BAB 2

LANDASAN TEORI

2.1 Latar Belakang Pengenalan Ucapan

Pengenalan ucapan atau pengenalan wicara dalam istilah bahasa inggirisnya *automatic speech recognition (ASR)* adalah suatu pengembangan teknik dan sistem yang memungkinkan komputer untuk menerima masukan berupa kata yang diucapkan. Teknologi ini memungkinkan suatu perangkat untuk mengenali dan memahami kata – kata yang diucapkan dengan cara digitalisasi kata dan mencocokkan sinyal digital tersebut dengan suatu pola tertentu yang tersimpan dalam suatu perangkat. Kata – kata yang diucapkan diubah bentuknya menjadi sinyal digital dengan cara mengubah gelombang suara menjadi sekumpulan angka yang kemudian disesuaikan dengan kode – kode tertentu untuk mengidentifikasi kata – kata tersebut, Hasil dari identifikasi kata yang diucapkan dapat ditampilkan dalam bentuk tulisan atau dapat dibaca oleh perangkat teknologi sebagai sebuah komando untuk melakukan suatu pekerjaan. Misalnya penekanan tombol pada telelpon genggam yang dilakukan secara otomatis dengan komando suara Lestary (2009).

Perkembangan teknologi dalam bidang speech recognition bertujuan untuk mewujudkan keinginan manusia dalam memaksimalkan fungsi PC sebagai alat yang mampu mempermudah pekerjaan manusia disegala aspek. Hal yang hendak dicapai adalah menciptakan PC yang mampu berinteraksi dengan manusia secara langsung menggunakan bahasa manusia sehari – hari sesuai tata bahasa yang berlaku, studi tentang pengenalan ucapan sudah dilakukan selama bertahun – tahun untuk mencapai sukses yang ideal. Tetapi hal tersebut belum juga dapat terpenuhi sampai saat ini. Masih perlu dilakukan penelitian dan peningkatan lebih lanjut terhadap metode pengenalan yang sudah ada.

Secara umum , proses pengenalan ucapan dimulai dengan meng-input-kan suara melalui microphone, sinyal suara yang menjadi input bersifat continue, untuk itu diperlukan pemrosesan awal (pre-processing) untuk mengubah sinyal tersebut menjadi discrete agar dapat diproses oleh komputer. Setelah itu sinyal tersebut akan

melalui proses ekstraksi ciri (*feature extraction)* untuk mendapatkan parameter khusus yang menjadi bahan pembanding dalam proses pencocokan pola. Pada tahap selanjutnya yaitu pencocokan pola, maka program akan membandingkan sinyal ucapan masukan dengan sinyal pembanding lalu program menentukan keputusan. Tahapan dalam speech recognition dapat dilihat pada gambar 2.1 Rachman (2006).

Sumber : Rachman (2006)

Gambar 2.1. Tahapan dalam *Speech Recognition.*

2.2 Suara Musik

Kehadiran musik sebagai bagian dari kehidupan manusia bukanlah hal yang baru. Setiap budaya di dunia memiliki musik yang khusus diperdengarkan atau dimainkan berdasarkan peristiwa – peristiwa bersejarah dalam perjalanan hidup anggota masyarakatnya. Ada musik yang dimainkan untuk mengungkapkan rasa syukur atas kelahiran seorang anak. Ada juga musik yang khusus mengiringi upacara – upacara tertentu seperti pernikahan dan kematian. Musik juga menjadi pendukung utama untuk melengkapi dan menyempurnakan beragam bentuk kesenian dan berbagai budaya.

Musik adalah suara yang disusun demikian rupa sehingga mengandung irama. Lagu, dan keharmonisan terutama suara yang dihasilkan dari alat- alat yang dapat menghasilkan bunyi – bunyian, walaupun musik adalah sejenis fenomena intuisi, untuk mencipta, memperbaiki dan mempersembahkannya adalah suatu bentuk seni , mendengar musik adalah sejenis hiburan. Musik adalah sebuah fenomena yang sangat unik yang bisa dihasilkan oleh beberapa alat musik.

