

KLASTERING SUARA BERDASARKAN GENDER DENGAN EKSTRAKSI CIRI BERBASIS DOMAIN WAKTU

Arif Setiawan, Putri Kurnia Handayani

Fakultas Teknik, Universitas Muria Kudus

E-mail : arif_plstuff@yahoo.com, pu3_kurnia@yahoo.com

ABSTRAK

Pengenalan suara merupakan salah satu bidang artificial intellegent. Klastering suara berdasar gender dibuat dengan tujuan agar komputer dapat membedakan suara pria dan wanita. Untuk mengenali sebuah pola suara, komputer memerlukan sebuah mekanisme standar dan logis. Permasalahan utama yang terjadi apabila hendak mengenali suatu pola tertentu adalah bagaimana proses akuisisi data dilakukan hingga menghasilkan sejumlah data numerik yang representatif dan konsisten terhadap sampel yang diberikan.

Untuk sistem pengenalan suara disini, digunakan teknik ekstraksi ciri berbasis domain waktu dengan dua metode yaitu short time energy dan zero crossing rate. Tahapan yang dilakukan adalah disiapkan data sampel 10 audio .wav kemudian dilakukan proses ekstraksi ciri berbasis time domain dengan metode short time energy dan zero crossing rate. Dari data yang telah diekstraksi berupa file pola.txt, dilakukan perhitungan ekstraksi ciri sinyal audio sehingga diperoleh data rata-rata masing-masing metode. Dari data pola.txt diubah ke dalam bentuk matrik, selanjutnya dilakukan klasifikasi data menggunakan fungsi K-means.

Dari hasil penelitian disimpulkan bahwa klastering suara berdasarkan gender dapat dilakukan dengan teknik ekstraksi ciri sinyal suara berbasis domain waktu dengan metode short time energy dan zero crossing rate.

Kata kunci : Klastering suara, Time domain, K-means

1. Pendahuluan

Secara awam, aspek yang cukup penting yang mendasari berbagai teori dalam AI adalah sistem pengenalan pola (*Pattern Recognizing*). Sistem pengenalan pola merupakan komponen penting dalam proses peniruan kemampuan inderawi manusia terutama penglihatan dan pendengaran. Sebagai contoh, untuk meniru indera pendengaran manusia, komputer harus mempunyai suatu mekanisme standar dan logis dalam mengenali pola yang ada pada suara yang sedang diproses. Dari sinilah diperoleh motivasi untuk mencoba suatu konsep sederhana untuk mengenali pola dari suara sehingga dapat diidentifikasi dengan baik oleh komputer.

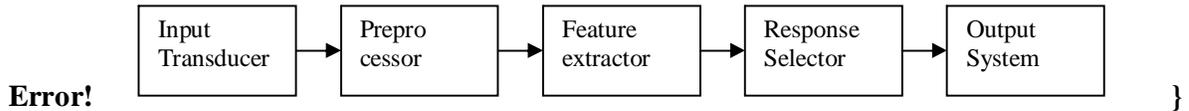
Permasalahan utama yang terjadi apabila hendak mengenali suatu pola tertentu adalah bagaimana proses akuisisi data dilakukan sehingga menghasilkan sejumlah data numerik yang representatif dan konsisten terhadap sampel yang diberikan. Dalam penulisan ini, kita mencoba mengaplikasikan suatu metode sederhana untuk mengenali suara dan mengklasifikasikannya berdasarkan gender sehingga dapat diidentifikasi dengan baik oleh komputer dengan memanfaatkan berbagai teori ekstraksi ciri untuk data audio.

Program bantu yang penulis buat untuk mensimulasikan metode ini dibuat dengan menggunakan Matlab 7.1 pada sistem operasi yang mendukung. Tujuan utama dari penulisan ini adalah untuk menganalisis dan membuktikan bahwa metode sederhana untuk mengekstraksi data suara dari sampel yang berupa input suara dari 10 orang, yang terdiri dari 8 pria dan 2 wanita dapat diimplementasikan dengan baik sesuai dengan tujuannya sehingga komputer dapat mengidentifikasi suara secara baik dan konsisten.

