Identifikasi Pembicara dengan Menggunakan Mel Frequency Cepstral Coefficient (MFCC) dan Self Organizing Map (SOM)

Inggih Permana[1], Benny Sukma Negara[2]
[1]Universitas Sultan Sarif Kasim Riau
[2]Universitas Sultan Sarif Kasim Riau
inggihjava@gmail.com[1], benny_sukma_negara@yahoo.co.id[2]

Abstrak

Studi ini telah menguji ketepatan identifikasi pembicara dengan menggunakan MFCC (Mel Frequency Cepstral Coefficient) dan SOM (Self Organizing Map). MFCC digunakan sebagai ekstrasi ciri sinyal suara sedangkan SOM digunakan untuk mengkluster vektor-vektor ciri yang didapat dari proses MFCC. Vektor-vektor bobot yang didapat dari pengklusteran akan digunakan sebagai codebook. Suara pengujian diukur kedekatannya dari codebook dengan menggunakan Euclidean Distance. Nilai terkecil dari hasil pengukuran tersebut merupakan vektor pemenang yang mewakili orang tertentu. Hasil pengujian memperlihatkan SOM bisa digunakan untuk identifikasi pembicara tetapi dengan rata-rata akurasi tertinggi untuk pembicara perempuan hanya 54.4% , pembicara laki-laki 75.6% dan untuk semua pembicara 62.5%.

Kata Kunci: *Codebook, Euclidean Distance, Identifikasi Pembicara, MFCC, SOM*

Abstract

This study tested the accuracy of speaker identification by using the MFCC (Mel Frequency Cepstral Coefficient) and SOM (Self Organizing Map). MFCC is used as feature extraction of voice signal. SOM is used to clustering the feature vectors that obtained from the MFCC. Weight vectors that obtained from the clustering will be used as a codebook. Voice testing is measured the proximity from the codebook by using the Euclidean Distance. The smallest value of the measurement results is the winner vector that representing a particular person. Test results show the SOM can be used for the identification of the speaker but with the highest average accuracy for female speaker only 54.4%, male speaker 75.6% and for all speaker 62.5%.

Keywords: *Codebook, Euclidean Distance, MFCC, SOM, Speaker Identification*

1. Pendahuluan

Identifikasi pembicara merupakan cabang dari pengenalan pembicara di bidang pengolahan suara. Identifikasi pembicara adalah menentukan pembicara dari data-data suara yang telah terdaftar sebelumnya [1]. Penelitian ini akan membahas identifikasi pembicara pada independent text dan data close-set.

Topik pengenalan pembicara ini diangkat mengingat keterbatasan manusia dalam mengenali suara manusia yang begitu banyak ragamnya serta banyak suara yang hampir sama antara manusia satu dengan manusia lainnya. Dalam kehidupan sehari-hari pengenalan pembicara juga sangat penting, contohnya dalam mengidentifikasi siapa yang berbicara pada rekaman percakapan yang dijadikan bukti pada suatu kasus di persidangan atau sebagai identitas dan akses kontrol untuk telephon banking, shopping banking, membuka komputer pribadi dan lain sebagainya.

