

PENGENALAN LAFAL HUKUM NUN MATI MENGUNAKAN HIDDEN MARKOV MODEL

Agus Jamaludin, Arief Fatchul Huda, dan Rini Sahyandari

Program Studi Matematika UIN Sunan Gunung Djati Bandung

Email: Agusjamaludin91@gmail.com

Abstract: Reading the Qur'an obligatory for every Muslim, as the word of Allah QS Al-Ankabut: 45. Additionally, the Qur'an has rules in reading. This rules relating to the pronunciation of the letter and recitation. The world of technology is growing so rapidly. One was the discovery of the speech recognition system in which a machine can understand the information conveyed by the human voice. Many methods are used in speech recognition systems: feature extraction or recognition method. Feature extraction methods are often used is Mel Frequency Cepstral Coefficient (MFCC). This method combines the linear and nonlinear because of human auditory perception are not on a linear scale in the form of frequency but measured in the form of mel-frequency scale. In the feature extraction stage, the sound signal is formed into a characteristic vectors. Therefore, in the next stage, it will be in the feature vector quantization or mapped into a codeword and collected into a set of codebook. Codebook is then used in the training process of the HMM models. In the HMM training process, the parameters of transition opportunities (A), opportunities of initialization (π), and observation opportunities (B) counted and selected the best parameters to form an optimum model and then this model will be used in the classification process. In this study, we apply the Hidden Markov Model (HMM) to recognize the *nun mati* law pronunciation. The best model in our experiment is obtained when the codebook are $M = 128$ and state $S = 6$ with 51,7% accuracy rate.

Keywords: Speech Recognition, Feature Extraction, MFCC, HMM, Vector Quantization

Abstrak: Membaca Al-Qur'an wajib bagi setiap umat muslim sebagaimana firman Allah SWT Q.S. Al-Ankabut ayat 45. Selain itu, Al-Qur'an mempunyai aturan dalam membacanya. Aturan tersebut berhubungan dengan pelafalan huruf atau makhrojul huruf dan hokum tajwid. Dunia teknologi berkembang begitu pesat. Salah satunya penemuan sistem pengenalan suara dimana sebuah mesin dapat memahami informasi yang disampaikan oleh manusia melalui suara. Banyak metode yang digunakan pada sistem pengenalan suara baik itu metode ekstraksi ciri ataupun metode pengenalannya. Metode ekstraksi ciri yang sering dipakai adalah *Mel Frequency Cepstral Coefficient* (MFCC). Metode ini menggabungkan cara linier dan non linier. Hal ini disebabkan oleh persepsi pendengaran manusia yang tidak berada pada skala linier dalam bentuk frekuensi melainkan diukur dalam bentuk skala frekuensi mel. Pada tahap ekstraksi ciri ini sinyal suara dibentuk menjadi vektor-vektor ciri, kemudian pada tahap berikutnya vektor ciri ini akan dikuantisasi atau dipetakan menjadi *codeword* dan dikumpulkan menjadi *codebook*. *Codebook* ini kemudian digunakan pada proses pelatihan model *Hidden Markov Model* (HMM). Pada proses pelatihan HMM, parameter peluang transisi (A), peluang inisialisasi (π), dan peluang observasi (B) dihitung dan di cari parameter yang palik baik sehingga membentuk sebuah model yang optimum. Kemudian model ini digunakan pada proses kualifikasi. Pada penelitian ini diterapkan metode HMM pada pengenalan lafal hokum nun mati. Model terbaik didapat pada

percobaan yang dilakukan adalah pada saat besar *codebook* $M=128$ dan banyak *state* $S=6$ dengan tingkat akurasi 51,7%.

Kata Kunci: Pengenalan Suara, Ekstraksi Ciri, Mel Frequency Cepstral Coefficient (MFCC), Hidden Markov Model, Kuantisasi Vektor

1. PENDAHULUAN

Al-Qur'an merupakan kitab umat islam dan wajib dibaca bagi setiap muslim. Sebagai mana firman Allah SWT :

أَتْلُ مَا أُوحِيَ إِلَيْكَ مِنَ الْكِتَابِ وَأَقِمِ الصَّلَاةَ ... ﴿٤٥﴾

Artinya: “*Bacalah apa yang telah diwahyukan kepadamu, yaitu Al-Kitab (Al-Qur'an) dan dirikanlah shalat*” (Q.S. Al-Ankabut :45).