2.3 Pengolahan Audio

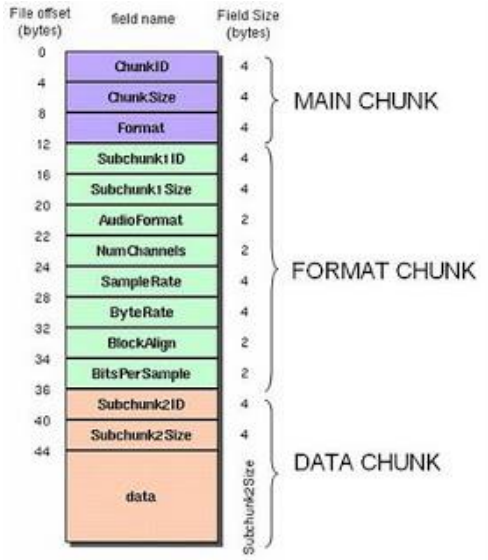
Suara adalah sebuah sinyal yang merambat melalui perantara, suara dapat dihantarkan dengan media air, udara, maupun benda padat, Dengan kata lain, suara adalah gelombang yang merambat dengan frekuensi dan amplitudo tertentu, suara yang dapat

didengar manusia berkisar antara 20 Hz sampai dengan 20 KHz, dimana Hz adalah satuan frekuensi yang artinya banyak getaran per-detik (cps / cycle per second) Rabiner (1993).

Perlengkapan produksi suara pada piano konvensional terdapat palu pemukul (*Hammer*),tuts (*Keys*), senar piano (*Strings*). Secara garis besar terdiri atas tuts ditekan oleh manusia, tuts ditekan menggerakan palu pemukul yang selanjutnya akan memukul strings yang ada. Setiap tuts berbeda bunyinya.

Format file ".WAV" merupakan bagian dari spesifikasi RIFF milik Microsoft yang digunakan untuk penyimpanan file-file mulitimedia. File wav dimulai dengan bagian header dan diikuti oleh rentetan data chunk. File wav terdiri dari 3 bagian, yaitu main chunk, format chunk, dan data chunk.

Sinyal suara yang direpresentasikan file WAV dalam bentuk discrete, berupa deret bilangan yang merepresentasikan amplitude dalam domain waktu. Pada bagian file header terdapat informasi tentang file WAV tersebut, diantaranya menyatakan nilai sample rate, jumlah channel, dan bit per sample. Dari keterangan pada file header tersebut dapat diketahui berapa sampel yang dicuplik dari sinyal analog tiap detik Wilson (2003). Struktur WAV dapat dilihatp ada gambar 2.2 :



Sumber: Wilson(2003)

Gambar 2.2 Struktur WAV

2.4 Mel Frequency Cepstrum Coefficient (MFCC)

Mel Frequency Cepstrum Coefficient (MFCC) merupakan salah satu metode yang banyak digunakan dalam bidang speech recognition. Metode ini digunakan untuk melakukan feature extraction, sebuah proses yang mengkonversikan sinyal suara menjadi beberapa parameter. Keunggulan dari metode ini adalah :

- Mampu menangkap karakteristik suara yang sangat penting bagi pengenalan suara atau dengan kata lain mampu menangkap informasi – informasi penting yang terkandung dalam sinyal suara
- Menghasilkan data seminimal mungkin tanpa menghilangkan informasi – informasi penting yang ada.
- Mereplikasi organ pendengaran manusia dalam melakukan persepsi sinyal suara.

Perhitungan yang dilakukan dalam MFCC menggunakan dasar dasar perhitungan short-term analysis. Hal ini dilakukan mengingat sinyal suara yang bersifat quasi stationary. Pengujian yang dilakukan untuk periode waktu yang cukup pendek (sekitar 10 sampai 30 milidetik) akan menunjukkan karakteristik sinyal suara yang stationary. Tetapi bila dilakukan dalam periode waktu yang lebih panjang, karakteristik sinyal suara akan berubah sesuai dengan kata yang diucapkan.

MFCC feature extraction sebenarnya merupakan adaptasi dari sistem pendengaran manusia, dimana sinyal suara akan di-filter secara linear untuk frekuensi rendah (dibawah 1000Hz) dan secara logaritmik untuk frekuensi tinggi (diatas 1000Hz), berikut blok diagram untuk MFCC Manunggal(2005).

2.4.1 DC- Removal

Remove DC Components bertujuan untuk menghitung rata-rata dari data sampel suara, dan mengurangkan nilai setiap sampel suara dengan nilai rata-rata tersebut. Tujuannya adalah mendapat normalisasi dari data suara input Putra(2011).

