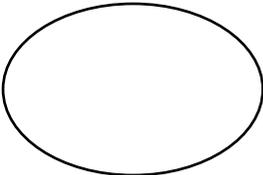


DAFTAR SIMBOL

Simbol	Nama Simbol	Kegunaan
	Simbol Titik Terminal	Menunjukkan permulaan (start) atau akhir (stop) dari suatu proses
	Simbol Proses	Digunakan untuk menunjukkan kegiatan yang dilakukan oleh suatu proses
	Simbol Arus	Simbol yang digunakan untuk menghubungkan antara simbol yang satu dengan simbol yang lain
	Simbol <i>Input</i> dan <i>Output</i>	Menunjukkan proses input-output yang terjadi tanpa bergantung dari jenis peralatannya

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi di masa saat ini merupakan hal yang sangat harus diterima bagi seluruh kalangan masyarakat, baik anak-anak hingga yang sudah dewasa. Khususnya pada bidang pendidikan, teknologi yang diterapkan pun berbagai macam mulai dari pembelajaran *online*, tes ujian *online*, dan masih banyak lagi, dengan tujuan untuk memudahkan dalam memberikan informasi serta pembelajaran.

Masa pengetahuan pada anak-anak yang masih dini memang sangat diperlukan, terlebih lagi pada saat memasuki umur-umur yang sudah harus mengenal berbagai macam kata-kata, huruf, dan bagaimana cara menyebutkan kata dan huruf tersebut. Menyebutkan huruf merupakan hal yang sangat dapat dilakukan bagi anak-anak usia dini. Teknologi menjadi suatu yang paling dekat dengan masyarakat saat ini, mulai dari kalangan anak-anak sampai dewasa serta macam-macam fungsinya. Menurut M.Maryono teknologi adalah pengembangan dan penerapan berbagai peralatan atau sistem untuk menyelesaikan persoalan-persoalan yang dihadapi manusia dalam kehidupan sehari-hari.

Dalam kehidupan sehari-hari perkataan dan ucapan merupakan hal yang sangat sering dilakukan oleh manusia, penggunaan suara merupakan salah satu media yang sangat tepat untuk digunakan sebagai media interaksi antara komputer dengan manusia untuk melakukan pembelajaran bagi anak-anak usia dini yang ingin membaca maupun mengetahui berbagai macam-macam huruf. Pengenalan suara sudah menjadi variasi dalam dunia teknologi, pengenalan suara mempunyai dua bagian meliputi ekstraksi dan klasifikasi. Pengenalan suara dilakukan dari setiap kata yang terucap akan memiliki informasi dengan bentuk sinyal suara, algoritma yang digunakan adalah metode MFCC dan *Dynamic Time Warping* (DTW). DTW

digunakan untuk mengukur 2 sekuens yang berbeda dengan mengukur kesamaan jenis suara berdasarkan waktu dan kecepatan, sedangkan MFCC merupakan tahap ekstraksi ciri yang digunakan pada suara.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka dapat ditulis bahwa rumusan masalah yang dilakukan pada penelitian ini adalah :

- a. Bagaimana *voice recognition* yang mengenali suara anak-anak dalam mengucapkan huruf menggunakan metode MFCC dan DTW.
- b. Bagaimana tingkat kesuksesan metode MFCC dan DTW pada *voice recognition*, dan dilakukan analisis.

1.3 Tujuan

Dapat dituliskan bahwa tujuan dari penelitian ini adalah :

- a. Suara huruf yang diberikan anak-anak dapat diketahui oleh system, dan pengenalan suara dapat dilakukan menggunakan metode MFCC dan menggunakan metode DTW.
- b. Menganalisis tingkat keberhasilan dari Metode MFCC dan DTW.

1.4 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah :

- a. Pengujian suara dilakukan pada anak usia 4-7 tahun.
- b. Suara yang digunakan berupa suara orang dewasa sebagai data latih dan anak-anak sebagai data uji.
- c. Metode yang digunakan adalah MFCC untuk ekstraksi fitur dan DTW sebagai pembandingan.
- d. Kesesuaian pada lafal yang diucapkan pada anak-anak disesuaikan dengan data yang tersimpan pada dataset.
- e. Pengucapan suara saat melakukan pengujian hanya mengucapkan 26 huruf.

