

KLASIFIKASI SUARA MANUSIA KE DALAM SOPRAN, MEZZO SOPRAN, ALTO, TENOR, BARITON, BASS DENGAN SELF ORGANIZING MAP

Andreas Saputra¹

Sri Suwarno²

Lukas Chrisantyo³

Abstract

Self Organizing Map adalah metode jaringan syaraf tiruan (Artificial Neural Network) yang biasa digunakan untuk melakukan proses klasifikasi dengan sifat unsupervised learning atau pelatihan tak terbimbing. Cluster yang akan digunakan akan ditentukan secara manual hanya saja dalam prosesnya data yang masuk akan dikelompokkan secara otomatis tanpa adanya intervensi dari sistem. Penelitian ini menerapkan Self Organizing Map untuk melakukan klasifikasi data berupa rekaman suara dengan format file WAV karena merupakan format audio yang belum terkompresi ke dalam sopran, mezzo sopran, alto, tenor, baritone, dan bass. Dalam pengambilan data untuk input melalui proses preemphasis, frame, blocking, dan windowing sebelum dirubah menjadi sinyal diskrit dengan Fast Fourier Transform. Data berupa rata-rata magnitude menjadi input dalam sistem klasifikasi Self Organizing Map. Dalam penelitian ini hasil yang didapat belum sesuai dengan harapan karena data tidak mengelompok dengan baik.

Keywords: *Artificial Neural Network, Self Organizing Map, WAV, Fast Fourier Transform.*

1. Pendahuluan

Setiap manusia memiliki jenis suara yang berbeda-beda, baik itu pria ataupun wanita memiliki jenis suara yang berbeda-beda. Untuk pria memiliki 3 (tiga) jenis suara yaitu tenor, baritone, dan bass. Wanita juga memiliki 3 (tiga) jenis suara yaitu *sopran*, *mezzo zopran*, dan *alto*. Dengan adanya jenis suara manusia yang beragam tersebut tidak semua orang dapat menentukan atau menggolongkan suara seseorang ke dalam jenis suara tertentu. Karena dalam menentukan dan menggolongkan jenis suara tersebut dibutuhkan pengetahuan dan keahlian dalam bidang music maupun olah vokal.

Selain itu saat ini perkembangan teknologi sangat cepat terutama teknologi informasi. Sudah banyak sistem maupun alat yang diciptakan untuk membantu pekerjaan manusia. Hampir semua kegiatan saat ini dapat dilakukan oleh komputer.

Dari permasalahan tersebut maka penulis akan melakukan untuk membuat sistem yang bisa membantu dalam pengelompokan suara manusia dewasa dengan ambitus suara masing-masing manusia. Sehingga dalam pengelompokan suara manusia dapat dipercepat oleh sistem dan juga dapat membantu pekerjaan guru music atau guru vokal maupun orang belum terlalu mengerti tentang vokal.

¹Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Duta Wacana.

²Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Duta Wacana.

³Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Duta Wacana.

2. Metode

Menurut Fausett (1994) *Self Organizing Map* merupakan salah satu model Jaringan Syaraf Tiruan yang menggunakan metode unsupervised dibekali dengan pengetahuan dasar (parameter-parameter jaringan). Proses pelatihan SOM adalah proses pencocokan bobot setiap *neuron output* yang ada.

Algoritma SOM :

- Step 0. Inialisasi bobot (bisa random)
Tentukan topologi serta parameter
Tentukan Learning rate
- Step 1. Bila kondisi terpenuhi lakukan step 2-8
- Step 2. Untuk setiap input vector lakukan step 3-5
- Step 3. Untuk setiap j hitung jarak dengan rumus
$$D(j) = \sum_i (w_{ij} - x_i)^2$$
- Step 4. Pilih index J dengan D(J) paling kecil
- Step 5. Update bobot j dan tetangganya dengan
$$w_{ij}(\text{new}) = w_{ij}(\text{old}) + \alpha[x_i - w_{ij}(\text{old})]$$
- Step 6. Update Learning Rate
- Step 7. Kurangi radius tetangga topologi
- Step 8. Berhenti

Dalam penelitian ini penulis mencoba melakukan klasifikasi jenis suara manusia kedalam *Sopran, Mezzo sopran, Alto, Tenor, Bariton, dan Bass*. Masukan yang digunakan berupa rekaman suara yang akan melalui proses-proses awal sebagai berikut:

2.1. Preemphasis

Proses ini adalah tahap awal setelah akuisi data atau proses *sampling* terhadap *input*. Tujuan dari proses ini untuk memperbaiki sinyanya dari gangguan *noise*. Dalam hal ini maka tingkat akurasi akan meningkat. Rumus diperoleh dari Rabinet (1993) dan digunakan dalam penelitian ini adalah sebagai berikut :

$$Y(n) = S(n) - \alpha S(n-1) \quad [1]$$

Keterangan :

Y(n) = Hasil *pre-emphasis* ke-n dari sinyal.

S(n) = input sinyal ke-n.

