

SISTEM IDENTIFIKASI SUARA PRIA DAN WANITA BERDASARKAN USIA MENGUNAKAN MEL FREQUENCY CEPSTRAL COEFFICIENT DAN LPC DENGAN METODE KLASIFIKASI SOM

Septya Erani¹, Bambang Hidayat², Suryo Adhi Wibowo³

¹Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom

Abstrak

Manusia mampu membedakan identitas seseorang yang mereka kenal, hanya dari suaranya. Proses ini dikenal sebagai **speaker recognition**. **Speaker recognition** atau pengenalan pembicara terdiri dari dua jenis, yaitu **speaker verification**, dan **speaker identification**. Verifikasi suara menentukan apakah suara yang didengar merupakan suara seseorang yang ia klaim. Sedangkan proses identifikasi mencoba menentukan dengan spesifik identitas pembicara. Sinyal suara merupakan sinyal yang kompleks. Pada sinyal suara terdapat banyak parameter, mulai dari parameter yang bersifat subjektif seperti logat, dialek hingga parameter yang dapat diukur secara akustik seperti formant, pitch dan energi spektral. Masalah yang timbul adalah bagaimana mengekstraksi ciri dari sinyal suara yang kompleks sehingga dihasilkan data baru yang lebih sederhana tetapi tetap mempertahankan karakter khas dari sinyal suara tersebut. Pada Tugas Akhir ini akan dirancang dan direalisasikan suatu sistem yang dapat mengidentifikasi suara manusia, untuk diketahui dari seorang pria atau wanita dan kemudian ditentukan usianya. Sistem identifikasi suara ini terdiri dari ekstraksi ciri dan pengklasifikasian suara.

Penelitian yang dilakukan menggunakan parameter akustik yang sifatnya lebih objektif, dengan menggunakan metode ekstraksi ciri Mel-Frequency Cepstral Coefficient (MFCC) dan Linear Predictive Coding (LPC), dan hasil ekstraksi ini akan dikuatkan ciri-cirinya dengan metode PCA (Principal Component Analysis). Metode klasifikasi yang digunakan yaitu SOM (Self Organizing Map).

Dalam sistem yang menggunakan metode JST-SOM ini diperoleh akurasi 100% saat pengujian dengan menggunakan data latih sebagai data uji, sedangkan saat data uji diberikan noise terdapat beberapa hasil akurasi. Sehingga dapat disimpulkan bahwa metode JST-SOM dapat digunakan sebagai salah satu metode pengklasifikasian suaranya manusia karena performansinya yang cukup bagus.

Kata Kunci : Mel-Frequency Cepstral Coefficient (MFCC), Linear Predictive Coding (LPC), SOM (Self Organizing Map)

Telkom
University

Abstract

Humans are able to distinguish the identity of someone they know, just from his voice. This process is known as speaker recognition. Speaker recognition or speaker recognition consists of two types, namely speaker verification and speaker identification. Voice verification to determine whether the voice is heard is the voice of someone who he claims. While the identification process of trying to determine the specific identity of the speaker.

Sound signals are complex signals. At the sound signal there are many parameters, ranging from the subjective parameters such as accent, dialect until the parameters that can be measured acoustically as formant, pitch and spectral energy. The problem that arises is how to extract the features of complex sound signals so that the data generated a new, more modest but still maintaining the distinctive character of the sound signal.

In the Final Project will be designed and realized a system that can identify the human voice, to be known from a man or a woman and then determined his age. Voice identification system consists of feature extraction and classification of sounds.

Research conducted using the acoustic parameters that are more objective, using the method of feature extraction Mel-Frequency Cepstral Coefficient (MFCC) and Linear Predictive Coding (LPC), and the results of these extraction characteristics will be amplified by the method of PCA (Principal Coefficient Analysis). Classification method used is SOM (Self Organizing Map) and the performance based on the level of accuracy.

In a system using ANN-SOM method is obtained an accuracy of 100% when testing using training data as test data, while when the noise test data is given there is some accuracy results. So it can be concluded that the ANN-SOM method can be used as one method of classifying human voice because it's a pretty good performance.

Keywords : Mel-Frequency Cepstral Coefficient (MFCC), Linear Predictive Coding (LPC), SOM (Self Organizing Map)

BAB I

PENDAHULUAN

1.1 Latar Belakang

Suara merupakan suatu bentuk lazim dari sebuah komunikasi. Perkembangan terakhir menciptakan kemungkinan bahwa suara dapat digunakan dalam sistem *security*. Dalam pengenalan suara, tujuan yang ingin dicapai adalah menggunakan sampel dari suara untuk menentukan identitas seseorang yang mengeluarkan suara tersebut diantara sejumlah orang yang berbicara. Teknik ini memungkinkan penggunaan suara seseorang untuk memverifikasi identitas dan mengendalikan akses terhadap suatu layanan tertentu misalnya panggilan suara (*voice dialing*), telepon *banking*, akses database, akses informasi, bahkan akses ke suatu tempat/ruangan tertentu.

Untuk mencapai riset pengenalan suara setidaknya diketahui terlebih dahulu bahwa suara tersebut berasal dari seorang pria atau wanita dan ditentukan usianya. Tugas akhir ini akan berusaha melakukan penelitian untuk dapat dijadikan langkah awal dari pembangunan sistem pengenalan suara yang handal. Dalam tugas akhir ini akan diteliti bagaimana membedakan suara pria dan suara wanita dengan menggunakan metode analisis MFCC dan LPC.

Sebelumnya sudah pernah dibuat Tugas Akhir yang menggunakan metode MFCC, akan tetapi metode pengklasifikasiannya yang berbeda. Untuk Tugas Akhir saya ini menggunakan metode MFCC ditambah dengan metode LPC sebagai ekstraksi ciri dan menggunakan metode SOM sebagai pengklasifikasiannya.

