Bab 2
Teori - teori

2.1 Penelitian Yang Berkaitan

Penelitian mengenai *Automatic Speech Recognition* (ASR) dimulai dengan pendekatan hibrida antara Hidden Markov Models (HMM), Connectionist Probability Estimator dan Multi Layer Perceptrons (MLPs) pada tahun 1990-an [1, 2]. Penelitian *Speech Recognition* dengan mencoba untuk mencari relasi antara audio signal dan fonem yang disebut model akustik / *Acoustic Model* [3, 4]. Berakar dari penilitian yang sudah ada, dalam penelitian Baidu pada tahun 2014, mencoba untuk membentuk sistem *end-to-end automatic speech recognition* yang disebut *Deep Speech* dengan menggunakan *Bi-directional Recurrent Neural Network* (Bi-RNN) dan *Connectionist Temporal Classification* (CTC). *Bi-directional Recurrent Neural* ini hampir sama dengan yang dilakukan oleh Hannun et al [6]. Dengan mengadopsi CTC yang diperkenalkan oleh Graves et al [5] untuk mengklasifikasi *acoustic model* yang dihasilkan oleh gabungan antara DNN dan RNN menggunakan *(Long-Short-Term Memory)* LSTM. Penelitian Baidu berhasil membuat model yang lebih simpel dengan hasil WER hingga mencapai 6.56, hasil ini lebih baik daripada Google, Bing, dan Apple. [7]

2.2 Tuna Rungu

Dalam Kamus Besar Bahasa Indonesia (KBBI) arti dari tuna rungu adalah tidak dapat mendengar; tuli. Tuli, tunarungu, atau gangguan dengar dalam kedokteran adalah kondisi fisik yang ditandai dengan penurunan atau ketidakmampuan seseorang untuk mendengarkan suara.

Tuli dalam kedokteran dibagi atas 3 jenis:

1. Tuli/Gangguan Dengar Konduktif adalah gangguan dengar yang disebabkan kelainan di telinga bagian luar dan/atau telinga bagian tengah, sedangkan saraf pendengarannya masih baik, dapat terjadi pada orang dengan infeksi telinga

tengah, infeksi telinga luar atau adanya serumen di liang telinga.

2. Tuli/Gangguan Dengar Saraf atau Sensorineural yaitu gangguan dengar akibat kerusakan saraf pendengaran, meskipun tidak ada gangguan di telinga bagian luar atau tengah.
3. Tuli/Gangguan Dengar Campuran yaitu gangguan yang merupakan campuran kedua jenis gangguan dengar di atas, selain mengalami kelainan di telinga bagian luar dan tengah juga mengalami gangguan pada saraf pendengaran.

(https://id.wikipedia.org/wiki/Ketulian)

2.3 Suara

Dalam Kamus Besar Bahasa Indonesia (KBBI) arti dari bunyi adalah 1 yang dikeluarkan dari mulut manusia (seperti pada waktu bercakap-cakap, menyanyi, tertawa, dan menangis); 2 bunyi binatang, alat perkakas, dan sebagainya; 3 ucapan; 4 bunyi bahasa (bunyi ujar); 5 sesuatu yang dianggap sebagai perkataan (untuk melahirkan pikiran, perasaan, dan sebagainya); 6 pendapat; 7 pernyataan (setuju atau tidak); 8 dukungan (dalam pemilihan).

Suara atau Bunyi adalah sebuah gelombang longitudinal yang merambat pada medium tertentu, medium bisa berupa padat, cair, dan gas. Kebanyakan suara adalah gabungan berbagai sinyal getar terdiri dari gelombang harmonis, tetapi suara murni secara teoritis dapat dijelaskan dengan kecepatan getar osilasi atau frekuensi yang diukur dalam satuan getaran *Hertz* (Hz) dan amplitudo atau kenyaringan bunyi dengan pengukuran dalam satuan tekanan suara desibel (dB).
(https://id.wikipedia.org/wiki/Bunyi)

2.4 Frekuensi

Dalam Kamus Besar Bahasa Indonesia (KBBI) arti dari frekuensi adalah 1 kekerapan: 2 jumlah pemakaian suatu unsur bahasa dalam suatu

teks atau rekaman; 3 jumlah getaran gelombang suara per detik; 4 jumlah getaran gelombang elektrik per detik pada gelombang elektromagnetik.

