

BIOMETRIK POLA SUARA DENGAN JARINGAN SARAF TIRUAN**Ina Agustina¹, Fauziah², Aris Gunaryati³**

- (1) Sistem Informasi Universitas Nasional (Contact : 021-7806700, agustina2008@gmail.com)
(2) Sistem Informasi Universitas Nasional (Contact : 021-7806700, fauziahkasyfi@yahoo.co.id)
(3) Teknik Informatika Universitas Nasional (Contact : 021-7806700, arisgunaryati@yahoo.co.id)

ABSTRAK

Metode yang digunakan dalam penelitian ini adalah biometrik pengenalan identitas berdasarkan karakteristik fisik. Pengenalan pola suara biometrik yang memiliki biaya rendah. Penelitian ini merancang pola penggunaan suara biometrik. sinyal input suara dalam bentuk suara yang direkam dengan durasi sekitar dua (2) detik pada sistem dan sinyal dekomposisi menggunakan transformasi wavelet diskrit. sistem verifikasi menggunakan ekstraksi karakteristik Koefisien Sub Band berdasarkan parameter cepstral (SBC) dengan menggunakan jumlah koefisien 12.

Kata kunci: *Biometrik, pola suara, transformasi wavelet, jaringan saraf*

ABSTRACT

Method man with Biometric identity recognition, based on its physical characteristics. The introduction of a biometric voice pattern that has a low cost. This study designed a biometric use voice patterns. Voice input signals in the form of recorded sound with a duration of approximately two (2) seconds on the system and signal decomposition using discrete wavelet transform. Verification system using characteristic extraction Sub Band Coefficient Based cepstral parameters (sbc) by using the number of coefficients 12.

Keywords: Biometrics, voice patterns, wavelet transformation, neural network

1. Pendahuluan

Metode pengenalan identitas manusia dengan Biometrik, berdasarkan karakter fisiknya, misalnya wajah, sidik jari, struktur telapak tangan, letak retina mata, suara/wicara. Verifikasi ini jauh lebih handal daripada dengan metode konvensional karena selain mudah untuk Tabel 1. Metode-metode yang digunakan Penelitian sebelumnya

diproses juga selalu dibawa oleh manusia tersebut. Dan yang paling utama kelebihan dari verifikasi dengan biometrik ini adalah bervariasinya untuk tiap-tiap orang dan tidak mudah ditiru oleh orang lain seperti halnya kode numerik atau alfabetik

Metode Pemrosesan Sinyal Suara	Tingkat Identifikasi (%)
LPC (Linier Predictive Coding)	90,61
LPCC	96,73
MFCC (Mel Frequency Cepstrum Coefficient)	97,55
PLP	86,12
NPC (random initialization)	61,63
NPC (linier initialization)	100

Sumber : Universitie Pierre & MarieCurie, LA Science A Paris, 2004

Dari penelitian terdahulu mengenai pemrosesan sinyal suara, metode yang telah digunakan *Linier Prediction, Coding (LPC), Frequency Cepstrum Coefficient (MFCC), Neural Predictive Coding NPC*, yang menggunakan transformasi fourier dan tingkat identifikasinya mencapai 100%.

Kelemahan metode menggunakan transformasi fourier antara lain kurang mampu memberikan informasi sinyal dalam waktu domain dan frekuensi secara bersamaan dan dalam menganalisis sinyal suara tidak stationer. Oleh karena itu pada penelitian ini dalam pemrosesan sinyal suara tidak menggunakan transformasi fourier.

Sebelumnya telah dilakukan penelitian mengenai Sistem Identifikasi pembicara menggunakan Transformasi Wavelet Diskrit berbasis orthogonal (symlet) dan Jaringan Saraf Tiruan Multi Layer Perceptron sebagai Pengenal Pola [1].

Dengan dasar penelitian tersebut dan studi literatur di atas, pada penelitian ini akan dibuat suatu sistem biometrika dan verifikasi menggunakan dua buah masukan (Multimodal) pengenalan pola suara dan pengenalan telinga. Dari perancangan ini diharapkan adanya peningkatan performansi sistem untuk meningkatkan tingkat keamanan dan akurasi dari proses verifikasi. Karena dari beberapa penelitian yang telah ada, sistem biometrika yang menggunakan satu masukan memiliki banyak kelemahan dibandingkan dengan multimodal biometrika.