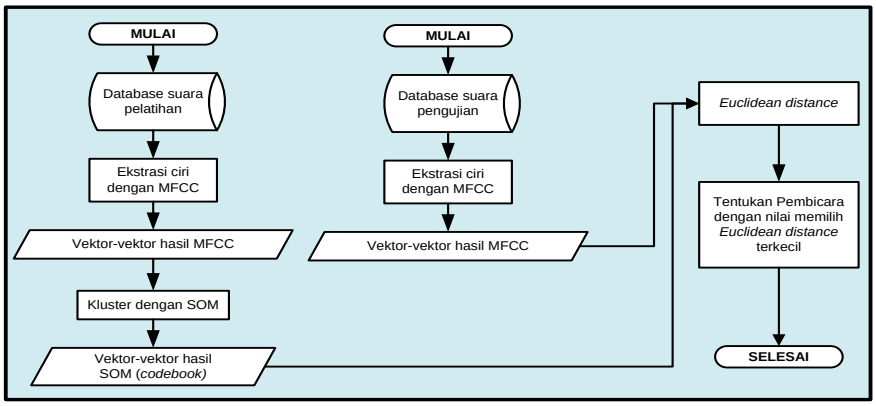
Penelitian ini akan menguji ketepatan identifikasi pembicara dengan menggunakan kombinasi metode MFCC (Mel Frequency Cepstral Coefficient) untuk ekstraksi ciri dan SOM (Self Organizing Map) untuk menghasilkan codebook. Codebook adalah representasi numerik dari ciri pada orang tertentu [2]. MFCC merupakan cara yang paling sering digunakan, karena cara kerjanya didasarkan pada perbedaan frekuensi yang dapat ditangkap oleh telinga manusia [3]. SOM merupakan salah satu tipe dari jaringan saraf tiruan yang mana sistem pembelajarannya menggunakan unsupervised learning. SOM merupakan suatu lapisan yang berisi neuron-neuron akan menyusun dirinya sendiri

berdasarkan input nilai tertentu dalam suatu kelompok yang dikenal dengan istilah cluster [4]. Vektor-vektor bobot dari hasil proses SOM tersebut akan dijadikan codebook.

2. Metode Penelitian

Penelitian ini akan menggunakan 10 orang pembicara yang terdiri dari 5 orang laki-laki dan 5 orang perempuan. Masing-masing orang akan diambil sample suaranya selama 20 detik sebanyak 2 kali. Sample suara pertama akan dijadikan data pelatihan sedangkan sample suara kedua akan dijadikan data pengujian. Kata-kata yang diucapkan pembicara bersifat bebas (independent textI) dan jumlah pembicara bersifat tetap (close-set).

Proses identifikasi pembicara pada penelitian ini secara garis besar terdiri dari tiga bagian, yaitu ekstrasi ciri, kluster dan pencocokan ciri. Pada ekstrasi ciri, data suara pelatihan masing-masing pembicara dilakukan ekstrasi ciri dengan metode MFCC. Vektor-vektor ciri yang didapat dijadikan sebagai vektor input pada proses pengklusteran dengan menggunakan SOM. Pada proses tersebut akan dihasilkan vektor-vektor bobot yang pada penelitian ini akan digunakan sebagai codebook. Pada proses pencocokan ciri, data suara pengujian pada masing-masing pembicara juga dilakukan proses MFCC. Setalah itu vektor-vektor hasil proses MFCC dari suara pengujian dihitung kedekatan jaraknya dengan codebook menggunakan Euclidean Distance. Jarak terkecil dari perhitungan tersebut merupakan vektor pemenang yang mewakili suara orang tertentu.

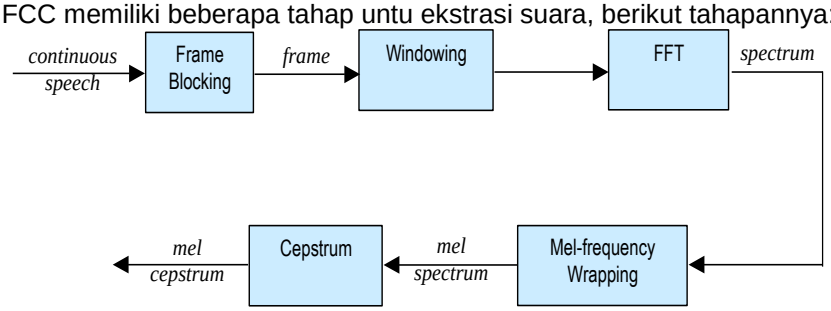


Gambar 1. Alur identifikasi pembicara

2.1. Ekstrasi Ciri (MFCC)

Bagian yang penting dari pengenalan suara adalah ektrasi ciri (feature extraction). Kegunaan dari ekstrasi ciri adalah mengurangi informasi-informasi yang tidak dibutuhkan dari sensor dan mengkonversi informasi-informasi yang penting dari signal untuk pengenalan pola agar menghasilkan format yang lebih sederhana dengan kelas-kelas yang jelas [5]. MFCC merupakan cara yang paling sering digunakan, karena cara kerjanya didasarkan pada perbedaan frekuensi yang dapat ditangkap oleh telinga manusia [3].

