BAB 2
DASAR TEORI

2.1 KAJIAN PUSTAKA

Penelitian Totok Chamidy pada tahun 2016 yang berjudul ”*Metode Mel-Frequency Cepstral Coeffisients* (MFCC) Pada klasifikasi *Hidden Markov Model* (HMM) Untuk Kata *Arabic* pada Penutur Indonesia” meneliti tentang deteksi penutur untuk kata *Arabic* menggunakan metode *Mel-Frequency Cepstral Coeffisients* (MFCC) pada klasifikasi *Hidden Markov Model* (HMM). Pengujian yang dilakukan dalam penelitian ini dititikberatkan pada tingkat akurasi dalam mengenali setiap data uji. Pengujian tingkat akurasi ini untuk mengetahui tingkat akurasi klasifikasi *hidden markov model* untuk mengenali ucapan dalam satu frekuensi cuplik, serta perbedaan dalam mengenali ucapan dalam tiga frekuensi cuplik yang berbeda. Kata yang diucapkan sebagai data uji dipilih secara acak yang dipergunakan dalam komunikasi sehari-hari. Dalam pengambilan data, kata tersebut ditulis dalam huruf latin menggunakan pedoman transliterasi dari kementrian agama Republik Indonesia dan penutur mengucapkannya sesuai dengan kemampuan masing-masing. Tidak ada kekhususan dalam pemilihan kata bahasa arab. Kata-kata tersebut adalah sebagai berikut: *Alhamdulillah; Bikhair; Ismi; Syukran; Afwan; Ahsanta; Na’am; La; Shahih; Shadaqta; Baarakallah; Syafakallah; Hafizhanallah; Hadaanallah; Tafadhdhal.* Dengan menggunakan metode MFCC dapat diperoleh hasil data dari sistem pengenalan ucapan menjadi lebih presisi dalam berbagai kondisi, suara penutur masih dapat dikenali dengan data latih menggunakan suara penutur arab. Penutur indonesia berusaha untuk menyesuaikan supaya pengucapannya sama dengan penutur arab. Namun sistem yang dirancang dengan menggunakan ekstraksi ciri MFCC, penutur Indonesia dalam pembacaan Al-Qur’an mampu dikenali dengan data latih menggunakan suara penutur Indonesia yang telah fasih dalam membaca Al-Qur’an[4].

Ayat Hafzalla Ahmed dan Sherif Mahdi Abdo pada tahun 2017 dengan penelitiannya berjudul “*Verification System for Quran Recitation Recordings*” membahas mengenai sistem verifikasi rekaman Khotbah Al-Quran dengan menggunakan metode MFCC, HMM dan *Artificial Neural Network* (ANN). Metode ANN di sini adalah model matematika berdasarkan jaringan saraf biologis. ANN

digunakan untuk mendapatkan pemahaman tentang jaringan saraf biologis, atau untuk memecahkan masalah kecerdasan buatan. Metode ANN adalah salah satu pendekatan kecerdasan buatan, yang mencoba mengkomputerisasi prosedur pengenalan. Sistem ini diimplementasikan dengan menggunakan *Spectral Subtraction* untuk pra pengolahan sinyal wicara, MFCC sebagai fitur ekstraksi dan ANN untuk pemodelan akustik dan pencocokan pola. Sistem ini mencapai akurasi 88% pada pengenalan kalimat. Namun, sistem ini dilatih hanya dengan kalimat yang berbeda dengan potongan ayat-ayat pendek [5].

Penelitian Afrillia, Yesy pada tahun 2017 yang berjudul "*Performance Measurement Of Mel-Frequency Ceptral Coefficient* (MFCC) *Method In Learning System Of Al- Qur'an Based In Nagham Pattern Recognition"* membahas mengenai pengukuran kinerja metode MFCC dalam pembelajaran AlQur'an berbasis Nagham sebagai sistem pengenalan suaranya. Ciri khas pola nagham Al-Quran jauh lebih kompleks dari pola *makhraj* dan *tajwid*. Dalam nagham gelombang suara memiliki lebih banyak variasi yang menyiratkan tingkat derau jauh lebih tinggi dan memiliki durasi suara lebih lama. Data pengujian dalam penelitian ini diambil lewat perekaman *real-time*. Pengukuran evaluasi dalam kinerja sistem pola Nagham Al-Quran didasarkan pada parameter deteksi benar dan salah dengan akurasi 80%. Untuk mengukur akurasi ini perlu memodifikasi MFCC atau memberi lebih banyak proses belajar data dengan lebih banyak variasi. Dalam sistem ini dengan menggunakan metode MFCC mampu menggunakan hanya satu buah data belajar untuk dibandingkan dengan banyaknya data uji [6].