2. Metodologi

2.1 Pengolahan Sinyal Digital

Konsep dasar pengolahan sinyal digital yang digunakan untuk penelitian ini terdiri atas *sampling, frame blocking, windowing, DFT, Power Spectral Substraction, Mel Filter Bank dan Discrette Cosine Transform*.

Sampling

Suara manusia akan menghasilkan sinyal analog yang terus kontinyu. Karena itu sinyal yang ada dipotong-potong dalam slot-slot interval waktu tertentu. Deret diskrit sample $x[n]$ diperoleh dari sinyal kontinu $x(t)$ dengan hubungan sebagai berikut,

$$x[n] = x(nT) \dots\dots\dots(1)$$

Di mana T adalah periode sampling dan $1/T = F_s$ merupakan frekuensi sampling dalam satuan sampel/detik. Nilai N merupakan jumlah sampel. Berdasarkan pada teori sampling Nyquist, maka syarat dari frekuensi sampling adalah minimal dua kali frekuensi sinyal asli.

Frame Blocking

Frame Blocking merupakan pembagian suara menjadi beberapa frame dan satu frame terdiri dari beberapa sampel. Proses ini diperlukan untuk membentuk sinyal suara yang non stasioner menjadi sinyal suara yang quasi-stasioner sehingga dapat dibah dari domain waktu ke dalam domain frekuensi dengan Transformasi Fourier.

Windowing

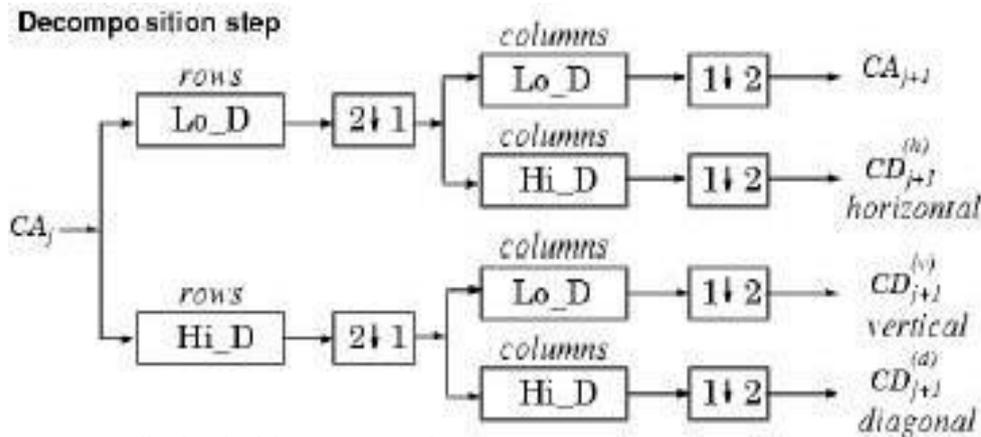
Sinyal suara yang dipotong-potong menjadi beberapa frame akan menyebabkan efek diskontinuitas pada awal dan akhir sinyal. Hal ini akan menyebabkan kesalahan data pada proses Transformasi Fourier . Windowing diperlukan

untuk mengurangi efek diskontinuitas potongan sinyal tersebut.

Transformasi Wavelet Diskrit secara umum merupakan dekomposisi citra pada frekuensi subband sinyal tersebut di mana komponennya dihasilkan dengan cara penurunan level dekomposisi. Implementasi transformasi wavelet diskrit dapat dilakukan dengan cara melewati sinyal frekuensi tinggi atau highpass filter dan

merupakan representasi dari jumlah klas dari hasil penghitungan Eigenface yang didapatkan.

Untuk Hidden layer pertama digunakan 40,80, 150 node dan hidden layer kedua sebanyak 25 node. Sedangkan untuk output layer yang akan digunakan 8 node disesuaikan dengan jumlah orang/identitas.



Gambar 1. dekomposisi wavelet frekuensi rendah atau lowpass filter.

Subband Based Cepstrum Coefficient (SBC)

Suara Manusia menyimpan informasi penting pada frekuensi tinggi. Informasi penting ini yang menentukan karakteristik suara manusia dan Mel scale digunakan untuk mengakomodasi karakteristik tersebut.