2. Dasar Teori

2.1 Pengenalan Suara

Pengenalan suara adalah proses secara otomatis mengenali siapa yang berbicara atas dasar informasi individu termasuk dalam gelombang suara. Suara dikenali melalui ciri-cirinya (features). Ciri-ciri tersebut digunakan untuk membedakan satu suara dengan suara lainnya. Ciri yang bagus adalah ciri yang memiliki daya pembeda yang tinggi, sehingga pengelompokan suara berdasarkan ciri yang dimiliki dapat dilakukan dengan keakuratan yang tinggi. Pengenalan suara dapat diklasifikasikan dalam identifikasi dan verifikasi [[HYPERLINK \l "Min03" 1](#)]. Identifikasi suara adalah proses pengenalan suara berdasarkan sampel suara yang telah diberikan, sedangkan verifikasi suara adalah proses penerimaan atau penolakan terhadap suara yang diberikan.



Gambar 1 Struktur Pengenalan Pola [2]

1. **Input Transducer**

Menganalisa sinyal elektronik yang diinput. Alat yang digunakan ; scanner

2. **Preprocessor**

Melakukan penambahan kondisi sinyal, termasuk fungsi penguatan sinyal, analisa spektrum dan konversi analog ke digital.

3. **Feature Extractor**

Sebagai distriminator, dengan membandingkan kecocokan template.

4. **Response selector**

Algoritma untuk memilih input pattern dengan data pattern yang sudah ada di software dengan menggunakan teknik pencarian, sortir, least-squares analysis.

5. **Output System**

Output system mungkin berupa suara, gambar atau video yang sesuai dengan hasil proses aplikasi.

2.2. **Klastering**

Salah satu teknik utama yang terkandung dalam sistem pengenalan pola adalah analisis cluster, yaitu identifikasi substruktur dalam data set berlabel [3]. Beberapa teknik klastering yang paling sederhana dan umum adalah klastering K-means. Secara detil teknik ini menggunakan ukuran ketidakmiripan untuk mengelompokkan objek. Ketidakmiripan dapat diterjemahkan dalam konsep jarak. Dua objek dikatakan mirip jika jarak dua objek tersebut dekat. Semakin tinggi nilai jarak, semakin tinggi nilai ketidakmiripannya. Secara ringkas, algoritma K-means dapat dijelaskan seperti dalam Gambar 2.

Algoritma 1. K-Means

1. Pilih jumlah klaster
2. Initalisasi k pusat klaster (diberi nilai-nilai random)
3. Tempatkan setiap data/obyek ke klaster terdekat, kedekatan dua obyek ditentukan bersarkan jarak kedua obyek tersebut. Jarak paling dekat antara satu data dengan satu klaster tertentu akan menentukan suatu data masuk dalam klaster mana.
4. Hitung kembali pusat klaster dengan anggota klaster yang sekarang. Pusat klaster adalah rata-rata semua data/obyek dalam klaster
5. Tugaskan lagi setiap obyek memakai pusat klaster yang baru. Jika pusat klaster sudah tidak berubah lagi, maka proses pengklasteran selesai.
6. Kembali ke langkah 3 sampai pusat klaster tidak berubah lagi.

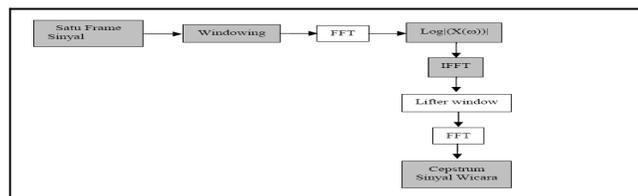
Gambar 2 Algoritma K-means [4]

2.2. **Sinyal Audio**

Tujuan melakukan estimasi dalam domain waktu adalah untuk mendapatkan nilai autokorelasi sinyal audio. Nilai auto korelasi suatu sinyal audio akan menunjukkan bagaimana bentuk gelombang itu membentuk suatu korelasi pada dirinya sendiri sebagai fungsi perubahan waktu ke waktu. Bentuk-bentuk yang sama atau mirip pada setiap *delay* waktu tertentu menunjukkan perulangan bentuk atau periodisitas pola sinyal audio. Dengan demikian akan dapat kita lakukan estimasi nilai frekuensi fundamentalnya.