$$y[n] = x[n] - \pi\,, 0 \leq n \leq N-1 \text{.....(1)}$$

dimana : $y[n]$ = sampel signal hasil proses DC removal

$x[n]$ = sampel signal asli

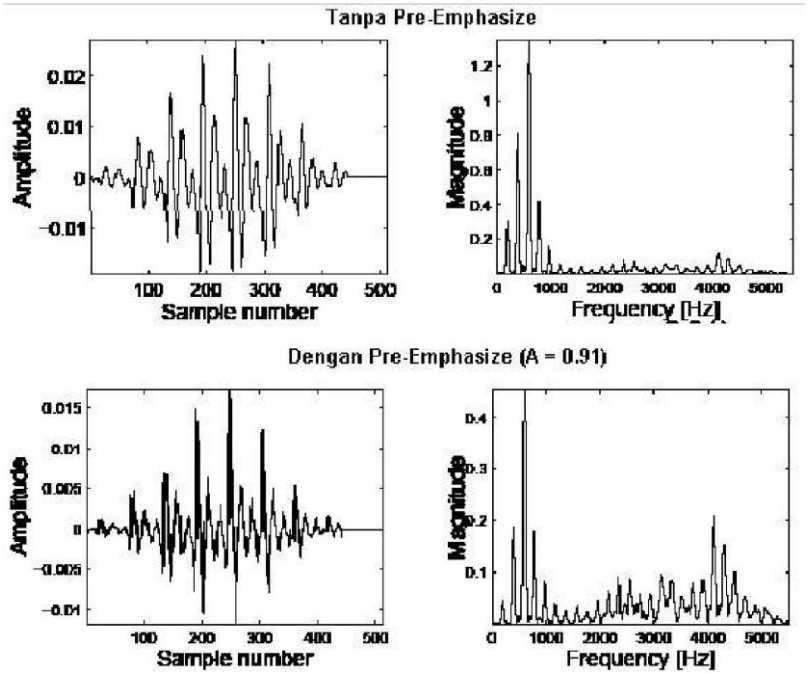
π = nilai rata-rata sampel signal asli

N = panjang signal

2.4.2 Pre – Emphasize Filtering

Pre – empahsize filtering merupakan salah satu jenis filter yang sering digunakan sebelum sebuah signal diproses lebih lanjut, Filter ini mempertahankan frekuensi – frekuensi tinggi pada sebuah spektrum, yang umumnya tereleminasi pada saat proses produksi suara. Tujuan dari Pre – emphasize Filtering ini adlaah (Manunggal , 2005)

a. Mengurangi noise ratio pada *signal*, sehingga dapat meningkatkan kualitas *signal*
b. Menyeimbangkan spektrum dari *voice sound*.



Sumber : Manunggal, 2005

Gambar 2.3 Contoh dari Pre-Emphasize pada sebuah frame

Pada gambar diatas terlihat bahwa distribusi energi pada setiap frekuensi terlihat lebih seimbang setelah diimplementasikan *pre-emphasize filter*. Bentuk yang paling umum digunakan dalam *pre-emphasize filter* adalah sebagai berikut.

$$y[n] = s[n] - \alpha\, s[n-1]\,, 0.9 \leq \alpha \leq 1.0 \ldots\ldots\ldots (2)$$

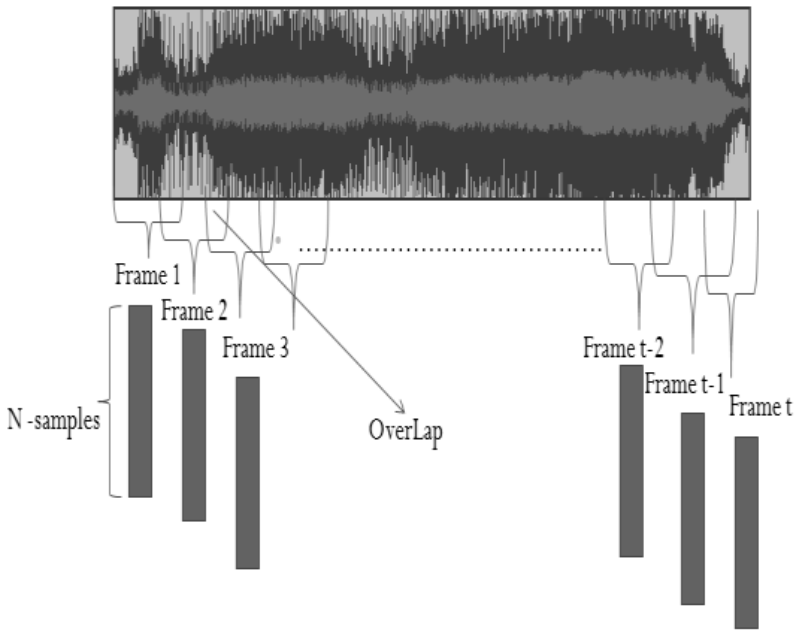
dimana :

$y[n]$ = *signal* hasil *pre-emphasize filter*

$s[n]$ = *signal* sebelum *pre-emphasize filter*

2.4.3 Frame Blocking

Karena sinyal suara terus mengalami perubahan akibat adanya pergeseran artikulasi dari organ produksi suara, sinyal harus diproses secara short segments (short frame). Panjang frame yang biasanya digunakan untuk pemrosesan sinyal adalah antara 10-30 milidetik. Panjang frame yang digunakan, sangat mempengaruhi keberhasilan dalam analisa spektral, di satu sisi, ukuran frame harus diperpanjang sepanjang mungkin untuk dapat menunjukkan resolusi frekuensi yang baik, tetapi dalam slain sisi, ukuran frame juga harus cukup pendek untuk dapat menunjukkan resolusi waktu yang baik Ridwan(2011).