Penelitian Chaerul Hadi dan Muhammad Rifqi Ma'arif pada tahun 2017 yang berjudul "Implementasi *Cosine Similarity* Dalam Aplikasi Pencarian Ayat Al-Qur'an Berbasis Android" membahas mengenai pencarian dan indeks tematik ayat Al-Qur'an dengan menggunakan metode *cosine similarity* untuk memaksimalkan hasil pencarian. *Cosine similarity* merupakan metode yang digunakan untuk menghitung *similarity* (tingkat kesamaan) antara dua buah objek. Algoritma pencarian *cosine similarity* diterapkan pada dokumen tema, bukan pada dokumen terjemahan Al-Qur'an. Sehingga apabila algoritma tersebut diterapkan pada dokumen terjemahan dalam kasus ini, maka dikhawatirkan terjadi klaim tafsir yang tidak sesuai terhadap ayat tertentu untuk tema tertentu. Hasil dari aplikasi dengan pencarian *cosine similarity* ini

didapatkan 70% *responden* yang dilibatkan dalam pengujian aplikasi menyatakan bahwa aplikasi yang dikembangkan dapat membantu mereka dalam pencarian ayat Al-Qur'an secara tematik. Pada sistem ini *cosine similarity* dapat digunakan sebagai pengukur kesamaan antara kemiripan dari pola *cepstrum* MFCC[2].

2.2 SINYAL WICARA DAN CARA PEMBACAAN AL-QUR'AN

Sinyal wicara penutur terbagi menjadi tiga bagian yaitu sinyal diam (*silence*), sinyal tidak berwicara (*unvoiced*) dan sinyal wicara (*voiced*). Sinyal *silence* merupakan sinyal ketika tidak ada wicara yang dihasilkan, sinyal *unvoiced* merupakan sinyal ketika pita wicara tidak bergetar sehingga menghasilkan bentuk gelombang wicara periodik atau aperiodik [7], sedangkan sinyal wicara secara umum bukanlah sebuah sinyal stasioner. Sinyal stasioner adalah sinyal yang di setiap rentang waktu memiliki kandungan informasi frekuensi sama. Akan tetapi sinyal wicara jika dipotong pada durasi yang cukup singkat sekitar 5 – 100 ms menampilkan bentuk sinyal stasioner [8]. Dengan demikian sinyal wicara digolongkan sebagai sinyal *quasi-stationary*. Jika durasi pemotongan lebih panjang maka karakteristik sinyal wicara berubah terhadap reflek sinyal wicara yang dihasilkan [9].

Al-Qu'ran merupakan kitab suci umat Islam, dimana didalamnya terdapat ajaran, perintah dan larangan dalam beribadah maupun menjalani kehidupan. Cara membaca Al-Qu'ran yang benar sangat diperlukan agar bacaan tersebut terdengar bagus, indah serta makna dari setiap ayat-ayat Al-Qur'an juga tersampaikan dengan benar [10]. Terdapat beberapa cara dalam pembacaan Al-Qur'an salah satunya adalah membaca secara *Murottal*. *Murottal* merupakan membaca Al-Qur'an yang memfokuskan pada dua hal yaitu kebenaran bacaan dan lagu Al-Qur'an. Tempo membaca secara *Murottal* dapat lambat, sedang, maupun cepat. Secara Bahasa *Murottal* membaca Al-Qur'an dengan *Tartil*, memperhatikan ilmu *tajwid* dan *makharijul* huruf. *Tartil* memiliki makna dibaca berdasarkan ilmu *tajwid*. *Tajwid* artinya setiap ayat-ayat Al-Qur'an dibaca dengan lafal yang bagus dan mengikuti dengan tepat di mana saja hendaknya berhenti ketika membaca Al-Qur'an. Sehingga membaca Al-Qur'an dengan sikap tenang, tidak terburu-buru dan setiap ayat demi ayat Al-Qur'an dibaca dengan jelas [11].

