

ANALISA TENTANG PENGARUH PANJANG SINYAL SUARA DATA PRIMER TERHADAP KINERJA SISTEM IDENTIFIKASI PEMBICARA

Nurul Hidayat ¹⁾, Aris Tjahyanto ²⁾

- 1) Jurusan Matematika – FMIPA, Institut Teknologi Sepuluh Nopember
- 2) Jurusan Sistem Informasi – FTIF, Institut Teknologi Sepuluh Nopember
Kampus ITS, Sukolilo – Surabaya 60111

ABSTRAK

Penelitian ini mencoba menganalisa pengaruh panjang sinyal data primer, yang diolah menggunakan metode estimasi trispektrum, terhadap kinerja sistem pembicara melalui sinyal suara yang diucapkan. Proses pengolahan data primer ini dilakukan untuk mendapatkan estimasi data magnitudo dan fase sinyal suara. Pada proses ini, sinyal suara yang berupa sebuah kata yang diucapkan, dibagi menjadi beberapa chunk. Setiap chunk diestimasi kemudian dicari rata-ratanya. Hasil estimasi ini, berupa data magnitudo dan fase sinyal, disimpan menjadi beberapa kategori, yakni disimpan seluruh data, separuh data, seperempat, dan seterusnya. Masing-masing dianalisa berdasarkan waktu eksekusi estimasi, jumlah data yang dihasilkan estimasi itu, dan pengaruhnya terhadap kinerja sistem pembicara.

Dari beberapa ujicoba menunjukkan bahwa ukuran panjang sinyal yang diestimasi memberikan pengaruh yang cukup signifikan terhadap kinerja sistem pada seperempat bagian atau yang lebih kecil lagi. Lama eksekusi proses estimasi untuk seperempat, seperdelapan, dan seperenambelas bagian sinyal tidak sampai satu detik. Sedangkan untuk satu bagian sinyal membutuhkan sembilan detik, dan untuk setengah bagian sinyal membutuhkan waktu eksekusi selama satu detik.

Kata kunci : Sistem identifikasi, estimasi trispektrum.

1. PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi dewasa ini mengalami kemajuan yang sangat pesat, khususnya yang berkaitan dengan alat-alat elektronis yang serba otomatis, praktis pemakaiannya, nyaman, memiliki portabilitas tinggi, dan harganya terjangkau untuk kalangan menengah atas. Berbagai produk teknologi informasi yang dipasarkan saat ini semakin memanjakan konsumen, misalnya pesawat telepon yang memiliki *memory*, pintu elektronis, mesin-mesin sensor, robot, dan lain-lain.

Salah satu bagian penting untuk dapat memberikan kontribusi terhadap perkembangan teknologi informasi adalah dengan menciptakan atau mengembangkan suatu sistem identitas (mesin) yang didasarkan pada suara ucapan. Sampai dengan saat ini, pengembangan sistem suara ini sudah banyak dilakukan, baik oleh kalangan akademisi maupun perusahaan-perusahaan yang memproduksi peralatan elektronik, namun belum bisa mencapai tingkat kinerja yang diharapkan, apalagi sistem identifikasi yang bersifat *real time*, khususnya di Indonesia. Oleh karena itu, saat ini berbagai kalangan saling berlomba melakukan penelitian untuk mengembangkan sistem suara (pembicara) yang memiliki kinerja tinggi, diukur dari

kemampuan mengidentifikasi suara pada suasana tanpa *noise* maupun dengan *noise*, dan bersifat *real time*.

Hal yang terpenting untuk membangun sistem identifikasi suara yang bersifat *real time* adalah dengan menekan waktu komputasi (*run time*) serendah mungkin. Dengan kata lain, melakukan analisa tentang pengaruh panjang sinyal suara yang dikomputasi terhadap kinerja sistem. Analisa ini diperlukan untuk mendapatkan informasi tentang panjang sinyal yang optimal yang harus dikomputasi yang relatif tidak menurunkan tingkat identifikasi sistem secara signifikan.

