MATICS : Jurnal Ilmu Komputer dan Teknologi Informasi ISSN : 1978-161X(p); 2477-2550(e)
Volume 9, No. 1(2017), pp 33-40 DOI : 10.18860/mat.v8i1.3482
Received : January25th 2016; **Accepted** : March 9th 2016 ; **Avalaible Online** : April 5th 2016

Metode *Mel Frequency Cepstral Coeffisients* (MFCC) Pada klasifikasi *Hidden Markov Model* (HMM) Untuk Kata *Arabic* pada Penutur Indonesia

Totok Chamidy

Abstract— Speech recognition is a system to transform the spoken word into text. Human voice signals have a very high of variability. Speech signals in the different pronunciation text, also resulting in distinctive speech patterns. This, furthermore, happens if the text is spoken by a speaker who is not the mother tongue of the speakers. For example, text Arabic words spoken by Indonesian speaker. In this study, Mel Frequency cepstral Coeffisients (MFCC) feature extraction techniques explored for voice recognition of the Arabic words for Indonesian speakers with data training using Arabian native speakers. Furthermore, features that have been extracted, classified using Hidden Markov Model (HMM). HMM is one of the sound modeling where the voice signal is analyzed and searched the maximum probability value that can be recognized, from the modeling results will be obtained parameters are then used in the word recognition process. Recognized word is a word that has the maximum suitability. The system produces an accuracy by an average of 83.1% for test data sampling frequency of 8,000 Hz, 82.3% for test data sampling frequency of 22050 Hz, 82.2% for test data sampling frequency of 44100 Hz.

***Index Terms*— Indonesian dialect, Hidden Markov Model, Mel Frequency cepstral Coeffisients**

I. INTRODUCTION

Al-Qur'an menggunakan Bahasa Arab yang dalam perkembangan agama Islam merupakan hal yang diperlukan karena Bahasa Arab dapat menampung makna al-Qur'an secara keseluruhan. Pada periode kepemimpinan Ustman bin Affan r.a. mulai semakin diketahui perbedaan dalam Qiro'ah sehingga dapat mengubah tulisan. Dalam perkembangannya, banyak sekali yang memeluk agama Islam, sehingga mengakibatkan perbedaan Qiro'ah masing-masing dalam pengucapan bahasa Arab dalam Al-Qur'an. Perbedaan qiraat ini diakibatkan oleh perbedaan logat atau dialek. Pada periode kepemimpinan Ustman bin Affan r.a. mulai dibukukan Al-Qur'an dengan qiraat berbasis pada quraisy pada saat itu [13].

Pada periode yang sangat jauh dari khalifah Ustman bin affan r.a. perkembangan agama Islam begitu pesatnya di seluruh dunia. Perkembangan ini mengakibatkan sangat beragamnya dialek bahasa arab yang bukan merupakan bahasa ibu pada setiap daerah masing masing, demikian juga dengan masyarakat Indonesia. Masyarakat Indonesia mempunyai dialek sendiri dalam pengucapan bahasa arab. Perkembangan teknologi komputer pada pengenalan ucapan saat ini membantu manusia untuk memeriksa kesesuaian dengan pengucapan bahasa arab dengan penutur Indonesia degan penutur aslinya. Pada pengenalan ucapan, lebih dititik-beratkan pada ekstraksi dari beberapa bagian informasi pesan yang di dalamnya terdapat teks yang diucapkan dalam bentuk lisan. Teks ini mengandung unit-unit linguistik terkecil yang disebut sebagai fonem yang dikenali melalui sinyal suara.

Sinyal suara manusia mempunyai tingkat variabilitas yang sangat tinggi. Sinyal suara yang dalam pengucapannya mengucapkan teks yang berbeda-beda, menghasilkan pola ucapan yang berbeda-beda pula. Hal ini juga terjadi jika pengucapan teks suatu bahasa diucapkan oleh orang yang bukan merupakan bahasa ibu orang tersebut. Misalnya teks kata arab diucapkan oleh orang Indonesia. Dalam penelitian ini, fitur teknik ekstraksi *Mel Frequency Cepstral Coeffisients* (MFCC) dieksplorasi untuk mendapatkan nilai kesesuaian pada penutur Indonesia terhadap penutur aslinya.

