

Sistem Pengenalan Dan Verifikasi Pembicara HMM

Budi Darmawan¹, Suthami Ariessaputra²

^{1,2}Jurusan Teknik Elektro, Fakultas Teknik, Universitas Mataram

^{1,2}Jl. Majapahit No. 62 Mataram NTB, 83125

¹budidarmawan@unram.ac.id, ²suthami@unram.ac.id

Abstract— the weakness of the HMM speaker recognition system is to group all inputs into one of the speakers contained in the database, although the voice input does not include the voice of one of the speakers listed in the system database, and this is a mistake. In this study, the HMM speaker recognition system will be integrated with HMM speaker verification system so that a person's non-voice signal from one of the speakers contained in the database will be rejected by the system. In this study will also be examined some number of states and some threshold value used to see the number of state and threshold value which is best for use on the system. Voice signal data is the result of voice recording obtained from twenty speakers. Ten speakers will be built HMM model and will be entered into the system database. Each of these ten pronouns the word "HADIR" 30 times. 10 pieces of voice cues will be used in the training process and 20 other voice cues are used for the testing process. As for the other 10 speakers say the same word 20 times that will be used in the testing process. The results of the experiment show that the best recognition system is in the HMM model with the number of state 4, the value of% threshold -4% and with an average accuracy of 73.75%.

Keywords-Hidden Markov Models; Speaker recognition; verification, MFCC.

I. PENDAHULUAN

Sistem pengenalan pembicara dan sistem verifikasi pembicara merupakan bagian dari cabang pengolahan sinyal atau isyarat suara dari seorang pembicara. Sistem pengenalan pembicara bertujuan untuk mengenali identitas seorang pembicara melalui masukan suara yang diberikan kepada sistem. Sedangkan sistem verifikasi pembicara bertujuan untuk menolak atau menerima masukan sinyal suara yang diberikan merupakan identitas seorang pembicara atau tidak. Sistem Pengenalan pembicara menggunakan HMM telah banyak diteliti oleh para peneliti antara lain [1], [2], [3], [4] dan [5] begitu juga sistem verifikasi pembicara juga telah banyak diteliti antara lain [6], [7] dan [8]. Sedangkan Pada penelitian ini akan digabungkan antara sistem pengenalan pembicara HMM dengan sistem verifikasi pembicara HMM.

Dalam sistem pengenalan pembicara HMM, sebuah isyarat suara dari seorang pembicara diekstraksi cirinya menggunakan salah satu metode ekstraksi ciri untuk mendapatkan runtun observasi. Runtun observasi ini akan akan dicocokkan dengan seruh model HMM suara pembicara yang terdapat dalam database. Pencocokan ini dilakukan dengan cara menghitung probabilitas runtun observasi terhadap seluruh model HMM suara pembicara dengan menggunakan algoritma viterbi[10]. Model HMM suara pembicara yang terpilih adalah yang mempunyai probabilitas yang paling besar.

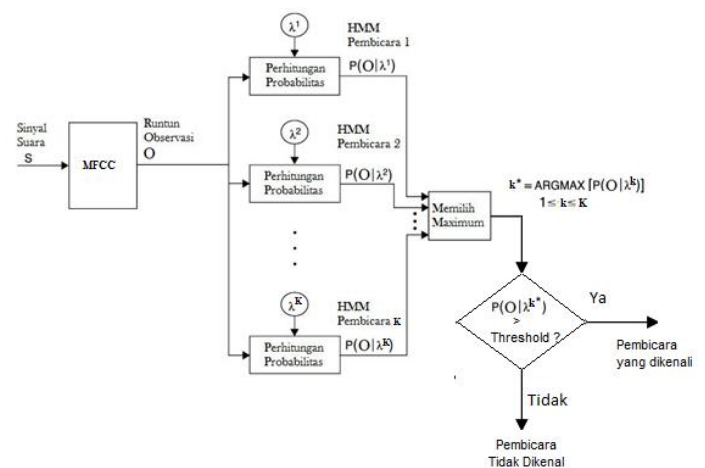
Begitu juga apabila sinyal suara tersebut bukan termasuk suara dari salah seorang pembicara yang terdapat dalam database, akan dilakukan hal yang sama sehingga sinyal suara tersebut dikenali sebagai suara dari salah seorang pembicara yang terdapat dalam data base. Dan ini merupakan salah satu kelemahan dari sistem pengenalan pembicara HMM.