1.2 Permasalahan

Permasalahan yang muncul di dalam melakukan optimalisasi panjang sinyal ucapan sebuah kata dapat dirumuskan sebagai berikut:

- a) Merumuskan algoritma estimasi trispektrum terhadap sinyal suara $X(t)$ untuk setiap bagian pertama subsinyal dari setiap panjang sinyal (lebar ucapan) $2^{-n} * \lambda$ ($n = 0, 1, 2, \dots$) bagian pertama dari setiap sampel sinyal $X(t)$. λ adalah panjang satu bagian penuh dari sinyal suara yang diucapkan.

- b) Mengestimasi data primer dengan panjang sinyal $2^n * \lambda$ ($n = 0, 1, 2, \dots$) bagian pertama dari setiap sampel sinyal $X(t)$.
- c) Mengimplementasikan dan mengujicoba hasil estimasi trispektrum pada sistem identifikasi suara.
- d) Menganalisa pengaruh panjang sinyal suara ucapan terhadap kinerja sistem, dengan parameter pembanding adalah panjang satu sinyal penuh.

1.3 Tujuan

Tujuan dari penelitian ini adalah untuk mendapatkan informasi tentang panjang optimal sinyal suara yang perlu diestimasi sehingga biaya komputasi dapat diturunkan sekecil mungkin tetapi kinerja sistem identifikasi masih tetap tinggi. Sebagai pembandingnya untuk mencapai tujuan ini adalah hasil estimasi panjang sinyal satu bagian.

1.4 Ruang Lingkup

Penelitian ini merupakan tindak lanjut dari beberapa penelitian yang telah peneliti lakukan, yang menggunakan estimasi trispektrum sebagai prapengolahan datanya.

Ruang lingkup penelitian meliputi transformasi fourier untuk mendapatkan domain frekwensi sinyal data primer, estimasi dan analisis trispektrum, implementasi estimasi dan ujicoba hasil estimasi pada sistem identifikasi suara menggunakan kuantisasi skalar sebagai ekstraksi cirinya dan jaringan syaraf tiruan propagasi balik sebagai pengklasifikasi. Semua implementasi ini menggunakan *Personal Computer* (PC) dengan sistem operasi Windows NT dan Linux.

2. TINJAUAN PUSTAKA

Umumnya, pengembangan sistem suara dipengaruhi oleh tiga status pengolahan datanya, yakni status pengolahan data primer (prapengolahan data), pengolahan data untuk mengekstraksi ciri (kuantisasi data), dan pengklasifikasi data. Sistem suara yang prapengolahan datanya menggunakan estimasi telah dilakukan, misalnya yang didasarkan pada analisa *power spectrum* [8], bispektrum [3,4], maupun trispektrum [5] dengan hasil masih jauh dari sempurna; yakni kemampuan identifikasinya masih dibawah 90%. Disamping itu, ketiga sistem ini mengabaikan biaya komputasi, karena yang menjadi target penelitian dari ketiga sistem itu adalah kemampuan mengidentifikasi. Pada hal, semestinya faktor biaya komputasi ini juga perlu diperhitungkan. Sebagai hipotesa, untuk menekan biaya komputasi pada pengembangan sistem suara

adalah dengan memperkecil durasi waktu dari sinyal (panjang sinyal) yang diestimasi.

Sistem yang telah dikembangkan pada [4, 8] memiliki tingkat ketahanan yang sangat rendah terhadap noise Gaussian aditif. Sedangkan sistem pada [5] walaupun memiliki tingkat ketahanan yang lebih tinggi (berifat *robust*) namun biaya komputasinya sangat tinggi sebagai akibat peningkatan dimensi estimasi dan kurangnya memanfaatkan sifat-sifat simetris yang dimiliki trispektrum.