Pada penelitian ini, dilakukan untuk mendapatkan tingkat kesesuaian metode yang diterapkan pada masukan sinyal suara pengucapan kata *arabic* pada

Manuscript received March 22, 2016 This work was supported in part by Informatics Engineering Department of Maulana Malik Ibrahim Islamic State University

Totok Chamidy is with the Infornatic Engineering Departement of UIN Maliki , Malang, Indonesia (to2k2013@ti.uin-malang.ac.id)

penutur Indonesia. Selanjutnya, fitur yang telah diekstraksi, diklasifikasi menggunakan *Hidden Markov Model* (HMM).

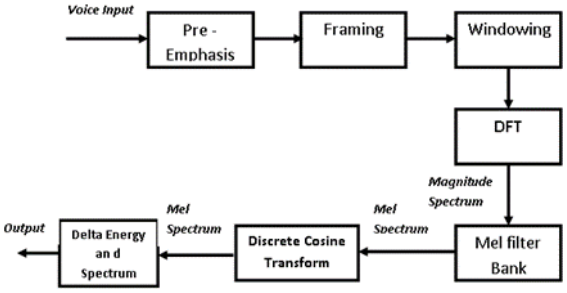
Hidden Markov Model (HMM) adalah metodologi statistik untuk pengenalan suara otomatis. HMM telah dicoba dan dibuktikan pada berbagai macam aplikasi. Parameter model HMM menggambarkan ciri atau perilaku ucapan segmen ucapan. Banyak algoritma heuristik dikembangkan dalam kerangka untuk mengoptimalkan model parameter dalam menggambarkan urutan yang terbaik dalam pengklasifikasian. Penelitian ini adalah untuk mengembangkan algoritma pengenalan suara dengan sistem yang telah ada menggunakan algoritma HMM dengan menggunakan MFCC.

II. STUDI PUSTAKA

2.1. Ekstraksi **Ciri** (Feature Extraction)

Ekstraksi ciri suara adalah untuk mengubah gelombang suara menjadi beberapa tipe representasi parametrik yang dapat diproses. Ada banyak cara untuk merepresentasikan suara secara parametris sehingga dapat diproses lebih lanjut. Salah Satunya adalah menggunakan *Mel Frequency Cepstral Coefficients (MFCC).*

Mel Frequency Cepstral Coefficients (MFCC) merupakan koefisien yang merepresentasikan audio. Metode ini diperkenalkan oleh Davis dan Mermelstein di tahun 1980-an. Ekstraksi ciri dalam proses ini ditandai dengan pengubahan data suara menjadi data citra berupa spektrum gelombang. Kebanyakan sistem pengenal ucapan saat ini menggunakan MFCC sebagai *feature* karena sistem pengenal ucapan menjadi lebih presisi dalam berbagai kondisi. Gambar 1 menunjukkan diagram blok ekstraksi ciri menggunakan metode MFCC.



Gambar 1. Ekstraksi menggunakan metode MFCC

Pre-emphasis

Pada langkah ini dilakukan filtering terhadap sinyal menggunakan FIR filter orde satu untuk meratakan spektral sinyal tersebut. Proses ini mencakup penambahan energi suara pada frekuensi tinggi. Secara matematis, dapat dirumuskan dalam bentuk berikut [6]:

$sp(n) = s(n) – 0.97\ s(n − 1)$

di mana sp(n) adalah sinyal yang ditekan, sedangkan s(n) adalah sinyal terdigitasi. Koefisien dengan nilai 0.97 menunjukkan sinyal yang diekstrak merupakan 97% sinyal aslinya

Frame Blocking (Tracking)

Pada langkah ini sinyal ucapan yang telah ter-emphasis dibagi menjadi beberapa frame (bingkai) dengan masing-masing frame memuat N sampel sinyal dan frame yang saling yang berdekatan dipisahkan sejauh M sample. Menyusun sinyal ke dalam bingkai yang lebih pendek. Panjang frame yang membagi setiap sample menjadi beberapa frame berdasarkan waktu terletak di antara 20 hingga 40 ms . Dengan asumsi bahwa frekuensi suara 18 kHz, maka sampel yang akan diekstrak adalah 0,025 detik * 18.000 Hz = 450 sampel.
Windowing

Windowing (Penjendelaan) merupakan proses pembobotan terhadap setiap frame yang telah dibentuk pada langkah sebelumnya menggunakan fungsi Window. Ada dua fungsi window yang biasa digunakan, yaitu Rectangular Window dan Hamming Window. Rectangular Window yang didefinisikan sebagai:

$$w_n = \begin{cases} 1 & 0 \le n < N \\ 0 & lainnya \end{cases}$$

Fungsi window ini menghasilkan sinyal yang diskontinyu. Salah satu cara untuk menghindari diskontinyu pada ujung window adalah dengan meruncingkan sinyal menjadi nol atau dekat dengan nol sehingga dapat mengurangi kesalahan.