1.2 Rumusan Masalah

Beberapa permasalahan pada tugas akhir dapat diformulasikan sebagai berikut :

1. Bagaimana melakukan analisis mel-frequency cepstral coefficient (MFCC) dan LPC untuk mendapatkan *feature* dari suara masukan.
2. Bagaimana melakukan identifikasi dari *feature* yang ada untuk menentukan apakah itu merupakan suara pria atau wanita dan menentukan berdasarkan tingkat umurnya.

1.3 Tujuan Penelitian

Tujuan tugas akhir ini dapat dirumuskan sebagai berikut :

1. Menerapkan algoritma pengklasifikasian dalam sistem pengenalan jenis suara manusia.
2. Merancang dan merealisasikan suatu sistem pengenalan jenis suara pria dan wanita.
3. Melakukan analisis kinerja sistem dalam membedakan atau mengenali suara pria ataupun wanita.
4. Menganalisis performansi sistem berdasarkan tingkat akurasi.

1.4 Batasan Masalah

Tugas akhir ini akan membatasi permasalahan pada poin-poin berikut ini :

1. Data suara merupakan file digital dalam format *.wav dengan panjang sinyal 2 detik dan frekuensi sampling 8 kHz.
2. Data suara yang akan dianalisa dan dideteksi diambil dari orang Indonesia dan dalam kondisi berbicara secara normal.
3. Usia dibedakan menjadi tiga jenis anak-anak, remaja dan dewasa.
4. Simulasi dilakukan dengan menggunakan software Matlab 7.8

1.5 Metode Penelitian

Beberapa langkah penelitian yang dilakukan dalam tugas akhir ini adalah:

1. Studi literatur

Tahap pertama yang dilakukan adalah studi literature, tahapan ini bermaksud untuk mencari dan mengumpulkan data-data dan mendapatkan informasi yang jelas yang dapat mendukung pembuatan dasar teori yang kuat dan metode yang akan digunakan.

2. Analisis dan Desain

Tahap kedua yang dilakukan adalah analisis dan desain, pada tahapan ini dilakukan analisa untuk mendesain perangkat lunak pengenalan suara dengan suatu metode pengklasifikasian.

3. Implementasi Sistem

Tahap ketiga adalah Implementasi Sistem, meliputi pembuatan sistem yang telah dirancang dan didesain pada tahap sebelumnya. Ditahapan ini diimplementasikan perancangan yang telah dilakukan sebelumnya menjadi system sebenarnya dengan menggunakan software Matlab.

4. Pengujian dan Analisis Hasil

Tahap keempat adalah pengujian dan analisis hasil, pada tahap ini dilakukan pengujian terhadap sistem yang telah dirancang dan dibentuk, kemudian dilakukan analisis terhadap system berdasarkan output yang dihasilkan dari system tersebut. Output dari sistem ini dianalisis keakurasiannya.

5. Penyusunan Laporan

Tahap terakhir yang dilakukan adalah penyusunan laporan, disini dilakukan penyusunan laporan dari hasil penelitian yang telah dilakukan dan membuat kesimpulannya.

1.6 Sistematika Penulisan

Tugas Akhir ini disusun berdasarkan sistematika sebagai berikut :

BAB 1 : Pendahuluan

Pada bab ini akan dibahas mengenai latar belakang penelitian, perumusan masalah, tujuan penelitian, batasan masalah, metodologi penelitian, hipotesis, dan sistematika penulisan tugas akhir.

BAB 2 : Dasar Teori

Pada bab ini akan dipaparkan berbagai dasar teori yang mendukung dan mendasari penulisan tugas akhir ini.

BAB 3 : Perancangan dan Realisasi Sistem Pengenalan Suara

Pada bab ini menjelaskan proses desain dan realisasi sistem.

BAB 4 : Pengujian Sistem dan Analisis

Pada bab ini akan dilakukan pengujian sistem dan analisis hasil

yang diperoleh dari tahap perancangan dan implementasi.

BAB 5 : Kesimpulan dan Saran

Pada bab ini akan diberikan kesimpulan mengenai permasalahan yang dibahas berdasarkan serangkaian penelitian yang dilakukan. Selain itu, akan diberikan saran untuk pengembangan selanjutnya.



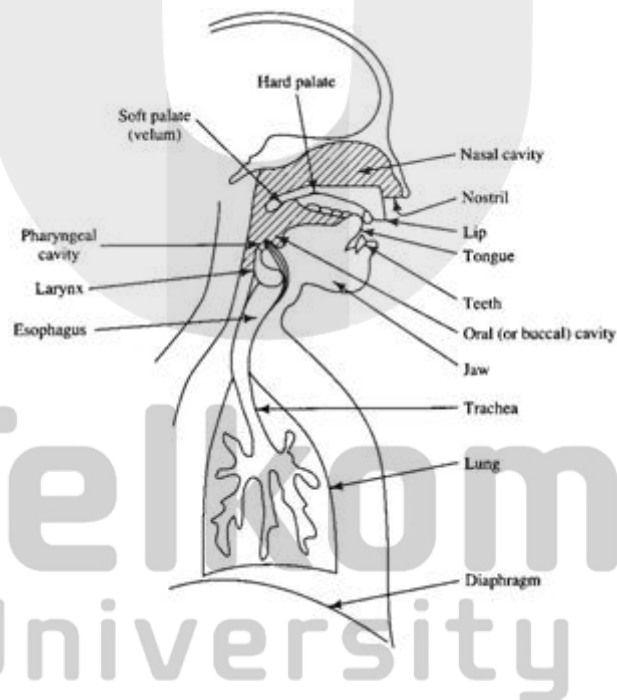
BAB II

DASAR TEORI

2.1 Proses Produksi Suara

Proses produksi suara pada manusia dapat dibagi menjadi tiga buah proses fisiologis, yaitu : pembentukan aliran udara dari paru-paru, perubahan aliran udara dari paru-paru menjadi suara, baik *voiced*, maupun *unvoiced* yang dikenal dengan istilah phonation, dan artikulasi yaitu proses modulasi/ pengaturan suara menjadi bunyi yang spesifik.