Setiap nada suara manusia dengan frekuensi aktual yang diukur dalam Hz nilai subyektif pitch dapat diukur dalam Mel Scale. Mel Scale merupakan ukuran atau skala persepsi manusia dari frekuensi suatu suara.

Pada proses SBC sebenarnya hampir sama dengan proses MFCC hanya saja bila pada MFCC sinyal mengalami proses DFT tetapi pada SBC digunakan Transformasi Wavelet Diskrit untuk menggantikan proses DFT.

2.2 Jaringan Saraf Tiruan

Dari data yang telah mengalami proses Eigenface dan subband decomposition sebelumnya masing-masing akan dilakukan pelatihan menggunakan Jaringan Saraf Tiruan (JST)..

JST yang pertama adalah untuk pengenalan telinga dengan menggunakan satu hidden layer dengan banyak node yang digunakan pada hidden layer adalah 60 node, jumlah node ini

Elemen pembobot W pada setiap *neuron* merupakan representasi dari kekuatan sambungan (*strength of synapse*) antar *neuron*.

JST yang kedua adalah untuk pengenalan pola suara dengan menggunakan Hidden Layer sebanyak 2 hidden layer dengan jumlah node pada hidden layer pertama sebanyak 40,80, 150 dan pada hidden layer kedua sebanyak 25 node dan dengan output layer sebanyak 10 disesuaikan dengan jumlah identitas.

Jaringan Saraf Tiruan merupakan salah satu representasi buatan dari otak manusia yang selalu mencoba untuk mensimulasikan proses pembelajaran pada otak manusia. Secara umum, meskipun Jaringan Saraf Tiruan banyak jenisnya akan tetapi terdapat persamaan pada komponen penyusunnya. Seperti juga otak manusia, Jaringan Saraf Tiruan juga terdiri dari neuron dan hubungan diantaranya. Neuron mengolah informasi yang masuk dan meneruskan ke neuron selanjutnya. Pada Jaringan Saraf Tiruan istilah untuk menggantikan hubungan/koneksi adalah bobot. Informasi berupa sinyal listrik disimulasikan sebagai harga yang spesifik pada bobot. Dengan cara mengubah-ubah harga bobot artinya kita juga mengubah-ubah struktur hubungan/koneksi antar neuron. Gambar dibawah ini adalah model / struktur neuron Jaringan Saraf Tiruan.

Model satu *neuron* dapat dilihat pada gambar 3. dimana *input* X_i masuk ke *neuron* (*neuron* dengan hirerarki yang lebih rendah, jika JST tersebut memiliki banyak *neuron* / multi *layer*). $F(x)$ = fungsi *output* dari jaringan

Pembobot W_{ij} pada model diatas menyatakan kekuatan hubungan antara *neuron* lapisan i ke lapisan j . Jika *neuron* pada lapisan i berjumlah n

sedangkan pada lapisan j berjumlah m , maka pembobot W_{ij} berbentuk $m \times n$. Penulisan dari subskrip pada W selalu dimulai dari lapisan dengan hierarki yang lebih tinggi baru kemudian ke hierarki yang lebih rendah. Sehingga model pembobot tadi ditulis W_{ij} .

Semua data *input* yang telah diboboti dengan kekuatan sambungannya akan dijumlahkan terlebih dahulu sebelum diolah menjadi sinyal *output* dan hasil penjumlahan ini, biasanya dikenal dengan NET.

Fungsi aktivasi *neuron* $F(x)$ adalah fungsi pengolahan dari *input* menjadi sinyal *output*. Mengaktifkan jaringan saraf tiruan berarti mengaktifkan neuron yang ada pada jaringan tersebut. Banyak fungsi yang dipakai sebagai

fungsi pengaktif, diantaranya fungsi hiperbolik, fungsi impuls, fungsi step dan fungsi sigmoid. Tetapi yang lazim digunakan adalah fungsi sigmoid karena fungsi ini mendekati kinerja sistem otak