Fungsi Hamming digunakan seperti bentuk jendela dengan mempertimbangkan blok berikutnya dalam rantai pemrosesan ekstraksi fitur dan memadukan semua garis frekuensi terdekat. Fungsi Hamming Window [4] didefinisikan sebagai:

$$w_n = \begin{cases} 0.54 - 0.46\cos(2\pi n/(N-1)) & 0 \le n < N \\ 0 & lainnya \end{cases}$$

Setelah itu, dikalikan dengan hasil persamaan jendela Hamming dengan sinyal masukan / input yang telah ditetapkan sebagai berikut:

Y(n) = y(n) * w(n)

di mana
N = Banyaknya sampel tiap frame
Y(n) = Sinyal Output
y(n) = Sinyal Input
w(n) = Jendela Hamming

Fast Fourier Transform (FFT).

Pada langkah ini setiap frame hasil dari fungsi window dikenai proses FFT. Fungsi FFT digunakan untuk mengubah sinyal yang semula merupakan time domain menjadi *frequency domain*. Langkah ini mengubah tiap frame N sampel dari domain waktu ke dalam domain frekuensi. Dalam pengolahan suara, Transformasi Fourier Cepat berguna untuk mengubah konvolusi getaran celah suara dan respon gelombang saluran suara dalam domain waktu.

Mel-scale dan *Filter Bank*

Pada tahap ini dilakukan *wrapping* terhadap spektrum yang dihasilkan dari FFT sehingga dihasilkan *Mel-scale* untuk menyesuaikan resolusi frekuensi terhadap properti pendengaran manusia.

Kemudian *Mel-scale* dikelompokkan menjadi sejumlah critical bank menggunakan *filter bank*. Jangkauan frekuensi dalam spektrum sangatlah luas dan sinyal suara tidak mengikuti skala linear. Sehingga setelah spektrum terkomputasi, data dipetakan dalam skala Mel menggunakan filter segitiga yang saling tumpang tindih.

2.2. *Hidden Markov Mode*l (HMM)

Hidden Markov Model (HMM) digunakan untuk menentukan parameter-parameter tersembunyi (hidden) dari parameter-parameter yang dapat diamati. Pada model Markov secara umum, keadaannya dapat secara langsung dapat diamati, oleh karena itu probabilitas transisi keadaan menjadi satu-satunya parameter. Di dalam HMM, keadaan-nya tidak dapat diamati secara langsung, akan tetapi yang dapat diamati adalah variabel-variabel yang terpengaruh oleh keadaan [9].

III. METODOLOGI PENELITIAN

3.1 Pengumpulan Data

a. Data uji coba dilakukan oleh penutur asli Indonesia sebanyak 3000 contoh suara. 150 contoh suara ini adalah sebanyak 5 orang penutur asli Indonesia yang tidak terlalu fasih berbahasa arab dan 5 orang penutur asli Indonesia yang lebih fasih berbahasa arab dari 5 penutur sebelumnya, setiap penutur mengucapkan 15 kata dalam bahasa arab dan dilakukan pengulangan sebanyak 20 kali pengucapan.
b. Data Uji menggunakan direkam menggunakan mikrophone pada laptop dengan frekuensi sampling 8000 Hz, 22050 Hz, 44100 Hz 16 bit PCM Mono dengan format suara .wav. Data uji didapat dengan cara merekam setiap kata dari setiap penutur sebanyak 20 ucapan dan direkam pada frekuensi sampling 44100 Hz. Data ucapan penutur ini disimpan dalam bentuk file sebanyak 20 file dengan format .wav untuk setiap penutur. Sehingga untuk frekuensi sampling 44100 Hz didapat data ucapan berbentuk file sebanyak 10 x 20 = 200 file.
c. Data uji dengan frekuensi sampling 22050 Hz didapat dari mengkonversi data file dengan frekuensi sampling 44100 Hz menjadi frekuensi sampling 22050 Hz dan disimpan dalam bentuk file. Data uji yang didapat dari hasil konversi ini juga sebanyak 200 file.
d. Data uji dengan frekuensi sampling 8000 Hz didapat dari mengkonversi data file dengan frekuensi sampling 22050 Hz menjadi frekuensi sampling 8000 Hz dan disimpan dalam bentuk file. Data uji yang didapat dari hasil konversi ini juga sebanyak 200 file.
e. Jumlah keseluruan data uji adalah sebanyak 200 x 3 data uji = 600 data uji untuk satu kata yang diucapkan oleh semua penutur.
f. Jadi data uji untuk 15 kata yang diucapkan oleh kesemua penutur adalam sebanyak 15 x 600 data uji = 9000 data uji.