Berbeda dengan bacaan lainnya, Al-Qur'an memiliki aturan dalam membacanya. Firman Allah SWT :

أَوْزِدْ عَلَيْهِ وَرَتِّلِ الْقُرْآنَ تَرْتِيلاً ﴿١﴾

Artinya: “*Atau lebih dari seperdua itu, dan bacalah Al-Qur'an itu dengan tartil*” (Q.S. Al-Muzzammil:4).

Pada ayat di atas Allah SWT memerintahkan setiap muslim untuk membaca Al-Qur'an dengan tartil. Secara umum tartil dapat diartikan perlahan-lahan dan tidak tergesa-gesa. Ilmu yang mempelajari aturan dalam membaca Al-Qur'an adalah tajwid. Banyak orang yang belum menerapkan tajwid dalam membaca Al-qur'an padahal menurut Imam Ibnu al-Jazari membaca Al-qur'an tidak dengan tajwidnya itu adalah dosa karena Allah SWT menurunkan Al-Qur'an dengan tajwidnya.

Pada jaman sekarang ini teknologi berkembang dengan pesat. Teknologi banyak membantu manusia dalam menyelesaikan masalah. Salah satu nya adalah teknologi pengenalan suara yang sudah banyak digunakan orang untuk menggali informasi dari suara. Untuk mendapatkan sistem pengenalan suara yang baik dibutuhkan metode ekstraksi ciri dan klasifikasi yang baik. Suara merupakan sebuah data yang berorientasi pada waktu. *Hidden Markov Model* (HMM) merupakan proses probabilistik yang juga berorientasi pada waktu dan bisa digunakan untuk klasifikasi. HMM terdapat pada hampir semua sistem pengenalan suara dan merupakan metode yang baik dalam klasifikasi sinyal suara [1]. Berhubungan dengan kasus pada pembacaan Al-Qur'an teknologi pengenalan suara bisa digunakan untuk mengenali pelafalan huruf hijaiyah dan hukum-hukum tajwid. Pada makalah ini HMM akan digunakan sebagai metode untuk mengklasifikasi atau mengenali lafal hukum nun mati.

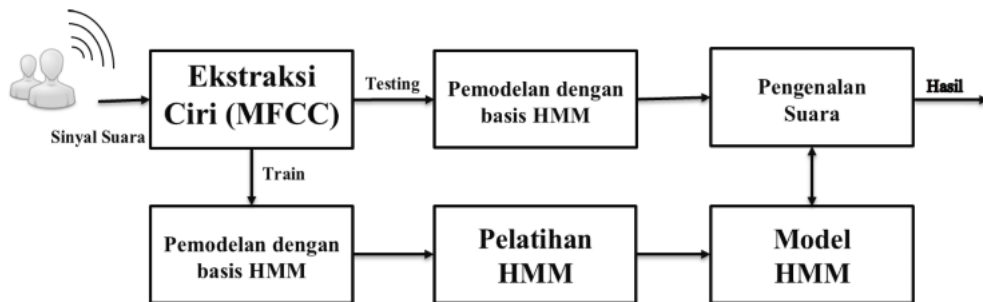
2. PENGENALAN SUARA

Pengenalan sinyal suara merupakan suatu proses dimana suara yang dihasilkan oleh pembicara direkam oleh computer, kemudian sinyal suara yang berbentuk sinyal analog dirubah menjadi sinyal digital agar bisa diproses kemudian di indentifikasi. Sinya suara

Pengenalan Lafal Hukum Nun Mati menggunakan Hidden Markov Model

biasanya diidentifikasi apa yang diucapkan yang biasa disebut *speech recognition* atau diidentifikasi siapa yang mengucapkan yang biasa disebut *speaker recognition*.

Pada penelitian ini penulis mencoba untuk mengidentifikasi apa yang diucapkan. Ucapan yang dikenal adalah pelafalan hukum nun mati. Berikut adalah alur dari sistem yang dibuat:

