-Project: An Automatic Speaker Recognition System*. Audio Visual Communications Laboratory, Swiss Federal Institute of Technology, Lausanne, Switzerland. http://lcavwww.epfl.ch/~minhdo/asr_project/asr_project.pdf. 1994. [December 12 2006]

[5] Taylor, H.M. and Samuel Karlin. An Introduction to Stochastic Modeling. Academic Press, Inc. Florida, 1984.

[6] Rakesh D. "A Tutorial on Hidden Markov Model. Technical Report, Departement of Electrical Engineering, Indian Institute of Technology, Bombay", 1996