α = alpha dimana nilai *default* yang digunakan adalah 0.97.

n = jumlah data sinyal *input*.

2.2. Frame Blocking

Frame Blocking adalah proses yang akan diterapkan pada sinyal atau data suara yang sebelumnya sudah dilakukan proses *preemphasis*. Dalam proses ini sinyal atau data akan dibagi menjadi beberapa potongan untuk memudahkan proses selanjutnya, potongan tersebut disebut *frame*. Setiap *frame* memiliki beberapa sampel tergantung berapa detik dan besar frekuensi samplingnya. Dalam kasus ini akan digunakan 1024 data untuk masing-masing *frame*.

2.3. Windowing

Dalam proses sebelumnya yaitu *frame blocking* akan menghasilkan efek sinyal *discontinue*. Oleh sebab itu perlu untuk menjadikan sampel suara tersebut menjadi *continue*, digunakanlah proses *windowing*. Jenis *window* yang akan digunakan adalah *Hamming*

Window. Rumus diperoleh dari Rabinet (1993) dan digunakan dalam penelitian ini adalah:

$$w(n) = 0.54 - 0.46 * \cos(2 * \pi * n / (N - 1)) \quad [2]$$

Keterangan :

w(n) : hasil ke-n dari *hamming window*.

n : sampel ke-n.

N : jumlah sampel tiap *frame*.

2.4. Fast Fourier Transform (FFT)

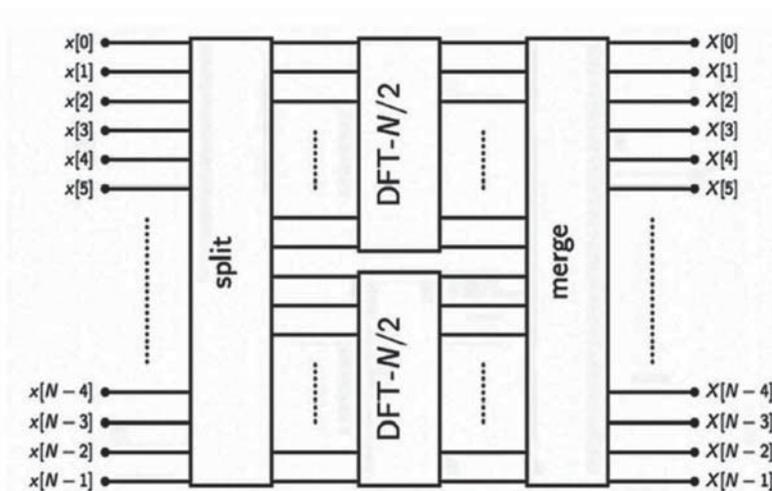
Sampel yang dihasilkan dari proses-proses sebelumnya adalah spectrum suara dalam domain waktu, sehingga harus dirubah menjadi sinyal frekuensi dengan menggunakan *Fast Fourier Transform*. Hasil dari proses FFT merupakan pendeteksian gelombang frekuensi domain dalam bentuk diskrit. Rumus FFT didapat dari Prandoni (2008) adalah :

$$X[k] = \sum_{n=0}^{N-1} X[n] W_N^{nk} \quad k = 0, \dots, N - 1 \quad [3]$$

dengan

$$W = e^{-\frac{j2\pi}{N}} \quad [4]$$

Fast Fourier Transform merupakan proses cepat perhitungan *Discrete Fourier Transform* (DFT), dengan mengubah bentuk DFT yang memiliki panjang N menjadi penjumlahan dari dua buah bentuk DFT dengan panjang masing-masing $N/2$, satu bagian terdiri dari elemen bernomor ganjil dan yang lain genap . FFT memiliki contoh skema seperti Gambar 1. dibawah ini:



Gambar 1. Skema FFT
(Dikutip dari : Chrisantyo, L. Pengolahan Sinyal Digital)

Dalam penelitian ini *input* yang dipakai adalah rata-rata nilai *magnitude* yang didapat dari proses FFT, sehingga pada kasus ini hasil yang didapat dari pengklasifikasian *Self Organizing Map* tidak cukup baik karena akurasinya sangat kecil.

3. Diskusi

Pada penelitian ini ada beberapa langkah utama dalam tahap mengambil *input* untuk proses klasifikasi SOM, yaitu *sampling*, *preprocessing* dan FFT.

Proses *sampling* dilakukan dengan cara membaca *header* dari struktur *file wav*, sehingga diperoleh nilai sampel-sampelnya berupa amplitudo.