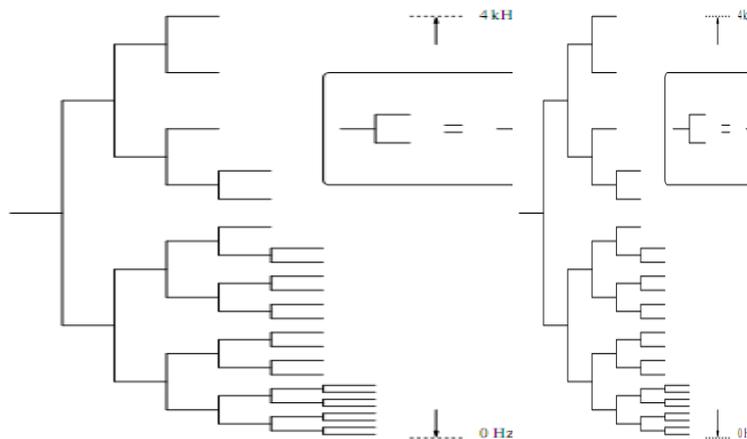
2.3. Identifikasi Pola Suara

2.3.1 Proses Identifikasi Pola Suara

1. Sampling Suara

Pada proses perekaman suara ini

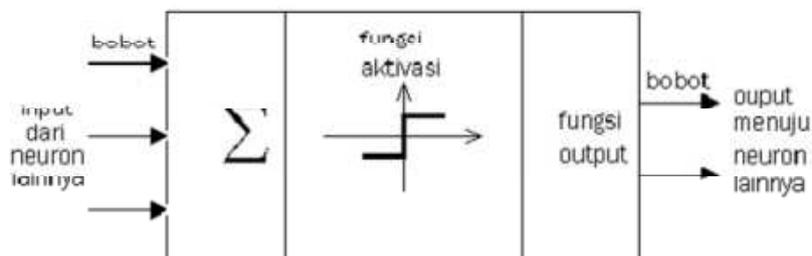
dengan aturan Nyquist.



Gambar 2. Konstruksi *filter bank*

digunakan alat tersendiri untuk merekam suara,

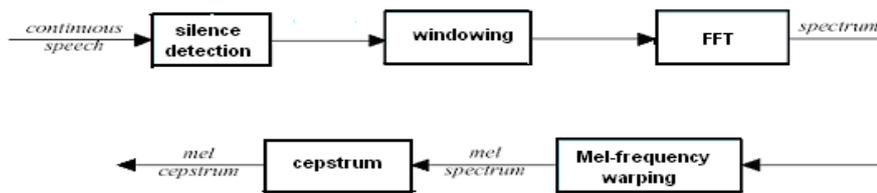
2. Proses Frame Blocking



Gambar 3. Struktur Neuron Jaringan Saraf Tiruan

yaitu menggunakan software free audio recorder. Untuk penyamplingannya dilakukan dengan frekuensi sampling sesuai

Sinyal yang telah disampling selanjutnya dibagi dalam frame dalam waktu tertentu.



Gambar 4. Proses Suara

3. Windowing Hamming

Untuk mengurangi diskontinuitas sinyal setelah proses frame blocking maka sinyal tiap-tiap frame

difilter dengan windowing.

4. Wavelet Subband Coding Based MFCC

Pada bagian ini dilakukan Subband memilah milah sinyal ke dalam 24 filter yang hampir sama dengan filterbank pada MFCC dengan melakukan dekomposisi sinyal wavelet.

3. Pembahasan

Pelatihan jaringan dilakukan dengan mengambil input dari pembicara sebanyak 8 orang (pembicara 1, pembicara 2, pembicara 3, pembicara 4 sampai pembicara 8) dimana masing-masing pembicara mengucapkan 10 buah pola kata yaitu 'nol', 'satu', 'dua' dan seterusnya sampai pola kata 'sembilan'. Pola kata dari masing-masing pembicara tersebut disimpan dan kemudian dilatihkan secara bersamaan ke dalam jaringan saraf tiruan. Pola kata tersebut dimasukkan secara urut mulai pembicara 1 dengan pola kata 'nol', 'satu', 'dua' dan seterusnya sampai pola kata 'sembilan', kemudian pembicara 2 dengan pola kata 'nol', 'satu', 'dua' sampai pola kata 'sembilan', demikian seterusnya sampai pembicara 5. Setelah semua data dimasukkan maka proses training dilakukan sampai error yang dihasilkan mencapai nilai yang telah ditentukan dimana pada proses ini digunakan nilai error 0,0001.