Sumber : Ridwan(2011)

Gambar 2.4 Contoh Frame Blocking

Proses frame yang dilakukan terus sampai seluruh sinyal dapat terproses. Selain itu, proses ini umumnya dilakukan secara overlapping untuk setiap frame-nya.

Panjang daerah overlap yang umum digunakan adalah kurang lebih 30% sampai 50% dari panjang frame.

2.4.4 Windowing

Proses framing dapat menyebabkan terjadinya kebocoran spektral (Spectral leakage) atau aliasing, aliasing adalah timbulnya sinyal baru dimana memiliki frekuensi yang berbeda dengan sinyal aslinya. Efek ini dapat terjadi karena rendahnya jumlah sampling rate, ataupun karena proses frame blocking dimana menyebablan sinyal menjadi discontinue, Untuk mengurangi kemungkinan terjadinya kebocoran spektral maka hasil dari proses framing harus melewati proses window.

Ada banyak fungsi window, w(n), seperti yang ditabel 2.1 sebuah fungsi window yang baik harus menyempit pada bagian main lobe, dan melebar pada bagian side lobe-nya.

Berikut ini adalah representasi fungsi window terhadap sinyal suara yang di-inputkan.

$$x(n) = xi(n)\, w(n) n = 0,1, \ldots\ldots, N-1 \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

$x(n)$ =Nilai sampel sinyal
$xi(n)$ =Nilai sampel dari frame sinyal ke i
$w(n)$ =Fungsi window
N =Frame size, merupakan kelipatan 2

Setiap fungsi windows mempunyai karakteristik masing-masing, diantara berbagai fungsi window tersebut, Blackman windows menghasilkan sidelobe level yang paling tinggi (kurang dari -58 dB). Tetapi fungsi ini juga menghasilkan noise paling besar (kurang dari 1,73 BINS), Oleh karena itum fungsi ini jarang sekali digunakan baik untuk speaker recognition maupun speech recognition.

Fungsi Rectangle window adalah fungsi window yang paling mudah untuk diaplikasikan , Fungsi ini menghasilkan noise yang paling rendah yaitu sekitar 1.00 BINS. Tetapi sayangnya fungsi ini memberikan sidelobe level yang paling rendah.

Sidelobe level yang rendah tersebut menyebabkan besarnya kebocoran spektral yang terjadi dalam proses feature extraction.

Fungsi window yang paling sering digunakan dalam aplikasi speech recognition adalah Hamming window. Fungsi window ini menghasilkan sidelobe level yang tidak terlalu tinggi (kurang dari -43 dB), selain itu noise yang dihasilkan pun tidak terlalu besar (kurang lebih 1.36 BINS) Darmawan (2011).

Tabel 2.1 Fungsi – fungsi window dan Formulanya

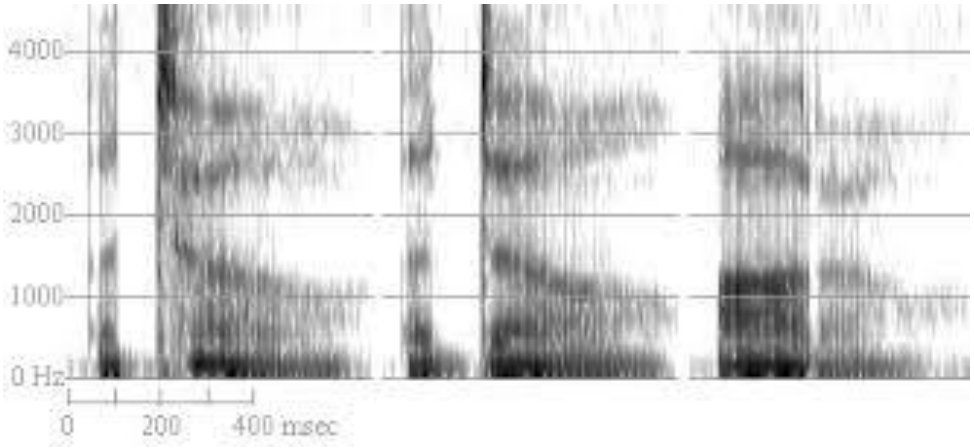
Nama window	Sequence domain waktu
Rectangular	1
Bartlett	$w(n) = a_0 - a_1 \left\| \frac{n}{N-1} - \frac{1}{2} \right\| - a_2 \cos\left(\frac{2\pi n}{N-1}\right)$
Hanning	$w(n) = 0.5 \left(1 - \cos\left(\frac{2\pi n}{N-1}\right)\right)$
Hamming	$w(n) = \alpha - \beta \cos\left(\frac{2\pi n}{N-1}\right),$
Blackman	$w(n) = a_0 - a_1 \cos\left(\frac{2\pi n}{N-1}\right) + a_2 \cos\left(\frac{4\pi n}{N-1}\right)$
Tukey	$w(n) = \begin{cases} \frac{1}{2}\left[1 + \cos\left(\pi\left(\frac{2n}{\alpha(N-1)} - 1\right)\right)\right] & 0 \leqslant n \leqslant \frac{\alpha(N-1)}{2} \\ 1 & \frac{\alpha(N-1)}{2} \leqslant n \leqslant (N-1)(1-\frac{\alpha}{2}) \\ \frac{1}{2}\left[1 + \cos\left(\pi\left(\frac{2n}{\alpha(N-1)} - \frac{2}{\alpha} + 1\right)\right)\right] & (N-1)(1-\frac{\alpha}{2}) \leqslant n \leqslant (N-1) \end{cases}$