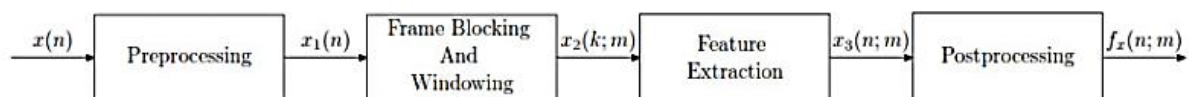


Gambar 1. Desain Sistem

Setiap data masukan diekstraksi ciri menggunakan ekstraksi ciri MFCC. Data penelitian akan masuk ketahap pelatihan. Data dimodelkan dengan basis HMM yaitu data masukan dikuantisasi berdasarkan *codebook* yang dibuat sebelumnya. Kemudian data pelatihan dilatih menggunakan algoritma *baum-welch*. Hasil pelatihan dari HMM ini adalah model-model HMM. Model-model ini disimpan sebagai *dictionary* pada proses pengenalan. Data yang merupakan data pengujian akan masuk pada proses pengujian dan dikuantisasi berdasarkan *codebook*. Kemudian masuk proses pengenalan suara menggunakan algoritma *forward* berdasarkan model yang dihasilkan dari proses pelatihan. Hasil dari pengenalan ini adalah tulisan berdasarkan kelas yang dikenali oleh sistem.

3. EKSTRAKSI FITUR

Ekstraksi fitur merupakan suatu proses untuk mengekstraksi informasi inti dari sebuah sinyal suara. Ekstraksi fitur ini bertujuan agar sinyal mudah dikenali pada saat proses pengenalan suara oleh sistem. Langkah-langkah utama dari ekstraksi fitur adalah *preprocessing*, *frame blocking and windowing*, ekstraksi fitur, dan *post processing* [2].



Gambar 2. Bagan proses utama ekstraksi fitur.

Dari sampel sinyal suara $x(n)$ menghasilkan vektor ciri $f(n; m)$, dengan $m = 0, 1, \dots, M - 1$ dan $n = 0, 1, \dots, N - 1$, artinya terdapat M vektor dengan ukuran masing-masing vektor sebesar N.

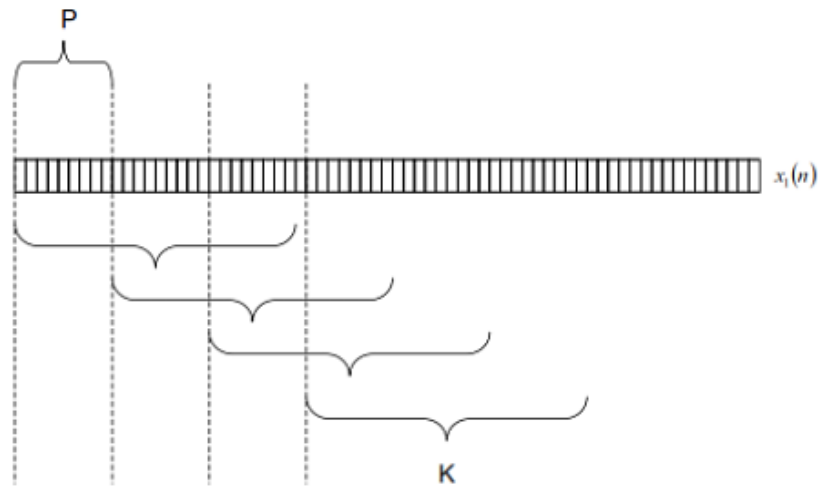
3.1. PREPROCESSING

Preprocessing adalah langkah pertama untuk membuat vektor ciri. Tujuan dari *preprocessing* adalah untuk memodifikasi sinyal suara, $x(n)$, sehingga akan lebih mudah

untuk dianalisis. Hal yang biasanya dilakukan pada *preprocessing* diantaranya, *noise cancelling*, *preemphasis*, *Voice Activation Detection (VAD)*

3.2. FRAME BLOCKING DAN WINDOWING

Selanjutnya adalah membagi sinyal suara ke dalam frame-frame suara dan dilakukan *windowing* untuk setiap framenya. Setiap frame mempunyai panjang K sampel, dengan frames yang berdekatan dipisahkan oleh P sampel [2].



Gambar 3. *Frame Blocking*

Biasanya nilai untuk K dan P adalah 320 sampel dan 100 sampel koresponden dengan 20 ms frame, dipisahkan oleh 6.25 ms dengan sampling rate dari suara 16 KHz [2]. Selanjutnya yang harus dilakukan adalah *windowing* setiap frame dengan tujuan agar mengurangi diskontinuitas sinyal di kedua ujung blok. *Windowing* yang biasa digunakan adalah *Hamming Window* yang dihitung sebagai berikut [2]:

$$w(k) = 0.54 - 0.46 \cos\left(\frac{2\pi k}{K-1}\right), \quad (1)$$

Maka hasil dari *windowing*-nya adalah:

$$x_2(k;m) = x_1(k;m) \cdot w(k), \quad (2)$$

dengan $x_1(k;m)$ adalah sinyal suara sebelum *windowing*, dengan k adalah panjang frame dan m adalah frame ke- m .