Frekuensi memiliki satuan *Hertz,* artinya terjadi n getaran dalam 1 detik, sehingga dapat dinotasikan rumus sebagai berikut :

$$f = \frac{n}{t} \quad \text{…(2.4.1)}$$

Dimana,

n = jumlah getaran

t = waktu dalam detik diukur saat bergetar

2.5 Amplitudo

Dalam Kamus Besar Bahasa Indonesia (KBBI) arti dari amplitudo 1 simpangan yang paling jauh dari titik keseimbangan pada getaran; 2 selisih suhu tahunan atau suhu harian; 3 jarak antara puncak gelombang bunyi dan titik rata-rata.

2.6 Desibel

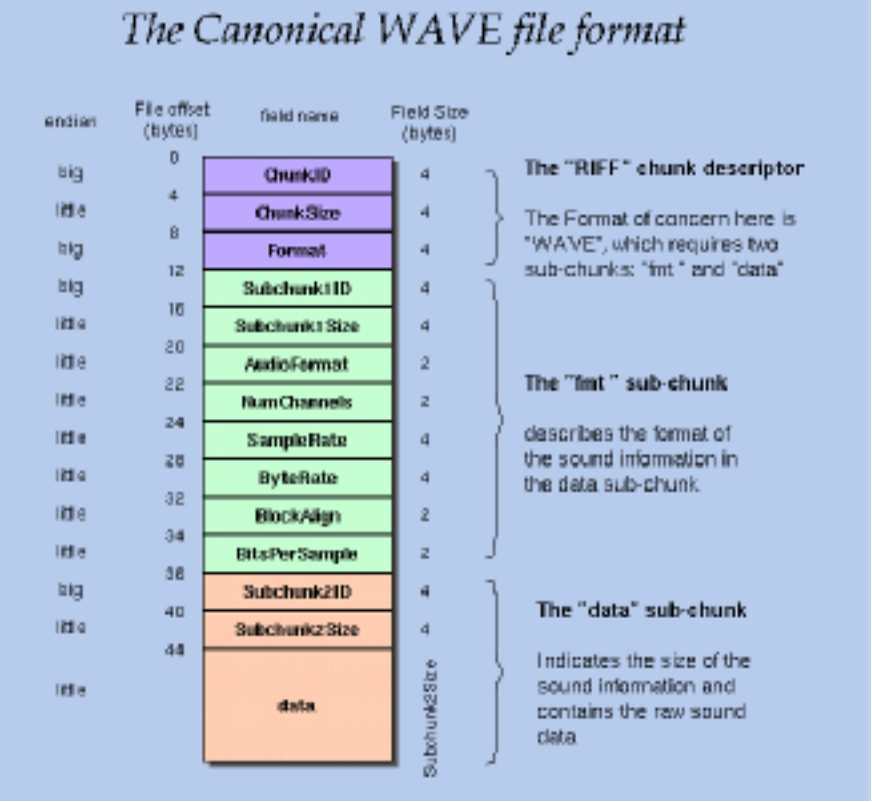
Dalam Kamus Besar Bahasa Indonesia (KBBI) arti dari desibel 1 satuan ukuran untuk mengukur kerasnya suara; satuan ukuran untuk mengukur ketajaman pendengaran; 2 satuan untuk mengukur rasio dua daya atau intensitas, atau rasio antara suatu daya dan daya acuan; satuan yang sama dengan 10 kali logaritme rasio tersebut; 3 satuan untuk menyatakan intensitas bunyi relatif pada skala dari 0 untuk rata-rata bunyi yang dapat didengar sampai 130 untuk rata-rata bunyi pada ambang pendengaran tertinggi.