Dalam penelitian ini, sistem pengenalan pembicara HMM akan dipadukan dengan sistem verifikasi pembicara HMM sehingga sinyal suara seseorang yang bukan suara dari salah seorang pembicara yang terdapat dalam database akan ditolak oleh sistem. Dalam penelitian ini juga akan diteliti beberapa jumlah state dan beberapa nilai *threshold* yang digunakan untuk melihat jumlah state dan nilai *threshold* yang mana yg paling baik untuk digunakan pada sistem.

II. METODOLOGI

A. Prinsip Kerja Sistem Pengenalan Pembicara HMM

Sistem yang dibangun pada penelitian ini adalah sistem pengenalan pembicara yang digabung dengan sistem verifikasi pembicara. Sistem ini bertujuan untuk mengenali identitas seorang pembicara dari sinyal suara yang dimasukkan kedalam sistem sekaligus melakukan verifikasi apakah sinyal suara tersebut benar-benar suara pembicara yang terdapat dalam database atau tidak. Gambar 1 menunjukkan sistem yang dibangun pada penelitian ini.



Gambar 1. Sistem Pengenalan dan Verifikasi Pembicara HMM

Berdasarkan gambar, maka algoritma sistem verifikasi pembicara dengan HMM adalah sebagai berikut:

- Untuk setiap pembicara k yang dimasukkan kedalam database harus dibangun sebuah model HMM (λ).
- Sinyal suara yang akan dikenali terlebih dahulu dilakukan ekstraksi ciri untuk mendapatkan runtun observasi O .
- Kemudian melakukan perhitungan probabilitas observasi (B) dengan menggunakan persamaan:

$$b_j(O_t) = \frac{1}{1 + d(O_t, \mu_j)} \quad (1)$$

dengan

$$d(O_t, \mu_j) = \sqrt{\sum_{k=1}^M (o_{tk} - \mu_{jk})^2} \quad (2)$$

dimana $d(O_t, \mu_j)$ merupakan jarak euclidean antara runtun observasi pada waktu t dengan rata-rata runtun observasi pada state j . Dan M adalah jumlah elemen runtun observasi pada state j .

- Menghitung probabilitas model yang ada, $P(O | \lambda^k)$ untuk $1 \leq k \leq K$, dengan menggunakan algoritma viterbi.
- Pembicara yang dikenali adalah pembicara yang mempunyai probabilitas paling tinggi yang diestimasi dengan cara berikut

$$k^* = \arg \max_{1 \leq k \leq K} [P(O | \lambda^k)] \quad (3)$$

- Sebelum diputuskan bahwa k^* merupakan pembicara yang dikenali terlebih dahulu $P(O | \lambda^{k^*})$ dibandingkan dengan sebuah nilai *threshold*. Jika $P(O | \lambda^{k^*})$ lebih besar daripada nilai *threshold* yang ditentukan maka sinyal suara tersebut diterima sebagai suara dari pembicara k^* , dan apabila sebaliknya maka sinyal suara tersebut tidak dikenali sebagai salah satu suara pembicara yang terdapat dalam database sistem.

Untuk setiap pembicara yang terdapat dalam database ditentukan nilai *threshold* tersendiri yang dapat dicari dengan menggunakan persamaan:

$$Threshold = P(\min) + (ab/100) \quad (4)$$

dengan:

$$a = P(\max) - P(\min) \quad (5)$$

Dimana:

$P(\max)$: nilai probabilitas dari salah satu runtun observasi data latih terhadap model HMM yang menghasilkan nilai probabilitas yang paling besar.