Organ tubuh yang terlibat pada proses produksi suara adalah : paru-paru, tenggorokan (trachea), laring (larynx), faring (pharynx), pita suara (vocal cord), rongga mulut (oral cavity), rongga hidung (nasal cavity), lidah (tongue), dan bibir (lips), Gambar 2.1 menunjukkan organ produksi suara pada manusia.



Gambar 2.1. Organ Produksi Suara Pada Manusia

Organ tubuh ini dapat dikelompokkan menjadi tiga bagian utama, yaitu : *vocal tract* (berawal di awal bukaan pita suara atau glottis, dan berakhir di bibir), *nasal tract* (dari velum sampai nostril), dan *source generator* (terdiri dari paru-paru, tenggorokan, dan larynx). Ukuran *vocal tract* bervariasi untuk setiap individu, namun untuk laki-laki dewasa rata-rata panjangnya sekitar 17 cm. Luas dari *vocal tract* juga bervariasi antara 0 (ketika

seluruhnya tertutup) hingga sekitar 20 cm^2 . Ketika velum, organ yang memiliki fungsi sebagai pintu penghubung antara *vocal tract* dengan nasal *tract*, terbuka, maka secara akustik nasal *tract* akan bergandengan dengan *vocal tract* untuk menghasilkan suara nasal.

Aliran udara yang dihasilkan dorongan otot paru-paru bersifat konstan. Ketika pita suara dalam keadaan berkontraksi, aliran udara yang lewat membuatnya bergetar. Aliran udara tersebut dipotong-potong oleh gerakan pita suara menjadi sinyal pulsa yang bersifat quasi-periodik. Sinyal pulsa tersebut kemudian mengalami modulasi frekuensi ketika melewati *pharynx*, rongga mulut ataupun pada rongga hidung. Sinyal suara yang dihasilkan pada proses ini dinamakan sinyal *voiced*. Namun, apabila pita suara dalam keadaan relaksasi, maka aliran udara akan berusaha melewati celah sempit pada permulaan *vocal tract* sehingga alirannya menjadi turbulen, proses ini akan menghasilkan sinyal *unvoiced*. Ketika sumber suara melalui *vocal tract*, kandungan frekuensinya mengalami modulasi sehingga terjadi resonansi pada *vocal tract* yang disebut formants. Apabila sinyal suara yang dihasilkan adalah sinyal *voiced*, terutama vokal, maka pada selang waktu yang singkat bentuk *vocal tract* relative konstan (berubah secara lambat) sehingga bentuk *vocal tract* dapat diperkirakan dari bentuk spektral sinyal *voiced*.

2.2 Prinsip Pengenalan Suara

Metode pengenalan suara dapat di bagi menjadi *text-independent* dan *text-dependent*. Pada metode *text-independent*, sistem menangkap karakteristik suara pembicaraan seseorang tanpa mempedulikan apa yang dibicarakan oleh orang tersebut. Pada metode *text-dependen*, sistem menangkap karakteristik suara berdasarkan apa yang dikatakan orang tersebut. Semua teknologi pengenalan suara baik *text-independent* dan *text dependent* memiliki masing-masing kelebihan dan kekurangan. Pemilihan teknologi yang digunakan akan sangat tergantung dari aplikasi apa yang ditangani. Pada level yang paling tinggi, secara umum semua teknologi pengenalan suara akan terdiri dari dua bagian besar berupa modul *feature extraction* dan modul *feature matching* atau *classification*.

Tugas akhir ini merupakan sistem *text-independent* karena hanya akan menentukan suara pria atau wanita dari input suara yang diberikan.

2.3 Mel Frequency Cepstral Coefficient ^[2]

Mel frequency Cepstral Coefficient adalah sebuah koefisien audio, dalam beberapa studi menunjukkan bahwa persepsi pendengaran manusia tidaklah dalam skala yang linier

dalam bentuk frekuensi melainkan diukur dalam bentuk skala mel-frekuensi. Sehingga *Mel-Scale* merupakan skala yang diambil berdasarkan pendekatan terhadap pendengaran manusia. Dimana penggunaannya adalah untuk mengestrak data sinyal suara. Tahapan proses dalam *Mel Frequency Cepstral Coefficient* adalah;

1. *Frame Blocking*

Dalam proses *frame blocking* sinyal suara yang masuk akan diblok menjadi frame-frame dengan jumlah N sampel dan jarak antar sampel dengan spasi M dimana $M < N$. Nilai dari N biasanya adalah 256 dengan spasi $M=100$

2. *Windowing*

Proses *windowing* dilakukan pada setiap frame dengan tujuan untuk meminimumkan diskontinuitas antar dua frame yang adjacent, khususnya pada bagian awal dan akhir.

3. FFT (*Fast Fourier Transform*)

Tahapan berikutnya adalah FFT, disini FFT akan mengubah masing-masing frame N sampel dari domain waktu menjadi domain frekuensi.