3.3. MEL FILTER BANK

Bagian ini adalah satu bagian yang paling penting yaitu untuk mendapatkan informasi yang relevan dari blok ucapan. Banyak metode yang digunakan pada tahap ini. Namun, pada penelitian ini digunakan metode *Mel Frequency Cepstral Coefficient (MFCC)*. Pada metode MFCC sinyal yang telah melalui proses *frame blocking* dan *windowing* kemudian dihitung koefisien mel cepstrum dari sinyal tersebut dengan menggunakan rumus sebagai berikut [2]:

Pengenalan Lafal Hukum Nun Mati menggunakan Hidden Markov Model

$$m_k = \sum_{n=1}^{N-1} |x_2(n; m)| H_k^{mel}(n), \quad (3)$$

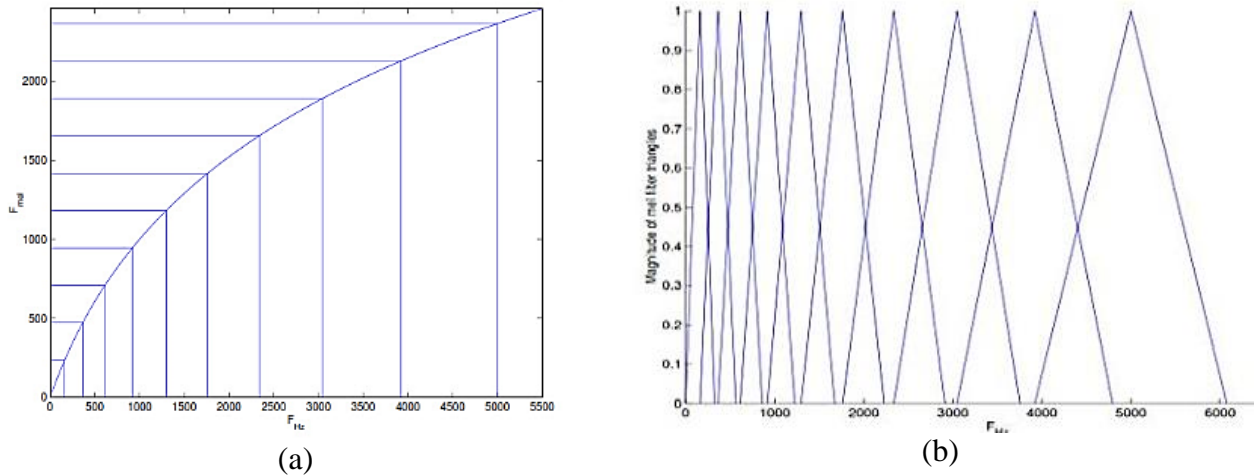
dengan N merupakan panjang frame, m_k adalah koefisien mel cepstrum ke- k .

$H_k^{mel}(n)$ adalah sebuah filter triangular. Untuk mendapatkan filter triangular kita harus mengubah suara dengan frekuensi Hz menjadi frekuensi mel (F_{mel}). Persepsi manusia untuk sinyal suara tidak mengikuti skala linier. Jadi untuk suara dengan frekuensi yang sebenarnya diukur dalam Hz. Skala frekuensi mel merupakan frekuensi linier dengan batas bawah 1000 Hz dan logaritmik dengan batas atas 1000 Hz. 1 Hz biasanya dikatakan juga 1000 mels. Sebagai transformasi frekuensi nonlinier, rumus berikut digunakan [2];

$$F_{mel} = 2595 \cdot \log_{10} \left(1 + \frac{F \text{ Hz}}{700} \right), \quad (4)$$

Setelah frekuensi mel (Gambar 6) didapat kemudian dibagi menjadi K bagian. Setelah membaginya menjadi K bagian sekarang kita bisa menemukan triangular filter dengan menghitung F_{Hz} sebagai berikut [2]:

$$F_{Hz} = 700 \cdot 10^{\frac{F_{mel}}{2595}} - 1, \quad (5)$$



Gambar 4. (a) Skala Mel, (b) Mel Scale filter bank.