3.1. Penentuan Struktur Jaringan Saraf Tiruan

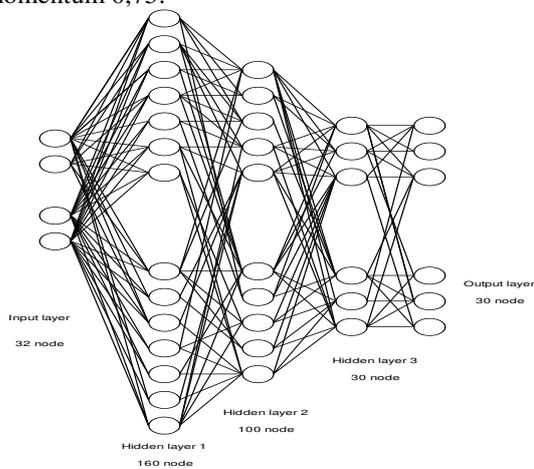
Pertama kali akan ditentukan jumlah *node hidden layer* dengan menggunakan satu *hidden layer*, nilai *learning rate* 0,5 momentum 0,5 serta input 32 data. Dari hasil pengujian didapat bahwa makin banyak jumlah *node hidden layer* yang digunakan maka akan menghasilkan error yang kecil dalam iterasi yang makin singkat, sampai mencapai suatu nilai tertentu dimana perubahan jumlah *node* hanya mengakibatkan sedikit perubahan pada jumlah iterasi. Dari pengujian tersebut maka didapat bahwa jumlah *node hidden layer* yang optimal ialah 160 buah. Penambahan jumlah *node* lebih besar dari

160 tidak menghasilkan penurunan jumlah iterasi yang berarti. Makin banyak jumlah *node* yang digunakan akan memakai memori komputer makin besar sehingga jika dipilih jumlah *node* diatas 160 akan terdapat pengorbanan pada jumlah memori yang digunakan tanpa diiringi perubahan jumlah iterasi yang berarti.

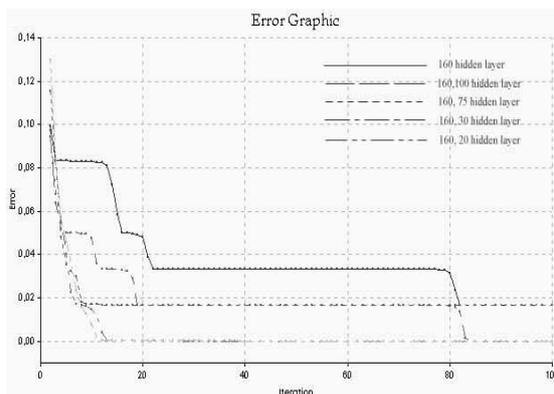
Setelah diketahui jumlah *node hidden layer* yang optimum, maka kemudian dilanjutkan untuk menentukan nilai *learning rate* yang optimum. Untuk itu jaringan saraf tiruan akan diuji dengan menggunakan 1 *hidden layer* dengan 160 *node*, serta momentum 0,9, 0,75, 0,5, 0,25 dan 0,1. Dari pengujian dapat diambil kesimpulan bahwa makin besar nilai *learning rate* yang digunakan, maka jumlah iterasi yang dibutuhkan untuk mencapai error yang kecil makin sedikit. Tetapi penggunaan nilai *learning rate* yang terlalu besar akan memperbesar kemungkinan error yang terjadi. Sehingga nilai *learning rate* yang baik tercapai pada nilai yang tidak terlalu besar ataupun terlalu kecil. Dari hasil eksperimen, diambil nilai *learning rate* 0,5 sebagai nilai yang terbaik.

Selain menentukan nilai *learning rate*, maka perlu ditentukan pula nilai momentum yang optimum. Dari pengujian yang dilakukan didapat bahwa makin kecil nilai momentum maka makin banyak iterasi yang dibutuhkan untuk mencapai error yang kecil. Untuk itu perlu diambil nilai momentum yang optimum dimana dalam program ini diambil nilai momentum 0,75. Hal ini disebabkan karena dengan menggunakan momentum 0,75 akan diperoleh error yang kecil dengan jumlah iterasi yang tidak terlalu banyak ataupun terlalu sedikit.