MFCC memiliki beberapa tahap untu ekstrasi suara, berikut tahapannya:



Gambar 2. Blok diagram tahapan MFCC [5]

Langkah ke 1 : Frame Blocking

Frame blocking adalah membagi sinyal yang masuk kedalam beberapa frame. Pada tahap ini kita melakukan overlopping. Biasanya overlopping dimulai dari ukuran frame dibagi dengan dua yang disebut dengan istilah hopsize [5]. Untuk lebih jelasnya perhatikan gambar 3. pada **Langkah ke 2**.

Langkah ke 2 : Windowing

Langkah selanjutnya adalah menghaluskan masing-masing frame untuk meminimalkan sinyal yang tidak kontinu pada awal dan akhir masing-masing frame. Hal ini dilakukan dengan Hamming Window, berikut rumusnya [5].

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

$$y_l(n) = x_l(n)w(n), \quad 0 \le n \le N-1 \tag{2}$$

Hamm

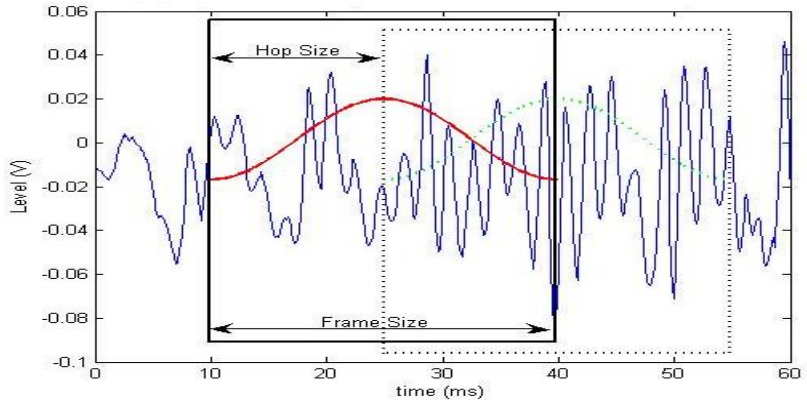
Dimana :

w(n) = Hamming window

$y_l(n)$ = Output signal

$x_l(n)$ = Input signal

N = jumlah sample pada masing-masing frame.



Gambar 3. Frame blocking [5]

Gambar 3. memperlihatkan contoh frame blocking. Sumbu horizontal Menerangkan suara yang masuk berdasarkan satuan waktu (milli secon). Sedangkan sumbu vertikal menunjukan level (volt) dari suara yang masuk. Pada gambar tersebut terlihat frame size adalah 30 ms dan hop size nya adalah 15 ms. Pada gambar tersebut juga terlihat garis melengkung yang disebut Hamming Window.

Langkah ke 3 : Fast Fourier Transform

Langkah selanjutnya adalah Fast Fourier Transform (FFT). Hal ini dilakukan untuk mengubah frame dari domain waktu ke domain frekuensi. FFT adalah implementasi dari Discrete Fourier Transform (DFT), berikut rumusnya.[6].

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j2\pi kn/N}, \qquad k = 0,1,2,\ldots,N-1 \tag{3}$$

Dimana :

Xk = Output signal

N = jumlah sample pada masing-masing frame.