Setelah koefisien spektrum mel dihitung tahap selanjutnya adalah *Discrete Cosine Transform* (DCT). Pada tahap ini akan dikonversi spektrum mel ke dalam domain waktu. DCT ini sama dengan IFFT atau invers dari FFT. Hasilnya adalah :

$$c_s(n; m) = \sum_{k=0}^{N-1} \alpha_k \cdot \log(m_k) \cdot \cos \left(\frac{\pi(2n+1)k}{2N} \right), \quad n = 0, 1, \dots, N-1 \quad (6)$$

dengan

$$\alpha_k = \begin{cases} \alpha_0 = \sqrt{\frac{1}{n}} \\ \alpha_k = \sqrt{\frac{2}{n}}, 1 \leq k \leq N-1 \end{cases} \quad (7)$$

3.4. POSTPROCESSING

Langkah terakhir dari proses ekstraksi fitur adalah *postprocessing*. Pada proses ini biasanya suara dinormalisasi, yaitu dengan mengurangi sampel suara dengan rata-rata ($f_{\bar{\mu}}(n)$) dari 2] seluruh sampel suara, atau bisa dirumuskan sebagai berikut [2]:

$$f_{\bar{\mu}}(n) = \frac{1}{M} \sum_{m=0}^{M-1} x_3(n; m), \quad (8)$$

dengan M adalah banyak frame untuk cepstrum ke- n . Sampel suara dinormalisasi dengan cara [2]:

$$f_x(n; m) = x_3(n; m) - f_{\bar{\mu}}(n), \quad (9)$$

dengan $x_3(n; m)$ adalah cepstrum sebelum normalisasi hasil dari DCT dan $f_x(n; m)$ adalah cepstrum hasil normalisasi.

4. HIDDEN MARKOV MODEL

Markov model adalah salah satu metode probabilistik yang digunakan untuk memprediksi kejadian berikutnya berdasarkan kejadian sebelumnya atau yang biasa disebut dengan rantai markov dimana kejadian sebelumnya teramati. Misalkan, terdapat N kejadian (q) yang berbeda dengan setiap kejadiannya memiliki indeks waktu $t = 1, 2, \dots, N$ atau kejadian-kejadian itu biasa dinotasikan dengan q_t . Kejadian-kejadian tersebut dihubungkan dengan peluang transisi. Rantai markov didefinisikan sebagai berikut :

$$P(q_t = j | q_{t-1} = i, q_{t-2} = k, \dots) = P(q_t = j | q_{t-1} = i), \quad (10)$$

Persamaan di atas biasa disebut juga dengan *first order markov assumption*. Peluang transisi biasanya dinotasikan dengan a_{ij} , atau bisa ditulis juga sebagai berikut :

$$a_{ij} = P(q_t = j | q_{t-1} = i), i \leq i, j \leq N, \quad (11)$$

dimana a_{ij} memenuhi sifat peluang, yaitu :

$$a_{ij} \geq 0 \forall i, j, \quad (12)$$

$$\sum_{j=1}^N a_{ij} = 1, \forall i, \quad (13)$$

Pengenalan Lafal Hukum Nun Mati menggunakan Hidden Markov Model

atau bisa dituliskan dalam sebuah matriks :

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{bmatrix} \quad (14)$$

Selain itu, pada rantai markov terdapat kejadian awal yang disebut dengan *initial state*. Kejadian awal ini biasa didefinisikan sebagai suatu vektor :

$$\pi = \begin{bmatrix} \pi_1 = P(q_1 = 1) \\ \vdots \\ \pi_N = P(q_1 = N) \end{bmatrix} \quad (15)$$

Contoh penerapan metode ini adalah pada kasus cuaca. Misalkan suatu hari terjadi cuaca yang bisa kita amati seperti pada gambar 6.

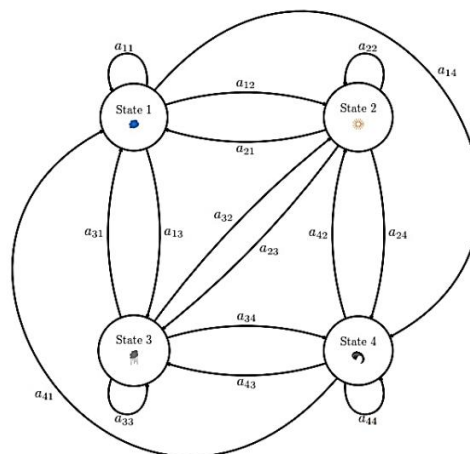
Dengan model pada gambar 6 kita bisa memprediksi urutan cuaca yang terjadi. Misalkan, terdapat urutan kejadian pada 6 hari berturut-turut sebagai berikut :

$$O = (\text{cerah, hujan, cerah, berangin, berawan, berawan}) = (2, 3, 2, 4, 1, 1)$$