2.3. **Analisa Spectral Sinyal Audio**

Proses ekstraksi ciri sinyal audio didasarkan pada sebuah diagram blok yang cukup populer seperti berikut.



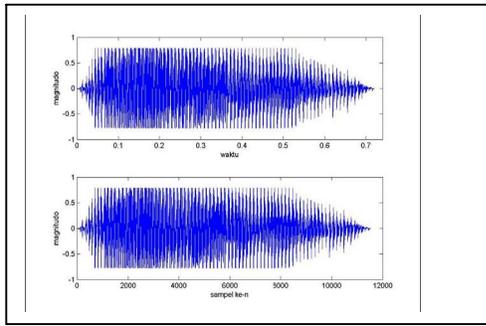
Gambar 3 Diagram blok ekstraksi ciri sinyal audio
INFRM 365

Dengan mengikuti diagram blok diatas, kita akan mendapatkan langkah-demi langkah ekstraksi ciri. Pada sub bab ini pembicaraan berkisar pada proses melihat *short time fourier analisys* dari sinyal audio, atau yang juga kita kenal sebagai *power spectral density* (PSD) sebuah sinyal audio pada durasi atau frame tertentu. Dengan mengetahui bentuk *power spectral density* (PSD) sinyal audio kita akan mampu melakukan ekstraksi ciri sinyal audio tersebut. Sebelum proses pada gambar diatas dilakukan ada baiknya kita melihat gambaran sebuah sinyal audio yang telah kita simpan dalam bentuk file “a.wav”. Setelah kita dapatkan bentuknya dalam domain waktu seperti pada Gambar 4 bagian atas, selanjutnya kita coba melihatnya sebagai fungsi dari *sampling*. Dalam hal ini kita lihat bentuk sinyal audio sesuai dengan urutan sampel yang ada. Seperti kita lihat bahwa untuk nilai sampel ke-700 sampai dengan sampel ke-8200, menunjukkan nilai magnitudo sinyal yang relatif stabil.

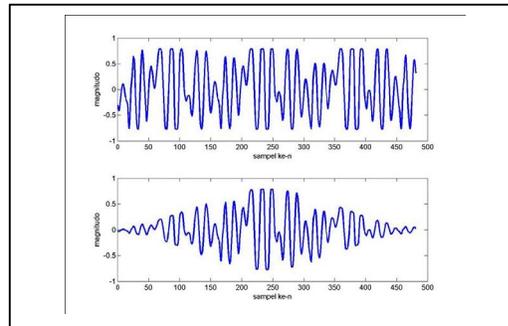
Kita lanjutkan dengan melakukan pembentukan *frame* sebuah sinyal audio seperti pada Gambar 5 bagian atas. Dengan melakukan *windowing* kita akan mendapatkan bentuk frame sinyal audio terwindow seperti pada Gambar 5 bagian bawah. Sudah tentu kita paham untuk apa proses *windowing* dilakukan disini. Dengan demikian tidak salah apabila kita mengambil satu frame sinyal dari sampel ke-2000 sampai dengan sampel ke 2480. Karena dalam satu frame kita bentuk dari:

$$\begin{aligned} \text{Sampel/frame} &= (\text{sample/detik}) * (\text{detik/frame}) \\ &= 16000 * 0,06 \\ &= 480 \text{ sampel/frame} \end{aligned}$$

Hal ini dilakukan dengan menetapkan bahwa satu frame sinyal audio sepanjang 50 ms.