Langkah 4 : Mel Frequency Wrapping

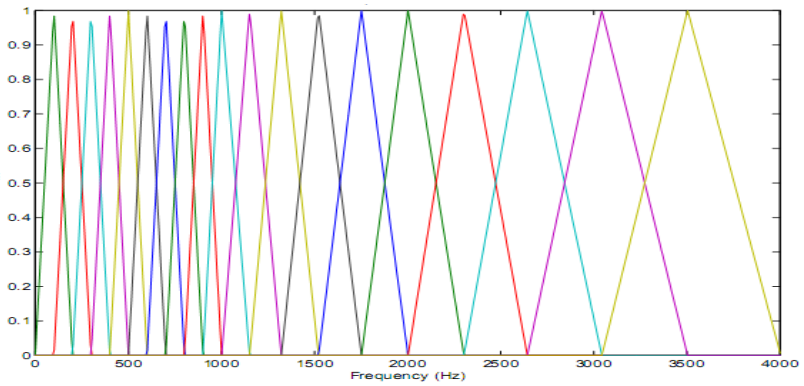
Untuk menggunakan metode ini frekuensi dijadikan dalam skala Mel, berikut rumus-rumusnya [7].

$$Mel(f) = 2595 \log 10(1 + \frac{f}{700}) \quad (4)$$

$$Freq(m) = 700(10^{\frac{m}{2595}} - 1) \quad (5)$$

$$\tilde{S}(l) = \sum_{k=0}^{N/2} S(k) M_l(k) \quad (6)$$

Persamaan (4) dan (5) adalah untuk mendapatkan mel filter bank. Setelah itu, sinyal suara hasil dari **Langkah 3** di proses dengan persamaan (6). Mel filter bank diperlihatkan pada gambar 4.



Gambar 4. Mel filter bank [6]

Terlihat pada gambar 4. mel filter bank merupakan segitiga tumpang tindih. Pada gambar tersebut sumbu horizontal adalah mewakili frekuensi dan sumbu vertikal mewakili db sinyal suara.

Langkah 5 : Cepstrum

Langkah terkahir dari proses ini adalah mengembalikan sinyal suara dari domain frekuensi ke domain waktu. Hasil dari langkah ini disebut MFCC. Berikut rumus untuk mencarinya [6].

$$\tilde{c}_n = \sum_{k=1}^{K} (\log \tilde{S}_k) \cos\left[n\left(k - \frac{1}{2} \right) \frac{\pi}{K} \right], \qquad n = 0,1,\ldots, K\text{-}1 \quad (7)$$

2.3. Euclidean Distance

Euclidean distance, yaitu mengukur jumlah kuadrat perbedaan nilai masing-masing variabel [8].

$$d = \sqrt{\sum_{i}^{n} (Wi - Xi)^2} \quad (8)$$

Dimana:

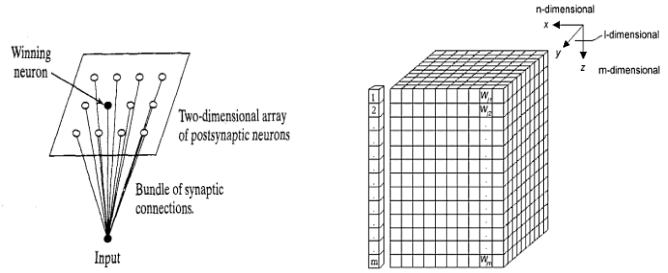
d = jarak Euclidian
Wi= vektor bobot ke-i
Xi= vektor input ke Xi

Semakin kecil nilai d, semakin besar kesamaan antara kedua obyek atau kasus tersebut, dan sebaliknya, semakin besar nilai d, semakin kecil kesamaan antara dua obyek.

2.2. SOM

Self Organizing Maps (SOM), atau Kohonen, merupakan salah satu tipe dari artifcial neural network yang mana sistem pembelajarannya menggunakan unsupervised learning untuk menghasilkan dimensi rendah (pada umumnya dua dimensi). Unsupervised learning adalah input disini tidak terdefinisikan. model ini tidak membutuhkan inputan yang selalu ada, sehingga tidak mempengaruhi output. Jadi jika input dari model ini hilang, tidak akan mempengaruhi variabel karena tidak memiliki target.