3.2. Pengumpulan data yang digunakan untuk training.

Pengumpulan data yang digunakan untuk data training adalah suara kata dalam bahasa arab yang diucapkan oleh penutur asli penutur arab. File suara yang didapat didapat dari youtube pada pembelajaran bahasa arab oleh instruktur penutur arab. Data training sebanyak 90.

3.3. Pengujian dan evaluasi

Pengujian dilakukan menggunakan kata bahasa arab sebanyak 15 kata dalam bentuk file wav dan diucapkan oleh 10 penutur Indonesia. Setiap penutur melakukan pengulangan kata yang sama sebanyak 20 kali. Ucapan yang terbanyak yang dikenali oleh sistem dianggap sebagai hasil akhir. Perekaman ucapan penutur indonesia menggukan mikrophone yang terdapat dalam komputer Netbook Aspire One Intel Celeron 1,6 GHz.

Evaluasi yang dilakukan adalah melihat tingkat akurasi pengucapan kata bahasa arab oleh penutur indonesia pada klasifikasi Hidden Markov Model. Evaluasi juga dilakukan dengan membandingkan akurasi sistem pada 3 frekuensi sampling yang berbeda.

IV. HASIL DAN PEMBAHASAN

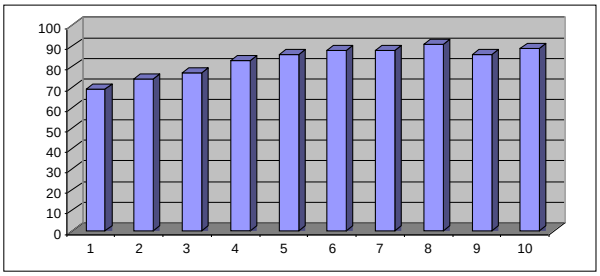
Pengujian ini untuk mengetahui kemampuan sistem untuk mengenali 2700 kata bahasa arab dalm bentuk file yang telah didapat dan direkam dari 10 penutur indonesia. Nilai parameter standar yang dimasukan dalam sistem yang diberlakukan untuk semua data uji adalah sebagai berikut:

pre-emphasis: 0.95, jumlah filterbank: 20, dan
jumlah koefisien cepstral MFCC: 20.

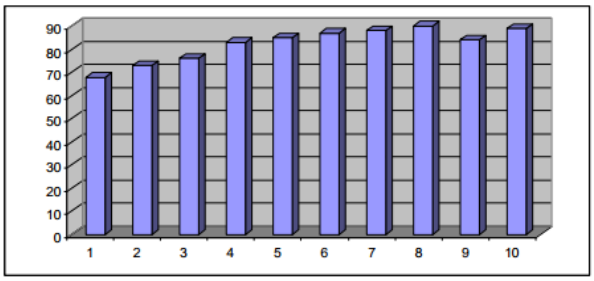
Pengujian pada sistem ini menggunakan bantuan perangkat keras berupa satu unit laptop Aspire One Intel Celeron 1,6 GHz. Laptop ini dilengkapi dengan Microphone Realtek High Definition untuk merekam data uji, serta perangkat lunak Matlab R10 dan perangkat lunak perekam data suara yaitu Audacity.

Pengujian yang dilakukan dititik beratkan 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 sampling, serta perbedaan dalam mengenali ucapan dalam tiga frekuensi sampling yang berbeda.

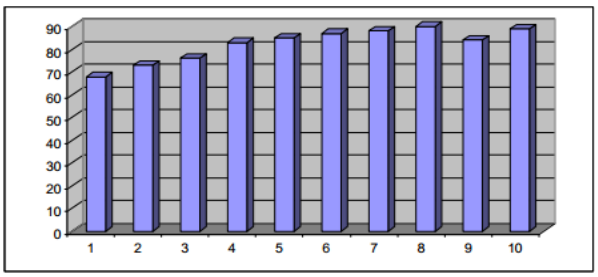
Kata yang diucapkan sebagai data uji dipilih secara acak yang dipergunakan dalam komunikasi sehari hari. Dalam pengambilan data, kata tersebut ditulis dalam huruf latin menggunakan pedoman transliterasi dari kementrian agama Republik Indonesia dan penutur mengucapkannya sesuai dengan kemampuan masing-masing. 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.