Gambar 4

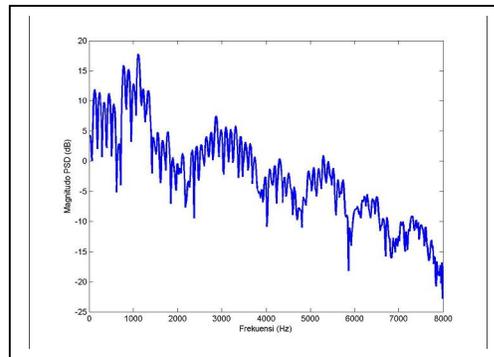


Gambar 5

Gambar 4 Sinyal audio dalam domain waktu dan sebagai fungsi sample ke-n [5]

Gambar 5 Satu frame sinyal audio dalam domain waktu [5]

Proses dilanjutkan dengan melakukan transformasi sinyal ke dalam domain frekuensi. Dengan menggunakan *fft* dan proses logaritmik akan kita dapatkan nilai *power spectral density* (PSD) sinyal audio seperti pada Gambar 6.



Gambar 6 Power spectral density sinyal audio

3. Ekstraksi Ciri

Untuk mendapatkan data yang akurat dan konsisten dari setiap sampel, digunakan suatu metode ekstraksi ciri sinyal suara dengan time domain [6]. Time domain adalah pemaparan teknik sinyal audio dasar, dimana sinyal digambarkan sebagai amplitudo dengan satuan waktu, sinyal dapat bernilai positif atau negatif tergantung pada tekanan suara. Pada makalah ini penulis menggunakan dua metode, yaitu *sort time energy* dan *zero crossing rate*. Adapun metode yang digunakan adalah sebagai berikut :

a. Sort Time Energy

Menandakan kekerasan suara pada waktu yang pendek

$$STE = \frac{1}{N} \sum_{n=1}^N X(n)^2 \quad (1)$$

Keterangan:

STE = Sort time energy

N = Jumlah Sampel

X(n) = Nilai sinyal dari sampel

b. Zero Crossing Rate

Sampel berurutan pada sebuah sinyal digital memiliki perbedaan tanda, ukuran dari noise sebuah sinyal pada fitur domain

$$ZC = \frac{\sum_{n=1}^N |sgn x(n) - sgn x(n-1)|}{2N} \quad (2)$$

Keterangan:

ZC = Zero Crossing Rate

sgn x(n) = nilai dari x(n), bernilai 1 jika x(n) positif, -1 jika x(n) negatif

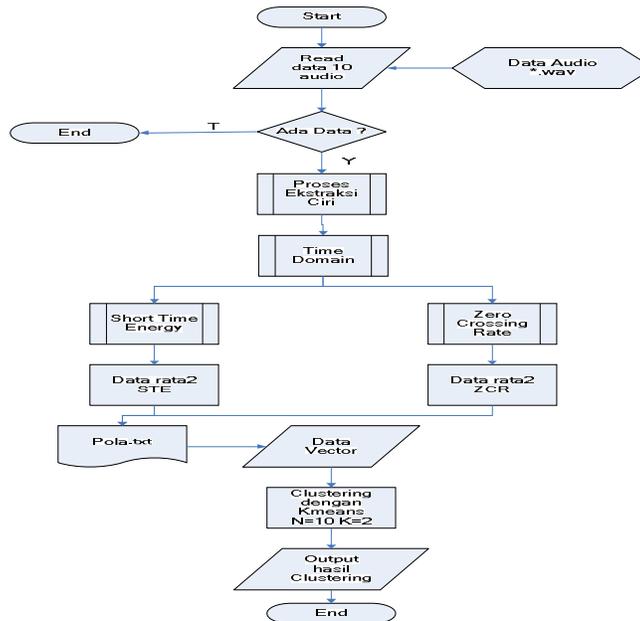
N = jumlah sampel

Setiap metode ciri diambil rata-rata nya, menggunakan standart deviasi, berikut adalah tabel penggunaan rata-rata untuk tiap-tiap metode ciri.

Tabel 1. penggunaan rata-rata untuk tiap-tiap metode ciri.