Jaringan SOM Kohonen terdiri dari dua lapisan (layer), yaitu lapisan input dan lapisan output. Setiap neuron dalam lapisan input terhubung dengan setiap neuron pada lapisan output. Setiap neuron dalam lapisan output merepresentasikan kelas dari input yang diberikan [10].



Gambar 5. Kohonen Model [10]

Setiap neuron output mempunyai bobot untuk masing-masing neuron input. Proses pembelajaran dilakukan dengan melakukan penyesuaian terhadap setiap bobot pada neuron output. Setiap input yang diberikan dihitung jarak Euclidian-nya dengan setiap neuron output, kemudian cari neuron output yang mempunyai jarak minimum. Neuron yang mempunyai jarak yang paling kecil disebut neuron pemenang atau neuron yang paling sesuai dengan input yang diberikan.

Algoritma pengelompokan pola jaringan SOM adalah sebagai berikut [11] :

1. Inisialisasi
   a. Bobot w_{ji} (acak)
   b. Laju pemahaman awal dan faktor penurunannya
   c. Bentuk dan jari-jari topologi sekitarnya
2. Selama kondisi penghentian bernilai salah, lakukan langkah 2-7
3. Untuk setiap vektor masukan x, lakukan langkah 3-5
4. Hitung d dengan Euclidean Distance, perhitungan dilakukan untuk semuan j
5. Tenntukan indeks j sedemikian hingga d(j) minimum
6. Untuk setiap unit j disekitar J modifikasi bobot :
   $w_{ji}^{baru} = w_{ji}^{lama} + \alpha(x_i - w_{ji}^{lama})$
   (9)
7. Modifikasi laju pemahaman
8. Uji kondisi penghentian

Apabila semua w_{ij} hanya berubah sedikit saja, berarti iterasi sudah mencapai konvergensi sehingga dapat dihentikan.

3. Hasil dan Analisa

Pengujian dilakukan dengan merubah nilai beberapa parameter. Hal ini untuk menunjukan pengaruh perubahan parameter-parameter tersebut terhadap ketepatan pendeteksian pembicara. Parameter-parameter yang akan diuji adalah pengaruh perubahan iterasi pada SOM, perubahan koefisien MFCC, serta kombinasi perubahan koefisien MFCC dan iterasi pada SOM yang digunakan. Penelitian ini menggunakan 512 byte per frame, 60 codebook dan 80 filter bank.

1. Pengaruh Perubahan Iterasi SOM

Tabel 1. menunjukkan pengaruh perubahan iterasi SOM terhadap akurasi identifikasi pembicara. Pengujian pada tabel tersebut menggunakan 30 koefisien MFCC.

Tabel 1. Pengaruh Perubahan Iterasi SOM

Jumlah Iterasi	Jumlah Terdeteksi	Rata-Rata Akurasi Identifikasi Perempuan	Rata-Rata Akurasi Identifikasi Laki-Laki	Rata-Rata Akurasi Identifikasi Semua Pembicara
1000	10 Orang	49.60%	74.20%	61.90%
2000	10 Orang	48.80%	73%	60.90%
3000	10 Orang	50.40%	74.60%	62.50%
4000	10 Orang	50%	72%	61%
5000	10 Orang	47.60%	73.60%	60.60%
6000	10 Orang	47.60%	73.60%	60.60%
7000	10 Orang	48.80%	73%	60.90%
8000	10 Orang	49.20%	72.40%	60.80%
9000	10 Orang	48.40%	72.40%	60.40%
10000	10 Orang	47.80%	72%	59.90%

Rata-rata akurasi terendah Rata-rata akurasi tertinggi

Tabel 1. memperlihatkan bahwa pada semua iterasi berhasil mendeteksi semua pembicara tetapi dengan akurasi yang berbeda. Dari tabel tersebut terlihat bahwa kenaikan iterasi tidak selalu meningkatkan akurasi dalam identifikasi pembicara. Tabel tersebut juga memperlihatkan bahwa rata-rata identifikasi pembicara perempuan lebih rendah dibandingkan pembicara laki-laki. Rata-rata akurasi tertinggi pembicara perempuan 50.4 % (iterasi ke 3000), pembicara laki-laki 74.6% (iterasi ke 3000), semua pembicara 62.5% (iterasi ke 3000).