$P(\min)$: adalah nilai probabilitas dari salah satu runtun observasi data latih terhadap model HMM yang menghasilkan nilai probabilitas yang paling kecil.

b : % *threshold* Selisih nilai dari $P(\min)$ (dalam persen(%))

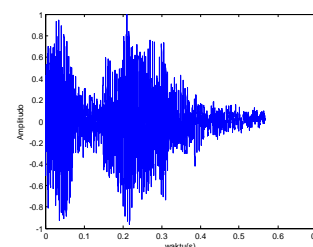
Pada penelitian ini diuji nilai b atau %*threshold* dari -30% sampai dengan 30% untuk melihat nilai *threshold* yang paling memberikan hasil akurasi yang paling baik pada sistem.

B. Data Isyarat Suara

Data isyarat suara merupakan hasil perekaman suara dari dua puluh orang pembicara. Sepuluh orang pembicara akan dibangun model HMMnya dan akan dimasukkan kedalam database sistem. Masing-masing dari sepuluh orang ini mengucapkan kata "HADIR" sebanyak 30 kali. 10 buah isyarat suaranya akan digunakan pada proses pelatihan dan 20 buah isyarat suaranya yang lain digunakan untuk proses pengujian. Sedangkan untuk 10 orang pembicara yang lainnya mengucapkan kata yang sama sebanyak 20 kali yang akan digunakan pada proses pengujian. Sehingga didapatkan data untuk proses pengujian berupa:

- 200 data isyarat suara pembicara yang terdapat dalam database untuk menguji apakah sistem mampu untuk mengenali semua data tersebut. Data ini yang selanjutnya akan disebut dengan "data isyarat suara pembicara".
- 200 data isyarat suara selain pembicara diluar database untuk menguji apakah sistem mampu untuk menolak semua data tersebut. Data ini yang selanjutnya akan disebut dengan "data isyarat suara selain pembicara".

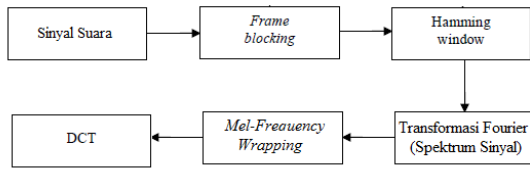
Isyarat suara direkam pada frekuensi pencuplikan 44100 Hz dengan resolusi 32 bit. Gambar 2 memperlihatkan salah satu bentuk isyarat suara dari salah seorang pembicara.



Gambar 2. Bentuk salah satu isyarat suara dari salah seorang pembicara

C. Ekastraksi Ciri

Penelitian ini menggunakan ekstraksi ciri *Mel-Frequency Cepstrum Coefficients (MFCC)*. Diagram blok dari MFCC dapat dilihat pada Gambar 3.



Gambar 3. Blok diagram MFCC

Pada proses ekstraksi ciri ini, setiap isyarat suara direkam dengan durasi waktu 0,56 detik, setelah itu dilakukan praproses framing dengan lebar frame 23,2 ms dengan overlapping sebesar 11,7 ms. Sehingga diperoleh 47 buah frame setiap kata. Selanjutnya dilakukan perhitungan koefisien MFCC untuk setiap frame, yang dalam hal ini diambil 12 koefisien MFCC untuk setiap frame. Sehingga setiap suara akan dikonversi menjadi 47 observasi berdimensi 12. Setelah itu data hasil ekstraksi cirri akan dibagi berdasarkan jumlah state yang digunakan dalam model HMM.