Setelah menentukan parameter-parameter untuk satu *hidden layer*, maka sekarang akan diuji respon sistem jika menggunakan *hidden layer* lebih dari satu, dengan menggunakan nilai *learning rate* dan momentum yang didapat dari pengujian di atas. Dari pengujian ini (gambar 5 dan Gambar 6) didapat bahwa sistem akan optimal jika memakai struktur tiga *hidden layer* dengan konfigurasi 160, 100 dan 30, dimana dengan struktur tersebut dicapai error yang kecil dalam iterasi yang singkat. Jadi dari semua pengujian yang dilakukan didapat struktur jaringan saraf tiruan yang optimal untuk sistem pengenalan suara ini ialah menggunakan 3 *hidden layer* dengan jumlah *node* 160, 100 dan 30 serta nilai *learning rate* yang digunakan 0,5 dan momentum 0,75.



Gambar 5. Struktur Jaringan Saraf Tiruan



Gambar 6. Grafik Error dengan Satu dan Dua Hidden Layer

3.2. Uji Pengenalan Suara

Pada tahap awal uji pengenalan dilakukan terhadap sinyal suara yang sama persis dengan yang telah ditrainingkan (*training data set*) dan didapat hasil bahwa error yang terjadi sebesar 12,5 % atau

dengan kata lain keakuratan sistem untuk mengenali pola *training data set* mencapai 87,5 % (Tabel 2).

Kemudian dilakukan pengujian terhadap sinyal suara secara langsung dari *microphone* oleh orang yang sama dengan yang telah dilatihkan (pembicara1 s/d pembicara 8) ataupun oleh orang yang belum pernah dilatihkan sebelumnya yaitu pembicara 6 s/d 8 (*blind data set*). Dari proses pengujian ini didapat error rata-rata sebesar 10 % atau dengan kata lain keakuratan sistem untuk pengenalan pola *blind data set* mencapai 90 % (Tabel 3). Di sini tampak bahwa untuk pengenalan kata pada pembicara 1 s/d 5 terdapat error yang rendah, sedangkan pada pembicara 6 s/d 8 tampak error lebih tinggi, namun masih tetap bisa dikenali dengan kesalahan sekitar 20 %.

Tabel 2. ERROR RATE PADA PENGUJIAN DENGAN TRAINING DATA SET

Pembicara	Error Rate
Pembicara 1	0%
Pembicara 2	0%
Pembicara 3	0%
Pembicara 4	10%
Pembicara 5	0%
Pembicara 6	0%
Pembicara 7	0%
Pembicara 8	0%
Error rata-rata	12,5 %

Tabel 3.. ERROR RATE PADA PENGUJIAN DENGAN BLIND DATA SET

Pembicara	Error Rate
Pembicara 1	0 %
Pembicara 2	0 %
Pembicara 3	10 %
Pembicara 4	10 %
Pembicara 5	0 %
Pembicara 6	20 %
Pembicara 7	20 %
Pembicara 8	20 %
Error rata-rata	10 %

4. Kesimpulan

Berdasarkan hasil analisis terhadap data yang telah diperoleh pada penelitian ini, maka diambil kesimpulan sebagai berikut :

1. Keakuratan sistem pengenalan suara untuk pengenalan *training data set* mencapai 87,5 % dan untuk pengenalan *blind data set* mencapai 90 %.
2. Kesalahan pengenalan yang terjadi diakibatkan adanya perbedaan yang terlalu besar antara sinyal suara yang hendak dikenali dengan sinyal suara yang dilatihkan, hal ini dapat diatasi dengan menambahkan/memperbanyak berbagai variasi pola kata pada saat pelatihan dengan demikian sistem jaringan lebih diperkaya pengetahuannya.
3. Terbuka penelitian lanjutan untuk memperbesar jumlah perbendaharaan kata, dan penggunaan metode hibrid lainnya sehingga pengenalan kata bersifat speaker independent.