Didasarkan pada sifat *robust* yang dimiliki oleh sistem yang didasarkan pada analisa trispektrum, penelitian ini dilanjutkan dengan mencoba menganalisa pengaruh panjang sinyal suara pembicara terhadap kinerja sistem. Caranya adalah demikian: Panjang satu sinyal dari suatu ucapan didefinisikan sebagai durasi waktu yang diperlukan untuk mengucapkan satu kata. Misalkan pada durasi waktu t panjang sinyal suara yang direkam adalah λ . Pada proses prapengolahan data primer ini, mula-mula data yang diestimasi sepanjang durasi waktu t , kemudian dilanjutkan beturut-turut sebesar $\frac{1}{2} t$, $\frac{1}{4} t$, ..., $2^{-n} t$. Jelas bahwa biaya komputasi akan turun untuk n yang membesar.

Hasil dari masing-masing estimasi pada proses pengolahan data primer selanjutnya dikelompokkan menggunakan metode kuantisasi skalar untuk mendapatkan skalar perwakilan yang nantinya menjadi masukan (*input*) pada jaringan syaraf tiruan propagasi balik. Dari beberapa eksperimen yang akan dilakukan pada masing-masing estimasi ini diharapkan akan diketahui nilai optimal dari panjang sinyal data primer yang akan dikomputasi tanpa mempengaruhi kinerja sistem.

3. METODOLOGI

Metode yang digunakan dalam penelitian ini terdiri dari beberapa tahap (Lihat Gambar 1):

Tahap 1: Akuisisi data sinyal suara (data primer).

Akuisisi data primer berupa sinyal suara kata "maju" dilakukan menggunakan Personal Computer (PC) berprosesor pentium, perangkat lunak *Sound Forge* dan *Gold Wave*.

Sinyal suara kata "maju" direkam dari 10 orang berlainan, enam berjenis kelamin laki-laki dan empat perempuan, berusia 5 tahun sampai 50 tahun. Perekaman suara dilakukan pada suasana sunyi dengan durasi waktu untuk setiap suara kata "maju" sekitar 1.28 sampai dengan 1.30 detik. Hasil perekaman berupa file-file *.raw.

Tahap 2: Rancangan algoritma estimasi trispektrum.

Algoritma estimasi trispektrum pada penelitian ini menggunakan asumsi bahwa sinyal suara ucapan manusia merupakan sinyal ergodik dan bersifat

stasioner. Sinyal ergodik adalah suatu sinyal dimana nilai rata-rata dari beberapa sinyal sama dengan nilai dari satu sinyal. Sedangkan sinyal stasioner adalah bahwa nilai sinyal di suatu titik waktu tertentu bersifat tetap.

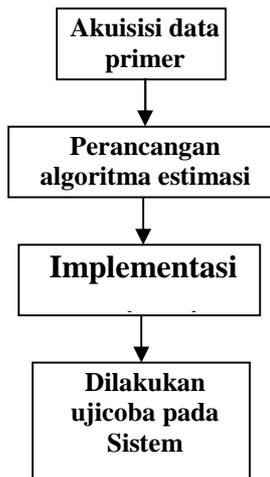
Tahap 3: Implementasi (*coding*) dari algoritma estimasi trispektrum

Algoritma estimasi ini diimplementasikan menggunakan bahasa pemrograman C dengan kompilator GNU C pada PC yang berbasis Linux.

Hasil estimasi terhadap sinyal suara berupa data magnitudo dan fase sinyal untuk masing-masing suara ucapan. Untuk data magnitudo disimpan ke dalam file-file *.tmg, sedangkan data fase disimpan ke dalam file *.tph

Tahap 4: Ujicoba pada Sistem

Pada tahap ini dilakukan *running* program dan ujicoba hasil estimasi pada sistem sebagai proses validasi untuk digunakan menganalisa dan menentukan panjang sinyal data primer yang optimal.