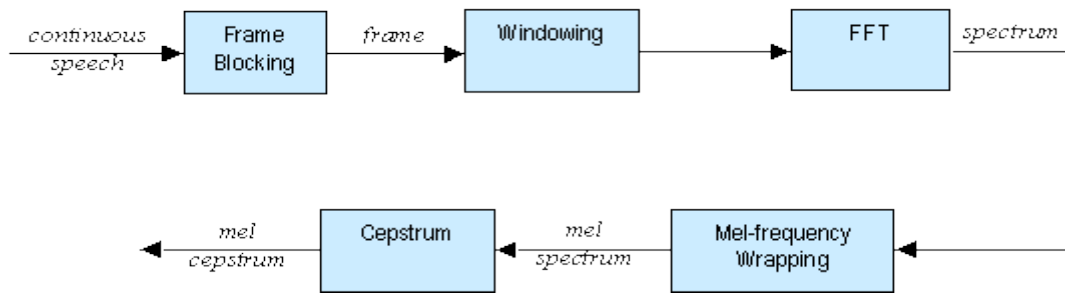
4. *Wrapping*

Skala mel-frekuensi adalah frekuensi linier berada dibawah 1000 Hz dan bentuk logaritmik berada diatas 1000 Hz. Pendekatan persamaan untuk mengitung mel dalam frekuensi f (Hz) adalah;

$$\text{Mel} (f) = 2595 \times \text{Log}_{10}(1+ f / 700) \dots\dots\dots (2.1)$$

5. *Cepstrum*

Langkah terakhir adalah mengubah spektrum log-mel menjadi domain waktu. Hasil dari proses inilah yang disebut dengan *Mel Frequency Cepstral Coefficient*. Representasi *cepstral* dari spektrum sinyal suara berasal dari sifat spektrum sinyal untuk analisa frame yang ada.



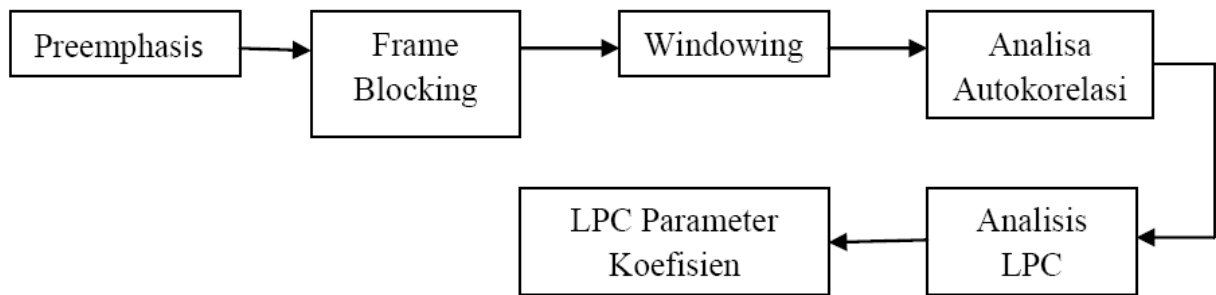
Gambar 2.2 Block Diagram MFCC

Mel-Frequency Cepstrum Coefficients (MFCC) merupakan salah satu teknik ekstraksi ciri yang telah luas dipakai pada pemrosesan sinyal suara, terutama pada pemrosesan pengenalan pembicara.. Sejak diperkenalkan oleh Davis and Mermelstein pada tahun 1980, maka teknik MFCC ini telah mengalami berbagai pengembangan dan variasi. Bagian utama yang biasanya mengalami modifikasi adalah pada pengembangan filter, seperti jumlah, bentuk, cara membuat sekatan, serta lebar filter yang dibuat. Sedangkan dari sisi masukan tetap sama, yaitu spektrum energi dari input yang diperoleh melalui transformasi Fourier.

2.4 Linear Predictive Coding

Analisis sinyal adalah kegiatan melakukan ekstraksi terhadap informasi yang terdapat di suatu sinyal. Linear Predictive Coding (LPC) merupakan salah satu teknik analisis sinyal percakapan yang paling powerful dan menyediakan ekstraksi fitur yang berkualitas baik dan efisien untuk digunakan dalam perhitungan. LPC pertama kali digunakan pada tahun 1978 untuk membuat alat sintesis sinyal percakapan.

LPC melakukan analisis dengan cara memperkirakan formant, memisahkan formant dari sinyal, yang dinamakan proses inverse filtering, lalu mengestimasi intensitas dan frekuensi dari sinyal percakapan yang tersisa, yang disebut residue.¹⁶ Karena sinyal percakapan bervariasi seiring waktu, estimasi tersebut dilakukan untuk setiap potongan kecil dari sinyal, yang dinamakan frame. Prosedur untuk mendapatkan koefisien LPC diperlihatkan pada gambar dibawah.



Gambar 2.3 Block Diagram LPC

Gambar diatas menunjukkan blok diagram LPC yang biasa digunakan dalam sistem pengenalan suara manusia. Langkah-langkah dasar yang harus dilakukan mengenai blok diagram tersebut adalah

1. *Preemphasis* : Proses dimana sinyal / speech ucapan sirubah menjadi sinyal
2. *Blocking into frames* : Pada tahap ini sinyal yang telah di preemphasis, diblok menjadi beberapa bagian dengan jumlah sample N , dan tiap bagian dipisahkan dengan sejumlah M sample.
3. *Frame Windowing* : Tahap berikutnya adalah melakukan proses *window* pada setiap bagian sinyal yang telah dibuat sebelumnya. Hal ini dilakukan untuk meminimalkan pada bagian awal dan akhir sinyal. Jika definisikan sebuah *window* $w(n)$ dan sinyal tiap bagian adalah $x(n)$ maka sinyal hasil proses *windowing*.
4. *Auto Correlation Analisis* : Tiap bagian yang telah diberi *window* kemudian akan dibentuk autokorelasinya.
5. *Analisa LPC* : Langkah berikutnya adalah analisa LPC dimana semua nilai autokorelasinya yang telah dihitung pada tahap sebelumnya akan diubah menjadi parameter LPC
6. *Pengubahan parameter LPC menjadi koefisien cepstral* : Parameter LPC yang sangat penting yang bisa diturunkan dari koefisien LPC adalah koefisien *cepstral* LPC, $c(m)$.