Daftar Pustaka

- [1] Agustini, Ketut, 2006. Biometrika Suara dengan Transformasi Wavelet Berbasis Orthogonal Daubenchies, Manajemen Informatika, Fakultas PTK Undiksha Singaraja
- [2] Arun A. Ross, Karthik Nandakumar dan Anil K. Jain, 2006, Handbook of Multibiometrics, Springer Science+Business Media, New York
- [3] Bhatnagar, B.E., Mukul, 2002. A Modified Spectral Subtraction Method Combined With Perceptual Weighting For Speech Enhancement, Thesis, Master of Science In Electrical Engineering, The University of Texas Dallas
- [4] Bromba GmbH, Bioidentification Frequently Asked Questions. Updated 2003-09-12 [Retrieved October 28, 2003] From: <http://www.bromba.com/faq/biofaq.htm>.
- [5] Burge, M. and Burger, W. Ear Biometrics. In A. Jain R. Bolle and S. Pankanti, editors, BIOMETRICS: Personal Identification in a Networked Society, pp. 273-286. Kluwer Academic, 1998.
- [6] Burrus, CS, Gopinath, RA and Guo, H. 1998. Introduction to Wavelets and Wavelets Transforms A Primer. International Edition. New Jersey : Prentice-Hall International, Inc.
- [7] Campbell, JPJR, 1997, Speaker Recognition: A Tutorial. Proc IEEE, Vol.85, no.9 pp 1437-1462
- [8] Centigul H.E, 2005, Multimodal Speaker/Speech Recognition Using Lip Motion, Lip Texture And Audio Turkey : College of Engineering, Koc-University, Sariyer, Istanbul 34450
- [9] F. Itakura, 2006, Fundamentals of speech analysis and synthesis and its application to speech coding
- [10] Gonzales RC, and Wood, RE, 2002, Digital Image Processing, Second Edition, New Jersey : Prentice-Hall, Inc.
- [11] Jain, Anil K., Ross, Arun ., Pankanti, Sharath., IEEE Transactions on Information Forensics and Security, Volume 1, Issue 2, June 2006, pp. 125 - 143
- [12] J. Fraden and M.R. Neumann, "QRS wave detection", *Med. Biol. Eng. Comput.*, vol.18, pp. 125-132, 1980.
- [13] Jain, A., Hong, L., Pankati, S. Biometric Identification. Communications of the ACM, February 2000/Vol. 43, No. 2, pp. 91-98.
- [14] Mihelic France, 2008. Speech Recognition Technologies and Applications. I-Tech Education and Publishing KG. Croatia
- [15] Mudry, AH, 1997, Speaker Identification using Wavelet Transform, Tesis Master of Engineering, Ontario : Ottawa-Carleton Institut of Electrical Engineering
- [16] Proakis, JG dan Manolakis, DG, 1997. Penrosesan Sinyal Digital. Edisi Bahasa Indonesia jilid 1, Jakarta : Prenhallindo
- [17] Rabiner, L.R, Juang BH, 1993, Fundamentals of Speech Recognition, New Jersey : Prentice Hall. ISBN 013-015157-2
- [18] Ratha, N.K., Senior, A., Bolle, R.M. Automated Biometrics in Proceedings of International Conference on Advances in Pattern Recognition, Rio de Janeiro, Brazil, March 2001.
- [19] Sarikaya Ruhi, 1998, Wavelet Packet Transform With Application To Identification, Robust Speech Laboratory, Duke University, Durham
- [20] Freeman, James.A. *Neural Network Algorithms, Applications, and Programming Techniques*, Addison-Wesley Publishing Company, Inc., 1991.
- [21] Fausett, Laurene. *Fundamentals Of Neural Network*. Englewood Cliffs, New Jersey : Prentice-Hall, Inc., 1994.
- [22] Rabiner, L.R., Juang, B.H. *Fundamentals Of Speech Recognition*. Englewood Cliffs, New Jersey : Prentice-Hall, Inc., 1993.
- [23] Oppenheim, Alan.V. *Discrete-Time Signal Processing*. Englewood Cliffs, New Jersey : Prentice-Hall, Inc., 1989.

- [24] Orfanidis, Sophocles.J. *Optimum Signal Processing*. Singapore : McGraw-Hill Book Co., 1990.
- [25] Eberhart, Russell.C. *Neural Network PC Tools*. San Diego, California : Academic Press. Inc., 1990.
- [26] Todd, Bill., Kellen, Vince. *Delphi A Developer's Guide*. New York : M&T Books, 1995.
- [27] David, Frederikus. B.S. *Penggunaan Prosesor Sinyal Digital Keluarga TMS320 Sebagai Alat Pengenalan Suara Manusia Dengan Algoritma DTW (Dynamic Time Warping)*. Surabaya : Universitas Kristen Petra, 1996.