Sumber : Darmawan(2011)

2.4.5 Analisis Fourier

Analisis Fourier adalah sebuah bentuk yang memungkinkan untuk menganalisa terhadap spectral propertis dari sinyal yang diinputkan. Representasi dari spectral propertis disebut sebagai spectogram Gambar 2.5 merupakan contoh dari spectogram.

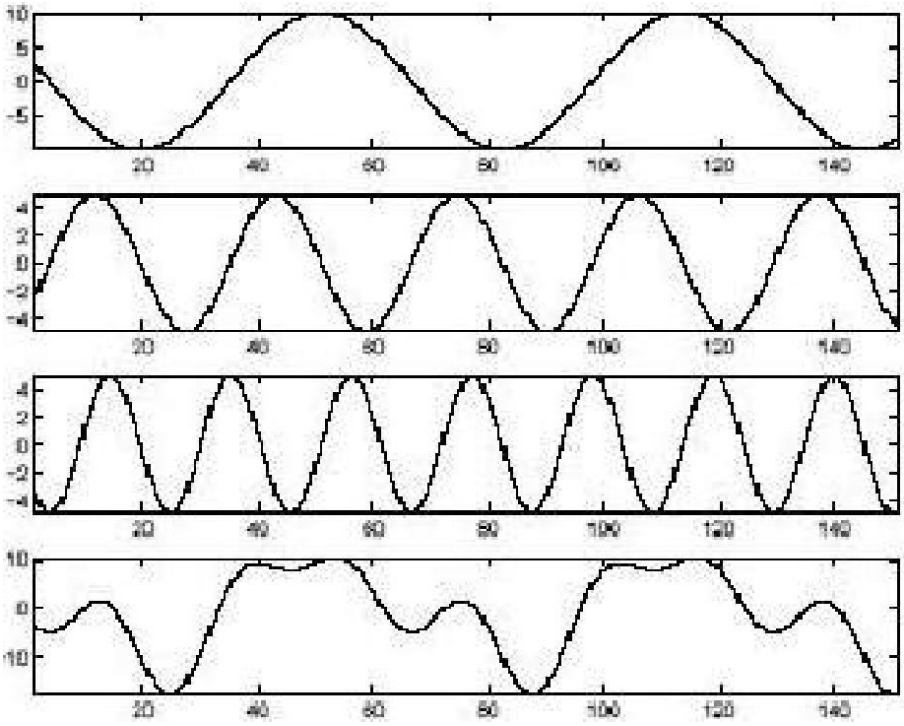
Dalam spectogram terhadap hubungan yang sangat erat antara waktu dan frekuensi. Hubungan antara frekuensi dan waktu adalah hubungan berbanding terbalik. Bila resolusi waktu yang digunakan tinggi. Maka resolusi frekuensi yang digunakan akan semakin rendah. Kondisi seperti ini akan menghasilkan narrowband spectogram, sedangkan wideband spectogram adalah kebalikan dari narrowband spectogram



Sumber : Darmawan (2011)

Gambar 2.5 Contoh dari Spectogram

Inti dari transformasi fourier adalah menguraikan sinyal ke dalam komponen-komponen bentuk sinus yang berbeda – beda frekuensinya. Gambar 2.11 menunjukkan tiga gelombang sinus dan superposisinya. Sinyal semula yang periodik dapat diuraikan menjadi beberapa komponen bentuk sinus dengan frekuensi berbeda, jika sinyal semula tidak periodik maka transformasi fourier-nya merupakan fungsi frekuensi yang continue, artinya merupakan penjumlahan bentuk sinus dari segala frekuensi, jadi dapat disimpulkan bahwa transformasi fourer merupakan representasi domain frekuensi dari suatu sinyal. Representasi ini mengandung informasi yang tepat sama dengan kandungan dari sinyal semula Darmawan (2011).