Adapun langkah-langkah analisis LPC untuk pengenalan suara adalah sebagai berikut.

1. *Preemphasis* terhadap cuplikan sinyal dengan persamaan *preemphasizer*:

$$s(n) = s(n) - as(n-1) \tag{2.2}$$

dengan $s(n)$ adalah sampel ke- n dan harga a yang paling sering digunakan ialah 0.95.

- Membagi hasil preemphasis $s(n)$ ke dalam *frame-frame* yang masing-masing memuat N buah sampel yang dipisahkan sejauh M buah sample. Semakin $M < N$ semakin baik perkiraan spektral LPC dari *frame* ke *frame*.
- Melakukan *windowing* terhadap setiap *frame* yang telah dibentuk untuk meminimalkan diskontinuitas pada ujung awal dan ujung akhir setiap *frame* dengan persamaan *Hamming Window* untuk sampel ke- n adalah:

$$W(n) = 0.54 - 0.46 \cos(2\pi n / N - 1), \quad 0 \leq n \leq N-1 \quad (2.3)$$

- Analisis autokorelasi terhadap setiap *frame* hasil *windowing* $x_1(n)$ dengan persamaan:

$$r_1(m) = \sum_{n=1-m}^{N-1-m} x_1(n)x_1(n+m) \quad (2.4)$$

dengan m dimulai dari 0 dan nilai tertinggi dari $m = p$ adalah orde LPC yang biasa bernilai 8 - 16

- Mengubah $p + 1$ buah hasil autokorelasi pada masing-masing *frame* menjadi koefisien LPC $a_m = a_m(p)$ untuk $m = 1, 2, \dots, p$ dengan persamaan dibawah ini:

$$E(0) = r(0) \quad (2.5)$$

$$k_m = \{r(m) - \sum_{j=1}^{m-1} a_j(m-1)r(|m-j|) / E(m-1)\}, \quad 1 \leq m \leq p$$

$$a_m(m) = k_m$$

$$a_j(m) = a_j(m-1) - k_m a_m - j(m-1), \quad 1 \leq j \leq m-1$$

$$E(m) = (1 - k_m^2) E(m-1)$$

dengan $r(0)$ adalah hasil autokorelasi, $E(m)$ adalah *error*, k_m adalah koefisien pantulan, $a_j(m)$ adalah koefisien prediksi untuk $1 \leq j \leq m$.

- Mengubah parameter LPC a_m ke koefisien cepstral c_m untuk mendapatkan kinerja yang lebih baik dan tahan terhadap *noise*, yaitu dengan persamaan:

$$c_m = a_m + \sum_{k=1}^{m-1} (k/m) c_k a_{m-k}, \quad 1 \leq m \leq p \quad (2.6)$$

$$c_m = \sum_{k=1}^{m-1} (k/m) c_k a_{m-k}, \quad m > p$$

koefisien *cepstral* ini adalah koefisien dari representasi pada *spectrum logaritmis*.

2.5 Klasifikasi ^[2]

Dengan dihasilkannya vector akustik dari suatu suara masukan, kita kemudian membangun database ciri untuk dua jenis suara masukan yang berbeda beserta masing-masing vector akustiknya. Semua vektor akustik suara pria dikumpulkan sebagai satu kelompok, demikian juga semua vektor akustik dari suara wanita. Kedua kelompok vektor akustik

tersebut membentuk database yang akan digunakan untuk mengklasifikasikan suatu masukan baru yang tidak diketahui apakah merupakan suara pria atau wanita.

2.6 Topologi JST-SOM (Fungsi Topologi) ^[3]

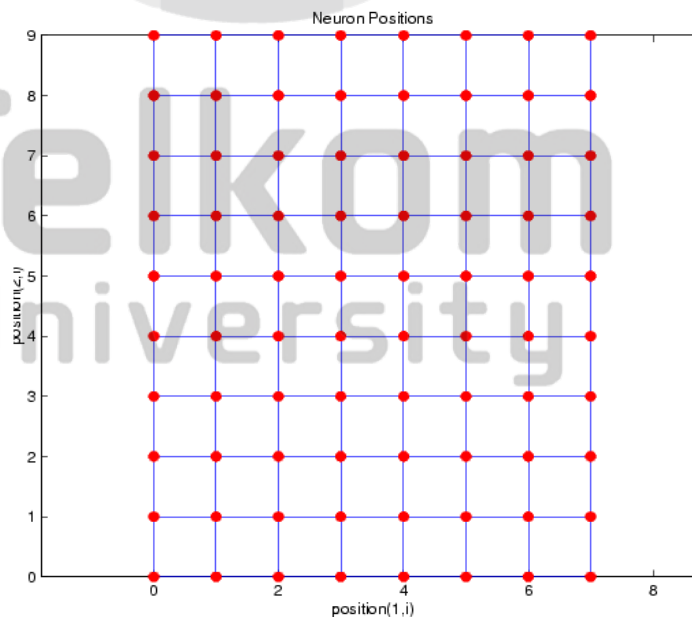
Algoritma *Self Organizing Map* (SOM) merupakan suatu metode jaringan saraf tiruan yang diperkenalkan oleh Professor Teuvo Kohonen pada tahun 1980an. SOM merupakan salah satu bentuk topologi dari *Unsupervised Artificial Neural Network* (*Unsupervised ANN*) dimana dalam proses pelatihannya tidak memerlukan pengawasan (*target output*). SOM digunakan untuk mengelompokkan (*clustering*) data berdasarkan karakteristik / fitur-fitur data .