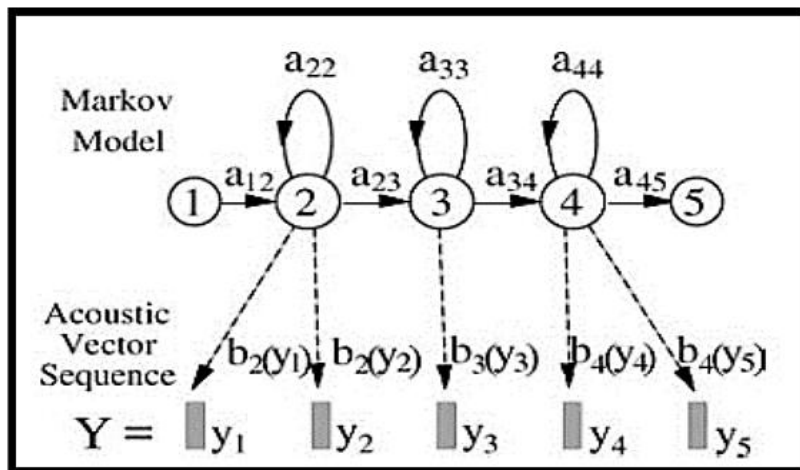
Maka peluang cuaca pada hari ke-7 sebagai berikut :

$$\begin{aligned} P(O|A, \pi) &= P(2,3,2,4,1,1|A, \pi) \\ &= P(2) \cdot P(3|2) \cdot P(2|3) \cdot P(4|2) \cdot P(1|4) \cdot P(1|1) \\ &= \pi_2 \cdot a_{23} \cdot a_{32} \cdot a_{24} \cdot a_{41} \cdot a_{11} \end{aligned}$$

Hidden Markov Model (HMM) atau model markov tersembunyi adalah model markov dengan asumsi bahwa kita tidak mengetahui kejadian yang terjadi tapi bisa memprediksi dengan melihat fenomena yang terjadi. Contohnya adalah pada kasus cuaca sebelumnya, ketika kita berada di dalam ruangan kita tidak bisa mengetahui cuaca diluar. Tetapi, kita bisa mengetahuinya dengan melihat fenomena yang terjadi. Misalnya, dengan melihat orang yang masuk ke ruangan apakah memakai payung atau tidak. Gambar 7 adalah arsitektur dari HMM.



Gambar 6. Model markov pada kejadian cuaca dengan *state*: 1: berawan, 2: cerah, 3: hujan 4: berangin [5].



Gambar 7. Arsitektur HMM [8].

B merupakan peluang kemunculan peubah yang terobservasi pada suatu *state*. Sedangkan Y merepresentasikan notasi untuk nilai teramati pada waktu t . Suatu HMM dapat dilambangkan dengan :

$$\lambda = (A, B, \pi), \quad (16)$$

HMM memiliki tiga masalah mendasar, yaitu [6]:

1. Mencari $P(O|\lambda)$, peluang dari barisan observasi $O = \{o_1, o_2, \dots, o_T\}$ jika diberikan HMM; $\lambda = (A, B, \pi)$. Peluang ini dapat ditentukan secara induksi dengan menggunakan algoritma *forward*.
2. Mencari barisan *state* yang optimal $Q^* = \{Q_1^*, \dots, Q_T^*\}$ jika diberikan barisan observasi $O = \{o_1, o_2, \dots, o_T\}$ dan model $\lambda = (A, B, \pi)$. Barisan *state* terbaik yang akan ditentukan yaitu berupa lintasan tunggal yang mungkin. Untuk menyelesaikan masalah ini digunakan algoritma viterbi.

Mengatur parameter model $\lambda = (A, B, \pi)$ agar $P(O|\lambda)$ maksimum jika diberikan sebuah HMM, λ , dan barisan obserbasi O. Untuk menyelesaikan masalah ini digunakan algoritma *Baum-Welch*.