Sumber : darmawan(2011)

Gambar 2.6 Tiga Gelombang Sinusoidal dan Superposisinya

a. Discrete Fourier Transform (DFT)

DFT merupakan perluasan dari transformasi fourier yang berlaku untuk sinyal – sinyal diskrit dengan panjang yang terhingga. Semua sinyal periodik terbentuk dari gabungan sinyal – sinyal sinusoidal yang menjadi satu dalam perumusanya dapat ditulis:

$$S[k] = \sum_{n-0}^{n-1} s[n] e^{-i2\pi nnk/n} , 0 \leq k \leq n-1 \ldots\ldots\ldots\ldots\ldots (4)$$

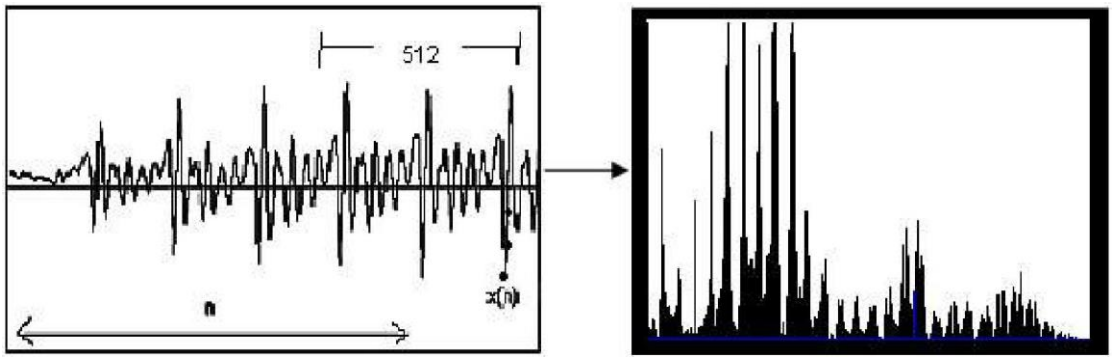
N = Jumlah sampel yang akan diproses

$S(n)$ = Nilai Sampel sinyal

K = Variabel frekuensi discrete, dimana akan bernilai ($k = \frac{N}{2}, k \in N$)

Dengan rumus diatasa. Suatu sinyal suara dalam domain waktu dapat kita cari frekuensi pembentuknya. Hal inilah tujuan dari penggunaan analisa Fourer pada data suara, yaitu untuk mengubah data dari domain waktu menjadi data spektrum di

domain frekuensi, Untuk pemrosesan sinyal suara, hal ini sangatlah menguntungkan karena data pada frekuensi dapat diproses dengan lebih mudah dibandingkan data pada domain waktu, karena pada domain frekuensi, keras lemahnya suara tidak seberapa berpengaruh.



Sumber : darmawan(2011)

Gambar 2.7 Domain waktu menjadi domain frekuensi

Untuk mendapatkan spektrum dari sebuah sinyal dengan DFT diperlukan N buah sampel data berurutan pada domain waktu, yaitu data x[m] sampai dengan x[m+N-1]. Data tersebut dimasukkan dalam fungsi DFT maka akan menghasilkan N buah data, Namun karena hasil DFT adalah simetris, maka hanya N/2 data yang diambil sebagai spektrum Darmawan (2011).

2.4.6 Fast Fourier Transform

Perhitungan DFT secara langsung dalam komputerisasi dapat menyebabkan proses perhitungan yang sangat lama. Hal itu disebabkan karena DFT, dibutuhkan N^2 perkalian bilangan kompleks. Karena itu dibutuhkan cara lain untuk meghitung DFT dengan cepat. Hal itu dilakukan dengan menggunakan algoritma Fast Fourier Transform (FFT) dimana FFT menghilangkan proses perhitungan yang kembar dalam DFT. Algoritma FFT hanya membutuhkan $N \log_2 N$ perkalian kompleks. Berikut ini menunjukkan perbandingan kecepatan antara FFT dan DFT

Algoritma recombine (DFT) melakukan N perkalian kompleks, dan dengan metode pembagian seperti ini. Maka terdapat $\log_2(N)$ langkah perkalian kompleks. Hal ini berarti jumlah perkalian kompleks berkurang dari N^2 (pada DFT) menjadi N $\log_2(N)$.

Hasil dari proses FFT adalah simetris antara indek $0 - (N/2 - 1)$ dan $(N/2) - (N - 1)$. Oleh karena itu , umumnya hanya blok pertama saja yang akan digunakan dalam proses-proses selanjutnya.