Langkah-langkah dalam ekstraksi ciri MFCC adalah sebagai berikut :

- 1) *Frame blocking*: pada langkah ini, isyarat suara dibentuk dalam frame N cuplikan dengan frame yang berdekatan terpisah dengan jarak M cuplikan. Nilai N yang digunakan disini adalah 1024, sedangkan nilai M yang digunakan adalah 512.
- 2) *Windowing*: window yang digunakan pada penelitian ini adalah Hamming Window. Untuk setiap isyarat yang sudah dibentuk dalam frame dikalikan dengan hamming window.
- 3) *Transformasi fourier*: transformasi fourier dilakukan untuk setiap frame yang telah dibentuk. Tujuannya adalah untuk merubah isyarat dari domain waktu menjadi domain frekuensi. Karena hasil dari transformasi fourier ini adalah simetris, maka hanya diambil setengah dari hasil transformasinya saja. Setelah itu hasil ini dikalikan dengan konjugatannya untuk mendapatkan spektrum isyarat.

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi kn/N} \tag{6}$$

$$Spektrum(k) = X(k) * conj(X(k)) \tag{7}$$

- 4) *Mel-Frequency Wrapping*: pada tahap ini, spektrum isyarat ditapis menggunakan M buah tapis segitiga dengan tinggi satu. Pada penelitian ini digunakan 32 buah tapis segitiga. Tapis ini dibuat dengan mengikuti persepsi telinga manusia dalam

menerima suara. Persepsi ini dinyatakan dalam skala 'mel' yang mempunyai hubungan tidak linear dengan frekuensi suara. Langkah-langkah untuk pembuatan tapis segitiga adalah sebagai berikut :

- a) Menentukan frekuensi tertinggi isyarat suara yaitu setengah dari frekuensi pencuplikan.
- b) Menghitung frekuensi tertinggi isyarat suara dalam frekuensi mel menggunakan persamaan:

$$f_{mel} = 2595 * \log(1 + \frac{f}{700}) \tag{8}$$

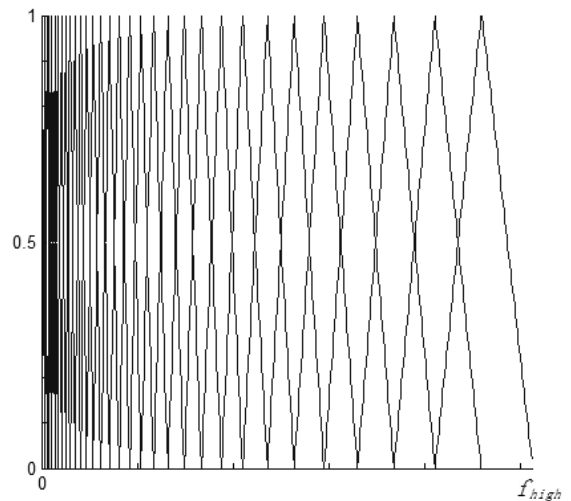
- c) Menghitung titik pusat dari tapis segitiga dengan menggunakan persamaan

$$pusat = \frac{f_{mel}^{high}}{jumlah_filter + 1} \tag{9}$$

- d) Merubah setiap frekuensi pusat tapis segitiga dalam skala mel menjadi frekuensi dengan menggunakan persamaan

$$f_i = 700 * (10^{f_{mel}/2959} - 1) \tag{10}$$

- e) Membuat tapis segitiga dengan pusat f_i dengan skema yang saling tumpang tindih seperti yang diperlihatkan pada Gambar 4.



Gambar 4. Tapis segitiga

- f) *Menghitung koefisien spektrum mel*: Nilai koefisien spektrum mel (mel spectrum coefficients) adalah jumlah dari hasil pennapisan sesuai persamaan :

$$Y_i = \log \left(\sum_{k=0}^{N-1} |Spektrum(k)| * H_i(k) \right) \tag{11}$$

Dimana:

- $i=1,2,3, \dots ,M$; dengan M adalah jumlah tapis segitiga; N adalah jumlah koefisien spektrum isyarat dalam suatu frame;

X adalah spektrum isyarat suara dan H_i adalah adalah tapis segitiga ke-i.