Dalam jaringan SOM, neuron target tidak diletakkan dalam sebuah baris seperti layaknya model jaringan syaraf tiruan yang lain. Neuron target diletakkan dalam 2 dimensi yang bentuk/topologinya dapat diatur. Dalam Matlab disediakan fasilitas untuk mendefinisikan topologi jaringan yang berbeda-beda. Topologi yang berbeda akan menghasilkan neuron sekitar neuron pemenang yang berbeda sehingga bobot yang dihasilkan juga akan berbeda.

Ada 3 macam topologi yang dapat dibentuk yaitu:

1. Gridtop

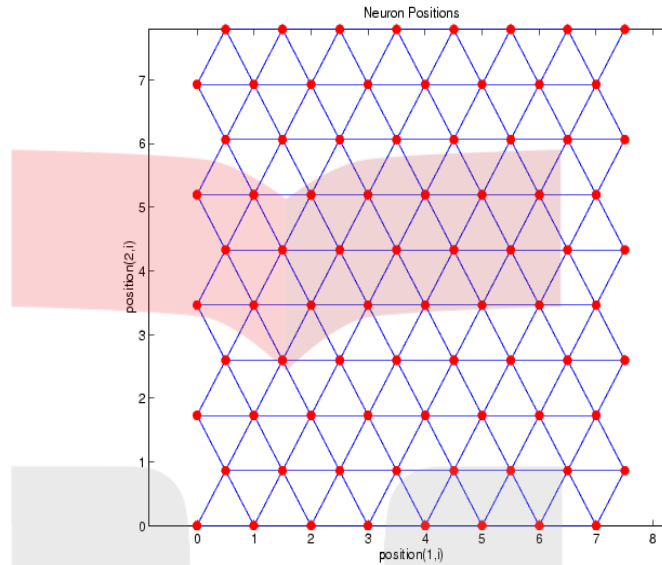
Dalam gridtop neuron disusun dalam array 2 dimensi dengan bentuk persegi.



Gambar 2.4 Topologi Gridtop

2. Hextop

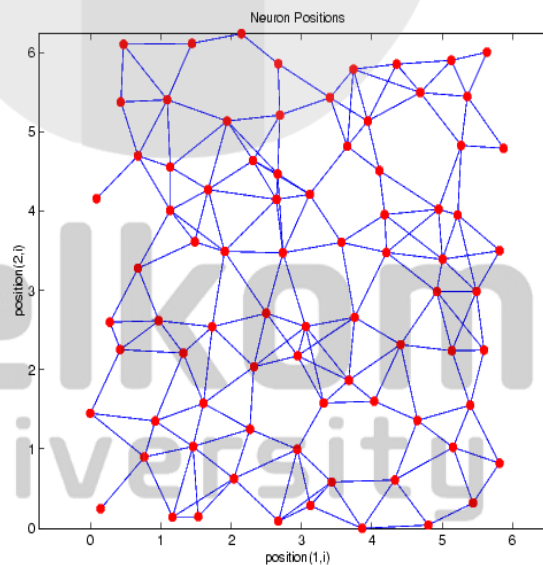
Dalam hextop neuron disusun dalam array 2 dimensi dengan bentuk heksagonal.



Gambar 2.5 Topologi Hextop

3. Randtop

Dalam randtop neuron disusun dalam array 2 dimensi dengan susunan acak.



Gambar 2.6 Topologi Randtop

2.6.1 Jarak Antar Neuron (Fungsi Jarak) ^[6]

Dalam SOM, perubahan bobot tidak hanya dilakukan pada bobot garis yang terhubung ke neuron pemenang saja, tetapi juga pada bobot garis ke neuron-neuron

di sekitarnya. Neuron di sekitar neuron pemenang ditentukan berdasarkan jaraknya dari neuron pemenang. Dalam fungsi Matlab, default neuron tetangga=1.

Ada 4 macam definisi jarak antara 2 neuron yaitu:

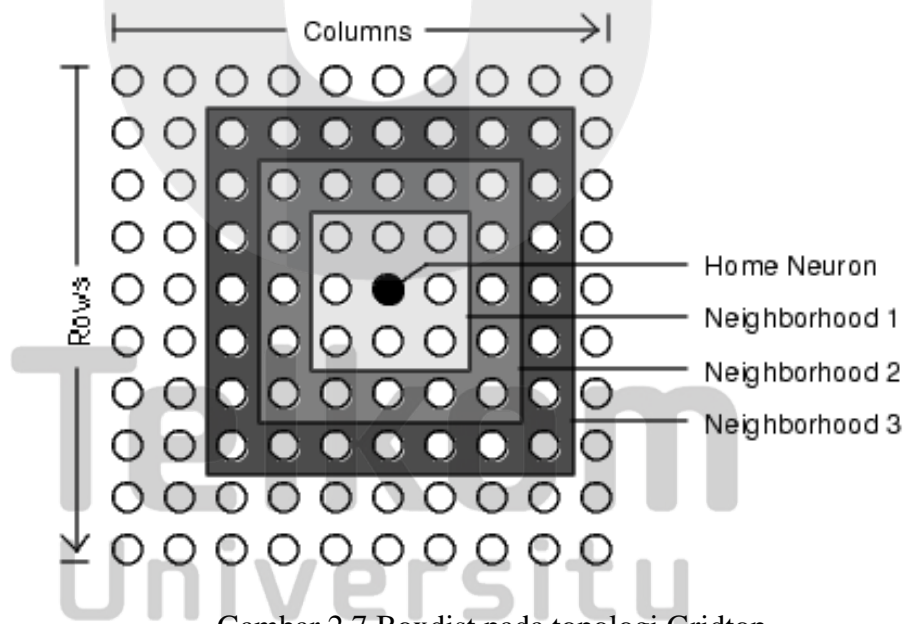
1. Jarak Euclidist (dist)