2.3 PRA PENGOLAHAN

2.3.1 *Centering*

Proses *centering* adalah tahapan pra-pengolahan yang dilakukan untuk membuang nilai bias (*offset*) agar *baseline* sinyal berada pada sumbu-x *origin*. Dengan kata lain untuk membentuk sinyal *zero-mean.* Pada tahapan ini data masukan diolah sehingga didapatkan rata-rata (*mean*) vektor sinyal. Vektor sinyal tersebut yang kemudian akan diolah ke tahapan selanjutnya [12]. Formula *centering* dapat dirumuskan sebagai berikut:

$$X' = X - \mu(X) \tag{2.1}$$

dimana:

X' = Matriks baru / tanpa *mean*

X = Matriks lama / mengandung *mean*

$\mu(X)$ = Rata-rata data sinyal

Dengan melakukan proses perata-rataan data yang akan diproses memiliki nilai referensi (acuan) yang sama yaitu di sumbu horizontal x = 0.

2.3.2 Normalisasi

Normalisasi adalah proses penskalaan pada nilai amplitudo tiap data sinyal sesuai skala yang diinginkan. Proses ini dilakukan agar rentang nilai amplitudo pada tiap data sinyal yang akan diolah bernilai sama. Besarnya nilai amplitudo sinyal wicara manusia saat melakukan pengucapan selalu berbeda-beda, sehingga penskalaan nilai amplitudo sinyal terhadap acuan skala yang diinginkan sangat diperlukan [13]. Proses normalisasi amplitudo diperoleh dengan membagi semua nilai sampel *digital* dengan nilai mutlak (*absolute)* maksimum dari sampel sinyal tersebut. Formula normalisasi adalah sebagai berikut [14]:

$$x'(n) = \frac{x(n)}{max(|x|)}, 1 \leq n \geq -1 \tag{2.2}$$

dimana:

$x'(n)$ = Hasil Normalisasi

$x(n)$ = Sinyal asli

2.3.3 *End Point Detection*

End Point Detection (EPD) dibutuhkan untuk mendapatkan posisi awal dan akhir yang tepat dari sinyal wicara. Sehingga bagian sinyal sebelum awal dan setelah akhir sinyal yang mengandung tutur akan dihilangkan. Hal ini perlu dilakukan karena sinyal yang direkam mengandung bagian tanpa tutur yang bisa muncul saat proses perekaman (proses persiapan), atau bisa juga karena lama perekaman lebih panjang dibandingkan dengan tutur yang direkam. Proses ini juga menjadikan ukuran sinyal menjadi lebih kecil. Di dalam penelitian ini, proses EPD menggunakan metode *Short Term Fourier Transform* (STFT) dengan nilai ambang batas [15] dan *Zero Crossing Rate* (ZCR) [7].

2.3.3.1 *Discrete Fourier Transform*

Discrete Fourier Transform (DFT) adalah suatu metode yang digunakan untuk mentransformasikan sinyal dari kawasan waktu ke kawasan frekuensi, formula dari DFT adalah sebagai berikut:

$$X(k) = \sum_{n=0}^{N-1} x(n) W^{\frac{k}{N}} \tag{2.3}$$

$k = 0, \dots, N-1$

dimana:

$X(k)$ = Transformasi *fourier* dari sinyal wicara

$x(n)$ = Representasi sinyal sampel n

N = Jumlah sampel dalam setiap *frame*

Formula DFT di atas menyatakan bahwa sinyal akan periodik pada setiap N [1]. Kekurangan dari DFT adalah waktu yang dibutuhkan untuk mengubah sinyal dari kawasan waktu ke kawasan frekuensi itu sangatlah lama yaitu sebesar $\mathbf{O}(N^2)$. Teknik yang digunakan untuk mendapatkan koefisien frekuensi dengan proses yang cepat adalah *Fast Fourier Transform* (FFT). Didalam penelitian ini untuk mengubah sinyal dari kawasan waktu ke kawasan frekuensi yang digunakan adalah FFT.

2.3.3.2 *Short Term Fourier Transform*

Short Term Fourier Transform (STFT) merupakan pengembangan dari FFT, dimana pada STFT sinyal diolah *frame* demi *frame* yang memiliki jumlah sampel

tertentu. *Frame* demi *frame* tersebut oleh STFT diterjemahkan ke dalam kawasan frekuensi. Dengan melakukan alihragam sinyal *frame* demi *frame*, maka posisi dari waktu terhadap frekuensi akan dengan mudah diketahui [16]. Secara matematis STFT dapat dirumuskan sebagai berikut [17]:

$$X(k,m) = \sum_{i=0}^{N-1} w(i)x(m(N-D)+i)e^{-j\frac{2\pi}{N}ik}, \quad (2.4)$$

$k = 0, \dots, N-1$

$i = 0, \dots, N-1$

$m = 0, \dots, N-1$

dimana:

$X(k,m)$ = Representasi spektral-waktu dengan indeks bingkai m diskrit dan diskrit frekuensi indeks k

N = Jumlah sampel dalam setiap *frame*

D = Panjang *overlapping* dari *frame*

w = *Hamming Window*

x = Sampel dalam *frame*

STFT berbasis nilai ambang batas, dimana nilai magnitudo yang lebih besar dari nilai ambang batas akan dianggap sebagai *voice*, sedangkan nilai magnitudo kurang dari nilai ambang batas akan dianggap sebagai *unvoiced*/*silenced*. Sinyal *voice* akan dilabelkan sebagai '1', dan sinyal *unvoiced*/*silenced* dilabelkan sebagai '0' [18].

Teknik penjendelaan (*windowing*) adalah sebuah teknik manipulasi amplitudo sinyal dengan menggunakan formula matematis yang sesuai dengan jenis windownya. Teknik *windowing* yang digunakan adalah *hamming window*. Fungsi *window* ini menghasilkan *sidelobe* level yang tidak terlalu tinggi (kurang lebih -43 dB), selain itu derau yang dihasilkanpun tidak terlalu besar. *Hamming window* dapat dirumuskan sebagai berikut [19]:

$$w[n] = 0.54 - 0.46\, cos\left(\frac{2\pi n}{N-1}\right), (0 \le n \le N-1) \quad (2.5)$$

dimana:

$w[n]$ = Nilai *window* tiap *frame*

N = Jumlah sampel dalam setiap *frame*

2.3.3.3 *Zero Crossing Rate*

Zero Crossing Rate (ZCR) merupakan fitur utama yang berguna dalam analisa sinyal wicara. Tingkat dari ZCR pada *frame* yang sempit mengisyaratkan tinggi rendahnya frekuensi sinyal pada potongan *frame* tersebut. Jika dianalogikan dengan partikel yang bergerak pada gelombang yang dibentuk oleh seutas tali, maka ZCR merupakan ukuran jumlah partikel bergerak melewati titik nol dari satu simpul tali yang positif ke simpul negatif dalam ukuran *frame* tertentu. Metode ini juga merupakan fitur temporal (analisis yang dilakukan dalam kawasan waktu) yang berguna dalam analisa tutur, mengacu pada berapa kali sampel tutur mengubah tanda dalam sebuah *frame* [7]. Metode ini dirumuskan sebagai berikut:

$$Z(m) = \frac{1}{L} \sum_{n=m-L+1}^{N} \left| \frac{sgn\big(s(n)\big) - sgn\big(s(n-1)\big)}{2} \right| \tag{2.6}$$

$$sgn\big(s(n)\big) = \begin{cases} +1, s(n) \geq 0 \\ -1, s(n) < 0 \end{cases}$$

$$m = 0, \dots, N-1$$

dimana:

sgn = Tanda bilangan pada sampel tertentu dengan ketentuan

$Z(m)$ = Nilai ZCR *frame* demi *frame*

N = Jumlah sampel dalam setiap *frame*

$1/L$ = Tingkat partikel melewati sumbu 0

2.4 PEMOTONGAN AYAT DAN KATA

Envelope dapat digunakan untuk mengurangi efek gema (reverberasi) dan derau yang disebut juga sebagai osilasi pada sinyal. Amplitudo dari osilasi bervariasi, dan bentuk variasi waktu lambat disebut sebagai *envelope*. Sinyal *envelope* ini mengandung informasi penting dari tutur. Metode untuk melakukan deteksi *envelope* salah satunya dapat menggunakan perhitungan nilai maksimum *absolute*. Cara kerja deteksi *envelope* dengan perhitungan nilai maksimum *absolute* ini memiliki konsep yang serupa dengan konvolusi, namun nilai sinyal wicara akan dikalikan dengan *frame* yang berukuran 1×1501 dan memiliki nilai *logical* 1. Nilai *envelope* yang telah didapatkan nantinya dikenakan nilai ambang batas untuk memisahkan bagian sinyal wicara dan nonwicara. Nilai *envelope* yang memiliki nilai lebih besar dari nilai

ambang batas akan dilabelkan sebagai ‘1’, sedangkan nilai *envelope* lebih kecil dari nilai ambang batas dilabelkan sebagai ‘0’. Nilai ‘0’ yang saling berdekatan dan memiliki jumlah terbanyak diasumsikan sebagai jeda antar ayat dan antar kata. Dengan demikian jika jeda antar ayat dan kata dihilangkan, maka hanya didapatkan bagian wicara.