1.5 Metodologi Penelitian

Metodologi dalam proses penyelesaian masalah penelitian ini terdiri dari beberapa tahap yaitu:

a. Tahap Studi Literatur

Pada tahap ini merupakan tahap dimana dilakukan beberapa pengumpulan berbagai data yang diperlukan pada penelitian ini, baik dalam perancangan maupun saat implementasi. Data dapat di dapat dalam berupa jurnal, buku elektronik, maupun internet, dengan bertujuan untuk membantu penulis dalam mengerjakan Tugas Akhir ini dengan baik.

b. Analisis dan Perancangan Kebutuhan Sistem

Tahap ini merupakan tahap perancangan yang dibuat sesuai dengan kebutuhan, dengan membuat program dengan fungsinya untuk melakukan *voice recognition* sehingga suaranya dapat direkam dan dilihat kesesuaiannya menggunakan metode MFCC dan DTW.

c. Implementasi Sistem

Tahap Implementasi Sistem ini dilakukan untuk melakukan implementasi dan menerapkan pada system yang telah dirancang.

d. Pengujian sistem dan analisis

Pada tahap ini dilakukan pengujian pada system yang telah dirancang dan dibuat sebelumnya, melihat bagaimana tingkat keakurasian pada *voice recognition* yang digunakan, kemudian bagaimana suara dapat diterima dengan melihat nilai yang ada.

e. Penyusunan Laporan Tugas Akhir

Pada tahap ini dilakukan pembuatan laporan Tugas Akhir dengan mengumpulkan hasil, dokumentasi, dan penulisan yang. Pada tahap ini dapat menjelaskan dari awal hingga akhir bagaimana penelitian dibuat sehingga terbuatnya laporan tugas akhir, laporan harus mengikuti format yang ada dan sesuai ketentuan.

1.6 Sistematika Penulisan

Dalam penyusunan laporan Tugas Akhir ini disusun secara baik dan secara terstruktur, diantaranya sebagai berikut :

a. BAB I PENDAHULUAN

BAB I berisi tentang latar belakang, rumusan masalah, tujuan, Batasan masalah, metodologi penelitian dari penelitian Tugas Akhir.

b. BAB II LANDASAN TEORI

BAB II berisi landasan teori dari Media Pembelajaran Interaktif, *Voice Recognition*, Metode MFCC dan Metode DTW.

c. BAB III PERANCANGAN SISTEM

BAB III berisi mengenai perancangan dan penjelasan umum system yang akan dibuat , data-data, dan metode yang digunakan.

d. BAB IV IMPLEMENTASI DAN PENGUJIAN

BAB IV berisi tentang hasil pengujian penelitian serta analisis penelitian.

e. BAB V KESIMPULAN DAN SARAN

BAB V berisi kesimpulan dari keseluruhan hasil penelitian dan saran untuk penelitian selanjutnya agar lebih baik.

BAB II

LANDASAN TEORI

2.1 Media Pembelajaran

Menurut Gagne dan Briggs (1975) dalam Arsyad (2013: 4) secara rinci mengatakan bahwa media pembelajaran meliputi alat yang secara fisik digunakan untuk menyampaikan isi materi pengajaran. Media pembelajaran dapat berupa visual, audio, dan elektronis yang tujuannya agar informasi dan materi dapat sampai dengan cara lebih mudah. Media pembelajaran yang baik akan memberikan materi yang lebih bermanfaat dan cepat mengerti antar pengajar dan penerima.

Media pembelajaran yang dilakukan dengan suara maupun audio yang dibuat oleh suatu program atau computer, memungkinkan pembelajaran yang lebih menarik karena bentuk media yang berbeda. Pembelajaran interaktif ini dimaksudkan sebagai banyaknya berbagai interaksi yang dilakukan dari berbagai arah, dilakukan dengan 2 pihak maka interaksi dapat dilakukan dengan baik [6]. Kegunaan media audio suara dapat membuat keinginan anak-anak untuk belajar menjadi tinggi, materi yang diberikan pada media interaktif juga dapat memuat respon lebih pada anak-anak.