Ciri	Statistik
Sort Time Energy	Standart Deviasi (std)
Zero Crossing Rate	Standart Deviasi (std)

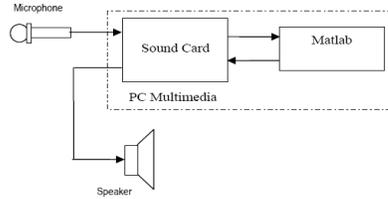
4. Algoritma Program



5. Implementasi

5.1 Persiapan Peralatan

Sebelum melakukan perekaman suara harus dilakukan penataan seperti pada gambar berikut ini.



Gambar 7 Desain perangkat perekaman pengukuran energi sinyal audio

PC harus dilengkapi dengan peralatan *multimedia* seperti *sound card*, *speaker active* dan *microphone*. Untuk *microphone* dan *speaker active* bias juga digantikan dengan *head set* lengkap. Sebelum anda memulai praktikum, sebaiknya anda tes dulu, apakah seluruh perangkat *multimedia* anda sudah terintegrasi dengan PC.

5.2 Perhitungan Ekstraksi Ciri Sinyal Audio

Data hasil perhitungan ekstraksi ciri sinyal audio yang diperoleh dari 10 sampel suara dan direkam pada file "pola.txt", data yang disimpan adalah data rata-rata dari masing-masing metode ekstraksi ciri.

Tabel 2. Data rata-rata dari metode ekstraksi ciri

No	Nama Mahasiswa	Nama File	Short Time Energy	Zero Crossing Rate
1	Desi	nesi1.wav	0.0538388	0.02373
2	Ramadan	ednofri4.wav	0.0730546	0.0305071
3	David	david4.wav	0.0662328	0.023269
4	Sofyan	sofyan4.wav	0.137326	0.0199186
5	Gunawan	gunawan4.wav	0.0923871	0.0449648
6	Herlambang	herlambang4.wav	0.112719	0.0181066
7	Hendy	hendy4.wav	0.115847	0.0283285
8	Susapto	susapto1.wav	0.129537	0.0441993
9	Arif	Arif4.wav	0.0664739	0.0190522
10	Retno	retno1.wav	0.0866736	0.0283818

Pemanggilan data dari pola .txt diubah ke bentuk matrik. Matrik M yang diperoleh:

M=

0.0538388	0.02373	0.0893531	0.13844
0.0730546	0.0305071	0.123814	0.473905
0.0662328	0.023269	0	0.511881
0.137326	0.0199186	0	0.732067
0.0923871	0.0449648	0.124977	0.483487
0.112719	0.0181066	0.181973	0.586998
0.115847	0.0283285	0.116744	0.540119
0.129537	0.0441993	0	0.655706
0.0664739	0.0190522	0.173884	0.409927
0.0866736	0.0283818	0.126805	0.313418

5.3 Klasifikasi Data Menggunakan Fungsi K-Means

Dari hasil klasifikasi menggunakan fungsi K-means, diperoleh hasil:

maxRow = 10

maxCol = 4

c =

0.0538 0.0237 0.0894 0.1384

0.0731 0.0305 0.1238 0.4739

X =

0.0538	0.0237	0.0894	0.1384	1.0000
0.0731	0.0305	0.1238	0.4739	2.0000
0.0662	0.0233	0	0.5119	2.0000
0.1373	0.0199	0	0.7321	2.0000
0.0924	0.0450	0.1250	0.4835	2.0000
0.1127	0.0181	0.1820	0.5870	2.0000
0.1158	0.0283	0.1167	0.5401	2.0000
0.1295	0.0442	0	0.6557	2.0000
0.0665	0.0191	0.1739	0.4099	2.0000
0.0867	0.0284	0.1268	0.3134	1.0000

Hasil Clustering

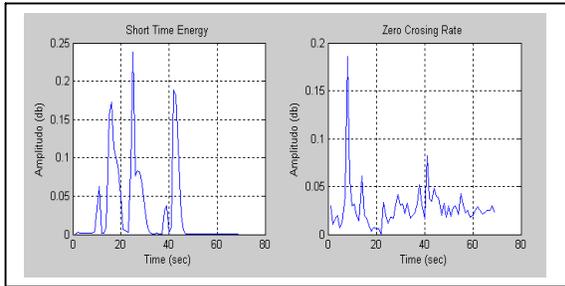
Atau dengan algoritma K-means bawaan matlab:

$X =$
1 2 2 2 2 2 2 2 1

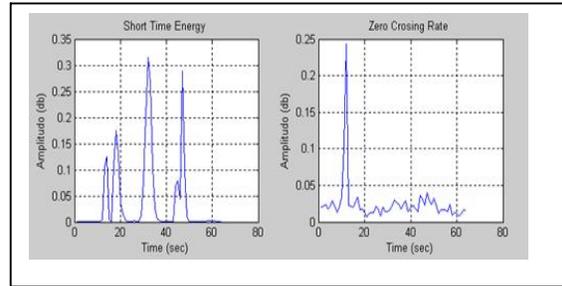
Keterangan :

1 = Wanita
2 = Pria

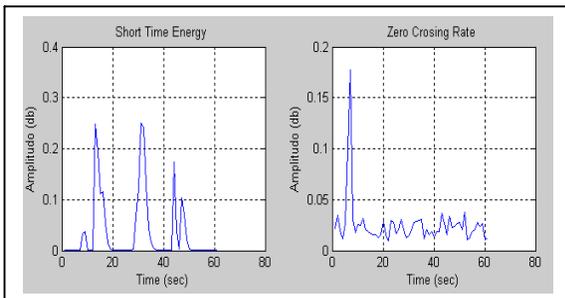
6. Hasil dan Pembahasan



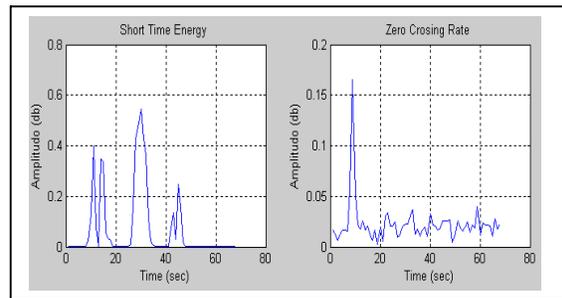
Gambar 8 Grafik ciri suara Desi dengan file desi1.wav



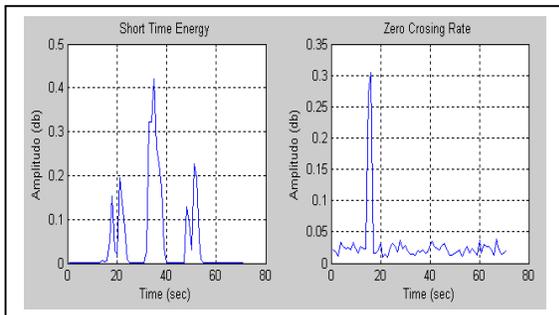
Gambar 9 Grafik ciri suara Ramadan file ramadan4.wav



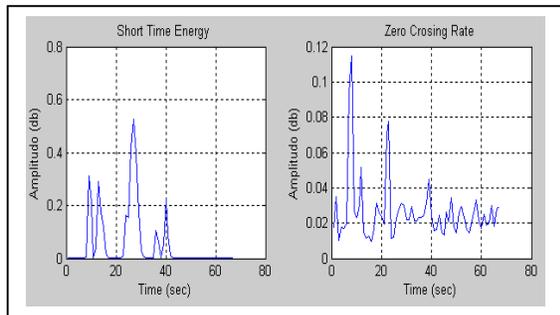
Gambar 10 Grafik ciri suara David dengan file david4.wav



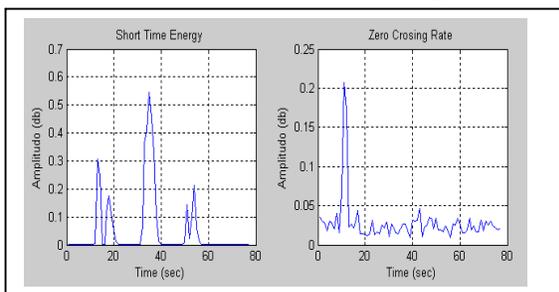
Gambar 11 Grafik ciri suara Sofyan dengan file sofyan4.wav