2. Pengaruh Perubahan Koefisien MFCC

Tabel 2. menunjukkan pengaruh perubahan koefisien MFCC terhadap akurasi identifikasi pembicara. Pengujian pada tabel tersebut menggunakan 1000 iterasi.

Tabel 2. Pengaruh Perubahan Koefisien MFCC

Koefisien MFCC	Jumlah Terdeteksi	Rata-Rata Akurasi Identifikasi Perempuan	Rata-Rata Akurasi Identifikasi Laki-Laki	Rata-Rata Akurasi Identifikasi Semua Pembicara
10	10 Orang	43.60%	57.80%	50.70%
20	10 Orang	52.60%	66.60%	59.60%
30	10 Orang	49.60%	74.20%	61.90%
40	10 Orang	47.20%	71.80%	59.50%
50	10 Orang	47.20%	72.20%	59.70%
60	10 Orang	46.60%	71.60%	59.10%
70	10 Orang	45.00%	71.80%	58.40%

Rata-rata akurasi terendah Rata-rata akurasi tertinggi

Tabel 2. memperlihatkan bahwa pada semua koefisien MFCC berhasil mendeteksi semua pembicara tetapi dengan akurasi yang berbeda. Dari tabel tersebut terlihat bahwa kenaikan koefisien MFCC pada rentang tertentu dapat meningkatkan tingkat akurasi. Pada pembicara perempuan rata-rata akurasi meningkat dalam rentang koefisien MFCC 10 sampai 20. Sedangkan pada pembicara laki-laki rata-rata akurasi meningkat dalam rentang koefisien MFCC 10 sampai 30. Pada semua pembicara rata-rata akurasi meningkat dalam rentang koefisien MFCC 10 sampai 30. Tabel tersebut juga memperlihatkan bahwa rata-rata identifikasi pembicara perempuan lebih rendah dibandingkan pembicara laki-laki. Rata-rata akurasi tertinggi pembicara perempuan 52.6 % (20 koefisien MFCC), pembicara laki-laki 74.2% (30 koefisien MFCC), semua pembicara 61.9% (30 koefisien MFCC).

3. Pengaruh Kombinasi Perubahan Iterasi SOM dan Koefisien MFCC

Tabel 3. menunjukkan pengaruh perubahan iterasi SOM dan koefisien MFCC secara bersamaan terhadap akurasi identifikasi pembicara.

Tabel 3. Pengaruh kombinasi Perubahan iterasi SOM dan Koefisien MFCC

Koefisien MFCC	Jumlah Iterasi	Jumlah Terdeteksi	Rata-Rata Akurasi Identifikasi Perempuan	Rata-Rata Akurasi Identifikasi Laki-Laki	Rata-Rata Akurasi Identifikasi Semua Pembicara
10	1000	10 Orang	43.60%	57.80%	50.70%
20	2000	10 Orang	53.60%	67.60%	60.60%
30	3000	10 Orang	50.40%	74.60%	62.50%
40	4000	10 Orang	48.60%	72.60%	60.60%
50	5000	10 Orang	48.60%	71.80%	60.20%
60	6000	10 Orang	54.40%	70.60%	62.50%
70	7000	10 Orang	47.00%	70.40%	58.70%