Gambar 1. Diagram metodologi penelitian

4. ESTIMASI TRISPEKTRUM

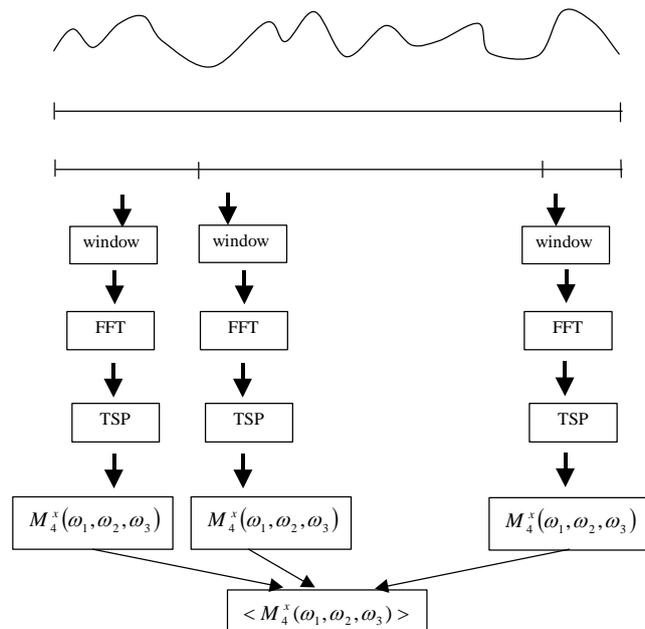
Fokus bahasan penelitian ini sesungguhnya terletak pada proses prapengolahan data primer, yakni mengestimasi sinyal suara untuk mendapatkan data magnitudo dan fase sinyal suara. Panjang sinyal yang diestimasi dapat bervariasi, mulai dari satu bagian, setengah bagian, seperempat bagian, dan seterusnya. Yang perlu dianalisa adalah pengaruh panjang sinyal yang diestimasi ini terhadap kinerja sistem identifikasi. Hasil yang optimal adalah bila panjang sinyal yang diestimasi pendek yang secara signifikan dapat menurunkan waktu eksekusi pada proses estimasi tetapi tidak mempengaruhi kinerja sistem.

Proses estimasi yang digunakan pada penelitian ini menggunakan analisis order empat yang dikenal sebagai estimasi trispektrum, dengan rumus:

$$M_T(f_1, f_2, f_3) = \frac{1}{T}$$

$$[X_T(f_1) \cdot X_T(f_2) \cdot X_T(f_3) \cdot X_T^*(f_1 + f_2 + f_3)]$$

dimana $X_T(f_i)$ adalah spektrum Fourier berdurasi T pada frekwensi f_i , dan $X_T^*(f_i)$ adalah konjugasi kompleks dari $X_T(f_i)$ [2].



Gambar 2. Diagram alir estimasi trispektrum sinyal suara.

Proses estimasi trispektrum terhadap sinyal suara dapat dilihat pada Gambar 2. Sedangkan ruang estimasinya berupa setengah kubus dengan jumlah

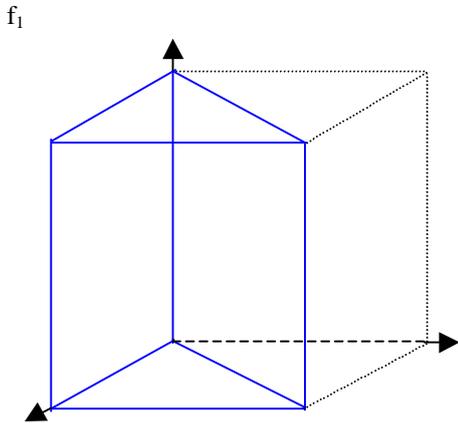
data sebanyak $\frac{N * N}{2} * (N + 1)$. Untuk satu

bagian sinyal, N bernilai 64. Lihat Gambar 3.