2.5 *PRE-EMPHASIZE*

Proses penapisan sinyal wicara diperlukan setelah proses perekaman. Tujuan dari penapisan adalah untuk mendapatkan bentuk *spectral* frekuensi sinyal wicara yang lebih halus. Filter *pre-emphasize* didasari oleh hubungan *input*/*output* dalam kawasan waktu [6]. Dengan memberikan sinyal wicara $s(n)$, *pre-emphasize* dilakukan dengan filter satu tahap yang memiliki fungsi $(1 - \alpha z^{-1})$, dan koefisien penekanan, α mendekati 1. Nilai khas yang digunakan dalam penelitian adalah $\alpha = 15/16 = 0{,}9375$. Setiap sampel tutur *pre-emphasized* $s'(n)$ berasal dari sampel masukan saat ini dan setelahnya yang dirumuskan oleh filter FIR berikut [20]:

$$H(z) = 1 - \alpha z^{-1} \tag{2.7}$$

$$s'(n) = s(n) - 0{,}9375 \times s(n-1) \tag{2.8}$$

dimana:

$s(n)$ = Sinyal sebelum *pre-emphasize*

$s'(n)$ = Sinyal hasil *pre-emphasize*

Hasil *filtering pre-emphasize* menyebabkan frekuensi tinggi muncul dan frekuensi-frekuensi rendah yang lemah telah dihilangkan. Filter *pre-emphasize* ini memiliki konsep yang sama dengan *High Pass Filter* (HPF), dimana filter ini menggunakan nilai koefisien penekanan 0,9375. Maka 93,75% dari suatu sampel sinyal wicara diduga berasal dari sampel sinyal wicara sebelumnya sehingga mampu mendapatkan bentuk *spectral* frekuensi sinyal wicara yang lebih halus.

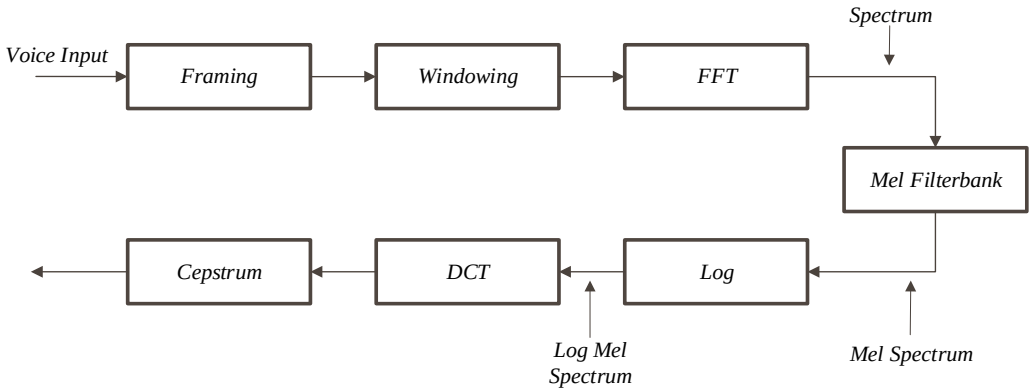
2.6 EKSTRAKSI CIRI

Ekstraksi ciri bertujuan untuk mengubah sinyal wicara ke dalam bentuk vektor yang berisi koefisien ciri dari sinyal tersebut. Koefisien ciri berisi informasi yang diperlukan untuk identifikasi tutur yang diberikan. Karena setiap tutur memiliki atribut yang unik dan berbeda yang terkandung dalam kata-kata yang diucapkan.

Selain itu, tujuan ekstraksi ciri adalah untuk mereduksi ukuran data tanpa mengubah karakteristik dari sinyal wicara dalam setiap *frame* yang dapat digunakan sebagai ciri. Ciri didapat dari mengubah bentuk sinyal wicara ke dalam bentuk baru yang dihasilkan dari kombinasi parameter matematis [6]. Sebuah metode ekstraksi ciri untuk sinyal wicara harus memenuhi kriteria tertentu: 1) metode ekstraksi ciri sinyal wicara harus sederhana; 2) metode ekstraksi ciri harus konsisten dengan waktu; 3) dan harus kebal terhadap derau. Metode ekstraksi ciri *Mel-Frequency Cepstral Coefficients* (MFCC) merupakan salah satu metode yang memenuhi kriteria tersebut dan umum digunakan untuk pengenalan wicara [21].