2.2 Voice Recognition

Voice Recognition merupakan suatu teknologi yang memungkinkan mesin untuk merespon secara baik dan benar terhadap suara manusia . Secara umum *voice recognition* dapat dibagi menjadi dua tahapan, yakni tahap pembelajaran pola dan tahap pengenalan suara melalui perbandingan pola [1][13].

Voice recognition dibagi menjadi dua jenis menjadi *speech recognition* dan *speaker recognition*. *Speech recognition* biasanya digunakan untuk mengenali suara yang asalnya dari mulut manusia, baik berupa perkataan atau hanya sebatas ejaan, tujuan *speech recognition* adalah untuk membuat komputer agar dapat mengenali perkataan manusia seperti manusia lain mengenali nya [2].

2.2.1 *Speech Recognition*

Pada penggunaan pengenalan suara adalah proses mengkonversi sinyal akustik, pada pengenalan suara dapat berfungsi sebagai salah satu input satu pemrosesan bahasa alami. Terdapat berbagai proses pada *speech recognition* seperti *pre-processing*, *feature extraction*, dan *classification*. pada tahap awal dilakukan *pre-processing* dengan menyetarakan sinyal suara dengan sinyal noise yang ada. Pada tahap kedua dilakukan proses ekstraksi fitur, dan terakhir dilakukan klasifikasi untuk mengetahui jenis sinyal pada suara [3]. Kecocokkan akurasi dan skor yang didapat tergantung dari metode yang digunakan dan data latih yang digunakan.

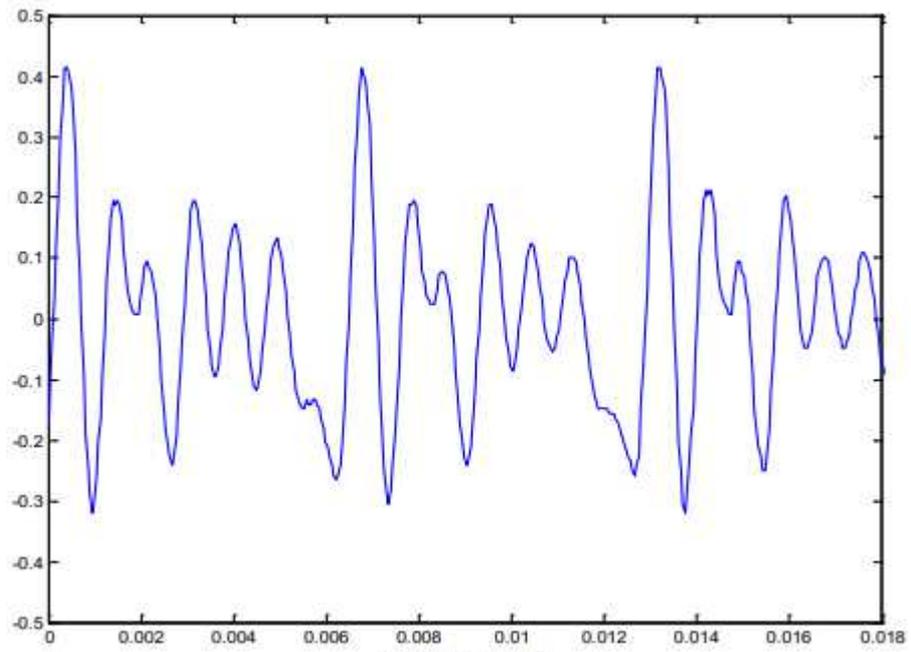
2.2.2 *Speaker Recognition*

Speaker recognition merupakan proses yang dilakukan untuk mengenali identitas dari pembicara, *speaker recognition* biasanya diklasifikasikan sebagai identifikasi dan verifikasi. Sistem pengenalan speaker pertama kali mencoba memodelkan karakteristik saluran vokal seseorang. *Speaker Recognition* merupakan model matematika dari suatu sistem yang menghasilkan ucapan manusia atau hanya model statistik dengan karakteristik keluaran yang serupa dengan saluran vokal manusia [15].