Jarak Euclidist adalah jarak antara 2 titik dalam posisi berbeda yang biasa kita kenal. Missal (x1,y1) dan (x2,y2) adalah koordinat 2 neuron, maka jarak kedua neuron tersebut didefinisikan sebagai:

$$d = \sqrt{(x2 - x1)^2 + (y2 - y1)^2} \dots\dots\dots(2.7)$$

2. Jarak persegi (boxdist)

Jarak persegi adalah jarak langsung antara neuron pemenang dengan neuron tetangganya secara langsung. Sebagai contoh apabila topologi neuron adalah gridtop, maka paling banyak terdapat 8 buah neuron tetangga dengan boxdist=1, dan terdapat paling banyak 16 tetangga jika boxdist=2. Jika neuron pemenang terletak di pinggir maka hanya ada 5 neuron tetangga dengan boxdist=1.

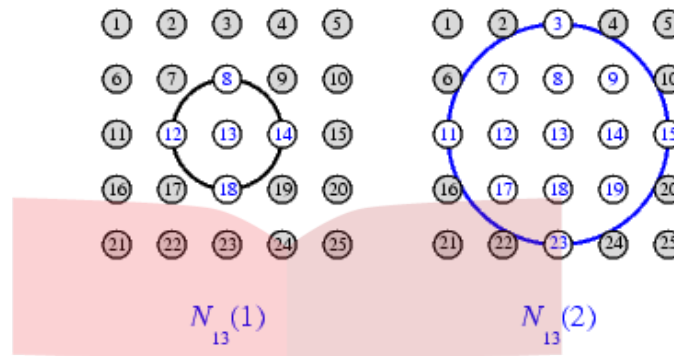


Gambar 2.7 Boxdist pada topologi Gridtop

3. Jarak link (linkdist)

Jarak link adalah jumlah langkah yang dibutuhkan untuk menuju neuron tersebut. Jika dalam jaringan SOM menggunakan topologi gridtop dengan linkdist=1, berarti hanya neuron-neuron yang berhubungan langsung dengan

neuron pemenang saja yang diubah bobotnya (jumlah neuron tetangga=4).
 Jika linkdist=2 maka jumlah neuron tetangga=12.



Gambar 2.8 Linkdist pada topologi Gridtop

4. Jarak Manhattan (mandist)

Jarak Manhattan antara vektor $x(x_1, x_2, \dots, x_n)$ dan vektor $y(y_1, y_2, \dots, y_n)$ didefinisikan sebagai:

$$sum(abs(x - y)) = \sum_{i=1}^n |x_i - y_i| \tag{2.8}$$

Jika $x=(x_1, x_2)$ dan $y=(y_1, y_2)$ menyatakan koordinat neuron yang dibentuk melalui topologi tertentu, maka jarak Manhattan antara neuron x dan y adalah:

$$D = |x_1 - y_1| + |x_2 - y_2| \tag{2.9}$$

2.7 **Principal Component Analysis (PCA)**

PCA merupakan sebuah metode untuk menemukan pola dari data berdimensi tinggi. Melalui PCA, data yang berdimensi tinggi dapat direduksi menjadi data berdimensi lebih rendah, namun dengan memilih fitur-fitur penting dari data yang akan diolah sehingga informasi penting dapat diambil dan informasi yang tidak penting dapat dibuang.

PCA akan bekerja melalui beberapa langkah :

a. Normalisasi Data

Mula-mula terdapat sekumpulan N data latih $X = [x_1, x_2, x_3, \dots, x_N]$ berukuran $m \times n$ maka representasi masing-masing data latih diubah menjadi sebuah vector atau matriks kolom berordo $mn \times 1$ yakni dengan menggabungkan masing-masing elemen baris secara terurut dalam satu kolom.

Setelah masing-masing data latih x_i direpresentasikan sebagai vector Γ_i , hitung rata-rata Ψ :

$$\Psi: \Psi = \frac{1}{N} \sum_{i=1}^N \Gamma_i \quad (2.10)$$

Langkah selanjutnya adalah mengurangi masing-masing vector data latih dengan vector data rata-rata.

$$\Phi_i = \Gamma_i - \Psi \quad (2.11)$$

Kemudian gabung masing-masing Φ_i menjadi sebuah matriks $A = [\Phi_{i1} \Phi_{i2} \Phi_{i3} \dots \Phi_{in}]$ berordo $m \times n$.

b. Mencari Matriks Kovarian

Kovarian merupakan sebuah ukuran yang menggambarkan hubungan antara dua variable. Matriks kovarian C dihitung melalui persamaan :

$$C = \frac{1}{M} \sum_{n=1}^N \Phi_n \Phi_n^T = AA^T \quad (2.12)$$

c. Mencari Vektor Eigen dan Nilai Eigen dari Matriks Kovarian

Untuk langkah selanjutnya akan dicari vector eigen dari matriks kovarian dan vector eigen dari AA^T :

$$Cv = \lambda v \quad (2.13)$$

Skalar λ dinamakan nilai eigen (*eigenvalue*) dari C dan v dikatakan vektor eigen yang bersesuaian dengan λ . Selanjutnya matriks kumpulan vektor eigen dinyatakan dalam V, dimana setiap kolom V adalah vektor eigen. Sedangkan matriks kumpulan nilai eigen yang bersesuaian dengan masing-masing kolom V dinyatakan dalam matriks kolom E, dimana setiap elemennya merupakan nilai eigen yang bersesuaian dengan setiap kolom V.

d. Pemetaan Data ke Ruang Eigen

Vektor eigen yang diperoleh dari proses sebelumnya dapat diseleksi sesuai dengan kebutuhan pengguna. Penyeleksian dilakukan berdasarkan nilai eigen dari vektor eigen yang berkorelasi. Semakin besar nilai eigen, maka semakin besar hubungan antara vektor eigen yang berkorelasi dengan dimensi lama. Hal ini menyebabkan vektor eigen yang dipilih ialah yang memiliki nilai eigen besar. Jumlah vektor eigen yang dipilih dapat disesuaikan dengan kebutuhan pengguna.