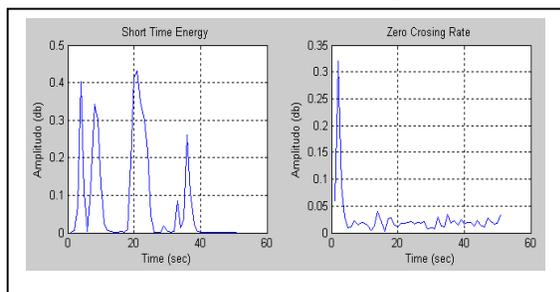
Gambar 12 Grafik ciri suara Gunawan file gunawan4.wav



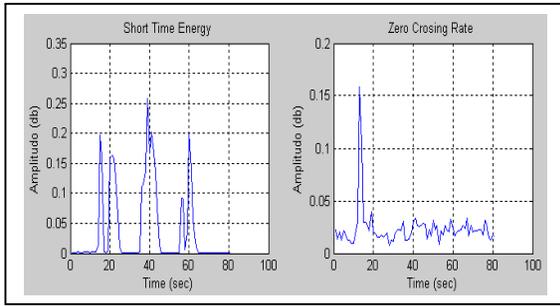
Gambar 13 Grafik ciri suara Herlambang file herlambang4.wav



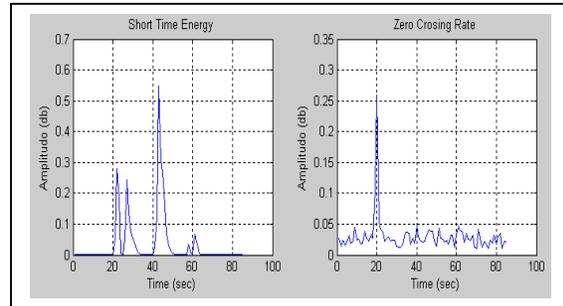
Gambar 14 Grafik ciri suara Hendy file hendy4.wav



Gambar 15 Grafik ciri suara Susapto file susapto1.wav



Gambar 16 Grafik ciri suara Arif dengan file arif4.wav



Gambar 17 Grafik ciri suara Retno file retno4.wav

7. Kesimpulan

Dari hasil makalah yang penulis susun, dapat disimpulkan bahwa pengklasifikasian suara berdasarkan jenis kelamin atau gender dapat dilakukan dengan suatu metode ekstraksi ciri sinyal suara berbasis domain waktu dan domain frekuensi. Ekstraksi ciri yang digunakan bisa bermacam-macam, diantaranya adalah dengan Sort Time Energy, Zero Crossing Rate, Spectral Centroid, dan Spectral Flux. Nilai rata-rata dari masing-masing ciri dihitung dengan standart deviasi, untuk memperoleh rata-rata nya, lalu diolah untuk pengklasifikasian.

DAFTAR PUSTAKA

- [1] Minh N. Do, "Digital Signal Processing Mini-Project "An Automatic Speaker Recognition System", " 2001.
- [2] Agus Sasmito Aribowo, "MODEL PENELUSURAN CITRA DIGITAL PADA DATABASE CITRA , " *Seminar Nasional Informatika (semnasIF)*, May 2009.
- [3] Richard J. Hathaway and James C. Bezdek, "Recent Convergence Results for the Fuzzy c-Means," *Journal of Classification*, vol. 5, pp. 237-247, 1988.
- [4] Budi Santoso, *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu, 2007.
- [5] E. Darren Ellis, "Design of a Speaker Recognition Code using MATLAB," May 2001.
- [6] Olivier Lartillot and Petri Toiviainen, "A MATLAB TOOLBOX FOR MUSICAL FEATURE EXTRACTION FROM AUDIO," in *Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07)*, Bordeaux, France, 2007.