

## 2.3 Teori Pendukung

### 2.3.1 Metode Ekstraksi Ciri MFCC

Metode Melf-Frequency Cepstral Coefficients (MFCC) merupakan ekstraksi ciri yang paling sering digunakan dalam pemrosesan suara, karena dapat mempresentasikan sinyal dengan baik. Langkah-langkah proses MFCC berdasarkan pada perbedaan dari frekuensi yang terdengar oleh pendengaran manusia melalui panca indranya, sehingga dapat seperti layaknya manusia merepresentasikan sinyal suara. Proses ekstraksi ciri MFCC [14] adalah sebagai berikut:

#### a. Preemphasis

Preemphasis filter pada proses pengolahan sinyal suara diperlukan setelah proses pengolahan data sampel didapatkan. Tujuan dari proses penyaringan ini adalah untuk mendapatkan bentuk spectral dari frekuensi sinyal suara yang lebih bagus. Dimana bentuk spectral yang memiliki nilai tinggi untuk daerah kecil akan cenderung turun secara tajam pada daerah dengan frekuensi >2000 Hz. Penyaringan pre-emphasis didapatkan oleh hubungan masukan ataupun luaran dalam domain waktu, seperti yang ditunjukkan pada persamaan 1:

$$y(n)=x(n)-ax(n-1) \quad (1)$$

Dimana alpha (a) adalah sebuah ketetapan filter pre-emphasis, dengan nilai  $0.9 < a < 1.0$ .

#### b. Frame Blocking

Pada proses frame blocking, sinyal suara yang telah di dapatkan, akan disegmentasikan menjadi beberapa frame yang saling bertumpang tindih (overlap). Proses ini akan meminimalisir sinyal yang terganggu atau hilang (deletion). Dimana proses ini akan terus berlanjut hingga seluruh sinyal yang ada telah masuk ke dalam frame.

#### c. Windowing

Sinyal suara yang telah di proses pada proses sebelumnya dibaca setiap framenya dan dalam setiap frame-nya akan dilakukan proses windowing dengan menggunakan fungsi window tertentu. Proses windowing digunakan untuk mengurangi sinyal yang tidak

berkelanjutan diawal dan akhir setiap frame. Luaran yang dihasilkan dari proses windowing adalah sinyal, dengan bentuk

persamaan seperti persamaan 2:

$$y1(n) = x1(n)w(n), 0 \leq n \leq N-1 \quad (2)$$

Dimana  $w(n)$ , menggunakan fungsi window Hamming, sehingga bentuk persamaan menjadi:

$$w(n) = 0.54 - 0.46 \cdot \cos(2\pi n / N - 1), 0 \leq n \leq N-1 \quad (3)$$

d. Fast Fourier Transform (FFT)

FFT digunakan untuk mengubah setiap frame dengan nilai N sampel data dari domain waktu menjadi bentuk domain frekuensi, seperti pada persamaan sebagai berikut:

$$Xn = \sum_{k=0}^{N-1} xk e^{-2\pi jkn/N} \quad (4)$$

Dimana  $n = 0, 1, 2, \dots, N-1$  dan  $j = \sqrt{-1}$ .

e. Mel-Frequency Wrapping

Persepsi cara pendengaran pada manusia terhadap sebuah frekuensi suara yang terdengar

tidak bisa diukur kedalam skala linear. Untuk setiap nada suara yang merupakan frekuensi aktual,  $f$ , memiliki bentuk satuan Hertz (Hz), “mel” adalah sebuah tinggi rendah nada dalam suara yang dapat diukur dalam sebuah skala. Skala mel-frequency adalah sebuah frekuensi rendah yang dengan ukuran  $<1000$  Hz bersifat linear dan sebuah frekuensi tinggi yang dengan ukuran  $>1000$  Hz bersifat logaritmik. Hubungan skala mel dengan frekuensi dalam Hz dapat ditunjukkan pada persamaan berikut ini:

$$F_{mel} = \begin{cases} 2595 \cdot \log_{10}(1 + F_{HZ}/700), & F_{HZ} > 1000 \\ F_{HZ}, & F_{HZ} < 1000 \end{cases} \quad (5)$$

Proses mel-frequency wrapping terhadap sinyal suara yang berada dalam domain frekuensi, menggunakan persamaan berikut:

$$Xi = \log_{10}(\sum_{k=0}^{N-1} |X(k)|^2 Hi(k)) \quad (6)$$

### 2.3.2 Pengenalan Pola DTW

Dynamic time warping(DTW) adalah algoritma yang menghitung warping path yang optimal antara dua sekuens numerik data sehingga output adalah nilai-nilai warping path dan jarak diantara kedua sekuens numerik tersebut. Algoritma DTW disebut juga sebagai non-linear sequence alignment, sehingga algoritma ini lebih realistis untuk digunakan dalam mengukur kemiripan suatu pola (pattern/template matching).

algoritma DTW adalah menghitung jarak local (local distance) antar elemen dari kedua sekuens numeric menggunakan teknik penghitungan jarak yang berbeda, biasanya yang digunakan adalah Euclidean Distance. Local distance dihitung dengan mencari nilai absolute dari selisih kedua sekuens numerik data yang dirumuskan dengan :

$$d_{ij}=|x_i-y_j|$$

$$|, i=1:n \text{ dan } j= 1:m \dots 1$$

Jika dalam matriks maka dapat ditulis dengan memiliki n garis dan m kolom.

$$x_{ij}=d_{ij}+\min(x_{i-1,j-1}, x_{i-1,j}, x_{i,j-1}) \dots 2$$

Dimana x merupakan jarak minimal antara subsekuens. Langkah selanjutnya

merupakan point yang harus kita lewati untuk menemukan warping path. Warping path adalah sebuah path yang melewati jarak matrik minimum dari elemen  $x_{ij}$  ke  $x_{11}$ . Ongkos warping path secara global dari dua sekuens: Dimana  $w_i$  adalah elemen yang dimiliki warping path dan  $p$  adalah jumlahnya. Penghitungannya dibuat untuk dua sekuens.

$$GC = \frac{1}{p} \sum_{i=1}^p w_i \cdot 3$$

### **2.3.3 Speech Recognition**

Speech Recognition adalah proses identifikasi suara berdasarkan kata yang diucapkan dengan melakukan konversi sebuah sinyal akustik, yang ditangkap oleh audio device (perangkat input suara). Speech Recognition juga merupakan sistem yang digunakan untuk mengenali perintah kata dari suara manusia dan kemudian diterjemahkan menjadi suatu data yang dimengerti oleh komputer.

Kata – kata yang ditangkap dan dikenali bisa jadi sebagai hasil akhir, untuk sebuah aplikasi seperti command & control, penginputan data, dan persiapan dokumen. Parameter yang dibandingkan ialah tingkat penekanan suara yang kemudian akan dicocokkan dengan template database yang tersedia. Sedangkan sistem pengenalan suara berdasarkan orang yang berbicara dinamakan speaker recognition. Algoritma yang akan diimplementasikan pada bahasan mengenai proses speech recognition ini adalah algoritma FFT (Fast Fourier Transform), yaitu algoritma yang cukup efisien dalam pemrosesan sinyal digital (dalam hal ini suara) dalam bentuk diskrit. Algoritma ini mengimplementasikan algoritma Divide and Conquer untuk pemrosesannya. Konsep utama algoritma ini adalah mengubah sinyal suara yang berbasis waktu menjadi berbasis frekuensi dengan membagi masalah menjadi beberapa upa masalah yang lebih kecil. Kemudian, setiap upa masalah diselesaikan dengan cara melakukan pencocokan pola digital suara.