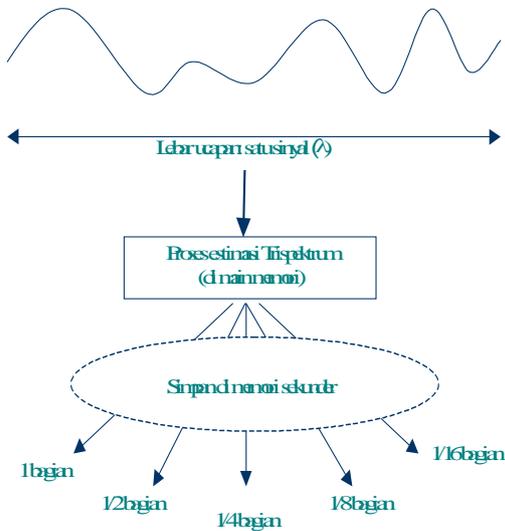
Ada dua pendekatan yang bisa dilakukan untuk menganalisa lebar ucapan ini, yaitu:

- Pendekatan I: Lebar ucapan saat proses estimasi di *main memory* berupa satu sinyal penuh, tetapi yang disimpan di memori sekunder bervariasi, mulai dari satu bagian, 1/2 bagian, 1/4 bagian, 1/8 bagian, dan 1/16 bagian. Lihat Gambar 4. Pendekatan kedua ini digunakan mengingat pada estimasi terdapat suatu transformasi yang bersifat otokorelasi pada domain frekwensi, sehingga supaya tidak kehilangan informasi pada bagian yang diestimasi.

f_3



Gambar 3. Ruang estimasi trispektrum berupa setengah kubus berukuran 32x32x33.



Gambar 4. Pendekatan I dari estimasi trispektrum.

- Pendekatan II: Pada pendekatan ini, bagian sinyal (lebar ucapan) yang diestimasi benar-benar diambil bervariasi, mulai dari satu bagian, 1/2 bagian, 1/4 bagian, 1/8 bagian, dan 1/16 bagian saja. Karena estimasi trispektrum ini terdapat otokorelasi pada domain frekuensi, maka selang frekuensi yang diolah hanya sepertiga dari selang yang seharusnya untuk setiap bagian sinyal yang diestimasi. Misalkan lebar sinyal yang diestimasi adalah l , maka $f_1 + f_2 + f_3 \leq l$. Mengingat $f_1, f_2, dan f_3$ berasal dari domain yang sama, maka $f_1, f_2, f_3 \leq \frac{1}{3}l$. Jadi, dengan pendekatan ini ada informasi yang hilang pada bagian yang diestimasi. Hasil estimasi dari masing-masing bagian ini kemudian disimpan di memori sekunder.

Proses estimasi pada penelitian ini hanya menggunakan pendekatan I saja. Rancangan algoritmanya adalah sebagai berikut:

- (1) Sinyal suara $X(t)$ disampling pada frekuensi 11 kHz pada durasi waktu 1.28 detik sampai dengan 1.30 detik, mempunyai $N = 11000 * 1.28$ titik sampel (simpangan). Sinyal ini dibagi menjadi K chunk (potongan) yang masing-masing mempunyai M titik sampel. Misalkan masing-masing chunk ini dinamai λ_i , dengan $i = 0, 1, 2, \dots, 109$.
- (2) Setiap nilai sinyal pada masing-masing chunk “dimasukkan” ke dalam window Hamming :

$$d(u) = \begin{cases} 0.54 + 0.46 \cos(\pi u), & |u| \leq 1 \\ 0, & \text{untuk } u \text{ lainnya} \end{cases}$$

dengan tujuan untuk mereduksi terjadinya frekuensi tinggi di titik-titik perpotongan chunk.