5) *Discrete Cosinus Transform*: Pada tahap ini dilakukan konversi dari koefisien spektrum mel kembali ke domain waktu menggunakan transformasi kosinus berikut:

$$C_j = \sum_{i=1}^M Y_i * \cos\left(\frac{j(i-1)}{2 * \pi / M}\right) \quad (12)$$

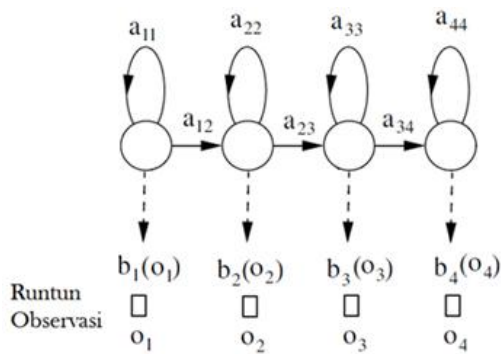
Dengan :

$j=1,2,3, \dots, k$; dimana k merupakan jumlah koefisien MFCC yang diinginkan;

M adalah banyaknya tapis segitiga; Y_i adalah koefisien spektrum mel. Dalam hal ini C_j disebut dengan koefisien ke j dari Mel Frequency Cepstrum Coefficients (MFCC).

D. Membangun Model HMM

Tipe HMM yang digunakan pada penelitian ini adalah HMM tipe kiri ke kanan dengan jumlah state yang diuji adalah 2 sampai dengan 5 buah state. Gambar 5 menunjukkan model HMM dari sistem yang dibangun dengan 4 buah state.



Gambar 5. Model HMM tipe kiri-kanan dengan 4 state

Parameter-parameter yang digunakan untuk membangun sebuah model HMM adalah sebagai berikut:

$$\lambda = (A, \pi, \mu) \quad (13)$$

dimana :

A merupakan matriks probabilitas transisi state.

π merupakan distribusi probabilitas state awal.

μ merupakan rata-rata dari runtun observasi. Pada penelitian ini μ dicari dengan merata-ratakan runtun observasi dari 10 buah data isyarat latih.

E. Pelatihan Model HMM

Pelatihan model HMM dilakukan dengan menggunakan algoritma Baum-Welch reestimation [10] untuk mendapatkan nilai probabilitas transisi state yang optimum. Proses pelatihan ini menggunakan 10 buah data isyarat latih.

F. Algoritma viterbi

Langkah-langkah dalam algoritma viterbi adalah sebagai berikut

- Inisialisasi

$$\delta_1(i) = \pi_1 b_1(o_1), 1 \leq i \leq N \quad (14)$$

$$\psi_1(i) = 0 \quad (15)$$

- Rekursi

$$\delta_t(i) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) * a_{ij}] * b_j(o_t) \quad (16)$$

$$\psi_t(i) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) * a_{ij}] \quad (17)$$

Untuk:

$$2 \leq t \leq T$$

$$1 \leq j \leq N$$

- Terminasi

$$P(O | \lambda) = \max_{1 \leq i \leq N} [\delta_T(i)] \quad (18)$$

$$q_t^* = \arg \max_{1 \leq i \leq N} [\delta_T(i)] \quad (19)$$

- Penjejukan balik lintasan

$$q_t^* = \psi_{t+1}(q_{t+1}^*) \quad (20)$$

Untuk :