2.6.1 *Mel-Frequency Cepstral Coefficient*

Mel-Frequency Cepstral Coefficients (MFCC) merupakan proses pengambilan ciri yang berdasar pada transformasi *fourier* diskrit. MFCC merupakan salah satu metode ekstraksi ciri dan cara yang paling sering digunakan pada berbagai bidang area pengolahan sinyal wicara, karena dianggap cukup baik dalam mempresentasikan ciri sebuah sinyal wicara. Cara kerja MFCC didasarkan pada perbedaan frekuensi yang dapat ditangkap oleh telinga manusia sehingga mampu mempresentasikan ciri sinyal wicara sebagaimana manusia mempresentasikan [6]. Diagram proses MFCC ditunjukan pada gambar berikut:



Gambar 2.1 Diagram Proses MFCC

2.6.1.1 *Framing*

Dalam proses *framing* sinyal wicara kontinyu diblok menjadi *frame* - *frame* N sampel, dengan *frame* - *frame* saling berdekatan dengan jarak M ($M < N$). *Frame* pertama terdiri dari N sampel pertama, *frame* kedua dengan M sampel setelah *frame*

pertama dan *overlap* dengan N - M sampel. Dengan cara yang sama, *frame* ketiga dimulai dari 2M sampel setelah *frame* pertama (atau M sampel setelah *frame* kedua) dan *overlap* dengan N - 2M sampel. Proses ini berlanjut hingga semua sinyal wicara dihitung dalam satu atau banyak *frame*. Tipikal nilai yang digunakan adalah M = 100 dan N = 256 [9].

2.6.1.2 *Windowing*

Proses *windowing* yaitu proses *filtering* tiap *frame* dengan mengalikan setiap *frame* tersebut dengan fungsi *window* tertentu yang ukuranya sama dengan *frame*. *Windowing* juga digunakan untuk memastikan kelanjutan tutur dari *frame* awal hingga akhir [21]. Fungsi *window* yang baik harus menyempitkan pada bagian *main lobe* dan melebar pada bagian *side-lobe* nya. Pada dasarnya, *frame* wicara dibangun dengan fungsi *window*. Fungsi dasar *window* adalah *window rectangular*, *window rectangular* dapat dirumuskan sebagai berikut:

$$w[n] = 1, (0 \leq n \leq N - 1) \tag{2.9}$$

dimana:

$w[n]$ = Nilai *window* tiap *frame*

Ada banyak fungsi *window*, salah satunya adalah *Hamming window*. *Hamming window* merupakan salah satu metode yang memenuhi kriteria tersebut dan umum digunakan untuk pengenalan wicara.

2.6.1.3 *Fast Fourier Transform*

Fast Fourier Transform (FFT) adalah salah satu metode mentransformasi sinyal dari kawasan waktu ke kawasan frekuensi. FFT bertujuan mendekomposisi sinyal menjadi sinyal sinusoidal, terdiri atas dua unit yaitu unit *real* dan unit imajiner. FFT digunakan untuk menganalisis frekuensi, sehingga dapat mempermudah pemrosesan pengenalan wicara [6]. Waktu yang dibutuhkan untuk mengevaluasi DFT pada komputer sangat bergantung dari jumlah perkalian yang terlibat. DFT membutuhkan penggandaan $\mathbf{O}(N^2)$ sedangkan FFT membutuhkan $\mathbf{O}(N \log_2 N)$ [22]. FFT dapat dirumuskan sebagai berikut:

$$X[k] = \sum_{n=0}^{\frac{N}{2}-1} x(n) W_{\frac{N}{2}}^{k} \tag{2.10}$$

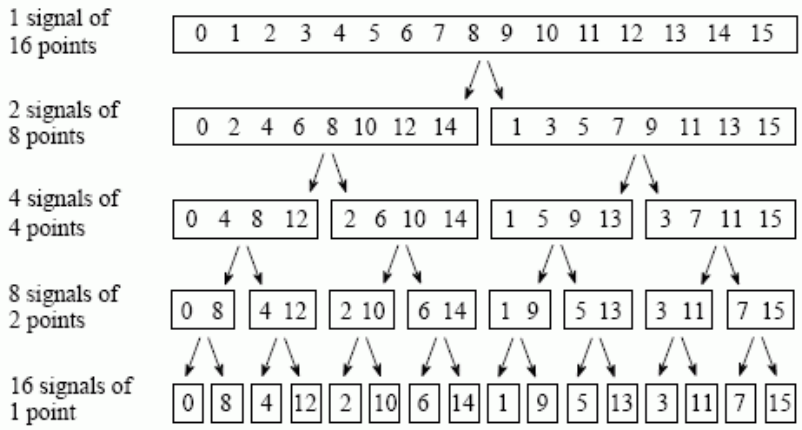
dimana:

$X[k]$ = Deretan periodik dengan nilai N

$x(n)$ = Sinyal wicara

N = Jumlah sampel dalam setiap *frame*

FFT beroperasi dengan mendekomposisi sinyal domain titik waktu N menjadi sinyal domain waktu N yang masing-masing terdiri dari satu titik. Langkah kedua adalah menghitung spektrum frekuensi N yang sesuai dengan sinyal domain waktu N ini. Selanjutnya nilai N disintesis menjadi spektrum frekuensi tunggal yang dapat dilihat seperti pada Gambar 2.2.



Gambar 2.2 Dekomposisi FFT

Dari gambar di atas sinyal titik N didekomposisi menjadi sinyal N yang masing-masing mengandung satu titik. Setiap tahap menggunakan dekomposisi *interlace*, memisahkan sampel genap dan ganjil [23].

2.6.1.4 *Mel-Frequency Wrapping*

Mel-Frequency Wrapping secara umum menggunakan *Filterbank*. *Filterbank* merupakan salah satu bentuk filter yang dilakukan dengan tujuan untuk mengetahui ukuran energi dari frekuensi *band* tertentu dalam sinyal wicara. Untuk keperluan MFCC, *filterbank* diterapkan dalam kawasan frekuensi. *Filterbank* menggunakan representasi konvolusi dalam melakukan filter terhadap sinyal. Konvolusi dapat dilakukan dengan melakukan multiplikasi antara *spectrum* sinyal dengan koefisien

filterbank, dimana nilai koefisien *filterbank* ini didapatkan dari nilai *band pass filter* yang digunakan [6] [24]. Perhitungan *Mel-filterbank* dapat dirumuskan sebagai berikut:

$$Y[f] = \sum_{j=1}^{n} S[j]H_i[j] \tag{2.11}$$

dimana:

n = Jumlah magnitudo *spectrum* ($N \in N$)

$S[j]$ = Magnitudo *spectrum* pada frekuensi j

$H_i[j]$ = Koefisien *filterbank* pada frekuensi ($1 \leq i \leq M$)

M = Jumlah *frame* dalam *filterbank*

Frekuensi dalam sebuah sinyal akan diukur manusia secara subyektif dengan menggunakan skala *Mel*. Skala *Mel-frequency* adalah skala frekuensi *linear* pada frekuensi di bawah 1000 Hz, dan merupakan skala logaritmik pada frekuensi di atas 1000 Hz. Berikut ini adalah formula untuk menghitung skala *Mel* [25]:

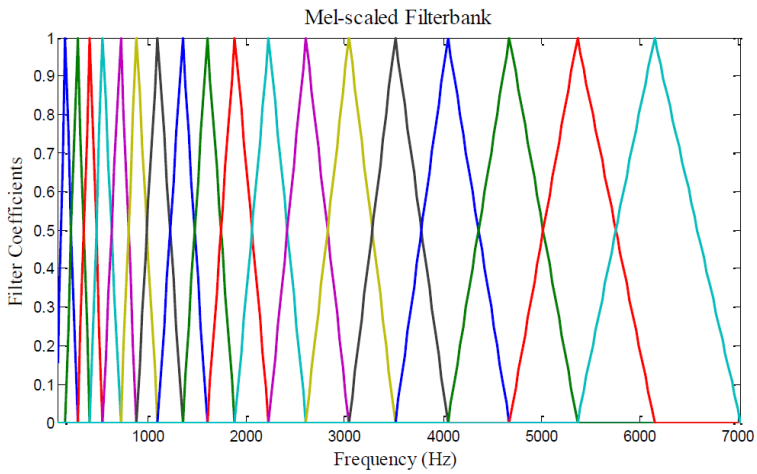
$$F(Mel) = 2595 \times log_{10}(1 + \frac{f}{700}) \tag{2.12}$$

dimana:

$F(Mel)$ = Frekuensi skala *Mel* (Hz)

f = Segmen wicara

Komponen *spectrum* yang didapatkan mendekati skala *Mel*. *Respons* masing-masing filter diberikan oleh besaran frekuensi dalam bentuk segitiga seperti pada Gambar 2.2.