- (3) Pada setiap chunk dilakukan transformasi Fourier menggunakan algoritma FFT.
- (4) Untuk masing-masing chunk dilakukan estimasi trispektrum menggunakan rumus:

$$M_4^x(\omega_1, \omega_2, \omega_3) = X(\omega_1)X(\omega_2)X(\omega_3)X^*(\omega_1 + \omega_2 + \omega_3)$$

dimana $X(\omega)$ adalah hasil transformasi Fourier dari $X(t)$

$$\omega_i = 2\pi f_i \quad (f_i = \text{frekuensi sinyal})$$

- (5) Terakhir, masing-masing hasil estimasi ini dicari rata-ratanya yang disimpan dalam sebuah file magnitude dan sebuah file fase. Pada proses penyimpanan nilai-nilai hasil estimasi ini dibuat berubah-ubah mengikuti rumus $2^{-k} * \lambda_i$ bagian pertama, dengan $k=0, 1, 2, 3, 4$ dan λ_i adalah panjang sinyal setiap chunk. Panjang λ_i ini sesungguhnya merepresentasikan satu bagian sinyal, karena hasil dari proses estimasi satu bagian sinyal dari sebuah kata “maju”. estimasi trispektrum terhadap sinyal ditentukan dengan mengambil rata-ratanya.

5. HASIL PENELITIAN

Pada penelitian ini, proses estimasi dilakukan terhadap 5 bagian sinyal, yakni satu bagian, setengah bagian, seperempat bagian, seperdelapan, dan seperenam belas bagian sinyal. Hasil estimasi ini kemudian dilakukan ekstraksi ciri terhadap masing-masing bagian; dan dilanjutkan proses .

Setelah dilakukan beberapa kali ujicoba pada sistem suara, maka didapatkan hasil penelitian sebagai berikut:

Tabel 1. Hasil ujicoba pengenalan sistem untuk beberapa bagian sinyal hasil estimas.

No	Bagian sinyal yg diestimasi	Waktu Estimasi (detik)	Jumlah data	Tingkat pengenalan (%)
1	1	9	133,120	81.625
2	½	1	16,898	81.5
3	¼	0	2,176	34
4	1/8	0	288	27
5	1/16	0	40	12.5

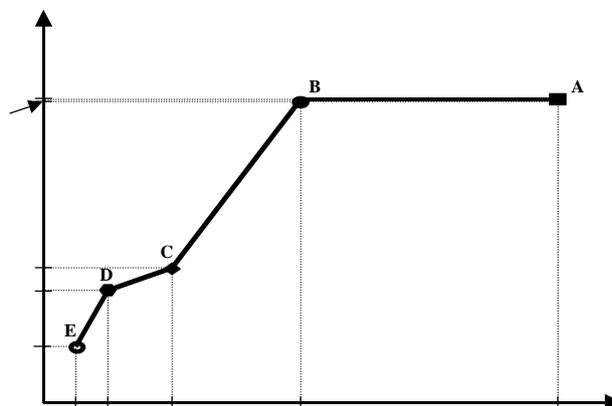
6. PEMBAHASAN

Pembahasan atau analisis terhadap hasil penelitian ini dilakukan dengan menggunakan tiga parameter, yakni tingkat pengenalan sistem terhadap suara pembicara (prosentase pengenalan), lama eksekusi (waktu) saat estimasi, dan jumlah data magnitude dan fase hasil estimasi. Kriteria yang digunakan untuk menentukan bagian sinyal yang paling baik pada proses estimasi adalah tingkat pengenalan yang tinggi, waktu eksekusi rendah, dan jumlah data yang kecil. Dari ketiga parameter itu, tingkat pengenalan sistem merupakan parameter yang paling penting. Artinya, jika tingkat pengenalannya rendah, maka parameter lainnya menjadi tidak berguna.

Pada Tabel 1 tampak bahwa untuk seper-empat bagian sinyal atau yang lebih kecil, walaupun waktu eksekusi proses estimasi yang diperlukan mendekati nol detik dan jumlah data yang dihasilkan juga kecil, tetapi prosentase pengenalan sistem relatif rendah dibanding pada satu bagian sinyal. Sedangkan untuk setengah bagian sinyal, prosentase pengenalannya hampir sama dengan satu bagian sinyal, atau hanya mengalami penurunan sebesar 0,15% saja. Waktu eksekusi estimasi untuk setengah bagian ini jauh lebih cepat atau terjadi penurunan sebesar 88,89%, dan jumlah data yang dihasilkan mengalami penurunan yang cukup signifikan, yakni sebesar 87,31%.