$$t = T - 1, T - 2, \dots, 1$$

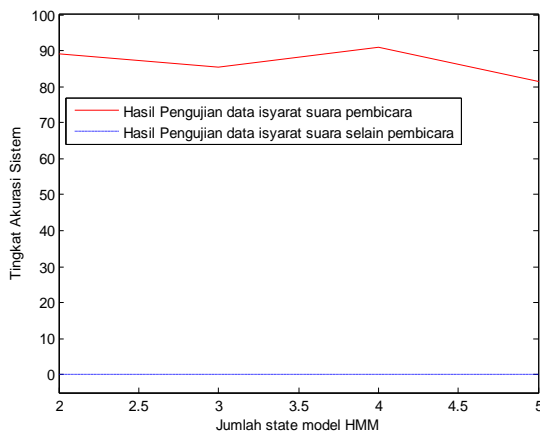
III. HASIL DAN PEMBAHASAN

Gambar 6 memperlihatkan grafik hasil pengujian sistem pengenalan pembicara HMM sebelum ditambahkan sistem verifikasi pembicara. Dapat dilihat bahwa akurasi tertinggi (91%) terdapat pada model HMM dengan jumlah state 4 untuk pengujian dengan data isyarat suara pembicara. Sedangkan untuk pengujian dengan data isyarat suara selain suara pembicara, maka akurasi sistem adalah 0% karena setiap data dikenali sebagai salah seorang dari pembicara yang terdapat dalam database sistem.

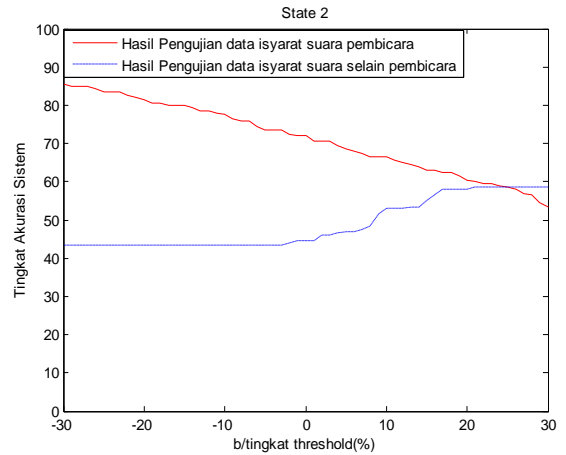
Gambar 7 sampai dengan Gambar 10 menunjukkan grafik hasil pengujian sistem pengenalan pembicara HMM dengan Jumlah state 2 sampai dengan 5 state yang telah digabungkan dengan sistem verifikasi pembicara. Dari gambar grafik dapat terlihat bahwa setelah ditambahkan dengan sistem verifikasi pembicara, akurasi sistem pengenalan pembicara HMM semakin menurun. Hal ini disebabkan karena ada sebagian suara pembicara yang telah berhasil dikenali oleh sistem pengenalan pembicara HMM secara tepat namun tertolak oleh sistem verifikasi pembicara karena tidak dapat melewati nilai *threshold*.

Dari Gambar 7 sampai dengan Gambar 10, untuk pengujian dengan data isyarat suara pembicara dapat dilihat bahwa semakin tinggi nilai *threshold* maka akurasi sistem semakin berkurang. Hal ini disebabkan karena semakin banyak data isyarat pembicara yang tertolak oleh sistem karea tidak bisa melewati nilai *threshold*.

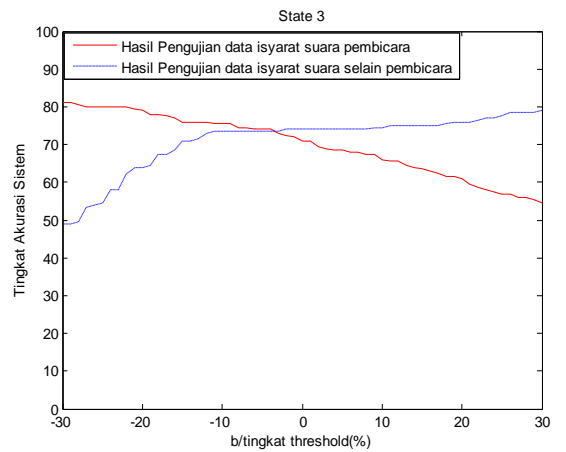
Dari Gambar 7 sampai dengan Gambar 10, untuk pengujian dengan data isyarat suara selain pembicara dapat dilihat bahwa semakin tinggi nilai *threshold* maka akurasi sistem semakin bertambah. Hal ini disebabkan karena semakin banyak data isyarat suara selain pembicara yang tertolak oleh sistem karena tidak bisa melewati nilai *threshold*.