2.2.3 Sinyal Suara

Sinyal suara bisa diartikan sebagai sebagai suatu sinyal yang asalnya dari suara, dimana sinyal terbentuk dari kombinasi antara frekuensi yang ada pada amplitud dan fasa. Sinyal suara yang berasal dari pengucapan

manusia biasanya dilakukan dengan konsep atau gagasan yang sudah dipikirkan sebelumnya pada pengucap untuk disampaikan kepada pendengar. Sinyal suara terjadi secara perlahan waktu variasi sinyal. Berikut adalah contoh sinyal suara :



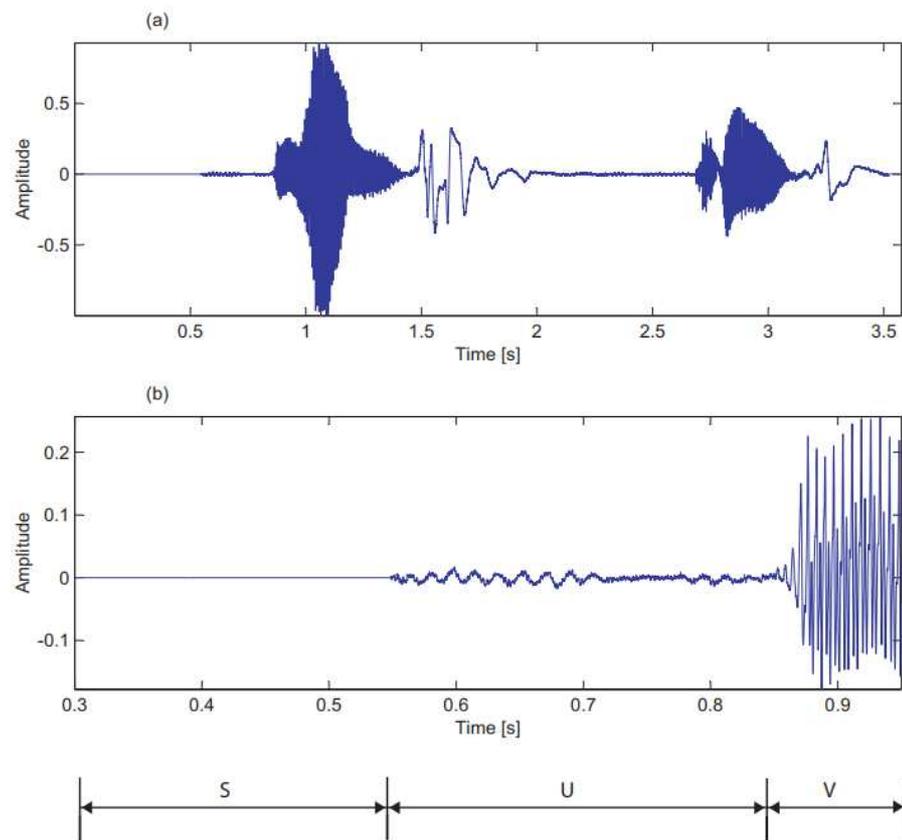
Gambar 2.1 Contoh sinyal suara [17]

Sinyal suara dapat direpresentasikan karakteristiknya dengan cara yang tepat dengan *three-state representation*, atau representasi *spectral*.