2.4.7 Mel Frequency Warping

Mel frequency Warping umumnya dilakukan dengan menggunakan filterbank. Filterbank adalah salah satu bentuk dari filter yang dilakukan dengan tujuan untuk mengetahui ukuran energi dari frequency band tertentu dalam sinyal suara. Filterbank dapat diterapkan baik pada domain waktu maupun domain frekuensi, tetapi untuk keperluan MFCC, filterbank harus diterapkan dalam domain frekuensi,

Filterbank menggunakan representasi konvolusi dalam melakukan *filter* terhadap *signal* konvolusi dapat dilakukan dengan melakukan multiplikasi antara spektrum *signal* dengan koefesien *filterbank*. Berikut ini adalah rumus yang digunakan dalam perhitungan *filterbanks.*

$$Y[t] = \sum_{f=1}^{n} S[j]Hi\,[j] \text{.....(5)}$$

N = Jumlah magintude spectrum ($N \epsilon N$)
$S[j]$ = Magnitude spectrum pada frekuensi j
$Hi[j]$ = koefesien filterbank pada frekuensi j ($1 \leq i \leq M$)
M = Jumlah channel dalam filterbank

Persepsi manusia terhadap frekuensi dari *signal* suara tidak mengikuti linear scale , frekuensi yang sebenarnya (dalam Hz) dalam sebuah signal akan diukur manusia secara subyektif dengan menggunakan *Mel scale, Mel frequency scale* adalah linear frekuensi *scale* pada frekuensi dibawah 1000 Hz, dan merupakan *logarithmic scale* pada frekuensi diatas 1000 Hz Putra(2011).

2.4.8 DCT

DCT merupakan langkah terakhir dari proses utama MFCC *feature extraction.* Konsep dasar dari DCT adalah mendekorelasikan *mel spectrum* sehingga menghasilkan representasi yang baik dari property spektral local. Pada dasarnya konsep dari DCT sama dengan *inverse fourier transform*. Namun hasil dari DCT mendekati PCA (*principle component analysis).* PCA adalah metode static klasik yang digunakan secara luas dalam analisa data dan kompresi. Hal inilah yang menyebabkan seringkali DCT menggantikan *inverse fourier transform* dalam proses MFCC *Feature Extraction.* Berikut adalah formula yang digunakan untuk menghitung DCT.

$$Cn = \sum_{k=1}^{k} (logSk)\, Cos\, [\, n(\, K - \tfrac{1}{2})\, \tfrac{\pi}{k}\,]\, 1\, n = 1.2 \ldots . K \ldots\ldots\ldots\ldots (6)$$

Sk = Keluaran dari proses *filterbank* pada *index* K

K = Jumlah koefesien yang diharapkan

Koefesien ke nol dari DCT pada umumnya akan dihilangkan, walaupun sebenarnya mengindikasikan energi dari frame signal tersebut. Hal ini dilakukan karena, berdasarkan penelitian – penelitian yang pernah dilakukan , koefesien ke nol tidak reliable terhadap speaker recognition Putra(2011).

2.5 Jaringan Syaraf Tiruan

Semakin berkembangnya teknologi komputer menyebabkan pemanfaatan teknologi jaringan syaraf untuk mempermudah manusia dalam memecahkan masalah tertentu semakin banyak diterapkan. Tetapi banyak masalah yang kelihatan mudah bagi manusia cukup sulit dilakukan oleh komputer, misalnya dalam pengenalan suatu tanda tangan yang telah dikenal sebelumnya. Kemudahan yang dirasakan oleh manusia tersebut disebabkan otak manusia memproses informasi yang didapat dengan menggunakan elemen-elemen yang saling terkoneksi dalam suatu jaringan yang disebut *Neuron*, sebaliknya jika masalah-masalah tersebut dipecahkan komputer, maka menimbulkan berbagai kesulitan (marimin,2002)

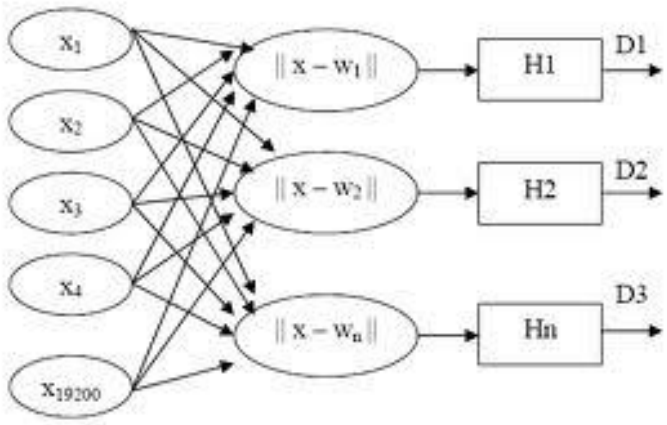
Didasar pada kemudahan otak manusia melakukan hal-hal tersebut, para ahli merancang suatu jaringan yang memiliki konsep menyerupai jaringan otak manusia dengan *neuron-neuron* dan hubungan-hubungannya. Jaringan tersebut dapat dilatih

sehingga dapat berpikir dan mengambil keputusan seperti yang dilakukan oleh otak manusia, jaringan tersebut disebut jaringan syaraf tiruan (JST).