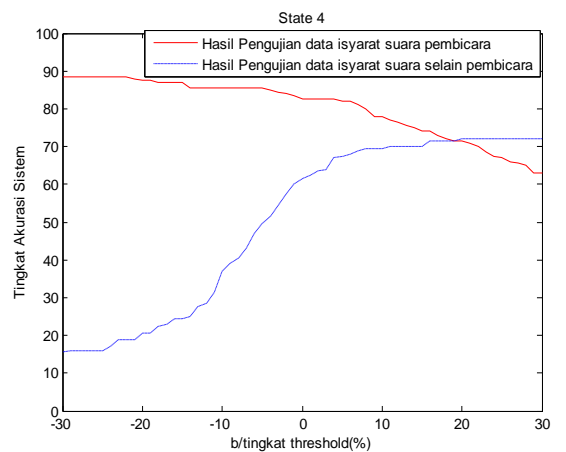
Gambar 6. Hasil pengujian Sistem Pengenalan pembicara HMM tanpa sistem verifikasi



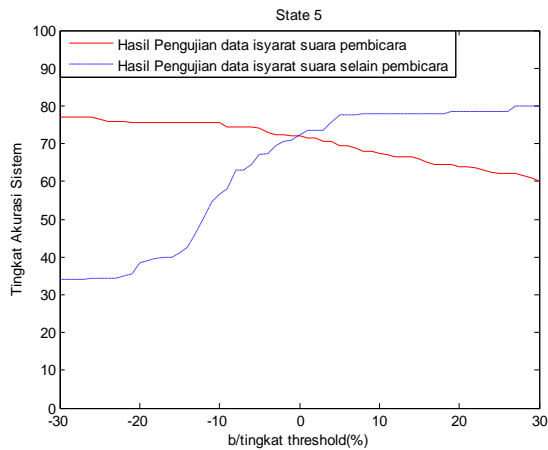
Gambar 7. Hasil pengujian Sistem Pengenalan pembicara HMM dengan sistem verifikasi dengan jumlah state 2



Gambar 8. Hasil pengujian Sistem Pengenalan pembicara HMM dengan sistem verifikasi dengan jumlah state 3



Gambar 9. Hasil pengujian Sistem Pengenalan pembicara HMM dengan sistem verifikasi dengan jumlah state 4



Gambar 10. Hasil pengujian Sistem Pengenalan pembicara HMM dengan sistem verifikasi dengan jumlah state 5

Nilai *threshold* yang paling baik digunakan untuk sistem terdapat pada titik potong kurva hasil pengujian data isyarat suara pembicara dengan kurva hasil pengujian data isyarat suara selain pembicara. Karena seimbang antara akurasi pengujian sistem dengan data isyarat suara pembicara dan akurasi pengujian sistem dengan data isyarat suaraselain pembicara.

Pada Gambar 7, untuk model HMM dengan jumlah state 2, titik potong kedua kurva terdapat pada tingkat *threshold* 25%, dengan rata-rata akurasi 58%.

Pada Gambar 8, untuk model HMM dengan jumlah state 3, titik potong kedua kurva terdapat pada tingkat *threshold* -4%, dengan rata-rata akurasi 73.75%.

Pada Gambar 9, untuk model HMM dengan jumlah state 4, titik potong kedua kurva terdapat pada tingkat *threshold* 19%, dengan rata-rata akurasi 71,75%.

Pada Gambar 10, untuk model HMM dengan jumlah state 5, titik potong kedua kurva terdapat pada tingkat *threshold* 0%, dengan rata-rata akurasi 72,25%.