1. *Three-state representation*

Three-state representation adalah salah satu cara yang digunakan untuk mengklasifikasikan kejadian pada suara, kejadian yang terjadi adalah sebagai berikut :

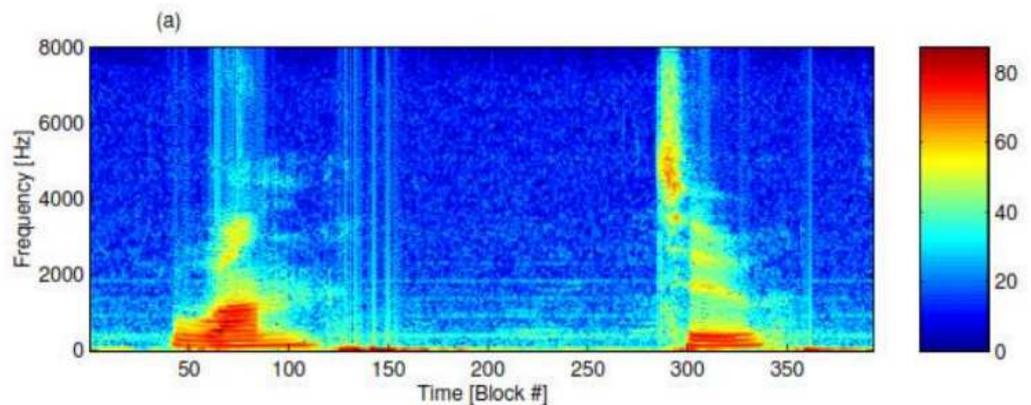
- *Silence (S)*, yaitu pada saat suara tidak digunakan atau tidak diproduksi.
- *Unvoiced (U)*, yaitu pada saat suara mengeluarkan huruf vocal namun tidak bervibrasi.
- *Voiced (V)*, yaitu pada saat suara mengeluarkan huruf vocal bervibrasi secara periodik.



Gambar 2.2 *Three-state representation* [19]

2. Representasi Spektral

Representasi spectral adalah representasi sinyal-sinyal pada suara dengan keterangan sebagai warna-warna. *Spectrogram* suara adalah representasi 3 dimensi dari sinyal suara, dengan keterangan pada sumbu x menunjukkan keterangan pada waktu, sedangkan pada sumbu y adalah memberikan representasi keterangan, warna yang ditampilkan menandai keluaran dari sinyal suara. Berikut adalah contoh representasi spektral :



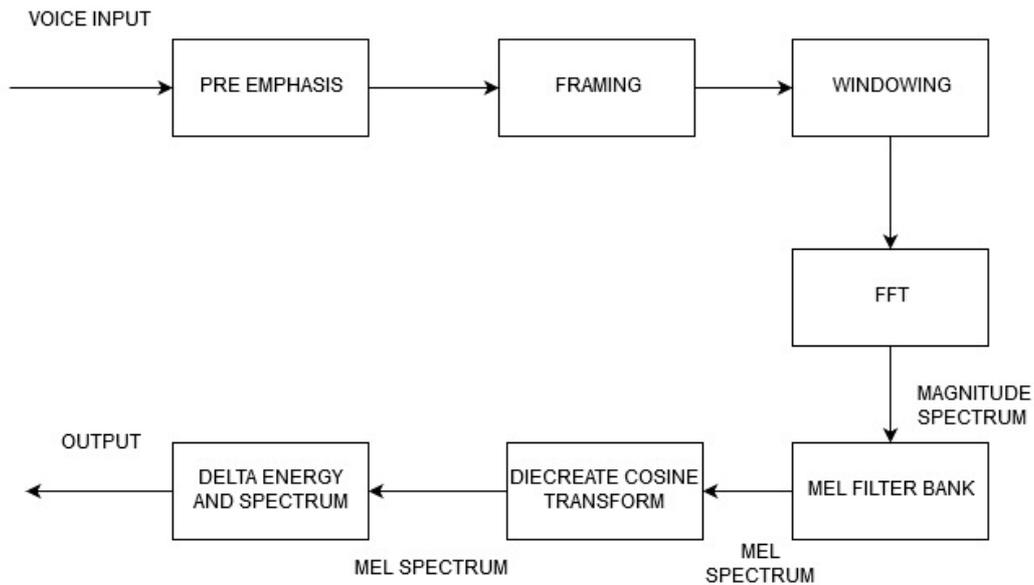
Gambar 2.3 Representasi spektral [19]

Bagian berwarna biru pekat pada gambar 2.5 menunjukkan bagian gelombang dusrara dimana suara tidak terjadi atau tidak diproduksi dan bagian yang berwarna merah menunjukkan dimana danya suara diproduksi [19].