Vektor Eigen yang dipilih ini dinamakan *principal component* yang akan digunakan untuk memetakan data.

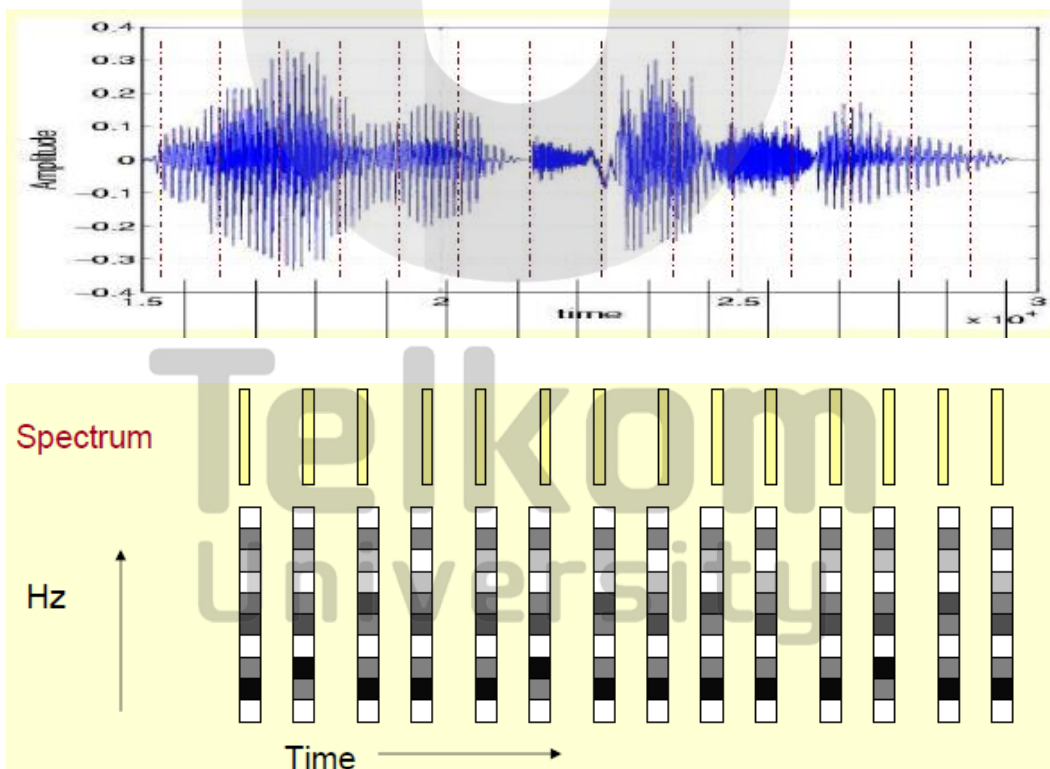
$$y = U^T D \tag{2.14}$$

Dimana,

- y : Data dengan dimensi baru (yang direduksi)
- U : Vektor eigen yang dipilih (*principal component*)
- D : Data hasil normalisasi

2.8 Spectrogram ^[2]

Untuk akusisi data suara yang lebih baik digunakan teknik *Spectrogram*. *Spectrogram* ini digunakan untuk menganalisis data suara. Format data suara yang akan dianalisis yang paling umum adalah berbentuk grafik dengan dua dimensi geometris yaitu sumbu horizontal mewakili waktu dan sumbu vertical mewakili frekuensi. Selain dua dimensi tersebut biasanya terdapat dimensi ketiga yaitu amplitude frekuensi tertentu pada waktu tertentu yang dimana diwakili dengan intensitas atau warna setiap titik digambar.



Gambar 2.9 Spectrogram

Kegunaan dari Spectrogram:

- Representasi waktu dan frekuensi dari sinyal suara
- Sebuah alat atau *tool* untuk mempelajari suara (*speech sound*)
- Spectrogram dapat digunakan untuk menganalisa sinyal keluaran dari prosesor sinyal seperti filter untuk melihat performa
- Berguna untuk evaluasi proses perubahan sistem teks ke suara
- Membantu mengatasi kerusakan data suara

2.9 Additive White Gaussian Noise (AWGN)

AWGN (Additive White Gaussian Noise) merupakan suatu proses stokastik yang terjadi pada kanal dengan karakteristik memiliki rapat daya spectral noise merata di sepanjang range frekuensi.

AWGN mempunyai karakteristik respon frekuensi yang sama disepanjang frekuensi dan variannya sama dengan satu.

Pada kanal transmisi selalu terdapat penambahan derau yang timbul karena akumulasi derau termal dari perangkat pemancar, kanal transmisi, dan perangkat penerima. Derau yang menyertai sinyal pada sisi penerima dapat didekati dengan model matematis statistik AWGN.

Derau AWGN merupakan gangguan yang bersifat Additive atau ditambahkan terhadap sinyal transmisi, dimodelkan dalam pola distribusi acak Gaussian dengan mean (m) = 0, standar deviasi (σ) = 1, power spectral density (pdf) = $N_0/2$ (W/Hz), dan mempunyai rapat spektral daya yang tersebar merata pada lebar pita frekuensi tak berhingga. Distribusi AWGN dengan pdf :

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-(x-m)^2 / 2\sigma^2\right]$$

dimana: $p(x)$ = probabilitas kemunculan derau

σ = standar deviasi

m = rata-rata (mean)

x = variable (tegangan atau daya sinyal)