2.5.1 Learning Vector Quantization (LVQ)

Menurut Jang, *et al.*(1997) LVQ merupakan metode klasifikasi data adaptif berdasarkan pada data pelatihan dengan informasi kelas yang diinginkan. Walaupun merupakan suatu metoda pelatihan *supervised* tetapi LVQ menggunakan teknik data *clustering unsupervised* untuk praproses set data dan penentuan *cluster center*nya. Arsitektur jaringan LVQ hampir menyerupai suatu jaringan pelatihan kompetitif kecuali pada masing-masing unit *output*nya yang dihubungkan dengan suatu kelas tertentu.

Kusumadewi dan hartai (2006) menyatakan LVQ merupakan metoda untuk melakukan pelatihan terhadap lapisan-lapisan kompetitif *supervised.* Lapisan kompetitif akan belajar secara otomatis untuk melakukan klasifikasi terhadap vektor *input* yang diberikan. Apabila beberapa vektor *input* memiliki jarak yang sangat berdekatan, maka vektor-vektor *input* tersebut akan dikelompokkan dalam kelas yang sama.



Sumber : Ridwan(2011)

Gambar 2.10 Arsitektur Jaringan LVQ

Jaringan LVQ terdiri atas 2 lapis yaitu lapis kompetitif dan lapis linear, Lapis kompetitif disebut juga *Self Organizing Map (SOM).* Disebut lapis kompetitif karena neuron – neuron berkompetisi dengan algoritma kompetisi yang akan menghasilkan neuron pemenang (*winning neuron*). Kelebihan dari LVQ adalah :

1. Nilai error yang lebih kecil dibandingkan jaringan syaraf tiruan seperti backpropagation.
2. Dapat meringkas data set yang besar menjadi vektor codebook berukuran kecil untuk klasifikasi.
3. Dimensi dalam codebook tidak dibatasi seperti dalam teknol nearest neighbour.
4. Model yang dihasilkan dapat diperbaharui secara bertahap.

Kekurangan dari LVQ adalah :

1. Dibutuhkan perhitungan jarak untuk seluruh atribut.
2. Akurasi model dengan bergantung pada inisialisasi model serta parameter yang digunakan (learning rate, iterasi dan sebagainya).
3. akurasi juga dipengaruhi distribusi kelas pada data training.
4. sulit untuk menentukan jumlah codebook vektor untuk masalah yang diberikan. Ridwan (2011).

2.6 Penelitian Terdahulu

Dibagian ini akan dijabarkan beberapa penelitian terdahulu. Saat ini sudah banyak penelitian yang berbasis pengenalan suara. Untuk lebih jelasnya. Pada table 2.2 berikut ini akan dijelaskan penelitian – penelitian yang telah dibuat sebelumnya.

Tabel 2.2 Penelitian terdahulu

No.	Judul	Tahun	Keterangan
1	Perbandingan pemodelan Wavelet dan MFCC sebagai ekstraksi ciri pada pengenalan fonem dengan teknik jaringan syaraf tiruan sebagai *classifier*	2011	Perbandingan dua metode ekstraksi ciri yang berbasis transformasi *Fourier* dan transformasi *Wavelet* pada pengenalan fonem serta penggunaan JST sebagai *Classifier*
2.	Pengenalan *Chord* pada alat musik gitar menggunakan teknik ekstraksi ciri MFCC	2010	Menerapkan metode *Codebook* dan teknik ekstraksi ciri MFCC dalam mengenali setiap *chord* yang dimainkan dengan alat musik gitar.

No.	Judul	Tahun	Keterangan
3.	Verifikasi biometrika suara menggunakan metode MFCC dan DTW	2011	Penggunaan metode MFCC untuk proses ekstarksi ciri dari sinyal wicara dan metode DTW (*Dynamic Time Warping*) untuk proses pencocokan.
4.	Pengenalan Suara Alat Musik Dengan Metode Jaringan Saraf Tiruan (JST) *Learning Vector Quantization* Melalui ekstraksi Koefisien Cepstral	2011	Pengenalan suara alat musik dengan menggunakan metode ekstraksi ciri Koeffisien Cepstral dan metode pencocokan adalah *Learning Vector Quantization*