Gambar 2.3 *Mel-filterbank* [26]

2.6.1.5 *Log* Frekuensi

Dengan menggunakan logaritma, efek multiplikasi besarnya FFT diubah menjadi penjumlahan. Mengurangi nilai *Mel-filterbank* dengan mengganti setiap nilai *log* dasarnya menggunakan perintah matlab "*log*" dari *segment Mel* yang telah difilter. Efek dari mengambil *log* dasar adalah mengurangi nilai-nilai *Mel-filterbank* [1].

2.6.1.6 *Discrete Cosine Transform*

Discrete Cosine Transform (DCT) adalah mendekorelasikan *Mel spectrum* sehingga menghasilkan representasi yang baik seperti sinyal wicara aslinya. Pada dasarnya konsep dari DCT sama dengan *inverse fast fourier transform* (IFFT). Berikut adalah formula yang digunakan untuk menghitung DCT [6]:

$$C_n = \sum_{n=1}^{K} (logS_k) cos\left[n\left(k-\frac{1}{2}\right)\frac{\pi}{k}\right]; n = 1,2, \dots K \qquad (2.13)$$

dimana:

C_n = Koefisien *cepstrum Mel-frequency*

S_k = Keluaran dari proses *filterbank* pada indeks *k*

K = Jumlah koefisien yang diharapkan

2.6.1.7 *Cepstrum*

Cepstrum adalah sebutan kebalikan dari spektrum. *Cepstrum* biasa digunakan untuk mendapatkan informasi dari suatu sinyal wicara yang diucapkan oleh manusia. Pada langkah terakhir ini, spektrum *log Mel* dikonversi menjadi spektrum menggunakan DCT yang merupakan nilai dari hasil *Mel-frequency* yang diubah menjadi kawasan waktu [6]. Nilai *cepstrum* digunakan dalam perhitungan jarak untuk mencari tingkat kesamaan antara data *template* dengan data latih maupun data uji.

2.7 *COSINE SIMILARITY*

Cosine similarity merupakan metode yang digunakan untuk menghitung *similarity* (tingkat kesamaan) antar dua buah objek. Secara umum perhitungan metode ini didasarkan pada vektor *space similarity measure*. Metode *cosine similarity* ini menghitung *similarity* antara dua buah objek (misalkan D1 dan D2) yang dinyatakan

dalam dua buah vektor dengan menggunakan kata kunci dari sebuah dokumen sebagai ukuran. Formula di bawah ini adalah perhitungan dari *cosine similarity* [2].

$$C = 1 - \frac{q_i \times d_i}{|q_i||d_i|} = 1 - \frac{\sum_{j=1}^{t}(d_{ij}.d_{ij})}{\sqrt{\sum_{j}^{t}1(q_{ij})^2.\sum_{j=1}^{t}(d_{ij})^2}} \quad (2.14)$$

dimana:

q_{ij} = *Term* ke-i untuk dokumen ke-j q

d_{ij} = *Term* ke-i untuk kueri ke-j (*keyword term*)

t = Jumlah istilah j pada q atau d

C = Hasil tingkat kesamaan antara dua data *cepstrum*

2.8 PENGUKURAN KUALITAS UNJUK KERJA

Pengukuran kualitas unjuk kerja digunakan sebagai pengukuran performasi dari sistem yang dibuat. Kinerja sistem klasifikasi menggambarkan seberapa baik sistem dalam mengklasifikasikan data. Untuk mengukur performasi sistem digunakan pengukuran *recall* dan *precision* dengan melibatkan nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN). Formula untuk mengukur *recall* dan *precision* adalah sebagai berikut [27]:

$$recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \times 100\% \quad (2.15)$$

$$precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \times 100\% \quad (2.16)$$

dimana:

TP = Data positif yang terklasifikasi positif oleh sistem.

FP = Data positif yang terklasifikasi negatif oleh sistem.

FN = Data negatif yang terklasifikasi positif oleh sistem.

Sedangkan TN adalah data negatif yang terklasifikasi negatif oleh sistem. Dalam pengukuran kualitas unjuk kerja pada sistem ini digunakan *recall* untuk mengukur ketepatan seluruh bacaan, sedangkan untuk mengukur ketepatan kata menggunakan *recall* dan *precision*.