Gambar 2 Grafik % Akurasi pada frekuensi sampling 8000Hz.



Gambar 3 Grafik % Akurasi pada frekuensi sampling 22050 Hz



Gambar 4 Grafik % Akurasi pada frekuensi sampling 44100 Hz

Dengan menggunakan metode MFCC ini, suara penutur masih dapat dikenali dengan data latih menggunakan suara penutur arab. Penutur indonesia berusaha untuk menyesuaikan supaya pengucapannya sama dengan penutur arab. Hal ini dapat dilihat pada gambar 2, 3, dan 4 pada penutur 6 sampai dengan 10 yang mempunyai % Akurasi diatas nilai 80. Frekuensi sampling yang semakin besar juga mengakibatkan penurunan tingkat akurasi.

V. KESIMPULAN

Dari hasil pengujian disimpulkan sebagai berikut :

1. Sistem menghasilkan nilai akurasi rata-rata sebesar 83,1% untuk frekuensi sampling data uji sebesar 8000 Hz, 82,3% untuk frekuensi sampling data uji sebesar 22050 Hz, 82,2% untuk frekuensi sampling data uji sebesar 44100 Hz.
2. Penutur yang fasih mengucapkan bahasa arab mempunyai tingkat keakurasian yang lebih tinggi daripada orang yang bisa bahasa arab tetapi belum fasih. Hal ini disebabkan karena dalam mengucapkan kata arab, penutur indonesia berusaha untuk mengucapkan bahasa arab sesuai dengan asli nya.
3. Dialek Indonesia juga mempengaruhi tingkat keakurasian sistem dalam mengenali ucapan. Hal ini dibandingkan dengan penelitian sebelumnya dengan data uji berasal dari dialek arab.

VI. DAFTAR PUSTAKA

[1] Alcaraz Meseguer, Noelia, "*Speech Analysis for Automatic Speech Recognition*", Norwegian University of Science and Technology Department of Electronics and Telecommunications, 2009.

[2] Aria, Muhammad, "*Sistem Pengenalan Kata Bahasa Indonesia Berbasis LabView untuk Pengendalian Peralatan Ruang Perkuliahan*", Universitas Komputer Indonesia, 2013.

[3] Ashikin Binti Norzain, Norul, "*Security System Using Biometric Technology: Voice Recognition*", Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 2014.

[4] Elkourd, Amer M., "*Arabic Isolated Word Speaker Dependent Recognition System*", Islamic University, Gaza, Palestine Deanery of Higher Studies Faculty of Engineering Computer Engineering Department, 2014.

[5] Elminir, Hamdy K. , Mohamed Abu ElSoud, L. M. Abou El-Maged, " *Evaluation of Different Feature Extraction Techniques for Continuous Speech Recognition*", International Journal of Information and Communication Technology Research, 2012.

[6] Hassine, Mohamed, Lotfi Boussaid, Hassani Massouad, "Hybrid techniques for Arabic letter recognition", International Journal of Intelligent Information Systems 2015.

[7] Holmes, J. & Holmes, W. , "*Speech Synthesis and Recognition*", 2th ed., Tailor & Francis, London, 2001.

[8] Ibrahim, Noor Jamaliah Binti, "Automated Tajweed Checking Rules Engine For Quranic Verse Recitation", Faculty Of Computer Science And Information Technology University Of Malaya Kuala Lumpur, 2010

[9] Rabiner, L.R. & Juang, B.H., 1993, 'Fundamental of Speech Recognition', Prentice Hall, New Jersey, USA.

[10] Rozaq, Ahmad, dkk, "Identifikasi Ciri Musik Dengan Menggunakan Mel-Frekuensi Cepstral Coefficien (MFCC)", Politeknik Elektronika Negeri Surabaya, 2010.

[11] Shah, Nazaruddin bin MD, "Voice Activation Switch", Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer Universiti Teknikal Malaysia Melaka, 2010.

[12] Sijabat, Davit Wasty, "Simulasi Pengenalan Chord Terisolasi Berbasiskan Speaker Dependent Dengan Metode Hidden Markov Model", Fakultas Teknik Universitas Indonesia, 2009.

[13] S. Y. Mark Gales, "*The Application of Hidden Markov Models in Speech* Recognition" dalam Foundations and Trends in Signal Processing, 2009.

[14] Taufik Adnan Amal, "*Rekonstruksi Sejarah Al-Qur'an",* Cet. I; Penerbit Forum Kajian Budaya dan Agama, Yogyakarta. h. 151, 2001.