2.3 MFCC (*Mel Frequency Cepstral Coefficient*)

MFCC (*Mel Frequency Cpestral Coefficient*) adalah metode yang biasanya digunakan pada pengaplikasian untuk *voice recognition* untuk merepresentasi perkataan manusia atau sinyal music yang dapat ditangkap dengan frekuensi dibawah 1Khz [14]. Terdapat berbagai metode yang digunakan untuk mengenali blok ucapan, MFCC lebih berguna bila digunakan untuk *speech recognition* [4]. MFCC lebih sering digunakan , MFCC memanifestasikan dirinya dalam sebuah *daya spectrum* waktu singkat dan kemudian MFCC akan secara akurat merepresentasi sebuah spectrum tersebut untuk suara dengan tipe pengucapan pada manusia dengan melakukan ekstraksi kepada suara [5]. MFCC diperoleh dengan menerapkan sejumlah filter berbentuk segitiga ke dalam sinyal untuk mendapatkan jumlah energi dari rentang frekuensi tertentu [6].

MFCC dapat dengan baik digunakan sebagai pengenalan suara dan *speaker recognition*, dengan menggabungkan fitur energi dan fitur delta. Dalam perhitungan MFCC *neural network* digunakan sebagai model pada pengenalan suara. Gambar 1 menampilkan alur ekstrasi pada fitur yang menggunakan MFCC.



Gambar 2.4 Diagram alur ekstrasi MFCC

Dalam melakukan ekstrasi pada MFCC suara yang didapat harus bebas dari *noise* atau suara yang tidak diinginkan, maka harus dilakukan proses filterasi sebelumnya, menghilangkan *noise* merupakan hal yang penting pada metode ini [7].

1) *Preemphasis*

Preemphasis adalah suatu proses yang filterasi yang biasa digunakan sebelum sinyal di proses lebih jauh [8]. Dengan dilakukannya *preemphasis* yang meningkatkan energi frekuensi menjadi lebih tinggi membuat informasi yang dikeluarkan menjadi lebih akurat, *preemphasis* dapat ditentukan dengan persamaan sebagai berikut :

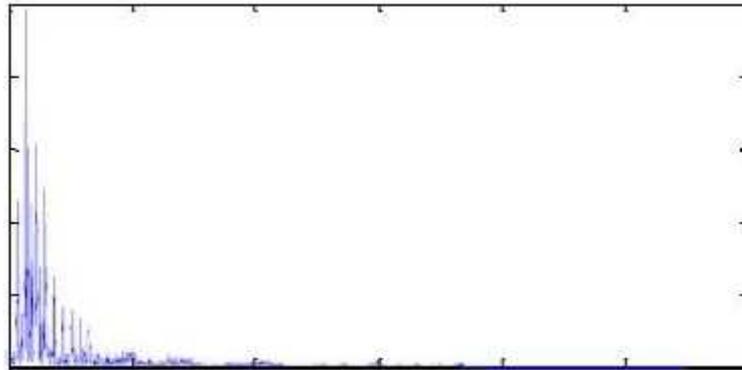
$$y(n) = x(n) - a * x(n - 1) \dots (2.1)$$

Dimana :

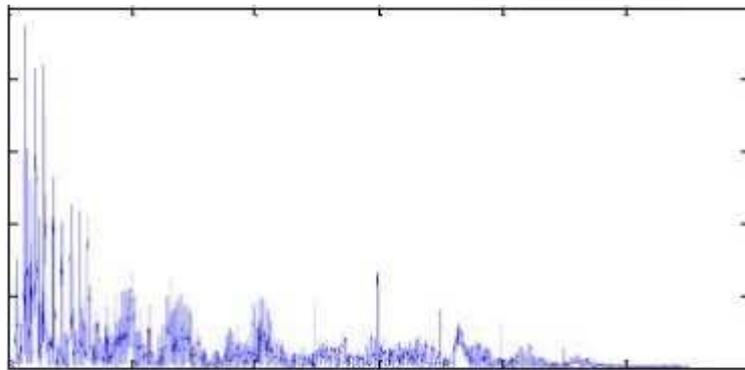
$y(n)$ = sinyal keluaran, biasanya nilainya berada diantara 0.9 – 1.0 Khz [9]

$x(n)$ = sinyal *speech* yang dikirim melalui *high pass filter*.