2.7 WAV

WAV adalah singkatan dari istilah dalam bahasa Inggris *waveform audio format* merupakan standar format berkas audio yang dikembangkan oleh Microsoft dan IBM. WAV merupakan varian dari format *bitstream* RIFF dan mirip dengan format IFF dan AIFF yang digunakan komputer Amiga dan Macintosh. Baik WAV maupun AIFF kompatibel dengan

sistem operasi Windows dan Macintosh. Walaupun WAV dapat menampung audio dalam bentuk terkompresi, umumnya format WAV merupakan audio yang tidak terkompres.
*(*https://id.wikipedia.org/wiki/WAV*)*

WAV memiliki header file untuk mendefiniskan konfigurasi data suara dalam file tersebut. Secara umum header file juga memuat konfigurasi audio seperti *sample rate*, jumlah *channel (stereo / mono)*, *bit ber sample (bit-depth)*. Total byte yang digunakan oleh header file ini adalah 44 *bytes*. Header file tersebut adalah sebagai berikut :



Gambar 2.7.a header wav file beserta kegunaannya

2.8 Pre Emphasis

Pre-Emphasis yang merupakan proses sistem yang dirancang untuk meningkatkan sebuah band frekuensi, langkah ini memproses lewat sinyal melalui filter yang menekankan frekuensi yang lebih tinggi. Tujuan dari pre-emphasis ini adalah mengurangi noise ratio pada sinyal dan

menyeimbangkan spectrum dari suara mikrofon. Proses ini akan meningkatkan energi sinyal pada frekuensi yang lebih tinggi.

Bentuk yang paling umum digunakan dalam Pre-Emphasis adalah sebagai berikut:

$$H(z) = 1 - \alpha z^{-1} \quad \text{(2.8.1)}$$

dimana $0.9 \leq \alpha \leq 1.0$, dan $\alpha \in R$. Sehingga fungsi tersebut diatas akan dijadikan sebagai *first order differentiator,* sebagai berikut dibawah ini:

$$y[i] = x[i] - \alpha x[i-1] \quad \text{(2.8.2)}$$

dimana x[i] adalah signal pada sample ke-i, y[i] adalah hasil signal pre-emphasis sample ke-i.

2.9 Hamming Window

Hamming Window adalah salah satu fungsi matematika dimana memiliki nilai 0 jika memasuki di luar rentang interval tertentu. Fungsi *Hamming Window* ini diperkenalkan oleh Richard W. Hamming. *Window* ini mengoptimasi sehingga meminimalisir nilai maksimum pada sisi lobus. Rumus *Hamming Window* dinotasikan sebagai berikut :

$$w(n) = \alpha - \beta cos(\frac{2\pi n}{N-1}) \quad \text{(2.9.1)}$$

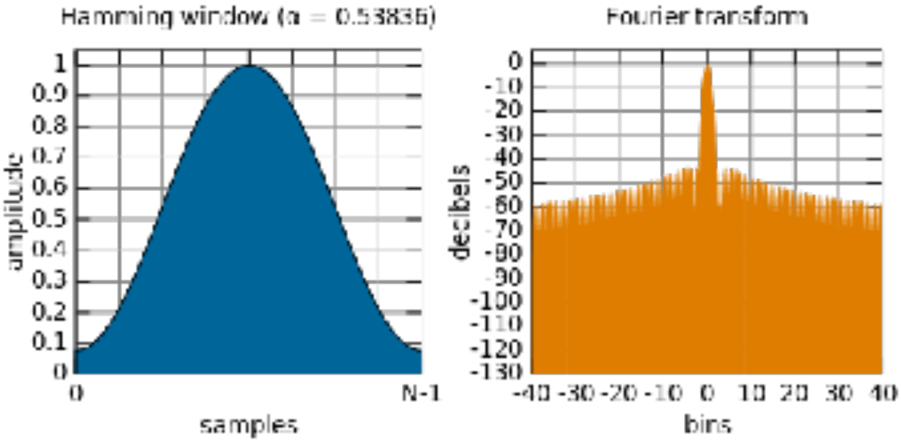
dimana,

w = hasil fungsi hamming window

$\alpha = 0.54$

$\beta = 1.0$

(https://en.wikipedia.org/wiki/Window_function)



Gambar 2.9.a grafik hamming window (sebelah kiri) hasil Fourier transform yang difungsikan terhadap Hamming Window (sebelah kanan)