KESIMPULAN

Dari penelitian yang telah dilakukan, dapat diambil beberapa kesimpulan yaitu :

A. Untuk pengujian sistem pengenalan pembicara sebelum ditambahkan sistem verifikasi pembicara

- Untuk pengujian dengan data isyarat suara pembicara didapatkan hasil akurasi tertinggi 91% pada model HMM dengan jumlah state 4.
- Untuk pengujian dengan data isyarat suara selain suara pembicara, maka akurasi sistem adalah 0% karena setiap data dikenali sebagai salah seorang dari pembicara yang terdapat dalam database sistem.

B. Untuk pengujian sistem pengenalan pembicara setelah ditambahkan sistem verifikasi pembicara

- Dari gambar grafik dapat terlihat bahwa setelah ditambahkan dengan sistem verifikasi pembicara, akurasi sistem pengenalan pembicara HMM semakin menurun. Hal ini disebabkan karena ada sebagian suara pembicara yang telah berhasil dikenali oleh sistem pengenalan pembicara HMM secara tepat namun tertolak oleh sistem verifikasi pembicara karena tidak dapat melewati nilai *threshold*.
- Dari hasil pengujian yang telah dilakukan, sistem pengenalan yang terbaik terdapat pada model HMM dengan jumlah state 4 dan dengan nilai %*threshold* -4%.

DAFTAR PUSTAKA

- [1] D. Handayana, "Perbandingan Metode Kuantisasi Vektor Dan Model Markov Tersembunyi Pada Pengenalan Pembicara Berbahasa Indonesia" Tesis, ITB, Juni 2016
- [2] D. Handaya, H. Fakhruroja, E. M. I. Hidayat, and C. Machbub, "Comparison of Indonesian Speaker Recognition Using Vector Quantization and Hidden Markov Model for Unclear Pronunciation Problem," Proceeding of IEEE 6th International Conference on System Engineering and Technology (ICSET), pp. 39-45, october 2016.
- [3] S. J. Abdallah, I. M. Osman, and M. E. Mustafa, "Text-Independent Speaker Identification Using Hidden Markov Model," World Comput. Sci. Inf. Technol. J., vol. 2, no. 6, pp. 203-208, 2012.
- [4] A. Buono, B. Kusumoputro, "Pengembangan Model HMM Berbasis Maksimum Lokal Menggunakan Jarak Euclid Untuk Sistem Identifikasi Pembicara," Proceedings of National Conference On Computer Science&Information Technology. 2007, pp.49-54.
- [5] B. Darmawan, R. hidayat, and A. Susanto, "Sistem Pengenalan Pembicara Menggunakan *Hidden Markov Models* dan MFCC," Proceeding of CITEE Sesi Nasional Indonesia, pp. 201-205, Juli 2011.
- [6] M. T. A. Zaen, A. S. Pardiyansyah, B. Darmawan, "Sistem Verifikasi Pembicara Menggunakan MFCC dan *Hidden Markov Models*" Proceeding of Seminar Nasional APTIKOM 2016, pp.14-20, Oktober 2016.
- [7] A. Charisma, "Sistem Verifikasi Penutur Menggunakan Metoda Mel Frequency Cepstral Coefficients-Vector Quantisation (Mfcc-Vq) Serta Sum Square Error (Sse) Dan Pengenalan Kata Menggunakan Metoda Logika Fuzzy" Jurnal Teknik Elektro ITP, Volume 2 No. 2, pp.71-75, Juli 2013
- [8] I.W.A. Resmawan "Verifikasi Suara Menggunakan Metode MFCC Dan DTW" Skripsi Jurusan Teknik Elektro, Fakultas Teknik Universitas Udayana, 2010
- [9] J. H. L. Hansen and T. Hasan, "Speaker Recognition by Machines and Humans," IEEE SIGNAL PROCESSING MAGAZINE, no. november, pp. 74-99, 2015.
- [10] L.R. Rabiner (1989), "A Tutorial on Hidden Markov Models and Selected applications in Speech Recognition," Proc. IEEE. vol. 77, 1989, pp.257-286.