5. KUANTISASI VEKTOR

Kuantisasi adalah proses aproksimasi sinyal amplitudo kontinyu dengan simbol diskrit. Kuantisasi dari nilai sinyal tunggal atau parameter disebut kuantisasi skalar. Sebaliknya, kuantisasi gabungan beberapa nilai sinyal atau parameter disebut kuantisasi vector. Sebuah *quantizer* vektor digambarkan oleh *codebook*, yang merupakan himpunan vektor prototipe tetap atau vektor reproduksi. Masing-masing vektor prototipe ini juga disebut sebagai *codeword*. Untuk melakukan proses kuantisasi, vektor masukan adalah pencocokkan ulang *codeword* di *codebook* menggunakan beberapa ukuran distorsi. Vektor masukan kemudian digantikan oleh indeks dari *codeword* dengan distorsi terkecil. Oleh karena itu, deskripsi dari proses kuantisasi vektor meliputi [10]:

- a. Mengukur distorsi,

Pengenalan Lafal Hukum Nun Mati menggunakan Hidden Markov Model

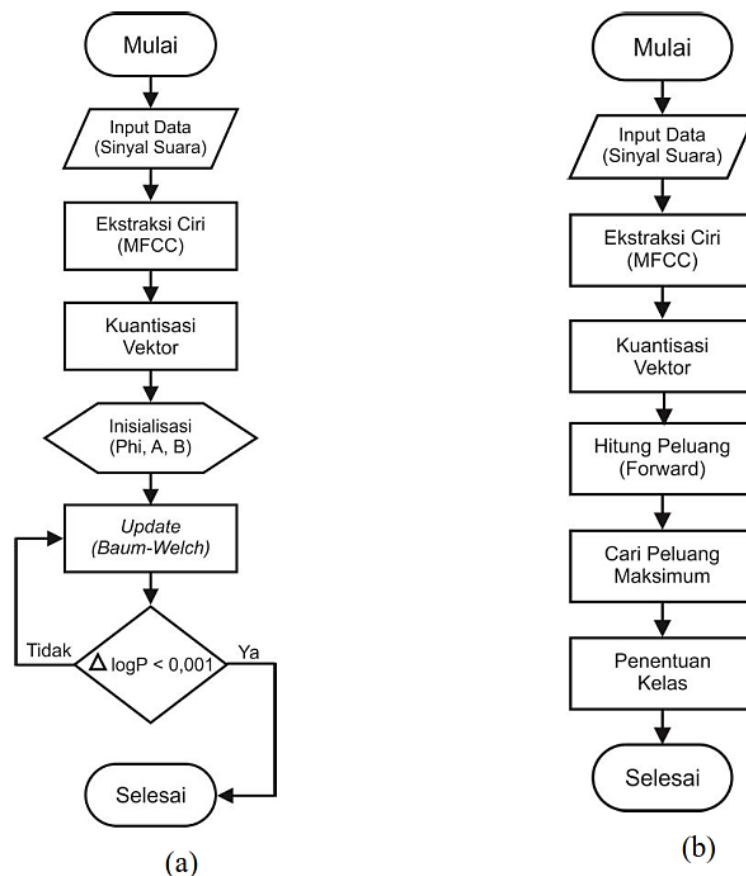
- b. Generasi setiap *codeword* di *codebook*.

Karena vektor diganti dengan indeks dari *codeword* dengan distorsi terkecil, data yang dikirimkan dapat dipulihkan hanya dengan mengganti urutan indeks kode dengan urutan *codeword* yang sesuai. Ini pasti menyebabkan distorsi antara data asli dan data yang ditransmisikan. Bagaimana untuk meminimalkan distorsi demikian tujuan utama vektor kuantisasi. Bagian ini menjelaskan beberapa tindakan distorsi yang paling umum. Ada beberapa algoritma yang digunakan pada proses vektor kuantisasi, yaitu k-means dan LBG. Pada percobaan ini digunakan algoritma LBG.

6. HASIL DAN ANALISIS PERCOBAAN

Pada penelitian ini dilakukan percobaan pengenalan lafal hukum nun mati dengan menggunakan data suara dari 5 orang pembicara. Setiap pembicara mengucapkan 28 lafahz yang merupakan potongan ayat dalam Al-Qur'an yang mengandung hukum nun mati. 1 lafahz diucapkan sebanyak 5 kali. *Sample rate* yang dipilih adalah 11025 Hz. Format data suara yang hasil perekaman adalah wav.