Berikut adalah salah satu contoh sinyal yang telah dilakukan *preemphasis* :



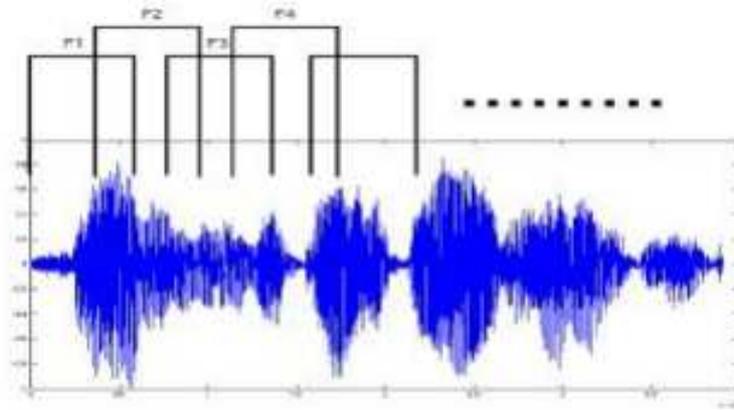
Gambar 2.5 Sinyal sebelum *preemphasis* [16]



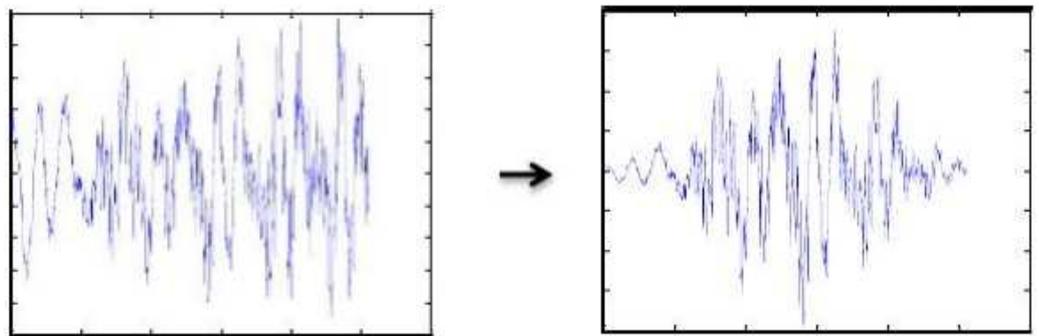
Gambar 2.6 Sinyal sesudah *preemphasis* [16]

2) *Framming and Windowing*

Framming dan *Windowing* adalah tahapan dimana sinyal di proses yang berada pada kisaran 20-40 ms [4], kemudian *overlapping* sebelumnya dapat dilakukan secara berlanjut dalam *frame*. Khusus pada *windowing*, digunakan untuk meminimalisir sinyal diskontinuitas pada awal dan akhir pada setiap bingkai *frame* [10]. Berikut adalah salah satu contoh sinyal yang di *farming* dan *windowing* :



Gambar 2.7 Frame blocking pada sinyal [16]



Gambar 2.8 Sinyal sebelum dan sesudah *windowing* [16]

3) *Fast Fourier Transform (FFT)*

Fast Fourier Transform merupakan suatu alur pada MFCC yang berfungsi sebagai konversi setiap *frame* yang berisi N sampel dari ranah waktu ke ranah frekuensi . FFT merupakan jenis komputasi yang bisa dibilang cepat untuk melakukan *Fourier transform* pada domain diskrit [4].

FFT mempunyai persamaan sebagai berikut :

$$Y(w) = FFT [h(t) * X(t)] = H(w) * X(w) \dots (2.2)$$

Dimana :

Y (w) = Output signal. Domain *frekuensi*

X (t) = input *signal*, domain waktu