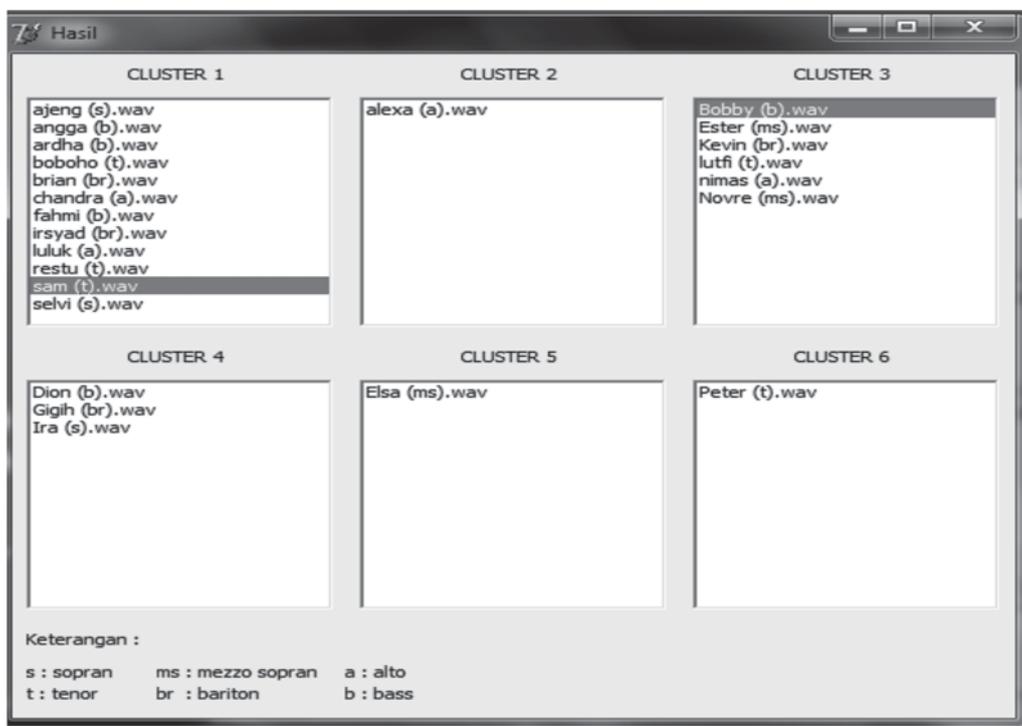
Setelah didapatkan amplitudo kemudian akan diolah untuk tahap pemrosesan awal yaitu *preemphasis*, *frame blocking* dan *windowing*. Akan diperoleh sinyal dari setiap frame.

Pada tahap akhir sinyal yang sudah didapatkan dari setiap frame tersebut akan melalui proses FFT yang akan merubah sinyal berbasis waktu menjadi sinyal berbasis frekuensi, kemudian dalam kasus ini dihitung nilai rata-rata dari *magnitude* tiap frame.

Hasil dari proses sebelumnya menjadi *input* untuk klasifikasi SOM, sehingga secara otomatis akan dikelompokkan menjadi 6 (enam) cluster yang sudah disediakan.

4. Hasil

Dalam penelitian ini proses klasifikasi menggunakan Jaringan Syaraf Tiruan yaitu *Self Organizing Map* (SOM). Setelah melakukan penelitian terhadap data suara *wav* dengan format PCM 44100 Hz 16 Bit Mono, dimana dalam prosesnya dilakukan terlebih dahulu pengambilan nilai sampel melalui *sampling*, kemudian dilakukan *preprocessing* berupa *preemphasis*, *frame blocking*, dan *windowing* sehingga didapatkan nilai sampel untuk diubah menjadi frekuensi dengan FFT. Setelah itu rata-rata *magnitude* yang didapatkan menjadi *input* untuk *Self Organizing Map*. Pada penelitian kali ini tingkat akurasi yang didapatkan tidak terlalu baik.



Gambar 2. Hasil Klasifikasi Sistem

Pada Gambar 2. Menunjukkan hasil klasifikasi dari sistem ,dengan kondisi *epoch* 100 dan *learning rate* 0.6.

Tabel 1.
Tabel hasil penelitian dengan learning rate 0.6

Nama	Cluster sesuai epoch			Jenis suara asli
	100	200	1000	
Ajeng	1	1	1	Sopran
Alexa	3	4	4	Alto
Tere	1	3	3	Sopran
Boboho	1	1	1	Tenor
Brian	2	2	3	Bariton
Chandra	1	1	1	Alto
Angga	2	3	3	Bass
Ardha	1	1	1	Bass
Fahmi	1	1	1	Bass
Luluk	1	1	1	Alto
Nimas	4	5	6	Alto
Selvi	2	3	4	Sopran

5. Kesimpulan

Berdasarkan penelitian dan pengamatan algoritma *Self Organizing Map* tidak cukup baik dalam mengklasifikasikan jenis suara dengan hanya menggunakan nilai rata-rata *magnitude* dari tiap frame pada masing-masing sampel suara.

Daftar Pustaka

- Chrisantyo, L. (2014). *Fast Fourier Transform*. Dipresntasikan pada kuliah Pengolahan Sinyal Digital TIP383, Universitas Kristen Duta Wacana.
- Fausset, L. (1994). *Fundamentals of Neural Networks, Architecture, Algorithm and Application*. Prentice Hall.
- Prandoni, P. & Vetterli, M. (2008). *Signal Processing for Communication*. EPFL Press
- Rabinet, L & Juang, B.H. (1993). *Fundamental of Speech Recognition*. Prentice Hall.