2.10 Fourier Transform

Fourier transform merupakan fungsi transformasi yang mengubah dari domain waktu menjadi domain frekuensi. Hasil dari *Fourier transform* adalah bilangan kompleks, dimana nilai real merepresentasikan jumlah frekuensi, dan nilai imajiner merepresentasikan fase pada grafik sinus pada frekuensi. Secara umum, *Fourier transform* dinotasikan rumus sebagai berikut :

$$X(f) = \int_{-\infty}^{\infty} x(t) * e^{-i2\pi ft} dt \quad \text{……………………………………………..(2.10.1)}$$

dimana,

X(f) = hasil transformasi

x(t) = signal kontinu pada waktu t

e = bilangan Euler

f = frekuensi

t = waktu

Fourier transform memiliki masalah jika harus berhadapan dengan data yang diskrit sehingga perlu adanya perhitungan pendekatan diskrit ke kontinu. Sehingga didapatkan rumus *Discrete Fourier transform (DFT)* sebagai berikut :

$$X_k = \sum_{i=0}^{N-1} x_i * e^{-i2\pi kn/N}, 0 \le k \le N-1 \quad \text{…………………………………..(2.10.2)}$$

dimana,

X(f) = hasil transformasi

x(t) = signal kontinu pada waktu t

e = bilangan Euler

f = frekuensi

t = waktu

N = jumlah sample signal

k = variable frekuensi diskrit ($k = \frac{N}{2},\ k \in N$)

2.11 Periodogram

Dalam *signal processing*, periodogram digunakan untuk meng-estimasi *spectral density (power spectrum)*. Istilah ini diciptakan oleh Arthur Schuster pada tahun 1898. Periodogram untuk mengestimasi power spectrum dapat dinotasikan sebagai berikut :

$$P_i(k) = \frac{1}{N} |S_i(k)|^2 \quad \text{.....(2.11.1)}$$

dimana,

$P_i(k)$ = hasil power spectrum index ke k

$S_i(k)$ = hasil signal transformasi DFT

N = jumlah sample signal

k = variable frekuensi diskrit ($k = \frac{N}{2},\ k \in N$)

2.11 Skala Mel

Skala mel adalah perseptual skala atas tangga nada suara dengan jarak yang antar tangga nada. Referensi posisi titik antara skala dan penghitungan frekuensi normal didefiniskan sebagai berikut :

$$m = 2595 log_{10}(1 + \frac{f}{700}) \quad \text{.....(2.11.1)}$$

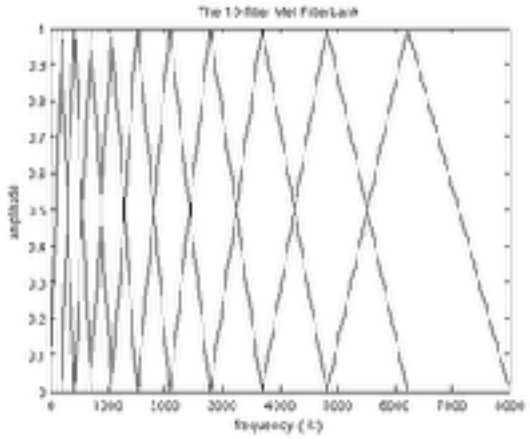
m = skala mel

f = frekuensi

Nama mel berasal dari kata melody untuk mengindikasikan skala berdasarkan perbedaan tangga nada.
(https://en.wikipedia.org/wiki/Mel_scale)

2.13 Mel Filter Bank

Filterbank adalah kumpulan daripada band-pass filter dimana akan memisahkan signal menjadi beberapa komponen, setiapnya mewakili satu frekuensi sub-band daripada signal aslinya. Mel filter bank menggunakan skala mel dalam menetapkan kumpulan band-pass filter, jumlah kumpulan biasanya 20 - 40 filterbank.



Gambar 2.13.a contoh 10-filter mel filterbank dengan 8000Hz frekuensi

2.12 Discrete Cosine Transform

Pada dasarnya konsep dari *Discrete Cosine Transform* (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. Oleh karena itu, masing-masing masukan ucapan berubah menjadi urutan vektor akustik *(acoustic model)*.