Berikut ini informasi tentang pengaruh panjang sinyal (bagian sinyal yang diestimasi) berturut-turut terhadap kinerja sistem. Perhitungan gradien pada setiap segmen garis dimaksudkan untuk melihat pengaruh yang signifikan akibat panjang sinyal yang diestimasi diperkecil, mulai dari satu bagian penuh sampai dengan 1/16 bagian saja. Pada Gambar 5 tampak bahwa gradien terkecil terjadi pada segmen garis BA, yaitu sebesar 0.25. Ini berarti bahwa pengaruh panjang yang diestimasi dari satu bagian

sinyal menjadi setengah bagian sinyal hampir tidak ada pengaruhnya terhadap kinerja sistem.



Gambar 5. Pengaruh lebar ucapan data primer (suara) terhadap kinerja sistem

Walaupun untuk panjang sinyal yang diestimasi dari setengah bagian menjadi seperempat bagian terjadi penurunan jumlah data maupun waktu estimasi, tetapi untuk bagian ini terjadi penurunan kinerja sistem yang cukup signifikan, sehingga lebar ucapan seperempat bagian maupun yang lebih kecil lagi menjadi tidak berarti.

7. KESIMPULAN

Setelah melakukan penelitian ini, maka dapat ditarik kesimpulan bahwa panjang sinyal yang optimal yang perlu diestimasi adalah setengah bagian saja. Dibanding panjang sinyal satu bagian, estimasi terhadap panjang sinyal setengah bagian ini dapat menurunkan waktu estimasi dan jumlah data hasil estimasi dengan prosentase penurunan berturut-turut mencapai 88.99% dan 87.31%. Sedangkan kinerja sistem hanya turun sebesar 0.58%.

DAFTAR PUSTAKA

- [1] Brigham, E. O. [1988], The Fast Fourier Transform and Its Applications, Prentice-Hall International, Inc..
- [2] Chester, M. [1993], Neural Networks : a tutorial, Prentice-Hall, International edition.
- [3] Fanany, Mohammad I., Kusumoputro, Benyamin [1998], Metode Kuantisasi pada Data Bispektrum untuk Pembicara Menggunakan Jaringan Syaraf Tiruan, Makalah, Program Studi Ilmu Komputer, Pascasarjana UI.
- [4] Fanany, Mohamad I. [1998], Bispectrum Pattern Analysis and Quantization to Speaker

- Identification, Thesis, Program Studi Ilmu Komputer UI.
- [5] Hidayat, Nurul [1999], Pengembangan Sistem Pengenal Suara Menggunakan Metode Estimasi Trispektrum dan Kuantisasi Skalar, Thesis, Program Studi Ilmu Komputer UI.
- [6] Nikias, Chrysostomos L., Petropouou, Athina P. [1993], Higher-Order Spectra Analysis : a nonlinier signal processing framework, Prentice-Hall, International edition.
- [7] Nikias, Chrysostomos L. [July 1987], Bispectrum Estimation : A Digital Signal Processing Framework, Proceeding of The IEEE Vol. 75 No. 7.
- [8] Supriyono, Ipon [1998], Penggunaan Sistem Penentu Lebar Ucapan, Penyaring Sinyal, dan Proses Filter Bank pada Prapengolahan Sistem Pengenal Kata Jaringan Neural Buatan dengan Metode Propagasi Balik, Skripsi, Fasilkom UI.
- [9] [Sept 1997], Speaker Recognition : A Tutorial, Proceeding of the IEEE, Vol. 85, No.9.
- [10] Triono, Adi [2000], Ekstraksi Ciri Pada Data Suara Menggunakan Spektra Orde Tinggi Dan Kuantisasi Vektor Untuk Pembicara Menggunakan Jaringan Neural Buatan, Thesis, Program Studi Ilmu Komputer UI.