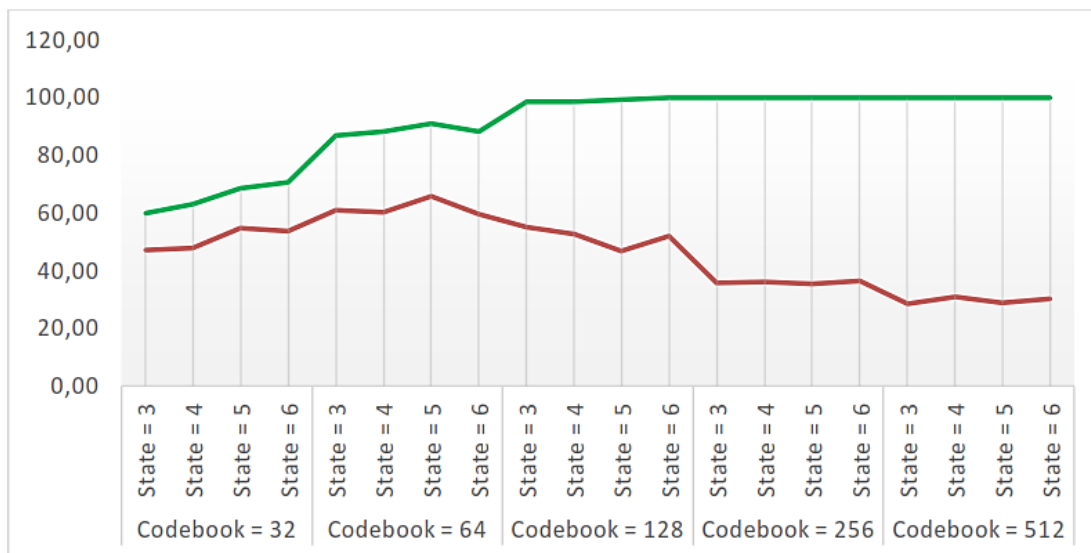
Data penelitian dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian. Data pelatihan digunakan untuk melatih sistem, sedangkan data pengujian digunakan untuk menguji tingkat pengenalan sistem. Adapun persentase pembagian data pelatihan dan data pengujian ditentukan berdasarkan hasil percobaan dengan hasil terbaik.



Gambar 8. (a) Alur proses pelatihan HMM, (b) alur proses pengujian HMM.

Gambar 8 adalah alur proses pelatihan dan pengujian HMM dan hasil dari percobaan ditunjukkan oleh Gambar 9.

Akurasi tertinggi didapat pada saat besar *codebook* $M=64$, terlihat jelas pada grafik bahwa untuk setiap *state* yang diambil menghasilkan tingkat pengenalan yang baik. Namun, kembali lagi pada hasil dari proses pelatihan bahwa untuk model dengan besar *codebook* $M=64$ memiliki rata-rata tingkat akurasi 90% sehingga modelnya masih diragukan. Jika kita perhatikan lagi akurasi pengenalan data pengujian yang paling tinggi dengan tingkat akurasi pengenalan data training 100% terdapat pada saat besar *codebook* $M=128$ dengan jumlah *state* $S=6$. Tingkat akurasi pada saat besar *codebook* $M=128$ dan jumlah *state* $S=6$ adalah 51,79%.



Gambar 9. Grafik rata-rata akurasi pengenalan data pengujian dan data pelatihan

7. KESIMPULAN

Hasil percobaan menunjukkan bahwa *codebook* untuk pelatihan HMM didapat saat besar *codebook* $M=128$ dengan jumlah *state* $S=6$ karena memiliki tingkat akurasi pengenalan data testing tertinggi dengan akurasi pengenalan data pelatihan 100%, yaitu 51,79%. Dengan tingkat akurasi pengenalan data pengujian 51,79% ini, sistem yang dibuat bisa dikatakan kurang baik atau sistem tidak bisa dengan tepat mengenali.

REFERENSI

- [1] Gales, Mark dan Steve Young. (2007). *The Application of Hidden Markov Models in Speech Recognition*. Hanover, USA: NOW.
- [2] Nillson, Mikael, Ejnarrson, Marcus. 2002. *Speech Recognition Using Hidden Markov Model Performance evaluation in noisy environment*. Blekinge Tekniska Hogskola, Sweden.
- [3] Firdaniza dkk. (2006). *Hidden Markov Model*.
- [4] Huang, Xuedong dkk. (2000). *Spoken Language Processing, A Guide to Theory, Algorithm, and System Development*. Redmond, WA: Prentice Hall.