DCT dinotasikan rumus sebagai berikut :

$$C_k = \sum_{i=0}^{N-1} (log(S_i))cos[\frac{\pi}{N}(i+\frac{1}{2})k], i = 0,1,2,..N-1 \ldots\ldots \textit{(2.12.1)}$$

dimana,

C_k = hasil *Discrete Cosine Transform* pada indeks ke k

S_i = signal hasil dari proses *log filterbank energies* pada indeks ke i

N = jumlah koefisien yang diinginkan

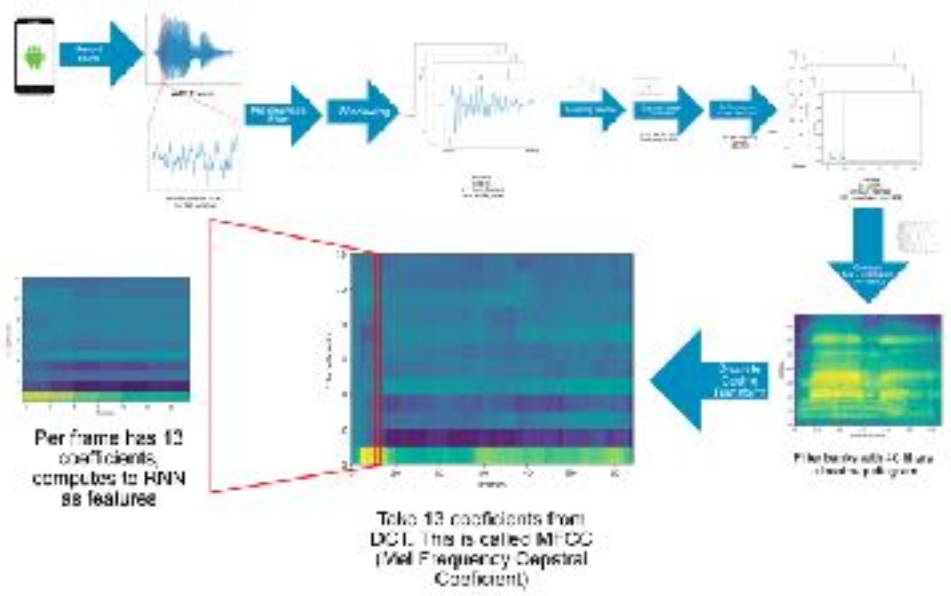
2.13 MFC (mel-frequency cepstrum)

MFC (mel-frequency cepstrum) adalah representasi *power spectrum* suara berdasarkan transformasi cosinus linear dari *log power spectrum* pada skala mel frekuensi.
(https://en.wikipedia.org/wiki/Mel-frequency_cepstrum)

2.14 MFCCs (mel-frequency cepstral coefficients)

MFCC (mel-frequency cepstral coefficients) adalah koefisien dari koleksi yang dihasilkan oleh MFC. MFCC menggunakan skala mel pada *frequency band*-nya daripada *normal cepstrum* karena *frequency band* pada mel scale meng-aproksimasi sistem respon suara dan pendengaran manusia.
(https://en.wikipedia.org/wiki/Mel-frequency_cepstrum)

Langkah - langkah dalam mengambil fitur ektraksi MFCC pada WAV :



Gambar 2.14.a flowchart / diagram alur pengambilan MFCC dari wav data file

Pada gambar 2.14.a, langkah untuk mendapatkan MFCC dimulai dengan inputan data WAV. Langka - langkah selanjutnya dirincikan sebagai berikut :

1. Dilakukan Pre-emphasis filter dengan alpha = 0.9.
2. Ambil frame sebanyak 0.02s * 16000 Hz = 320 sample, dengan *striding* 0.01s * 16000 Hz = 160 sample.
3. Dilakukan window filter menggunakan Hamming window.
4. Dilakukan Discrete Fourier transform dengan N = 512 sample.
5. Dilakukan periodogram untuk menghasilkan power spectrum setiap frame.
6. Hitung *mel-filterbank* dengan 40-*filter,* dan hitung log setiap *filter*-nya sejumlah 26 koefisien, maka dihasilkan 26 *log filterbank energies*.
7. Ubah dari domain frekuensi menjadi kembali ke domain waktu dengan Discrete Cosine Transform sejumlah n koefisien. (namun pada gambar mengambil 13 koefisien)