Rata-rata akurasi terendah Rata-rata akurasi tertinggi

Tabel 3. memperlihatkan bahwa pada semua iterasi SOM dan koefisien MFCC berhasil mendeteksi semua pembicara tetapi dengan akurasi yang berbeda. Dari tabel tersebut terlihat bahwa kenaikan keduanya tidak selalu meningkatkan akurasi dalam identifikasi pembicara. Seperti pengujian sebelumnya, tabel tersebut juga memperlihatkan bahwa rata-rata identifikasi pembicara perempuan lebih rendah dibandingkan pembicara laki-laki. Rata-rata akurasi tertinggi pembicara perempuan 54.4 % (iterasi ke 6000, 60 koefisien MFCC) dan pembicara laki-laki 74.6% (iterasi ke 3000, 30 koefisien MFCC). Sedangkan khusus rata-rata akurasi untuk semua pembicara adalah 62.5% yang terdapat pada dua buah pengujian yaitu iterasi ke 3000 dengan 30 koefisien MFCC dan iterasi ke 6000 dengan 60 koefisien MFCC. Dalam hal ini tentu saja yang terbaik adalah yang memiliki iterasi dan koefisien MFCC lebih rendah karena waktu proses yang dibutuhkan juga akan lebih sedikit.

4. Kesimpulan

Penelitian ini menunjukkan bahwa SOM konvensional bisa digunakan dalam identifikasi pembicara tetapi dengan rata-rata akurasi tertinggi untuk pembicara perempuan hanya 54.4% (iterasi ke 6000, 60 koefisien MFCC), pembicara laki-laki 75.6% (iterasi ke 3000, 30 koefisien MFCC) dan untuk semua pembicara 62.5% (iterasi ke 3000, 30 koefisien MFCC). Penelitian ini juga menunjukkan bahwa jumlah iterasi pada SOM tidak selalu memberikan akurasi yang lebih baik. Sedangkan peningkatan jumlah koefisien MFCC dapat memberikan peningkatan akurasi dalam rentang tertentu yaitu 10 sampai 30. Pada penelitian juga didapat bahwa rata-rata akurasi identifikasi pembicara perempuan selalu lebih rendah dari pada laki-laki. Oleh sebab itu, pada studi selanjutnya perlu diteliti mengapa pembicara perempuan lebih sulit diidentifikasi daripada pembicara laki-laki.

Referensi

[1] Alex SP. ASR Dependent Techniques for Speaker Recognition. MSc Thesis. Massachusetts: Department of Electrical Engineering and Computer Science of MIT. 2002.
[2] Thang Wee Keong. Voice Print Analysis for Speaker Recognition. BSc Thesis. Singapore: SIM University. 2009.
[3] Lindasalwa M, Mumtaj B, I Elamvazuthi. Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques. Journal of Computing. 2010; 2(3): 138-142.
[4] Sri K. Artificial Intelligence (Teknik dan Aplikasinya). Yogyakarta: Graha Ilmu. 2003.
[5] Abeer MAH. Text Independent Speaker Identification System. BSc Thesis. Nablus: Electrical Engineering Department An-Najah National University. 2010.
[6] Magnus N. Speaker Verification in Java. MSc Thesis. School of Microelectronic Engineering of Griffith University. 2001.
[7] Mikael L, Jens KR. JOPSPEECH Embedded Java Speech Recognition SDK. Denmark: Copenhagen Business School. 2006
[8] Gregorius SB, Liliana, Steven H. Cluster Analysis untuk Memprediksi Talenta Pemain Basket Menggunakan Jaringan Saraf Tiruan Self Organizing-Map (SOM). UK Petra. Surabaya. 2006
[9] Arum M. Clustering Specimen Daun Dikotiledon Dengan Menggunakan SOM. Bogor: Institut Pertanian Bogor. 2003
[10] Simon H. Neural Network A Comprehensive Foundation. Edisi 2. Singapore: Pearson Prentice Hall. 2005: 465-501.
[11] ong JS, Jaringan Syaraf Tiruan dan Pemograman Menggunakan MATLAB. Andi: Yogyakarta. 2005:143